MORPHOMETRIC IDENTIFICATION, ECOLOGY AND OLFACTORY MEDIATED SUGAR FEEDING BEHAVIOUR OF *CULEX PIPIENS S.S.* AND *CULEX TORRENTIUM* MOSQUITOES OF THE NORTH WEST MIDLANDS OF ENGLAND

RICHARD KENNETH HALFPENNY

A thesis submitted in partial fulfilment of the requirement of Staffordshire University for the degree of Doctor of Philosophy

AUGUST 2018

Acknowledgements

I would like to thank my supervisors Dr Angela Priestman, Dr Kevin Reiling and Professor John Dover, for their support and guidance throughout my PhD studies. I would also like to thank Staffordshire University for funding my PhD.

I am hugely grateful to the technical staff at the Staffordshire University Science Centre for their help and assistance; due to the range of methods used, I think you all helped at one time or another, so thank you very much!

Out in the field, I would like to acknowledge the cooperation and support from Staffordshire Wildlife Trust and Rupert Randall from Natural England for seeing the value in my work and arranging permits for me to carry out research on their excellent nature reserves. Also, Arthur and Vera Green I thank you for your services in checking that I made it safely from wetter parts of Wybunbury Moss each day!

Of course, my fellow PhD students, Elle, Trust, Kayleigh, Elecia, Udeshika, Dilhani, Paul and Fran, you helped make the more stressful parts of study bearable, the lonely bits less so, and many parts downright enjoyable, so thank you.

I'd like to acknowledge my inspirational friends. Andrew Turner for always being there and taking part in many hours of mosquito related chat and Dr Katy Vigurs for helping me believe that I was capable of a PhD in the first place.

Finally, and most importantly, I need to thank my wife Claire, children Luke and Neve and our parents. Without your love, patience and support, this journey would have been impossible.

1 Table of Contents

Abstract

Mosquito distribution data in the UK are patchy and much are too old to be useful to describe the current distribution. Compounding this, species that are particularly challenging to identify result in them often being collated as complexes rather than species. Within the *Culex pipiens* complex, members of which have demonstrated roles as vectors internationally, morphological discrimination between female *Culex pipiens* L. and *Culex torrentium* Martini cannot reliably be made without using molecular identification methods. This often results in species discrimination not being made. We also know relatively little about the sugar feeding preferences of these mosquito species, particularly in terms of males. Male mosquitoes are generally not considered as targets for control. However, recent research has demonstrated swarming male targeting strategies that show potential. Therefore, knowledge of these behaviours might lead to novel approaches for control interventions or improve the effectiveness of those in development.

In this thesis, current mosquito distribution data were investigated and fieldwork conducted to add to new species presence data for North Staffordshire and environs. A recently published wing morphometric technique for the discrimination of *Culex pipiens s.s.* and *Culex torrentium* was tested for suitability for locally collected specimens, and subsequently optimised and semi-automated using a new coded workflow. Behavioural assays were conducted to assay behaviour of male *Culex pipiens s.s.* and *Culex torrentium to* olfactory cues associated with sugar feeding from flowers of various species of Apiaceae. Subsequently, a synthetic chemical lure, derived from the odours of these flowers, was tested in the laboratory and the field.

Fieldwork surveys collected eleven of the thirty-four British mosquito species within the survey, adding eight more species to the local record.

The wing morphometric approach for species discrimination, following optimisation, led to the development of a new computerised approach and workflow that demonstrated 84% accurate species identification. This technique represents a viable, cost effective,

identification method for those without access to molecular methods, subsequently leading to better data relating to their distributions.

In behavioural assays, both *Culex pipiens s.s.* and *Culex torrentium* were shown to be attracted to Apiaceae flowers. However, there was no significant difference in preference to different flower species within the Apiaceae, nor between the two mosquito species. Following field and laboratory testing, the synthetic lures were found to have no significant influence on behaviour or detectable catch rate increase in traps in the field.

2 List of Tables

List of figures

1 General Introduction

1.1 Mosquitoes and Disease

Through their ability to transmit viral and parasitic pathogens to humans, mosquitoes cause millions of deaths every year and are considered one of the deadliest animals in the world (WHO 2016). Malaria alone is still one of the largest causes of death, despite the recent downward trend in human mortality due to the upscaling of mosquito control activities (Bhatt *et al.* 2016). Whilst the progress that is being made in the fight against malaria is welcome, the fight against mosquito-borne diseases is a continually evolving struggle. Emerging and re-emerging diseases continue to threaten human health, and mosquito-borne diseases feature strongly within the threat (Taylor *et al.* 2001). Dengue incidence has risen 30 fold over the past 50 years (WHO 2016). Similarly, Chikungunya and Zika are spreading geographically, with countries reporting first outbreaks of disease among their populations (WHO 2016).

Although the diseases mentioned above are mainly restricted to the tropical and sub-tropical regions (Bhatt *et al.* 2013; Nsoesie *et al.* 2016; Samy *et al.* 2016), temperate climes are also seeing increasing disease burdens from mosquito-borne diseases and mosquito-borne viruses (MOBVs) in particular. One striking example of this would be the emergence and rapid expansion of West Nile virus (WNV) in North America from an initial outbreak in New York in 1999 to a persistent continent-wide cause of morbidity and mortality, affecting 46 US states by January 2017 (CDC 2017). As all mosquito-borne disease incidence is a result of the interaction between the mosquito vector, the pathogen and the blood meal host (Dennison *et al.* 2014), the distributions of these diseases are typically constrained by the ecological habitat range of their competent vectors. The level of interaction is further complicated, and so too the disease distribution, in the case of zoonoses, those diseases which normally exist in animals but can infect humans.

1.2 British mosquitoes

Thirty-four mosquito species (Figure 1) are currently listed as present in the British Isles and are considered as permanent autochthonous residents comprising of six species of Anophelinae, all are genus *Anopheles* and twenty-eight species of Culicinae comprising of seven genera: three *Aedes* species, one *Coquillettidia* species, four *Culex* species, seven *Culiseta* species, one *Dahliana* species, eleven *Ochlerotatus species*, one *Orthopodomyia* species (Harbach & Howard, 2007; Medlock *et al.*, 2005; Medlock & Vaux, 2009; Reinert *et al.*, 2006, 2009).

Throughout this thesis, mosquito species naming conventions are based on Chandler (2017); this is the latest relevant checklist for British diptera and reflects most of the updated species naming conventions (Reinert *et al.* 2009), although it has not adopted the *Ochlerotatus* (*Aedes*) genus naming suggested in Wilkerson *et al.* (2015).

Due to the number of species, and the ability of mosquitoes to exploit ecological habitats, mosquitoes are common across most of Britain (Snow 1986). Urbanisation and changes in land use have changed much of the British landscape, often reducing biodiversity and removing habitat niches. However, these changes offer excellent opportunities to generalist species, such as representatives of the *Culex pipiens* complex (Vinogradova 2003), which can tolerate or adapt to change or move quickly to new habitat creation (Becker *et al.* 2010a). Urbanisation has been particularly important for mosquitoes (Vinogradova 2003); with aspects such as the urban heat island (UHI) effect increasing the temperature in cities in general, and at night in particular (Bohnenstengel *et al.* 2011). This effectively increases the habitable ranges for species near their northern range limit, and the length of the breeding season for others. Strategies such as the creation of urban wetlands to reduce the effect of the UHI can create new habitats for mosquitoes, and as such require careful consideration (Medlock and Vaux 2015a). Other factors associated with residential development, such as the use of artificial water containers for the collection and storage of rainwater, which increase the availability of predator-free breeding sites (Townroe and Callaghan 2014), encourages mosquito population growth for those species which can take advantage of these

features (Derraik 2005). Similarly, evidence suggests that climate change is increasing the length of the breeding season and therefore increasing mosquito abundance, and reducing duration and severity of winters increasing the overwintering survival rates of important species such as *Culex pipiens* s.l.*, Coquillettidia richiardii* and *Ochlerotatus punctor* (Snow and Medlock 2006; Bale and Hayward 2010; Townroe and Callaghan 2014; Medlock and Leach 2015; Ewing *et al.* 2016). These changes can result in increased numbers of mosquitoes directly adjacent to the human population causing increased interactions between them and therefore increasing the opportunity for vector-borne disease transmission (Reiter 2001).

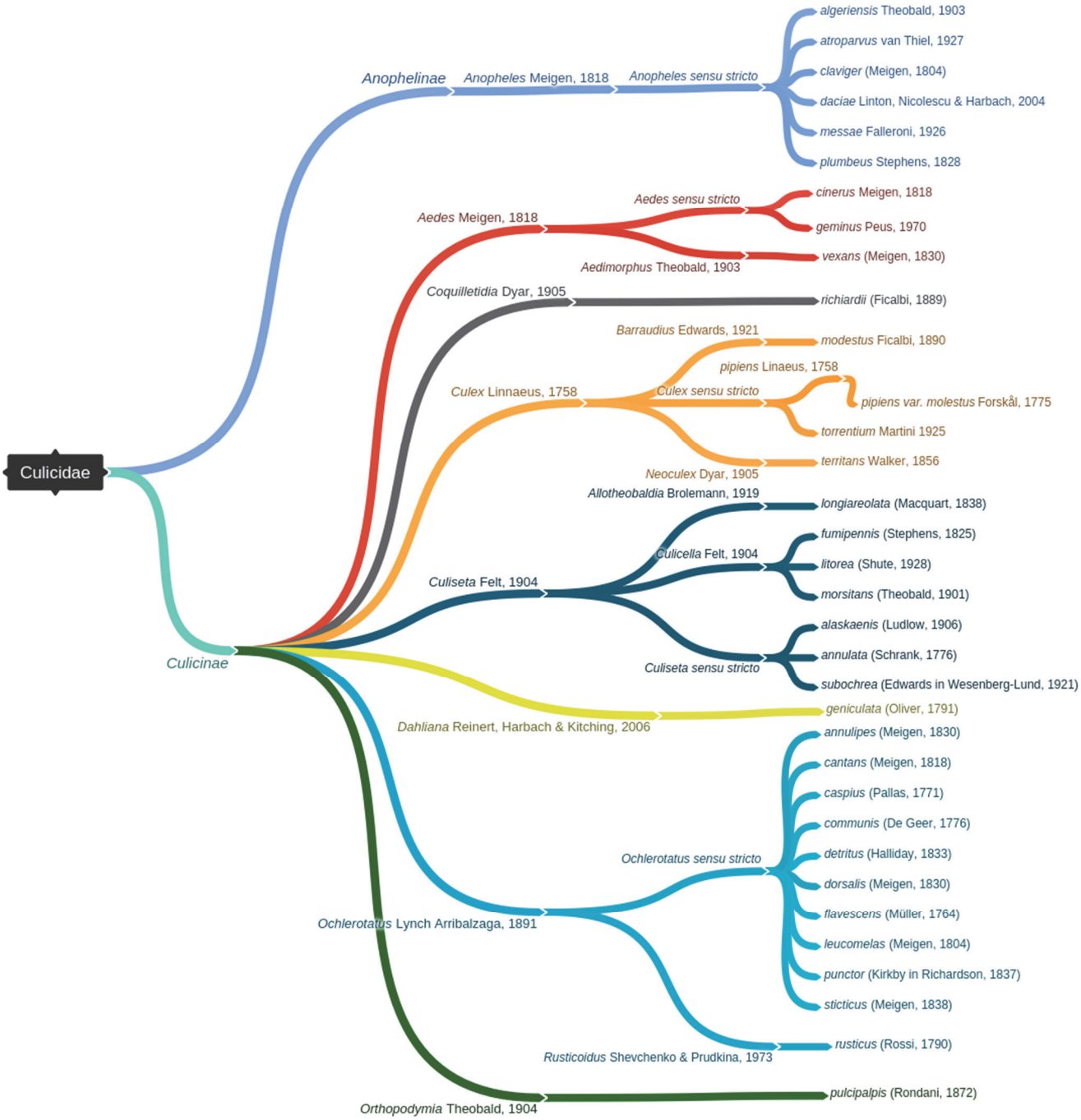


Figure 1 Taxonomic tree of the 34 British Culicidae species. This diagram is based on the taxonomic checklist from Chandler 2017 and forms the basis of the nomenclature used throughout this thesis

1.3 British Mosquitoes of potential medical importance

Included here is an overview of those mosquitoes which are permanent inhabitants of the British Isles and which may be expected to act as vectors in future disease outbreaks, and/or are demonstrated as acting as pests through nuisance biting. Currently, there is no evidence of autochthonous mosquito-borne transmission in humans in the UK (Medlock and Vaux 2016). Of the 34 species recorded in the British Isles most will bite humans and at least a dozen are associated with doing so at such a level as to cause a nuisance (Medlock and Hansford 2012).

British mosquitoes from the *Culex pipiens* complex: *Culex pipiens s.s.*, *Culex pipiens* var. *molestus* and *Culex torrentium* are potentially important as vectors of various diseases within wildlife reservoirs and as bridge vectors to humans (Fonseca *et al.* 2004; Hamer *et al.* 2008), this complex of species and subspecies is generally referred to as *Culex pipiens* s.l. in this thesis. Certain *Culex* species, particularly *Culex modestus* and *Culex pipiens* var. *molestus*, could act as direct vectors between humans if the disease was prevalent in humans; their direct threat to humans is modified by their host-seeking behaviour.

It is important to consider the research from Europe as it may be possible to make predictions of habitats for British mosquitoes from high quality studies from nearby countries such as Belgium (Versteirt *et al.* 2013) where species' associations with habitat types could be distinguished. Outside of the UK, *Culex pipiens* s.l. have been identified as confirmed or most likely vectors of West Nile Virus (WNV) in a number of outbreaks affecting humans and livestock. In 1996, an outbreak of WNV in south-eastern Romania caused 17 human deaths; in this outbreak 393 patients were identified (Tsai *et al.* 1998). An outbreak of equine encephalomyelitis, amongst racehorses caused by WNV occurred in the Tuscany region of Italy in 1998 (Romi *et al.* 2004). Whilst no WNV was recovered from mosquitoes following that outbreak, it was considered most likely that *Culex impudicus* Ficalbi (not found in the UK) acted as the enzootic vector, and *Culex pipiens* s.l. as the bridge from avian to equine hosts (Romi *et al.* 2004). More recently, in 2010 *Culex pipiens* s.l. was implicated in an outbreak WNV lineage 2 in Greece (Papa *et al.* 2011), in the article the author reported endophilic *Culex*

pipiens, collected by means of human bait method, as being the dominant mosquitoes in the rural region of Greece that was affected, this makes it very likely that the anthropophilic var. *molestus* sub-species of *Culex pipiens* was involved in the outbreak. *Culex pipiens var. molestus* is an aggressive biter of people (Vaux *et al.* 2015) as is the recently re-established *Culex modestus*; although currently restricted to the south of England, this mosquito species is aggressively anthropophagic and a principle vector of West Nile Virus in southern Europe (Golding *et al.* 2012). *Culex modestus* was identified as a principal vector of WNV in the Camargue in France (Balenghien *et al.* 2008). Magurano *et al.* (2012) reported an outbreak of WNV on the island of Sardinia, Italy, in 2011 that led to four human deaths; again, *Culex pipiens* s.l. mosquitoes were considered to be the most likely vector species.

Culex pipiens s.l. have been shown to bite humans, rabbits and birds (Service 1971a) and also frogs, lizards and snakes (Snow 1990), but are typically considered as being preferentially ornithophilic. In Italy, differential host feeding preference between bird species has been demonstrated, with different preferences in urban and peridomestic populations of *Culex pipiens* s.l. (Rizzoli *et al.* 2015). Interestingly, Rizzoli *et al.* (2015) also reported seasonal variation in avian host preference that was not explained by the change of bird species abundance caused by bird migration. Those species such as *Culex pipiens s.s.* and *Culex torrentium,* which are considered primarily ornithophilic, can have significant roles in the amplification, maintenance and dispersal of disease through avian hosts (Farajollahi *et al.* 2011; Leggewie *et al.* 2016) as they are competent vectors of several viruses, filarial worms and avian malaria (Farajollahi *et al.* 2011). *Culex pipiens s.s.* has been identified as a competent vector for WNV and a candidate bridge between avian and mammalian hosts (Turell *et al.* 2005; Kilpatrick *et al.* 2006).

Within the *Culex* genus, species diagnosis can be challenging. For example, discrimination between *Culex pipiens s.s.* and *Culex pipiens var. molestus* is not possible morphologically (Becker *et al.* 2010a) and for dead specimens requires identification by molecular methods such as those developed by Bahnck & Fonseca (2006). However, where specimens are reared on to adulthood from eggs or larvae, confidence in species identification, by observation of behaviour, increases due to *Culex pipiens* s.l. being eurygamous (requiring an ample space to mate) and anautogenic (requiring a blood meal for egg creation) and *Culex pipiens var.*

molestus being stenogamous (needing little space to mate) and autogenous (not requiring a blood meal before laying eggs); autogeneity is considered sufficient to discriminate between these subspecies (Farajollahi *et al.* 2011). Species discrimination between *Culex pipiens s.s.* and *Culex torrentium* is similarly difficult; males can be identified by morphological differences in their terminalia, but there are no such differences between the females (Becker *et al.* 2010a). Elsewhere in Europe *Culex torrentium* has been shown to be more common than *Culex pipiens s.s.* in Sweden (Hesson *et al.* 2014), and evidence suggests that it is the principal enzootic vector of Sindbis virus in Sweden (Hubálek 2008). Taken together, these factors provide support for the importance of carrying out more research into *Culex torrentium* distribution and behaviour as a species and not simply treating *Culex pipiens* s.l. as a homogenous species complex.

Ochlerotatus detritus is a persistent biter of humans and is locally common in parts of the UK (Snow 1990), being the most prevalent species in some coastal and estuarine habitat areas (Clarkson and Setzkorn 2011) and is associated with significant nuisance biting in these regions (Medlock and Hansford 2012). In addition to humans, it also feeds on bovids, pigs, other mammals and birds (Service 1971a) and may, therefore, act as a bridge vector between species. *Oc. detritus* has been shown, in the laboratory, to be a competent vector for West Nile Virus, but showed no evidence of competence for dengue or chikungunya viruses (Blagrove *et al.* 2016)

Anopheles mosquitoes are notoriously associated with malaria, and indeed the *Plasmodium vivax* strain of malaria was once autochthonous in Britain (Ramsdale and Snow 1995; Dobson 1998) up until the start of the 20th century, when changes in housing and farming practices including the separation of human and animal housing caused its incidental eradication (Ramsdale and Gunn 2005). Members of the *Anopheles maculipennis* Meigen complex are the dominant European vector of vivax malaria, three species of this groups have been reported in Britain: *Anopheles atroparvus* van Thiel, *Anopheles messae* Falleroni and *Anopheles daciae* Linton, Nicolesu and Harbach (Danabalan *et al.* 2014). However, vector competency for malaria is also apparent in other species, such as *Anopheles plumbeus* Stephens (Bueno-Marí and Jiménez-Peydró 2011).

1.4 What are Mosquito-borne organisms and competent vectors

Mosquito-borne organisms (MOBOs) are those organisms and viruses which have mosquitoes acting as the vector between hosts. MOBOs can be zoonotic, normally existing in and transferred between animals, and to humans from animals, by mosquitoes. They can also be transferred between humans by mosquitoes without an animal intermediary, other than the vector mosquito.

Competent vectors are, in general terms, organisms which can acquire, maintain and transmit microbial agents to other species. Not all competent vectors are blood feeders, freshwater snails and schistosomiasis for example, and not all blood feeders are competent vectors of all diseases (Kenney *et al.* 2017). Within this thesis, competent vector is used to refer to mosquitoes specifically. For mosquitoes to complete this role they must be able to ingest the organism/virus in the course of their blood feeding from an infected blood meal host, they then need to be able to pass the this on to further hosts during subsequent blood feeding (Chamberlain and Sudia 1961). For onward transmission, there needs to be a sufficient load of the pathogen transmitted to the host to cause infection in this new organism (Ahmed *et al.* 2007). As with all modes of disease transmission, an infection threshold must be overcome before a successful infection can take place, this requires sufficient inoculum of pathogen to allow the pathogen to persist in the vector (Hubálek 2008).

The vector competency of a mosquito species is not a constant, and may vary spatially and temporally (Vaidyanathan and Scott 2007; Leggewie *et al.* 2016), and in response to environmental conditions such as temperature (Leggewie *et al.* 2016; Vogels *et al.* 2017). The effect of temperature has been demonstrated experimentally to be sufficient to prevent transmission of WNV by southern European *Culex pipiens s.s.* after 14 days incubation at 18 °C, whereas transmission was observed at 23 °C and 28 °C by Vogels *et al.* (2017). However, Leggewie *et al.* (2016) found that northern European *Culex pipiens s.s.* and *Culex torrentium* had high rates of tissue dissemination of WNV by 28 days post infection at 18 °C, and found no significant difference between dissemination rates between these species at 18 °C and 25

°C. This suggested that there might be variation between populations, that the failure of transmission at 18 °C is related to factors other than viral dissemination in the mosquito or that a longer incubation period is required for transmission to occur. It should be noted that successful dissemination of virus does not necessarily mean that transmission will occur (Vogels *et al.* 2017).

Being a competent vector is only one of the required factors in vector incrimination. In addition to establishing a species as a competent vector, it is necessary to demonstrate synchrony between the mosquito species and the cases of diseases, whether human or animal, and the presence of the pathogen in the mosquito (Beier 2002). There should also be evidence of direct contact between the mosquito species and host, this is often identified by human landing/biting catches or by analysis of blood meals. In post outbreak investigations, it is important to find evidence of the pathogen in the local mosquito population. When all three states are met then incrimination can be made with high confidence (Beier 2002).

1.5 Are there mosquito-borne organisms in the UK?

Although there are at least 10 mosquito-borne viruses in Europe, belonging to 3 families – *Togaviridae* (Sindbis, Chikungunya), *Flaviviridae* (West Nile, Usutu, Dengue), *Bunyaviridae* (Batai, Ťahyňa, Snowshoe hare, Inkoo, Lednice) (Hubálek 2008), there is no published research suggesting that any MOBOs are currently endemic in the UK (Medlock *et al.* 2007).

However, (Buckley *et al.* 2003, 2006) provided evidence that at least three MOBOs West Nile virus (WNV), Usutu virus (USUV) and Sindbis virus (SINV), have been introduced to the UK resident bird population by vector activity of mosquitoes receiving viral infection from migratory birds. They also found that birds often had neutralising antibodies for combinations of the viruses discussed; suggesting that the viruses share overlapping habitats. Buckley *et al.* (2003) suggested two explanations regarding the lack of bird mortality due to WNV infections: first that the local bird populations have been exposed to these viruses for many years and have developed herd immunity, or second that the viral strains found in the UK are avirulent,

and that the strain in North America, with its resultant high avian mortality, is more virulent than that found in the UK (Buckley *et al.* 2003).

West Nile virus is a flavivirus and a member of the Japanese encephalitis group (Hubálek 2008), and originally isolated in 1937 in the West Nile district of Uganda (Smithburn *et al.* 1940). It is one of the most widespread of the Flaviviridae and is found through Africa, Asia, Europe, Australia and America (Hubálek 2008). The introduction of WNV to America in 1999 is, perhaps, one of the most studied and discussed MOBO introductions, with almost 13000 publications between 1999 and 2009 listed for the search "WNV America" on Google Scholar. The WNV outbreak in the USA would serve as a reminder, if one were needed, of the impact of a novel pathogen into a naïve system, and the necessity of horizon scanning research and surveillance.

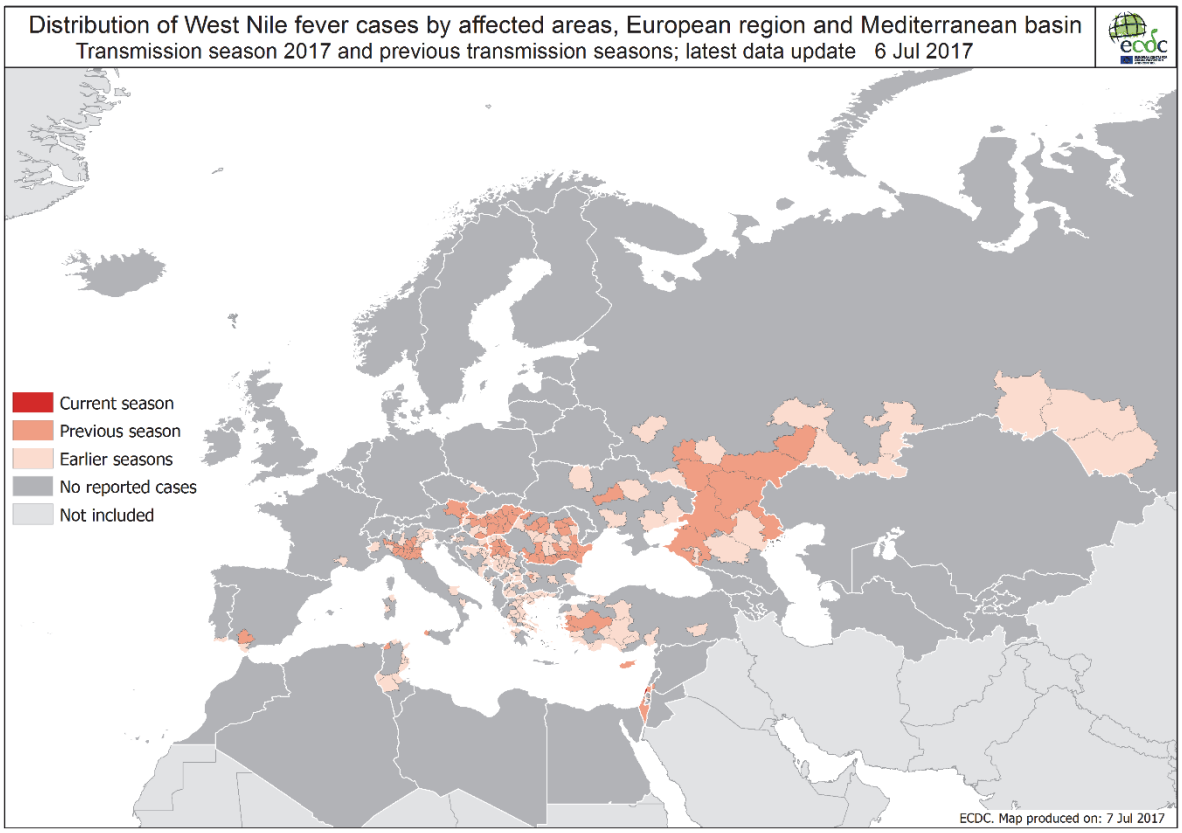


Figure 2 Distribution map of West Nile fever cases in Europe. As of 06/07/2017, no clinical cases had been reported for this season. However, we have yet to reach the peak of the mosquito season and previous West Nile fever reporting period which is typically in August. ECDC (2017)

Sindbis virus is an alphavirus and member of the American Western equine encephalomyelitis complex (Hubálek 2008). Sindbis virus is seen in northern Europe and Scandinavia as well as Africa, Asia and Australasia (PHE 2014). *Culex pipiens* and *Culex torrentium* are enzootic vectors for Sindbis virus infection in Sweden amongst and between migratory and non-migratory birds (Hesson, Verner-Carlsson, *et al.* 2015). *Culex torrentium* also has a greater prevalence than *Culex pipiens* in more northern regions (Hesson *et al.* 2014). In Sweden *Culex torrentium* are found to be more competent vectors of SINV than *Culex pipiens* (Hesson, Verner-Carlsson, *et al.* 2015). Buckley *et al.* (2003) found serological evidence of SINV in non-migratory wild birds and sentinel chickens (Buckley *et al.* 2006) in Britain.

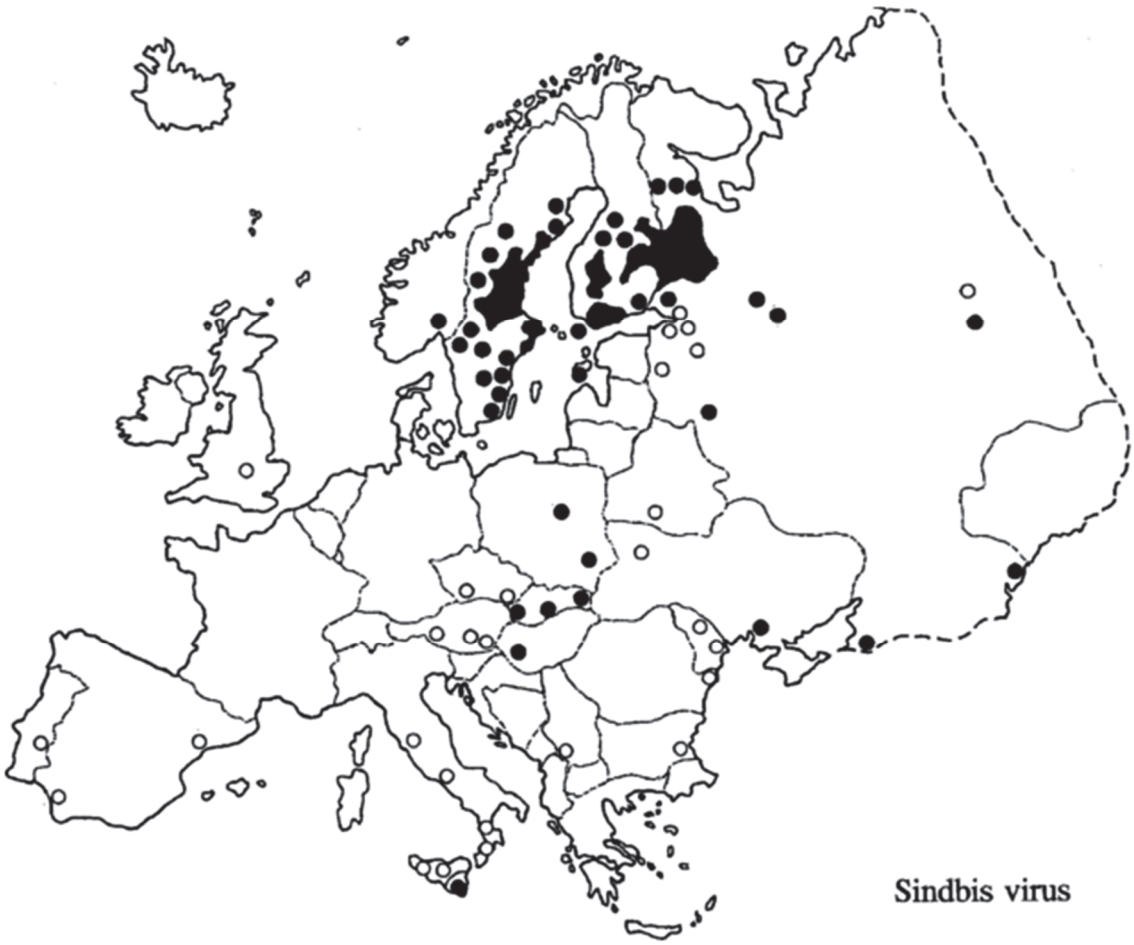


Figure 3 Geographic distribution of Sindbis virus in Europe. Black points indicate the locations of virus isolation; white circles indicate the locations where specific antibodies have been detected. Figure reproduced from Hubálek (2008).

Usutu virus is a flavivirus and member of the Japanese encephalitis group (Hubálek 2008), and was first isolated in South Africa in 1959 from *Culex univittatus* (Hubálek 2008). USUV was also found in British birds in the form of virus-specific neutralising antibodies in Buckley *et al.* (2003), it was subsequently found in sentinel chickens (Buckley *et al.* 2006); although at the time there had been no evidence of human morbidity associated with this virus in Europe. Rather it was related to significant bird mortality (Hubálek 2008; Ashraf *et al.* 2015). More recently, however, USUV has been demonstrated to cause human morbidity and to be neuro-invasive (Ashraf *et al.* 2015; Grottola *et al.* 2017) following the first record of human infection in Europe in 2009 (Pecorari *et al.* 2009).

1.6 Future threats: climate change and the impact of land use change

As discussed, the British Isles are not currently under pressure from autochthonous MOBOs, but in the face of increased human mobility and transportation of goods, climate change and land use change, it is, on the balance of probabilities, likely that MOBOs might be introduced in future (Vardoulakis and Heaviside 2012). In response to the changing and rising threat of mosquito-borne disease, the European Centre for Disease Control (ECDC) produced the *Guideline for the surveillance of native mosquitoes in Europe* (ECDC 2014).This document clearly describes the importance of surveillance and the generation of increased understanding of native mosquitoes in mitigation of the increasing mosquito-borne disease threat. Indeed, the document's executive summary states that "Vector-borne diseases, including those transmitted by native mosquitoes, should, therefore, be a high priority for public and veterinary health authorities across all of Europe." (ECDC 2014).

A significant body of work is being carried out to see which MOBOs could establish in the UK as a result of climate (Medlock *et al.* 2005, 2006; Lindsay *et al.* 2010; Medlock and Leach 2015) and land use changes (Lindsay *et al.* 2010; Medlock and Vaux 2011, 2015a) and findings suggest that risk will increase over time and that efforts to monitor the situation need to be continued. The changing nature of distributions is exemplified by the re-introduction, and apparent establishment, of *Culex modestus* (Golding *et al.* 2012), and the recent discovery of

visiting *Aedes albopictus* (Medlock *et al.* 2017), which is second only to *Aedes aegypti* in terms of vector competence for Dengue virus. It has also been demonstrated in the laboratory to be a competent vector for seven alphaviruses, eight bunyaviruses and three flaviviruses (Benedict *et al.* 2007). Post-invasion, the potential for establishment of *Aedes albopictus* estimates for the potential range within which it could overwinter under current temperature conditions have been predicted (Figure 4). These are based on two critical climate parameters requirements for overwintering of the species: the January temperature isotherm (TJan) which is related to the eggs surviving winter and annual rainfall (AR) determining the likelihood of inundation of oviposition sites. If the overwintering conditions for *Aedes aegypti* are indeed TJan > 0 °C and AR > 500 mm then most of the UK offers suitable conditions for overwintering which is considered the standard threshold (Medlock *et al.* 2006; Vardoulakis and Heaviside 2012). The ability to overwinter is critical for any invasive species establishment as the absence of this ability would be a barrier to dispersal (Kolar and Lodge 2001). Provided that egg survival rates are high enough to support year to year persistence then *Aedes albopictus* could become established and potentially increase its cold adaptation still further as has occurred in North American *Aedes albopictus* population (Hawley *et al.* 1989).

A persistent population of *Aedes albopictus* would present a significant increase in the risk of mosquito borne disease in the UK. This potential is exemplified by an outbreak of chikungunya virus in Italy in 2007 that caused symptomatic infection in 205 humans, and one fatality (Rezza *et al.* 2007). This outbreak was caused by the importation of the disease by a single infectious host visiting the area. Initial studies postulated that *Aedes albopictus* were the most likely vector in this outbreak (Rezza *et al.* 2007), this was subsequently confirmed during post outbreak investigations in which all other locally collected species were found to be free from the virus, but locally collected *Aedes albopictus* tested positive for chikungunya (Bonilauri *et al.* 2008). Italy was first colonised by *Aedes albopictus* in 1990 following the importation of scrap tyres containing eggs from the USA (Romi 2001).

As with their guidance for native mosquito surveillance, the ECDC has issued a technical report for the surveillance of invasive mosquito species and reinforcing the importance of this work (ECDC 2012).

Figure 4 Establishment zones for *Aedes albopictus*, based on overwintering survival requirements of (left) TJan 0 °C AR 500 mm, (middle) TJan 1 °C AR 600 mm, (right) TJan 2 °C AR 700 mm. Areas in grey represent areas for establishment Figure from Medlock *et al.* 2006.

1.7 Thesis rationale

Despite the future potential for mosquito borne disease in the British Isles, our understanding of native British mosquitoes is incomplete. The historical mosquito species record, and distribution data, appear limited and patchy. These data are important as these would form the basis for mosquito control interventions and planning in response to future disease outbreaks, they inform climate and land use modelling and good quality mosquito distribution data are essential for the basis of future ecological studies. The presence of cryptic, sibling, species in the region, such as *Culex pipiens s.s.* and *Culex torrentium*, causes a significant challenge to the generation of these data as it is often not possible to confidently identify which species specimens belong to using morphological methods. This has, historically, resulted in the use of species complexes in distribution data, which is unsatisfactory in the case of *Culex pipiens s.s*. and *Culex torrentium* as they have demonstrated differences in host preference and vector competencies (Fonseca *et al.* 2004; Hesson, Verner-Carlsson, *et al.* 2015).

Data related to the ecology, distribution and behaviour of target mosquito species form the basis of control measures (Becker *et al.* 2010a; Cameron and Lorenz 2013) and are typically targeted towards females as these are the disease vectors and so ultimately control must result in their reduction. Male mosquitoes have been largely ignored in mosquito control interventions as the link between the number of males is not likely to cause a linear reduction in the threat of disease transmission. However, if male populations can be effectively targeted and significant reduction of the male population achieved it has been demonstrated that a population threshold can be reached that causes significant mosquito population through the reduced availability of breeding age males (Sawadogo *et al.* 2017). This was achieved by the application of insecticides directly to swarms of males, but there should also be potential to achieve this by the deployment of lethal traps adjacent to swarming locations (Diabate and Tripet 2015). To achieve maximum effectiveness therefore, highly attractive lures should be employed to effect maximum collection rates by these traps. However, currently there is very little male specific data that can be applied in these areas of sugar feeding attraction, necessitating more study to be conducted in this aspect of behaviour to determine whether such highly attractive baits can be created.

1.8 Thesis aims

There are three aims of this thesis:

1) To investigate the current mosquito distribution data for North Staffordshire and its environs and to add new species occurrence data.
2) To investigate and optimise a non-molecular, wing morphometric, method for the diagnosis of species between female *Culex pipiens* s.s. and *Culex torrentium* mosquitoes.
    a. Can a wing morphometric analysis be used to discriminate between these female mosquitoes from the local mosquito population?
3) To examine the olfactory-mediated sugar feeding behaviour of wild male *Culex pipiens* s.s. and *Culex torrentium* to determine:
    a. Are these mosquito species attracted to odour emissions from inflorescences from selected Apiaceae species?
    b. Do preferences differ between mosquito and flower species?

c. Do the proposed synthetic blends based on floral emissions demonstrate similar preference behaviour to the natural emissions?
d. Do the proposed synthetic blends cause an increase in trapping efficiency in field testing?

1.9 Thesis structure

Chapter 2 describes the molecular methods of species identification used throughout the studies undertaken in subsequent chapters.

Chapter 3 reviews current British mosquito distribution data. It then describes the results of field survey work aimed at improving the local mosquito species record. Particular attention was applied to the identification of *Culex pipiens* s.s. and *Culex torrentium*, in light of their different potential roles in disease outbreaks in Britain, and the chronic underreporting of *Culex torrentium* numbers due to them being homogenised within the *Culex pipiens* complex.

In chapter 4.1, the method of discrimination between *Culex pipiens* s.s. and *Culex torrentium* using wing morphometrics proposed by Börstler *et al.* (2014) is validated using wildtype British specimens from the north Staffordshire region. Chapter 4.2 details the process of developing, optimising and testing a new wing morphometric method and presents a new software tool for the species diagnosis of *Culex pipiens* s.s. and *Culex torrentium*.

Chapter 5.1 relates to the analysis and quantification of VOCs emitted by flowers from British representatives of the Apiaceae family. Qualitative and quantitative analyses are conducted to compare the emission profiles between the species, and to identify the presence of compounds which have been shown to influence mosquito behaviour.

In chapter 5.2 the design, build and testing process for the dual choice olfactometer to be used for behavioural assays is presented. The chapter concludes with testing of the physical units for positional bias and confirmation of the suitability for use.

Chapter 5.3 examines the olfactory-mediated behaviour of male *Culex pipiens* s.s. and *Culex torrentium* in relation to selected Apiaceae species' flowers, as potential sugar sources. Preference was quantified by a novel and conservative preference index, the Expressed Preference Index (EPI) designed to avoid the exaggeration of preference which is possible in the high duration assays conducted here when using traditional preference indices.

In chapter 5.4, two synthetic chemical blends, are assayed against male *Culex pipiens* mosquitoes using the dual choice olfactometer and tested in the field as additional chemical lures to augment light trapping.

Chapter 6 contains the general discussion and conclusions for the thesis.

2 Molecular identification protocols

2.1 PCR Method for the Identification of *Culex pipiens s.s.* and *Culex torrentium*

The identification of many mosquito species can be conducted using traditional morphometric methods, using dichotomous keys such as those created by Cranston *et al.* (1987), Snow (1990) and Becker *et al.* (2010). However, many research situations obviate the use of morphological methods, either due to hybridisation between species (Fonseca *et al.* 2009), due to the presence of cryptic species, as is the case for female *Culex pipiens s.s.* and *Culex torrentium* (Dahl 1988; Becker *et al.* 2012) or cryptic sub-species such as *Culex pipiens s.s.* and the autogenous *Culex pipiens* var. *molestus* (Shute 1951; Dahl 1988) or the molecular forms of *Anopheles gambiae* (Diabaté *et al.* 2008). Challenging species discrimination scenarios such as these have often, historically, resulted in the formation of species complexes (Becker *et al.* 2012; Harbach 2012; Obenauer *et al.* 2013). Relatively recently, compared to the long history of mosquito research, new molecular methods have increased the diagnostic power of species identification, opening novel avenues for inter-species research into the behaviour and ecology of mosquitoes. These new approaches also inform taxonomy, which is still resulting in rearranging of taxonomic groupings and the identification of new species (Harbach 2007; Reinert *et al.* 2009; Wilkerson *et al.* 2015).

The focus species for the research within this thesis are *Culex pipiens s.s.* and *Culex torrentium*. Therefore, accurate identification of these species was critical. Morphological identification using dichotomous keys (Snow 1990; Becker *et al.* 2010a) could be confidently achieved to genus for all adult specimens; at this point, male species identity could be determined by examination of the terminalia. However, for females no further morphological discrimination beyond genus was possible. The potential for the use of the presence or absence of pre-alar scales for female species discrimination (Cranston *et al.* 1987) was considered as a method but discounted due to the unreliability of this approach (Onyeka 1982; Danabalan *et al.* 2012). Identification based upon 4th instar chaetotaxy, specifically the number of thoracic dorsal setae on each side and the characteristics of the inset setae on the anal segment (Becker *et al.* 2010a), appeared to have some merit. However, in my experience,

the observation of the presence of occasional bilateral asymmetry and intermediate forms ruled this method out due to reduced confidence in the accuracy of the method.

When considering which molecular approach to use for identification, three methods appeared to be the best candidates within the literature. All were based on DNA extraction and polymerase chain reaction (PCR) and had been shown to work well with for the identification of *Culex pipiens s.s.* and *Culex torrentium* (Smith and Fonseca 2004; Shaikevich 2007; Rudolf *et al.* 2013). Smith & Fonseca (2004), developed a series of PCR-based assays which exploit polymorphisms in the acetylcholinesterase-2 (ace-2) gene locus; they designed and tested multiple primer pairs to allow the identification of *Culex pipiens s.s.*, *Culex quinquefasciatus, Culex pipiens pallens, Culex australicus, Culex torrentium and Culex pervigilans*. Shaikevich (2007) proposed a method based on PCR-restriction fragment length polymorphism (PCR-RFLP). PCR was used to amplify a 603 bp fragment of the 5' end of the mitochondrial cytochrome oxidase C subunit 1 (COI) gene; differentiation assays between *Culex pipiens* and *Culex torrentium* by *Bc*1I restrictase, and *Culex pipiens* and *Culex pipiens* var. *molestus* by HaeIII restrictase. The final candidate method was developed by Rudolf *et al.* (2013) and was designed to be a high throughput real time-PCR (rtPCR) method to allow the identification of large numbers (up to 25) of specimens in batches. This was designed to work with a multiplex of species-specific primers and probes to facilitate the amplification and quantification of DNA respectively. Primers were designed based on microsatellite CQ11 gene loci for *Culex pipiens* and *Culex pipiens* var. *molestus* and on the ace2 gene for *Culex torrentium.*

Other molecular methods of species discrimination have been shown to work successfully with British mosquitoes. For example Hernández-Triana *et al.* (2017) used DNA barcoding methodology (Hebert, Cywinska, *et al.* 2003; Hebert, Ratnasingham, *et al.* 2003) to discriminate between broad species groups. This approach has the advantage of requiring the PCR amplification of the same 658 bp fragment of the 5' end of the COI gene for all mosquito species, which simplifies the method significantly. Subsequent matching of sequence data to the GenBank database allows species identification. However, the Hernández-Triana *et al.* (2017) approach, whilst powerful, does not discriminate between *Culex pipiens s.s.* and *Culex torrentium* and so was not suitable for use in this study.

Here we validate Smith's & Fonseca's (2004) 'Eurasian multiplex' primer mix which only contains primers for *Culex pipiens* and *Culex torrentium*, alongside the primers, minus the probes, for the same two species, devised by Rudolf *et al.* (2013). These were used in end-point PCR assays for species confirmation of *Culex* mosquitoes collected in North Staffordshire and environs, for use in ecological surveying and subsequent empirical study within this thesis. Shaikevich's (2007) method was not included in the testing and validation here as it was felt that the additional steps required to conduct the RFLP process was an unnecessary complication which would not be required should either of the other two methods be judged as suitable for the purpose of this research.

2.1.1 DNA extraction

All PCR and DNA extraction reagents were stored at -20 °C, except for 5 % Chelex which was stored at 4 °C. All pipette tips and tubes were autoclaved prior to use. All utensils and the microscope stage used in the specimen preparation were cleaned with 100 % ethanol prior to use and between each sample.

DNA was extracted using a Chelex® 100 and Proteinase-K master mix method adapted from (Walsh *et al.* 1991), and single legs from mosquito specimens. Due to the high viscosity of the Chelex® cut-off pipette tips are required when pipetting, and were used whenever necessary. A single leg was removed using fine forceps, sterilised with 100 % ethanol, from each locally collected mosquito specimen. For the purpose of this validation, each specimen had been previously identified, morphologically, as members of the *Culex* genus. Each leg was transferred to a labelled 0.2 ml clear tube (Bio-Rad TBS0201) with flat cap (Bio-Rad TCS0803). The single leg samples were used whole, without physical homogenisation. 5% Chelex® 100 stored in a bottle with magnetic flea was placed on a magnetic stirrer at room temperature. 30 µl per specimen of 5% Chelex® 100 was added to a 1.5 ml Eppendorf tube. Proteinase K (Bioline BIO-37037) was dissolved into PCR water at 10 mg/ml. 20 µl per specimen of 10 mg/ml Proteinase-K was added to the 1.5 ml Eppendorf previously loaded with Chelex® to complete the 50 µl per reaction extraction mix.

Before each DNA extraction, the extraction mix was vortexed to ensure the Chelex® remained suspended. 50 µl of master mix was added to each single leg specimen tube. Care was taken to ensure that the leg was submerged in the master mix and that there were no bubbles in the reaction volume. Samples were then placed in a thermocycler (Techne® 5Prime/02, Cole-Parmer) with the heated lid turned on at 105 °C, and run at 56 °C for 10 minutes then 96 °C for 10 minutes. The DNA extracts were then either used immediately or stored at -20 °C.

DNA extraction was tested using a Quant-iT^{TM} High-Sensitivity dsDNA Assay kit (ThermoFisher Scientific), and a Qubit fluorimeter. Two single leg extractions, from different specimens, and one double leg extraction from a single specimen were quantified, and both treatments showed sufficient DNA present for amplification by PCR. Single leg treatments showed 77 ng ml^{-1} and 173 ng ml^{-1}. Double leg treatment showed 342 ng ml^{-1}.

2.1.2 DNA amplification using PCR

DNA amplification was initially trialled using 2 different multiplex primer sets devised by Smith & Fonseca (2004) and Rudolf *et al.* (2013). The primers from Smith & Fonseca (2004) use two, species specific, forward primers, ACEpip and ACEtorr, and a single reverse primer, B1246s, which is common to both species. These primers target the acetylcholinesterase gene (ACE2). The Rudolf *et al.* (2013) primers have species-specific primer pairs which target loci on different genes, PIP F and PIP R target the CQ11 locus of *Culex pipiens s.s.,* and TORR F and TORR R, which target the ACE2 gene in *Culex torrentium*. Details of the primer sequences, melting points and expected fragment sizes are collated in Table 1, those from Smith & Fonseca (2004) are quoted from the original article, for Rudolf *et al.* (2013) these were calculated (Appendix 1) using Primer-BLAST (Ye *et al.* 2012). All primers were prepared by Eurofins Scientific. All primers were received as 100 picomole solutions and diluted 1:10 with PCR water.

Table 1 Primers used for discriminatory identification of *Culex pipiens s.s.* and *Culex torrentium* mosquitoes. *Fragment size from Smith & Fonseca (2004). ** Fragment length range for the target species, from NCBI PrimerBlast (Appendix 1).

Primer Name	Primer sequence	Melting point °C	Fragment size (bases)*
ACEpip	5' GGAAACAACGACGTATGTACT 3'	56.05	609*
ACEtorr	5' TGCCTGTGCTACCAGTGATGTT 3'	60.3	416*
B12346s	5' TGGAGCCTCCTCTTCACGG 3'	60.98	
PIP F	5' GCGGCCAAATATTGAGACTT 3'	56.48	163-171**
PIP R	5' CGTCCTCAAACATCCAGACA 3'	57.54	
TORR F	5' GACACAGGACGACAGAAA 3'	54.18	116**
TORR R	5' GCCTACGCAACTACTAAA 3'	51.66	

The PCR master mixes were optimised, by experimentation, to provide consistent results using DNA extracted from single legs. Table 2, shows the master mix concentrations for the PCR reaction using the primers designed by Smith & Fonseca (2004). Table 3, shows the master mix concentrations for the PCR reaction using the primers designed by (Rudolf *et al.* 2013). To ensure sufficient master mix volume, when preparing master mixes the volume calculations were based on n + 1 reactions or n + 10 %, where n is the number of specimens to process, whichever was the greater. Reactions for all primers took place in a Techne® 5Prime thermocycler using the protocol outlined in Table 4.

Table 2 Master mix details for the PCR reaction. This multiplex reaction uses primers from Smith & Fonseca (2004). The two forward primers are species specific, and the reverse primer B1246s is shared.

Component	Volume (µl)	Final concentration
PCR water	1.1	
5x reaction buffer	4	1.5 mM $MgCl_2$
ACE-pip	0.4	0.2 mM
ACE-torr	0.4	0.2 mM
B1246s	0.8	0.4 mM
dNTPs 2.5 mM of each	2	
GoTaq® G2 Colourless	0.2	
Bovine Serum Albumin (BSA)	0.3	
MgCl2 25 mM	0.8	1 mM

Table 3 Master mix details for the PCR reaction. This multiplex reaction uses primers from Rudolph et al. (2013).

Component	Volume (µl)	Final concentration
PCR water	1.1	
5x reaction buffer	4	1.5 mM $MgCl_2$
PIP F	0.4	0.2 mM
PIP R	0.4	0.2 mM
TORR F	0.4	0.2 mM
TORR R	0.4	0.2 mM
dNTPs 2.5 mM of each	2	
GoTaq® G2 Colourless	0.2	
Bovine Serum Albumin (BSA)	0.3	
MgCl2 25 mM	0.8	1 mM

Table 4 Protocol details for the PCR reaction using ACEpip, ACEtorr and B1246s, PIP F, PIP R, TORR F and TORR R primers. This program was run on a Techne 5Prime thermocycler, with heated lid at 105 °C.

Temperature	Time	Cycles
94° C	5 minutes	1
94° C	30 seconds	35
55° C	30 seconds	35
72° C	1 minute	35
72° C	5 minutes	1
4° C	Hold	Hold

2.2 Gel electrophoresis

2.2.1.1 Preparation of materials for gel electrophoresis.

A solution of 0.5 M ethylenediaminetetraacetic acid (EDTA) pH 8, was prepared as follows: 46.5 g of EDTA and 200 ml of distilled water were mixed in a 500 ml glass beaker with a magnetic flea and pH meter; this was then placed on a magnetic stirrer. 4 M NaOH was added until the pH was 8 and the EDTA was dissolved. The resultant 0.5 EDTA solution was autoclaved before use and stored at room temperature.

Next, 60.5 g of Tris base (Trizma® Base, Sigma-Aldrich), 15 ml glacial acetic acid and 25 ml of 0.5 M EDTA (pH 8) were added to a 250 ml volumetric flask and dissolved together. Distilled water was then added to a final volume of 250 ml. This was stored at room temperature.

20 ml of this 50x TAE buffer and 980 ml of distilled water were then mixed in a 1000 ml Duran® bottle, and stored at room temperature before use. 2.25 g of agarose gel powder was added to 150 ml of 1x TAE buffer, heated in a microwave oven for ~ 3 minutes before cooling in a water bath at 60 °C. 7.5 µl of ethidium bromide was added to this solution, as thoroughly swirled, before pouring the gel and allowing it to set. Resulting in a 1.5 % gel.

2.2.1.2 Sample loading process

5 µl of each DNA sample was mixed with 1 µl of loading dye in a 1.5 ml Eppendorf tube; this 6 µl was then loaded into the allocated well in a 1.5 % agarose gel. Depending on the number of samples being processed, a 100 base pair ladder was added to one or both end wells of the gel plate. The positive controls for *Culex pipiens s.s.* and *Culex torrentium* were also added to wells 2 and 3 of each gel. Gel electrophoresis was carried out at 80 V for ~ 1.75 hours.

2.2.2 Gel Imaging

Gels were imaged using a UVP GelDoc-ItTS2 (Ultra-Violet Products Ltd) imaging system. Images were stored digitally, and physical prints also collected for analysis.

2.2.3 Gel Analysis

Gel Analyzer (version 2010a) (Lazar 2010) was used to calculate the number of base pairs in target lanes, to check that the correct bands were being amplified from the specimens.

2.2.4 Validation of the approach

To demonstrate the effectiveness of the DNA extraction method, the primer pairings and their specific PCR cycles, specimens of *Culex pipiens* and *Culex torrentium*, previously identified by morphological features using established keys (Becker *et al*., 2010a; Snow, 1990), were selected as subjects for a confirmatory assay. These specimens were processed in accordance with the protocols described above.

To ensure that the bands present on the gel were those stipulated by the original primer design papers, and confirmed by NCBI PrimerBlast, the gel was processed using Gel Analyze (Ver. 2010a) (Lazar 2010). Figure 5 shows the highlighted target bands which correspond to those stated as diagnostic in the literature (Smith and Fonseca 2004; Rudolf *et al.* 2013). Table 5 is a comparison of the mean calculated fragment sizes to the expected fragment sizes which were diagnostic for the species. The fragment sizes for PIP F and TORR F are very close to the expected values, the fragment sizes for ACEpip and ACEtorr, however, are less close to the expected value; this is thought to be due to the reduced resolution of the Gel Analyzer software's approach at larger fragment sizes. The space between the 400-500 bp ladder and 600-700 bp ladder bands is significantly smaller than that between 100-100 bp. Therefore the

software has lower resolution. The small SDs indicates that the amplification product is similar in each case.

Table 5 Expected diagnostic fragment size for a primer/species combination, and the mean calculated fragment size resulting from the PCR assay. Standard deviation is included to show the consistency within the method. n=3.

Primer/species	Expected fragment size	Mean calculated fragment size	SD of the calculation
ACEpip/ *Cx. pipiens s.s.*	609	628	2.89
ACEtorr/*Cx. torrentium*	416	429	3.51
PIP F/ *Cx. pipiens s.s.*	163-171	166	2.31
TORR F / *Cx. torrentium*	116	116	4.04

The resulting, annotated, gel image (Figure 6) shows the species as determined by molecular methods. These species determinations agreed in all cases with the *a priori* identification made using dichotomous morphological keys (Becker *et al.*, 2010a). Equally important, is the agreement between the two different primer arrangements, giving further confidence in the approach, and meaning that either would be a viable method for further use in the study. The results using the ACEpip and ACEtorr primers, developed by Smith & Fonseca (2004) were much easier to reliably interpret by eye due to the ~200 bp difference in the fragment sizes.

Figure 5 This image has been processed using Gel Analyze software. Having trained the software with the 100 bp ladders, lanes 1 and 15, the software is able to calculate approximate bp numbers for any band present on the gel.

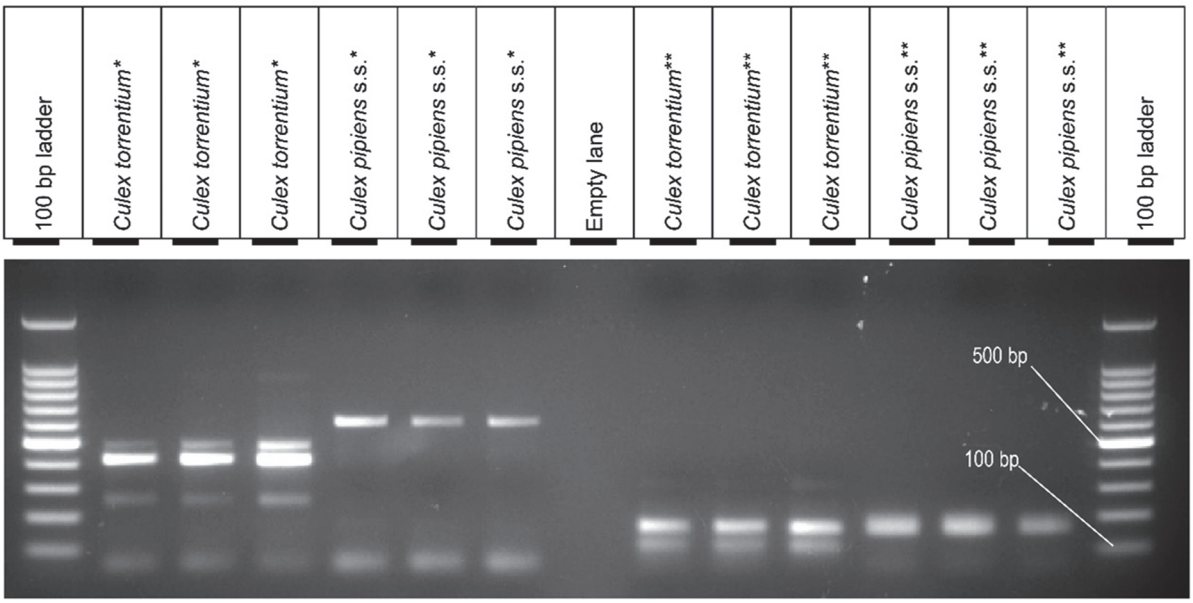


Figure 6 Annotated image of confirmatory PCR gel. *Specimens amplified using Smith & Fonseca (2003) primers. ** Specimens' DNA amplified using Rudolph et al. (2013) primers.

As a final confirmation, the protocol was used to identify a batch of "unknown" *Culex* mosquitoes which were a combination of wild caught adults; adults reared on from larvae collected in the wild and laboratory cultured mosquitoes of each species. The DNA extraction and amplification protocol were shown to work equally well with both the collected and

cultured specimens and, as seen in Figure 7, were sufficient to provide the discriminatory information required for each case.

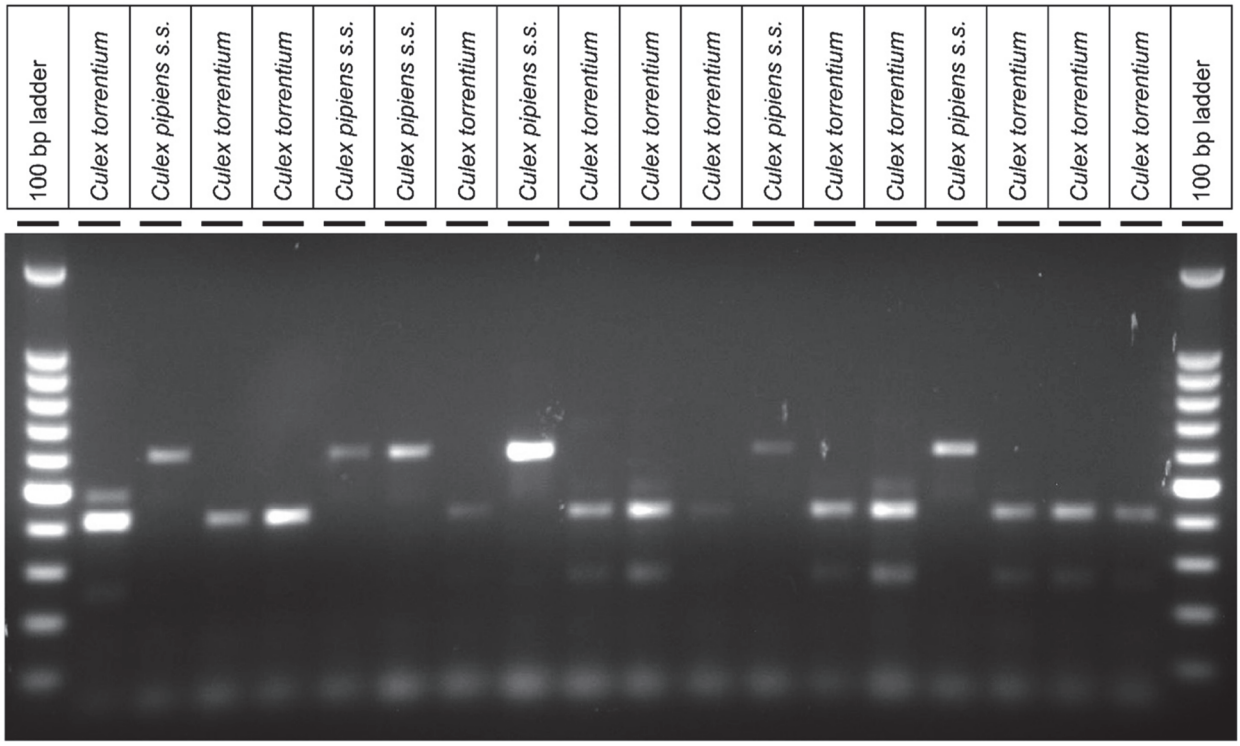


Figure 7 Example of a gel resulting from the PCR method in use to identify unknown *Culex* mosquitoes from field collection. The labels above show the species identifications confirmed by the molecular identification method.

2.3 **Discussion**

When researching possible problems with the chosen method, which might not have been apparent during the validation conducted here, it was noted that in geographical regions of hybridization between *Culex pipiens* and *Culex pipiens* var. *molestus* there could be the potential for misidentification of *Culex torrentium* as either of the forms of *pipiens* when conducting CQ11 based identifications (Danabalan *et al.* 2012). However, *Culex pipiens* var. *molestus* was considered as not being present in the populations under study due to there being no evidence of autogeny during any part of the mosquito culturing undertaken in this study.

It was decided that having optimised and validated the DNA extraction and PCR amplification methods, the Smith & Fonseca (2003) primers would be utilised for the end-point PCR

approach being employed. The main reason for this was that the fragment sizes were much easier to discriminate by eye using these primers than those designed by Rudolph *et al.* (2013). This is to be expected, as Rudolph *et al.* (2013) designed those primers to be used for real-time PCR (rtPCR), and would, therefore, use a fluorescent probe and be machine read in real time.

Throughout this method optimisation and validation, and the wider application of this approach in the broader study, there was no disagreement between morphological identification, using dichotomous keys, and the molecular identification. This was a critical validation of the methods of identification and the veracity of the chosen morphological keys when applied to local *Culex* mosquitoes.

3 Mosquito species present in North Staffordshire and environs

3.1 Introduction and aims of the fieldwork

The aim of the fieldwork was to investigate which mosquitoes are found in North Staffordshire and bordering areas of Shropshire and Cheshire to enable to local species record to be updated, and to investigate their ecology in terms of aquatic habitat preference and temporal abundance. Efforts were made to collect data from different habitat types typical of mosquitoes and to sample both the juvenile, aquatic, populations and the adult, terrestrial, populations.

Multiple sampling methods have been utilised during this study, each focussing on different aspects of mosquito behaviour and life stages. Thereby facilitating evaluation of the mosquito population and beginning to address the lack of spatially and temporally proximate data regarding mosquito abundance and distribution.

As was discussed in the main introduction to this thesis, *Culex pipiens s.s.* and *Culex torrentium* mosquitoes are of particular interest to this study, primarily due to the observed abundance of *Culex torrentium* in northern areas of Europe (Hesson *et al.* 2014), and its demonstrated competency as an enzootic vector of diseases of medical importance to humans and wildlife (Lundström *et al.* 1990; Hesson, Verner-Carlsson, *et al.* 2015). The likelihood that *Culex torrentium* has been historically overlooked, due to the difficulty of species diagnosis, and are therefore underrepresented in the species record for the region seemed very high.

3.1.1 Methods of sampling mosquitoes

Due to their complex life histories, there are multiple approaches to sampling mosquitoes each targeting a different life stage and/or behavioural drivers. Where some organisms have clear ‘best’ methods for sampling, aerial netting adult butterflies, or light trapping moths for example, there is no real equivalent for mosquitoes, with literature showing many methods being successfully applied (Silver 2008a; Becker *et al.* 2010a).

All mosquito sampling methods carry some bias towards a certain life stage or behaviour (Silver 2008a). For example, the use of CO_2 baited light traps is targeted towards and results in the capture of female mosquitoes engaged in host seeking (Muturi *et al.* 2007; Hapairai *et al.* 2013) whereas netting of mosquito swarms will typically collect more males (Knab 1906; Service 1994; Yuval *et al.* 1994). These factors were considered when designing the sampling strategy. To try and gather representative species data, therefore, required the application of various methods, each with different foci.

Broadly, mosquito sampling methods can be grouped by whether they target juvenile or adult specimens and whether they are active or passive approaches. Active sampling comprises the traditional adult trapping, larval dipping, baited landing catch and oviposition trapping methods that are targeted toward mosquito collection specifically (ECDC 2014; Kampen *et al.* 2015). Passive sampling is being used increasingly to avoid the cost, time and labour requirements associated with active sampling (Kampen *et al.* 2015). Passive sampling also includes broad spectrum trapping within which mosquitoes are collected but where mosquitoes were not the target species of other sampling effort (ECDC 2014), and increasingly through citizen science and community involvement (Kampen *et al.* 2015; Vaux and Medlock 2015).

3.1.2 Sampling the juvenile (aquatic) population.

Sampling the juvenile life stages of mosquitoes is a well-used approach for collecting data on mosquito populations. Because juvenile mosquitoes are found in water bodies selected for them by their parent, there is an element of habitat preference which can be investigated through these collections.

Sampling aquatic habitats by dipping with a vessel of known volume, a ‘dipper’, is amongst the most widely used methods for sampling and as such is included in the guidance offered by public health bodies (ECDC 2014), due to its proven efficacy and its low cost. As this is an active method, the amount of sampling possible is limited by the operator’s time and resources such as fuel needed for sampling visits to sites. Dipping has the advantage of being easy to standardise in terms of sampling effort. Another positive aspect of dipping is that it works regardless of the oviposition strategy employed by the adult mosquito, meaning that as the larvae are sampled directly, the observer can collect specimens whose egg may have

been laid above the waterline prior to rewetting, or may have been part of floating raft or laid singly on the water's surface.

Dipping is not without drawbacks however, and results can vary somewhat between operators due to different techniques which can be applied and their ability to avoid eliciting a 'flight' response in the larvae and pupae as they approach and walk at the edge of the water, therefore a time delay between arrival at the water's edge and the commencement of sampling is recommended (Silver 2008b). Different dipping techniques exist and can influence the species represented in a sample, as each method tends to be more effective at sampling the larvae of different mosquito species (Silver 2008b). For example, if the dipper is quickly submerged it will sample differently to a dipper which is slowly lowered into the water such that the water gently flows into the dipper; the slower method might result in bias toward shallow diving mosquitoes, such as *Anopheles*, where the faster method might sample more of the deeper diving species. Species of *Mansonia* and *Coquillettidia* mosquitoes require a particular dipping technique which scrapes the vertical surfaces of submerged vegetation when dipping if they are to be properly sampled (Batzer 1993), and may still be under-recorded by dipping using a conventional dipper as the various techniques devised specifically for sampling *Coquillettidia* and *Mansonia* involve removal or significant disruption of aquatic plants to dislodge larvae (Lounibos and Escher 1983). Another drawback is that dip sampling has potential for under-sampling of first and second instar larvae (Sulaiman 1983) in part due to the reduced time spent at the surface of the water for these early instars (Nielsen and Nielsen 1953) and potentially due seeing first instar larvae in very turbid water. These difficulties aside, dipping has been demonstrated to be a very powerful method for sampling a variety of habitats quickly and in a standardised manner which supports further analysis (Hutchinson 2004; Snow and Medlock 2008; Medlock and Vaux 2014). Standardisation of sampling is quite straightforward and generally requires the use of an identical dipping vessel and dipping technique be applied for the same number of dips at each sampling location.

Another approach to sampling juvenile mosquitoes is to sample eggs prior to hatching; this method is complicated by the various oviposition site preferences displayed between species of mosquitoes. Site preference is determined by a number of biological factors related to maximising embryo offspring survival (Refsnider and Janzen 2010), including avoidance of

predation (Gibson and Torr 1999; Kiflawi *et al.* 2003; Logan and Birkett 2007), selection based on the presence of conspecifics (Kiflawi *et al.* 2003; Sullivan *et al.* 2014) and sufficient nutrition for offspring such as bacteria and other microorganisms associated with decomposing vegetation (Hazard *et al.* 1967; Merritt *et al.* 1992). These factors manifest in a very broad range of oviposition site types across the various mosquito species including artificial containers, tree holes, ponds, marshes and plant axils (Bentley and Day 1989). Together these factors mean that for a comprehensive survey, multiple methods would need to be employed (Silver 2008a). For abundance studies, where species' presence has already been established, sampling methods should be tailored to ensure that they are likely to collect specimens of all species of interest, even if that means applying multiple methods. However, this can be labour intensive and requires a good deal of knowledge prior to the commencement of sampling. In the absence of the species record, best guesses might need to be made to facilitate egg collection. Often this is carried out using artificial oviposition sites, a deliberately placed water container configured to collect target species based on their reported behaviour. The artificial sites range in size, from including small tree hole analogues such as beer cans, plastic beakers and sections of bamboo (Silver 2008a) to larger vessels such as household waste bins (Townroe and Callaghan 2014) and mid-sized buckets (Silver 2008a; Becker *et al.* 2010a).

Other approaches to assessing preimaginal populations considered included collection using nets and visual searches. A recent study found that in circumstances such as sampling Anophelinae in open water, a net yielded better results than a dipper, and performed as well as a dipper for *Culex* species (Brisco *et al.* 2016).

In this thesis, due to the lack of data related to mosquito species in the region (discussed below), the application of narrowly focussed, species/genus specific, sampling was not possible and so the more general approaches were taken. Juvenile sampling consisted of larval dipping, visual searches for eggs, and the use of oviposition traps.

3.1.3 Sampling the adult population

Sampling the adult population offers many approaches, depending upon the ecological question being asked by a given study, each better suited to one situation than another, resulting in many different methods being employed across publications (Silver 2008a; Becker *et al.* 2010a). Sampling using emergence traps has been successfully used to examine the role of specific breeding habitats and the rate of emergence from them (Thullen *et al.* 2002; Fillinger *et al.* 2009; Hamer *et al.* 2011). Data from these studies could be considered more ecologically relevant to factors associated with adult mosquitoes, such as biting and disease, as only those individuals which have successfully completed the juvenile life stage can be included. Emergence traps are positioned on or just above the surface of the water and typically take either conical (Hamer *et al.* 2011) or pyramidal (Slaff *et al.* 1984) form with a killing/capture jar, or sticky capture surface, at the apex, although prism, cuboid and cylindrical shapes have also been used successfully (Silver 2008a).

Active, observer reliant, methods such as manually searching for and collection of resting adult mosquitoes have proven very useful, particularly for mosquitoes which are associated with human habitation, whether entering homes or making use of outbuildings for diel rest periods or seasonal diapause. Depending on the aims of the investigation, sampling via the collection of resting adults can provide lots of information. Mosquitoes spend much of their time resting (Jaenson 1988) in either natural settings (Service 1971b) or man-made shelters (Silver 2008a) as part of their diel cycle and during diapause. Because mosquitoes can be collected during their rest phase, which they spend much of their time and importantly will continue to participate throughout their adult lives. Therefore, it is possible to sample, males, blood fed and gravid females of certain species using one method. However, where mosquitoes use natural structures such as trees and ground vegetation, they become much harder to locate (Silver 2008a). Sweep netting can be utilised to sample mosquitoes resting on vegetation that is supple enough to facilitate the use of the technique (Service 1971b).

A traditional method of collecting host-seeking female mosquitoes human or animal bait trapping, including human landing collection (Silver 2008a). In human landing collection human acts as both the bait and the collector (Trpis *et al.* 1995) and would bare a patch of skin, and wait for the host-seeking female to land searching for a blood meal (Turell *et al.*

2008). Landing mosquitoes are collected using an aspirator and retained for identification. For human or animal, baited traps a means of mosquito interception and collection is typically placed around the bait, normally in the form of a net, cage or box (Silver 2008a). The bias inherent in these methods here is obvious, as it targets only actively host seeking female mosquitoes.

The use of traps for adult mosquitoes is commonplace, with a wide variety of traps available, each with their own advantages and limitations (Acuff 1976; Hapairai *et al.* 2013; Lühken *et al.* 2014). Most traps employ a lure of some kind to attract mosquitoes close enough to the trap to ensnare them. Because a lure or lures fitted to traps are designed to exploit behaviour, they apply a bias to the catch. The bias may be towards particular species or to one sex or the other. As an example, Mosquito Magnet™ traps emit 1-Octen-3-ol and generate CO_2 and heat through the combustion of propane gas. Mosquito Magnet™ traps are therefore very attractive to host-seeking female mosquitoes, which would be expected to feature in far greater numbers than the males in the catch by this device. This is perhaps unsurprising as Mosquito Magnet™ traps were designed as a pest control device rather than ecological survey apparatus; despite this, they are being used increasingly in mosquito ecological studies (Hutchinson 2004; Ritchie *et al.* 2008; Lühken *et al.* 2014; Vaux *et al.* 2015).

Even simple light traps depend on mosquitoes being attracted to the light source being used, and this appears to be affect some species which are underrepresented in catches using these traps. Light traps are often augmented with secondary lures, such as CO_2 emission devices to enhance their catch (Muturi *et al.* 2007), but again this can result in further reduction of species represented in the catch.

Attempts to carry out trapping without any type of lure have been utilised at various times, and typically takes the form of an unbaited net mounted on the roof of a vehicle and be driven around the region to be sampled (Sanders *et al.* 2012). These unbaited methods do tend to result in quite diverse catches including an increased number of species compared to baited traps (Bidlingmayer 1967). They may still be affected by biases such as only sampling the portion of the population which fly close to the trap, thereby missing the proportion of the population that are resting; or in surveys using vehicle mounted suction traps they might have to follow roads, therefore constraining sampling. Table 6 shows examples of the relative

abundance differences of mosquito species collected using different trapping method in previous studies; it clearly shows that trapping methods influence the samples collected. These biasing factors in unbaited trap sampling were grouped as Environmental, Biological and Operational by Bidlingmayer (1967).

Table 6 Relative abundances of various mosquito species collected using different adult collection methods. It can be seen that species composition can vary by both collection method and location.

Paper	Collection methods used	Location	Type of study	Species collected	Relative abundance in sample (by trap type if relevant)		
Acuff (1976)	CDC light trap + CO_2 (**CDC**) Vs. New Jersey light trap (**NJ**) Vs. Malaise trap (**MT**)	Iowa, USA	Trapping bias study		**NJ**	**MT**	**CDC**
				Aedes trivittatus	10.1%	6.8	23
				Aedes vexans	83.4%	86.4	70.4
Luhken *et al.* (2014)	Biogents Sentinel trap with BG Lure + CO_2 (**BGS**) Vs. CDC mini light trap + CO_2 (**CDC**) Vs. Mosquito Magnet Patriot trap + R-Octenol (**MM**)	Germany	Field evaluation of traps in central Europe		**BGS**	**CDC**	**MM**
				Aedes cinereus/geminus	20.6%	10.3%	26.2%
				Aedes vexans	11.1%	46.8%	25.9%
				Ochlerotatus cantans	16.0%	7.5%	2.1%
				Ochlerotatus rusticus	10.9%	6.2%	16.1%
				Culex pipiens/torrentium	18.5%	8.6%	0.7%
Bidlingmayer (1967)	Unbaited suction trap (**ST**) Vs. New Jersey Light trap (**NJ**) Vs. truck mounted trap (**TMT**)	Florida, USA	Comparison of trapping methods		**ST**	**NJ**	**TMT**
				Anopheles crucians	32.8%	81.2%	30.5%
				Anopheles quadrimaculatus	23.7%	7.4%	7.7%
				Culex nigripalpus	21.9%	3.8%	9.3%
				Mansonia perturbans	12.8%	2.4%	7.5%
				Psorophora confinnis	3.6%	2.5%	24.5%
				Aedes sticticus	3.1%	1.8%	2.9%
				Aedes taeniorhynchus	1.1%	0.5%	4.0%
				Aedes sollicitans	90.0%	0.3%	13.6%
Thullen, Sartoris and Walton (2002)	Emergence traps	California, USA	Effect of engineering intervention on mosquito production by aquatic habitats	*Culex tarsalis*	91.0%		
				Culex erythrothorax	8.0%		
				Culex quinquefasciatus	2.0%		
				Culex stigmatosoma	1.0%		
Fillinger *et al*. (2009)	Emergence traps	The Gambia	Investigation of most productive breeding sites for malaria mosquitoes	*Non-Culicidae insects*	92.4%		
				Culicinae	7.3%		
				Anophelinae	0.3%		
				Of the Anophelinae: Anopheles coustani group	66.0%		

Continues overleaf

Paper	Collection methods used	Location	Type of study	Species collected	Relative abundance in sample (by trap type if relevant)		
Fillinger *et al.* (2009) Continued				*Anopheles gambiae* s.l.	23.0%		
				Anopheles pharaoensis	4.0%		
				Unidentified Anopheles spp.	7.0%		
Hamer *et al.* (2011)	Emergence traps	Chicago, USA	Testing of novel design of emergence trap	*Culex pipiens*	84.4%		
				Culex restuans	15.6%		
Slaff *et al.* (1984)	Emergence traps	Florida, USA	Emergence trap modification testing	*Mansonia* spp.	95.0%		
				Coquilletidia perturbans	< 5%		
				Culex nigripalpus	< 5%		
				Culex salinarius	< 5%		
				Anopheles crucians	< 5%		
				Uranotaenia sp.	< 5%		
Jaenson *et al.* (1988)	Human Landing	Central Sweden	Diel activity of blood feeding activity of anthropophilic mosquitoes	*Aedes communis*	39.7%		
				Aedes punctor	23.4%		
				Aedes cinereus	12.0%		
				Aedes annulipes	6.9%		
				Coquilletidia richiardii	5.6%		
				Aedes cantans	3.6%		
				Aedes cataphylla	2.7%		
				Aedes excrucians	2.0%		
Turell *et al.* (2008)	Human Landing	Peru	Mosquito distribution and human attraction	*Anopheles darlingi*	49.7%		
				Culex spp.	12.7%		
				Aedes spp.	7.4%		
				Mansonia spp.	8.1%		
				Psorophora spp.	3.7%		
				Other species	3.2%		

3.1.4 Current distribution records

Given that the British Isles have been free of autochthonous mosquito-borne disease for a century (Hutchinson 2004), data related to the distribution of mosquitoes in England was found to be limited, and its collection was unevenly distributed. Using data from the National Biodiversity Network NBN Gateway (NBN) (Figure 8), it was apparent that the South East of the UK had the best data, and pockets of locally useful information existed for the Birmingham, North Wales and Sheffield areas. Categorising the data to view the major genera show different patterns of distribution among them. Figure 9, illustrates the distribution of *Anopheles* mosquitoes which, with the exception of Anglesey in North Wales, was relatively uniform with small aggregations of records across much of the country. This contrasts somewhat with Figure 10 which shows a much more clumped distribution for the *Aedes* and *Ochlerotatus* genera.

The distribution of *Culex* mosquitoes appeared to fit somewhere between the two (Figure 11), with some centres of increased abundance, or simply increased reporting, with a number of smaller distributed records. The *Culex* genus records submitted to the NBN contain specimens referred to as *Culex pipiens agg.* or only referred to as *Culex*, highlighting the difficulty in identifying specimens to species in this genus (NBN 2017). Interestingly, even those identified as *Culex pipiens* might leave the reader in doubt, as often there appeared to be some interchangeability in the use of terms, and the acceptance that within these data, in a UK setting, *Culex pipiens* denotes a complex, of species including *Culex pipiens s.s.*, *Culex pipiens var. molestus* and *Culex torrentium*. It was assumed, however, that within the NBN dataset, *Culex pipiens* means *Culex pipiens s.s.,* and that *Culex* and *Culex pipiens agg.* denote either incomplete identification beyond genus and the complex of three species (*Cx. pipiens s.s*, *Cx pipiens var. molestus*, *Cx. torrentium*) respectively. It should be noted that, when using a strict global taxonomic perspective on the members of the *Culex pipiens* complex/assemblage the three member species are *Culex pipiens*, *Culex pipiens* var. *molestus* and *Culex quinquefasciatus* Say (Harbach 2012).

The distribution of *Culex torrentium* (Figure 12) was surprising, with very few records. These records were limited, for the most part, to the south-east, with another cluster in Sheffield. Whether this is a result of actual absence or due to underreporting through misidentification,

or lack of imperative to discriminate between *Culex torrentium* and the other *Culex pipiens* s.l., is unknown, but special care was taken throughout the fieldwork to identify this species properly where possible. Not all data on British mosquito sampling has been uploaded to NBN, and a notable omission from the records is that of Service (1994) which related to the study of male *Culex torrentium* swarming near residences in and around West Kirkby, situated to the south west of Liverpool. It was also vital to use a clear naming convention throughout the study; therefore, *Culex pipiens s.l*. was used to denote the complex of species, *Culex pipiens s.s.* was used for the single species and is synonymous with *Culex pipiens pipiens* (Chandler 2017).

The area surrounding Stoke on Trent had very few records of mosquitoes (Figure 13). Interestingly, many of the records on NBN were for single specimens, and many of the entries were quite old, with the oldest dating back to 1827. In their paper discussing passive monitoring of mosquitoes, Kampen *et al.* (2015) reported only 3 species of mosquitoes for Staffordshire following their interrogation of the available species records and most of the sites visited and considered for inclusion this thesis showed that no mosquitoes had been added to the species record at the time of writing. The age of records was also particularly important when considering their value in ascertaining the current distribution of mosquitoes. The northern range boundary of a number of insect species falls within the British Isles (Hickling *et al.* 2005) and so it is highly likely that the same applies to mosquitoes. In the face of climate change, it must also follow that these northern range boundaries are in constant flux (Benedict *et al.* 2007; Campbell *et al.* 2015), causing the distribution of mosquitoes to increase northward and to increased altitudes (Calistri *et al.* 2010; Bessell *et al.* 2014); whilst also increasing the temporally active season of mosquitoes due to longer, warmer and wetter summers (Medlock and Leach 2015). This meant that it was not possible to assume that distribution data from other geographical areas may be suitable for superimposition onto similar ecotypes within this study area.

The paucity of local distribution data and the potential shift in ranges caused by climate change (Medlock and Leach 2012), increase urbanisation (Townroe and Callaghan 2014) and flood mitigation (Medlock and Vaux 2015b) demanded new and ongoing, ecological fieldwork to assess where mosquitoes are present and to try to evaluate habitat associations thereby

facilitating improved understanding of mosquito species' distributions. By the very nature of the fieldwork, and the lack of data in this geographical area it was an aim of the research to improve the local species record, by sharing findings with the NBN, either directly or via public engagement with passive surveillance through Public Health England's Mosquito Recording Scheme (MRS). Only through the sharing of data, will it be possible to generate a comprehensive mosquito distribution description for future use in response to outbreaks of disease or the arrival of new invasive species such as *Aedes albopictus* (Medlock *et al.* 2017).

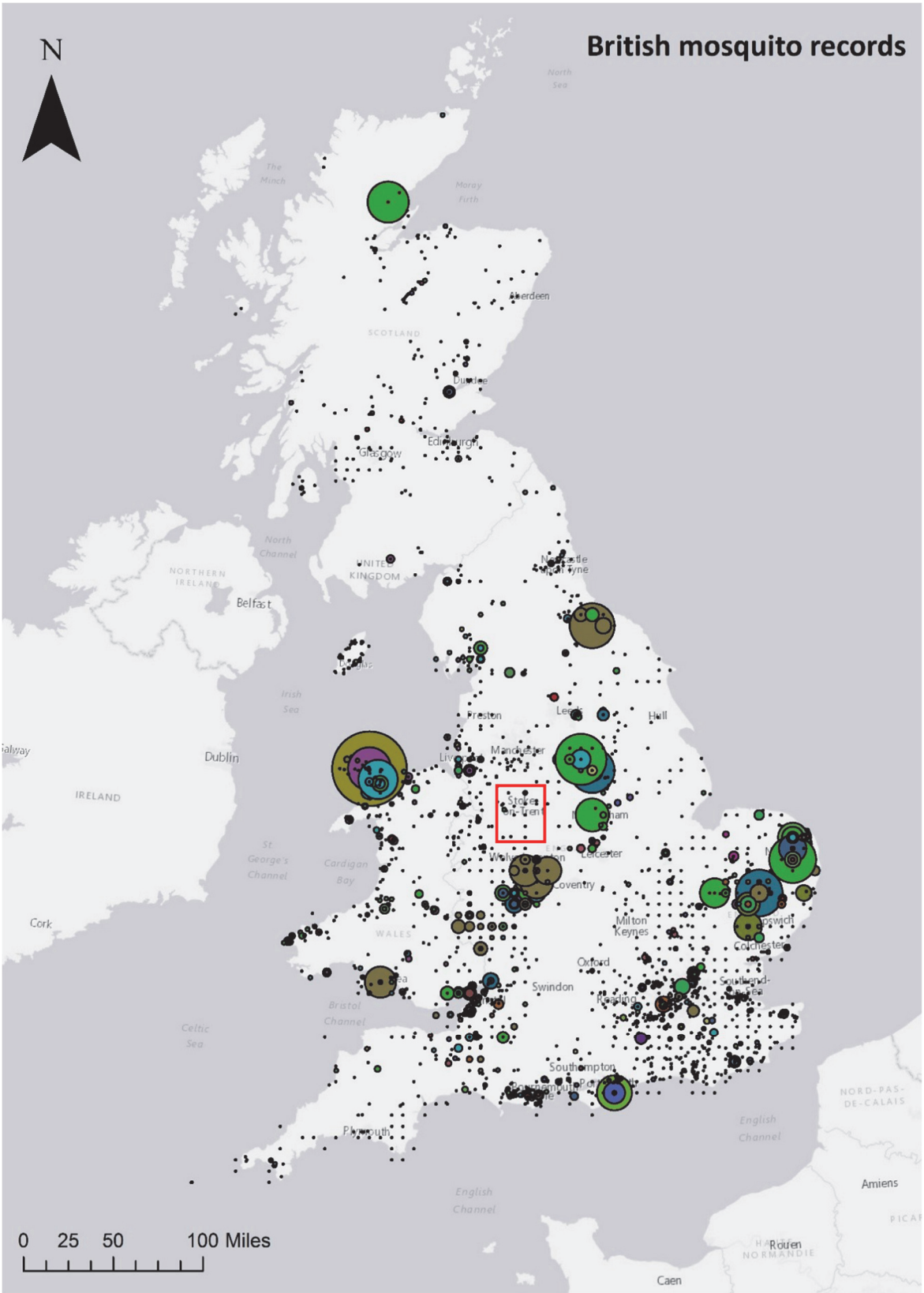


Figure 8 Locations of species recording data from NBN mosquitoes (As 09/01/2017). Different colours represent species and the size of the circle proportional to the number of records per taxon for each location. The smallest dots represent single specimens. Specimen dates range from 1827 to 2016. The red rectangle describes the area being studied in this thesis.

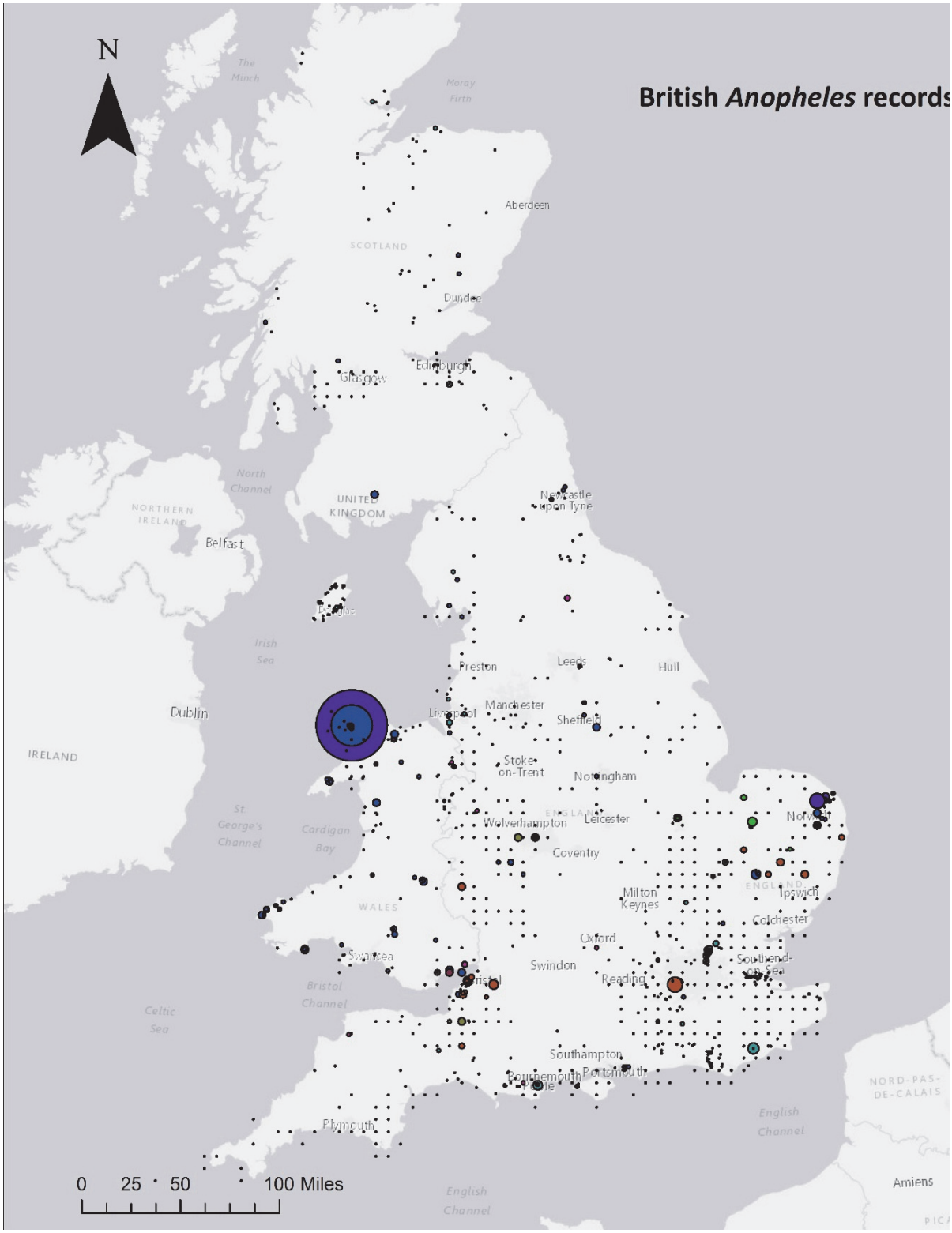


Figure 9 Locations of *Anopheles* mosquitoes from NBN (As 09/01/2017). Different colours represent species and the size of the circle proportional to the number of records per taxon for each location. The smallest dots represent single specimens. Specimen dates range from 1900 to 2015

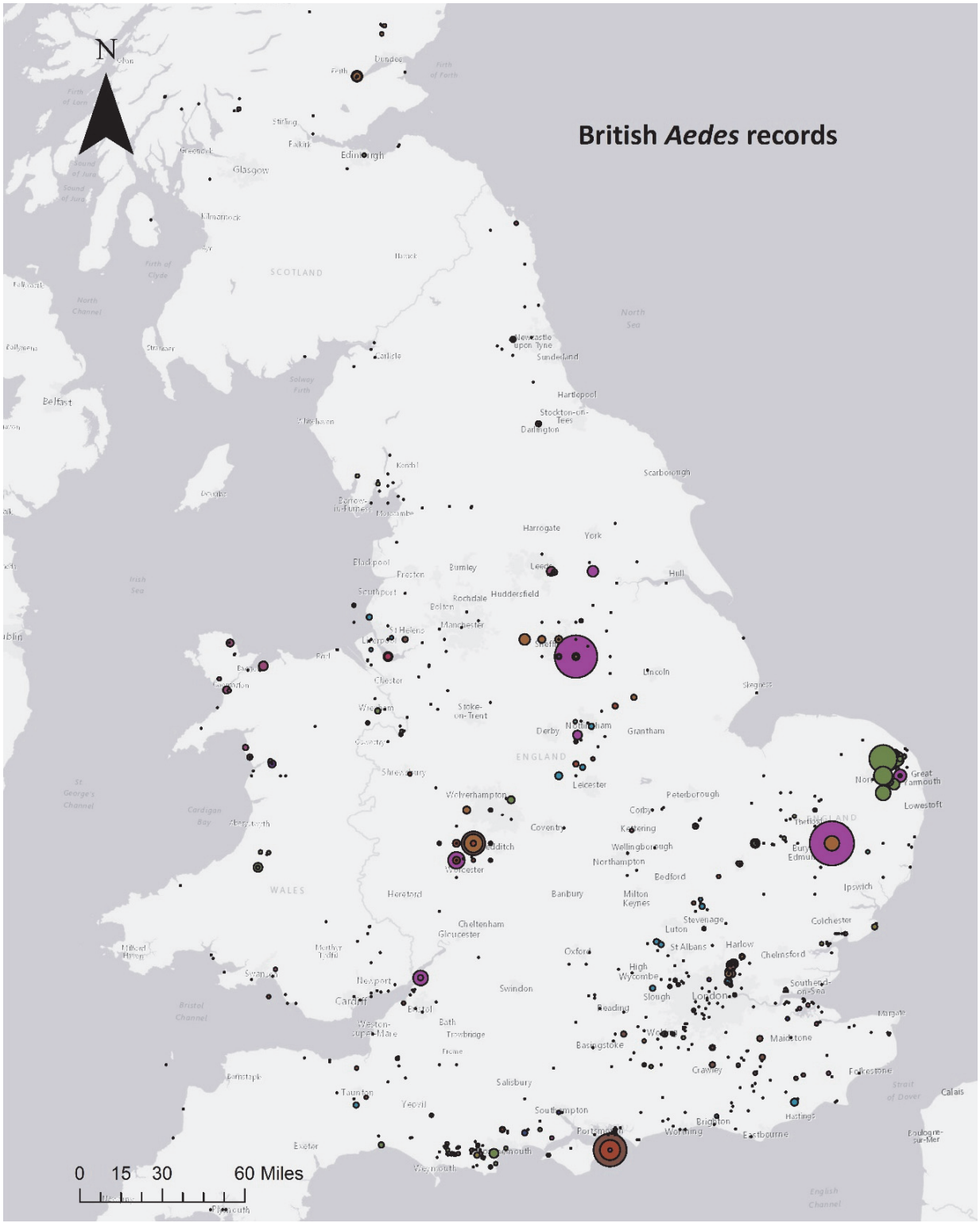


Figure 10 Locations of *Aedes* and *Ochlerotatus* mosquitoes from NBN (As 09/01/2017). Different colours represent species and the size of the circle proportional to the number of records per taxon for each location. The smallest dots represent single specimens. Specimen dates range from 1827 to 2016.

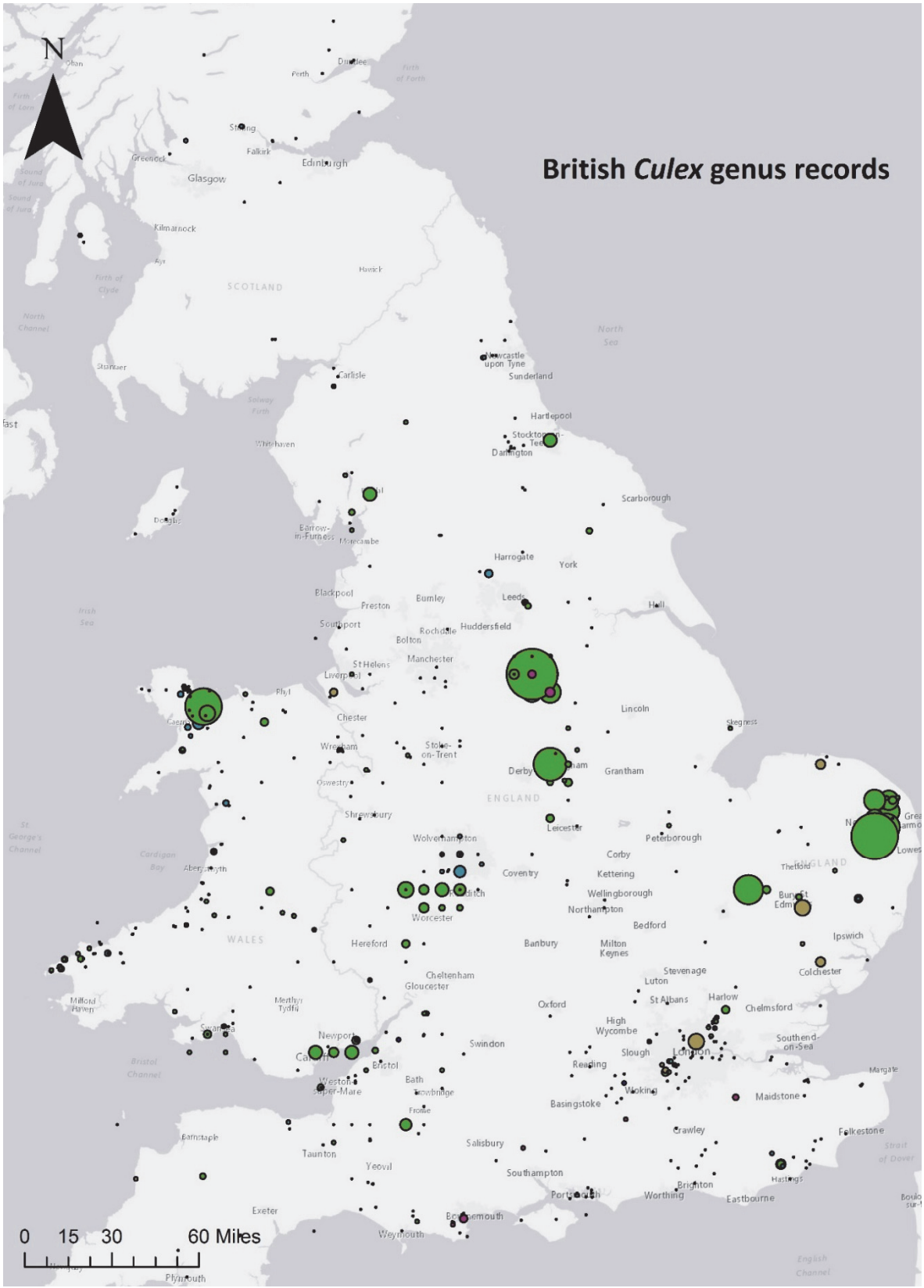


Figure 11 Locations of *Culex* mosquitoes from NBN (As 09/01/2017). Different colours represent species and the size of the circle proportional to the number of records per taxon for each location. The smallest dots represent single specimens. Specimen dates range from 1897 to 2016.

Figure 12 Locations of *Culex torrentium* from NBN (As 09/01/2017). As can be seen, there are very few records suggestive of a very restricted distribution and low abundance. The smallest dots represent single specimens. Specimen dates range from 1951 to 2014.

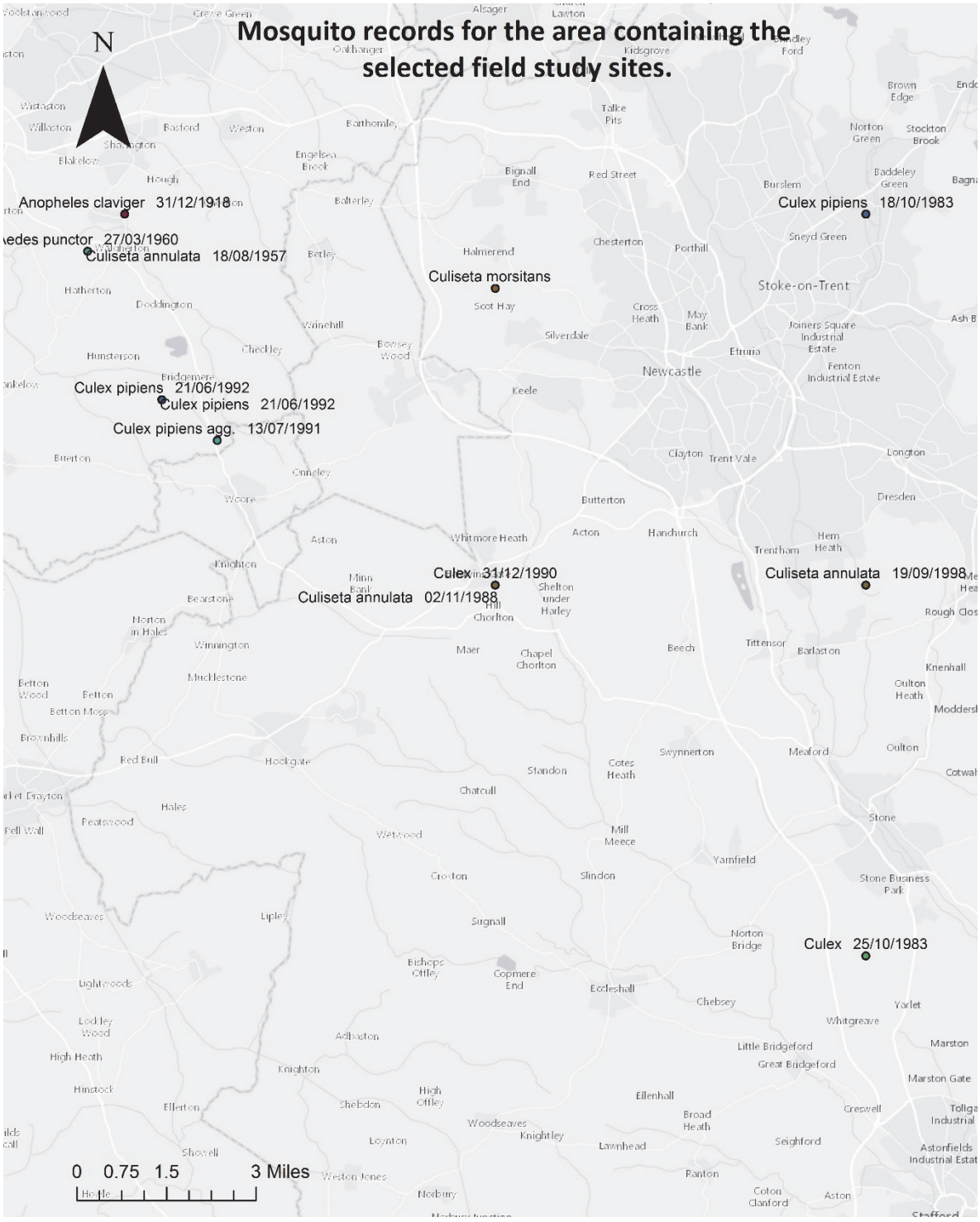


Figure 13 Mosquito species records listed, with the date of the record, on NBN for the geographical area which includes all the sites sampled during this study. None of the records that declare their date are in the 21st century.

3.2 An investigation of the mosquito species present in the North Staffordshire and environs

The principal aim of this chapter was to begin to address the paucity of temporally relevant data related to mosquito species occurrence in the North Staffordshire and its environs by surveying a variety of sites within the region. From the data from these surveys it was possible to conduct some preliminary investigations of other factors related to the mosquito populations surveyed.

These secondary aims included:

- Visualisation of mosquito temporal abundance in suburban settings.
- Was the spatial distribution of mosquitoes at Wybunbury Moss SSSI NNR evenly distributed?
- The fitting of a presence/absence analysis model for mosquito larvae in relation to rural habitat variables.
- A comparison of the mosquito sampling rate between rural and suburban settings using CDC 512 mini light traps. (Secondary data analysis)

The last of the analyses listed above should be considered as a secondary analysis only. The data used for this statistical investigation, whilst collected by the author, was not collected for the purpose of comparison between sites, rather it was part of the general survey data. Analysis was, therefore, a speculative treatment and limited by the imbalanced data collection regime and any outcomes should be considered with this limitation in mind.

For the purposes of this study, sites which were outside of the urban areas, those closely associated with the commercial and industrial centres of the city, and suburban areas, those typically forming the residential outskirts, of the City of Stoke on Trent and the Borough of Newcastle Under Lyme were considered to be rural.

3.2.1 Method

3.2.1.1 Rural Field Site selection

During February 2014, the process of identifying sites for inclusion in the ecological field sampling portion of this study took place. As previously discussed, there was a paucity of

recently published data for mosquito abundance locally and most potential study sites had no available records of mosquito presence and those that did have records tended to be of single individuals. Data were collected from the NBN Gateway for the SJ Ordnance Survey map square, and from the Chartered Institute of Environmental Health National Pest advisory Panel for the Midlands of England and for Wales (http://www.cieh.org/policy/npap_uk_sightings.html). These data were then visualised in ARC GIS (ESRI 2014) to allow identification of sites, as very often place names were not included in the data record, especially for older records which may have predated some of the more recently established nature reserves. The locations within the Chartered Institute of Environmental Health data were defined by postcodes which were converted to eastings and northings to facilitate mapping and visualisation.

The aim of the mosquito surveying was first and foremost to discover which species of mosquito were present in North Staffordshire and the immediate area, therefore site selection focussed on sites which would potentially have a diverse species composition. The study sites needed to offer diverse habitats, ideally a mosaic of identifiable habitat subtypes, which would potentially support a varied group of mosquito species. They would need to be suitable and safe for a lone-working arrangement to be put in place. As a result of these criteria, the sites considered most viable for study in the region were managed nature reserve. From those in the region, nine potential sites were then shortlisted, and following subsequent habitat assessment, securing of site sampling permits and risk assessment, the three rural sites to be included in the study were decided upon.

These sites were Jackson's Coppice and Marsh (378868 Eastings, 32977 Northings), Loynton Moss (378834 Eastings, 324217 Northings) and Wybunbury Moss (369638 Eastings, 350131 Northings), (Table 7). These sites offered the best balance of habitat suitability and variety, whilst still being accessible and safe for lone-working.

Table 7 An overview of the sites considered for inclusion in the field study, including pros and cons and the decision whether to include the location or not. NNR = National Nature Reserve, SWT = Staffordshire Wildlife Trust nature reserve, SSSI = Site of Special Scientific Interest, RAMSAR = Site is a RAMSAR nature reserve.

Place name	Location	Pros	Cons	Selected for inclusion in the study?
Prees Heath Common Reserve	Whitchurch, Shropshire	Excellent site access and easy parking	A parched and open site – prone to being windswept. Sandy soil is very well drained – pools even after rain. Few obvious oviposition sites	**Rejected**, too dry and windswept. Too few different habitat types
Fenn's, Whixall and Bettisfield Mosses NNR SSSI	Welsh Border, near Ellesmere and Wrexham	Rich, diverse site with many habitat types Good road links and parking. Many different mosquito oviposition options.	A large site, with some potentially dangerous peat bog drains. Relatively well studied by previous insect study groups. Patchy mobile phone reception	**Rejected.** Although an excellent site, its large size and dangerous features make it less suitable than other sites. Previous studies reduce the likelihood of finding new species for the site.
Doxey Marshes. Staffordshire Wildlife Trust nature reserve	Stafford	Wetland and grassland habitats Excellent road links and parking	A lot of flowing water, with relatively few pools suitable for mosquito oviposition. Very high footfall site, can't leave traps, etc. – theft and vandalism issues	**Rejected.** Very high footfall limits the sampling method availability too much. The relative lack of standing water reduces manual sampling too.
Bradwell Woods. Some local nature reserve,	Bradwell, Newcastle under Lyme	Easy access. Varied habitat, including some remnant ancient woodland and wetland.	Very hilly on much of the site. Lots of obvious vandalism to lighting and graffiti. People were walking around with long dogs and air rifles. Does not feel "safe" for lone working	**Rejected.** Lone working issues.
Black Firs & Cranberry Bog. SWT	Near Betley, Staffordshire	Peat bog, and wet woodland moss. Cranberry bog has lots of potential, due to high standing water year-round.	Parking is difficult, and access to Black Firs is very tricky through very overgrown paths.	**Rejected.** No access to Cranberry bog to the public by order of SWT risk assessment. Black Firs site not interesting enough on its own to justify travel
Bateswood Nature reserve	Leycett, Staffordshire	Grassland and woodland habitats. Access not difficult – requires access through locked gate Lower foot traffic real potential for leaving traps on site	Very large site meaning a lot of walking/carrying equipment between locations.	**Rejected.** Placed on the backup list. Not as good as selected sites.
Loynton Moss SSSI SWT	Woodseaves, Staffordshire	Good access and parking. Diverse habitat mosaic on a small site. Meadow, grassland, dry woodland, wet alder carr woodland	Relatively high footfall limits the use of "expensive looking" traps. Limits methods to manual sampling and cheaper traps which the project could afford to lose.	**Accepted.** An excellent diversity of habitats at this location. Good access to a safe site.
Jackson's Coppice and Marsh	Cop Mere, Staffordshire	2 small reserves separated by a short walk. Good access with a small carpark. The Marsh is an intermittently flooded wetland with the river Sow flowing through it. The Coppice is an ancient woodland remnant, in well-drained soil. This is high potential for tree hole dwelling mosquitoes	Few obvious cons. The Coppice has a large badger population and so care with footing is important when moving around the site	**Accepted.** Best potential for sampling tree hole mosquitoes. Good access
Wybunbury Moss NNR SSSI Ramsar	Wybunbury, Cheshire	A Schwingmoor, and one of only 3 known subsidence mires in the British Isles. Aside from the floating bog: the surrounding habitat mosaic includes, wet woodland, mixed woodland, fen and pasture.	Due to the nature of this site special access permission is required to carry out work on the majority of it. A safety briefing from the warden is also a requirement.	**Accepted.** The restricted access of the site means that more trapping methods can be used here than any of the other sites due to very low footfall and only pre-vetted persons having access to the sample area. The site is currently under-recorded with only 1 mosquito species on the record.

Wybunbury Moss National Nature Reserve (NNR) is managed by Natural England and covers approximately 16.5 hectares and is one of only three known ‘subsistence mires’, or Schwingmoor, in the British Isles. Located 2 miles south of Crewe in Cheshire, it consists of a water-filled basin, over 12 m deep, upon which a raft of peat floats. This peat raft is as little as 1 m thick in places (Figure 14), meaning that the ecological succession is limited by the ability of the raft to support larger plants and trees as their increased weight caused them to fall through the peat layer and drown, resulting in a quite unique landscape of isolated standing dead wood (Figure 15). The central basin is surrounded by woodland, reed swamp and flower rich meadows. As an NNR it is managed to stem the ecological succession and invasion of woodland and scrub around the edges of the peat raft and to increase the water level across the reserve by adding sluice boards to partially block the drains around the site which had been placed during past attempts at drying the reserve. At the interface between the Moss, the main drain and the wet woodland areas of the site (Figure 16) juvenile mosquito activity could be observed from the public footpath even during February. Finding mosquito activity in the dead of winter was highly encouraging and when considered with the other site features made this a very desirable place for study.

Figure 14 View across the floating peat layer at Wybunbury Moss is punctuated by the dead trunks of trees that have grown too large for the thin peat layer to support; resulting in them sinking through and drowning. This results in a halting of ecological succession and maintains high light levels for the ground flora.

Figure 15 At Wybunbury Moss there are several, apparently shallow, exposed pools that do not dry up even during summer. Presumably, the dip in the surface peat layer allows the underlying water to maintain the pool.

Figure 16 At Wybunbury Moss, the interface between the wet woodland, the floating peat layer and the main drain create a lush, warm and permanently wet region within which mosquitoes were observable even on the first site visits in February 2014. This photograph is from May 2014.

Loynton Moss nature reserve is managed by Staffordshire Wildlife Trust (SWT) and is located 6 miles to west of Stafford in Staffordshire and covers an area of 13.40 ha. It comprises a varied mosaic of habitats surrounding a central Moss, covered in tall reeds (Figure 17). Unlike Wybunbury Moss, the kettle hole is filled with peat and other soils, resulting in a more physically stable Moss, which is, therefore, more susceptible to ecological succession as it can support the weight of larger trees. Away from the central Moss, there are areas of very wet woodland dominated by alder and silver birch in peaty soils (Figure 18). Staffordshire Wildlife Trust's website gives this advice regarding this wet woodland area "Visitors might not welcome the attention of the Moss's other family of insects - mosquitos. During the summer months, these insects are on the wing deterring all but the most determined visitor from entering the darker, wetter areas of the Moss." (Anonymous 2017). Dry woodland on sandy soils make up the raised areas of the site and include many mature and over-mature beech and oak trees, and supports an extensive ground flora, including many wildflowers.

Figure 17 View across the central Moss at Loynton Moss is dominated by tall reeds. There is relatively little open water in this area, and once into summer, it is not unusual for it to dry up completely.

Figure 18 Wet woodland at Loynton Moss is very wet, with several areas as in this photograph where vegetation is limited to raised islands. Tree roots are very shallow here, and windthrow is very common and results in deeper pools which retain water when all else has dried.

Jackson's Coppice and Marsh is also an SWT nature reserve and is located 7 miles to the north west of Stafford in Staffordshire and covers an area of 8.10 ha. Jackson's Coppice and Marsh consists of two distinct habitat types adjacent to each other. Jackson's Coppice occupies a south facing hillside and is an ancient woodland remnant, with some other more exotic tree species added at the end of the 19th century (sweet chestnut, beech and hornbeam). The orientation, soils and deciduous flora allow for a huge number of bluebells in the spring; the bulbs of which are an important food source for the large badger sett that thrives on the Coppice, providing an additional source of mammalian blood for mosquitoes. The mix of trees on the site and the lack of ground water, due to the hilly nature and sandy soil, made this site the best candidate for finding juvenile mosquitoes in natural tree holes. The second part of the nature reserve, Jackson's Marsh, is an alder carr woodland where the river Sow splits into a delta before recombining downstream (Figure 19). A raised boardwalk facilitates access throughout most of the site, with much of the remaining area accessible with caution.

Figure 19 Marsh delta region of Jackson's Marsh. Fluctuation in water level of the River Sow results in the formation of many ephemeral pools; providing excellent potential habitats for mosquitoes whose eggs require wetting and rewetting to trigger hatching.

3.2.1.2 Site access and permission

Following consultation with Natural England, the relevant permits were secured for sampling access at Wybunbury Moss NNR for the whole duration of the project. Similar consultation with Staffordshire Wildlife Trust secured appropriate permits for sampling Jackson's Coppice and Marsh and Loynton Moss for the entire length of the project.

3.2.1.3 Suburban site selection

Several suburban locations were selected to sample the mosquito population, both for juvenile and adult mosquitoes. Suburban sites were selected as these were considered to be representative of most of the residential regions North Staffordshire, and enabled the study of peridomestic locations (those close to human habitation). Because the studies were to be longitudinal in nature and would require the placement of traps, it was important that each site was secure and have access arrangements facilitating study in the longer term. Suitable locations were selected in back gardens in Porthill, Stone, Ashley and Kidsgrove (Figure 20). These gardens were typical residential gardens, with a mixture of flowers, shrubs and trees; all, with the exception of the site in Stone which comprised hard standing and raised beds, had a significant portion set to lawn. Samples were also collected from the Staffordshire University grounds and from outdoor aquatic plant growing containers at Kidsgrove Aquatics in Kidsgrove (Figure 20).

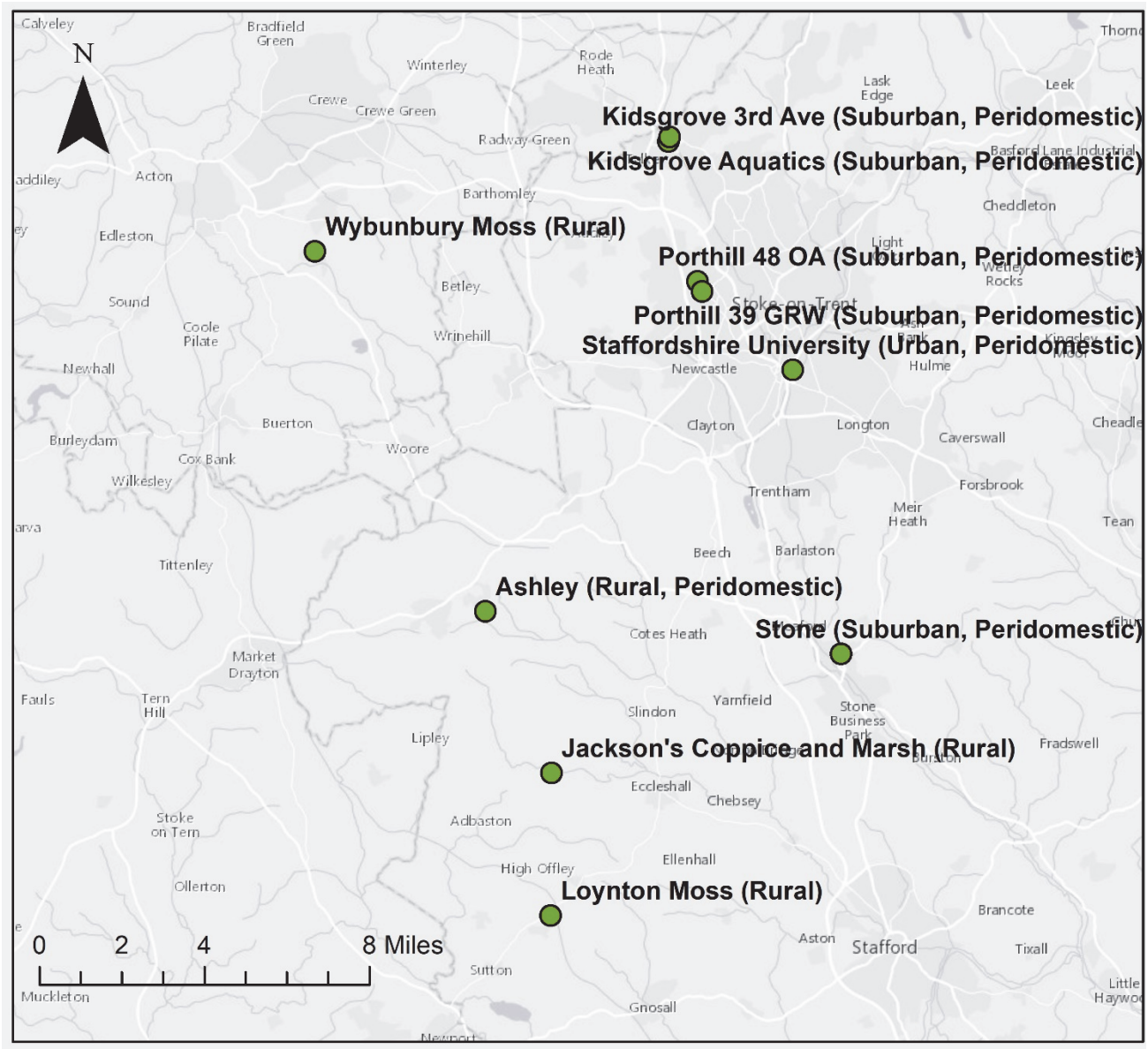


Figure 20 Sampling locations used in this study, and site types, as designated by the author.

3.2.1.4 Sampling methods

A rotational strategy for sampling work was scheduled into phases each focussing on different collection strategies across the sampling sites. Figure 21, gives an overview of the focus of the fieldwork with regards to the methods described in this chapter, across the three seasons available to the study.

Adult mosquito collection was conducted using CDC 512 mini light traps, without CO_2 bait, and Mosquito Magnet™ traps. These were selected based on their proven effectiveness in

sampling mosquitoes (Clarkson and Setzkorn 2011; Versteirt *et al.* 2013; Lühken *et al.* 2014) and their availability to the author. Sweep netting was considered but rejected as it could not be used at all the nature reserves being surveyed. This was primarily due to the presence of very rare invertebrate fauna located at Wybunbury Moss SSSI NNR and the risk of damaging their habitat (Piper and Compton 2013). There was little opportunity to effectively collect day resting adult mosquitoes as none of the rural sites had suitable buildings as have been successfully sampled in UK surveys (Danabalan 2010; Brugman *et al.* 2015), such as those associated with farming, and successfully locating mosquitoes on vegetation is very inefficient as their distributions are very localised (Service 1971b).

When considering the sampling effort required to gather representative samples using CDC light traps *a priori* sample size calculations were not conducted. This was largely due to the challenge in obtaining relevant data related to the efficiency of CDC light traps in comparable situations in the UK. Whilst CDC light traps have been used in a number of mosquito studies in the UK most have also employed CO_2 as an additional attractant (Hutchinson 2004; Hutchinson *et al.* 2007; Lindsay *et al.* 2010), fewer have used the light trap without the addition of CO_2 (Sulaiman 1982; Clarkson and Setzkorn 2011). Although Sulaiman (1982) does report data related to total catch, four mosquitoes over 13 nights, there is only this single iteration to utilise which is unlikely to be representative due to its short duration. Clarkson & Setzkorn (2011) by contrast generated far more data, but unfortunately collated the trapping data of both CDC light traps and Mosquito Magnet traps together meaning that it was not possible to derive the trap specific mean and standard deviation required to complete the calculation. Therefore, the decision was taken to complete as long a study as possible given the available resources.

A similar challenge was presented when determining the sample size required for the oviposition sampling. The most complete study using oviposition traps in the UK was perhaps that conducted by Yates (1979) which targeted *Dahliana geniculata* in the Monks Wood National Nature Reserve in Southern England. In this setting high numbers of eggs were collected using oviposition traps, however the mean per night catch rate and standard deviation are not reported. Similarly, the data needed to generate near approximations for these values were not reported, neither the actual number of trap nights per year nor an

indication of the daily variation were reported (Yates 1979). However, the sampling effort as designed here exceeds that used in this UK study.

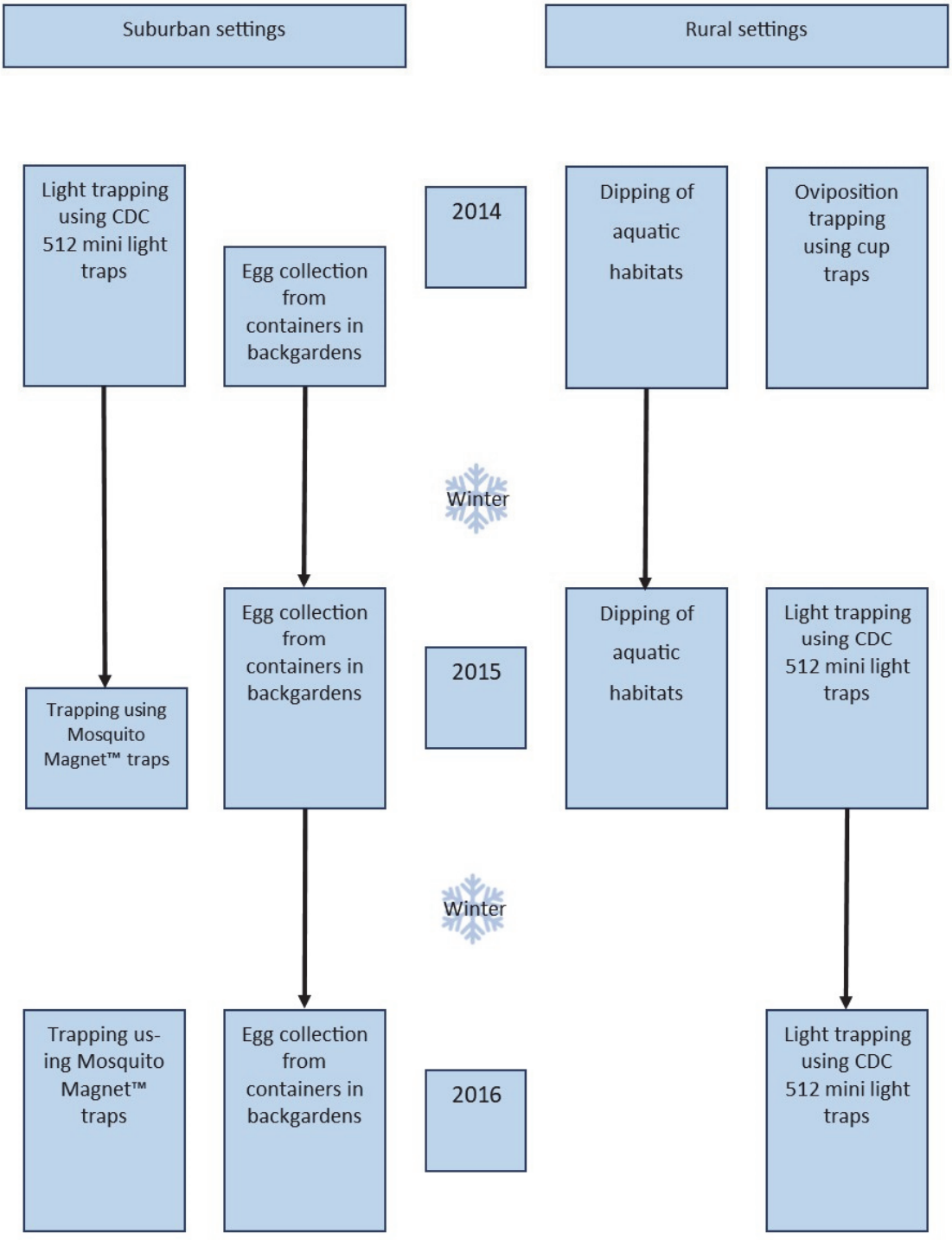


Figure 21 Overview of the surveying program, showing the method focus for each area each year.

3.2.1.5 Sampling using oviposition traps.

Due to the difficulties associated with locating and safely sampling from tree holes for mosquito juveniles, oviposition traps were used to sample tree hole and small container-dwelling mosquitoes. Historically, sections of bamboo have been used successfully to collect British mosquitoes (Service 1968). The oviposition traps used were very simple devices consisting of black plastic beakers which were tied to the trunks of trees (Figure 22). Each beaker had an 8 mm drain hole drilled below the upper lip to prevent overflow. A hardboard stick, ~ 2.5 cm wide and 15 cm long, was placed into the water as a site for oviposition for mosquito species which lay their eggs above the water line. The absorbent nature of the hardboard and rough surface make it an ideal oviposition site for a number of species (Silver 2008a).

On 23/04/14 and 24/04/14, 102 traps were attached to 51 trees which were selected as trap sites. These trees were located at Wybunbury Moss (Figure 23), Loynton Moss (Figure 25) and Jackson's Coppice and Marsh (Figure 24). Each tree had two traps attached; one trap ~30 cm from the ground and the second at ~140 cm from the ground. 50 % of the beakers were partly filled with water collected on site with the other 50 % left empty to allow rainwater to accumulate naturally; these starting conditions were randomly assigned at the outset of the study and remained in force for the duration. The intention was to explore whether there was any influence on the oviposition rate dependent upon the water source and the trap height.

Each trap was checked every 2 to 3 weeks when they were inspected for signs of oviposition and/or juvenile presence. The oviposition surface was removed from the water and inspected under a hand lens for eggs, and the surface of the water checked for egg rafts or individual eggs. The water was removed and checked for larvae, pupae and/or exuviae; following inspection the water was replaced facilitating maturation over time. Any such specimens were to be collected and returned to the insectary for identification and rearing on. After inspection, the water and hardboard were returned, unless the oviposition surface had eggs on it in which case it was returned to the laboratory for rearing on and replaced with a new one. These traps continued to be checked until 10th October 2014 when they were collected from all sites, giving a collection period of 5.5 months.

Figure 22 Top: An oviposition trap attached to a scots pine tree at Wybunbury Moss in April 2014. Bottom: a close-up of the attachment and the placement of the oviposition surface.

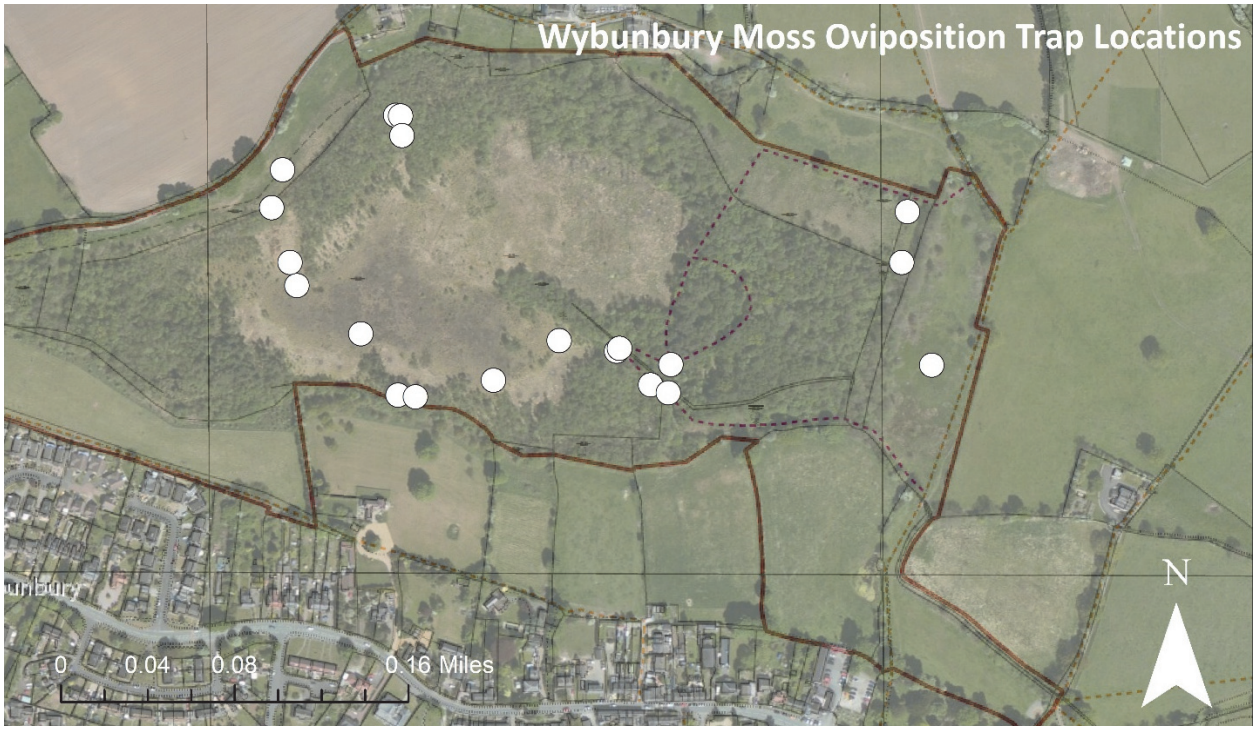


Figure 23 Locations of the oviposition traps placed at Wybunbury Moss. Each white circle represents the location of 2 traps. The red line indicates the extent of Wybunbury Moss NNR.

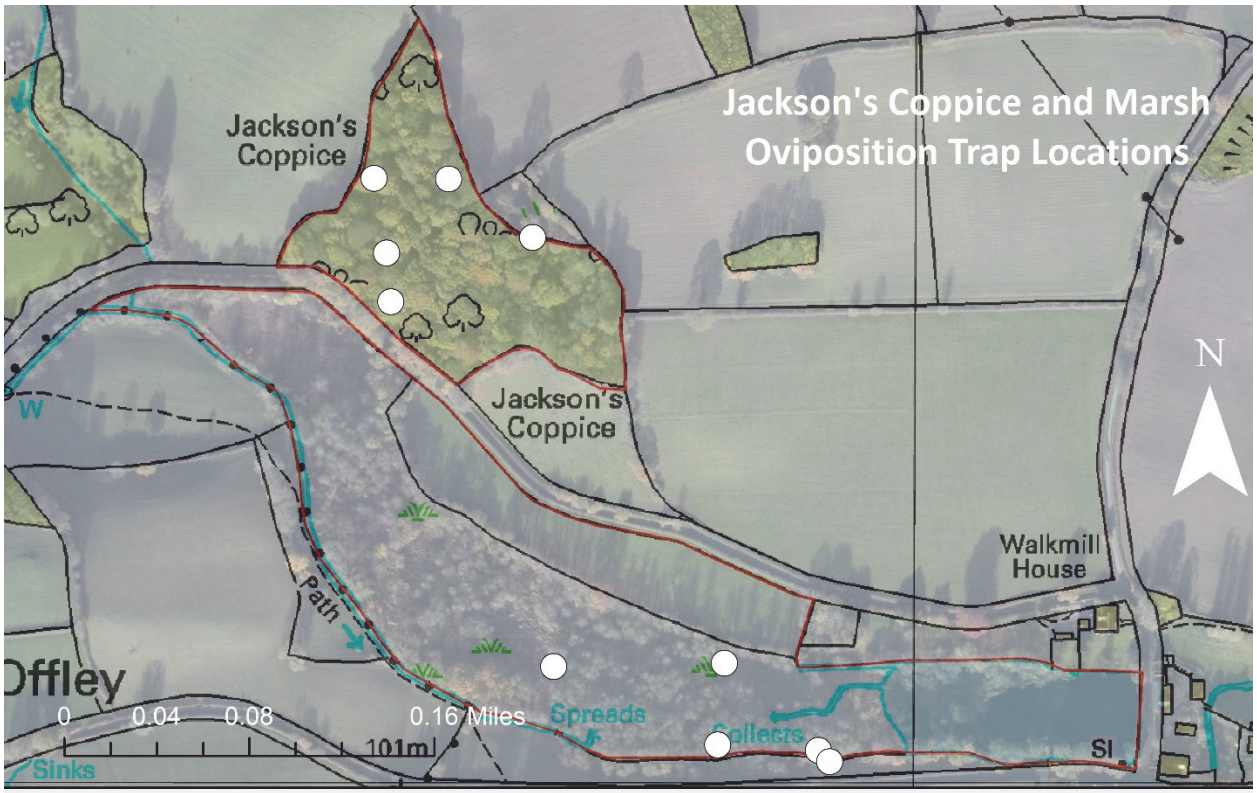


Figure 24 Locations of the oviposition traps placed at Jackson's Coppice and Marsh. Each white circle represents the location of 2 traps. The red line indicates the extent of Jackson's Coppice and Marsh NR.

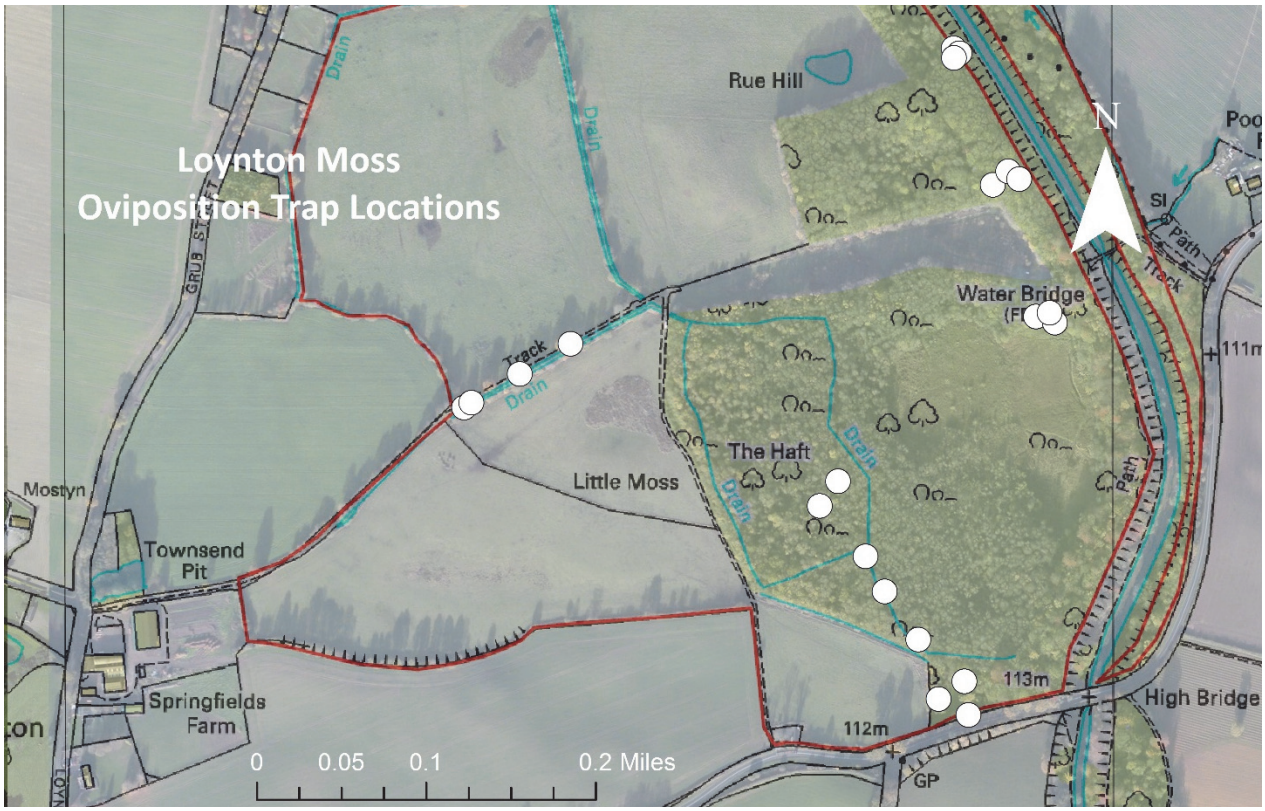
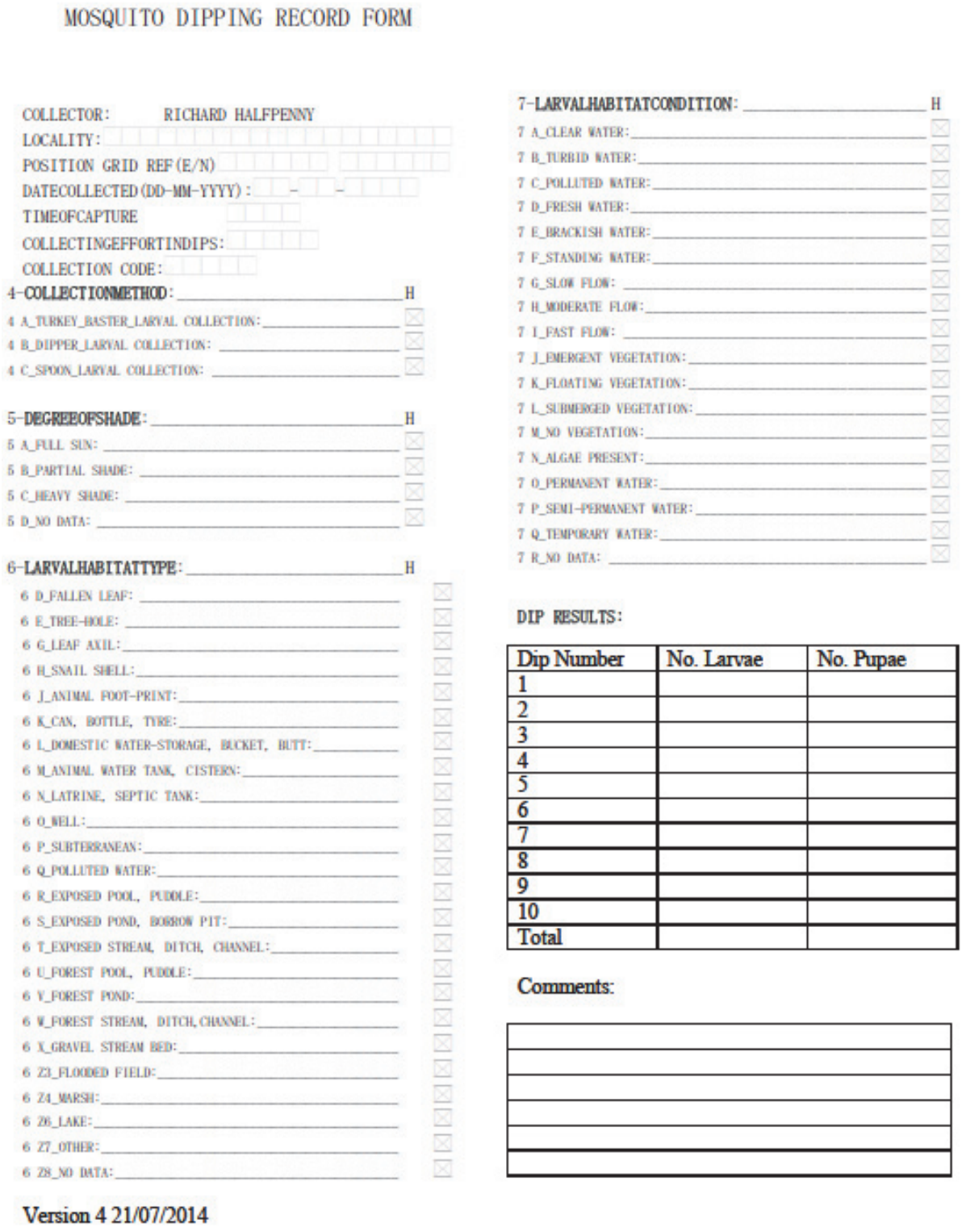


Figure 25 Locations of the oviposition traps placed at Loynton Moss. Each white circle represents the location of 2 traps. The red line indicates the extent of Loynton Moss NR.

3.2.1.6 Sampling of the larval population.

A dipper was used to sample larvae from water sources on each site. The dipper was made from a stainless-steel ladle (360 ml) attached to an extendable handle (minimum length 145 cm, maximum length 214 cm), facilitating safe access to water in ditches and drains.

Stratified random sampling was used to ensure that the different water bodies at each of the sampling sites were proportionally represented. Each dipping site was mapped, using a minimum 6 figure grid reference, photographed, and the key habitat data recorded on the dipping record form (Figure 26) based on that proposed by Gaffigan and Pecor (1997) .



MOSQUITO DIPPING RECORD FORM

COLLECTOR: RICHARD HALFPENNY
LOCALITY:
POSITION GRID REF(E/N)
DATECOLLECTED(DD-MM-YYYY):
TIMEOFCAPTURE
COLLECTINGEFFORTINDIPS:
COLLECTION CODE:

4-COLLECTIONMETHOD: H
4 A_TURKEY_BASTER_LARVAL COLLECTION:
4 B_DIPPER_LARVAL COLLECTION:
4 C_SPOON_LARVAL COLLECTION:

5-DEGREEOFSHADE: H
5 A_FULL SUN:
5 B_PARTIAL SHADE:
5 C_HEAVY SHADE:
5 D_NO DATA:

6-LARVALHABITATTYPE: H
6 D_FALLEN LEAF:
6 E_TREE-HOLE:
6 G_LEAF AXIL:
6 H_SNAIL SHELL:
6 J_ANIMAL FOOT-PRINT:
6 K_CAN, BOTTLE, TYRE:
6 L_DOMESTIC WATER-STORAGE, BUCKET, BUTT:
6 M_ANIMAL WATER TANK, CISTERN:
6 N_LATRINE, SEPTIC TANK:
6 O_WELL:
6 P_SUBTERRANEAN:
6 Q_POLLUTED WATER:
6 R_EXPOSED POOL, PUDDLE:
6 S_EXPOSED POND, BORROW PIT:
6 T_EXPOSED STREAM, DITCH, CHANNEL:
6 U_FOREST POOL, PUDDLE:
6 V_FOREST POND:
6 W_FOREST STREAM, DITCH,CHANNEL:
6 X_GRAVEL STREAM BED:
6 Z3_FLOODED FIELD:
6 Z4_MARSH:
6 Z6_LAKE:
6 Z7_OTHER:
6 Z8_NO DATA:

Version 4 21/07/2014

7-LARVALHABITATCONDITION: H
7 A_CLEAR WATER:
7 B_TURBID WATER:
7 C_POLLUTED WATER:
7 D_FRESH WATER:
7 E_BRACKISH WATER:
7 F_STANDING WATER:
7 G_SLOW FLOW:
7 H_MODERATE FLOW:
7 I_FAST FLOW:
7 J_EMERGENT VEGETATION:
7 K_FLOATING VEGETATION:
7 L_SUBMERGED VEGETATION:
7 M_NO VEGETATION:
7 N_ALGAE PRESENT:
7 O_PERMANENT WATER:
7 P_SEMI-PERMANENT WATER:
7 Q_TEMPORARY WATER:
7 R_NO DATA:

DIP RESULTS:

Dip Number	No. Larvae	No. Pupae
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
Total		

Comments:

Figure 26 Mosquito dipping form used for recording the habitat, location, time and juvenile mosquito dipping data

Where pools of water were too small for a dipper, e.g. trees holes or poached ground, a 30 ml nylon pipette (turkey baster) was used to collect specimens. 10 samples were then taken using the appropriate method, the dipper always being the preferred method if the water body was large enough to accommodate its use. Each sample was inspected using the naked eye, and all juvenile mosquitoes were collected using a disposable pipette, counted and placed in water tight specimen vials, containing water from the dipping location. These were then returned to the insectary for rearing on and identification.

To standardise the area available to be sampled at each sampling location the following rule was applied: if the water body was at a similar level to the surrounding ground level then the extendable pole was used at its shortest length, where the water was below the surrounding ground level, many field drains, for example, then the extendable pole would be utilised at its longest length. Any water, continuous with the target water body, which could be sampled without causing the operator to move their feet could be included in that location. All sampling was carried out by the same individual.

3.2.1.7 Rearing-on of wild mosquito collections

During this study, there were two circumstances which required the rearing of mosquitoes. These were, rearing-on of larvae collected as part of field collections where the principal concern was to ensure that life stage progressed to adulthood to facilitate species identification, in this case, the critical endpoint was the availability of a specimen, and as such there was no requirement to ensure that juvenile feeding was matched between individual mosquito specimens. The second circumstance was to rear-on specimens for use in behavioural assays, in this instance, the experimental design required specimens which were reared under identical feeding, lighting and temperature regimes to avoid unduly influencing feeding drive in the imagos (Gerberg 1970; Kiarie-Makara *et al.* 2016). All the mosquito stocks used in experiments were wild-type, meaning they were reared from egg rafts collected in the field. Therefore, establishment and maintenance of multigenerational species cultures were not required.

Where juveniles were retained for rearing on, for species confirmation, they were held in mesh topped vials (either 20 ml or 25 ml volume) half-filled with water collected from their

habitat at the time of sampling. In the case of late instars, additional food did not normally need to be added to these vials. For earlier instars, guinea pig food pellets (own brand guinea pig pellets, Pets at Home Ltd) were crushed using a pestle and mortar and passed through a 0.1 mm sieve before being fed sparingly, ~0.12 g l^{-1} to support development. Adults were allowed to emerge into the vial, from where they were collected for identification and storage.

Where *Culex* egg rafts were collected and reared for the purpose of generating adults for experimentation, plastic deli pots (450 ml) were filled with 300 ml aged tap water these were then topped with mesh. Tap water was aged for 2 days before use to allow off-gassing of chlorine were used. One egg raft was placed in each pot. Larvae were fed guinea pig food prepared as noted above at the daily rate of 0.12 g l^{-1} for 1st and 2nd instar larvae and 0.25 g l^{-1} for 3rd and 4th instars. Immediately prior to adult emergence, each container was placed into its own, mesh sided, insect rearing cage (32.5 cm x 32.5 cm x 32.5 cm) for emergence and identification of the adult males. Once identified, the cages were labelled with the species name, the unique collection code and the date of first imago eclosion.

3.2.1.8 Sampling of the adult population.

3.2.1.9 A peridomestic longitudinal study using light traps.

During 2014, CDC 512 mini light traps, equipped with "air-actuated gate systems" and LCS-2 photo switches (Hock 2016) were placed in two suburban back gardens in Porthill (39 GRW and 48 OA), Staffordshire. The traps used 6-volt rechargeable batteries (Yucel Y12-6L) rather than 4 D-cell batteries as supplied with the traps, to minimise costs and reduce waste. Two pairs of batteries were used in a 4 nights on and 4 nights off rotation to allow continuous 7-day operation. Traps were positioned with the fan at approximantely 1.5 m above the ground (Hutchinson *et al.* 2007).

The traps were set on 6th March 2014 and checked daily until 8th November 2014 for a total of 496 trap nights. All mosquito specimens were collected, labelled and stored at -20 °C until identification.

3.2.1.10 Wybunbury Moss (rural) study using light traps

During 2015, four CDC 512 mini light traps, configured as in the 2014 peridomestic longitudinal study, were used to sample the adult mosquito population at Wybunbury Moss. Traps were placed randomly across the 'no public access' portion of the site, and each sampling session ran for 4 nights. New trap locations were generated at each iteration. As with the peridomestic suburban study, light traps were positioned with the fan at approximantely 1.5 m above the ground (Hutchinson *et al.* 2007), although there was some limited variation of this height due to variation in anchoring positions at the random trap locations. The centre of the floating peat raft and reed marsh were excluded for safety reasons. Specimens were collected, labelled and stored at -20 °C until identification. A total of 12 such trapping iterations occurred for a total of 192 trap nights. Figure 27 shows examples of configuration and placement of the light traps ay Wybunbury Moss.

All traps were calibrated before use; such that the photo switches activated the fan motors within 1 minute of each other as the light level changed. This ensured similar sampling effort under similar sampling conditions. All specimens were collected and identified to species where possible*, using morphological identification keys (Snow 1990; Becker *et al.* 2010a).

* Females of *Culex pipiens s.s.* and *Culex torrentium* were morphologically indistinguishable using conventional methods; meaning that without access to molecular techniques, identification as *Culex pipiens s.l*. is typically applied. A potential solution to this problem using wing morphometric methods was applied and has been included in a separate section in the results.

Figure 27 Examples of light trap deployment at Wybunbury Moss. The top image is in the birch scrub encroaching onto the moss at the north of the site. The bottom image is in amongst the tall reeds at the edge of the safe area for sampling.

3.2.1.11 Sampling the adult population using Mosquito Magnet traps

To avoid reliance on one trapping mechanism, and to facilitate the utilisation of the CDC light traps elsewhere, Mosquito Magnet™ Executive (MM330) traps were used to collect adult mosquitoes in the Ashley, Stone, Kidsgrove and both Porthill (GRW and OA) back garden settings at the end of 2015 and through 2016.

Sampling using Mosquito Magnets™ was carried out from October 2015 until October 2016, with no sampling during the winter period: December 2015 through March 2016. During 2015 sampling was conducted for 7 days and nights per week every other week. This regime was altered for 2016 to be more efficient in terms of gas consumption. During 2016, for the early part of the sampling season, April to June, the traps were run every other week for 4 days and nights per week. During peak season, August through to October they were run every week for 4 days and nights per week to increase the resolution of the sampling. For 2015 there was a total of 56 trap nights and 2016 had a total of 276 trap night, yielding a combined total of 332 trap nights.

At the outset of sampling, traps were located in Ashley (Figure 29), Porthill (GRW)(Figure 32), Kidsgrove (Figure 28) and Stone (Figure 30). However, due to the structural failing of a wall directly adjacent to the trap location in Stone in July 2016, that trap location became unavailable on safety grounds, and consequently the trap needed to be removed; it was then placed in the second Porthill location (OA)(Figure 31); no trapping iterations were missed during this relocation. The Mosquito Magnet™ Executive traps that were used have some features to improve gas efficiency, some are user controlled whilst others are factory parameters. For example, the gas burn rate is set by the manufacturer as was assumed consistent between the traps; this was supported by the similar longevity of each gas cylinder used during the study. Another factory set feature is that the traps will turn off whenever temperatures are below 10 °C, automatically restarting once the temperature rises above this point. The main user adjustable setting on the Executive model relates to fuel saving, and this was utilised as follows: during sampling the traps were set up such that they ran for 16 hours per day, with a 4-hour period during the middle of the day and the middle of the night (fuel saving setting mode 4). These periods typically show less activity than dawn and dusk (Jaenson 1988; Yee and Foster 1992; Gary and Foster 2006) for mosquitoes and therefore give the best balance between fuel use and catch size.

Figure 28 Placement of the Mosquito Magnet™ Executive trap at Kidsgrove for 2015-2016. The trap was placed on gravel behind the greenhouse.

Figure 29 Mosquito Magnet™ Executive placement in the Ashley back garden. Top: close up showing sheds to the rear of the trap. Bottom: a wide view of the garden. A very small water feature is embedded in the raised bed ~ 5 m to the left of the trap.

Figure 30 Placement of the Mosquito Magnet™ trap in a back garden in Stone. This trap needed to be removed in July 2016 to allow the rebuilding of the adjacent wall. It was relocated to the second Porthill location (OA).

Figure 31 The placement of the Mosquito Magnet™ in the second Porthill back garden (OA). This placement replaced the Stone location after July 2016

Figure 32 Placement of the Mosquito Magnet™ at Porthill (GRW) at the base of a laurel bush.

Traps were placed in shaded locations where they were least likely to be disturbed during operation. Throughout the trapping periods, the gardens were under normal use, and no special rules for utilisation of the surrounding space were discussed with the property owners beyond those required for safety reasons.

Specimens were collected at the end of each trapping iteration by the removal of the trapping cage from the device, the door of which was then taped closed for transportation to the laboratory. Specimens were removed from the trapping cage using a battery-powered aspirator, prior to storage at -20 °C until identification.

3.2.1.12 Statistical analysis methods

Having collected various data regarding the local mosquito population, a range of statistical methods were applied to address the aims. Data were stored in an Access® database, spatial analyses and mapping were carried out using ARCGIS (Ver. 10.2.2) (ESRI 2014). All other statistical treatments were completed using R Statistical Software (R Development Core Team 2011).

Species occurrence was collated and reported for all specimens found during the study by all sampling methods. This comprehensive report includes all life stages and methods. These data were prepared ready for inclusion in the NBN Gateway dataset.

The data resulting from longitudinal adult mosquito trapping approaches, both light trapping and Mosquito Magnets™, were used to examine the temporal abundance of mosquitoes and to explore the relationship between weather and trapping rate. The weather data utilised in this study was collected at the Keele University Weather station, this weather station is located 4 miles from the centre of the city of Stoke on Trent. These data were visualised against time and mean temperature data.

Data collected using light trapping at Wybunbury Moss was analysed to test whether mosquitoes were evenly distributed across the sampled site, or whether they showed a clustered distribution. The Hot Spot analysis tool within ARCGIS was used to calculate the Getis-Ord Gi* statistic for each point (Anonymous n.d.). This statistic calculates the statistical significance of the number of mosquitoes collected at a single location and relates it to the nearby collection points. The local sum for a feature and its neighbours is then compared proportionally to the sum of all mosquito collections. If there is a difference, and that difference is too large to be caused by chance, then a statistically significant z-score is produced (Anonymous n.d.). The Getis-Ord Gi* statistic has the important advantage of neutralising the spatial distribution of the data points, meaning that the actual pattern of the data points will not bias the results (Getis and Ord 1992).The results of this analysis were displayed on a map of the site.

The relationship of juvenile mosquito presence or absence was modelled against some of the habitat features collected during sampling (using the mosquito dipping data across all sites. Only predictors from the “Degree of Shade” and “Larval Habitat Condition” were included in the modelling as there would be a significant interaction between these and those for “Larval Habitat Type”; for instance, a Tree Hole or a discarded bottle would be unlikely to hold emergent vegetation. Specifically, the relationship between vegetation and immature mosquitoes was investigated, as was the combined influence whether locations were under full sun, partial shade or full shade and observations of water flow.

For this modelling, the protocol proposed by Zuur *et al.* (2010) was applied to ensure that the correct model type was selected for these data. Having carried out the protocol, it was apparent that the data were not normally distributed, were overdispersed and zero-inflated. These data characteristics severely limited the number of potential models, and, following review, it was decided that either a Zero Inflated Negative Binomial Hurdle (ZINB) model or a Zero Adjusted Negative Binomial Hurdle (ZANB) model (Zuur *et al.* 2009) would be the most appropriate.

The ZINB and ZANB models differ in the way that they treat zeros within the data, and a logical decision needed to be made based on this and the model fit. Within mosquito dipping, specifically, zeros might be caused by:

1. Absence of mosquitoes due to the habitat being unsuitable
2. Absence of mosquitoes despite suitable habitat
3. Failure of the sampling method to collect mosquitoes that are present within the water body
4. Failure of the observer to notice, identify or record a mosquito which has been collected.

In a ZINB model, zeros caused by points 2, 3 and 4 are considered separately, from the Count model, and are instead used in the zero model. Zeros caused by point 1, which are particularly important for a habitat suitability model, are retained for inclusion in the Count model; this is the part of the model where zeros are calculated as being from point 1, or the count is > 0. In a ZANB model, zeros from all 4 points are treated by the Zero model, and only records where the count is > 0 are retained for inclusion in the Count model. The ZANB model makes no distinction between the different types of zeros (Zuur *et al.* 2009). Therefore, it was these 2 model types which were trialled with the data. Due to a better fit, in terms of reduced Akaike information criterion (AIC), and due to the biological interpretation of the likely source of the zeros, the ZANB models were chosen for inclusion. Following calculation of the initial saturated models, the models were optimised to the data using the stepAIC function from the MASS package.

3.2.1.13 Per night catch rate of CDC mini light traps: comparison between peridomestic and rural locations

Rural and peridomestic trapping data were compared to see if there was a significant difference in the mosquito capture rate in the different settings. This has several potential implications. The most obvious being that differences in capture rate could show differences in the prevalence of mosquitoes between rural and peridomestic suburban locations. It could also be associated with different levels of light pollution between rural and suburban settings affecting the conspicuousness of the bulb fitted to the light trap.

Because the data sets held different numbers of observations and ran over slightly different temporal windows, the peridomestic data was constrained so that it covered the same time period as the rural survey. A rolling mean was then applied to the peridomestic data such that each data point would be equivalent to the mean of the four surrounding points. Each rural data point was divided by 4 to make the value equal to the mean number of mosquitoes caught per night.

The rural dataset, therefore, consisted of 48 observations, each equal to the mean number of mosquitoes caught per trap night. The peridomestic dataset, after the treatment above, consisted of 298 observations, each equal to the mean number of mosquitoes per trap night. Welch's t-test was applied to the data, where a random sample of 48 observations was extracted from the peridomestic data set to compare to the 48 observations in the rural data set. This test was bootstrapped (10000 iterations), and the density plot of p-values was created to visualise the result.

3.2.1.14 General specimen handling and identification

All catches were sorted at the time of sampling whenever possible, and non-target organisms were replaced at the source of collection. Mosquito identification was conducted in laboratory conditions on the adult mosquitoes using the appropriate keys (Snow 1990; Becker *et al.* 2010a) and, subsequent to the development of the method, by shape based wing

morphometric analysis. All items that were taken onto study sites were removed completely following their use, leaving sites unaltered.

3.2.2 Results

3.2.2.1 Sampling using Oviposition traps

The 102 oviposition traps placed at Wybunbury Moss, Loynton Moss and Jackson's Coppice and Marsh during 2014 yielded no positive results. Despite being in situ from April 23rd/24th to October 10th, yielding >17000 trap nights, no evidence of mosquito activity was found in any of the traps over the whole duration of the study. Therefore, it was not possible to investigate either the effect of height nor presence or absence of organic material on oviposition rate. A selection of the hardboard oviposition sticks was treated to an additional series of drying, wetting and cooling events after the traps were removed (Silver 2008a), to ensure that Aedine eggs had not been overlooked in the field; these also resulted in no eggs being found.

3.2.2.2 Sampling the juvenile population

Larvae and pupae sampling, using a dipper or large pipette (where required), yielded larger numbers of mosquitoes than all other sampling methods and 2772 mosquitoes were sampled from Wybunbury Moss, Loynton Moss and Jackson's Coppice and Marsh. A broad range of habitat types was sampled, including, but not limited to, tree holes, land drainage channels (Figure 34), puddles, poached ground, animal drinking troughs (Figure 33) and discarded items such as buckets and a half barrel.

Aside from the prearranged sampling trips to the rural field sites other samples were collected when the opportunity presented itself. Samples were gathered from a single tyre awaiting refuse collection, and from tarpaulins laid on the ground during *Impatiens glandulifera* (Himalayan balsam) removal on the Staffordshire University grounds (Figure 33). Larval collections were also made at the suburban back garden locations when water had been seen to collect at those sites.

The summary of the juvenile collections can be seen in Table 8 and include the life stages and broad habitat types that they were collected from. Of the species that were regularly collected, few species seemed to have any strict associations, except *Dahliana geniculata* which was only found in tree holes. *Culiseta annulata*, which was common throughout many

of the sites tended to be found in naturally occurring water bodies, although eggs were collected from ornamental plant holding troughs at a pond retailer's premises.

Figure 33 Examples of manmade habitats within which mosquito juveniles were found. The top image is a puddle of water collected in a tarpaulin being used for collection, and the bottom a typical cattle water trough.

Figure 34 Not all naturally occurring habitats were utilised by mosquitoes during the sampling. The top image is one of the drainage channels at Wybunbury Moss - this held many mosquitoes. The bottom image is a drainage ditch at Loynton Moss from which no mosquitoes were sampled. Indeed, floating vegetation of this type seemed to have an adverse effect on the presence of mosquitoes.

Sampling location	Eastings/Northings (of site)	Species name	Life stages collected	Habitat, waterbody type/s
Porthill (suburban)	384744 - 348570	*Culex pipiens s.s.*	Eggs, larvae, pupae	Container (bucket) Paddling pool
Porthill (suburban)	384744 - 348570	*Culex torrentium*	Eggs, larvae, pupae	Container (bucket) Paddling pool
Staffordshire University (urban)	388276 - 345506	*Culex pipiens s.s.*	Larvae	Tyre
Staffordshire University (urban)	388729 - 345787	*Culex pipiens s.s.*	Larvae/pupae	Small puddle in tarpaulin
Staffordshire University (urban)	388729 – 345787	*Culex torrentium*	Larvae/pupae	Small puddle in tarpaulin
Wybunbury Moss (rural)	369888 – 350286	*Culiseta annulata*	Larvae/pupae	Groundwater pools
Wybunbury Moss (rural)	369888 – 350286	*Anopheles claviger*	Larvae/pupae	Groundwater pools
Wybunbury Moss (rural)	369888 – 350286	*Culex pipiens s.s.*	Larvae/pupae	Groundwater pools
Wybunbury Moss (rural)	369888 – 350286	*Culex torrentium*	Larvae/pupae	Groundwater pools
Loynton Moss (rural)	378824 – 324276	*Culex pipiens s.s.*	Larvae/pupae	Groundwater pools Natural pond Animal water troughs
Loynton Moss (rural	378824 – 324276	*Culex torrentium*	Larvae	Groundwater pools
Loynton Moss (rural	378824 – 324276	*Ochlerotatus punctor*	Pupae	Groundwater pools and puddles
Loynton Moss (rural	378824 – 324276	*Ochlerotatus annulipes*	Pupae	Groundwater pools
Loynton Moss (rural	378824 – 324276	*Culiseta annulata*	Larvae/pupae	Groundwater pools
Loynton Moss (rural	378824 – 324276	*Anopheles maculipennis*	Larvae	Natural pond
Kidsgrove (suburban)	383469 – 354591	*Culiseta annulata*	Eggs	Garden pond. Aquatic plant container in aquatic centre
Kidsgrove (suburban)	383469 – 354591	*Culex pipiens s.s.*	Eggs, larvae, pupae	Garden pond Aquatic plant container in aquatic centre
Kidsgrove (suburban)	383469 – 354591	*Culex torrentium*	Eggs, larvae, pupae	Garden pond Aquatic plant container in aquatic centre
Kidsgrove (suburban)	383469 – 354591	*Anopheles claviger*	Larvae/pupae	Garden pond
Jackson's Coppice and Marsh (rural)	378779 - 329997	*Dahliana geniculata*	Larvae/pupae	Tree holes
Jackson's Coppice and Marsh (rural)	378779 - 329997	*Anopheles claviger*	Larvae/pupae	Animal footprint/poached ground
Jackson's Coppice and Marsh (rural)	378779 - 329997	*Anopheles plumbeus*	Larvae	Tree holes
Jackson's Coppice and Marsh (rural)	378779 - 329997	*Culex pipiens s.s.*	Larvae/pupae	Tree holes, animal footprint/poached ground, groundwater pools
Ashley (rural)	376291 – 336083	*Culex pipiens s.s.*	Larvae/pupae	Container (bucket) Garden pond
Ashley (rural)	376291 – 336083	*Culex torrentium*	Larvae/pupae	Container (bucket) Garden pond

Table 8 Summary of the species collected as juveniles by sampling area and habitat/waterbody type. N.B. *Culex pipiens s.s.* was found at all locations. *Culex torrentium* was present at all sites except Jackson's Coppice and Marsh. Eastings and northings are for the wider site being sampled and do not represent the individual sampling points.

3.2.2.3 Modelling habitat preference of juvenile mosquitoes

A wide range of count values was returned at each dipping location (Figure 35). The high number of zeros and the overdispersed nature, where the variance is significantly greater than the mean, of the data, restricted the available statistical model choices somewhat.

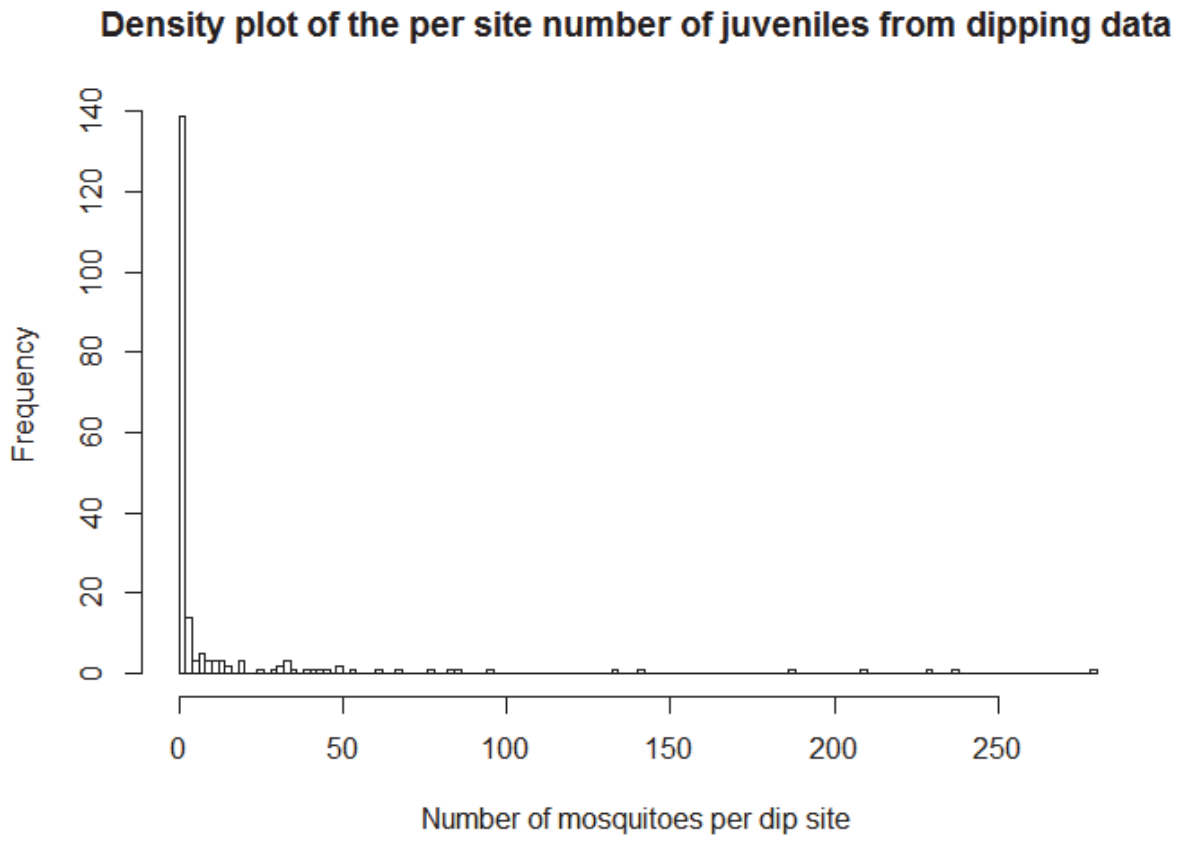


Figure 35 Shows the range of different numbers of mosquito juveniles collected at each dipping site and the frequency of those totals

The data were fitted to a zero adjusted negative binomial hurdle (ZANB) model (Zuur *et al.* 2009).

The **saturated model** for the relationship of juvenile mosquitoes to vegetation was:

- **Formula = Combined_total ~ Emergent_Vegetation * Floating_Vegetation * Submerged_Vegetation + Algae_Present * No_Vegetation | Emergent_Vegetation * Floating_Vegetation * Submerged_Vegetation + Algae_Present * No_Vegetation**

Where Combined_total is the number of juvenile mosquitoes sampled at that dipping site, the predictors relate to the vegetation check boxes on the mosquito dipping record form (Figure 26). Predictors before the | relate to the count part of the ZANB model and those after "|" relate to the zero-adjustment model.

The final **optimised model** for vegetation was:

- **formula** = **Combined_total ~ Emergent_Vegetation + Floating_Vegetation + Submerged_Vegetation + Algae_Present + No_Vegetation + Emergent_Vegetation : Floating_Vegetation | Emergent_Vegetation + Floating_Vegetation + Submerged_Vegetation + Algae_Present + No_Vegetation + Emergent_Vegetation : Floating_Vegetation**

The results of the model are reported in Table 9. The Zero model showed that none of the predictors had a significant influence on the presence or absence of mosquitoes for these data. The Count model, which describes the environmental influences in those cases where mosquitoes were found, showed that in water bodies with only single vegetation types there tended to be a negative impact on the number of mosquitoes collected by dipping. However, where both emergent and submerged vegetation was found, the number of mosquitoes might be expected to be 2.088 (+/- 0.968) times the mean number per sample.

Table 9 Results of the Zero Adjusted Negative Binomial Hurdle model for the relationship between juvenile mosquitoes and vegetation.

Model for the number and occurrence of juvenile mosquitoes in relationship to vegetation type

	Zero Adjusted Negative Binomial Hurdle Model
Count model: (Intercept)	5.122 (0.565)***
Count model: Emergent_Vegetation	-0.889 (0.548)
Count model: Floating_Vegetation	-2.022 (0.695)**
Count model: Submerged_Vegetation	-0.908 (0.395)*
Count model: Algae_Present	-1.732 (0.385)***
Count model: No_Vegetation	-1.698 (0.619)**
Count model: Emergent_Vegetation:Floating_Vegetation	2.088 (0.968)*
Count model: Log(theta)	-0.834 (0.304)**
Zero model: (Intercept)	-0.182 (0.537)
Zero model: Emergent_Vegetation	-0.355 (0.519)
Zero model: Floating_Vegetation	-0.532 (0.663)
Zero model: Submerged_Vegetation	0.289 (0.376)
Zero model: Algae_Present	0.522 (0.338)
Zero model: No_Vegetation	-0.679 (0.623)
Zero model: Emergent_Vegetation:Floating_Vegetation	-0.879 (0.862)
AIC	923.161
Log Likelihood	-446.581
Num. obs.	204

***p < 0.001, **p < 0.01, *p < 0.05

The saturated model for the relationship between juvenile mosquitoes and light levels and water flow was as follows:

- **Formula = Combined_total ~ Full_Sun + Partial_Shade + Heavy_Shade + Standing_Water + Slow_Flow | Full_Sun + Partial_Shade + Heavy_Shade + Standing_Water + Slow_Flow**

The final optimised model for levels and water flow was:

- **Formula = Combined_total ~ Full_Sun + Standing_Water | Full_Sun + Standing_Water**

The results of the model are reported in Table 10. The Zero model showed that where the water was static (standing water), this had a significant influence on the presence or absence of mosquitoes for these data; making it 1.786 (+/- 0.634) times more likely to find mosquitoes in standing water. The Count model, which describes the influence of the predictors on those cases where mosquitoes were found, showed that for water bodies subjected to full sunlight there was a significant ($p < 0.001$) negative influence on the number of mosquitoes collected by dipping, reducing the expected number of collected mosquitoes by 2.235 (+/- 0.63) times than the mean number collected.

Table 10 Results of the Zero Adjusted Negative Binomial Hurdle model for the relationship between juvenile mosquitoes and water flow and light levels.

Model for the number and occurrence of juvenile mosquitoes in relationship to water movement and light level

	Zero Adjusted Negative Binomial Hurdle Model
Count model: (Intercept)	2.403 (1.020)*
Count model: Full_Sun	-2.235 (0.630)***
Count model: Standing_Water	0.938 (1.019)
Count model: Log(theta)	-1.386 (0.426)**
Zero model: (Intercept)	-1.985 (0.615)**
Zero model: Full_Sun	-0.851 (0.440)
Zero model: Standing_Water	1.786 (0.634)**
AIC	924.819
Log Likelihood	-455.410
Num. obs.	204

***p < 0.001, **p < 0.01, *p < 0.05

3.2.2.4 Sampling the adult population

3.2.2.5 2014 Peridomestic longitudinal study using CDC light traps

Welch's t-test was conducted to compare the mosquito catch rate between the Porthill GRW (Mean = 0.093, SD = 0.52) and Porthill OA (Mean = 0.149, SD = 0.57). There was no significant difference between the mosquito catch rates at the 2 peridomestic sites (t = -1.1471, p = 0.2519).

Only two mosquito species were collected by the light traps in 2014, these being *Culex pipiens s.s.* and *Culex torrentium*. Traps at both sites collected specimens of each species with both species captured by the same trap on the same night on several occasions. The cumulative catch rate was plotted against mean high daily temperature (Figure 36) and mean low daily temperature (Figure 37). Collection rate reached its peak at the same time as the highest daily means were achieved for both metrics. Due to the low collection rate, and the lack of any significant difference between the collections, the data for both traps were combined for these plots (Figure 36, Figure 37).

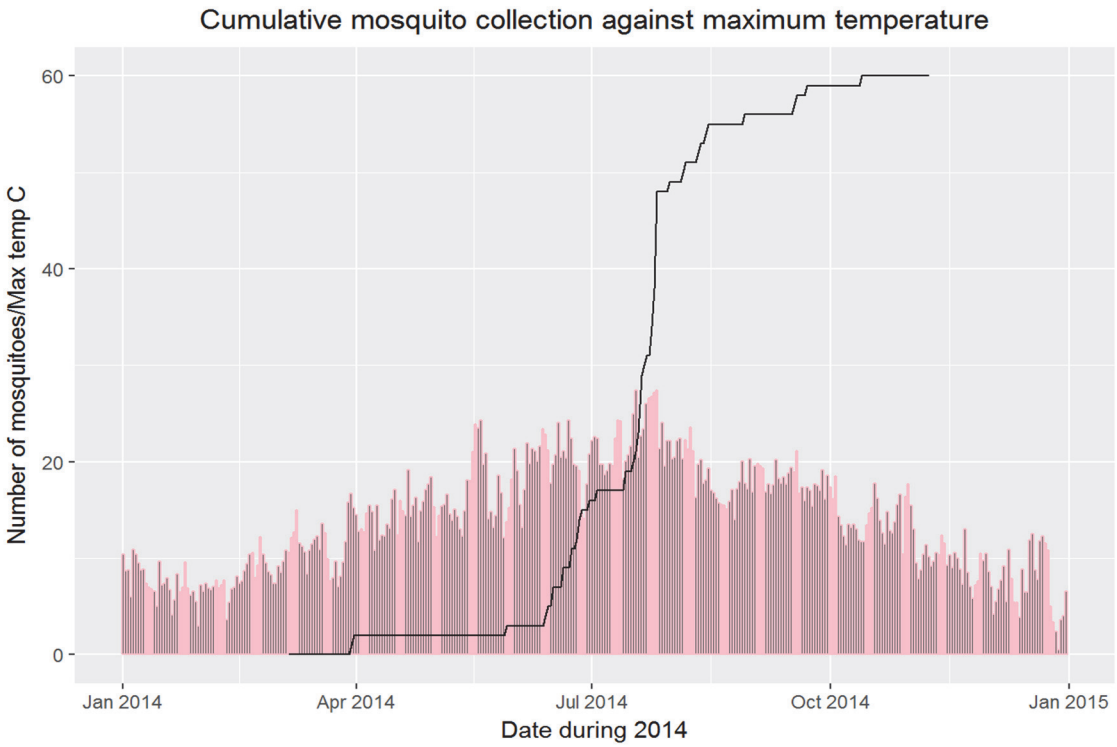


Figure 36 Total number of adult mosquitoes caught per night in CDC 512 mini light traps, plotted as a cumulative curve against the daily maximum temperature. These traps were located in 2 suburban back gardens.

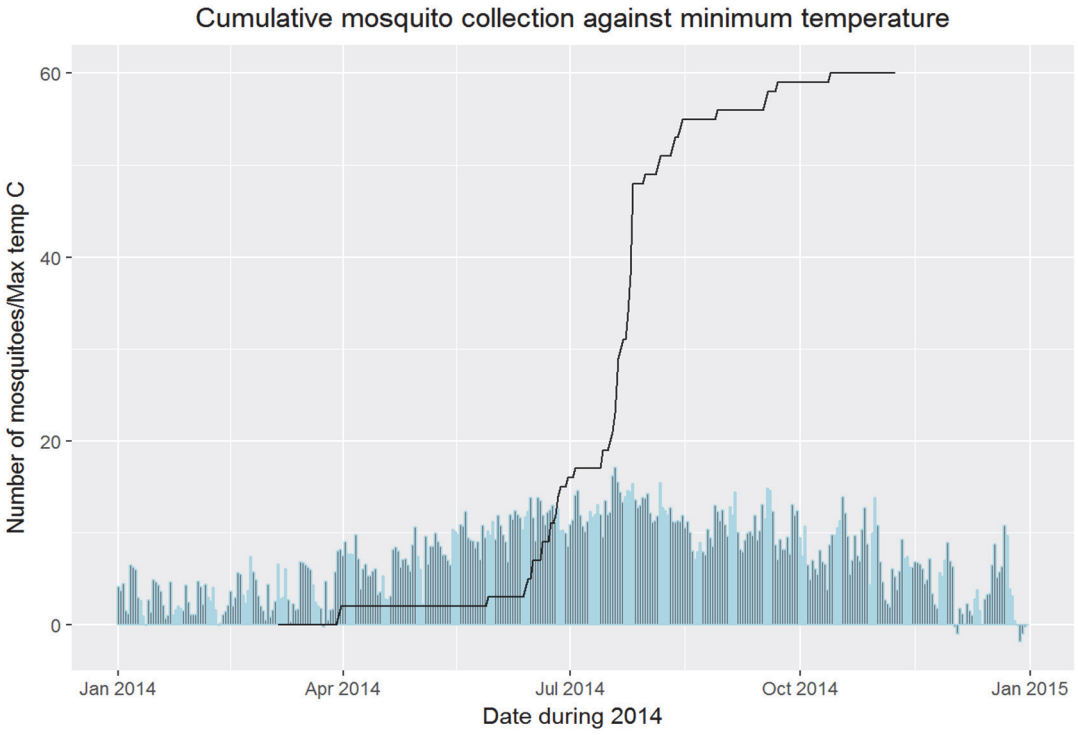


Figure 37 Total number of adult mosquitoes caught per night in CDC 512 mini light traps, plotted as a cumulative curve against the daily minimum temperature. These traps were located in 2 suburban back gardens.

3.2.2.6 2015 Wybunbury Moss (rural) study using CDC light traps

The species caught during this light trapping phase are displayed in Table 11, which also shows the number caught and the number of trapping iterations within which they were collected. Per trap data is presented in Figure 38, showing that the highest number of mosquitoes collected in one four-night trapping sessions was 5.

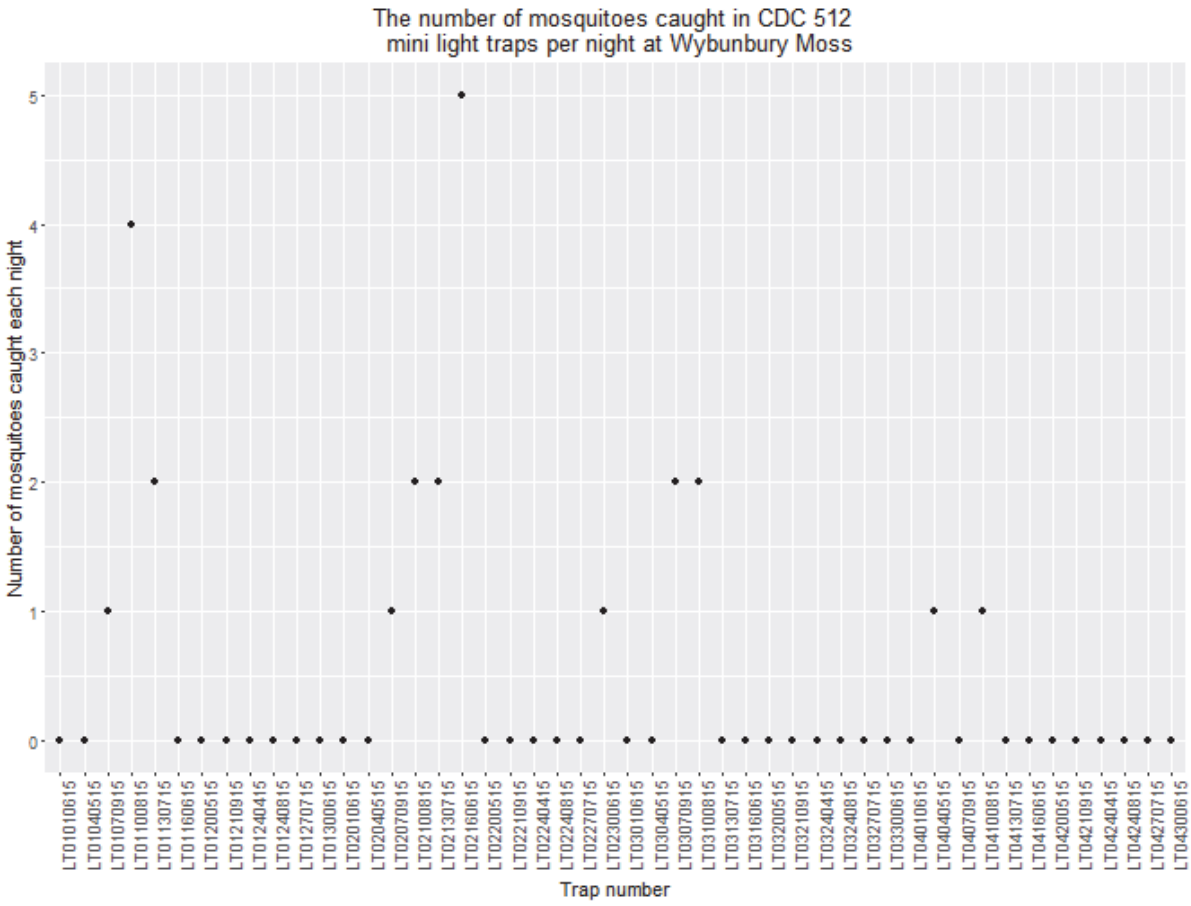


Figure 38 Number of mosquitoes caught per four-night light trapping session at Wybunbury Moss.

Table 11 Species sampled by light trapping at Wybunbury Moss in 2015. 48 trapping iterations of 4 nights duration (192 trap nights) were carried out for these samples. * Specimen too damaged to identify beyond genus.

Species	**Number of positive traps**	**Number of individuals**
Culex pipiens s.s.	3	4
Culiseta annulata	13	19
Culiseta fumipennis	1	1
Aedes spp. *	1	1

When considered spatially, there appeared to be clustering in the distribution of mosquito captures. Figure 39 shows the catch data and clustering analysis of the light trap collection data. Circle colour indicates the clustering score (GiZScore) and the concentric centre region communicates the level of significance. The two clustered regions, to the north and south-east, are associated very closely with the regions which held the highest numbers of larvae, suggesting limited movement away from the larval habitats by the adults.

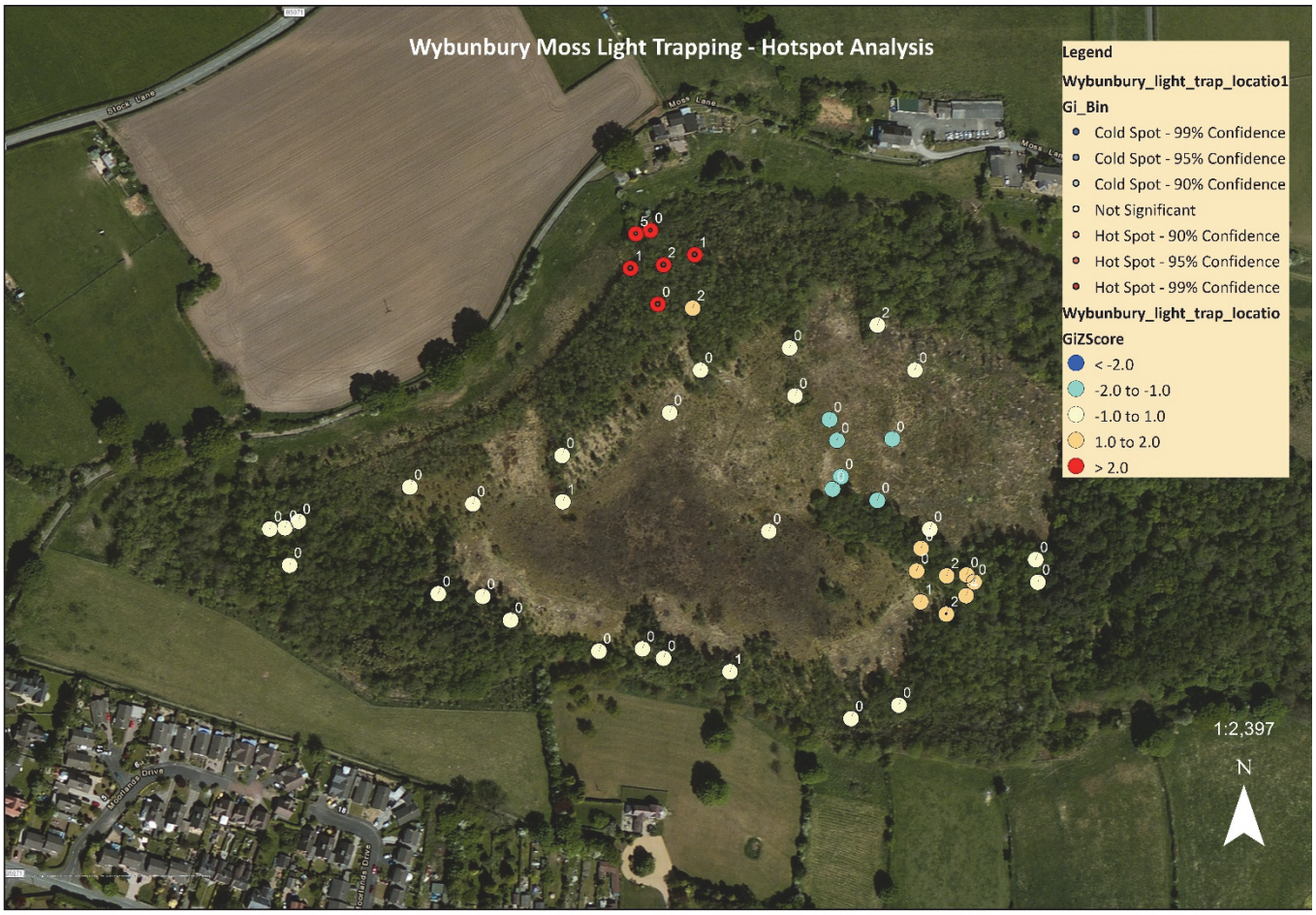


Figure 39 Light trap locations used during 2015 at Wybunbury Moss. Spatial relationships within the data have been modelled in ARCGIS 10.2.2 and show clear, significant clustering - red markers with centre circle.

3.2.2.7 Per night catch rate comparison between peridomestic and rural locations

The bootstrapped Welch's t-test provided a density plot (Figure 40) which clearly shows that there is no significant difference between the mosquito capture rate using CDC 512 mini light traps in rural or peridomestic settings. The red line on the graph shows where p = 0.05 occurs, and it is clear that for most samples these data the p-value is greater than 0.05. Therefore, there was no evidence to suggest that there was increased efficacy of these traps in either rural or suburban setting, nor did this provide support for a difference in abundance of mosquitoes within these locales.

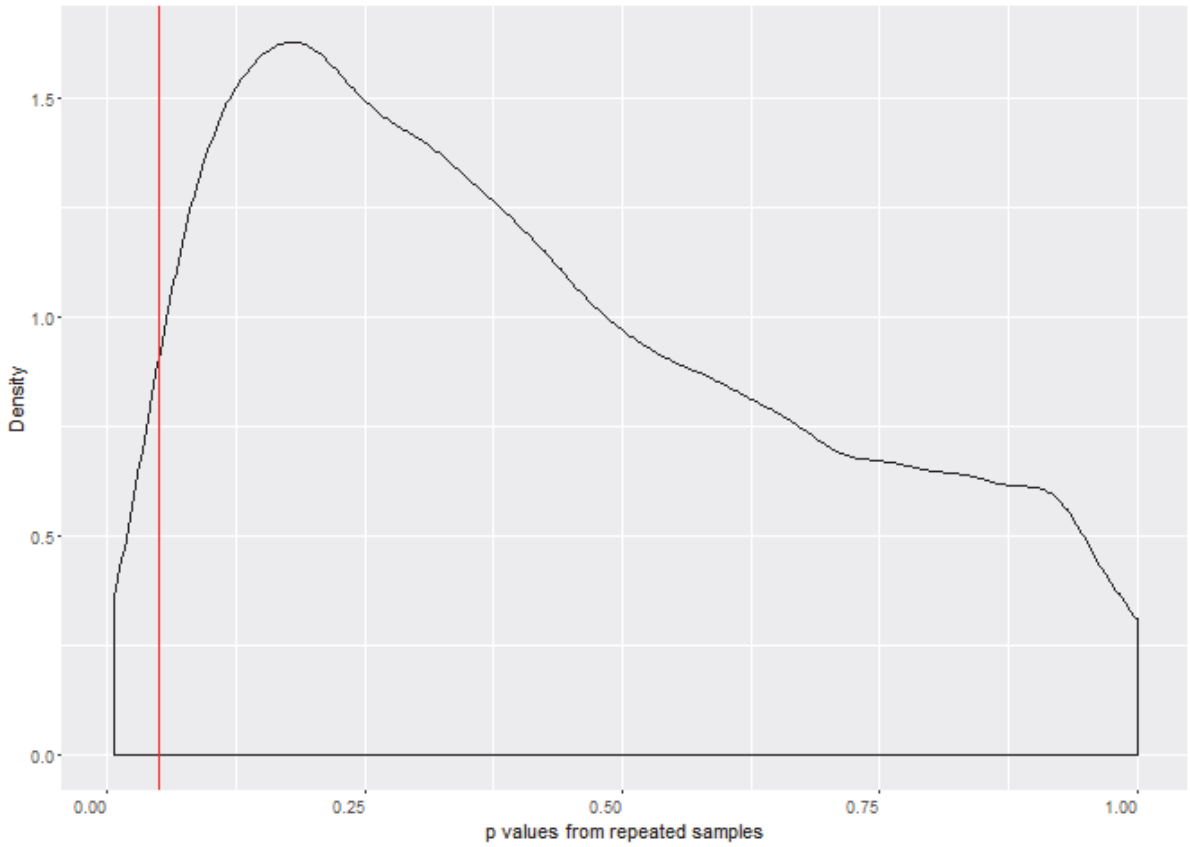


Figure 40 Distribution of p- values for bootstrapped Welch's t-test comparisons between rural and peridomestic catch rates using CDC 512 mini light traps.

3.2.2.8 2015/2016 Peridomestic study using Mosquito Magnet™ traps

Following the collection and identification of specimens collected using Mosquito Magnet™ traps during 2015 (56 trap nights) and 2016 (276 trap nights), the collection data were collated and are presented in Table 12.

As expected, female catch rates were very high for these traps, males were represented in the catch, with male *Culex pipiens s.s.* and *Culex torrentium* both represented at several of the locations. The predominance of female specimens was generally of little consequence in terms of identification. A key exception to this was for female *Culex pipiens* s.l. which are not identifiable to species by traditional morphological means. Therefore, the wing morphometric method devised in chapter three was applied to these individuals, the results of which are displayed in Table 13.

The cumulative collection data were plotted for each trap and were plotted as a time series (Figure 41) which clearly showed that differences between locations were apparent. Allied with the different species richness between sites, these cumulative curves suggest large localised differences between mosquito abundance and species richness.

N.B. No data for period 19/08/2016 to 23/08/2016 as the traps exhausted their gas in this time, with none available to immediately replace it; therefore, the data are excluded.

Table 12 Summary of the mosquitoes sampled using Mosquito MagnetTM traps in private back gardens over 332 trap nights. *Morphological identification of females, either *Culex pipiens s.s.* or *Culex torrentium*.

Trap location	**Species**	**Number of positive traps**	**Number of individuals**
Ashley (rural)	*Culex pipiens s.s.*	2	3
Ashley (rural)	*Culex pipiens* s.l. *	5	12
Ashley (rural)	*Culiseta annulata*	1	1
Kidsgrove (suburban)	*Culex pipiens s.s.*	8	30
Kidsgrove (suburban)	*Culex torrentium*	3	3
Kidsgrove (suburban)	*Culex pipiens* s.l. *	12	50
Kidsgrove (suburban)	*Culiseta annulata*	7	13
Kidsgrove (suburban)	*Ochlerotatus rusticus*	1	4
Kidsgrove (suburban)	*Ochlerotatus annulipes*	1	1
Kidsgrove (suburban)	*Anopheles claviger*	3	4
Porthill GRW (suburban)	*Culex pipiens s.s.*	5	8
Porthill GRW (suburban)	*Culex torrentium*	1	1
Porthill GRW (suburban)	*Culex pipiens* s.l. *	11	15
Porthill GRW (suburban)	*Culiseta annulata*	6	7
Porthill GRW (suburban)	*Anopheles plumbeus*	1	2
Porthill GRW (suburban)	*Coquillettidia richiardii*	1	1
Porthill OA (suburban)	*Culex pipiens s.s.*	3	3
Porthill OA (suburban)	*Culex pipiens* s.l. *	4	9
Porthill OA (suburban)	*Culiseta annulata*	7	11
Porthill OA (suburban)	*Anopheles plumbeus*	1	1
Porthill OA (suburban)	*Coquillettidia richiardii*	2	2
Porthill OA (suburban)	*Anopheles claviger*	1	2
Stone (suburban)	*Anopheles maculipennis*	1	1
Stone (suburban)	*Culex pipiens* s.l. *	1	2

Table 13 Further identification of female *Culex pipiens* s.l. specimens using wing morphometric analysis. 91.8 % of specimens were identified to species with the agreement of species identification for both wings. 8.2 % had an identification difference between wings and were, therefore, not allocated to species. Four individual's wings were too damaged to use.

Location	**Identified as *Culex pipiens s.s.***	**Identified as *Culex torrentium***	**Not identified to species – *Culex pipiens* s.l.**	**Specimens not tested using wing morphometrics**
Ashley (rural)	10	3	0	0
Kidsgrove (suburban)	39	5	5	1
Porthill GRW (suburban)	8	6	1	0
Porthill OA (suburban)	5	1	0	3
Stone (suburban)	1	0	1	0

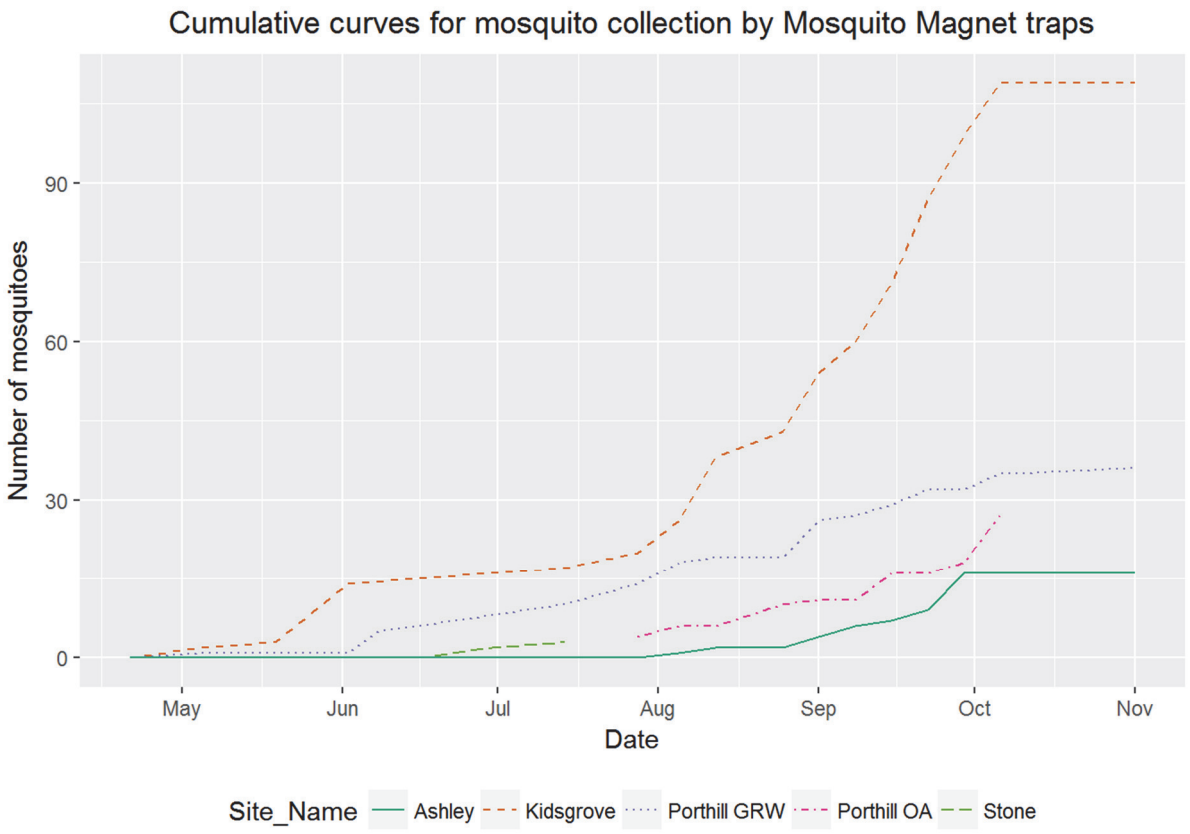


Figure 41 Cumulative capture curves for adult mosquitoes sampled using Mosquito Magnet™ traps during a sampling period.

3.2.3 Discussion

This study has achieved its aims of gathering new mosquito species distribution data to add to the national species record, an evaluation of the effect of vegetation on oviposition site selection in rural settings and mosquito temporal abundance.

With regard to the local species record, in their 2015 review of passive mosquito surveillance in the EU Kampen *et al.* reported the presence of three mosquito species and only three individual records in the region studied in this thesis (Kampen *et al.* 2015), and the NBN showed a few more records (Figure 13) most of the records were very old. This study facilitates the addition of 325 juvenile mosquito sampling sessions where mosquitoes were found, plus the adult collection data from light and Mosquito Magnet™ trapping, all of which contain habitat data and high-resolution spatial data (six-figure coordinates for eastings and northings). These data constitute a significant contribution to the field, and upon their inclusion within the public mosquito record will help generate more accurate predictive models of mosquito distribution and habitat preference (Golding 2013; Golding *et al.* 2015).

Sampling in 2014 using oviposition traps proved to be highly disappointing. These oviposition traps were designed to be attractive to both container dwelling mosquitoes and those which favoured tree holes, but no mosquitoes were recorded using them. However, the traps were not devoid of life by any means, being visited by slugs, spiders, woodlice, and often oviposited by drone flies (*Eristalis Spp.)* as evidenced by the presence of rat-tailed maggots. Whether there was something fundamentally wrong with the construction of the oviposition traps employed in this study, such as chemicals leaching from the plastic cups into the water, it was hard to say; although this was unlikely due to the vessels being used being catering quality products. It was more likely that the traps were less desirable oviposition sites than the naturally occurring ones, for example pools, puddles and tree holes found on the site. Typically, the tree holes sampled during this study dried up for at least some time during the year, were often partially filled with decomposing leaves, the olfactory products of which could be smelled by the operator when sampling them. These observations could be critical to the trap's lack of success. Periods of desiccation to facilitate oviposition by members of genera such as *Ochlerotatus* and *Dahliana*, which deposit their eggs above the waterline in anticipation of rising water (Snow 1990; Crans 2004; Khatchikian *et al.* 2010), olfactory cues

(Mokany and Shine 2003; Allan *et al.* 2005) and visual cues (Bidlingmayer 1975; Allan *et al.* 1987) would be more closely associated with other natural oviposition sites available to the mosquitoes rather than the artificial traps. However, there is potential for continued use of these traps, but they might be better suited to an urban setting where container breeding is the norm. Indeed, in many cases the oviposition traps have been used primarily in an urban setting where the mosquito species assemblage may be more attuned to opportunistic container oviposition rather than being highly selective towards specific criteria (Kampen *et al.* 2015; Vaux and Medlock 2015; Medlock *et al.* 2017).

Alternatively, changing the container from plastic to cut sections of bamboo could yield better results, indeed traps using cut bamboo have been successfully used to sample oviposition of *Dahliana geniculata* in the south of England (Yates 1979). In Yates' study eggs were collected from June through to November in 1972 (28 oviposition traps) and 1973 (54 oviposition traps), indicating a 19 week oviposition season for *Dahliana geniculata*. During 1972, 13889 mosquito eggs were collected, and 6332 collected in 1973 (Yates 1979). Synthetic oviposition traps made from beer cans, with similar dimensions and volume to the cups used here, trialled by (Loor and DeFoliart (1969) in Wisconsin, USA, demonstrated the viability of non-organic containers and highlighted the importance of dark coloration and the presence of organic debris for the attractiveness of these oviposition traps. When using dark coloured cans, organic debris or a combination of both, 18 of 21 sampling iterations were positive for oviposition activity, with as many as 1237 eggs being collected in a single week (Loor and DeFoliart 1969). As all the containers used in this study were black in colour and 50 % of the traps were filled with water containing organic matter from the trap location, and all of the sampling sites were shown by other means to have resident mosquito populations, it was surprising that there was no evidence of their utilisation by mosquitoes. A possible further modification would be to remove the hardboard oviposition surface in favour of a cloth one which has been shown to be very effective (Lenhart *et al.* 2005).

These traps, or variations thereon, have been shown to work internationally, particularly for Aedine mosquitoes (Fay and Eliason 1966; Carrieri *et al.* 2011). They are also currently being used in the South East of England by Public Health England (PHE) to survey for invasive *Aedes* species (Kampen *et al.* 2015). These PHE traps also yielded negative results with no

oviposition recorded during their first seasons (Vaux and Medlock 2015), although more recently invasive *Aedes albopictus* eggs have been collected in one of them (Medlock *et al.* 2017).

As expected, sampling the larval/pupal population using a dipper proved to be a highly efficient method in terms of mosquito species sampled. It also has the advantage over trapping methods in terms of low financial cost. The approach to dipping taken in this study was to carry out detailed surveys at the sites requiring multiple visits to locations over an extended period of time. This methodology allowed the collection of species distribution data plus data relating to temporal variation helping generate a better understanding of mosquito ecology at the sites sampled, and was, therefore, the correct approach for this study. An alternative strategy might have been to make single visits to a wider variety of locations, thereby creating a mosquito snapshot for each site. Whilst this would certainly have facilitated the collection of a more spatially diverse dataset, making single visits to sites would fail to record a number of species which were actually present. For example, some univoltine (those having one brood of offspring per year) species have a relatively short period of time within which to complete their development from egg to oviposition of the next generation(Snow 1990) prior to summer diapause (Clements 1999) and overwintering as eggs.

Modelling of habitat association to juvenile mosquito catch data was successfully carried out and fitted the data quite well. As with all models, it is only representative of the data included in the calculation, and so when considering application as a predictive tool this should be borne in mind, however the results of the model do describe the sampling experience quite well. The high adverse effect of habitats with only floating vegetation might be explained by sampling of land drains which were heavily covered with *Lemna spp.* (duckweed) which never resulted in positive collections of mosquitoes. This relationship is interesting as it runs counter to some of the literature (Golding *et al.* 2015), although it is important to understand that the models presented here do not include data related to the co-occurrence of predators in each habitat, which could be a driver of mosquito absence in these mature land drains.

When sampling the adult population, the CDC 512 mini light traps proved to be suitable devices, but, perhaps, are not as effective in the UK as in other locations globally due to the

reduced tendency of British mosquitoes to enter dwellings (Silver 2008a). For this study, the light traps were used with no bait other than the standard fitted lamp. Often CO_2 is added as an additional bait has been shown to bias collections (Acuff 1976; Muturi *et al.* 2007; Hesson, Ignell, *et al.* 2015) and reduce the diversity of the sample (Silver 2008a). Augmentation using a floral volatile organic chemical (VOC) based bait might help significantly increase the catch of male specimens, and potentially increase the targeted nature of the trap. Mosquitoes were always outnumbered in light trap catches by non-target invertebrates, including microlepidoptera, Chironomidae and Ceratopogonidae. Occasionally the catch also included more robust flies, such as *Lucilia sericata* (Meigen).

Initially, it was thought that the poor performance of light traps in 2014 might have due to light pollution in the suburban setting, especially since the traps are only equipped with a small 6-volt bulb. If this were the case, then it would have been expected that there would be an increase in the number of mosquitoes caught using these traps in a rural setting where there would be significantly reduced light pollution, thereby increasing the relative brightness of the bulb in the trap. This did not turn out to be the case, with differences between the two settings shown to be statistically non-significant. This calculation did assume that populations of mosquitoes which might be attracted to the traps at each location were equal, which may or may not be the case. This comparison between the traps' effectiveness was based on the analysis of data collected for another purpose and the sampling effort applied to each setting was not balanced. Therefore, caution must be applied when considering the outcome of this analysis.

Light trapping collected a very low number of species, with only two species collected in suburban settings and four species collected at Wybunbury Moss. Subsequent use of Mosquito Magnet™ traps, by contrast, collected more mosquito species and collected fewer bycatch species. Mosquito Magnet™ traps have over recent years become increasingly important tools for the collection of high quality mosquito data and have been used in many studies (Hutchinson 2004; Deblauwe *et al.* 2014; Medlock and Vaux 2015b; Vaux *et al.* 2015), and have been shown to outperform CDC light traps (Hoel *et al.* 2009; Sant'Ana *et al.* 2014) as seen in this study. However, they do have drawbacks associated with their use. They are expensive when compared to light traps, and because of this there may be a barrier to

purchasing the equipment, and there may be increased security concerns when deploying them onto public land, indeed, for this study that was deemed too great a risk and restricted their use to suburban back garden settings. Running costs are also much higher than for light traps as they require propane gas to operate. However, the increased effectiveness of these tools as sampling devices outweighs their potential limitations.

That *Culex torrentium* were collected throughout the study as all life stages, and at almost all sites runs counter to the latest literature for British *Culex* mosquitoes, which suggests that *Culex torrentium* are a rural species, not featuring in urban collections at all due to their being more cold adjusted and therefore repelled by increased temperature caused by the urban heat island effect (Townroe and Callaghan 2014). For their research, Townroe and Callaghan (2014) studied mosquitoes in and around Reading, which is ~ 130 miles south-east of the region studied here. Presumably, this distance is sufficient to cause a reduction in temperature such that *Culex torrentium* range is not restricted by higher temperatures in urban settings at this latitude; the yearly average maximum temperature (1981-2000) for the nearest weather station to this study (Keele) was 12.6 °C whereas for Reading University's weather station it was 14.5 °C ('Keele climate' 2017).

Aside from the confirmatory records of new species found, perhaps the most important finding of the field surveys carried out has been that *Culex torrentium* is found commonly in urban, suburban and rural settings in the area around Stoke on Trent. It is a good example of the unreliability of the English species record for mosquitoes in general and highlights how the species record in its current state provides poor information regarding the distribution of major potential enzootic disease vector. Only by continued research, complete specimen identification and sharing of findings can the species record be improved such that it be a representative description of the distribution of mosquitoes in general and of potential disease vectors in particular.

4 Investigation into wing morphometry as a means of species discrimination of *Culex pipiens s.s.* and *Culex torrentium*

4.1.1 Introduction

As previously discussed in chapter 1, British *Culex pipiens* complex mosquitoes are competent vectors of a variety of mosquito borne organisms (MOBOs), with the potential to cause disease in humans and animals. Within the complex, the different species each have different vector competences (Fonseca *et al.* 2004; Hesson, Verner-Carlsson, *et al.* 2015) and apparent differences in geographical distribution (Townroe and Callaghan 2014; Werblow *et al.* 2014). It would therefore be very useful to be able to easily and cheaply discriminate between the species within the complex, to facilitate the collection of ecological and behavioural data that will help to increase understanding at a species level.

Culex pipiens L. and *Culex torrentium* Martini are sympatric sibling species which are morphologically cryptic. As imago, only the males of each species can be discriminated morphologically. This involves dissection of the male terminalia to examine the characteristic differences of the structure of the dorsal arm of the aedeagus, and ventral arm of the paraproct (Figure 42) are visible under low power, ~40x magnification, light microscopy (Snow 1990; Becker *et al.* 2010b); this is a well-established method and has been shown to be genuinely discriminatory across many studies (Service 1968; Dahl 1988), and were found to be 100% consistent with identification by molecular means throughout this study (Chapter 2). In the larvae both sexes can be identified to species, when in the 4th instar stage only, by comparing the branching of setae (spiny hairs) on abdominal segments III-V and the presence of branching of the saddle setae (1-X) (Becker *et al.* 2010a), although previous studies of British *Culex* mosquitoes did not find the branching of saddle setae to be discriminatory (Service 1968).

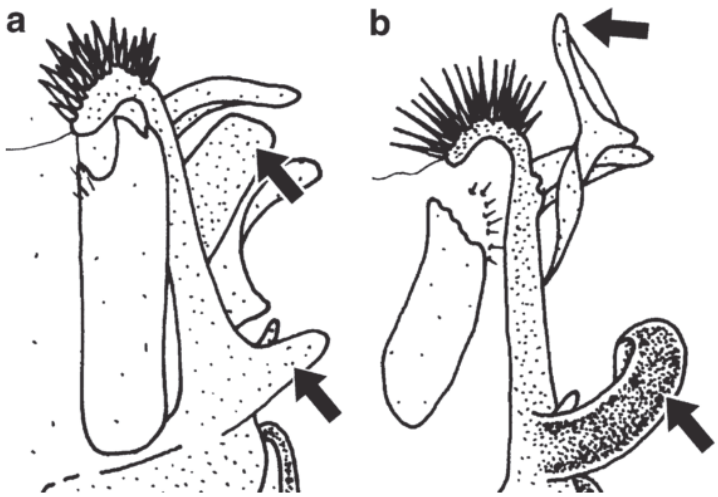


Figure 42 Differences in the terminalia of *Culex pipiens s.s.* and *Culex torrentium.* (a) *Culex pipiens s.s.* showing the truncated apex of the dorsal arm of the aedeagus, and weakly developed ventral arm of the paraproct and *Culex torrentium* (b) showing the twisted and pointed apex of the dorsal arm of the aedeagus, and strongly sclerotised ventral arm of the paraproct. Diagram adapted from Becker *et al.*, 2010a

Traditional morphological identification of the female imago of these species is not reliable (Service 1968; Jupp 1979; Onyeka 1983) as it relies on the presence of pre-alar scales on *Culex torrentium*; these scales are very easily dislodged and are therefore often missing, and have been observed as not present even in pristine *Culex torrentium* specimens (Service 1968). Consequently, female imago identification needs to be carried out using molecular methods, for which there are existing DNA primers (Bahnck and Fonseca 2006; Rudolf *et al.* 2013). Although these molecular methods work well, they might not always be available to entomologists, require expensive equipment and can be time-consuming. Studies have, therefore, been ongoing to find and demonstrate the consistency of a simple method of species discrimination, that does not demand access to equipment and apparatus beyond the reach of most amateur entomologists. The important role played by expert amateur entomologists (Morris 1987; Hopkins and Freckleton 2002) in supplying species distribution data to recording centres which then inform resources such as the National Biodiversity Atlas (NBN 2017) should not be overlooked (Pearson and Shetterly 2006).

4.1.1.1 Introduction to wing morphometric identification

Wing vein measurements have been successfully demonstrated in various areas of entomology as species discrimination tools such as with screw worm flies (Hall *et al.* 2014) and honey bees (Meixner *et al.* 2013; Oleksa and Tofilski 2015). Other interesting applications, which underline the potential power of wing morphology to convey genetic differences include the identification of Africanised honey bees (Francoy *et al.* 2008) and gender identification in bees (Francoy *et al.* 2009).

Wing morphology is seen to be well conserved within species (Garcia-bellido and Celis 1992; Birdsall *et al.* 2000) and a rich source of variation between species. Wing size and shape vary between individuals of the same species in response to several factors associated with larval development, these include temperature (Debat *et al.* 2003), nutritional availability (Garcia-bellido and Celis 1992) and larval density (Alto *et al.* 2012). Bilateral differences in wing morphology in individuals can also exist and often these are related to environmental stress (Parsons 1990); typically these take the form of fluctuating asymmetry (FA), which are small random deviations in paired bilateral traits (Hosken *et al.* 2000) normally distributed R-L values about the mean (Parsons 1990), and are often related to temperature and larval diet. Therefore, these asymmetries are often related to species populations near to their range boundary (Costa *et al.* 2015) or in regions subject to habitat disturbance (Lens *et al.* 1999). Importantly, the magnitude of the changes caused by fluctuating asymmetry is less than that generated by species. Two other forms of asymmetry are directional asymmetry (DA) where the asymmetry favours one side over the other in a population, and so the average deviation from symmetry is other than zero, and antisymmetry (AS) where one side is usually greater than the other, but the position of the larger side varies randomly in a population, generating a bimodal distribution (Parsons 1990; Costa *et al.* 2015). AS and DA are not related to developmental factors in the same way as FA and do not necessarily indicate that developmental instability is occurring due to physiological stress (Palmer and Strobeck 2003).

Culicine mosquito wing venation consists of mainly straight veins with several cross-linking veins; the arrangement and the standard vein nomeclature is shown in Figure 43 (Becker *et al.* 2010a). To date the features of interest for discrimination between *Culex pipiens s.s.* and

Culex torrentium have been related to the length of vein R2+3 (Börstler *et al.* 2014), and or the ratio of the lengths of R2+3:R2 or R2+3:R3 ((Mohrig (1969) in Börstler *et al.* (2014)).

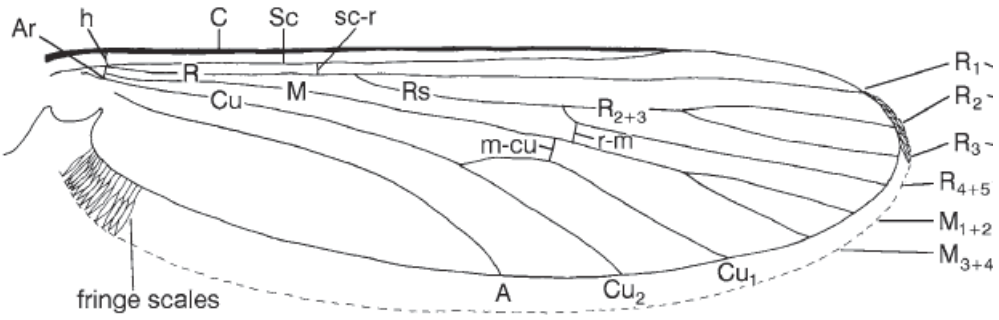


Figure 43 Wing vein nomenclature as applied to Culicine mosquitoes. C = costa, Sc = Subcosta, R = radius, Rs = radial sector, M = media, Cu = cubitus, A = anal, Ar = arculus, h = humeral. After Becker *et al*. (2010b).

Traditionally, wing features used for species identification have been based simple measures of vein lengths or the ratio of two wing veins. Typically, measurements are taken with a rule under a suitable magnification, and then calculations performed to determine which species' range the result falls into. For example, within the *Willistoni*, a group of sibling *Drosophila*, a number of different wing vein indices were diagnostic of species (Burla *et al.* 1949). However, there are many cases, particularly with closely related species, where there might not be a suitable univariate (single vein length, or section length) or bivariate (ratio of two vein lengths) means to discriminate between species. Until relatively recently, analysis of wing morphology would subsequently be ruled out as a method for species discrimination. Increasingly however, Shape Analysis (Adams *et al.* 2004), also referred to 'new morphometrics' (Rohlf and Marcus 1993) has been used as a multivariate method to find sufficient information to discriminate between species where traditional methods have failed. Shape analysis is a hugely powerful, multivariate, approach to describing the morphology of any object. In this context shape has a specific definition:

> *Shape is all the geometrical information that remains when location, scale and rotational effects are filtered out from an object* (Stegmann and Gomez 2002).

Figure 44 shows an example of four copies of the same shape, whilst they are located, rotated and scaled differently, they are still the same shape. The removal of scale is particularly important when applied to mosquito identification as it effectively removes the allometric growth effect of differential larval feeding introduced earlier. The removal of location and rotation data is useful as it makes the orientation of specimens for image collection much simpler, obviating the need to accurately match positions of specimens.

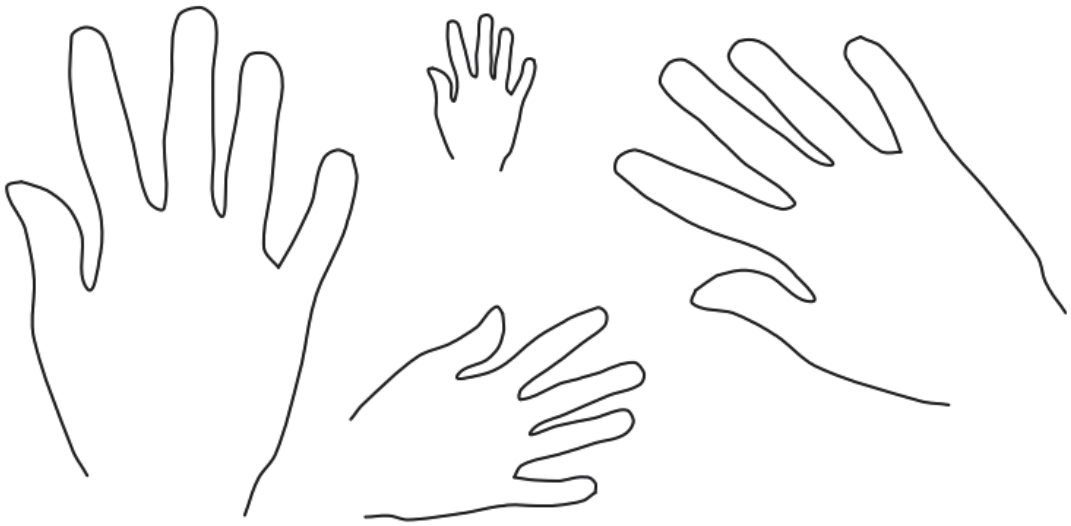


Figure 44 Once rotation, scale and location values are removed, these four shapes would be identical when analysed under the definition of a shape under shape analysis. after Stegmann & Gomez, 2002.

When the mathematics of shape analysis are applied to biological morphology comparisons, this becomes Geometric Morphometrics (GM). From early exploratory and method development studies (Rohlf and Archie 1984; Rohlf 1993) through to cutting edge applications (Sumruayphol *et al.* 2016; Wilke *et al.* 2016), mosquitoes, such is their importance for study, have always been candidates for new methods of investigation and species discrimination. Indeed, wing morphometric identification using modern methods has been used to successfully discriminate between mosquitoes of different genera (Wilke *et al.* 2016) and also to identify species within the same genus (Börstler *et al.* 2014; Sumruayphol *et al.* 2016).

An alternative approach to computational identification based on morphological shape characteristics is the application of artificial neural network (ANN) methods. A technique from

the field of computational intelligence, ANN is less rigid than traditional statistical techniques and can be applied to data where a proper analytical or statistical model can be found (Marcondes and Borges 2000). In a study by Lorenz, Sergio and Suesdek (2015), based on mosquito wing shape characters, it was demonstrated to be a powerful tool, showing increased species identification accuracy when compared to results from GM discriminant analyses when applied to 17 mosquito species from three different genera. The application of ANN to sibling and/or cryptic species requires further investigation (Lorenz *et al.* 2015), this however, is beyond the scope of this study.

Other shape comparison studies in fields outside of mosquito ecology (Oleksa and Tofilski 2015; Van Cann *et al.* 2015) have resulted in some exciting approaches and the development of new techniques for processing shape data (Jombart 2008; Jombart and Devillard 2010; Adams and Otárola-Castillo 2013), particularly when using software solutions such as R Statistical Software (R Core Team 2015).

Although easy to overlook, the ease of capture and storage of high quality digital imagery has also had complimentary effect on the continued development of these approaches. The ability to store images, effectively without limit if cloud storage resources are utilised, provides exciting opportunities for the creation of wing image databases for interrogation for identification purposes.

4.1.1.2 Wing morphometrics using shape analysis

For the purposes of this document, landmark-based Shape Analysis will be discussed as a two-dimensional method where insect wings are the target of the analysis. It can, however, due to its pure math origins, be applied to the analysis of any two or three-dimensional shape or structure where the specifics of that shape need to be interpreted.

Landmarks are those points of interest on the wing which will form the apices of the shape being analysed. Insect wings are well suited to this landmark-based approach as the intersection of wing veins with each other, or the wing edge is very easy to plot precisely, and as they are fixed anatomical features, they are related to species ontogeny (Comstock and Needham 1898). Landmarks are selected based on criteria: previous publications (to ease comparison), apparent differences between species and developmental or functional

reasons. Landmarks must sample sufficient morphological data that if the species can be discriminated, they will be; collecting too few landmarks at this point might result in a failure to find discriminatory variables that do exist (Zelditch *et al.* 2004). A balance must be struck however, between the effort required to collect the data and function, if a tool is too complex it might not be used.

In most shape analysis methods referred to in this thesis, more landmarks are collected than are used in the final analysis (Wilke *et al.* 2016). This is because at the time of landmark collection, the final variables diagnostic for species may not yet be known (Zelditch *et al.* 2004) and will be identified by dimension reducing tools such as Principle Components Analysis (PCA). Indeed, dimension reduction will reveal the existence of potential univariate or bivariate measures of discrimination when analysing new species or applying this approach to a species combination for the first time.

4.1.1.3 Tools

Numerous software tools now exist, facilitating the process of shape analysis, tools such as Collecting Landmarks for Identification and Characterization (CLIC) software (Dujardin 2016) and the R Statistical Software (R Core Team 2015) additional packages Geomorph (Adams and Otárola-Castillo 2013) and Adegenet (Jombart 2008) take on the mathematical burden of analysing shape data leaving the focus on developing the best approach to answering specific biological questions, and to developing workflows which simplify and foster future exploration.

4.1.1.4 Collection of landmarks

There are many approaches to landmark collection from images. Some software solutions are written specifically for this purpose, such as MOME-CLIC (Dujardin 2016), TpsDig2 (http://life.bio.sunysb.edu/morph/morphmet/tpsdig2w32.exe) and the Geomorph package for R (Adams and Otárola-Castillo 2013); other software has functions which allow for the landmark collection which is incidental to its original purpose, ImageJ (https://imagej.nih.gov/ij/) and Adobe Photoshop for example.

The key attribute for the collection software is that it is capable of accurately recording the coordinates of the landmarks and can output these coordinates in a manner that facilitates

analysis; typically, in the form of a text file. As data need to be accurately collected it is important that the input image is of sufficiently high quality that the operator can distinguish the morphological features being used.

4.1.1.5 Processing landmark data

The shape of the wing is described by the arrangement and position of the landmarks collected. These landmarks in their raw form contain information other than that needed for analysis. These extraneous data are related to rotation, size and position (Zelditch *et al.* 2004; Dujardin 2011). Whilst at first these seem necessary, they may be artefacts generated outside the actual shape of the wing. For example, if there is any difference in the magnification used to capture the image, even if slight, then this would be included in the analysis. Similarly, these features may be biologically irrelevant to species discrimination; as wing size is correlated to the body weight of the imago (Packer and Corbet 1989), which is strongly influenced by preimaginal feeding (Alto *et al.* 2012) as well as species, causing a broad range of overlapping values. The rotation will result from the difficulty inherent in perfectly aligning the wing on the slide and the slide with the camera. Therefore, these variables need to be removed using Procrustes superimposition on a consensus configuration before shape analysis can be performed (Zelditch *et al.* 2004; Dujardin 2011).

There are multiple approaches to the Procrustes superimposition (Zelditch *et al.* 2004), but Generalised Procrustes Analysis (GPA), also referred to as Generalised least squares Procrustes superimposition (GLS) (Rohlf 1999) but in this document, hereafter, GPA, is currently the most commonly used (Dujardin 2011). GPA calculates a consensus shape from within the available sample data and then transforms specimen's shape coordinate data relative to the consensus; the differences between the residual coordinates of each specimen and the residual coordinates of the consensus configuration are the 'Procrustes residuals' (Dujardin 2011). Because the post GPA shape data is relative to the consensus shape if new specimens are added to the data the GPA must be carried out again to generate a new consensus and relative shapes (Rohlf 1999). GPA results in a transformed dataset with a curved, non-Euclidean geometry, meaning that a further transformation needs to be applied to change the data to a Euclidean geometry such that the data may be studied using classical statistical tools (Dujardin 2011). This is achieved by projecting the GPA shape data into a linear

tangent space, yielding Kendall's tangent space coordinates (Rohlf 1999). In the linear tangent space, the distance between specimens is approximately the same as the Procrustes distances between the original pairs of landmark configurations for that individual (Adams *et al.* 2004).

4.1.1.6 Sensitivity and specificity in the discriminant wing morphometric context

The terms sensitivity and specificity are applied to diagnostic tests to report the effectiveness of the test. They can be defined as (modified from (Lalkhen and McCluskey 2008)):

- Sensitivity is the power of the test to correctly identify true positives i.e. the proportion of individuals that have the state being tested and are reported correctly by the test as having that state.
- Specificity is the test's ability to correctly report true negatives i.e. the proportion of individuals that do not have the state being tested and are reported correctly by the test as not having that state.

Often applied to diagnostic testing of disease in medical settings they help in the diagnosis of illness by testing whether an individual has a disease or not. Some tests are described as having a high sensitivity and so seek to identify whether an individual has the disease state. Alternatively, a test that has high specificity can be applied; in this case seeking to demonstrate that the individual does not have a disease state.

Both situations are somewhat different to the challenge of the wing morphometric challenge presented here. This is because in the medical testing examples used above the tests identify either the presence or absence of the disease state but offer no guidance beyond that; the next stage in diagnosis requires the clinician to make a judgement as to which disease to test for next. This is not the case in the analyses of wing morphometrics using a discriminatory approach as conducted here. Because all the specimens tested have been previously identified using reliable morphological methods as being either *Culex pipiens s.s.* or *Culex torrentium*, meaning that in all cases specimens can only be one of these species, an individual sensitive result for *Culex pipiens s.s.* must also be a specific result for *Culex torrentium.*

For the reasons outlined above, within the field of species discrimination using wing morphometrics it is challenging to apply the terms sensitivity and specificity to the results due to the inherent perspective associated with these terms. In a two species test a correct

classification against the true species identity is sensitive in one perspective and specific in the other perspective. Due to this, and in agreement with other investigations in this field (Lorenz *et al.* 2012, 2017; Börstler *et al.* 2014; Wilke *et al.* 2016), the reporting of the discriminatory power of the wing morphometric analyses presented here are in terms of the accuracy of the classification given by the approach against the true classification established using molecular means. The second metric reported in the results section is the precision, that being a measure of the consistency of the result provided by the model being tested when applied repeatedly to the dataset.

4.1.2 Assessment of the application of Börstler's *et al.* (2014) wing morphometric identifcation method to British *Culex* mosquitoes

To assess the applicability of wing morphometric shape analysis as a method for the diagnosis of species a study was conducted based upon the methods and findings of Börstler *et al.* (2014). The study also aimed to investigate any drawbacks of the methods proposed by Börstler *et al*. (2014) and consider areas of the process which might be simplified for increased ease of use.

Börstler *et al.* (2014) tested univariate, bivariate and multivariate methods for species identification of female *Culex pipiens* and *Culex torrentium*. They used the open source software package MOME CLIC to map and analyse the distribution of wing vein junctions to create a species-specific wing "shape" template for multivariate analyses using R Statistical software (R Core Team 2015). For their study, they used *Culex pipiens s.s.* and *Culex torrentium* mosquitoes collected between 2008 and 2012, predominantly from a range of habitats and at varying life stages. Specimens were collected from various sites across Germany, some of which had multiple collections at the same time, therefore their specimens will have included some reproductively isolated groups (either temporally or geographically) and other which may not be isolated (Slatkin 1987). The classification accuracy of the discriminant model within the training data using the multivariate approach was > 97% accurate. Using their findings, they went on to devise a manual method of measurements. Their method was very successful and gave a high proportion of correctly identified test individuals with over 91 % accuracy. Inter-species variation was sufficiently high to allow Börstler *et al*. 2014 to use a simple ratio of the length of wing vein R2+3:R3 to correctly identify the species of over 90 % of their specimens. However, the results from Börstler *et al.*(2014) run counter to those of Service (1968), who focussed on British *Culex pipiens s.s.* and *Culex torrentium* specimens, and a study of Russian *Culex pipiens s.s.* and *Culex torrentium* mosquitoes (Fedorova and Shaikevich 2007), neither of which were able to find reliable methods using the wing vein index R2+3/R2 or R3. Neither Service (1968) nor Fedorova & Shaikevich (2007) carried out a multivariate analysis of wing shape. Therefore, confirmation of the applicability of this method needed to be carried out using specimens from the UK, specifically from North Staffordshire and environs, before it could be used to identify specimens collected in the field for the wider ecological study presented within this thesis.

Whether British *Culex pipiens s.s.* and *Culex torrentium* are more like British *Culex* from 1968, or present-day Germany needed to be determined through new observations and analysis. Only then can the use of wing morphometric features, to discriminate between these species, be adopted or rejected.

4.1.2.1 Method

For this study, the method was closely based upon that of Börstler *et al.* (2014) to ascertain whether their approach would have similar discriminatory power described in that paper, using German *Culex pipiens s.s.* and *Culex torrentium*, when dealing with British specimens.

All mosquitoes were collected from the field, in the North Staffordshire region of England throughout 2014, as larvae and eggs prior to rearing on in the insectary, as detailed previously (3.2.1.7). Specimens did not include any laboratory cultured mosquitoes; they should, therefore, be considered morphologically representative of the wild populations sampled. 25 male *Culex pipiens s.s* and 25 male *Culex torrentium* were used as type specimens to train the algorithm. It should be noted that this is a deviation from the method used by Börstler *et al.* (2014), where they trained their algorithm with 42 *Culex torrentium* and 40 *Culex pipiens s.s.* specimens; the reduction in training specimens was due to availability of suitable specimens during this initial investigation. Each individual was identified morphologically by dissection of the genitalia, with reference to keys (Snow 1990; Becker *et al.* 2010b). The left wings of each mosquito were removed and placed on temporary dry slide mounts before being viewed using a 40x magnification dissecting microscope. Images were then captured using a 12-megapixel digital camera viewing through the microscope eyepiece.

Images were then processed using the COO module of the Collecting Landmarks for Identification and Characterization (CLIC) software package (Dujardin 2016). The user interface presents the user with tools, such as zoom, landmark editing including deletion and replacement allowing the user to review all landmark positions prior to saving the coordinate data to a text file. Thirteen landmark features were plotted at selected wing vein intersections (Figure 45). The specific order, and therefore subsequent numbering, of the landmarks, must be replicated on each wing being processed; failure to do so would effectively generate a different shape for the wing resulting in gross inaccuracies due to not transforming and comparing like with like (Rohlf 1993; Zelditch *et al.* 2004; Börstler *et al.* 2014).

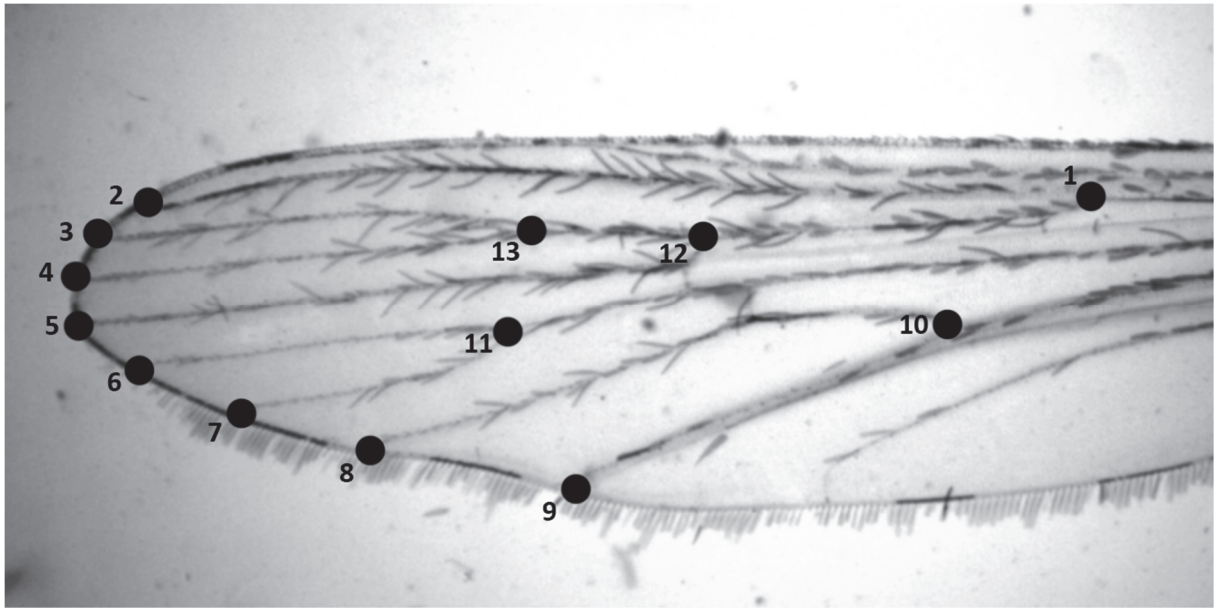


Figure 45 A *Culex* sp. mosquito wing image showing the 13 specific wing landmark locations are plotted using the MOME-CLIC interface

Once all 50 training specimens were digitised and landmarked the resultant text file was imported into the MOG package of CLIC for shape data preparation. The user interface for MOG is quite straightforward and leads the user through the steps of shape analysis. One challenge in learning to use this software is that the various supporting texts and reports frequently change language between English, French and Spanish. As previously discussed, the collated raw landmark data needs to be transformed to remove data related to location, scaling and rotation, the data transformation of which is shown in Figure 46.

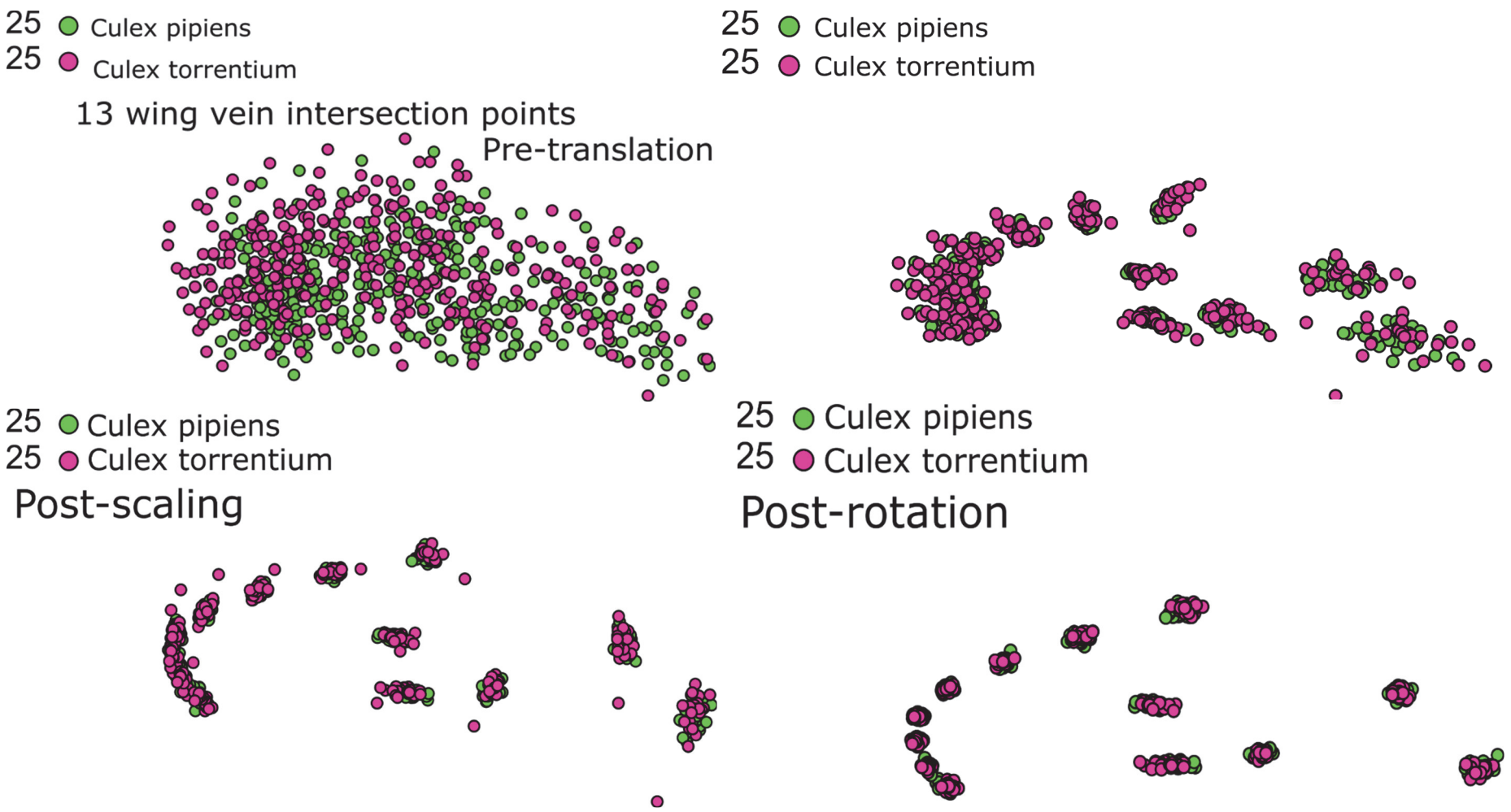


Figure 46 *Culex pipiens s.s.* and *Culex torrentium* wing landmark features from the training data in the four stages of transformation during Procrustes superimposition. Top left: data in their raw form. Top right: data post translation. Bottom left: the data post scale removal. Bottom right: data with variation from rotation removed. As can be seen, following the final step of the Procrustes superimposition the pattern of landmark collection is readily visible.

4.1.2.1.1 Identification of unknown specimens

Following preparation and analysis of training specimen landmarks, the software was ready to receive unknown specimens for discriminatory processing, which is loaded following the cues in the MOG package. To test the effectiveness of the software in discriminating between *Culex pipiens s.s. and Culex torrentium*, "known unknown" specimens were used. These were mosquitoes that had been identified in the same manner as the training specimens. The known unknowns did not include any of the specimens used for training and had their landmarks plotted after the training data was collected, in a separate processing session; replicating the typical scenario for an applied setting. For this pilot study testing, all known unknowns were male *Culex torrentium* (n = 10). Identification of unknown specimens was calculated using two different algorithms built into the MOG package, one based on the Procrustes distance and one on Mahalanobis distance.

4.1.2.2 Results

The training data confirmed that the two species groups had morphological differences which could be used to assign unknown specimens to groups. Figure 47 shows the grouping of principal components 1 and 2 within the training data. There was incomplete separation of the species groups when plotted against the first two principal components, this however

does not include the influence of the remaining components which are available to calculations related to identification of unknown mosquito specimens (Table 14).

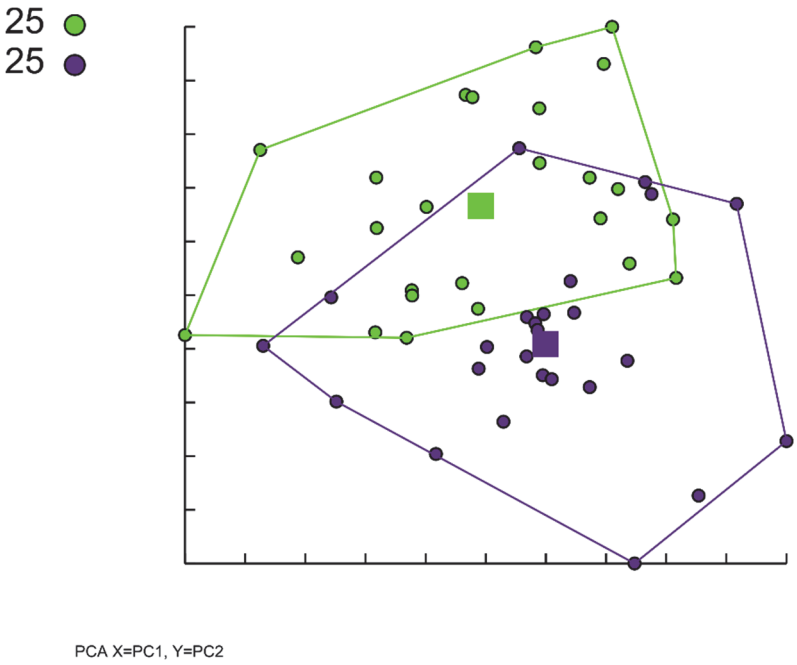


Figure 47 Principal component analysis of the wing morphometric landmarks for the training data. It is important to note that the discriminatory analysis uses all the features for discrimination and not just the principal component. The dots represent individual specimens and the squares represent the species centroids.

Table 14 Principal components that contributed to the difference between the two species groups and the magnitude of that contribution.

PC No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Contribution	0.31	0.26	0.10	0.08	0.06	0.04	0.04	0.03	0.02	0.02	0.01	0.01	0.01	0.01

Procrustes classification showed a 90% accuracy when identifying the *Culex torrentium* known unknowns, whereas the Mahalanobis classification was only 50% accurate (Table 15).

Table 15 Results from the wing morphometric identification process. All specimens in this test were *Culex torrentium*. The Procrustes analysis was 90% accurate, compared with 50% for the Mahalanobis classification approach.

Unknown specimen number All *Culex torrentium*	Procrustes Identification	Mahalanobis Identification
1	*Culex torrentium*	*Culex pipiens*
2	*Culex torrentium*	*Culex pipiens*
3	*Culex torrentium*	*Culex pipiens*
4	*Culex torrentium*	*Culex torrentium*
5	*Culex torrentium*	*Culex torrentium*
6	*Culex pipiens*	*Culex pipiens*
7	*Culex torrentium*	*Culex torrentium*
8	*Culex torrentium*	*Culex torrentium*
9	*Culex torrentium*	*Culex pipiens*
10	*Culex torrentium*	*Culex torrentium*

4.1.2.2.1 Can the wing vein R2+3/R3 index be used to discriminate between *Culex pipiens s.s.* and *Culex torrentium* reliably?

In this section, the data were checked to see if it is possible to use a reduced dimension approach against the data already collected. The ratio between wing vein R2+3 and R3 (Figure 48) was investigated to determine whether it was discriminatory between locally collected *Culex pipiens s.s.* and *Culex torrentium.*

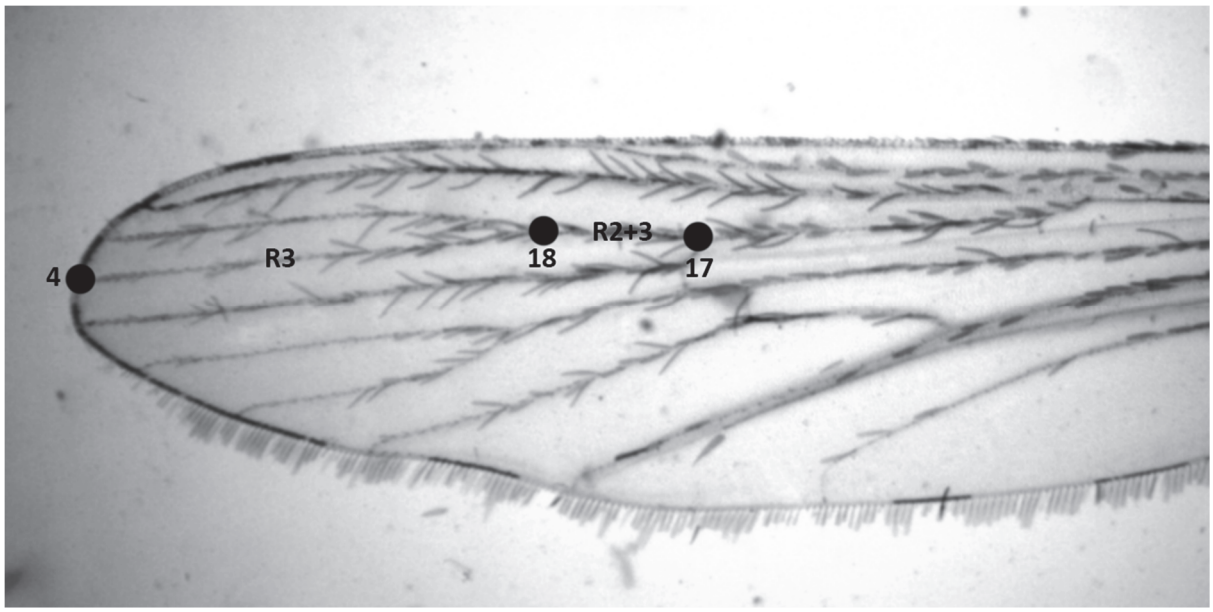


Figure 48 The 3 landmarks being called from the database, using an SQL query, to facilitate the calculation of the ratio of wing vein R2+3 and R3. This is a putative discriminating feature for *Culex pipiens s.s.* and *Culex torrentium.*

There was a significant difference between the distribution of the ratios of R2+3/R3 between the species (Figure 49). However, there was an overlap between the ratios (Table 16 and Figure 50) which weakened the discriminatory power of this metric. This was exemplified when using the wing vein ratio R2+3/R3 to predict species identification of both wings of the known unknown data, 50 of 77 wings were assigned to the correct species (64.94 %). However, if only the right wings from the known unknown data were used in the unknown data, then 12 of 38 were assigned to the correct species (31.58 %). Whether there is a measurable difference between left and right wings should therefore be investigated.

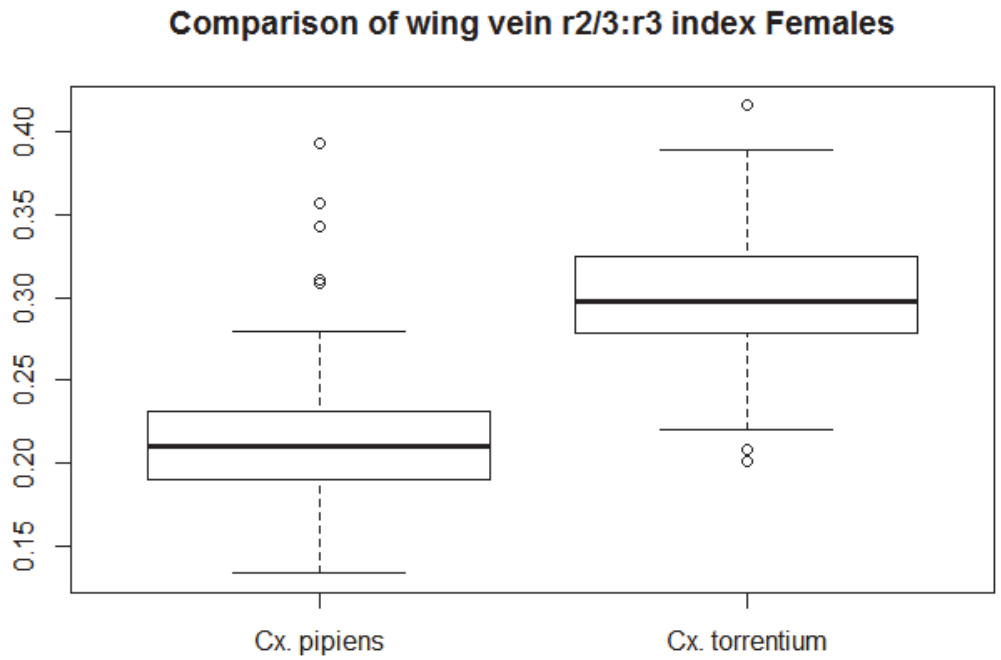


Figure 49 Wing vein ratio R2+3 to R3 for *Culex pipiens s.s.* and *Culex torrentium*. ANOVA showed that the difference between the species was significant (p = < 0.001)

Table 16 The mean ratios of wing vein R2+3/R3 for *Culex pipiens s.s.* and *Culex torrentium*. Whilst there is a significant difference between the species, there is also sufficient overlap that a bilateral measurement cannot be used to differentiate an individual mosquito's species.

Species	Mean R2+3/R3 ratio	Standard Deviation	Minimum	Maximum
Culex pipiens s.s.	0.2166	0.0404	0.1337	0.3930
Culex torrentium	0.2995	0.0395	0.2015	0.4157

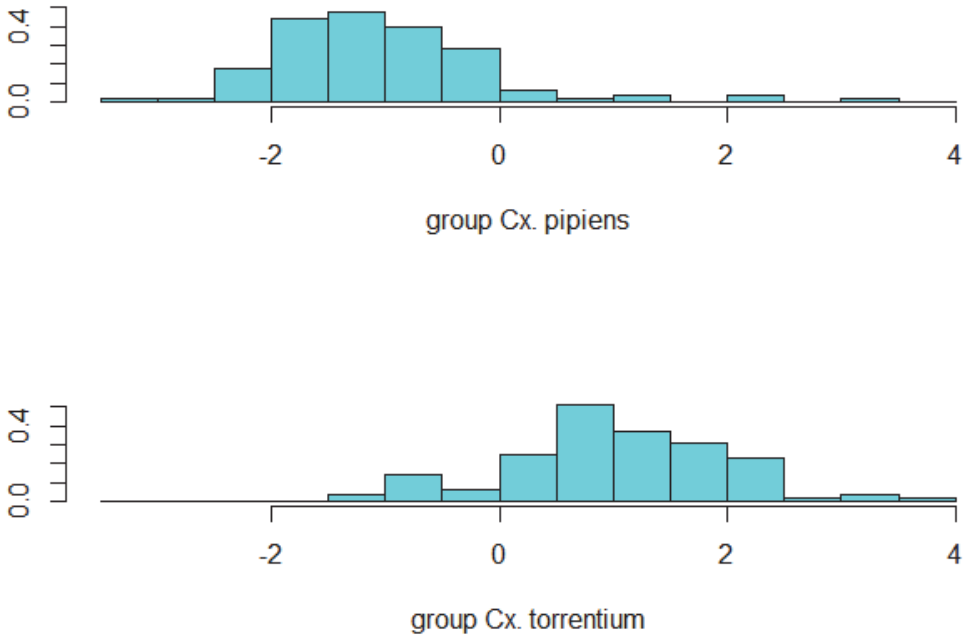


Figure 50 Linear discriminant analysis plot of the R2+3/R3 wing vein index of *Culex pipiens s.s.* and *Culex torrentium* shows that the species are only partly separated by this ratio.

4.1.2.2.2 Discussion

Results showed that the Börstler *et al.* (2014) method had mixed results with mosquitoes collected in North Staffordshire; significant multivariate and bivariate differences between the species were found, but identification of "unknown" specimens showed mixed success. Procrustes analysis gave the most accurate method for identification, with Mahalanobis distance being less accurate. That there was such disagreement between the two approaches was somewhat frustrating and reduced confidence in the power of the model to identify new specimens i.e. those not involved in creating the model. Dujardin *et al.* (2010) found that Mahalanobis identification was very powerful but sensitive to outliers, perhaps with the small data set used in this validation study this sensitivity was seen and resulted in the incorrect identifications.

Using the wing vein ratio, R2+3/R3 was found to be unsuitable for the identification of *Culex pipiens s.s.* and *Culex torrentium* for these locally collected specimens. The overlap between the ratios was considerable, therefore ruling out this approach of species discrimination. These findings are quite different to those of Börstler *et al.* (2014) who found quite disparate ratios and a very precise tool for species discrimination; they are, however, similar to those of Service (1968), who studied British specimens and of Fedorova & Shaikevich (2007), who

worked with Russian examples. It appeared, therefore, that British *Culex pipiens s.s.* and *Culex torrentium* were more similar morphologically than those found in Germany by Börstler *et al.* (2014) and is supportive of the notion that more localised studies are required before any general approach can be assumed, even in these cryptic, sibling, sympatric species.

Despite these mixed results, there was some discriminatory power using wing morphometry and therefore potential in the approach to be applied to British *Culex pipiens s.s.* and *Culex torrentium* mosquitoes. However, a more in-depth analysis would be needed to ascertain the best approach to take with regard to identification of which features the principal components relate to, and to optimisation of landmark collection to ensure all between species variance is captured. By conducting further study the user can be informed of the accuracy of wing morphometric discrimination between *Culex pipiens s.s.* and *Culex torrentium* and choose whether or not it is sufficient to justify the use of this method.

When considering a complete study, it was important to reflect on the work involved in processing the wing data used in the pilot. The process seemed quite user intensive in places. Whilst the method of collecting the wing morphology landmarks using the COO module of CLIC was quite intuitive and user-friendly, the subsequent data handling and processing using the MOG module of CLIC could be somewhat frustrating with a series of button clicks required. A more automated approach might have been preferable. Result output was considered to be appropriate for the communication of results for a single treatment but lacked the flexibility needed to compare different approaches and landmark collection regimes. For example, if the user wanted to try only 12 of the landmarks rather than all 13, or, perhaps, wanted to reorder them, then the user was tasked with either recollecting the landmarks from the images or painstakingly editing many text files. This was unsatisfactory and it afforded ample opportunity for operator measurement or data manipulation error which could influence the comparison between approaches. Therefore, a new post-processing procedure written specifically to facilitate the manipulation of variables applied to calculations, whilst leaving the original data pristine for subsequent tests is required.

For ease of identification male mosquitoes were used for this study; the final application, however, will be for the identification of female *Culex pipiens s.s.* and *Culex torrentium*. The results from this study can only be considered as suggestive of a potential sufficient disparity

between species wing shapes for this method to be useful, and that the correct identification rates may vary from those seen in male mosquitoes. Part of further studies should explore whether there is a significant difference between the male and female wing morphology, as has been shown to exist in other insects (Francoy *et al.* 2009), and other mosquito species (Christe *et al.* 2016). During this study, only morphological identification of mosquitoes needed to be carried out as these had shown to be accurate for males of these species. However, further studies using female mosquitoes require the morphological identifications to be supported by molecular identification using PCR. These DNA identifications will allow the use of female wild caught mosquitoes to be used in the shape analysis training data, and tailor the approach to the specific need in the field – the discrimination of female *Culex pipiens s.s.* from female *Culex torrentium* mosquitoes, as it is here that we lack any other reliable morphological feature based approach (Becker *et al.* 2010a).

In this study, wings were removed, flattened and placed on temporary dry slide mounts prior to being imaged; this approach has the advantage of being quick and very easy, it did, however, prevent the retention of reference materials which it might be desirable to retain as reference materials to underpin the approach. Additionally, there is the potential for some movement of the wing in the vertical axis, potentially allowing parallax errors to cause a slight variation in apparent values in the horizontal plane. Therefore, for all subsequent work, permanent slide mounts will be prepared and stored as reference specimens, in case further replication or verification be required.

The results of the complete study, following the adjustments suggested above, will then be quantified, analysed and reported. If the approach can be shown to function as intended, then a simplified workflow to use as an identification tool should be created, based on the best configuration from the findings, for utilisation in the wider study within this thesis, and should also be made available to other entomologists to help cheaply and quickly discriminate between female *Culex pipiens s.s.* and *Culex torrentium*. In an effort to further reduce the equipment needed and time taken to capture the wing images, thereby facilitating the use of this method by amateur entomologists, the further study will also investigate the use of a flatbed scanner to capture the wing venation data from the prepared slides.

4.1.3 Wing morphometric species diagnosis between British *Culex pipiens s.s.* and *Culex torrentium*

In chapter 4.1.2 it was demonstrated that the wing morphometric approaches, using either the ratio of veins R2+3 and R3 or the length of R2+3 alone, described by Börstler *et al*. (2014) could not be applied directly to British specimens due to overlapping value ranges between *Culex pipiens s.s.* and *Culex torrentium*. As noted by Börstler *et al.* (2014), overlaps of this nature had been described previously in Britain (Service 1968), and it was interesting to see that they still persisted even after almost 50 years. However, despite the overlap in this measure, which effectively ruled out the use of univariate and bivariate methods of species discrimination, there was sufficient potential in the multivariate analysis, as shown in the partially successful identification of unknown specimens, to conduct a full analysis. In this chapter, this analysis is conducted.

Within this chapter, a method of accelerated collection of digital images of mosquito wings using a flatbed scanner is tested against using a microscope mounted digital camera. The expectation was that the resultant method would be made more cost efficient; by replacing the necessity of an expensive microscope and camera with and cheaper digital scanner. Budget microscopes of this kind cost over £850 in the UK (Cole-Parmer 2018), compared with under £60 for the scanner used in these experiments (Canon 2018). This, therefore, would be more accessible to amateur entomologists. The ability to capture multiple images at once using a scanner would also facilitate higher throughput, making it a more attractive alternative for higher volume sampling.

Having determined the appropriate method of image collection, a new analysis pathway was coded in R Statistical Software (R Core Team 2015) and used to test the capability of multivariate analysis of shape data related to wing morphometry to discriminate between female *Culex pipiens s.s.* and *Culex torrentium*. The relationship of other variables was also explored in this section, including male versus female wings, to investigate whether gender dimorphism existed in these species (Francoy *et al.* 2009; Christe *et al.* 2016) and left versus right wings to observe whether evidence of directional asymmetry (DA) exists. The analyses being employed would not be sensitive to fluctuating asymmetry (FA) (Parsons 1990) or as calculations are based on consensus values of multiple individuals and not at the individual

level. Similarly, it would not be expected to detect antisymmetry (AS) as the random distribution of the symmetries in the population would be hidden by the even distribution about the mean. However, it would detect DA if the magnitude of the asymmetry was large enough to move the shape centroid of one wing compared to the other.

Subsequently, the optimal number and identity of wing landmarks were explored, and the accuracy of the method analysed for data within the identification model and for new specimen data. Finally, a new software tool to simplify the analytical process, such that it could be used as a tool for species discrimination between female *Culex pipiens s.s.* and *Culex torrentium* was designed and is presented.

4.1.3.1 Testing the use of a flatbed digital scanner to collect wing images

4.1.3.1.1 Introduction and Aims

Capturing high-quality images of mosquito wings from slides individually using a microscope can be time-consuming. Therefore, it was hoped that it might be possible to capture multiple slides at once using a flatbed scanner. The idea being that multiple slides could be placed on the scanner at one time, and after scanning at high resolution, the resultant image would be of high enough quality to be suitable for the wing detail to be used for identification purposes. Flatbed scanners have been shown to be useful in other biological digitisation settings (Sullivan *et al.* 2012), and could be applied for this purpose. The intention of using a flatbed scanner was that this would save time by being able to capture many images at once and make the technique and data available to amateur entomologists who may not have access to microscopes with mounted cameras. There are purpose built commercial solutions to microscope slide digitisation (García Rojo *et al.* 2003), but these are not being considered here as they are primarily designed for clinical settings and would be even less likely to be available to entomologists that molecular methods.

4.1.3.1.2 Method

4.1.3.1.2.1 Preparation of permanent slide mounts

All mosquito specimens were stored at -20 °C from death until the time of use, although card mounted or pinned specimen preservation would provide equally viable specimens (Walker and Crosby 1988). After defrosting at room temperature for at least two hours, each wing was carefully removed from the thorax using fine-tipped forceps under a dissecting microscope under low magnification (~20x). Both wings were placed face down on a flat glass microscope slide, with the wing/body joint at the centre, so that when viewed from above the left wing will be on the left of the slide, to avoid any confusion. If the wings were particularly heavily scaled, such that wing vein intersections were obscured, then scales were removed by gently brushing away with a fine paintbrush. Wings were then inspected, if damage or distortion was observed then, the wings and specimen were discarded. A small piece of 1 mm graph paper was placed adjacent to the wings as a permanent scale reference.

After inspection, four drops of D.P.X. (Sigma-Aldrich) slide mountant (~ 0.09 g) was added, and the wings flattened again, if necessary, before being covered with a glass coverslip. Each wing slide was labelled, with the specimen's unique reference number, the sex, and

morphological identification, if known (Figure 51); species DNA identification and verification (Chapter 2) took place after the process of wing removal. The slides were then left to cure at room temperature for at least 24 hours before any imaging took place. Post-curing, slides were placed in a slide box for long term storage.

The method of slide mount preparation was the same for the training data specimens, and the "known unknown" test specimens used later for predictive testing.

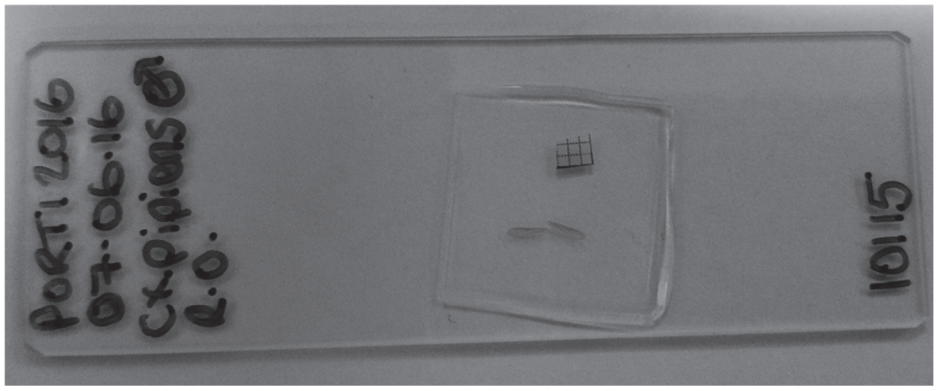


Figure 51 A permanent slide mount of a pair of *Culex* wings. Labels include source, date of preparation, species identification, gender and a unique slide reference number.

The now wingless specimens were then transferred to 1.5 ml Eppendorf tubes, labelled to correspond to the slide and quickly returned to the freezer at -20 °C for storage ahead of molecular identification. The molecular methods employed have been described chapter 2.

A Canon CanoScan LIDE 110 2400 dpi flatbed scanner was selected to test this image capture approach; this is a modern scanner designed and priced for the home office user. A previously prepared wing slide was placed on the scanner and scanned, following the manufacturer's instructions, and using the manufacturer's supplied software. Scans were captured at 300, 600, 1200, 2400 and 4800 dpi and the resultant files saved as uncompressed .tif files. While scanning, the slide was placed face down on the scanner to avoid transposing the wings left to right due to scanning through the bottom of the slide.

As a comparison, the same slide was imaged using a Meiji compound microscope with a 10x eyepiece and 4x objective lens. Images were captured at 1024 x 768 resolution, using a Microtek IS300 USB camera connected to ISCapture software (Ver. 4.1.3 Tucsen Photonics

Co., Ltd.) running on Windows 8.1. Post processing of the images was carried out in Adobe Photoshop CC (Ver. 2015.5) and was limited to cropping of the background enlargement, rotation of the whole picture, and adjustment of brightness and contrast. The resulting images were then inspected visually for suitability for use in plotting wing landmarks and compared with the other scanner images, and the digital camera image.

4.1.3.1.3 Results

An example of the region of interest on the scanned images is shown in Figure 52; at 300 dpi in this case. At the native resolution, the image shows some detail, but granularity is already visible; post zoom there is a significant loss of visual quality and detail.

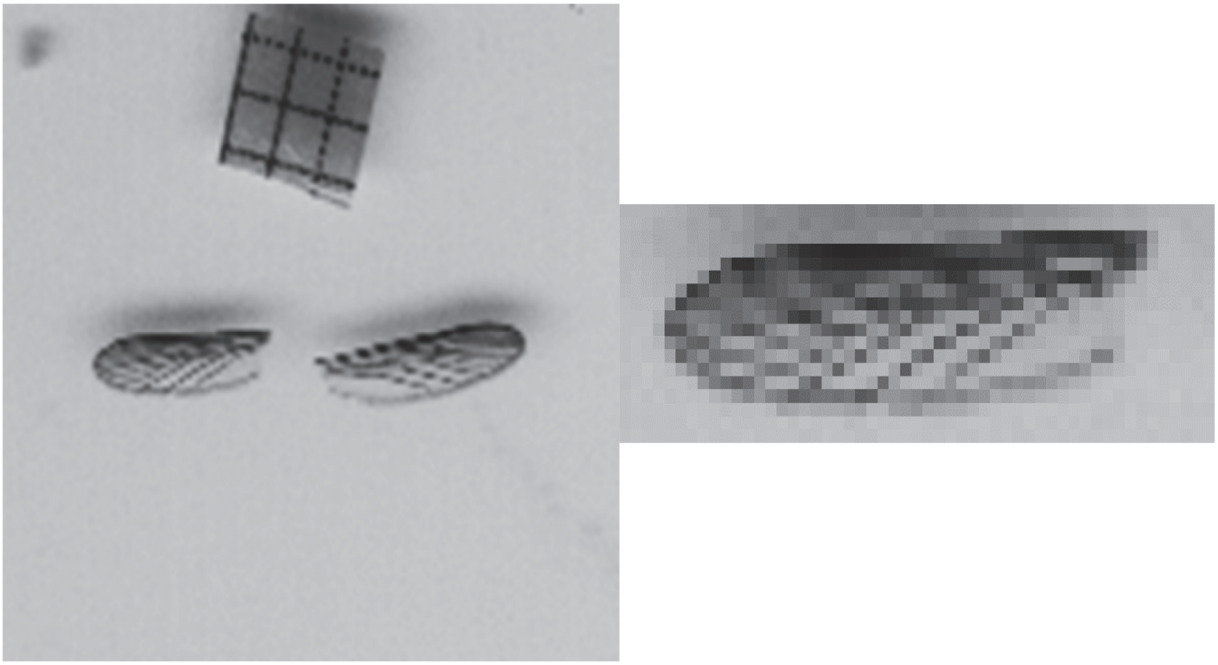


Figure 52 On the left is the .tif image as captured at 300 dpi using the CanoScan LIDE 110. The image to the right is the same image post enlargement and shows significant detail loss.

Comparing the partially zoomed images side by side, this loss of detail is evident (Figure 53). As dpi increases, the image quality gets much better. However, this improvement ceases above 2400 dpi, and it can be seen, in Figure 53, that 4800 dpi is no more detailed than 2400 dpi. When viewing the best scanner images (>= 2400 dpi) alongside the digital camera image the difference in acuity is clear at low zoom (Figure 53) and becomes increasingly so at a level

of enlargement which could be used while capturing landmarks for wing morphometric treatments (Figure 54).

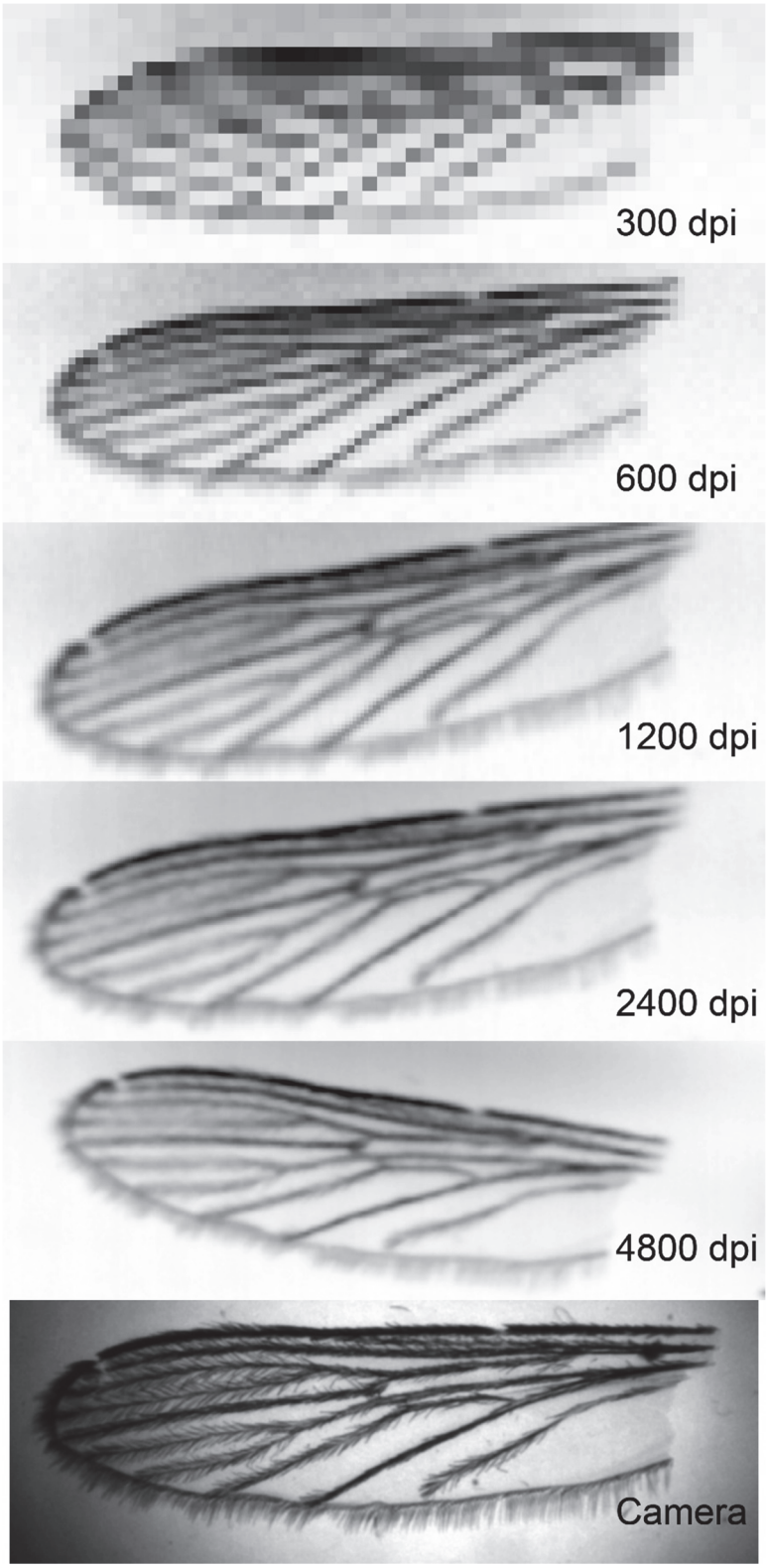


Figure 53 Enlarged images of the scanned wing slides at different resolutions (first five images) above the digital camera image show a clear difference in image quality. These images would be further enlarged before capturing landmarks for wing morphometrics.

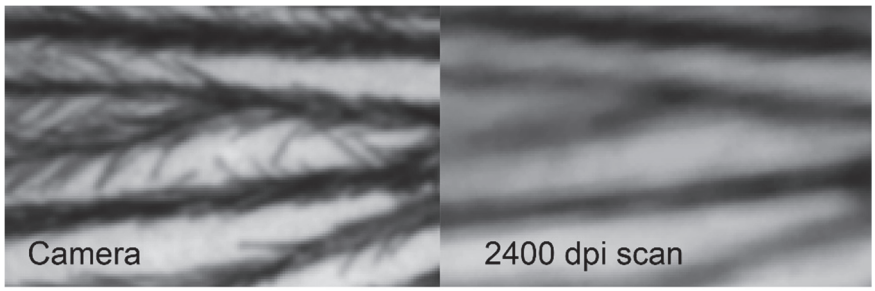


Figure 54 The intersection of wing veins R2+3, R2 and R3; displayed at the size used for landmark collection. The camera image (left) clearly shows much more detail than the scanned image (right).

4.1.3.1.4 Discussion

Results using a flatbed scanner were somewhat disappointing; even at the highest dpi setting (using a typical, modern, home use flatbed scanner), once enlarged the pixelation was too high to be confident discerning the exact position of wing vein intersections. There are higher dpi scanners available to the professional market, but these are currently prohibitively expensive; presumably, these will, over time, become more affordable and then this approach could be re-evaluated.

That there was no obvious image quality improvement beyond 2400 dpi was unsurprising as the scanner's optical resolution is 2400 dpi; resolution above 2400 dpi is achieved by a software solution, in other words, the scanner does not capture more data, rather the pixels captured are made smaller. Therefore, although the scanner claims to be able to capture images up to 19200 dpi, the actual image data captured and therefore available to the analysis does not increase. The high level of detail of the camera image is of particular importance when processing wings which still have many scales on the veins; the detail is sufficiently high that the underlying vein can be discerned as a darker band beneath the more diffuse scaled area. Although the use of the flatbed scanner has been shown to be substandard for use with mosquito wings, it should be borne in mind, however, that these are small wings (~ 4.5 mm). When working with other, larger, insect wings, it could be a viable approach and should be evaluated for use in those cases.

The main finding of this experiment was that the utilisation of this type of scanner to capture the images for wing morphometric analysis could not be recommended. The picture quality was not sufficiently good for the landmarks to be plotted with enough confidence to justify its use. Whilst there are commercial systems for the digitisation of prepared slide data (García Rojo *et al.* 2003; Gavrielides *et al.* 2014), they tend to be designed for pathology settings, e.g. the Hamamatsu NanoZoomer range (Anon 2017) and are beyond the scope of use for the application being discussed as they are prohibitively expensive. For the remainder of this study, this method was discarded in preference of using a digital camera mounted on a compound microscope and linked to a PC.

4.1.3.2 Exploration of the relationships between biological variables and wing morphometric discrimination

4.1.3.2.1 Introduction

The approach used here was designed to facilitate analysis of the ability of species discrimination, between *Culex pipiens s.s.* and *Culex torrentium* mosquitoes, using wing morphology alone.

The study carried out previously (4.1.1.4) found that the process of collecting landmark data was relatively straightforward. However, if the user wanted to test different landmark combinations, for example, their order or number of landmarks used then the process was too unwieldy and provided too much opportunity for human error; due to the need for editing large text files or requiring the user to collect landmarks from images again. Particular attention has been given to making analysis, post landmark collection, as straightforward and powerful as possible; thereby removing barriers, such as ease of use, reproducibility and standardisation, to this method's use in further studies. The ability to streamline the process and quantify the discriminatory power of the methods being tested required the use of technology which supports ease of data manipulation and allows for complete reproducibility. For this reason, data were stored in an Access® database; facilitating retrieval using Structured Query Language (SQL). All data handling and calculations were carried out using R Statistical Software (R Core Team 2015), and was fully coded such that the original data were left pristine and all calculations performed live, and so that all steps of the analysis were applied in the exact same way for each iteration and test permutation.

Several issues needed to be investigated to determine which wings could be included in the main data set. For example, the left wings should be compared to the right wings to ensure that they are not significantly different as might occur in situations related to genetic, environmental and stress (Palmer 1994; Hosken *et al.* 2000; Costa *et al.* 2015). Male and female wings then needed to be checked to test the assumption that they are different as seen in other insect species (Francoy *et al.* 2009; Christe *et al.* 2016); if they were not significantly different, then both sexes could be included in the training dataset used for the identification.

4.1.3.2.2 Aims

The primary aim was to determine whether wing morphometric analysis has sufficient discriminatory power to be used as a tool for species discrimination between *Culex pipiens s.s.* and *Culex torrentium mosquitoes.* If successful, the study should quantify the confidence in the discriminatory power of the method, such that this can be clearly communicated when this approach is used for identification.

A secondary aim was to find the optimal number of landmarks needed to make this species discrimination. It was imperative that the system for analysis used in the investigations facilitated the editing of landmarks without requiring the reacquisition of landmark data between treatments. Therefore, the collected data needed to be unaltered by the analysis. The intention of the process was to optimise the collection of the smallest number of landmarks possible whilst maintaining the discriminatory power of the method.

Finally, if possible, a pared-down version of the R computer code (R Core Team 2015) required for species discrimination should be created which would serve as a simple tool for the discrimination of these species in other studies, or could be distributed to entomologists for their use.

4.1.3.2.3 Method

Permanent slide mounts of pairs of wings from 263 mosquitoes were prepared, using the method previously described, with an even split between the sexes and *Culex pipiens s.s.* and *Culex torrentium*. The mosquito specimens utilised for this were collected from a range of locations across North Staffordshire and environs, from locations discussed in the Fieldwork chapter. Specimens were identified using PCR, as per the method described in chapter 2, or were specimens from a single egg raft which had specimens from the same egg raft already identified by PCR.

Each wing was individually photographed using a Meiji compound microscope with a 10x eyepiece and 4x objective lens. Images were captured at 1024 x 768 resolution, using a Microtek IS300 USB camera connected to ISCapture software (Ver. 4.1.3 Tucsen Photonics Co., Ltd.) running on Windows 8.1. Images were stored in .tif format. Images were then processed using the Collecting Landmarks for Identification and Characterization (CLIC)

software package (Dujardin 2016). Eighteen landmark features were plotted at the major wing vein intersections (Figure 55). Following landmark collection, these coordinates were saved in a text file (…_DB.txt), before being imported into an Access® database table.

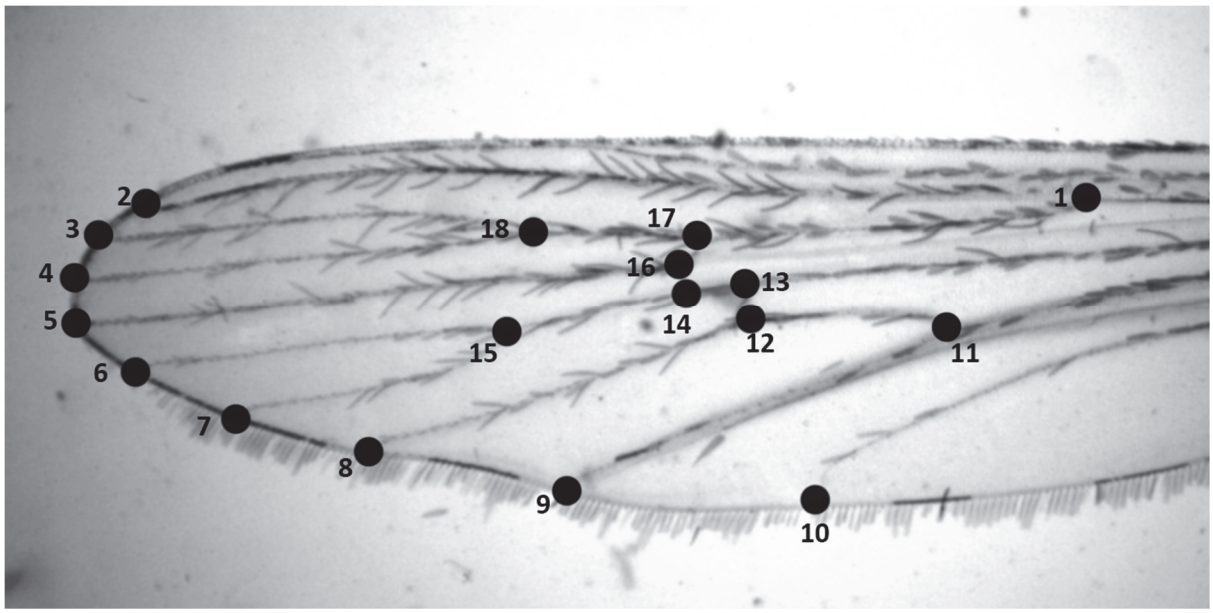


Figure 55 A *Culex* sp. mosquito wing image showing the 18 wing landmark locations are plotted using the MOME-CLIC interface

For each wing image the following fields were collected and stored in the Access® database:

- A unique database entry ID
- The path to the original image source being handled
- The image name – including the specimen's unique identifier.
- The species identification determined by PCR and or morphological methods
- The sex of the specimen
- Whether this is the left or right wing
- A scale value (set to not applicable (NA) for this study)
- X and Y coordinates for the 18 landmarks

4.1.3.2.3.1 Data analysis

R Statistical Software was used for all subsequent data analysis, including retrieval from the database using SQL via the RODBC (Ver. 1.3-13) package (Ripley and Lapsley 2016). Data manipulation, including selection and filtering of observations, used the dplyr (Ver. 0.5.0) package (Wickham and Francois 2016). Processing of shape data used the Geomorph (Ver. 3.0.2) package (Adams and Otárola-Castillo 2013) and Adegenet (Ver. 2.0.1) package (Jombart 2008). An example of the analysis code is included in Appendix 2 for reference, and step by step process detail.

4.1.3.2.3.2 Establishment of the starting conditions for species identification - Comparison between effect of biological variables

To ensure that the identification protocol being developed was discriminating by species related differences and not on other factors several potential causes of biological variance were examined. Specifically, the following were investigated for evidence of disparity between groups:

- *Culex pipiens s.s.* and *Culex torrentium* when using combined male and female data.
- Gender when analysed using combined *Culex pipiens s.s.* and *Culex torrentium* data.
- *Culex pipiens s.s.* and *Culex torrentium* females only, both wings.
- Wing asymmetry within the female-only data.

The eighteen-landmark dataset for all the training data was retrieved from the database using SQL and loaded into R. To aid visualisation of results the coordinates were inverted, *i.e*. all values * -1, such that the subsequent display outputs showed the wing the right way up. The data were then filtered to retain the variable combination which was to be investigated for each treatment using the filter() function of dplyr. Mosquito sex and wing choices (left or right) were specified e.g. (filter(fromdb1, COMMENT == "Female", VAR=="Right")).

GPA with partial Procrustes superimposition was then carried out using the gpagen function from Geomorph. The results were then plotted to ensure the data appeared correct. From Geomorph the plotOutliers and other associated functions were used to calculate and visualise any outliers from the main data and remove extremes if considered as erroneous. If

any data were removed at this stage, then GPA was carried out again. The morphol.disparity function, using 10000 iterations, calculated the difference between species for the given variables on the basis of Procrustes variance and then test the significance of any observed difference using an ANOVA. The tangent space of the variables being investigated was plotted, using the plotTangentSpace function from Geomorph, to visualise the distribution of individuals of the groups against PC1 and PC2. The weighting of the principal components was also plotted, as was the deformation from the consensus landmark configuration of specific examples of each variable type, using plotRefToTarget from Geomorph, to better understand the data and visualise the direction/s of the shape change. This process was repeated for each of the 4 variable combinations being tested.

4.1.3.2.4 Results

4.1.3.2.4.1 Whole group Culex pipiens s.s. versus Culex torrentium. 18 landmark data.

Following GPA, the transformed, rotated and rescaled data were visualised to show the difference mean landmark position for the whole group data and the variance between individuals within the population (Figure 56). It was immediately apparent that the variance in the different landmarks was not evenly distributed. For example, there was more visible variance in landmark 18 (see Figure 55 for the landmark mapping) than in landmark 4. It was also apparent that the variance in landmarks 18 and 15 appears to be mostly horizontal, particularly when compared to landmark 8 which shows both vertical and horizontal variance.

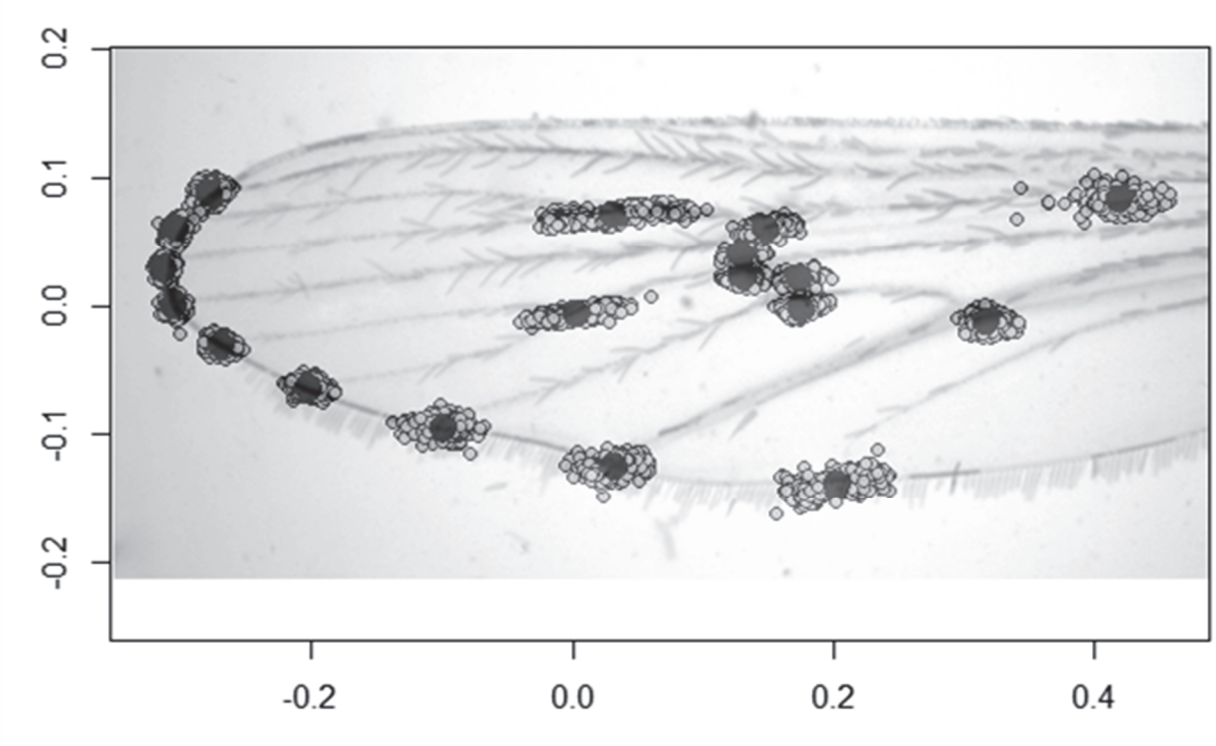


Figure 56 Landmark data, for the whole dataset, after Generalised Procrustes Analyses. Black spots are the mean position for each landmark, grey points are individual landmarks and show the variance in the data

There are some points in Figure 56 which appeared to be further away from their respective consensus landmark; landmark 1 had 2 points which might be outliers to the rest of the data. However, when the data was considered using the Procrustes distance from the mean to find potential extreme individuals that might bias the findings, those wings above the upper quartile were still a continuous and closely matched group (Figure 57) and as such most likely representative of variance found in the population. Inspection of these wings showed

deformations which appear similar to the expected ranges seen in Figure 56; an example output for an individual deformation plot is shown in Figure 58. Because, after inspection, the deviation seen in these points was shown to be distributed through many landmarks, as opposed to a unique landmark having a considerable deviation from the mean, no outliers were removed from the data in this treatment.

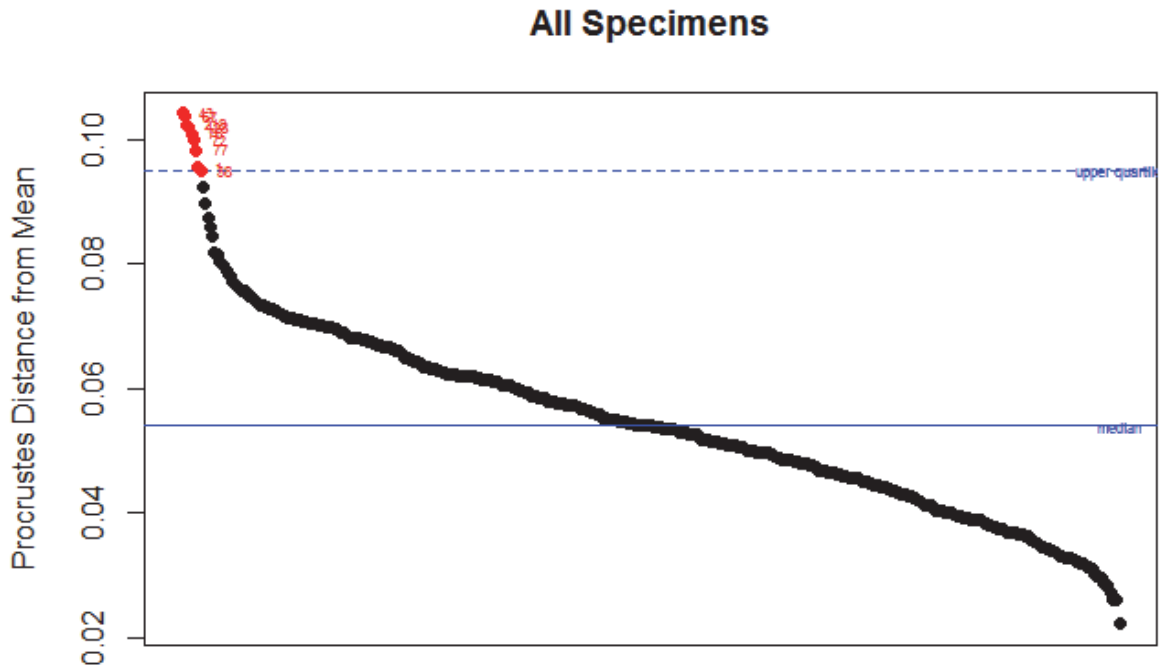


Figure 57 Procrustes Distance from the mean. The dotted line shows the upper quartile, and the red dots above it warrant inspection to check for possible outliers.

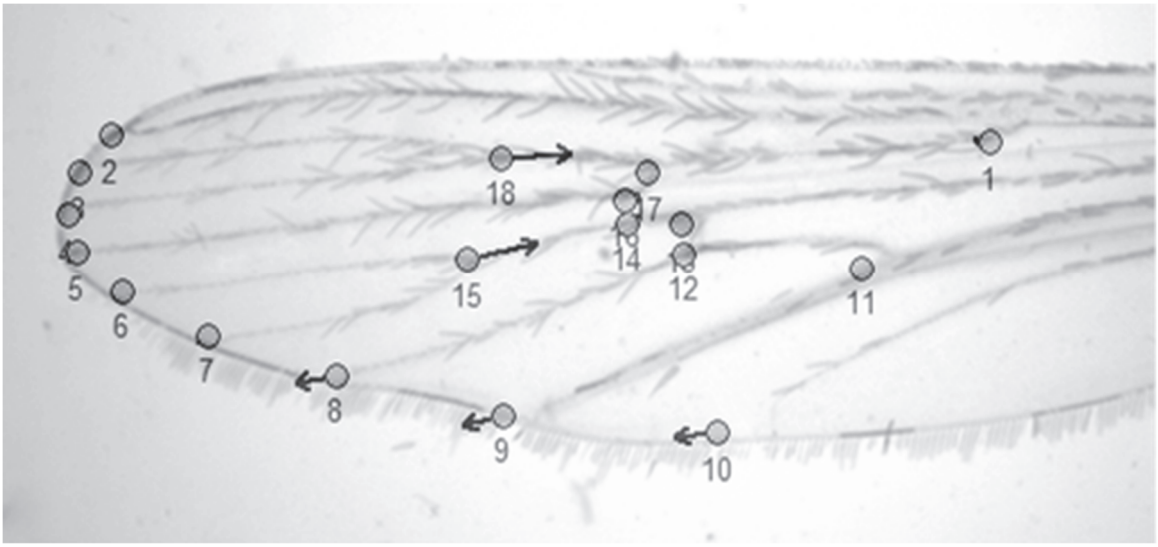


Figure 58 Exemplar deformation plot used as part of the analysis of outliers. Grey points are the mean group position; arrows show the direction and magnitude of the deviation of the individual being plotted.

The calculated Procrustes variances for the *Culex pipiens s.s.* = 0.00343 and *Culex torrentium* = 0.00300, giving an absolute difference of 0.00430, with a p-value of 0.0171. Therefore, there was a significant difference between the 2 species was observed in the data. Inspection of the influence of the Principal Components showed that PC1 explained ~70 % of the variance seen in the data. However, when the data were plotted against PC1 and PC2 (Figure 61) it became apparent that, due to the even distribution of both species along PC1, that PC1 was not related to the variance on a species basis but more related to sex.

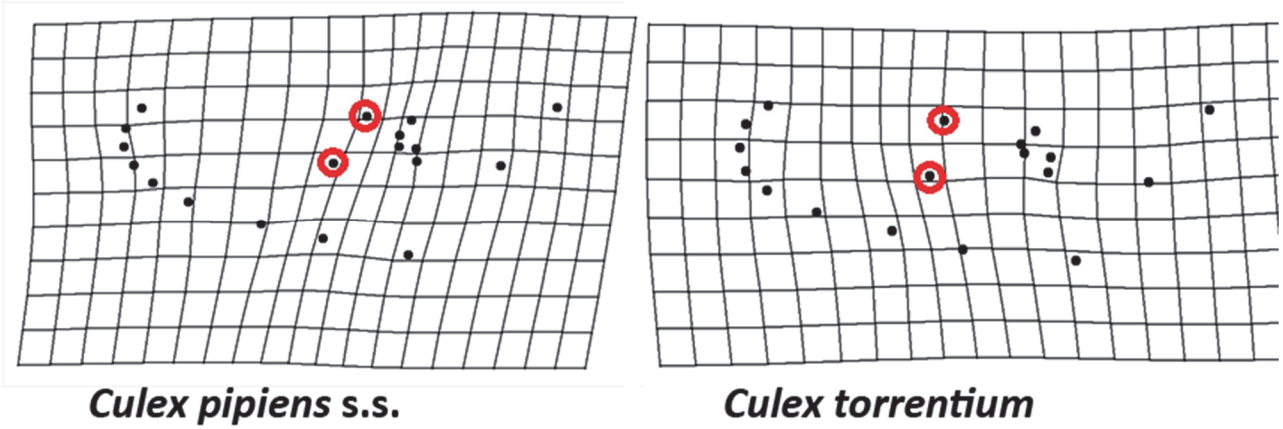


Figure 59 Magnitude of the horizontal difference in location of the species consensus points for landmarks 15 and 18, circled in red, between *Culex pipiens s.s.* and *Culex torrentium* when the analysis was conducted on the complete dataset, including male and female mosquitoes.

From Figure 56 and Figure 59 there appeared to be a large horizontal variation in landmarks 15 and 18 between the species. Therefore, the calculation was conducted again with landmarks 15 and 18 removed (Figure 60) to determine whether their assumed role was indeed significant. This resulted in Procrustes variances for *Culex pipiens s.s.* = 0.00161 and *Culex torrentium* = 0.00163 with an absolute difference of 2.501×10^{-5} and a p-value of 0.758. Therefore, the importance of these landmarks was confirmed.

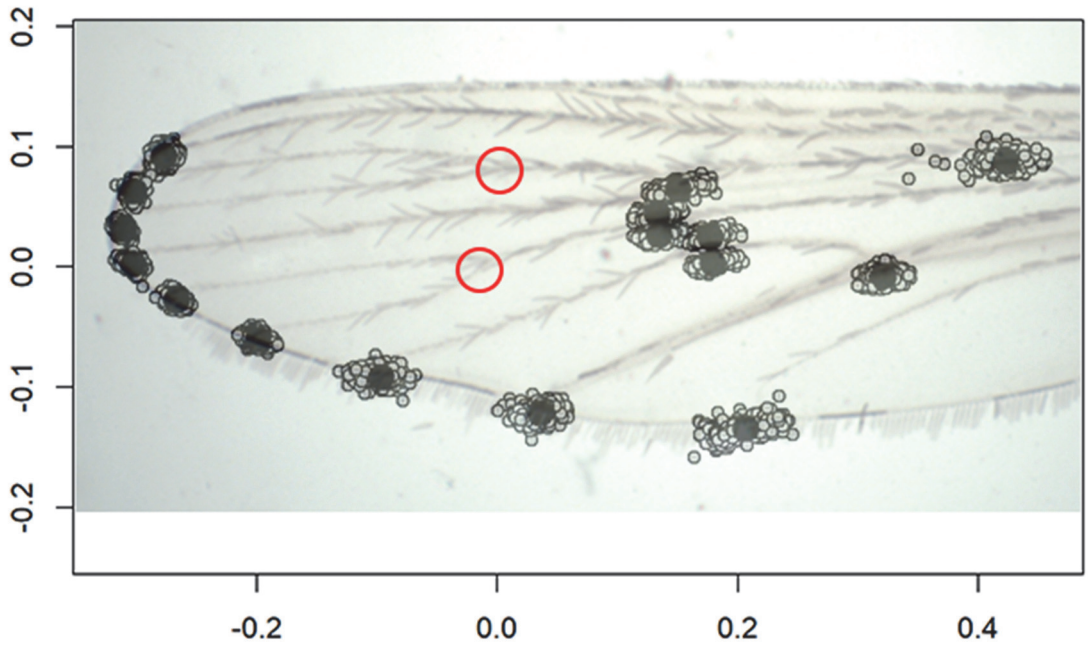


Figure 60 Landmark data, for the whole data set following the removal of landmarks 15 and 18 (red circles), after Generalised Procrustes Analyses. Black spots are the mean position for each landmark, grey points are individual landmarks and show the variance in the data.

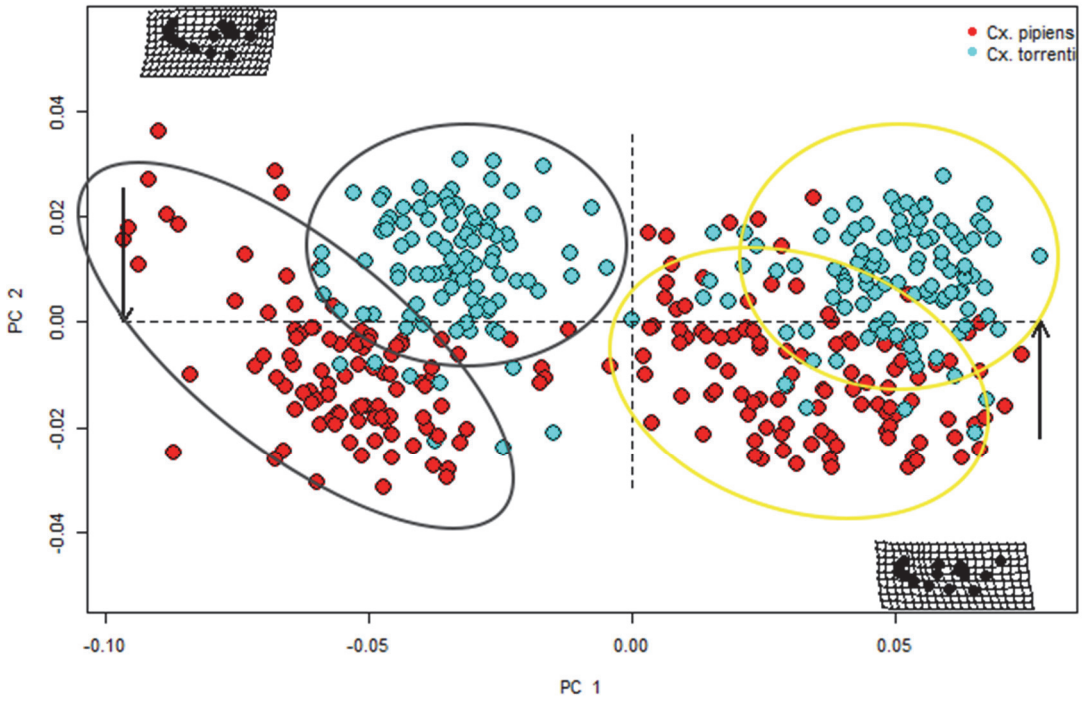


Figure 61 Species distributions when plotted against PC1 and PC2. The distribution of species is approximately even along PC1, suggesting it is not discriminatory for species; more likely to be discriminatory of sex. Grey ellipses suggest female groupings; yellow ellipses suggest males grouping.

4.1.3.2.4.2 Discrimination of gender

To inspect whether PC1 from the previous analysis was indeed related to sexual dimorphism the Procrustes variances of the sexes were calculated. Procrustes variance for females was 0.00328 and for males was 0.00313, giving an absolute difference of 0.00015 with a p-value of 0.3934. This insignificant result was, potentially, a result of having the both species data combined, particularly when Figure 62 is considered; here the separation of the sexes by PC1 is evident. However, when gender discrimination was analysed for species specific data the insignificance remained, although there is some suggestion of different levels of gender dimorphism between the species; in *Culex pipiens s.s.* the absolute Procrustes variance difference between genders was 0.00009 with a p-value of 0.7256. For *Culex torrentium*, the absolute Procrustes variance difference between genders was 0.0017 with a p-value of 0.285.

Figure 63 shows the thin plate spline projection of the deformation of wing vein intersection shape, comparing males and females within the data. When considering Figure 63, it is important to bear in mind that PC1 is a single component of all the apparent deformation of

the grid, and its large relative magnitude might sit within the context of an overall small, non-significant, gender difference. The presence of less obvious dimorphic factors, than PC1, was not assessed further, but that the all sexual dimorphism is captured in its entirety by PC1 seemed unlikely. Therefore, the decision was taken that for further analyses, the male data should be removed from the training data to avoid sexual dimorphic features influencing the model.

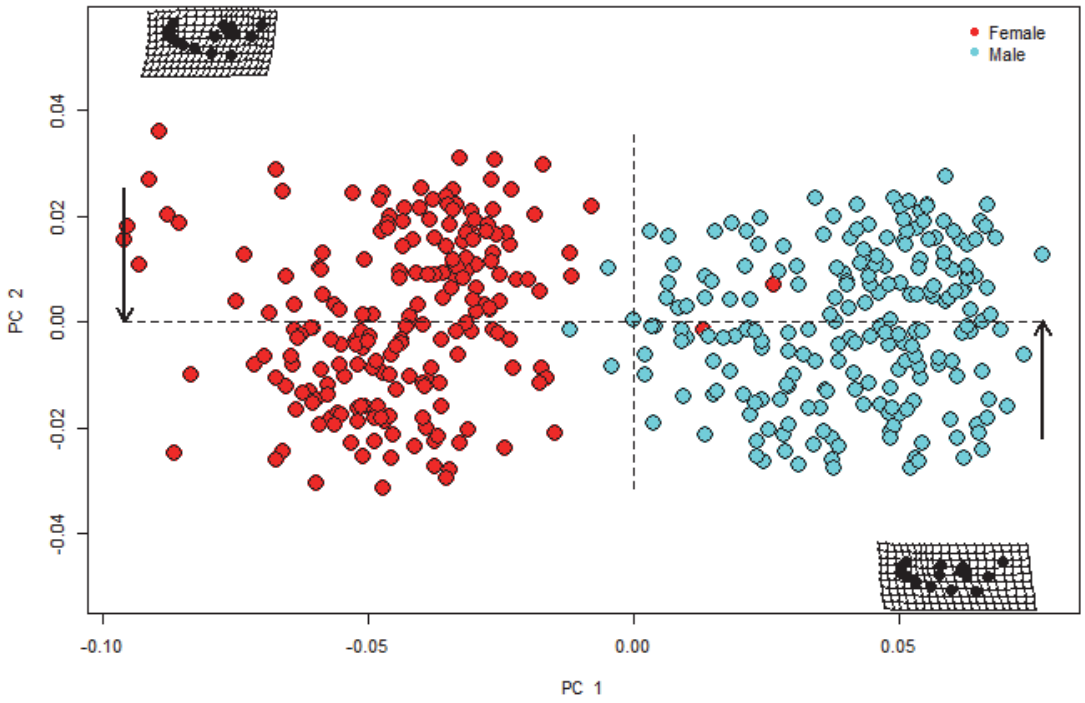


Figure 62 Distribution of the sexes when plotted against PC1 and PC2. PC1 is evidently an important measure of sexual dimorphism.

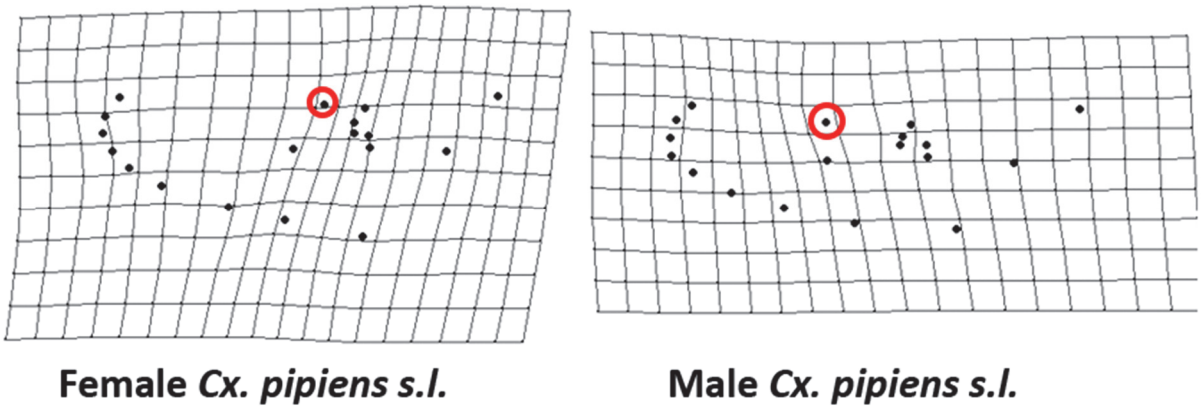


Figure 63 Thin plate spline visualisation of an individual example of each sex. The deformation of the grid shows the variation between the morphology of these individuals of each species and the mean wing shape for the whole data. The largest difference between male and female wings was the horizontal position of LM 18, which was found to represent PC1 in analyses for each species.

4.1.3.2.4.3 Left wing versus right wing

Within the female only data and with both *Culex pipiens s.s.* and *Culex torrentium* combined, Procrustes variance for the left wings was 0.00136 and for the right wings was 0.00135, giving an absolute difference of 1.44e-05 with a p-value of 0.9221. This insignificant result meant that where possible both wings were to be included in the training data set. Principal components analysis did not show any separation of the right and left wings (Figure 64), and therefore further supported the inclusion of both wings in the training dataset and the utilisation of the method for identification of individuals using either right or left wing with equal discriminatory power.

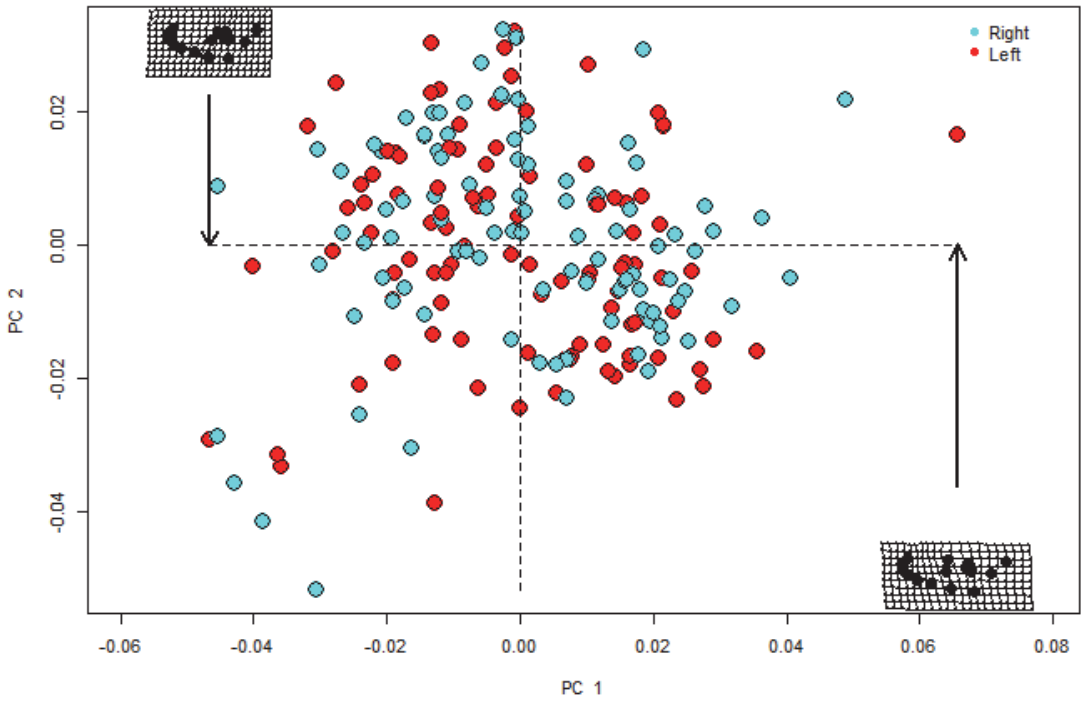


Figure 64 There was no separation between the right and left wings by PC1 and PC2. This, supported by the lack of significant difference between the right and left wings in bootstrapped ANOVA of Procrustes variance, support the inclusion of both wings in the training data.

4.2 Optimisation and accuracy quantification of wing morphometric discrimination of female *Culex pipiens s.s.* and *Culex torrentium*

4.2.1 Introduction

Having identified that landmarks collected from left and right wings should both be included in the training dataset, and that male data should be excluded, a method to test which landmark combinations would offer the best discriminatory power was devised. Exploration and optimisation of the method was also conducted; a poorly optimised method, using too many Principal Components (PCs) would create a model which fitted the training data very well, but may be unstable when used to predictively, becoming an overfitted model (James *et al.* 2013), conversely, using too few PCs could result in the discarding of significant discriminatory power.

4.2.2 Method

Three optimisation approaches were used:

- First, an iterative process using Cross Validation of Discriminant Analysis of Principal Components was used to find the number of PCs which best explain the difference between the species in the training data set.
- Second, an approach using alpha score optimisation was be used to measure the ratio of a successful group assignment to the training data and values obtained through iterative testing of each PC number. It is based on the proportion of successful reassignments corrected for the number of retained PCs. Multiple DAPC analyses were carried out using random groups; alpha scores were calculated for each group as were average alpha scores (Jombart and Devillard 2010; Jombart 2012).
- Third, the number of PCs were chosen by the operator following the inspection of the loadings data and the PC scree graph.

Different starting conditions were applied to the data treatment in the form of using an 18 landmark configuration (Figure 55), and the 13 landmark configuration as used by (Börstler *et al.* 2014); opportunities for optimisation were assessed throughout the process.

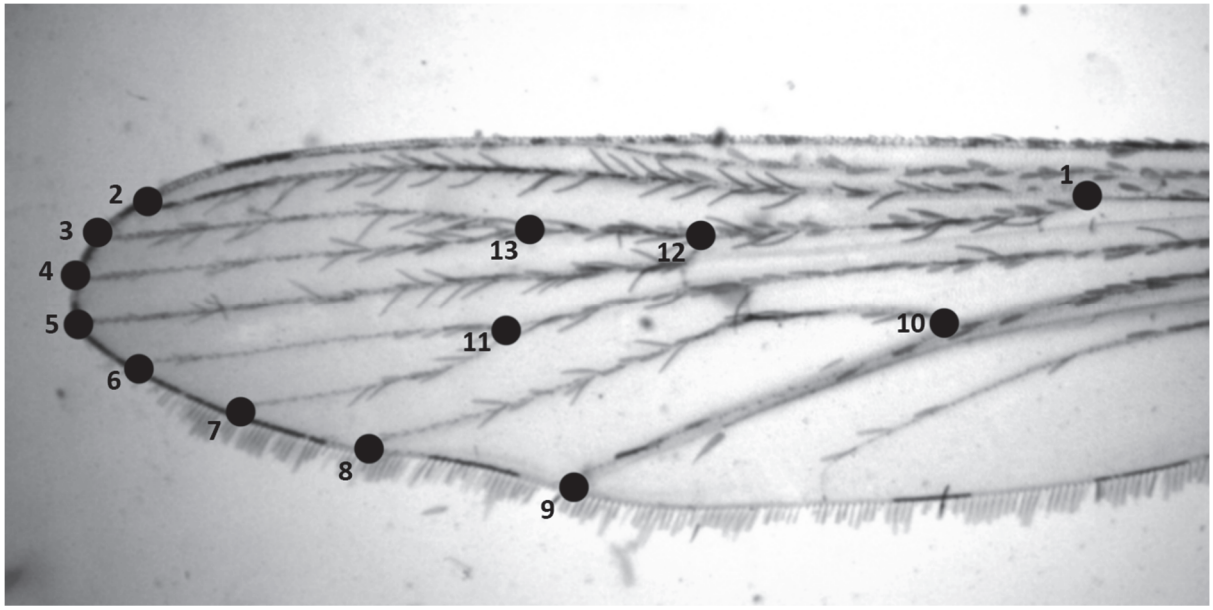


Figure 65 The 13 landmark configuration employed by Börstler *et al.* (2014) was used as one of the testing starting conditions for optimisation of the identification method.

For this section the data were filtered to retain both wings for female mosquitoes using the filter() function of dplyr, *i.e*. (filter(fromdb1, COMMENT == "Female")). Male mosquito data were excluded due to the difference discovered between sexes. GPA and morphol.disparity were conducted as before in the "Establishing starting conditions" method. The data were then examined by the operator at this stage to estimate the ideal number of PCs to retain; for approach 3 of the optimisation of the method. The xvalDapc() function was used, with 1000 iterations, to carry out cross-validated Discriminant Analysis of Principal Components to determine how many principal components was optimal for further analysis. This function was coded so that its findings are dynamically used in section 12. To enhance reproducibility set.seed(123) is used to establish starting conditions for "random" functions. The optim.a.score function from the Adegenet package was trialled as a method for optimising the number of PCs being retained for Discriminant Analysis of Principal Components. It should be noted that the optim.a.score function was still under development by the author and so was used as a guide only, any findings from this approach were tested using other means.

The following section was repeated for each of the 3 approaches above:

- The dapc function from the Adegenet package, using the number of PCs determined by each approach above, was used to check the fit of the training data to their identified species.
- Group membership of the individual specimens in the training data was visualised with a heat map and visualised the overlap of the species groups when plotted against the discriminant function, using compoplot and scatter.dapc, respectively, from Adegenet.

Those specimens which did not fit their actual group with at least 90 % confidence were recorded and visualised. Control data was then imported from the database, using SQL. These control data were pre-identified females of both species and acted as “known unknowns”, facilitating the testing of each model’s ability to identify the species of new specimens. These control specimens were grouped with the training data landmark data, and GPA carried out. The outputs of GPA are relative to a consensus and so for direct comparison all of the data needed to have values relative to the same consensus. DAPC was carried out again on the training portion of the data, following the consensus update, and the control data separated. The predict() function was applied with calculated DAPC model to the control data. Species identification of the “known unknown” data was calculated and visualised, before the proportion of correct species assignment was calculated, reported and sent to the Access® database for storage.

The collated results from the method above were then analysed to support and inform the design of the final approach proposed by this research.

4.2.3 Results

4.2.3.1 Culex pipiens s.s. *female versus* Culex torrentium *female exploration of discriminatory power using different numbers of landmarks*

4.2.3.1.1 Analysis of discriminatory power using the 18 landmark collection regime (as Figure 55)

After inspection of the female distribution of Procrustes distances from the mean curve, specimens 113 and 46 were considered outliers and removed from the analysis. Within the female only data, the Procrustes variance for *Culex pipiens s.s.* was 0.00144 and for *Culex torrentium* was 0.00115, giving an absolute difference of 0.00292 with a p-value of 0.0073; this significant result confirmed that there is sufficient difference between the groups in the data that it can be used for discriminatory purposes.

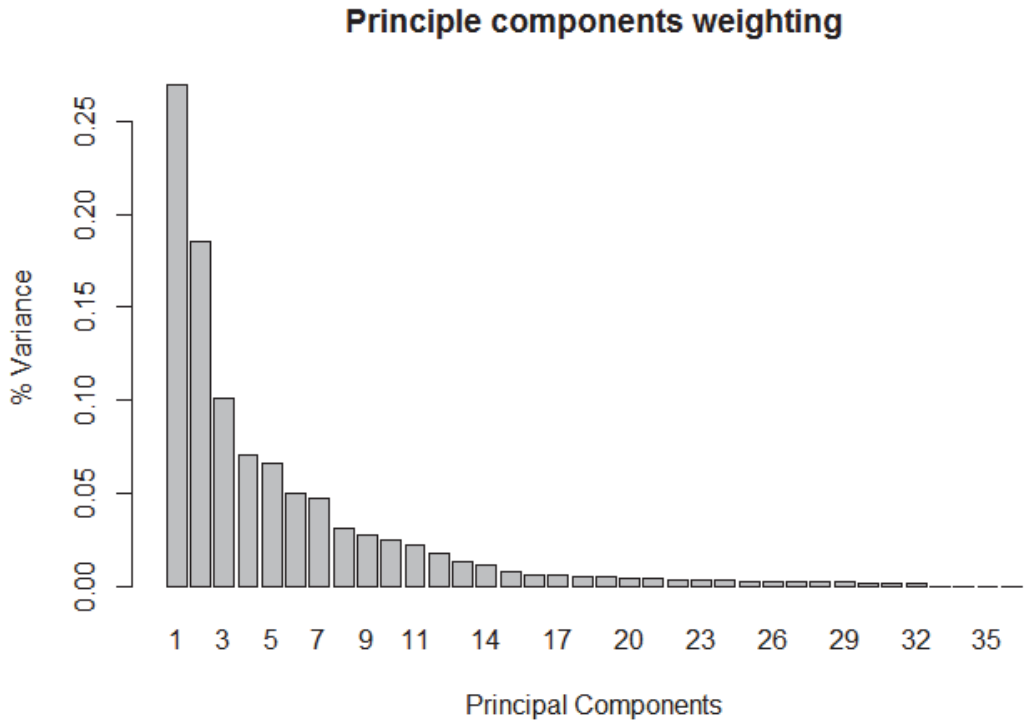


Figure 66 Principal component weighting for species discrimination, for female 18 landmark data. PC1 explains ~30% of the variance in the data, with multiple other PCs making significant contributions.

PCA showed that 16 components were required to explain 95 % of the total variance between groups, with PC1 only accounting for 30 % of the variance (Figure 66), which suggests that a multivariate approach was most likely to provide most discriminatory power. Cross-validated

DAPC retained 22 PCs with a mean successful species assignment of 94.90 % (Figure 67). Alpha score optimisation of the selection of PCs suggested the retention of only 1 PC. For the manually selected number of PCs, 9 PCs were chosen. This offered, subjectively, the best trade-off between variance explained and PC reduction; 9 PCs explained 84.7 % of the variance, only 10% less variance explained having excluded 7 PCs.

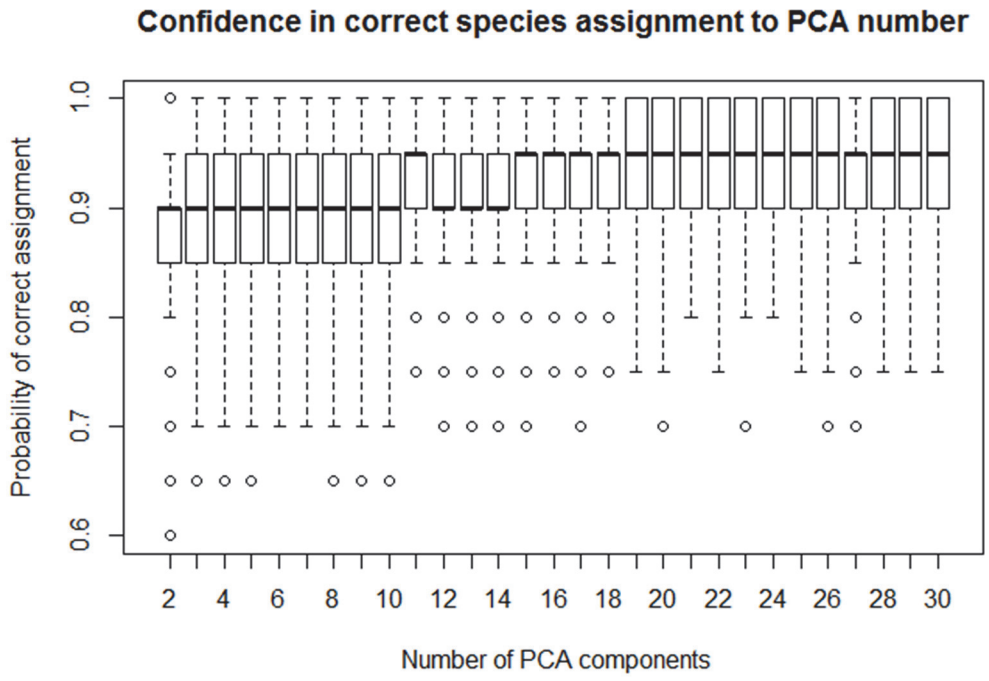


Figure 67 Correct species assignment, as calculated by cross-validated DPAC, for 2:30 PCs retained. (Here 22 PCs returned the best results with 94.90 % mean correct species assignment.)

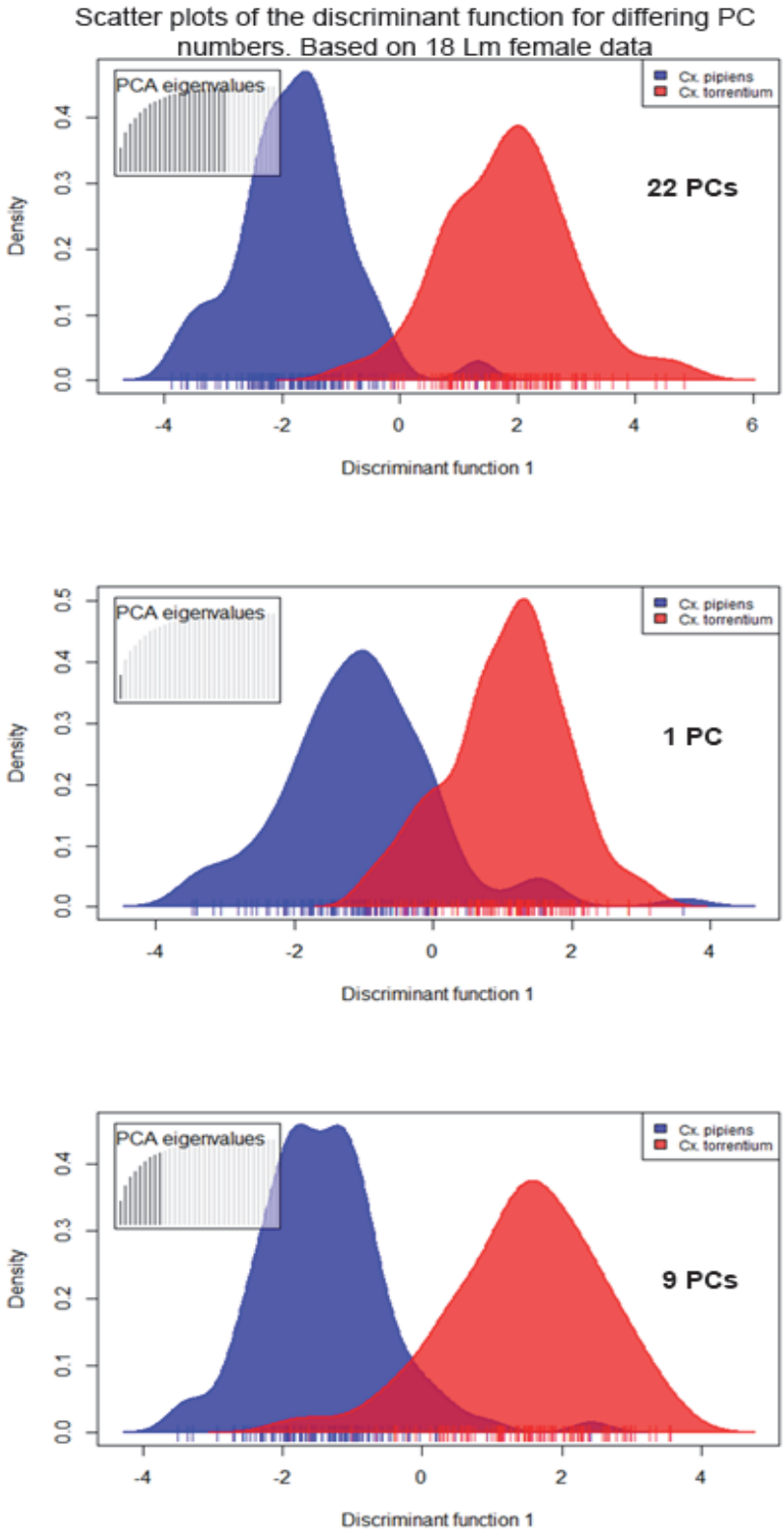


Figure 68 Comparison between the scatter plots of the species groups membership when using the different numbers of PCs, determined by xvalDAPC, alpha score optimisation and operator choice. Scree plot facets indicate proportion of data being using in analysis

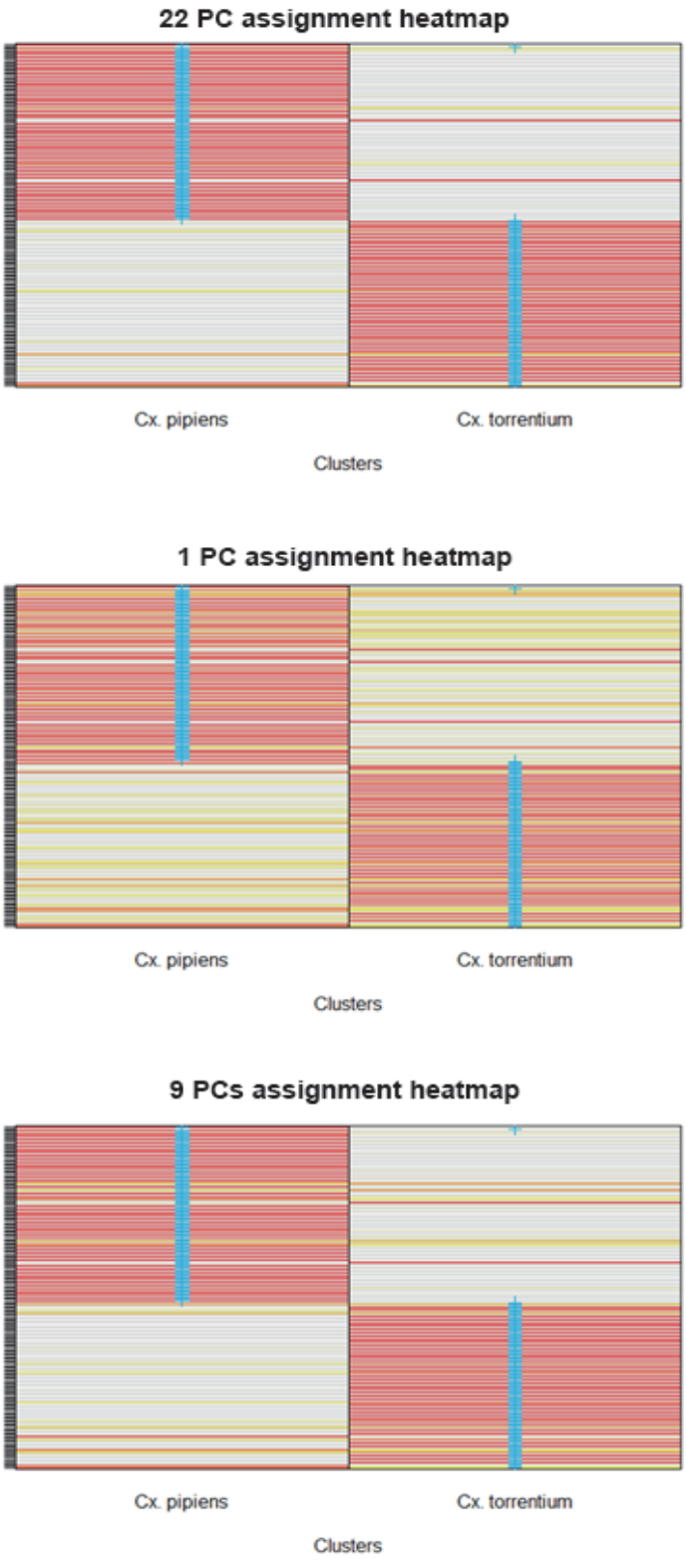


Figure 69 Species assignment heat maps for the 3 different numbers of PCs retained. Columns indicate the species. Blue crosses indicate prior, confirmed, ID for specimens. Darker regions = higher confidence of the identification by the analysis. 18 Landmark data were used to generate these heatmaps.

Figure 68 shows the comparison between the different numbers of PCs retained by the methods employed. Unsurprisingly, the higher the number of PCs the smaller the overlap between the species. A similar trend is displayed in Figure 69, which shows heat maps of the confidence in the group membership of the training data when the models were tested against the data. Considering Figure 68, Figure 69 and Table 17, it is apparent that the data do have the power to discriminate very well between these species when carrying out analysis within the training data itself.

Predictive power test results are displayed in Table 17 and show that the 9 PCs approach was the most accurate and precise trialled here. The "known unknown" mosquitoes used for this analysis consisted of 77 right wings for female mosquitoes, 40 *Culex pipiens s.s.* and 37 *Culex torrentium*.

Table 17 Accuracy and precision of the correct species identification of *Culex pipiens s.s.* and *Culex torrentium* mosquitoes using 18 landmarks, with different numbers of PCs retained for the analysis. The mean correct identification (training) results relate to the bootstrapped analysis within the training data, with a new random split of the dataset 90% training and 10% testing for each iteration. The mean correct identification (test) reports the accuracy of the identification of unknown specimens and is based on 15 iterations to capture any variation in the multivariate analysis.

Number of PCs used for the model	Mean % correct identification (training) Reps =1000	Mean % correct identification (test)	Standard Deviation	Max	Min
32 (all)	93.57	NA	NA	NA	NA
22	**94.89**	74.116	0.333077	75.32	74.03
1	50.00	58.44	0	58.44	58.44
9	90.82	**79.22**	0	79.22	79.22

4.2.3.1.2 Analysis of discriminatory power using 13 landmark collection regime *(as Börstler et al. (2014))*

After inspection of the female distribution of Procrustes distances from the mean curve, specimens 113 and 46 were removed from the analysis. Within the female only data, the Procrustes variance for *Culex pipiens s.s.* was 0.00130 and for *Culex torrentium* was 0.00106, giving an absolute difference of 0.00240 with a p-value of 0.0281; this significant result confirmed that there is sufficient difference between the groups in the data that it can be used for discriminatory purposes.

PCA showed that 12 components were required to explain 95 % of the total variance between species, with PC1 only accounting for 33.95 % of the variance (Figure 71), which suggests that a multivariate approach was most likely to provide most discriminatory power. Cross-validated DAPC retained 16 PCs with a mean successful species assignment of 94.49 % (Figure 70).

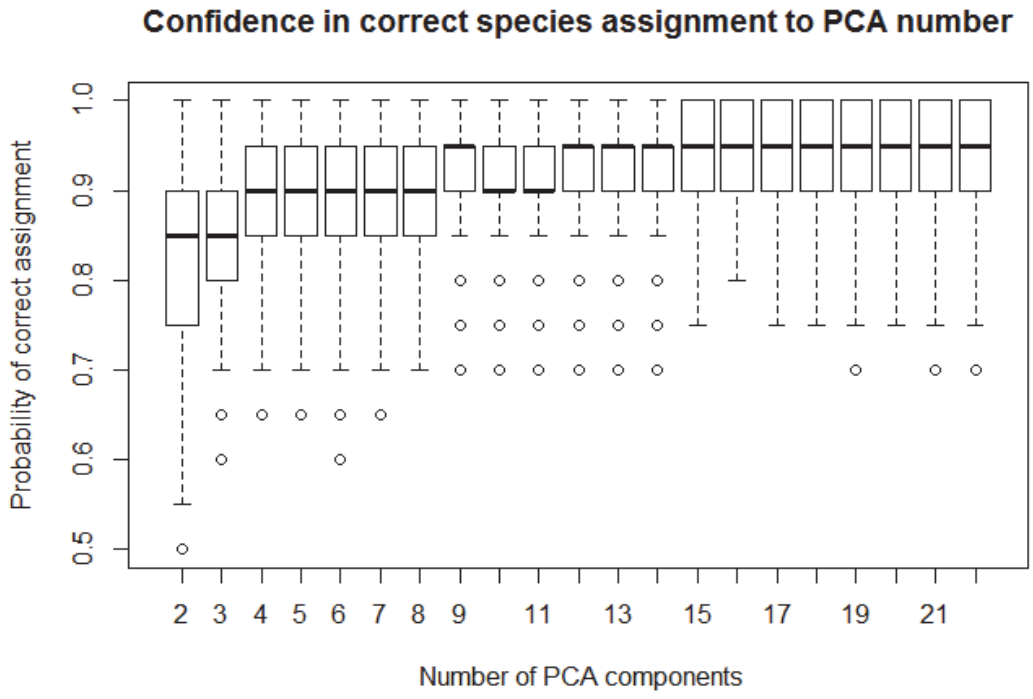


Figure 70 Correct species assignment, as calculated by cross-validated DPAC, for 2:30 PCs retained. Here 16 PCs returned the best results with 94.49 % mean correct species assignment.

Alpha score optimisation of the selection of PCs suggested the retention of 14 PCs. For the manually selected number of PCs, 9 PCs were chosen. This offered, subjectively, the best

trade-off between variance explained and PC reduction; 9 PCs explained 91.37 % of the variance, compared with 98.14 % explained by 16 PCs.

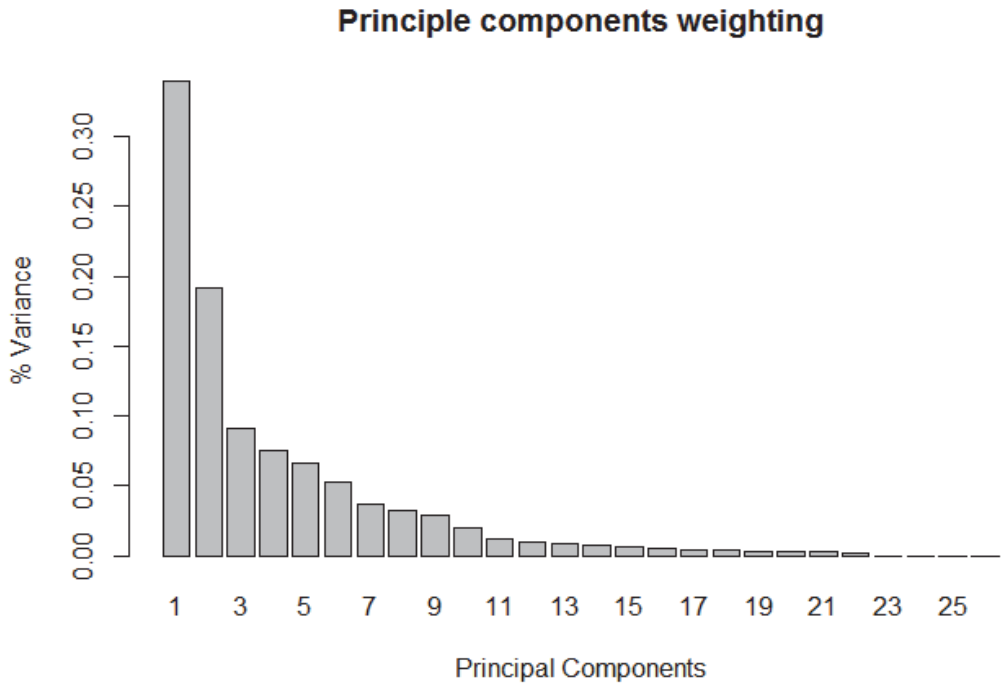


Figure 71 Principal component weighting for species discrimination, for female 13 landmark data. PC1 explains 33.95 % of the variance in the data, with multiple other PCs making significant contributions.

Figure 72 shows the comparison between the different numbers of PCs retained by the methods employed. Unsurprisingly, the higher the number of PCs the smaller the overlap between the species. A similar trend is displayed in Figure 73, which shows the confidence in the group membership of the training data when the model was tested against the data. Considering Figure 72, Figure 73 and Table 18, it is apparent that the data do have the power to discriminate very well between these species when carrying out analysis within the training data itself.

Predictive power test results are displayed in Table 18 and show that the 9 PCs approach was the most accurate and precise trialled here. The “known unknown” mosquitoes used for this analysis consisted of 77 right wings for female mosquitoes, 40 *Culex pipiens s.s.* and 37 *Culex torrentium*. Comparison between Table 17 and Table 18 shows that the predictive power of the model is actually highest when using the 13 landmark data with 9 PCs; This 13 landmark model explained the species difference in the training data less well than the 18 landmark

model, but in this applied setting the overfitted model appears to lack the stability required for predictive purposes.

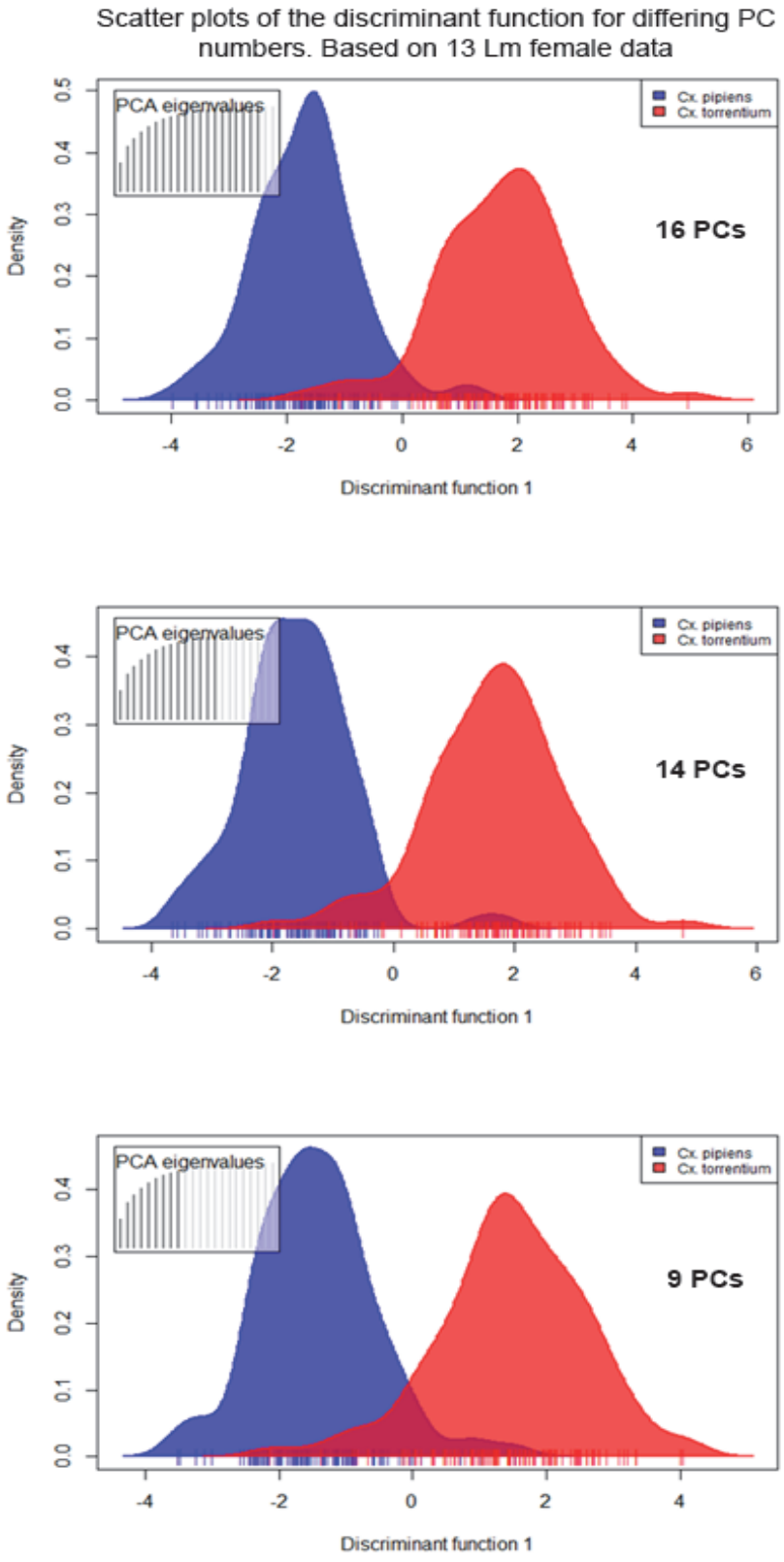
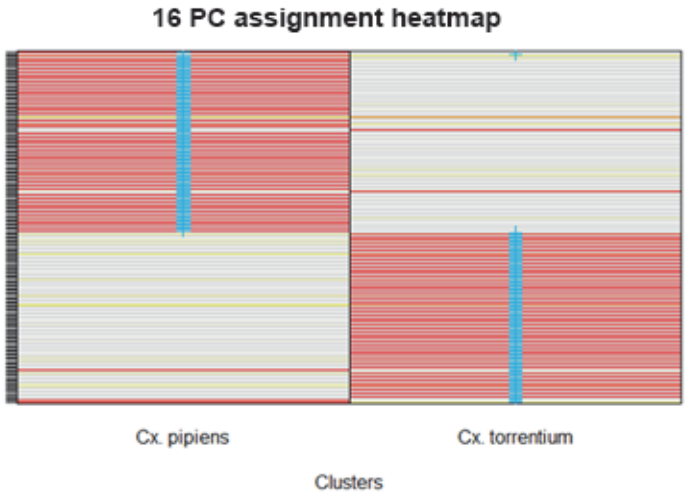
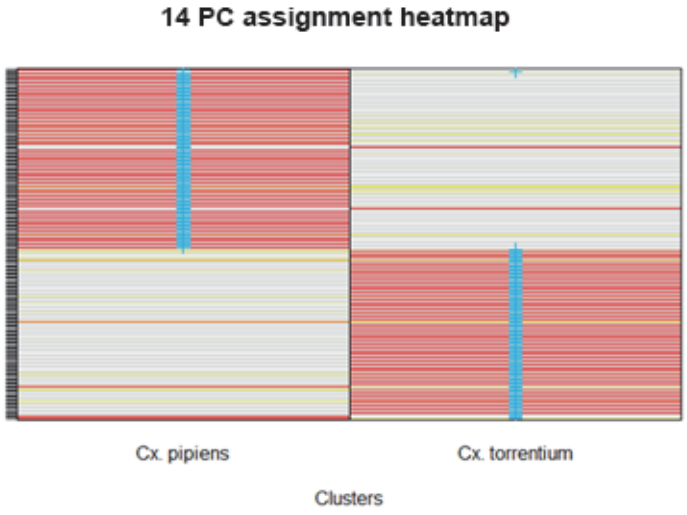


Figure 72 Comparison between the scatter plots of the species groups membership when using the different numbers of PCs, determined by xvalDAPC, alpha score optimisation and operator choice. Scree plot facets indicate proportion of data being using in analysis





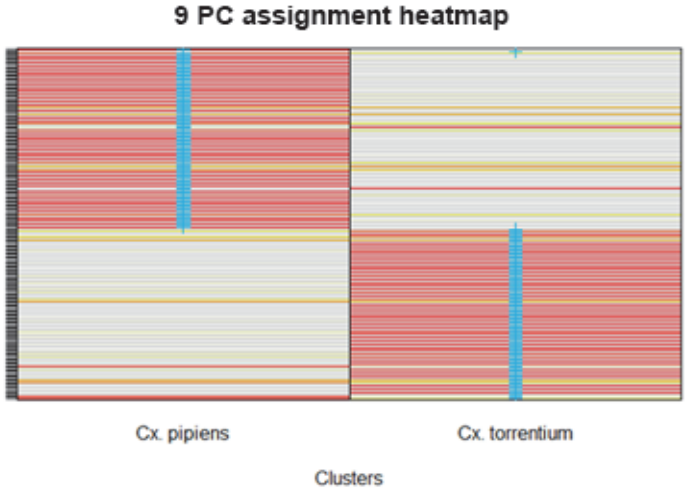


Figure 73 Species assignment heat maps for the 3 different numbers of PCs retained. Columns indicate the species. Blue crosses indicate prior, confirmed, ID for specimens. Darker regions = higher confidence of the identification by the analysis. 13 Landmark data were used to generate these heatmaps.

Table 18 Accuracy and precision of the correct species identification of *Culex pipiens s.s.* and *Culex torrentium* mosquitoes using 13 landmarks, with different numbers of PCs retained for the analysis. The mean correct identification (training) results relate to the bootstrapped analysis within the training data, with a new random split of the dataset 90% training and 10% testing for each iteration. The mean correct identification (test) reports the accuracy of the identification of unknown specimens and is based on 15 iterations to capture any variation in the multivariate analysis.

Number of PCs used for the model	Mean % correct identification (training) reps=1000	Mean % correct identification (test)	Standard Deviation	Max	Min
22 (all)	93.49	NA	NA	NA	NA
16	**94.38**	68.83	0	68.83	68.83
14	94.11	70.13	0	70.13	70.13
9	91.92	**84.42**	0	84.42	84.42

4.2.3.1.3 Method optimisation by targeted landmark retention

Having established that model overfit might result in reduced predictive power, attention was turned to reducing the sampling effort by reducing the number of landmarks requiring collection for each specimen. Using 9 PCs, selected from the 13-landmark data, was shown to be the most accurate and precise method for correctly identifying the unknown mosquitoes. The loadings of these PCs were analysed to ascertain which landmarks were the most discriminatory (Figure 74). The landmarks which contributed most to the discrimination between species were, in descending order of contribution: LM18, LM9, LM8, LM17, LM7, LM3 and LM2.

Having identified those landmarks which contain the data required for discrimination, the analyses were carried out again, this time retaining only the seven landmarks which contributed to the 9 PCs. These seven landmarks and the distribution of the data about the mean for each landmark is shown in Figure 75.

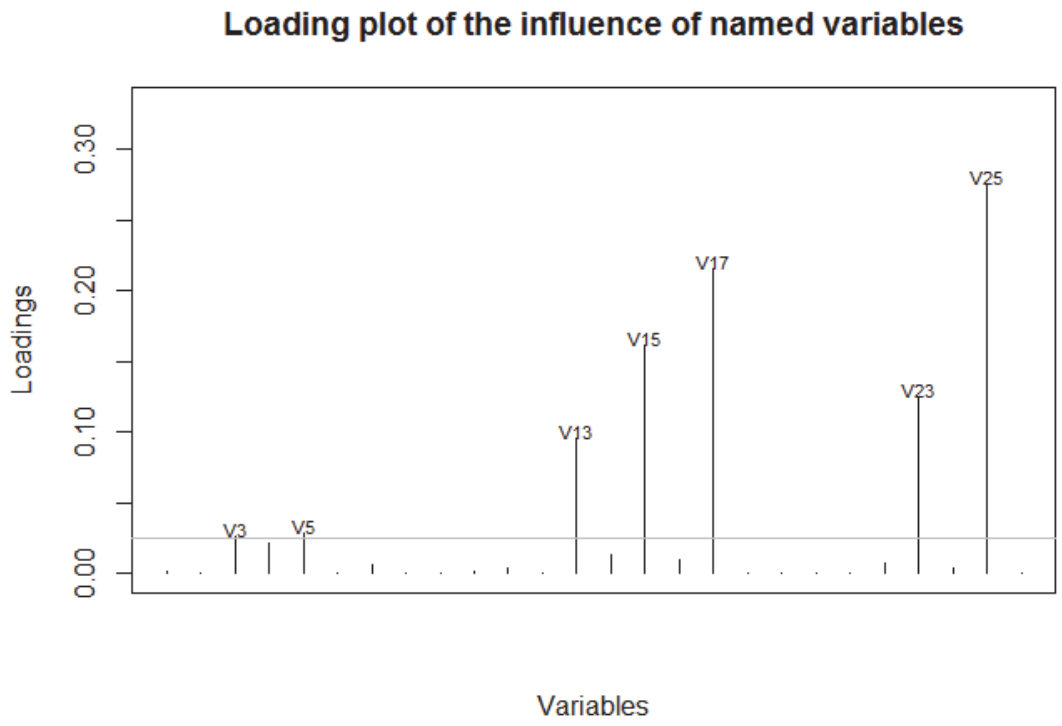


Figure 74 Loadings for the specific landmarks which have the most influence on discriminatory power. In descending order, they relate to LM18, LM9, LM8, LM17, LM7, LM3 and LM2. By plotting only these landmarks, it should be possible to retain all the discriminatory power of this method.

Within this female only data, where only 7 landmarks were collected, Procrustes variance for *Culex pipiens s.s.* was 0.00240 and for *Culex torrentium* was 0.00181 giving an absolute difference of 0.00059 with a p-value of 0.0101; this significant result confirmed that there was still sufficient difference between the species in this reduced data set such that it could be used for discriminatory purposes.

The PCA tangent plot (Figure 76) shows considerable species separation on PC1, accounting for 44.5 % of the variance between species. On PC2 the separation is less visually apparent, even though this accounted for 27.8 % of the variance between species. The 9 PCs retained for the analysis account for 99.3 % of the variance between species.

The thin plate spline transformation of exemplars of each species to the population mean show subtle differences between the species, as evidenced by the relatively small distortion of the grid upon which it is projected (Figure 77).

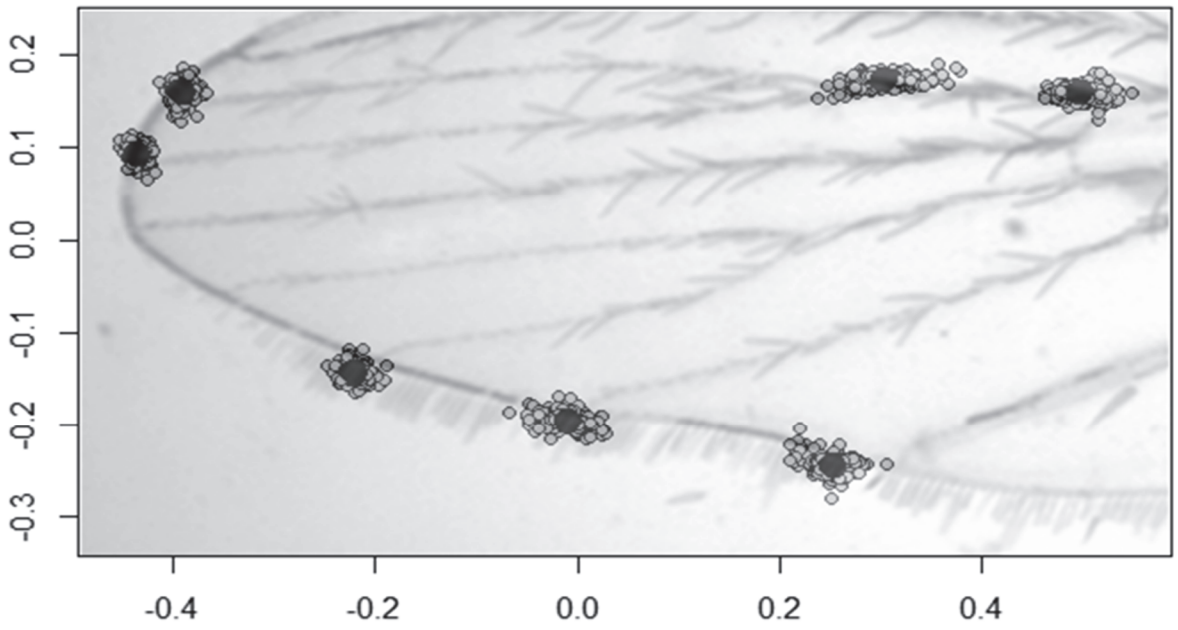


Figure 75 Using these 7 Landmarks retained all the discriminatory power of the model. Thereby requiring the collection of fewer landmarks by the user to achieve the same outcome.

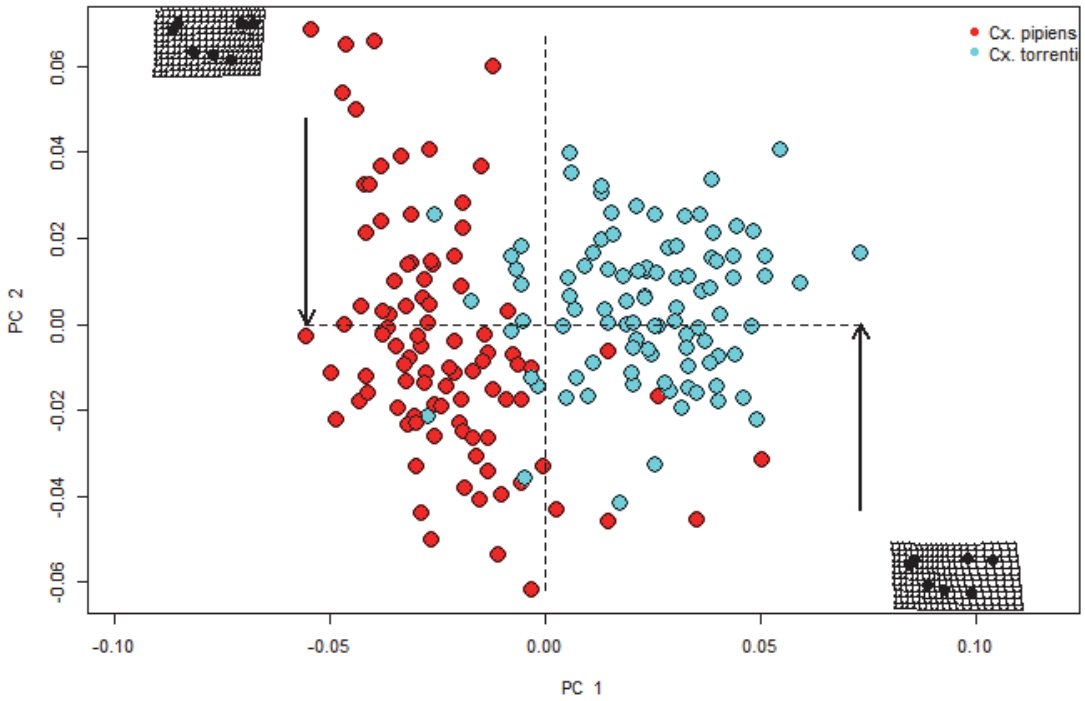


Figure 76 For the 7 Landmark data, there is still a visible separation between the *Culex pipiens s.s.* and *Culex torrentium* by PC1. The influence of PC2 is less obvious but still, account for 27.8 % of the variance between species.

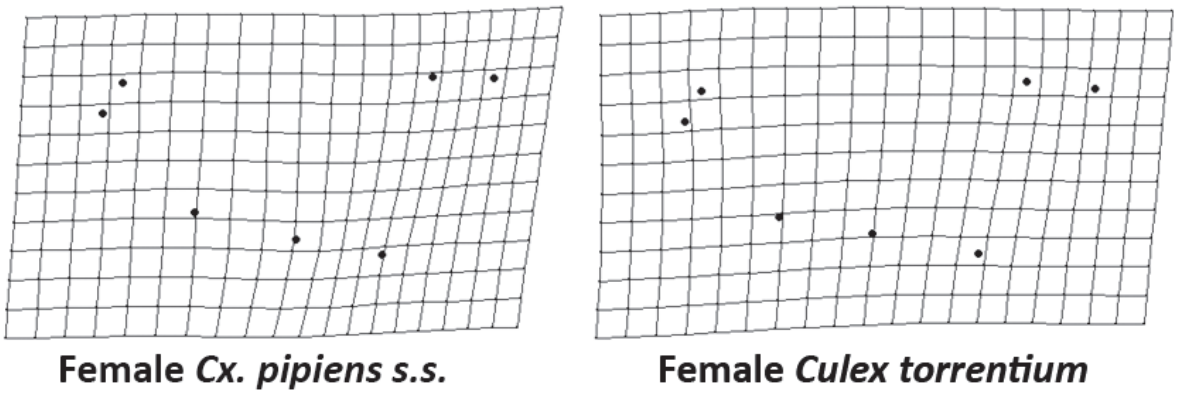


Figure 77 Thin Plate Spline transformation plot of the two species shows that differences from the mean consensus landmark positions are quite subtle and suggest a multivariate approach as no one feature is sufficiently different to be discriminatory.

Table 19 Accuracy and precision of the correct species identification of *Culex pipiens s.s.* and *Culex torrentium* mosquitoes using 7 landmarks, with different numbers of PCs retained by the analysis. The mean correct identification (training) results relate to the bootstrapped analysis within the training data, with a new random split of the dataset 90% training and 10% testing for each iteration. The mean correct identification (test) reports the accuracy of the identification of unknown specimens and is based on 15 iterations to capture any variation in the multivariate analysis.

Number of PCs used for the model	Mean % correct identification (training) reps=1000	Mean % correct identification (test)	Standard Deviation	Max	Min
9	**92.34**	83.12	0	83.12	83.12
8	91.84	83.12	0	83.12	83.12
7	92.05	**84.42**	0	84.42	84.42
6	91.75	83.12	0	83.12	83.12
5	91.48	84.42	0	84.42	84.42
4	89.69	84.42	0	84.42	84.42
3	89.39	81.82	0	81.82	81.82

Following permutation testing, it was found that when using all the landmark data in the 7-landmark treatment, a mean correct identification rate of 84.42 % was achieved using several different PC retention regimes. This matches the best mean correct identification rate from the prior method using 13 landmarks with principal components analysis to reduce the overfitting of the discriminatory model. Although this result is referred to as the mean correct identification, as it is based on 15 repeats, the precision was such that there was no deviation from the mean in any of the repetitions (Table 19). 7 PCs offered the best balance of fit of the identification model to the training data and predictive power when applied to the "known unknowns", and was, therefore, selected to be used in the optimised approach.

4.2.4 Proposed tool for the wing morphometric discrimination of female *Culex pipiens s.s.* and *Culex torrentium*

Before moving on to this development, a new training dataset was created by augmenting the previous training data with the addition of the, until now, separate control specimen data. As previously discussed, the addition of new specimens will result in an increased representation of the wild population and enhance the veracity of the results returned.

Based on the prior findings, a new tool has been coded in R to facilitate the discrimination of female *Culex pipiens s.s.* and *Culex torrentium* specimens. This tool is titled 'Final ID code.Rmd'. It is predominantly unsupervised, after setting starting conditions and initial run to visualise the training data and remove any outliers as necessary. After initial configuration, only the SQL query for the new specimens to be identified needs to be changed. The output from the tool can take the form of .html, Word document or .pdf facilitating the communication and collation of results.

4.2.4.1.1 Describing the editable regions of the tool 'Final ID code.Rmd'

The title, author and date at the top of the document can be amended to keep track of when identifications were carried out.

verbose - Enter the value TRUE for full output. This includes descriptive statistics for the training data being utilised, including how well it fits the data. A value of FALSE returns the identification of unknown specimens only.

con <- odbcConnect("EDIT") - Replace EDIT with the name of the ODBC identifier for the database connection on the computer being used.

remoutliers <- -c(x, y) - Replace x,y with the numbers of any outliers which need to be removed from the training data. This only needs to be changed after the first verbose = TRUE run of the code when the training data have been inspected.

qry1 <-" SQL query here" Enter the SQL query for the training data

qry2 <-" SQL query here" Enter the SQL query for the specimens to be identified

When using 7 landmarks and 7 PCs no further editing needs to be carried out.

4.2.4.1.2 Proposed identification workflow using discriminatory software

Unknown specimens should have slides of their wings prepared and imaged as previously described. Landmarks, as in Figure 78, should then be collected using the CLIC package of Mome-Clic, selecting the option to create a "_DB.txt" upon exit; these should then be imported into a new Access® database table.

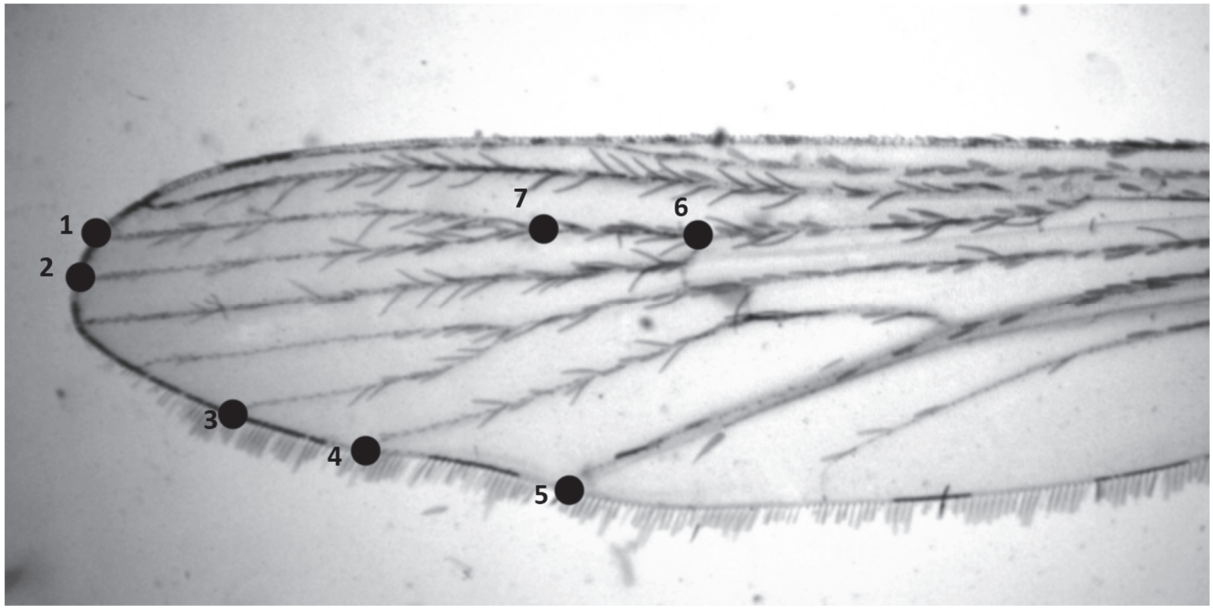


Figure 78 These seven landmarks should be plotted in this order when they are being collected. These 7 landmarks carried sufficient data to diagnose *Culex pipiens s.s.* or *Culex torrentium*, with >84 % accuracy as an identification tool when tested on new specimens

Open the file 'Final ID code.Rmd' in RStudio (RStudio Team 2015) and update the editable variables outlined above. For the first run with a training data set the presence of outliers will not be known, so remoutliers should read remoutliers <- -c() and verbose = TRUE. Once the output has been observed, then the user may choose to accept the report as is or enter the numerical value/values of outliers from the outlier plot. Verbose can now be set to FALSE and the code run once more, this time returning only the specimen identification output based on the training data. Provided that the training data are not changed, this code can be used for all subsequent identifications in a non-verbose mode requiring only the update of the SQL code for the new unknown specimens. An example of the output from the tool in "verbose mode" is included in Appendix 3, and the non-verbose mode identification output is available in Appendix 4.

4.2.5 Discussion

Through this analysis, it has been possible to evaluate and justify the use of shape-based wing morphometric analysis, and subsequently refine the method of, the diagnosis of species for British *Culex pipiens s.s.* and *Culex torrentium*. A time reduction was conferred by optimisation of the method requiring the collection of only 7 landmarks to achieve >84 % accurate species discrimination, equal to the results of the best alternative method tested here based on the digitisation of the 13 landmarks per wing proposed by Börstler *et al*. (2014). Although ultimately, this study did not achieve the >97 % discriminatory power seen by Börstler *et al.* (2014) with mosquitoes from the Germany within their training data, the difference appears to be related to differences in the respective populations (Service 1968; Börstler *et al.* 2014) and not due to any weakness in the method applied. It was observed that the model that was most accurate within the training data set did not necessarily yield the best model when used predictively with new specimen's data due to model overfitting (Jombart and Devillard 2010). Indeed, the most powerful model for correctly predicting unknown specimen's species was not the most powerful when tested within the training data alone. The most powerful model was based on a seven landmark collection regime and achieved 84.42 % accurate discrimination between unknown *Culex pipiens s.s.* and *Culex torrentium* specimens and 92.05 % accuracy within the training data, using 7 of the 14 available principal components in the model. Conversely, the most powerful model in terms of discrimination between species within the training data achieved 94.89 % accuracy, but only 74.12 % accuracy and introduced imprecision (0.33 % SD) when tested with unknown specimens, using the 18 landmark regime and 22 PCs of 36.

As stated, the shape-based wing morphometric method optimised here achieved > 84 % accuracy for unknown specimens and therefore it strongly supports the use of multivariate wing morphometrics to discriminate between female *Culex pipiens s.s.* and *Culex torrentium* from the British population. The use of this approach in ecological surveys will help reduce the chronic underreporting of *Culex torrentium* (discussed further in chapter 3) and reduce the reliance on the use of *Culex pipiens* s.l. to describe the females from this complex in the UK. The wing morphometric analysis tool developed here would be particularly useful when used to triage collections to determine where further effort should be applied to confirmatory species identification methods such as rearing-on of juveniles to imago for morphological

species identification of male terminalia using keys (Becker *et al.* 2010b), or PCR methods (Smith and Fonseca 2004).

The PCR method applied to the specimens used in this study showed 100 % agreement with the morphological identifications made of male *Culex pipiens s.s.* and *Culex torrentium* specimens, indicating 100% specificity of the primers designed by Smith and Fonseca (2004) within the population sampled; therefore that molecular method still represented the gold standard for discrimination between female *Culex pipiens s.s.* and *Culex torrentium*.

Despite their power, PCR based approaches are not available to all and do have a relatively high cost. In this study the sample handling and analysis time requirement was very similar for both wing morphometrics and PCR methods and as such the cost of this is not compared here. In a laboratory that already has the required equipment, the PCR method using Chelex based extraction and multiplex primers (Smith and Fonseca 2004) in this study cost £0.73 per sample based on 20 lane gel electrophoresis with all lanes utilised, an additional £0.69 per sample would be added when DNA quantification was required. The wing morphometric technique consumables used for slide preparation cost £0.11 per sample showing a significant per sample saving of up to £1.32 per sample. The cost incurred to purchase the required apparatus to conduct PCR was £11133, based on the purchase of entry level equipment from in the UK (Fischer-Scientific 2018). In contrast, to purchase both a dissecting microscope, and compound microscope with a digital camera attached would cost £1300 (Cole-Parmer 2018). Including equipment purchasing costs and based on the processing of 1000 samples, the PCR cost would be £12.32 per sample (based on 50 samples with DNA quantification as part of quality assurance) and the wing morphometric method would cost £1.41 per sample offering a cost efficiency.

Having access to a quick, low cost, identification method to discriminate between these two potential important vector species (Lundström *et al.* 1990; Hesson, Verner-Carlsson, *et al.* 2015; Leggewie *et al.* 2016) in the UK will help increase the understanding of the abundance of and interaction between the distributions of them both. When compared with the relative cost and complexity of molecular methods of identification, there is no doubt that shape analysis of wing morphometric features provides a useful and much more accessible tool for those without access to molecular means; even those with access to molecular methods will

find this method quicker and cheaper. Although permanent slide mounted wings and microscopes were used to capture the wing images in this study, any means of collecting good quality images of flattened mosquito wings would be equally viable. Whilst the use of an office flatbed scanner did not provide the image resolution required for accurate landmark collection, other options are viable alternatives; for example macro photography using a regular camera or a smartphone with a clip-on macro lens (Rousseau 2016) do provide sufficient detail, from personal experience capturing mosquito images in the field (Appendix 5) and as proposed in other fields (Myung *et al.* 2014). To increase the user friendliness of the proposed identification approach, a simplified procedure was developed using the best practice established in this study. The intention is that this method, and the wing image/landmark database that supports it, should be made available to other British entomologists and will hopefully facilitate a better understanding of the distribution of *Culex pipiens s.s.* and *Culex torrentium* by enabling easier specimen identification.

The identification tool presented here makes the post-processing of species identification very straightforward. It requires fewer landmarks to be collected for an equally accurate species diagnosis, thereby saving time and reducing the opportunity for measurement error caused by extended periods of landmark collection by the operator. However, because the tool uses SQL to import landmark data from a database, and because the order of the commands in the SQL query determines the column order of the resultant data table in R, data that were collected using other configurations can be utilised by any code savvy operator, provided that the 7 specific landmarks required by this tool are represented somewhere in their data. For example, in the exemplar (Appendix 3), the data that was imported from the database for the training data and the unknown specimens are from the original 18 landmark collection configuration, reordered to match the new 7 landmark configuration. This flexibility is necessary as it obviates the need to recollect landmark data where it already exists.

When considering the process used to optimise the method, it was observed that having selected the number of PCs using cross-validated discriminant analysis of principal components, which bootstraps the selection of PCs by randomly splitting the training data into 90 % training and 10 % control groups 10000 times (Jombart and Devillard 2010), the number of principal components returned tended to overfit the data (James *et al.* 2013) when

the model was run on unknown specimens. This resulted in reduced ability to discriminate between species accurately when tasked with the predicting species membership of individuals within new data. This reinforced the importance of testing discriminatory models against new 'known unknown' data which was not part of the training dataset, and not simply relying on finding the best fit for the training data whenever possible. The approach tested here relies upon the training data being representative of the wild populations of *Culex pipiens s.s.* and *Culex torrentium* from which they were sampled. All mosquitoes were either wild caught adults or reared on from wild collected eggs; therefore, these do not represent lab adjusted strains and can be considered as genuine representatives of the wild population. As with all model based approaches, adding more wild specimens' wing data to the training dataset will further increase the representativeness of the model, a process which could theoretically continue until the intra-species landmark variance stabilises, at which point that diminishing returns would result from further specimen addition. The inclusion of both wings in the training data set acted as a reinforcement of the data and because there was seen to be no significant difference between right and left wings could be thought of as a pseudo-replication of the training data; helping to reduce the measurement error which is inevitable even when a single operator is responsible for all collections (Dujardin *et al.* 2010). Collecting data for both wings effectively halved the possible influence of any single measurement error by the user, when collecting landmark data, and should be encouraged where possible (Dujardin 2011).

The analysis here supports Service's (1968) findings that significant overlap in wing vein ratios, R2+3:R2 or R2+3:R3, between these species in Britain, and suggest more similar wing morphology to Russian specimens (Vinogradova and Shaikevich 2007) than German ones (Börstler *et al.* 2014). When considering the apparent lack of species discriminating power of these wing vein ratios, there is potential that effects associated with gender are influencing matters. The most influential wing feature in the apparent, albeit nonsignificant, difference genders was the horizontal point of intersection between veins R2+3, R2 and R3; therefore, any gender related variation of this location will directly influence the ratio of these vein lengths potentially mask species related variation

It was interesting that there were several specimens within the data which defied correct classification by wing morphometric methods. Having checked that the prior identification was accurate and appropriately recorded, these specimens were apparently genuine outliers to the population sample being used. These non-conforming specimens tended to be much more similar to the opposite species than their own, and would not, therefore, be correctly identified by this method even with a large training data set as the tails of the species distributions overlap, even with multiple PCs being applied. Börstler *et al.* (2014) reported finding similar 'aberrant' specimens within their central European mosquito population; a population with a greater disparity between wing features than that of the British mosquito population examined here. Due to the identification error declared for this approach (15.58 %) when applied to British mosquitoes, it will be more reliable when applied to more than one individual to increase confidence of species representation within a sample from a locale, a similar approach is suggested for bee identification when considering cryptic sub-species discriminations (Meixner *et al.* 2013).

As an operational note for the use of the identification tool, because the tool uses a whole data GPA with partial Procrustes superimposition, it would be good practice to restrict the number of unknown specimens being processed in any single processing run. If a substantial number of unknown specimens were processed at one time, then they may skew the consensus configuration far enough from the consensus of the training data alone, and therefore could be somewhat unpredictable (Dujardin *et al.* 2010). In ad-hoc tests of the tool, groups of twenty specimens generated identical classification results to individual identifications. As the training data set grows, becoming more robust, this effect will become less pronounced, due to the proportion of new data to training data, and increased numbers of specimens could be processed at one time.

4.2.6 Conclusion

The identification of British female *Culex pipiens s.s.* and *Culex torrentium* was > 84 % accurate using the described shape based wing morphometric approach and is therefore recommended for use with for these species. However, bivariate and univariate approaches were not nearly as accurate (< 70 %), with significant overlap in the wing vein R2+3/R3 ratio between the species, and, therefore, their use was not supported by this study's findings.

The multivariate identification tool created in this study is designed to make the identification of specimens as simple as possible, requiring a little equipment as possible; certainly, far less than that required for PCR identification. Even where entomologists have access to PCR methods, cost concerns and resource availability might result in only a subsample being identified using PCR, this then being extrapolated to the whole sample and subsequently the population. Wing morphometric identification methods can be applied to whole samples for a fraction of the cost of molecular methods and require fewer new skills to be learned by the operator.

It is hoped that this method and the identification tool can help remove cost and equipment access barriers which have led to the paucity of data regarding the actual distribution of *Culex pipiens s.s.* and *Culex torrentium*. It has proved suitable for the inclusion as an identification method in the field work chapter of this thesis.

5 Olfactory mediated behaviour of *Culex pipiens* and *Culex torrentium*

5.1 Introduction

Within the knowledge base related to mosquitoes, there is a broad variation in the level of the known, the putative, the assumed and the unknown. Some research areas are, relatively, straightforward to research and offer new knowledge, with high confidence; investigating the presence of a mosquito in a region is often a case of applying appropriate resources to the problem and declaring a digital response. However, once the research aim shifts through declarative, to descriptive, and ultimately causal questions (Odom and Bell 2011) then the resources required to reach similar confidence increase dramatically and reduces the amount of generalisation which can be applied confidently to research outcomes. Research into mosquito behaviour requires research input to address all these question types and whilst there are many well-understood behaviours, it continues to be an area of very active research, particularly into the causal factors of observed behaviour. These areas are of significant interest as many of our current, and potential future, mosquito control tools (Becker *et al.* 2010a; Cameron and Lorenz 2013) exploit mosquito behaviour such that mosquitoes can be caused to assist in their own demise.

In insects, olfaction is their most important sense (Yin *et al.* 2015), and having an acute olfactory system is critical to their reproduction as it has a significant impact on their ability to find food, mates and oviposition sites (Leal 2012). Olfactory mediated behaviours in insects, in general, are critical to understanding the interactions between them and their environment; developing this understanding is more pressing when considering insects which directly affect our lives as pollinators, agricultural pests or as vectors of disease (Leal 2013). Understanding the importance of olfaction to the success of insects in general and mosquitoes specifically, has driven much research in these fields and has resulted in the development of a significant body of literature. In this section, the interactions between male *Culex pipiens s.s.* and *Culex torrentium* mosquitoes and floral nectar sources are examined in the laboratory.

5.1.1.1 Sugar feeding in mosquitoes

Adult mosquitoes of both sexes require sugars which they feed on from various sources including flowers and extrafloral nectaries of plants, fruit and honeydew (secretion from aphids) (Foster 1995a; Müller *et al.* 2011). Despite the ubiquitous need to utilise this energy source, this area of behaviour has traditionally been less studied than behaviours related to finding blood meal hosts. Mosquitoes species have been shown to exhibit differential preference for certain sugar sources (Mauer and Rowley 1999; Manda *et al.* 2007; Chen and Kearney 2015; Ding *et al.* 2016) and that feeding on preferred sugar sources conveys survival and fecundity advantages over less preferential sugars (Manda *et al.* 2007). This level of discrimination is similar to that exhibited towards blood meal host species preference (Service 1971a; Takken and Knols 1999; Cooperband *et al.* 2008; Takken and Verhulst 2013), and also preference towards certain individuals within species (Kelly 2001; Logan 2008). Preferential selection of blood meal host selection is also shown to increase fecundity in mosquitoes (Richards *et al.* 2012). Sugar feeding on preferred sources, as with blood feeding, therefore directly contributes to survival and the inheritance of genetic information, driving further evolution of these behaviours. There is also evidence of gender based differential selection of sugar sources in the field (Grimstad and DeFoliart 1974) and in laboratory experiments using *Culex pipiens pallens* (Ding *et al.* 2016), although in laboratory experiments no such gender differential was reported in *Aedes aegypti* (von Oppen *et al.* 2015). It is possible that as male *Aedes aegypti* do not form mating swarms instead mating close to blood meal hosts (Hartberg 1971), and so do not have such high energetic demands of sustained flight (Nayar and Van Handel 1971) associated with swarming (Yuval *et al.* 1994) as *Culex pipiens pallens* that do form mating swarms.

Male mosquitoes feed on sugar sources throughout their lives (Foster 1995a), although different species feed at various times during the diel cycle depending upon the species (Yee and Foster 1992; Yee *et al.* 1992). Sugar feeding occurs during the typical active hours for a species, therefore day flying species such as *Aedes albopictus* sugar feed diurnally (Revay *et al.* 2014) and nocturnally active species sugar feed during scotophase and/or in crepuscular conditions (Yee and Foster 1992; Yee *et al.* 1992; Gary and Foster 2006; Rund *et al.* 2016). It

is suggested that, in females, sugar feeding is very common soon after imago eclosion (Haramis and Foster 1990; Foster 1995a) and that they rarely blood feed until after a sugar meal (Bowen 1992a). Subsequently, female sugar feeding may decline, in favour of blood feeding following mating, but may reoccur in multivoltine species between gonotrophic cycles (Foster 1995a), although more recent studies suggest that in certain conditions sugar feeding may continue unabated (Gary and Foster 2006). Males have been demonstrated to feed on sugar significantly more often than females in experiments by Gary and Foster (2006) using *Anopheles gambiae* mosquitoes, they found that males sugar fed twice daily on average and females only needing to sugar feed every four days if they had access to blood meals; interestingly, in the absence of blood meals they would feed daily on sugar (Gary and Foster 2006). The necessity of the higher sugar feeding rate of male mosquitoes is largely explained by the energy expenditure associated with nightly swarming for mating purposes; male swarming is has been shown to consume over 50 % of a male's available calories (Yuval *et al.* 1994).

The different ways that male and female mosquitoes utilise energy gathered from sugar feeding activity have dramatic impacts on their relative longevity. The ability of female mosquitoes to convert sugars into triglyceride fat for storage enabled extended life compared to the males that cannot carry out this conversion (Van Handel 1984). The ability to build up fat reserves also facilitated winter diapause for those mosquitoes of temperate climes, such as *Culex pipiens s.s.* and *Culex torrentium*, which overwinter as females (Snow 1990). This ability potentially has a role in providing an overwintering reservoir for disease, and viruses have been recovered from females in winter diapause (Bailey *et al.* 1978; Nasci *et al.* 2001). However, typically, *Culex* mosquitoes that enter diapause after taking a blood meal or parous females are less likely to survive the winter than those which do not (Jaenson 1987). Indeed, Jaenson (1987) gathered field data that firmly supports the hypothesis that female *Culex pipiens* and *Culex torrentium* mosquitoes mainly or only survived the winter if they entered diapause as nulliparous, inseminated and non-bloodfed. *Culex* mosquitoes which complete their juvenile life stages in late summer or early autumn, respond to the shortening day length by feeding solely on sugars after adult emergence to build sufficient fat reserves to be able to overwinter in diapause (Spielman and Wong 1973; Takken and Knols 2007). In *Culex pipiens,* this is caused by the downregulation of genes that encode enzymes which enable the

digestion of blood meals, and the upregulation of a gene associated with the accumulation of lipid (Robich and Denlinger 2005). Towards the end of diapause, the genes for blood meal digestion begin to be expressed once more, enabling immediate blood feeding upon the resumption of post diapause activity (Robich and Denlinger 2005).

Male mosquitoes' utilisation of sugar is rather more simple because as adults they cannot store the energy as fat that was obtained by sugar feeding (Van Handel 1984), and due to their high energy utilisation are required to feed often. Once ingested sugar meals are stored in the mosquito's crop, and is immediately available as an energy source for flight(Foster 1995a). Most nectars consist of a mixture of glucose, fructose and sucrose, with some other trace level complex sugars (Van Handel 1972). These sugars are rapidly converted to glycogen by both sexes and concurrently converted to lipids by females (Foster 1995a). The main sugar in mosquito haemolymph is trehalose, a disaccharide, the level of which is maintained by the conversion of glycogen.

There is very little literature which refers directly to the characteristics of male sugar source location behaviour. However, given the importance and regularity of sugar feeding by male mosquitoes discussed above there must be a robust mechanism which allows individual males to locate sources of sugar. Yuva*l et al*. (1994) found that, for smaller male *Anopheles freeborni* mosquitoes at least, those males which took part in swarming had more energy reserves, than those that continued to rest at peak swarming time. This suggested that swarming in *Anopheles freeborni* would not happen if energy reserves were below a certain level. The very smallest of these mosquitoes never swarmed at all (Yuval *et al.* 1993), although this size would be related to juvenile nutrition rather than the ability to locate sugar as an imago. As swarming is critical to mating, the ability to gather sufficient sugar to increase mating opportunity directly influences fitness. Sugar feeding behaviour may be similar to that displayed by female mosquitoes in search of a blood meal host but responding to different cues.

If so, then it would share the same three components (modified from Sutcliffe 1987; Montell and Zwiebel 2016):

1. Appetitive search – hunger driven random/meandering flight to locate sugar associated stimuli, such as visual cues or odorant plume.
2. Activation – the switch over from random searching to directed stimuli source location
3. Attraction – the act of locating the source of the stimuli, at long range most likely an olfactory mediated response to a volatile organic chemical (VOC) plume, moving from lower to higher concentrations of the stimuli. At closer range attraction may shift towards visually mediated location, or short range olfactory stimuli to locate the specific feeding location; the actual nectary of a flower for example

5.1.1.2 Volatile organic chemicals from plants as mosquito kairomones

The volatile organic chemicals generated by plants act as long-range stimuli that aid the location of a dominant source of sugar for many mosquitoes, floral nectar (Foster 1995b). Their role can be considered analogous to that taken by CO_2 in long range blood meal host location in female mosquitoes. Due to the interspecific interaction between VOCs, and CO_2, and mosquitoes, these chemicals are termed kairomones (Logan *et al.* 2013) a type of semiochemical (a chemical which conveys a signal from one organism to another which can modify behaviour of the recipient) and have been demonstrated to be able to elicit the three behaviours above in mosquitoes and other insects in many studies (Takken and Knols 1999; Larsson *et al.* 2003; Otienoburu *et al.* 2012; Takken and Verhulst 2013; Nishida 2014; Pitts *et al.* 2014; Van Loon *et al.* 2015).

There is a very high number of VOCs associated with floral scent, with over 1700 identified from the analyses of 991 species (Knudsen *et al.* 1993, 2006). These compounds were classified by Knudsen *et al.* (2006) within seven classes of compounds: aliphatics, benzenoids and phenylpropanoids, C5-branched compounds, terpenoids which included mono-, sesqui-, di- and irregular – terpenes, nitrogen containing compounds, sulphur containing compounds and miscellaneous cyclic compounds. The most common compounds in floral scents were the monoterpenes limonene, (E)-β-ocimene, myrcene, linalool, α- and β-pinene, and the

benzenoids benzaldehyde, methyl salicylate, benzyl alcohol and 2-phenyl ethanol which occurred in 54 to 71 % of the families investigated by (Knudsen *et al.* 2006).

Floral VOCs can be sampled using a number of methods including steam distillation, solvent extraction and static or dynamic headspace sampling (Raguso and Pellmyr 1998; Tholl *et al.* 2006; Demeestere *et al.* 2007). However, they are generally sampled using headspace analysis where the chemical composition of the gaseous atmosphere around a flower, plant section or whole plant, is sampled using an adsorbent material before the desorption of the sample into an analytical system. Headspace analysis has the advantage of sampling those VOCs which are emitted 'naturally' by the plant rather than those which are liberated by the application of active extraction methods such as steam distillation, and should, therefore, be seen as being more representative (Tholl *et al.* 2006). Headspace is described as static when the sample is placed into an airtight fixed volume container along with the adsorbing matrix, without the addition of any form of continuous air stream (Tholl *et al.* 2006). The use of solid phase micro extraction (SPME) is very well suited to the entrainment of VOCs from static headspace (Herrmann 2010; Torto *et al.* 2013) and can collect samples at detection limits in the parts per billion by volume range (ppbv) (Tholl *et al.* 2006).

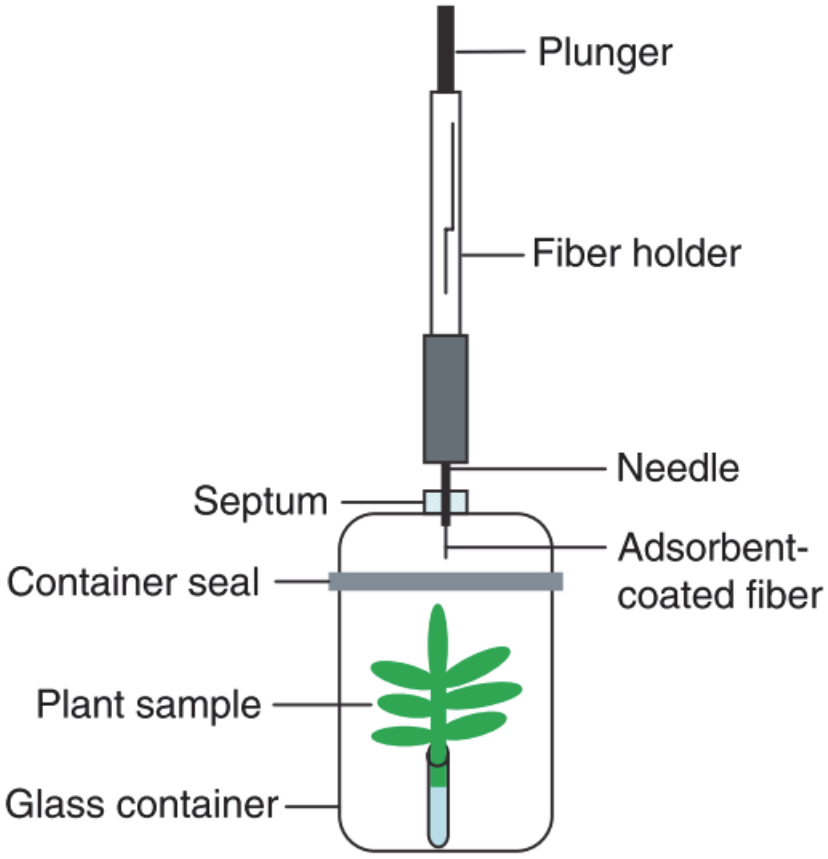


Figure 79 Solid phase micro extraction (SPME) being used to sample the static headspace around a plant sample. The ability to securely seal the vessel to allow headspace equilibration prior to introducing the SPME fibre by piercing the septum makes collection very straightforward. After Tholl *et al*. (2006)

Dynamic headspace sampling represents the most frequently used approach to sampling VOCs from plants (Tholl *et al.* 2006), and have continuous airflow through the sampling chamber, and through the adsorbent matrix. Configurations are tailored to specific applications, but may be open systems where the incoming air is drawn, unfiltered, to the sample or pre-filtered ones which have an inline filter conditioning the air before it reaches the headspace to prevent interference from VOCs from external sources (Figure 80) (Tholl *et al.* 2006).

For dynamic headspace sampling, a number of different compounds have been used as the adsorbent matrix, including charcoal, Tenax and Porapak Q (Raguso and Pellmyr 1998; Tholl *et al.* 2006). Each of these has different affinities for VOCs and therefore can generate slightly

different chemical profiles of the same headspace and so care would be required when comparing results between methods (Tholl *et al.* 2006). Having entrained a sample of the VOCs from within the headspace onto the adsorbent matrix, the sample is then eluted using a solvent to provide a liquid sample which can be employed for analysis of the sample. As with the choice of adsorbent material, there are multiple solvents which can and have been successfully utilised for this method (Raguso and Pellmyr 1998; Tholl *et al.* 2006). In their review of the methods applied to the analysis of VOCs emitted by plants Raguso and Pellmyr (1998) found that the combination of Porapak Q as the adsorption matrix and hexane as the eluting solvent provided the best results.

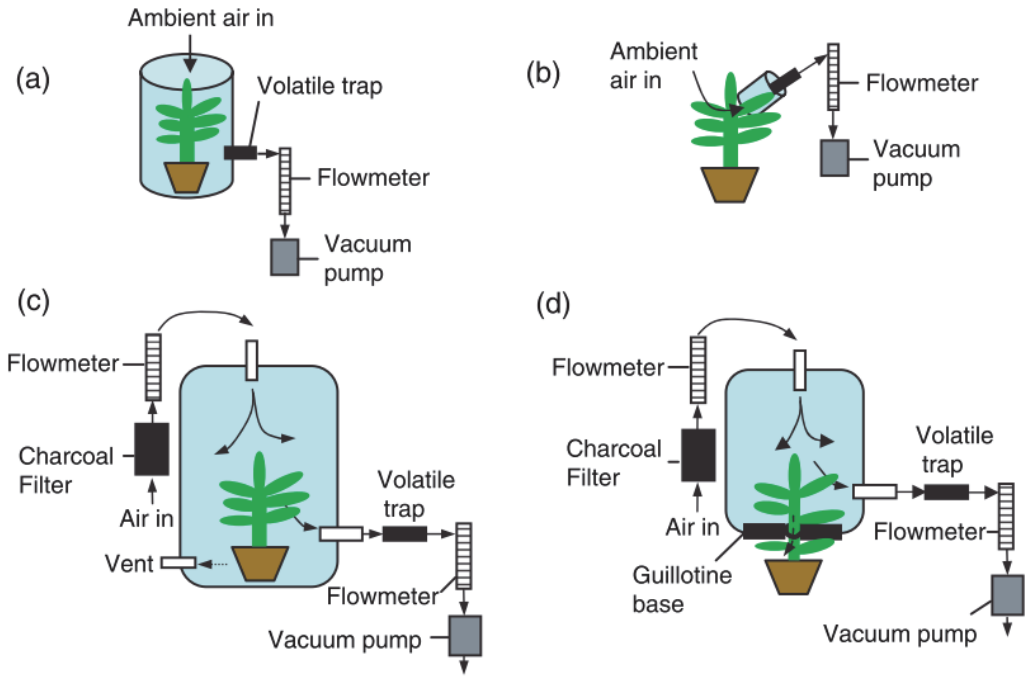


Figure 80 Four examples of dynamic headspace collection systems. a) is a simple pull system where the air is drawn over the whole plant, through an open top container, before being drawn through the VOC collecting adsorbent trap. b) is as (a) but for collection from a single leaf c) an example of a push-pull system when filtered air is pushed into the container and drawn out through the adsorbent trap at a defined rate d) is as (c) but shows a configuration suitable for collection of VOCs from only a section of the plant. After Tholl *et al.* (2006)

Typically for headspace analysis, the analytical system takes the form of gas chromatography (GC) which enables the qualitative visualisation of the sample, reporting the retention times and relative quantities of the various chemicals in the sample. These data can be made semi-

quantitative by the addition of a known amount of an internal standard (IS), such as 1-bromodecane or *n*-decane (Tholl *et al.* 2006) for comparative calculation. GC may be linked to mass spectrometry (MS) to allow the identification of the constituent chemical compounds by comparison to a chemical database.

5.1.1.3 The physiology of mosquito olfaction

Having established the importance of olfaction in mosquitoes, and the presence of kairomones associated with sources of energy, it important to consider the physiology which facilitates the acute and selective power of the olfactory sensory modality. Due to advances in the methodologies, particularly those associated with genetic modification of the expression of genes, such RNA interference (RNAi), within organisms, there have been great advances in the understanding of odorant reception in insects (Leal 2013). Within insects, olfaction is coordinated across four levels of interaction (Leal 2013):

1. The initial reception of the semiochemical at the peripheral sensory system
2. Processing of signals at the antennal lobes
3. Integration of olfactory signal with other modalities (vision or gustation for example) in the brain
4. The translation of olfactory signals into behaviour

For this study, the assays and research are related to how sugar meal associated VOCs influence male mosquito behaviour, and so for the purpose of this thesis this review focusses on point 1 in the list above, and assays related to the assays related to the resultant behaviours, point 4. External reviews of the processing of signals at the antennal lobes, point 2, are available (Christensen and Hildebrand 2002; Ghaninia *et al.* 2007; Olsen *et al.* 2007) and the integration and interpretation of sensory modalities, point 3, are reviewed in detail (Wessnitzer and Webb 2006; Martin *et al.* 2011; Leonard and Masek 2014).

When considering the role of VOCs in the scope of mosquito olfactory mediated behaviour they should correctly be seen as part of the wider nomenclature as odorants (Hudson 2000). The reception and detection of odorants is conducted primarily by odorant receptors (ORs) and ionotropic receptors (IRs) which are housed in the dendritic membrane of olfactory receptor neurons (ORNs) that are distributed across the main olfactory sensory organs of the

mosquito: the proboscis, the maxillary palps and the antennae (Montell and Zwiebel 2016). Each OR forms a binding site together with a conserved coreceptor (ORCO) (Vosshall and Hansson 2011) that does not bind odorants but is necessary for localisation of ORs to ORN membranes and are therefore essential for odorant detection (Stengl and Funk 2013). Odorants in the atmosphere which come into contact with mosquito sensilla, the cuticular surfaces of which are hydrophobic (Steinbrecht 1997), can enter the sensillar lymph via pores in the external wall (Step 1 in Figure 81). Most odorants are hydrophobic organic chemicals (Steinbrecht 1997; Montell and Zwiebel 2016) and so are not able to readily able to enter or cross the aqueous sensillar lymph fluid which surrounds the ORN dendrites (Leal 2012). To facilitate transit to the OR/ORCO units embedded in the membrane of the ORN dendrites, the odorant ligand binds to soluble odorant binding proteins (OBPs) (Step 2 in Figure 81) for transport across the lymph to activate the OR as either an OBP+odorant complex as seen in the case of the LUSH(DmelOBP76a) OBP in *Drosophila melanogaster* (Laughlin *et al.* 2008) and its orthologs identified in *Culex pipiens quinquefasciatus* (Pelletier and Leal 2009) (Step 3 in Figure 81), or by releasing the odorant to directly activate the OR (Step 4 in Figure 81). The release of the odorant is controlled by conformational changes to the OBPs at lower pH. The sensillar lymph close to the surface of ORN dendrites has been demonstrated to be lower pH than the remainder of the lymph, resulting in reduced affinity for the odorant ligand and subsequent release near to the ORs. Following the reception of the odorant at the OR, 'spent' odorants (Step 5 in Figure 81) are inactivated very rapidly. Here there are competing hypotheses. The first being that odorant degrading enzymes (ODEs) bind and remove the odorant (Ishida and Leal 2005, 2008) (Step 6 in Figure 81) and the second hypothesis is that signals are terminated by an unknown 'molecular trap' (Ishida *et al.* 2004) (Step 7 in Figure 81).

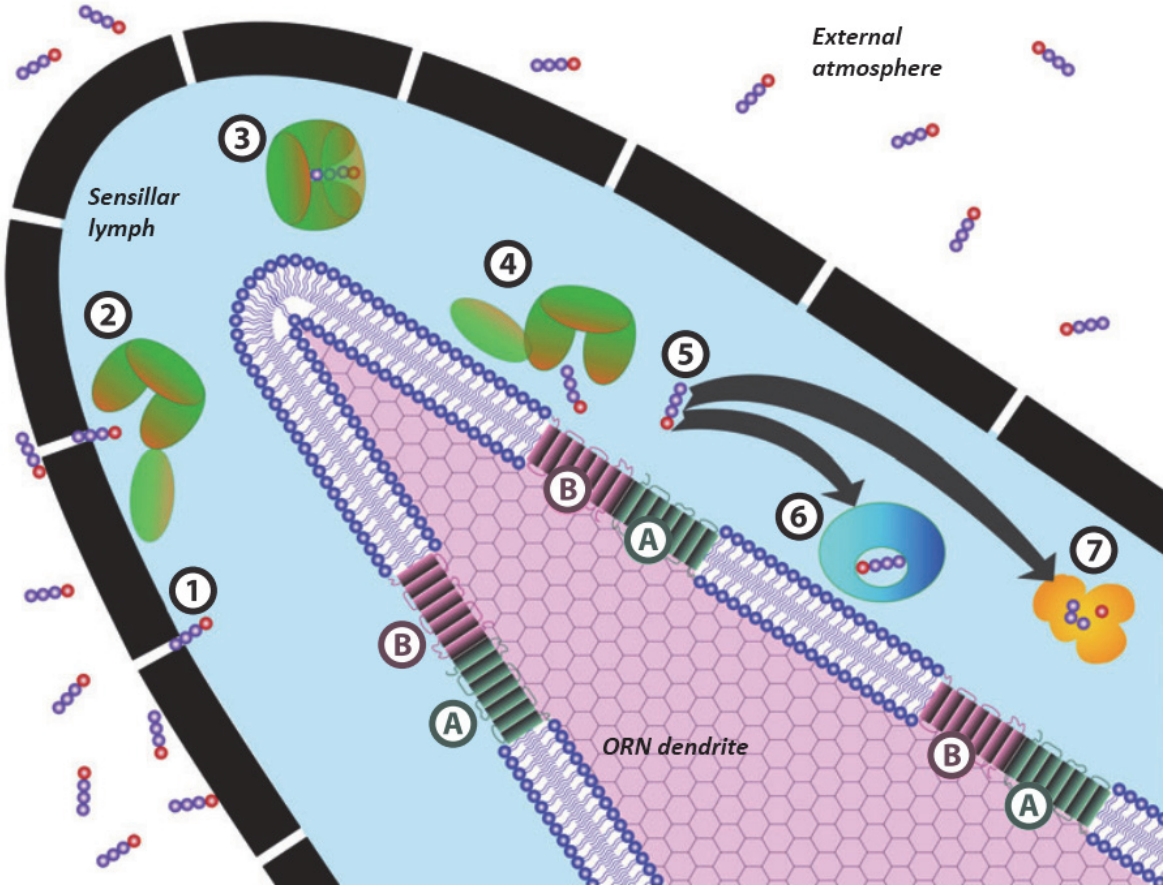


Figure 81 Systematic view of odorant binding, release and inactivation within mosquito sensilla. 1) Odorants reach the sensilla lymph from the atmosphere via pore tubules, 2) they bind to odorant binding proteins (OBPs), 3) the OBP+odorant complex is transported through the sensillar lymph and either activates receptors while bound (LUSH model) of it releases the odorant ligand at the receptor for direct activation (moth and mosquito OBPs). 4) Odorant receptors (ORs) work as heteromers (A and B) with a binding unit (B) and a well-conserved coreceptor (ORCO) (A). 5) Stray odorants are thought to be inactivated by either 6) a currently unknown molecular trap and/or by 7) the rapid action of odorant-degrading enzymes. Modified after Leal 2013.

Selectivity for specific odorants has been demonstrated to occur by the affinity of OBPs and ORs. Experiments to determine specific role of ORs have resulted in the deorphanisation (identification of odorants that activate a test receptor) of combinations of host specific odorants and ORs in *Aedes aegypti* (Bohbot *et al.* 2011, 2013; McBride *et al.* 2014), *Anopheles gambiae* (Bohbot *et al.* 2011; Hughes *et al.* 2014) and *Culex quinquefasciatus* (Hughes *et al.* 2010; Pelletier, Hughes, *et al.* 2010). Similar studies have indicated specificity within the OBPs (Biessmann *et al.* 2010; Pelletier, Guidolin, *et al.* 2010). By comparison to ORs, IRs have been understudied thus far and only one IR has been characterised leading to the deorphanisation

of AgIr76b by its response to butylamine in larval *Anopheles gambiae* (Liu *et al.* 2010). Whilst a number of OBPs and ORs have been found to be highly selective, others are more general and have an affinity for a more broad range of compounds Physiological selectivity of the peripheral sensory assemblage can be adjusted based on the needs of the individual mosquito, for example, post mating change in the expression rates of OBPs and/or ORs in *Anopheles gambiae* increases sensitivity towards host and oviposition specific odorants (Klowden 1999, 2006; Sirot *et al.* 2009). The genetic basis of odorant selectivity and the selection pressure applied by the ability to find adequate nutrition, including sugar meals also offers a clear mechanism for the development of localised behaviour of mosquito populations dependent upon the ability to best exploit the available floral assemblage.

The ability of peripheral sensory organs of mosquitoes and other insects have been assayed in many studies using electroantennography (EAG) (Bjostad and Roelofs 1980; Bjostad 1988) and single sensillum recordings (SSR) (Davis and Sokolove 1976; Bowen 1990, 1992b). These methods assess the physiological capability of the tissue to 'detect' a known odorant stimulus by making the sensory structure part of an electrical circuit. If the tissue can detect the odorant, there will be a synchronous depolarisation in the tissue at the time of stimulus release that will induce an action potential that generates an electrical signal (Zwiebel and Takken 2004). This electrical signal will be visible as a voltage change in the electrical circuit, and confirm the reception of the odorant by the tissue being test. By pairing these methods with GC/MS systems, it is possible to assay complete odorant profiles associated with whichever ecological scenario or investigation was under investigation (Baker 1995; Qiu *et al.* 2004; Gouinguené *et al.* 2005).

Over the past ~40 years of electrophysical recordings of mosquito odorant reception an extensive library of kairomones related to plant VOCs, host odours, oviposition sites (Logan and Birkett 2007; Logan *et al.* 2008; Carey *et al.* 2010; Cook *et al.* 2011) and one conspecific pheromone (mosquito oviposition pheromone (MOP)) (Laurence and Pickett 1982; Sullivan *et al.* 2014) has been generated. This library of known detectable compounds has enabled their use in further experimentation, for example, Carey *et al.* (2010) collated a panel of 110 known odorants associated with oviposition and host seeking to assay the selectivity of *Anopheles gambiae* ORs. By comparison, the library of plant related VOCs contains far fewer

confirmed odorants, and in their review of the field, Nyasembe and Torto (2014) found reports of only 29 compounds. Similarly, a comparatively small number of compounds (n = 38) demonstrated as detectable by male mosquitoes have been identified (Pitts *et al.* 2014).

5.1.1.4 The study of adult mosquito behaviour

The study of adult mosquito behaviour is an established field of research, with various approaches being employed (Takken and Knols 1999); from observations of wild swarms (Knab 1906; Lang 1920), through to carefully constructed experiments, in the field (Krebs *et al.* 2014), semi-field (Seyoum *et al.* 2002; Herrera-Varela *et al.* 2014) and laboratory settings (Posey *et al.* 1998; Verhulst *et al.* 2009). For the most part, laboratory behavioural studies have typically focussed on the host feeding choices of female mosquitoes (Takken and Knols 1999; Smallegange *et al.* 2010; Takken and Verhulst 2013) and been directly associated with the mosquito's role as the vector of diseases or as pest species causing nuisance through their biting habits. Some researchers, however, have carried out mosquito sugar feeding studies in attempts to more fully understand the olfactory mediated feeding and foraging behaviour of mosquitoes (Foster 1995a; Müller and Schlein 2006; Gouagna *et al.* 2010). Several have made connections between behaviour and plants, for example Jhumur *et al.* (2008) investigated the attractiveness of *Silene otites* L. to *Culex pipiens* var. *molestus*, and Mauer and Rowley (1999) examined the attraction of *Culex pipiens s.s.* to *Asclepias syriaca* L. (common milkweed), *Leucanthemum vulgare* Lamarck (ox-eye daisy), *Solidago canadensis* L. (Canada goldenrod) and *Achillea millefolium* L (yarrow). Others researchers have proposed several specific compounds found in floral odours as attractants which might be employed as optimisations of attractive toxic sugar bait (ATSB) methods for mosquito control measures (Schlein and Müller 2008; Gouagna *et al.* 2010; Müller, Junnila, *et al.* 2010; Otienoburu 2011; Qualls *et al.* 2014; Ding *et al.* 2016; Fikrig *et al.* 2017; Scott-Fiorenzano *et al.* 2017).

Despite the recent advances in the field of phytochemical VOCs and their interaction with mosquitoes, there are still many areas which are not entirely understood. The research effort has largely focussed on *Anopheles gambiae*, *Aedes aegypti* and *Culex quinquefasciatus* and this lack of breadth of species included reduces the likelihood of finding and testing the limits of useful generalisations related to olfactory mediated behaviours. Male mosquito behaviour

has been largely ignored (Pitts *et al.* 2014) and represents a significant gap in the knowledge of mosquito chemical ecology and behaviour.

For a number of reasons, male mosquito behaviours have been, historically, less studied relative to those of the female. As males do not blood feed, they are not directly responsible for spreading disease and have until relatively recently been overlooked as an effective avenue for mosquito control intervention (Lees *et al.* 2014; Pitts *et al.* 2014). Many control interventions based upon female mosquitoes has focused on behaviour associated with blood-meal host-seeking (Bernier *et al.* 2000; Osorio *et al.* 2012; Börstler *et al.* 2016) and oviposition site seeking (Fay and Eliason 1966; Bentley and Day 1989; Herrera-Varela *et al.* 2014) that is likely to elicit the strongest responses and therefore be the preferred behaviours for exploitation in control methods. However, when considering the use of behaviour as part of control measures, it is sensible to try and explore all avenues. Therefore, a study of olfactory-mediated activity in relation to sugar sources by males is conducted in the laboratory before field and laboratory trials of potentially attractive synthetic chemical blends. The decision was made to focus on the males of the species as males have been chronically understudied (Pitts *et al.* 2014), and it is, perhaps, these which might be more likely to be targeted using knowledge related to this aspect (Lees *et al.* 2014). For example, species-specific floral lures could be added to ATSBs to further enhance efficacy (Müller, Beier, *et al.* 2010; Beier *et al.* 2012) and could reduce non-target species casualties (Qualls *et al.* 2014). Further into the future, floral lures might be used to attract male mosquitoes to feeding stations where foods laced with orally delivered ribonucleic acid interference (RNAi) (Wuriyanghan *et al.* 2011; Coy *et al.* 2012; Baum and Roberts 2014; Ghosh *et al.* 2017) compounds are offered, facilitating the manipulation of fertility and mating competitiveness (Whyard *et al.* 2015), without requiring large scale breeding and release of males currently being used in systems such as the Release of Insects carrying a Dominant Lethal (RIDL) system (Alphey 2009, 2014; Alphey *et al.* 2013).

Recently, the viability of specifically targeting male *Anopheles gambiae* s.l. has been demonstrated (Sawadogo *et al.* 2017). They showed that it was possible to achieve up to an 80% reduction in mosquito population, a significant decrease in the number of mated females and a shift in the age structure of the male population so that many more in swarms were not

yet capable of mating. The authors described this as a 'paradigm shift' away from current control measures that focus on killing female mosquitoes (Sawadogo *et al.* 2017). That is, perhaps, a bold claim and only further research and application will support or refute their findings. There is very little research into the effects of mass killing of male mosquitoes, but Sawadogo *et al.* (2017) supports the position that a comprehensive understanding of all aspects of mosquito behaviour can help to highlight opportunities for the application of control measures that exploit specific behaviours.

Many of the behavioural studies cited above have made extensive use of choice chambers. Predominant amongst them are those that are modifications of Hancock's and Foster's (1993) "dual-port airflow olfactometer", which itself was based on previous designs (Schreck *et al.* 1967; Klowden and Lea 1978). A dual choice olfactometer has become an especially valuable and powerful tool for the assaying of behaviour in response to odour stimuli. The aim of this section of the study was to design and build an olfactometer suitable for the collection of behavioural data of adult British mosquitoes in relation to odour stimuli. The olfactometer needed to be robust enough to withstand many repetitions of assays, be easy to clean and prepare for assay iterations and allow the control of as many variables as possible thereby ensuring the behaviour recorded relates to the mosquito's response to the olfactory stimuli presented to them. Assay outcomes must be easy to record, generating discrete data for analysis accurately. The design of the olfactometer was strongly influenced by those used by Verhulst *et al.* (2009) and Pates *et al.* (2001) which were successfully used for the analysis of preference of mosquitoes to the odours of cultured human skin microbiota and those from humans and cows respectively. Additionally, the design and build of the dual choice olfactometer, and purchase of all air supply hardware, filtration and fittings, needed to be low cost and hand built from readily available parts.

5.1.1.5 Wind tunnels versus flight arenas

The olfactometers that have been employed in previous studies have fallen into two broad categories: Wind tunnels (Geier and Boeckh 1999; Omrani *et al.* 2010) and flight arenas (Hancock and Foster 1993; Posey *et al.* 1998). Other approaches do exist, beyond these broad categories, perhaps most notably the design of Dogan and Rossignol (1999) developed a vertically arranged chamber consisting of two conical sections joined at their base and stimuli

offered at opposite ends of the device, rather than the more typical horizontal arrangement with stimuli offered at the same end of the apparatus, which was demonstrated to be particularly well-suited to studies of repellent stimuli.

For the purpose of this study, wind tunnel olfactometer designs are those that do not have a distinct flight arena and so maintain a similar cross-sectional diameter throughout their length. Therefore they include Y-tube olfactometers such as that used by Cook *et al.* (2011) to investigate the enantiomeric selectivity towards 1-octen-3-ol in *Aedes aegypti*, as well as more simple single tube apparatus (Geier *et al.* 1999) exploit mosquito's upwind flight response to promote activity in the specimen/s, whilst ensuring the odour variables are presented to them. Flight arena designs, whilst still ensuring that the odour variables are correctly presented, may utilise lower airflow rates and be of a larger volume.

From a design point of view, wind tunnel designs typically have smaller cross-sectional areas, and higher air flow rates, when compared to the larger cross-sectional area and lower airflow rate employed by flight arenas, although the definitions described here are flexible, and many designs which have features from both groups exist. As this olfactometer was being designed specifically for a sugar feeding study, it was deemed best to the opt for a flight arena style, utilising a lower airflow rate and larger cross-sectional area.

5.1.1.6 The variables used in the behavioural assays conducted in this thesis.

In this thesis, the olfactory-mediated behaviours of two sibling *Culex* species, found to be common in this study's fieldwork chapter (Chapter 3) were assayed against the flowers of sympatric plant species. These British mosquitoes are not currently communicating disease to and between humans in the British Isles (Medlock and Leach 2015), although evidence exists for autochthonous communication of disease within avian hosts (Buckley *et al.* 2003). The lack of ongoing autochthonous disease outbreaks provides an opportunity to develop a better understanding of their ecology and behaviour to facilitate the development of improved control measures.

These behavioural assays also presented a chance to test whether there were any differences in preference to floral volatile organic chemicals (VOCs) between *Culex pipiens s.s.* and *Culex torrentium*. Any differences in olfactory-mediated sugar feeding behaviour would increase

the niche separation between these species, effectively adding another axis to the interaction between them (Lühken *et al.* 2015). This would be useful as, currently, there does not appear to be a comprehensive explanation of the species' relationship in their strongly overlapping distributions. Some niche separation trends seem to be relatively stable, but their interactions appear complex and distributions controlled by multiple factors such as temperature and overwintering refuge choice which facilitates some geographic separation (Hesson *et al.* 2014; Townroe and Callaghan 2014). There was also an apparent temporal prevalence difference, with *Culex torrentium* pre-imaginal stages more prevalent earlier in the season compared to *Culex pipiens s.s.* which was observed during specimen collection for this research and mentioned in Lühken *et al.* (2015). Further niche speciation is offered by differences in hibernation location which is well understood in *Culex pipiens s.s.* (Service 1968; Jupp 1979; Börstler *et al.* 2014) but as yet remains undefined in *Culex torrentium* (Börstler *et al.* 2014), and no records of the two species sharing hibernation locations in the UK. However, *Culex pipiens* and *Culex torrentium* were found sharing overwintering habitats in a study in Sweden (Jaenson 1987).

5.1.1.7 Flower species involved in the assays

Species from within the family Apiaceae were chosen as the sugar associated floral stimuli for this study. Apiaceae species with white flowers were selected for inclusion in the study. This choice was made as, although in assays visual cues were obscured, pale coloured flowers have been seen to be preferred in feeding observations of night feeding mosquitoes (Grimstad and DeFoliart 1974; Magnarelli 1977; Gadawski and Smith 1992). Therefore, if these plant species were regularly utilised in the field, then the mosquitoes would be attuned to detection of their odorants through the expression of OBPs and ORs (Leal *et al.* 2013) which maximise feeding success and therefore fitness (Hassell and Southwood 1978).

The selected species list consisted of:

- *Angelica sylvestris* (L.) (wild angelica)
- *Anthriscus sylvestris* (L.) Hoffm.. (cow parsley, wild chervil)
- *Conopodium majus* (Gouan) Loret (pignut)
- *Daucus carota ssp. carota* (L.) (wild carrot, Queen Anne's lace)
- *Heracleum sphondylium* (L.) (common hogweed)

These species have a widespread distribution in the UK and a temporal floral abundance pattern (Lovett-Doust *et al.* 1982) which means that members of this family are in flower near to sites of high mosquito abundance throughout the typical active season for British *Culex* mosquitoes. Indeed, examples of these species were observed across the rural sites, and near to the suburban sites used throughout the ecological study described in the fieldwork chapter of this thesis. They share similar appearance (Figure 82), with simple white flowers arranged in umbels. All of the species utilised ranged between 1 – 2 m in height with the exception of *Conopodium majus* which reaches ~ 60 cm in height. Species identification can be made on the basis of several features, including the arrangement of the leaves, characteristics of the stems, the arrangement of flowers in the umbels and features of the fruit (Rose 2006; Poland. and Clement. 2009).

Figure 82 All of the Apiaceae species utilised in the study have similar gross structure, with umbels of simple white flowers. *Heracleum sphondylium*, (pictured) *Angelica sylvestris*, *Anthriscus sylvestris* and *Daucus carota* are tall plants typically between 1 - 2 m in height. *Conopodium majus* is shorter, reaching ~ 60 cm in height.

Other species of Apiaceae have been shown to be sugar hosts to many insects (Zych 2007; Niemirski and Zych 2011), and have varied floral odours between species (Borg-Karlson *et al.* 1993). They have simple flower structures (Tollsten *et al.* 1994), produce superficial nectar (Borg-Karlson *et al.* 1993) meaning that access to nectar is very easy for mosquitoes which, being generalist sugar feeders, do not have specialised mouthparts suited to accessing complex nectaries. These factors combine to create a potentially naturally occurring association which may have resulted in *Culex pipiens s.s.* and *Culex torrentium* mosquitoes developing an attraction to the VOCs emitted by these species (Schiestl and Dötterl 2012).

Thus far there have been no behavioural studies of male *Culex pipiens s.s.* or *Culex torrentium* mosquitoes in relation to specified plants species. *Culex torrentium* has hitherto been omitted from sugar feeding behaviour assays altogether, the only study of their sugar feeding being a

field based analysis of general nectar feeding occurrence and temporal pattern (Andersson and Jaenson 1987).

In addition to their use in behavioural assays, the VOC emissions of these species were analysed (5.1.2.4), to explore the variation of the VOC profiles compared to a panel of 50 compounds collated from Nyasembe and Torto (2014), Otienoburu *et al.* (2012), Pitts *et al.* (2014) and Verhulst *et al.* (2011) which have been demonstrated to be detected by male mosquitoes. Visual representations of the chemical 'fingerprint' of each species' are generated in relation to these know behaviour modifying compounds. The results of the chemical analysis and behavioural assays then inform the creation of two synthetic blends of chemicals which are assayed in the laboratory and field to test whether the synthetic blends have a similar response to the flowers they mimic (5.4).

5.1.2 Chemical analysis of floral odours

This sub-chapter is split into two parts, a method development section and the subsequent application of these methods to characterise and quantify specific panel of 'target' volatile organic chemicals (VOCs) emitted by the plants that have been demonstrated to cause a behavioural response in mosquitoes (Verhulst *et al.* 2011; Otienoburu *et al.* 2012; Nyasembe and Torto 2014; Pitts *et al.* 2014).

In addition to the flower species discussed in the introduction to this part of the thesis (5.1.1.7), *Hedera helix* (L.) (common ivy) was used as an aid to method development as it continues to flower late in the year, after the candidate Apiaceae have finished flowering.

5.1.2.1 Methods development

This section outlines important aspects of the method development and optimisation process required for the investigation and characterisation of volatile floral emissions from the flowers selected for this study.

- Headspace sampling of floral volatiles
- Qualitative gas chromatography and mass spectrometry
- Quantitative gas chromatography and mass spectrometry
- Data analysis approaches

5.1.2.1.1 Aims

The aim of the method development chapter was to develop a VOC entrainment approach suitable for sampling the chosen flowers using headspace analysis, and subsequent optimisation of a gas chromatography with mass spectrometry (GCMS) method suitable for processing these headspace samples. The resultant data then needed to be compared to the panel of 'target' compounds derived from the literature (Verhulst *et al.* 2011; Otienoburu *et al.* 2012; Nyasembe and Torto 2014; Pitts *et al.* 2014). This comparison process needed to be developed in such a way that the emergence of further behaviour modifying compounds in future literature, or other sources, could be included in the analysis without requiring a complete manual reworking of the data analysis. This functionality is very important as more plant derived VOCs will, in all likelihood, be found to cause behavioural changes in

mosquitoes, and so the ability to re-evaluate the data collected in this thesis in the light of future knowledge is highly desirable.

5.1.2.1.2 Headspace sampling of floral volatiles

Three methods for the entrainment of VOCs were trialled based on the literature (Raguso and Pellmyr 1998; Tholl *et al.* 2006; Demeestere *et al.* 2007), these being field-based collection using SPME, laboratory collection using SPME and laboratory based collection using sorbent tubes with a dynamic headspace. For each of the three methods described below, five replicates were carried out as per the guidelines from Tholl *et al.* (2006). The final method would need to be robust, repeatable, and sensitive as it was anticipated that due to mosquito's highly attuned olfactory sensory organs, very small quantities of chemicals might be necessary for mediating behaviour (Riffell *et al.* 2014).

Although considered for inclusion in the study, the use of thermal desorption tubes, was rejected without testing; the GC with thermal desorption capabilities was not linked to MS, and as floral VOCs were being compared to a list of target chemicals, then the MS output was particularly important.

5.1.2.1.3 Field based static headspace analysis

Preliminary attempts used a divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS) (50/30 µm) SPME fibre (Supelco, Bellefonte, PA, USA) as the sample collecting media. This type of SPME fibre samples a broad range of compounds, due to its mixed chemistry and has been used successfully in previous floral VOC analyses (Otienoburu *et al.* 2012), meaning that it is well suited to applications where broad sampling is preferred. This SPME fibre type was used in all cases where SPME is discussed in this document.

In the field, whole plant sections/flower heads were encased in a nylon bag before being tied around the plant stem with string (Stewart-Jones and Poppy 2006). Static headspace collections were made in ambient conditions for temperature and light. The bag was left to equilibrate for 30 minutes, before piercing the bag with a pin and inserting the SPME holder through the hole and exposing the fibre for 30 minutes. After exposure, the fibre was returned to the laboratory for immediate processing using GC/MS

5.1.2.1.4 Laboratory-based static headspace analysis

Sample florets were collected and placed into 6 ml crimp top GC vials in the field. Samples were returned to the laboratory and placed in an oven at 30 °C. The use of small vials at an increased, but biologically viable temperature would enable a more concentrated sample to be formed. Each vial was allowed to equilibrate for 30 minutes before the septum was pierced and sampled for 30 minutes using SPME in the oven at 30 °C. The SPME sample was then immediately processed using GC/MS.

5.1.2.1.5 DHS under constant air flow rate using sorbent tubes and solvent extraction

The dynamic headspace method developed here was based on Raguso and Pellmyr (1998). Sorbent tubes were assembled using PoraPak™ Q 80-100 mesh porous polymer adsorbent, glass gas chromatography tubes and glass wool (Figure 83). PoraPak has the advantage of producing no artefacts on GCMS that would interfere with the target VOCs being investigated herein (Sturaro *et al.* 1992). 150 mg of PoraPak™ was trapped between wads of packed glass wool in each of three tubes. Each of the identically prepared tubes had a predefined role. One to pre-filter the air entering the dynamic headspace chamber, one as the main capture tube, this being the principal target of analysis, and one as a breakthrough tube. The term 'breakthrough' refers to an analyte passing through the capture tube and escaping the sample (Woolfenden 2010a). The function of the breakthrough tube was to collect any VOCs which overfill the adsorbent space available for that compound type in the capture tube, whilst there was no reason to assume that breakthrough would occur, having this mechanism in place demonstrated that the whole sample was captured by the capture tube. Prior to first use the sorbent tubes were pre-washed with copious analytical grade hexane, >20 ml per tube, and dried by passing nitrogen through the tubes. These washed and dried tubes were individually wrapped in aluminium foil and stored in a refrigerator at 4 °C until use.

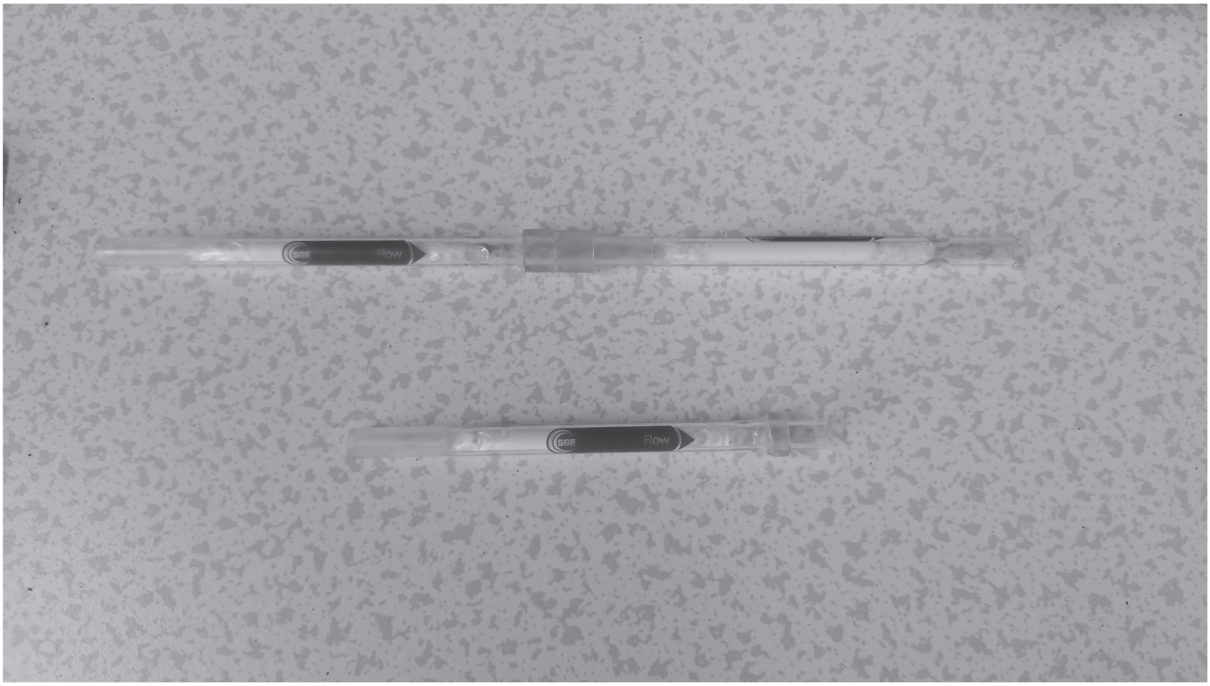


Figure 83 PoraPak Q 80-100 mesh packed tubes. The bottom tube is the pre-filter. The top tubes are the capture and breakthrough tubes which are connected in series.

The main body of the DHS chamber consisted of a glass desiccator with a single port at the top into which the Drechsel bottle head was fitted (Figure 85). The air was drawn through the system using a Vacuubrand CVC 3000 vacuum pump, with the following settings; Mode: Pump down, Speed = 95%, minimum = 450 mbar, delay = off and duration = 1200 minutes. This resulted in a mean flow rate of 355 ml/min, sd 12.3 ml/min, yielding ~ 426 L of air sampled per 20 hr sampling period. This flow rate was considered ideal for the collection method and was in line with many of the studies reviewed by Raguso and Pellmyr (1998) in their methodological comparison for this type of sampling. During sampling, the in-chamber pressure ranged between 932 and 972 mbar depending on external atmospheric conditions, whilst these showed a reduction from the atmospheric pressure, their effect is likely limited to a small increase in overall volatility of VOCs and therefore a similar increase in quantity of floral emissions.

Floral samples were collected from the field and placed in plastic sealable containers for transit to the laboratory. In the laboratory, the flower heads were cut, excess foliage removed and placed in a foil topped 20 ml glass vessel filled with distilled water (Figure 84). Sorbent tubes were attached to the DHS chamber using short sections of rubber. The pre-filter tube

was mounted in the mouth of the glass inlet tube, and the capture and breakthrough tubes mounted in series in the outlet tube (Figure 85).

Figure 84 A prepared floral sample before being placed into the dynamic headspace chamber. Excised plant sections were placed in a glass vessel containing distilled water and capped with foil for the duration of the sampling.

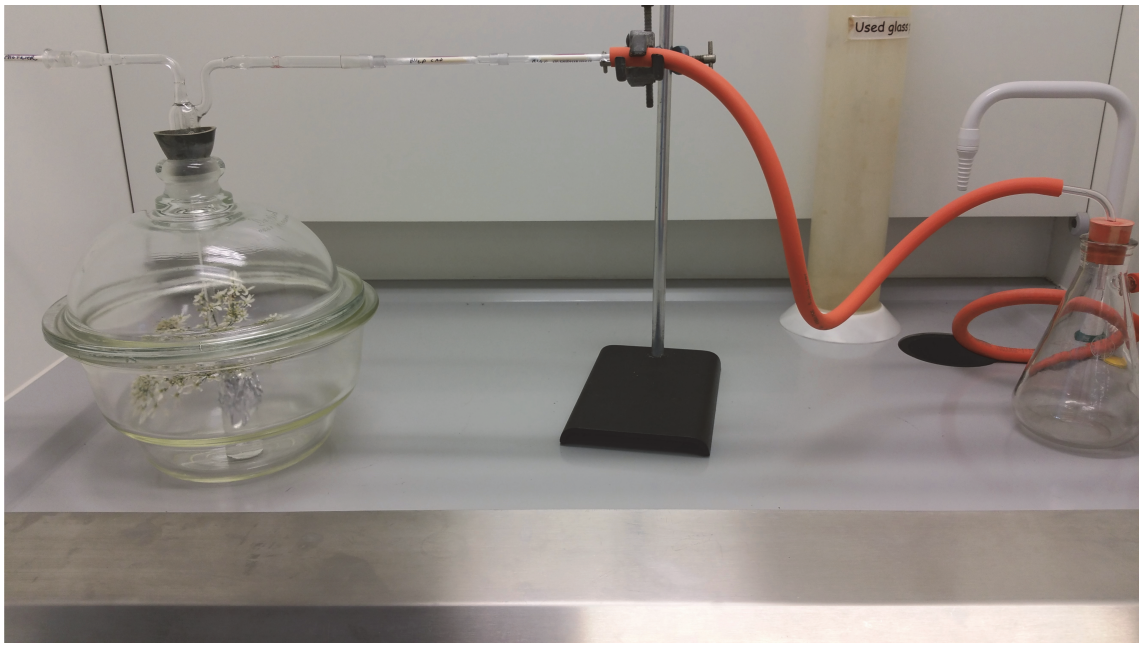


Figure 85 the assembled DHS arena and vacuum trap system. Air flow is from left to right and is drawn through the pre-filter before entering the DHS chamber. Air is then extracted through the capture tube, this is primary sampling tube, before finally passing through the breakthrough tube.

Each assay ran overnight for 20 hours, to facilitate collection of VOCs through all photophases, prior to the collection of the sorbent tubes for solvent desorption. Following sampling, the three sorbent tubes were removed from the DHS chamber and vacuum assembly and wrapped in aluminium foil before immediate processing. Upon completion of each sampling event the desiccator was opened and the flowers were inspected and observed to be in good condition. Upon opening, there was a detectable faint floral smell present. During methods development five repeats for *Heracleum sphondylium*, *Angelica sylvestris* and *Hedera helix* were conducted, in addition to empty chamber negative control.

For the solvent desorption process, separation funnels were used to deliver 3 ml of HPLC grade hexane (Sigma-Aldrich) to each tube. The product was collected in 4 ml glass gas chromatography vials (Chromacol Ltd) before being immediately concentrated to ~ 200 µl by evaporation under nitrogen, supplied via a 6-port micro manifold. 175 µl of this concentrated sample was pipetted into appropriately labelled GC autosampler vials (12mm x 32 mm x 6 mm, brown glass, Chromacol Ltd) with 0.3 ml inserts and PTFE/silicone septa.

5.1.2.1.6 GCMS method validation

All collected samples were processed on a PerkinElmer Clarus 500 Gas chromatograph with paired PerkinElmer Clarus 500 mass spectrometer. The GCMS method was initially based on the program utilised in Otienoburu *et al.* (2012), with several optimisations to ensure peak fidelity and suitability to the specific column and equipment being used. The optimisations were based on preliminary testing in the laboratory. The changes applied were: the increase the oven starting temperature from 25 °C to 40 °C to ensure the completion of the elution of solvent prior to initialisation of MS recording, reduction of the ramp rate from 15 °C per minute to 10 °C per minute to increase peak separation, and the change of the MS detector's range from 19-350 m/z to 40-350 m/z to improve the signal to noise ratio.

Ultimately, the GCMS method used for solvent (hexane) desorbed samples used 1 ul of the sample was delivered into an SLB™ - 5ms (DB-5) fused silica capillary column, 30 m x 0.32 mm x 0.25 µm film (Sigma-Aldrich- 42131-04B) in splitless mode, using the GC's attached autosampler. The oven was held at 40 °C for 3 minutes and then ramped at 10 °C per minute to 250 °C where it was held for 5 minutes. For solvent desorbed samples, a 3-minute solvent delay was added to the MS method, to allow the Hexane solvent to elute. The carrier gas was

helium at 1.5 ml min^{-1}. Mass spectra were recorded in the electron impact mode, and the mass selective detector's scanning range was set to 40-350 m/z.

For SPME samples the GC method was the same but required manual injection of the sample. The MS method was also similar, but lacked the 3-minute solvent delay, starting immediately instead.

5.1.2.1.7 Method discussion

5.1.2.1.8 Comparing GCMS results to the panel of target compounds

It was important to be able to calculate the expected retention time (RT) for each target compound, to facilitate the optimisation of the GCMS method, and check for co-eluting compounds. RT is system-specific, and is related to variables related to the GC method, column type and detector settings; it is therefore non-portable. However, the linear retention index (RI) of a compound is a system-independent constant normalised to the retention times of the nearest eluting n-alkanes (Van Den Dool and Kratz 1963). To transform RI to RT for a different GC system, alkane standards need to be analysed on that method and their RT substituted into the calculation. Whilst the calculation is simple, processing it for many compounds for every method adjustment is time consuming and a possible source of error.

Knowing the approximate range of expected RTs, and their proximity to one another would enable the development of a GC method which offered sufficient peak separation, and add additional validation to MS compound identifications. Therefore, a retention index calculator was written in R (R Core Team 2015) to calculate the RI derived from Kovat's retention index (KRI), by Van Den Dool and Kratz (1963) for each of the target compounds :

$$RI = 100n + 100\frac{RT(peak\ of\ interest) - RT(nearest\ preceding\ alkane)}{RT(following\ alkane) - RT(nearest\ preceeding\ alkane)}$$

This calculator is broadly similar in function to that developed by Lucero *et al.* (2009), but was better suited to this project as it integrated directly with the Microsoft Access database that houses all of the GCMS data gathered in this research and the panel of target compounds

data table; it also outputs its results directly to an MS Word document, portable data format file (.pdf) or HTML page as required. Thereby, eliminating multiple opportunities for human error in data entry, calculation and reporting. Reference Retention Index (RI) data for the target compounds was collected and collated from Babushok *et al.* (2011) and *The Pherobase* (El-Sayed 2014), which is the world's largest database of semiochemical and pheromones, and was based on data from GC machines using a DB-5 column similar to that employed in this research.

5.1.2.1.9 Calculating the expected Retention times for target compounds

The correlation of RI (RI = Cn * 100) (Van Den Dool and Kratz 1963) and measured RT for the alkane standards C8 to C20 was calculated using Pearson's product-moment correlation, and there was a positive correlation between the two variables, $r = 0.993701$, $df = 11$, $p < 0.01$. These data were plotted to create a calibration curve (Figure 86) for the calculation of the expected RT of the target compounds. The RT calculator then computed the expected RT for all the target compounds, based on their reported RI by NIST and Pherobase (Babushok *et al.* 2011; El-Sayed 2014) and Figure 86 relate to the expected RTs in the final GC method.

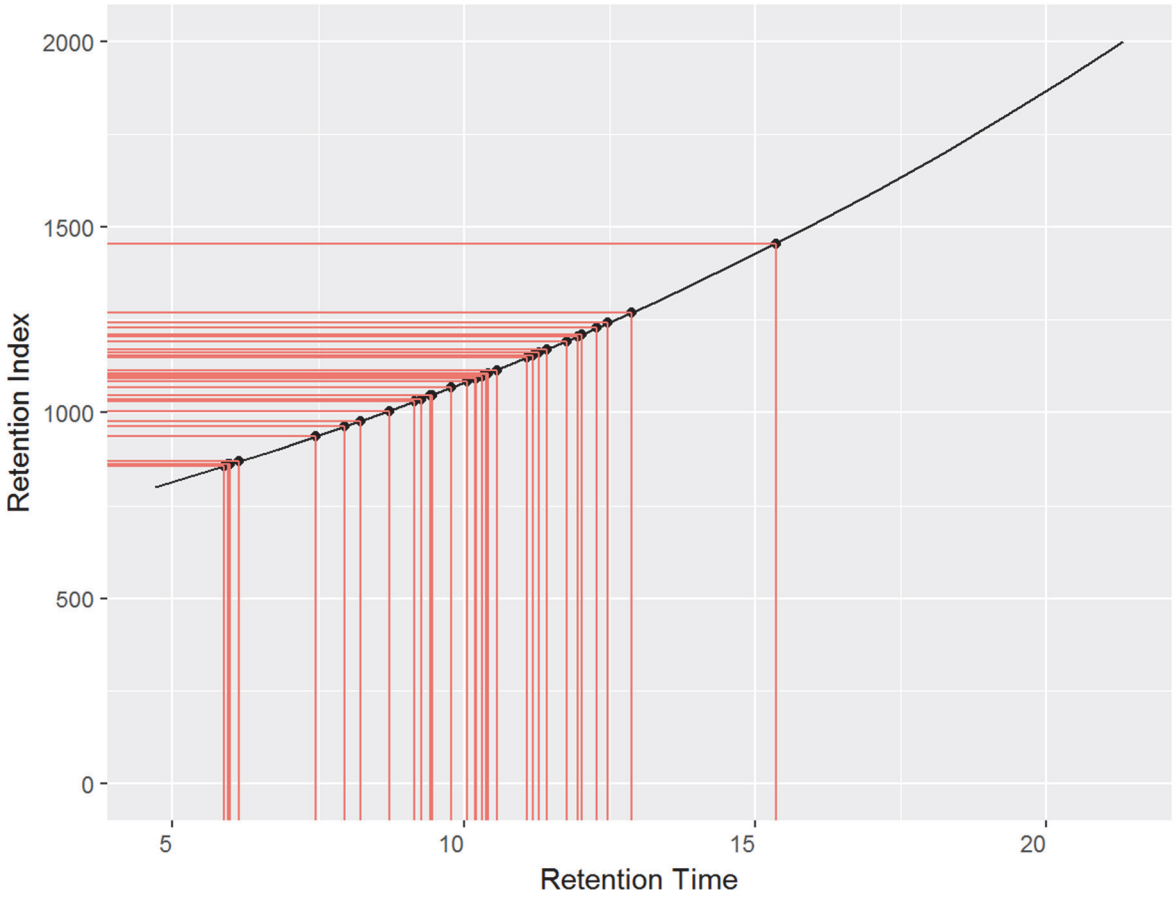


Figure 86 Calibration curve for the calculated linear retention index for the GC program being used for these analyses. The black line is the curve described by the retention times of alkane standards C8 to C20. Red lines indicate the intercept of the curves and x-axis for the target compounds with published RI > 800. This graph shows that the region of interest on chromatograms for the selected GC method is wholly contained between 5 and 16 minutes.

Table 20 The panel of target compounds identified as being plant emitted VOCs that are detectable by mosquitoes (Verhulst *et al.* 2011; Otienoburu *et al.* 2012; Nyasembe and Torto 2014), or compounds that are specifically detectable by male mosquitoes (Pitts *et al.* 2014). The average RI as reported by NIST (Babushok *et al.* 2011) and Pherobase (El-Sayed 2014), and the calculated expected RT if found using the final GCMS method developed for use in this study.

Compound name	Molecular mass	Average RI	RT Calculated from published RI
DEET	191.2695	0.0	NA*
DDT	354.4860	0.0	NA*
Acetaldehyde	44.0526	500.0	NA**
Hexanal	100.1589	799.9	NA**
(Z)-3-Hexen-1-yl alcohol	100.1589	856.6	5.8777
Hexan-1-ol	102.1748	869.7	6.1479
α-Pinene	136.2340	936.1	7.4517
Benzaldehyde	106.1219	962.7	7.9518
β-Pinene	136.2340	977.7	8.2338
(Z)-3-Hexen-1-yl acetate	142.1956	1004.0	8.7204
Limonene	136.2340	1029.5	9.1498
Benzyl alcohol	108.1378	1036.9	9.2744
Phenylacetaldehyde	120.1485	1045.9	9.4260
β-Ocimene	136.2340	1047.7	9.4563
Acetophenone	120.1485	1067.4	9.7880
(E)-Linalool oxide (furanoid)	170.2487	1083.3	10.0558
Dehydrolinalool	152.2334	1091.0	10.1854
o-Guiaiacol	124.1372	1092.1	10.2040
Linalool	154.2493	1099.0	10.3202
Nonanal	142.2400	1103.3	10.3877
Thujone	152.2334	1105.1	10.4153
Phenylethyl acohol	122.1644	1114.9	10.5659
Veratrole	138.1638	1148.1	11.0758
Lilac aldehyde A	168.2328	1155.0	11.1818
Lilac aldehyde B	168.2328	1155.0	11.1818
Lilac aldehyde C	168.2328	1155.0	11.1818
Lilac aldehyde D	168.2328	1155.0	11.1818
(E)-2-Nonenal	140.2227	1162.2	11.2924
Linalool oxide (pyranoid)	170.2487	1171.0	11.4276
Methyl Salicylate	152.1473	1192.9	11.7639
Verbenone	150.2176	1206.2	11.9615
Lilac alcohol A	170.2487	1211.0	12.0301
Lilac alcohol B	170.2487	1211.0	12.0301
Lilac alcohol C	170.2487	1211.0	12.0301
Lilac alcohol D	170.2487	1211.0	12.0301
Nerol	152.2400	1228.9	12.2857
Neral	152.2400	1242.1	12.4742
Geranial	152.2400	1270.3	12.8769
(E)- β -Farnesene	204.3511	1455.9	15.3499

*At the time of writing reference Kovat's index values were not available for DEET or DDT in NIST (Babushok *et al.* 2011) or Pherobase (El-Sayed 2014) on DB-5 column hence the NA in the table.**Hexanal and Acetaldehyde have retention indices lower than the minimum alkane in the standards ladder, and therefore the complete calculation cannot be made, hence their NAs.

5.1.2.1.10 Discussion

Both of the static headspace approaches using SPME, field collections in bags and laboratory collections from vials, proved to be relatively insensitive, generating fewer peaks than dynamic headspace sampling using sorbent tubes. It was imperative that the VOC sampling method utilised could capture a high-fidelity sample of the floral headspace for analysis. When considered in the context of insect olfactory behaviour, minor constituent odorants may contribute to attraction alone or synergistically with major constituents (Birkett *et al.* 2004; Piñero and Dorn 2007; D'Alessandro *et al.* 2009) and that the ratio of odorants has been shown to be critical between mosquitoes and blood meal hosts (Takken *et al.* 1997; Qiu and van Loon 2010) and potentially critical for insect-plant interactions (Bruce *et al.* 2005). It has been demonstrated in Lepidoptera that significant deviation in the ratio of VOCs in synthetic chemical blends mimicking naturally attractive plants resulted in reduced attractiveness (Tasin *et al.* 2006). The amount of plasticity of these ratios is unclear, but it is thought that some must exist to allow for phenotypic variation in plant odours across distribution ranges (Najar-Rodriguez *et al.* 2010). It should be borne in mind that overall ppm of a VOC will be relative to the distance from the source, but ratios between VOCs can be considered as a plant signature or fingerprint.

Following the method testing, headspace sampling using SPME in the field was ruled out due to the large daily and seasonal variance in environmental variables such as sunlight and temperature, both variables critical to the emission and retention of VOCs (Tholl *et al.* 2006), which were introduced which could be easily controlled across the whole sampling season were collections to be made in the laboratory. However, field sampling of headspace was found to be adequate for a simple qualitative snapshot of VOCs emitted under the environmental conditions recorded at the time.

The SPME static headspace sampling method increased the control of the variables light and heat by allowing the sampling to occur under constant light and temperature. This enhanced reproducibility and simplified comparison between samples compared to the field based method. The headspace vessel was significantly reduced in size to try and concentrate the VOC entrainment, by reducing the time taken to equilibrate the headspace, and fitted with a crimp top lid and septum. It showed improved detection of peaks compared to the field

collections, but less good detection than from dynamic headspace. Although it improved the quality of the VOC sample, the small vessel size introduced two issues: it effectively limited the size of flowers or florets that could be analysed, and meant that many cut stems were included in the headspace effectively simulating herbivory and the potential additional VOCs so liberated (Paré and Tumlinson 1999).

The dynamic headspace sampling approach, using sorbent tubes, has fewer drawbacks and many more advantages over the other methods trialled here. Twenty-hour dynamic headspace sampling included all light phases and took place under consistent, reproducible, temperature and humidity conditions; meaning that conditions were reproducible and sample VOCs entrained would include all those available during the normal, active period of *Culex* mosquitoes being used for the behavioural assays.

The continuous airflow used in this method prevented headspace partition equilibrium and therefore facilitated emission of VOCs in a similar way to the air movement in field settings. Also, because air being drawn into the headspace chamber was filtered using an identical sampling tube as the capture tube, confidence was very high that any chemicals collected in the capture tube were emitted by the flower; this confidence was shown to be well founded in the subsequent analysis in this chapter (see results). The inclusion of a breakthrough tube enabled the complete confidence that the all the VOCs which could be sampled by PoraPak Q were indeed collected by it; if a VOC was present on the capture tube and not on the breakthrough tube, then it could be quantified by the capture tube. If a breakthrough tube was not employed, then it would be impossible to know whether VOCs had saturated the capture tube and then passed out of the experiment uncollected.

In DHS, the only cut to each plant specimen was a single stem cut excising the flower from the main plant, to reduce the potential impact of increased herbivory-related VOCs being released to the sampled headspace. Because this cut was submerged in the water beneath the foil used to cap the vial, it was judged that any impact would be minimal and acceptable when considered with respect to the practical benefits of using excised inflorescences rather than whole plants. This use of foil to cover the water, in which plant stems are placed, intended for exclusion from analysis has been previously used successfully (McCallum *et al.* 2011).

In conclusion, the pairing of 20-hr dynamic headspace analysis using PoraPak Q sorbent tubes with the GCMS method described, modified from Otienoburu *et al.* (2012) was deemed as the appropriate approach for the collection and analysis of floral VOCs, and the results discussed in the following sections all relate to the application of this approach.

5.1.2.2 Developing a computational approach for the comparison of multiple GCMS outputs to a panel of 'target' compounds

Liquid samples collected using the DHS method were processed using GCMS using the method discussed above to investigate the VOCs present and compare them to the panel of target compounds previously introduced

Preliminary inspections of GCMS outputs showed that the target VOCs peak areas were relatively small, in terms of area under the peak, and so were not suitable for analysis using typical top-ten or top-twenty MS report styles which reported other, larger, incidental peaks rather than those of interest. Also, due to the high number of target VOCs (n= 50), and their close eluting times it was not practical to confine the report to any small region of the chromatograms. Therefore, an alternative approach was needed. Preliminary testing showed that generating MS reports of the top 150 peaks offered the optimal report, in terms of providing high identification confidence and not missing small peaks above the baseline. However, a 150-peak report is 150 pages long, and the contingency table for analysis was also quite long; this meant that a programmatical, automated, approach to the collation of and, parts of, the analysis was preferable.

5.1.2.2.1 Aim

To develop a computational method for the analysis of GCMS outputs of headspace samples collected from botanical specimens, suitable for the investigation of the presence of a panel of target VOCs (Table 20) and the variation of this presence between candidate flower species.

5.1.2.2.2 Method

The qualitative report from the mass spectrometer was set to return identification for the top 150 peak matches for each sample. To handle these data a two-step process was designed and written using R statistical software (R Core Team 2015). The data and analytical flow through the tool is charted in Figure 87.

The first step was to extract relevant data from multiple MS report files and import it directly into a Microsoft Access database. The data retained from the MS report for each sample included: Sample file name, peak number, retention time, peak height, integrated peak area,

peak area percentage, normalised percentage. Appended to the peak characteristics were the library match results including: Library name, match score, compound name and CAS number for the five matches returned by MS. The tool was configured so that multiple MS reports can be imported in one go, by importing all of the reports that are placed in a folder on the user's computer. The import process can create a new data table within the database or append data to an already existing one. After the data have been imported from the reports they are moved to a second folder to prevent accidental re-importation.

The second step was to retrieve and interrogate the database of MS results to compare the different sorbent tube samples, pre-filter, capture and breakthrough, for the plant species to the 'target' VOC panel (Figure 88). The number of target compounds was compared and visualised between the tube types for each plant species, before interspecies comparison and visualisation (Figure 89). Next, for MS data with an included internal standard for semi-quantification, the data were normalised to the internal standard and the headspace concentrations, in ppm, of each compound in the samples were calculated and visualised for each species. For samples without an internal standard, the ratio of compounds was calculated relative to a compound which was found to be common in all samples within the data set under analysis (Figure 90). Finally, a visual representation of the pattern of the odorant profile of each of the flower species was generated. These odorant 'fingerprints' were designed to be easy to interpret and allow the visual comparison between species' floral odours constrained by a panel of 'target' compounds known to be of biological importance to the question being investigated.

The tool is written in such a way that should new compounds be identified as drivers of male mosquito behaviour they can be added to the panel of target compounds at any time. Following the addition of new compounds, the program can be re-run over the already collected data with minimal effort, generating all visualisations again based on the new panel.

The R code for the first task was written in Rscript (Appendix 6) and the code for the second and third steps was combined into a single document (Appendix 7), written using R Markdown language to allow the results to be output directly to a web page or Microsoft Word document.

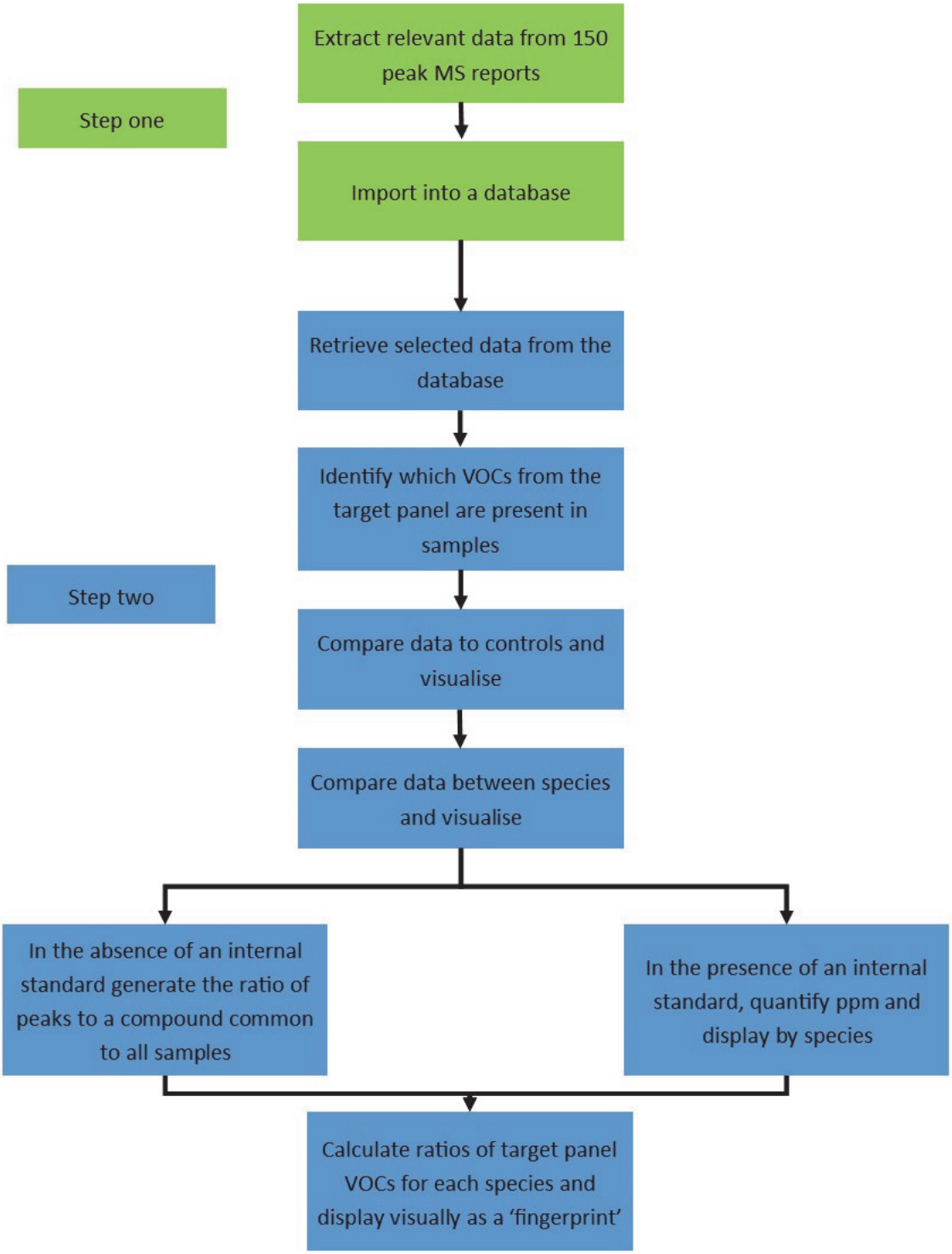


Figure 87 Data handling and analytical flow within the tool developed to analyse multiple 150 peak MS reports from multiple species, including the normalisation and semi-quantification of data to produce visual representations of floral odorants which match a panel of 'target' compounds.

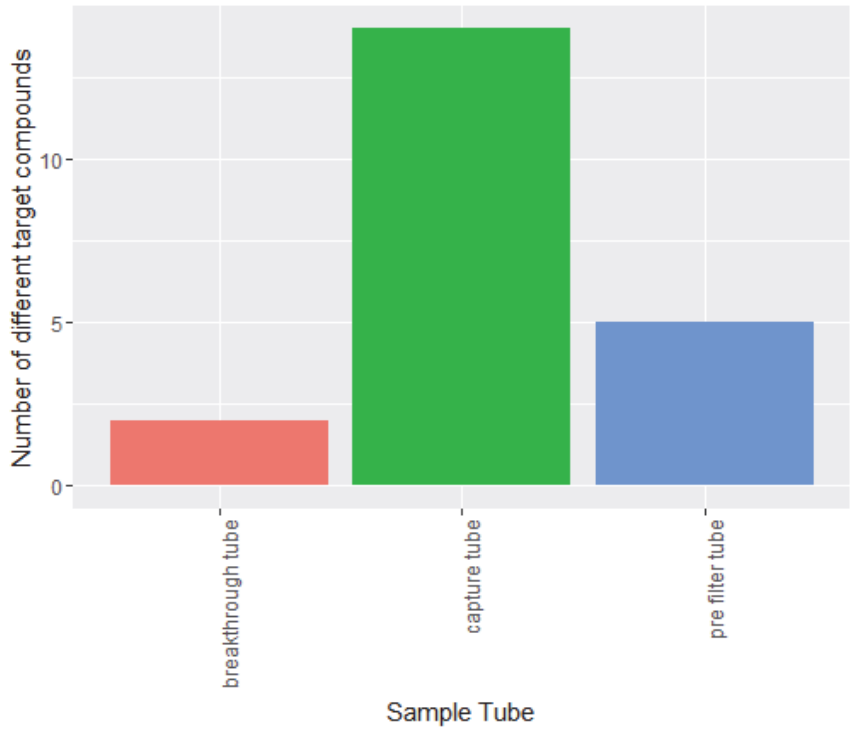


Figure 88 Example output from the analysis tool for the comparison of the number of target VOCs found by the different sorbent tube positions. This visualisation was typical of many samples and was indicative of correct function of the DHS and GCMS methods.

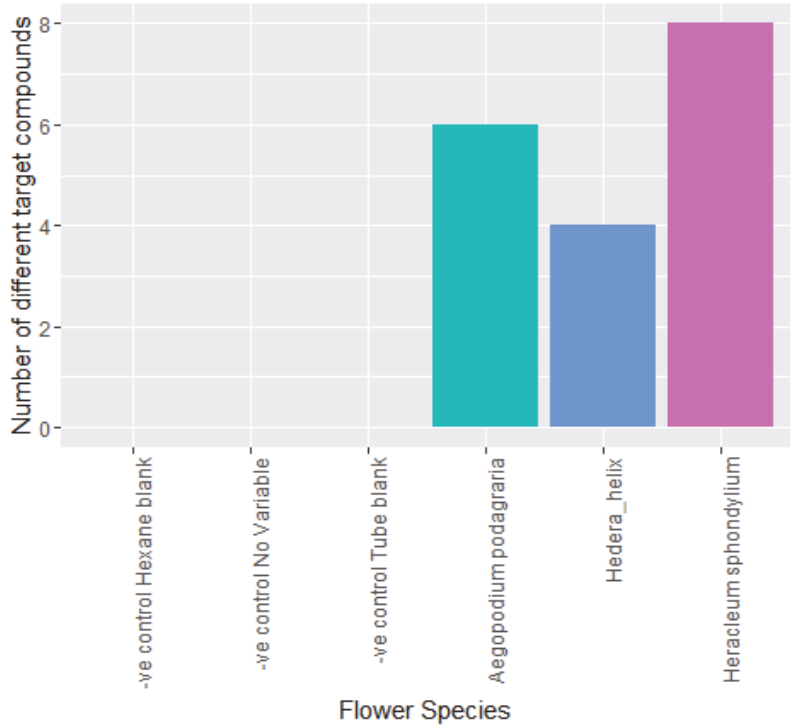


Figure 89 Example of the comparison of the number of target VOCs from the three different flower species and the negative controls.

The species shown here in the examples from the method development phase showed visually different floral 'fingerprints' (Figure 90) when presented graphically, however please note that these data should be considered qualitative as there was no internal standard used in these preliminary trials. Whilst the pattern of the ratios of target VOCs to β-pinene (a compound common to all the samples) can be seen to vary between flower species, the y-axis values cannot be compared between species as they did not share a common range. For semi-quantitative data, normalised against known quantity of an internal standard, they y-axis would be constrained for more accurate comparison.

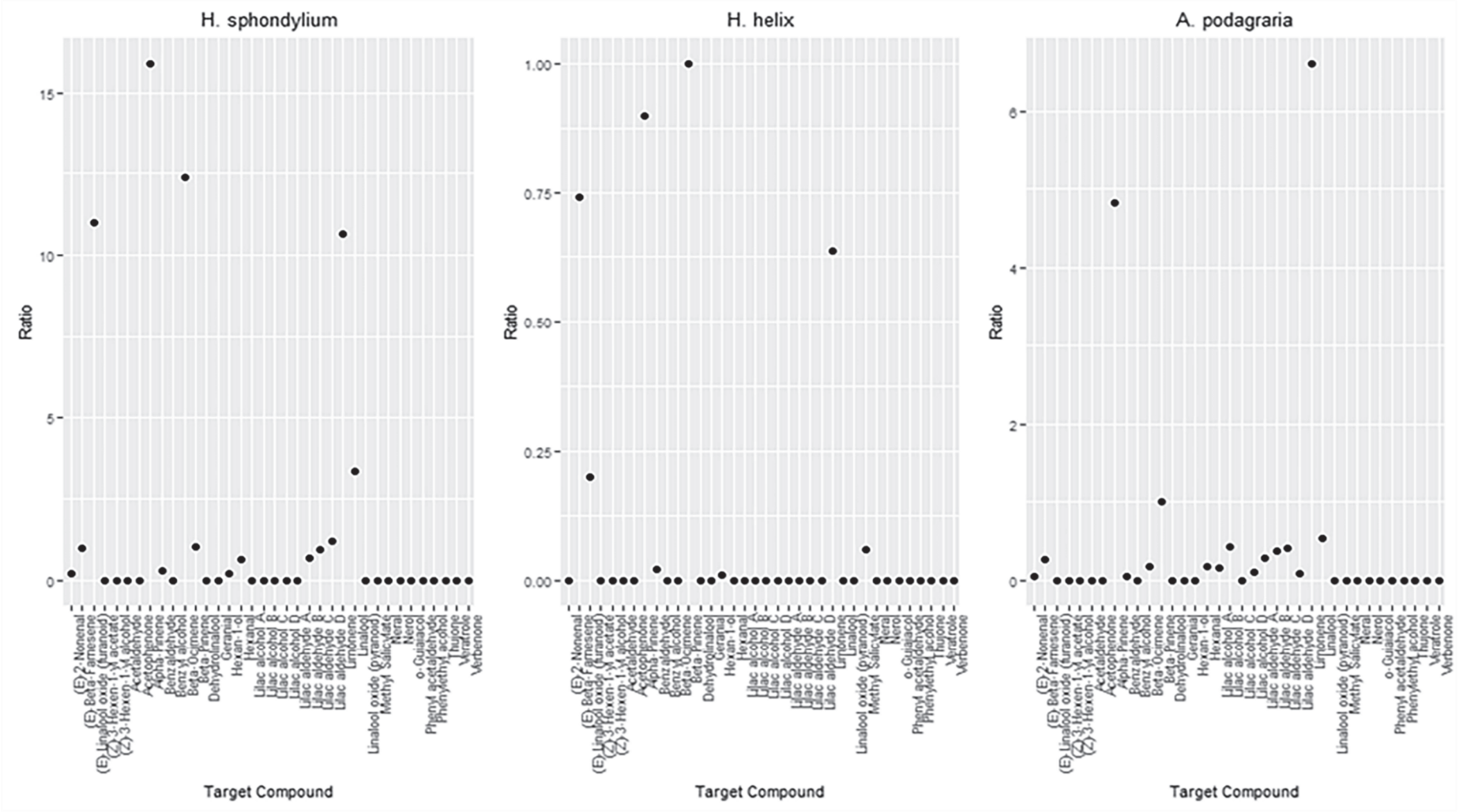


Figure 90 Example of the representation of ratios of target VOCs used to create a "fingerprint" of the emitted VOCs for each plant species. Note that y-axes are not the same scale and that all ratios are against the β-pinene median within the species samples

5.1.2.3 Quantitative Gas chromatography

Quantification of GC is generally achieved by the inclusion of an internal standard, of known concentration, and which elutes close to the region of interest without obscuring any of the target peaks on the chromatogram (Tholl *et al.* 2006). Adding a compound of known concentration to samples would allow the actual concentration of target VOCs, within a single specimen, to be calculated (Herrmann 2010). This is achieved by calculating the ratio of the area under peak value of the target compound to that of the internal standard. The inclusion of an internal standard also facilitates comparisons between samples; in this case, the internal standard allows for the normalisation of the samples. Where the same methods have been applied, then the value for the internal standard should be the same.

The aim of this section of the study was to create a method which used an internal standard, and then to develop the post analysis procedures needed to normalise the data and properly align the peaks for later statistical testing.

5.1.2.3.1 Methods

DHS sampling was carried out using the same method as that development section and the GCMS method was also as previously described.

5.1.2.3.1.1 Addition of an internal standard

Following the DHS process, the inflorescences were removed from the headspace collection vessel and dried to constant weight to control to allow any variance in specimen mass to be included in calculations. 2 µl of 0.0047 M 1-bromodecane (molarity by serial dilution), was added to each sample, post solvent desorption, but prior to the evaporation of excess hexane under N_2. The 1-Bromodecane volume used as the internal standard was arrived at by empirical testing until the area under peak on chromatograms was within a range of the majority of the target VOCs detected by the GC method. 1-bromodecane was an ideal internal standard as it eluted at ~14.14 minutes using the GC/MS method utilised, which placed it within the region of interest of the other compounds without masking any of the target VOC peaks. 1-bromodecane has been used successfully as an internal standard in several other analyses of floral odorants (Tholl *et al.* 2006; Heil and Silva Bueno 2007; Ballhorn *et al.* 2013).

5.1.2.3.1.2 Additional data handling for quantitative analysis of chromatograms – remedying time shift, peak alignment and normalisation

Due to the large number of chromatograms created during the analyses, and the need to normalise these by the included internal standard, it was desirable to used computational means of processing the chromatograms to analyse the differences between the species. This did, however, introduce several challenges; principal among these was the marked time shift between samples.

5.1.2.3.1.3 Pre-processing to remove chromatogram tail

Raw x,y coordinate chromatographic data were collected for each sample from the GC and saved as appropriately labelled .csv files. The .csv files needed to be the same length, however, the GC/MS machine being utilised delivered the data in two slightly different time frames. Critically, this varied between samples, not within a sample, and so could be fixed in post processing. The time shift was apparent because samples with the same run time would have chromatograms containing different numbers of data points - e.g. 6474 6442 6457 6456 were all dataset lengths for the same time period. At the most extreme this represented a slight discrepancy: 6474 - 6442 = 32, 32 * 0.004 (GC frequency resolution) = 0.128 minutes or 7.68 seconds over the whole 30-minute recording period; however, despite being relatively minor, it still required correction before analysis.

As has already been shown in the Retention Index calculator, none of the target compounds has expected RTs of greater than 16 minutes, at which point the oven temperature is 170 °C. Therefore, to ensure that the data had equal numbers of data points a custom R script, "Chromatogram pre trimmer.R", was created and was passed over the data files first. This was used to cut the chromatograms at 4070 data points which was equivalent to ~20 minutes. The actual recorded time at 4070 data points was 20.043 minutes in the shortest files (6442 data points in the original data) and 19.958 minutes in the longest files (6474 data points in the original data) - giving an error of 0.085 minutes or 5.1 seconds over the 20 minutes retained. This was not expected to present a problem as it only equated to 0.00125 seconds' error per data point.

5.1.2.3.1.4 Pre-processing to carry out peak alignment

To remedy the slight peak shift which remained following the previous trimming step peak alignment needed to be performed. The method devised by Vu *et al.* (2011) was used to align peaks within the data. It uses a novel peak alignment algorithm, called hierarchical Cluster-based Peak Alignment (CluPA). Although originally written for nuclear magnetic resonance (NMR) data it handled the chromatography data equally well. Vu and Laukens (2013) created the R package "speaq" to enable a simplified method of application of their process to data. Custom R code was written using the speaq package and passed over the data to fully align the peaks.

5.1.2.3.2 Results

5.1.2.3.2.1 Peak alignment

The data alignment pre-processing proved to be robust when trialled using *Hedera helix* L. samples. Figure 91 shows the level of misalignment which is present prior to pre-processing. These unaligned chromatograms cannot easily be processed using quantitative statistical methods. After processing, the peaks aligned properly along the x-axis facilitating further analysis using quantitative statistical methods (Figure 92). Although this alignment is only shown for a short window of the chromatograms this same transformation was carried out for all peaks along the chromatograms.

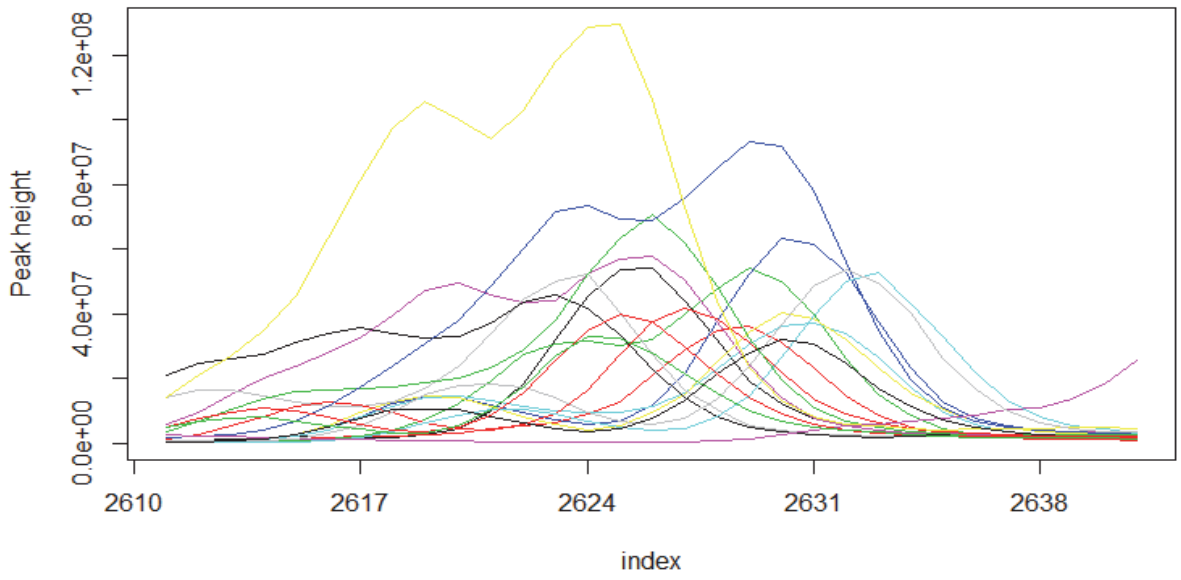


Figure 91 Non-aligned chromatograms of the 1-bromodecane region of the outputs. The shift between peaks is evident and problematic for direct quantitative analysis of the chromatogram data. The x-axis is the index of data points in the data. Coloured lines represent different replications of *Hedera helix* dynamic headspace collections.

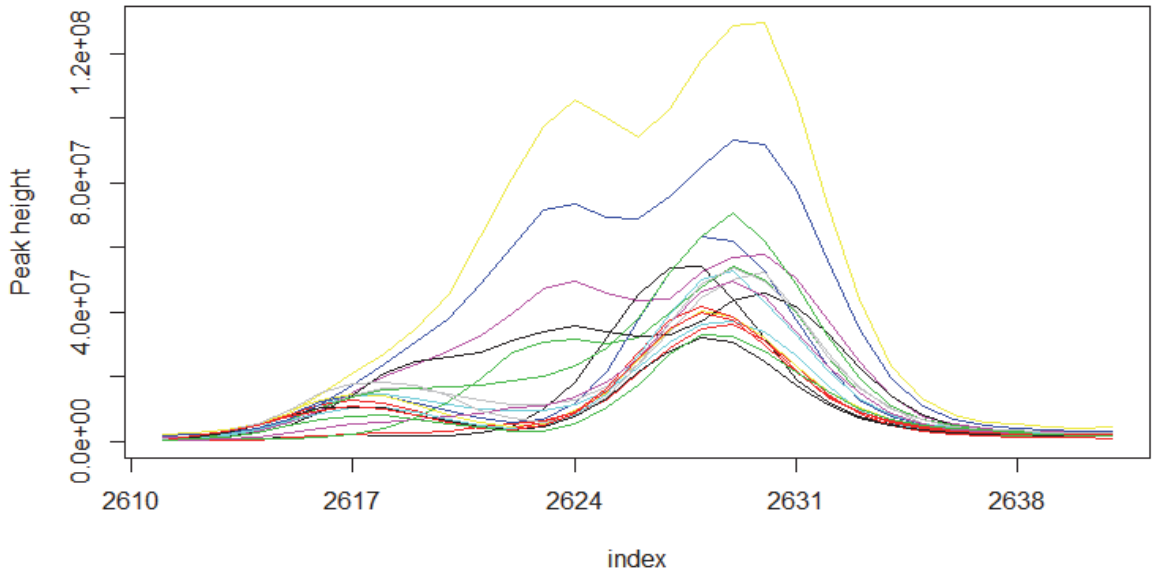


Figure 92 Detail of the 1-bromodecane region of the chromatograms post alignment pre-processing. Now that the peaks are more closely aligned with the x-axis, analysis of the chromatograms is facilitated. The x-axis is the index of data points in the data. Coloured lines represent different replications of *Hedera helix* dynamic headspace collections.

5.1.2.3.2.2 Normalisation using 1-bromodecane as an internal standard

The aligned but non-normalised 1-bromodecane region of the *Hedera helix* chromatograms is shown in Figure 93. As can be seen, the area under peak values for the 1-bromodecane internal standard varied widely, showing that there was variation present which needed to be accounted for by normalisation.

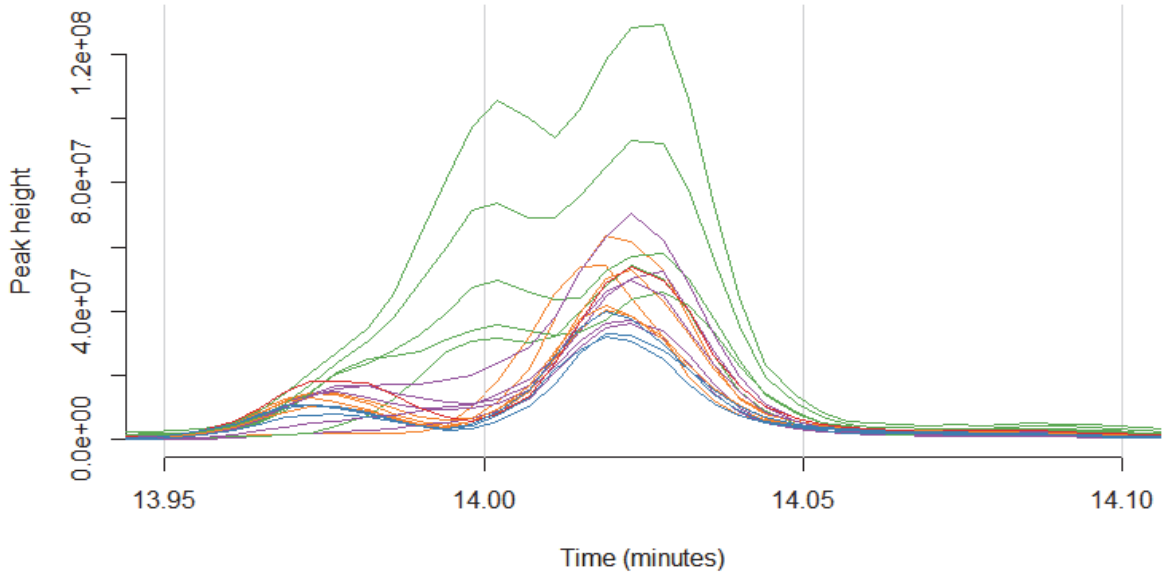


Figure 93 Aligned non-normalised chromatograms are showing the 1-bromodecane region. As can be seen, the peaks are of different areas. Coloured lines represent different replications of *Hedera helix* dynamic headspace collections.

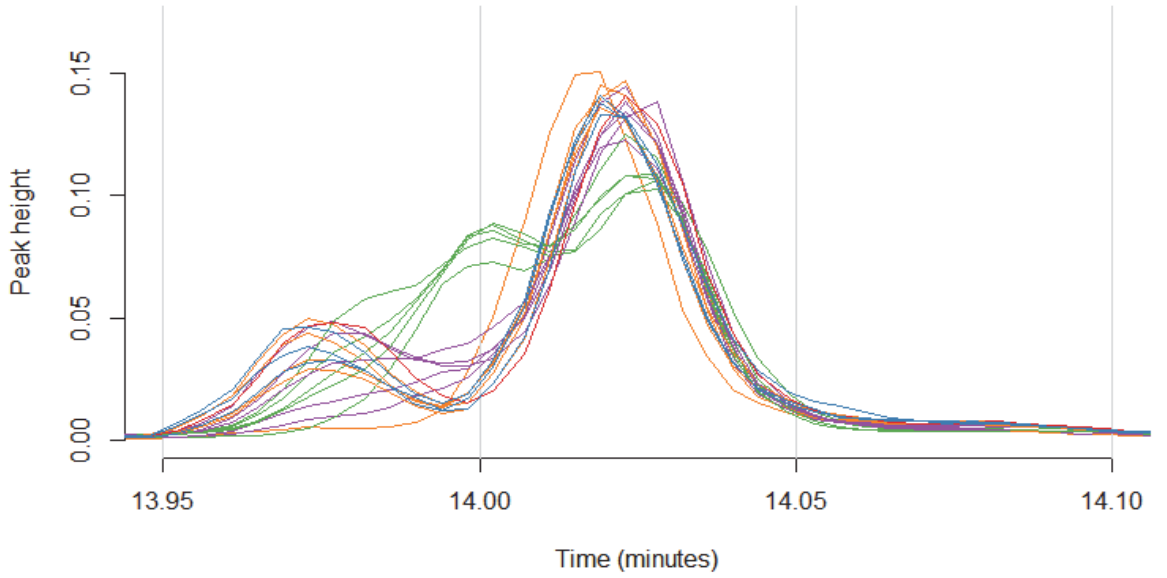


Figure 94 Aligned and normalised 1-bromodecane regions of multiple chromatograms. The peak areas now match between the chromatograms. Coloured lines represent different replications of *Hedera helix* dynamic headspace collections.

Figure 94 shows the result of the normalisation process on the 1-bromodecane region of the chromatograms. As this region contains the internal standard, the peaks are very similar as

must be the case to evidence the normalisation process working properly. The level of transformation applied to each chromatogram depended upon the ratio of the 1-bromodecane within and between samples.

When the full chromatograms were plotted on the same x-axis, following alignment and normalisation, it was evident that there were far more peaks visible in the capture tube chromatogram when compared to any of the negative controls (Figure 95), supporting the method application and suitability so far.

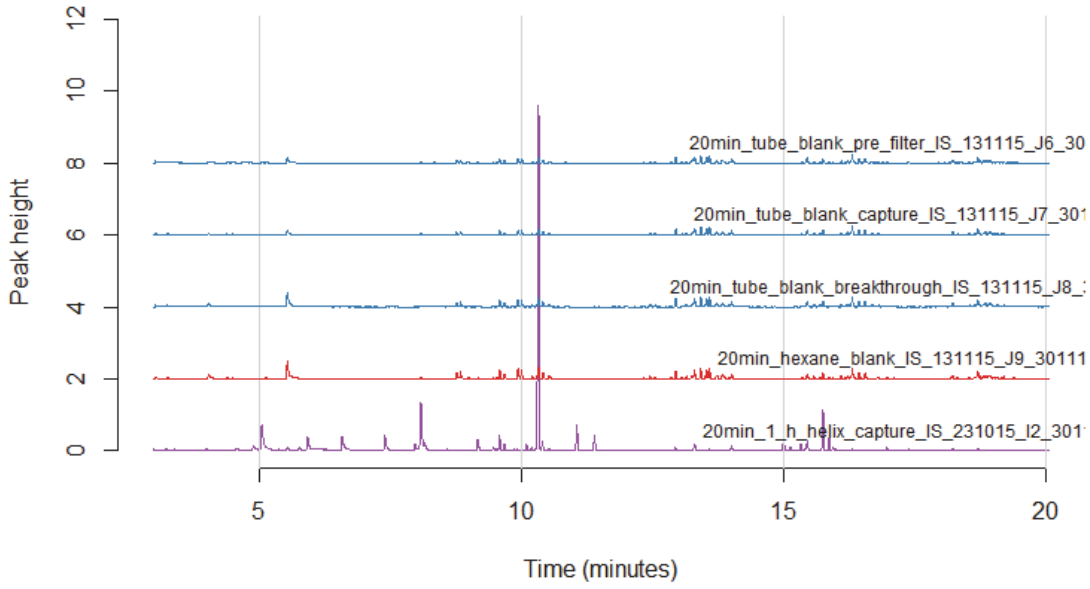


Figure 95 Aligned and normalised capture tube compared with the negative controls used in these analyses. Baselines are shifted on the y-axis for clarity. The additional peaks present in the bottom, non-control, chromatogram of *Hedera helix* DHS are obvious and must be caused by the emission of VOCs from the plant, as they are not present on negative controls.

5.1.2.3.3 Discussion

As can be seen in Figure 92 the peak alignment dos not look exactly the same; this was presumably a result of the slight differences in peak shape, and because the peak alignment pre-processing using R and the speaq package was carried out using lossless mode. This mode means that even very small peaks would not be lost during the alignment process, this effectively reduced the magnitude of any possible shift distance which can be applied. Although an absolutely aligned set of chromatograms would be desirable, the loss of small peaks in the broader context of the study is not acceptable. Looking at the final results of the normalisation and alignment in Figure 94 there does not appear to be any significant change

to the overall shape of the peaks as evidenced by the retention of the shoulder at 14 minutes prior to the main 1-bromodecane peak.

An advantage of the normalisation using 1-bromodecane was that it controlled any effect of differing concentration of the final sample following the evaporation of hexane solvent under nitrogen gas. The evaporation process endpoint was largely operator determined but because the 1-bromodecane was added prior to evaporation any difference so caused will be mitigated against and accounted for during the normalisation process.

Once the GC data had been processed using the tools laid out in this method they were ready for quantitative statistical analysis once the additional plant species were included in the sampling program. During the method development process, only *Hedera helix* specimens had been sampled using the quantitative DHS method with 1-bromodecane as an internal standard, and so statistical testing was not carried out.

5.1.2.4 Application of the devised methods to British Apiaceae

5.1.2.4.1 Introduction

Having previously, in the chapter introduction, made the ecological case for the use of several members of the Apiaceae family in the *Culex pipiens s.s.* and *Culex torrentium* olfactory-mediated behaviour assays; it was essential to characterise the VOC emissions from the flowers of the candidate floral species. Preliminary testing, during method development, confirmed that the methods employed were sufficient to discover the presence of target VOCs (Table 20) and to quantify them in terms of ppm in the collected sample.

This chapter utilises the methods developed above to analyse the VOC emissions by the flowers of selected British Apiaceae.

5.1.2.4.2 Aims

- Investigate whether the candidate Apiaceae species emit VOCs known to influence male mosquito behaviour.
- To examine the extracted material for any unknown compounds that could potentially be putative VOCs of biological importance to male mosquito behaviour.
- If present, quantify these emissions.
- If present, determine whether the different candidate Apiaceae emission characteristics appear to have sufficient inter-species difference that might cause the species to elicit different behaviour in mosquito behavioural assays.

5.1.2.4.3 Methods

The Apiaceae chosen for analysis were:

- *Angelica sylvestris* L.
- *Anthriscus sylvestris* (L.) Hoffm.
- *Conopodium majus* (Gouan) Loret
- *Daucus carota subsp. carota* L.
- *Heracleum sphondylium* L.

In addition, the *Hedera helix* data from the method development section was used as positive control. Negative controls were: hexane solvent only, hexane washed through clean sorbent tubes, and hexane washed tubes which had been processed through a 20 hr DHS cycle with no variable added to the headspace chamber.

All flower materials were collected as inflorescences from the field on the day of sampling. Florets were excised and prepared by cutting to a size suitable for placement in the DHS chamber and weighed before being placed in a 20 ml foil topped glass vial (Figure 84). Foil topping was used to isolate the excised stem headspace from the main floral headspace (McCallum *et al.* 2011).

Sorbent tubes were assembled using PoraPak™ Q 80-100 mesh porous polymer adsorbent, glass gas chromatography tubes and glass wool (Figure 83). 150 mg of PoraPak ™ was trapped between wads of packed glass wool in each of three tubes. Each of the identically prepared tubes had the following predefined roles (Woolfenden 2010a):

- One to pre-filter the air entering the dynamic headspace chamber
- One as the main capture tube, this being the principal target of analysis, and
- One as a breakthrough tube to collect any VOCs which overfill the adsorbent space available for that compound type in the capture tube.

Prior to first use the sorbent tubes were pre-washed with copious analytical grade hexane, >20 ml per tube, and dried by passing nitrogen through the tubes (Woolfenden 2010b). These washed and dried tubes were individually wrapped in aluminium foil and stored in a refrigerator at 4 °C until use.

The main body of the DHS chamber consisted of a glass desiccator with a single port at the top into which the Drechsel bottle head was fitted (Figure 85). The air was drawn through the system using a Vacuubrand CVC 3000 vacuum pump, with the following settings; Mode: Pump down, Speed = 95%, minimum = 450 mbar, delayed start = off and duration = 1200 minutes. The resulted in a mean flow rate of 355 ml/min, sd 12.3 ml/min and an in-chamber pressure of between 932 and 972 mbar. Sorbent tubes were attached to the DHS chamber using short sections of rubber. The pre-filter tube was mounted in the mouth of the glass inlet tube, and the capture and breakthrough tubes mounted in series in the outlet tube (Figure 85). This provided suitable conditions for the DHS in terms of air flow rate (Raguso and Pellmyr 1998) and presentation of the sorbent tubes (Tholl *et al.* 2006; Woolfenden 2010a)

Each DHS sampling session ran for twenty hours prior to the collection of the sorbent tubes for solvent desorption. Following sampling, the three sorbent tubes were removed from the DHS chamber and vacuum assembly and wrapped in aluminium foil. If the samples could not be processed immediately, the sorbent tubes were stored refrigerated at + 4 ° C, individually wrapped in foil; however, priority was placed on processing samples immediately. After each DHS session, the floral specimen was dried to constant weight and reweighed. Five replicates of each candidate flower species were conducted.

For the solvent desorption process, separation funnels were used to deliver 3 ml of HPLC grade hexane to each tube (). The product was collected in 4 ml glass gas chromatography vials. The internal standard, 2 µl of 0.0047 M 1-bromodecane (molarity by serial dilution), was added to each sample and vortexed for 15 seconds at 1400 rpm, post solvent desorption but prior to the evaporation of excess hexane under N_2. 1-Bromodecane has a molecular weight of 221.18, therefore, the ppm of 1-bromodecane added is 2.385 ppm. It follows that, therefore, the ppm of any other compound in the sample is the ppm of 1-bromodecane multiplied by the ratio of their integrated peak to that of 1-bromodecane (Herrmann 2010). Samples were then immediately concentrated to ~ 200 µl by evaporation under nitrogen,

supplied via a 6-port micro manifold (). 175 µl of this concentrated sample was pipetted into appropriately labelled GC autosampler vials (12mm x 32 mm x 6 mm, brown glass) with 0.3 ml inserts and PTFE/silicone septa. Following solvent desorption of the sample, the sorbent tubes were washed through with copious HPLC grade hexane (~20 ml), dried by evaporation under N_2 and wrapped in aluminium foil and stored at 4 °C until the subsequent sampling event.

Samples were processed using a PerkinElmer Clarus 500 Gas chromatograph with paired PerkinElmer Clarus 500 mass spectrometer. The GCMS injected 1 µl of each sample into an SLB™ - 5 ms (DB-5) fused silica capillary column, 30 m x 0.32 mm x 0.25 µm film (Sigma-Aldrich- 42131-04B) in splitless mode, using the GC's attached autosampler. The oven was held at 40 °C for 3 minutes and then ramped at 10 °C per minute to 250 °C where it was held for 5 minutes. A 3-minute solvent delay was added to the MS method, to allow the Hexane solvent to elute. The carrier gas was helium. Mass spectra were recorded in the electron impact mode, and the mass selective detector's scanning range was set to 40-350 m/z. GC/MS reports were collected digitally and stored on PC. The MS reports were imported into a Microsoft Access database using custom written R code (Appendix 6).

Data preparation was conducted in R Statistical Software (R Core Team 2015), and included peak alignment between chromatograms using the speaq package (Vu *et al.* 2011), to address the time shift problem already discussed, and normalisation of the chromatograms to the internal standard using the ChemoSpec package (Hanson 2014).

Data analysis again utilised R Statistical Software (R Core Team 2015), and the ChemoSpec package (Hanson 2014) for analysis of chromatographic data, including HCA. Data analysis of the data from MS analysis was analysed using code written specifically for that purpose (Appendix 7).

5.1.2.4.4 Results

Following peak alignment of the GC chromatograms, and normalisation to the 1-bromodecane internal standard, of the peaks in the GC and MS data, visualisation and analysis was carried out.

5.1.2.4.4.1 Aligned chromatograms, normalised to 1-bromodecane internal standard.

The peak aligned and normalised chromatograms of the five Apiaceae species were plotted together, and clearly show differences between the species (Figure 96). Whilst the larger peaks are clipped in this representation the plot still serves to display the considerable variation in the distribution and magnitude of the small peaks. Figure 97 shows the same chromatograms but with the y-axis adjusted to facilitate the comparison between the larger peaks. As already identified by the retention time calculator, the region of interest for the target VOCs is that less than 15 minutes 45 seconds and therefore peaks beyond this point on the chromatograms are unlikely to be involved in mosquito behaviour mediation due to their reduced volatility, but are indicative of the inter-species variation within this flower family. Conversely, using *H. sphondylium* as an example (Figure 98), the intra-species variation appears to be much less pronounced, with broadly similar peaks throughout the sample.

More detailed visualisations of the individual species chromatograms are shown in the species-specific chromatogram plots (Figure 100-25) and confirm that these differences continue to be apparent within the larger peaks also; Figure 99 shows the chromatogram of the negative control including 1-bromodecane internal standard, which showed a relatively flat baseline in comparison to the sample chromatograms. The apparent differences in the chromatograms were highly suggestive that substantial differences between the species were present.

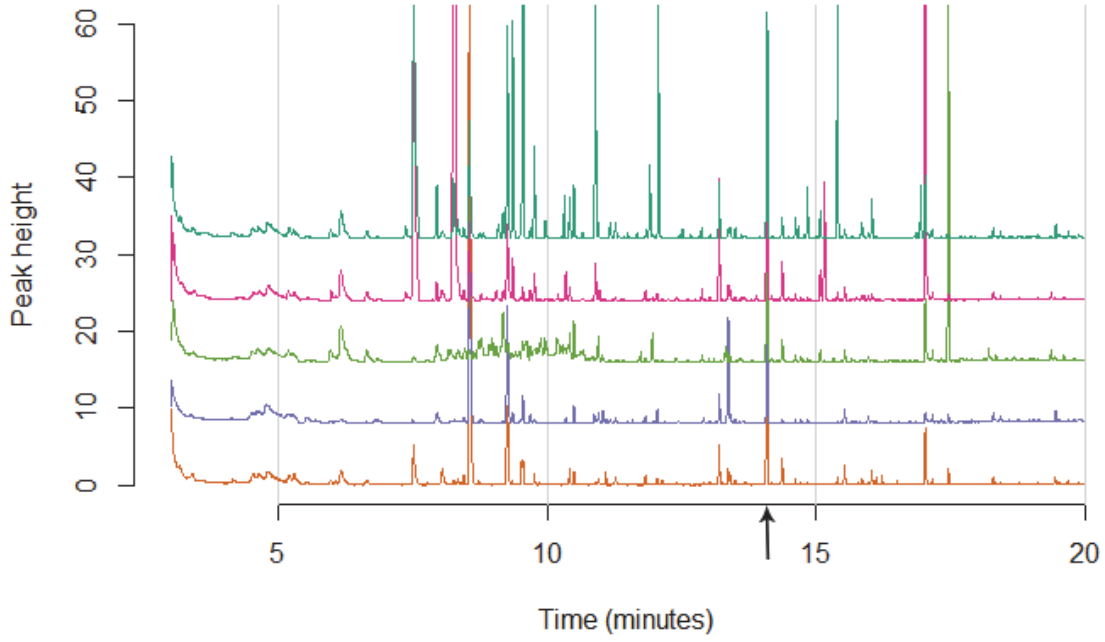


Figure 96 Offset chromatograms of the five Apiaceae species used in quantitative analysis. Chromatograms from top to bottom: *H. sphondylium, D. carota, C. majus, Ant. sylvestris, Ang. sylvestris*. peaks are truncated, but peak variation is clearly visible along the baseline. The arrow shows the location of the IS.

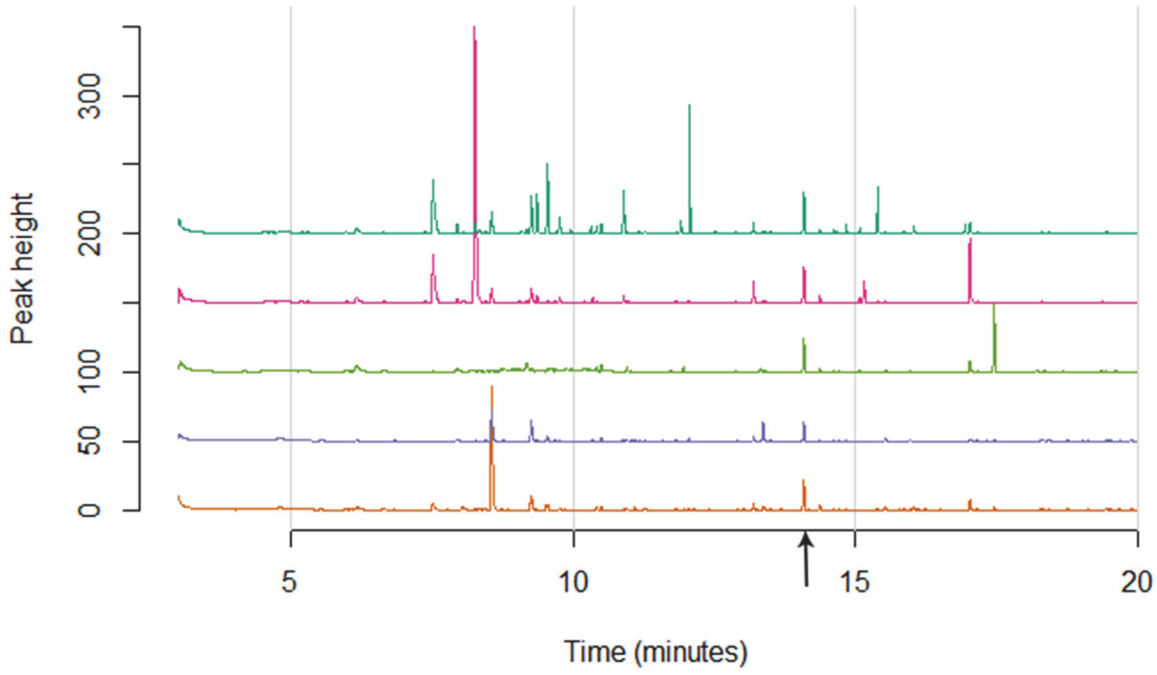


Figure 97 Offset chromatograms, zoomed out to allow comparison of larger peaks, of the five Apiaceae species used in quantitative analysis. Chromatograms from top to bottom: *H. sphondylium*, *D. carota*, *C. majus*, *Ant. sylvestris*, *Ang. sylvestris*. The arrow shows the location of the IS.

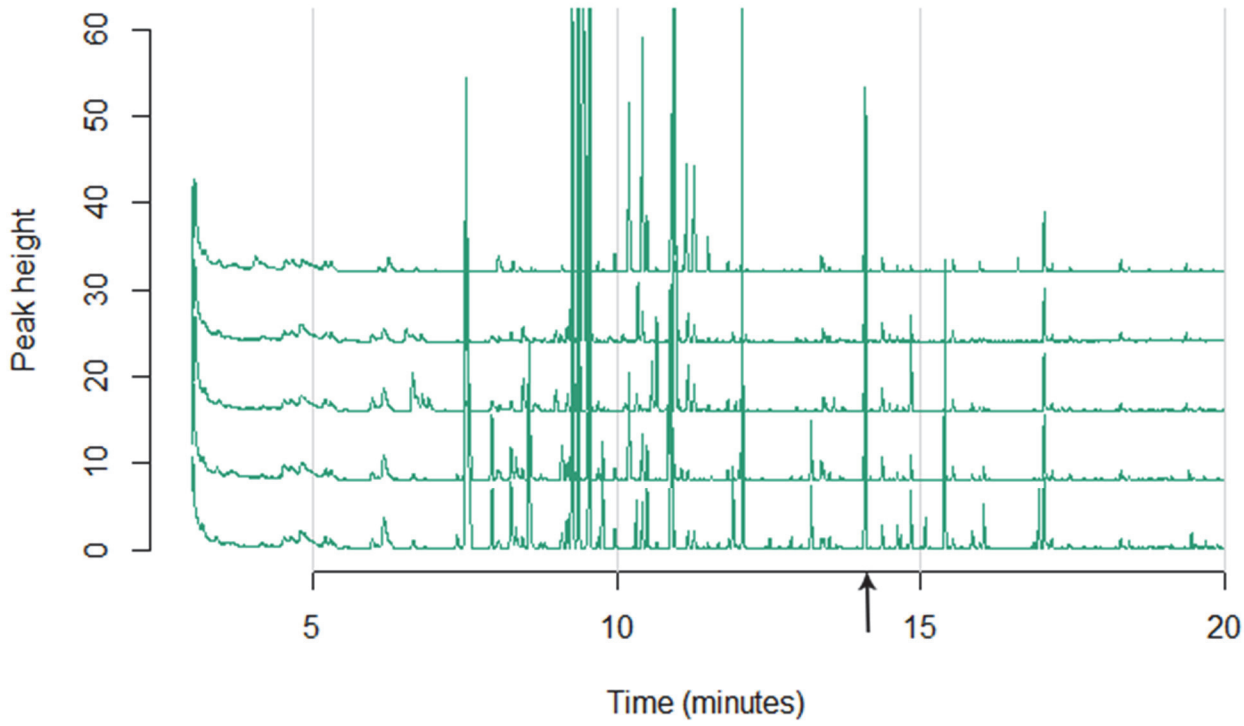


Figure 98 Offset chromatograms of the five repeats of *H. sphondylium*. The largest peaks are truncated, but the intra-species differences in the presence of peaks is much reduced compared to that seen in the inter-species graph. The arrow shows the location of the IS.

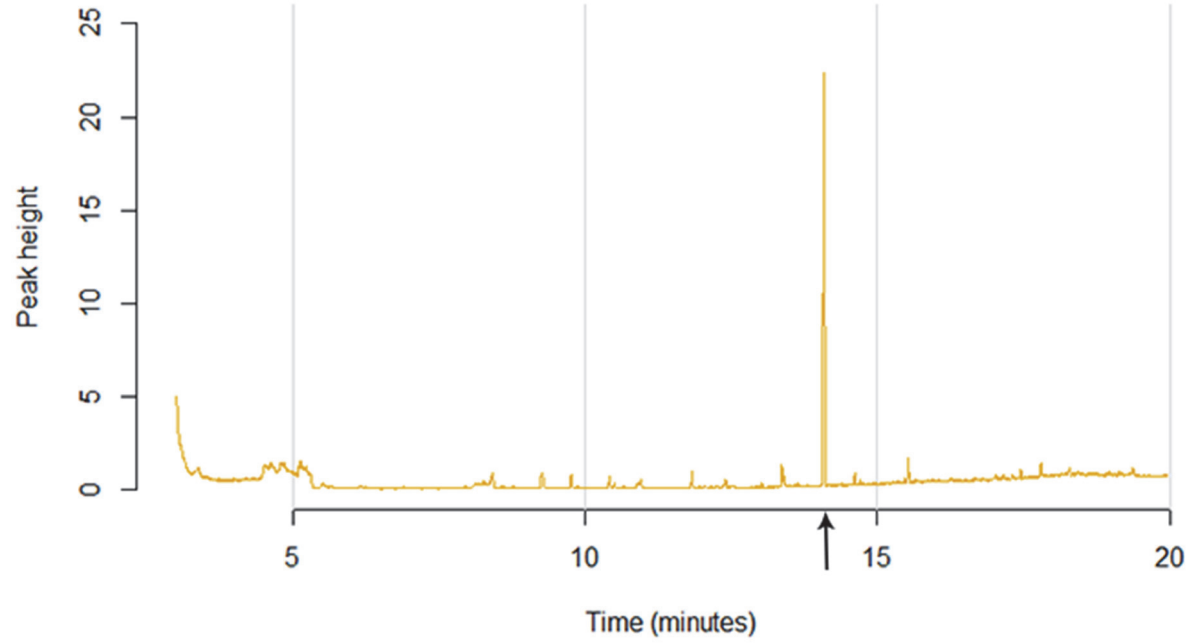


Figure 99 Chromatogram of the hexane blank negative control, normalised against the internal standard (1-bromodecane @ 2.385 ppm) peak at 14.15 minutes. The arrow shows the location of the IS.

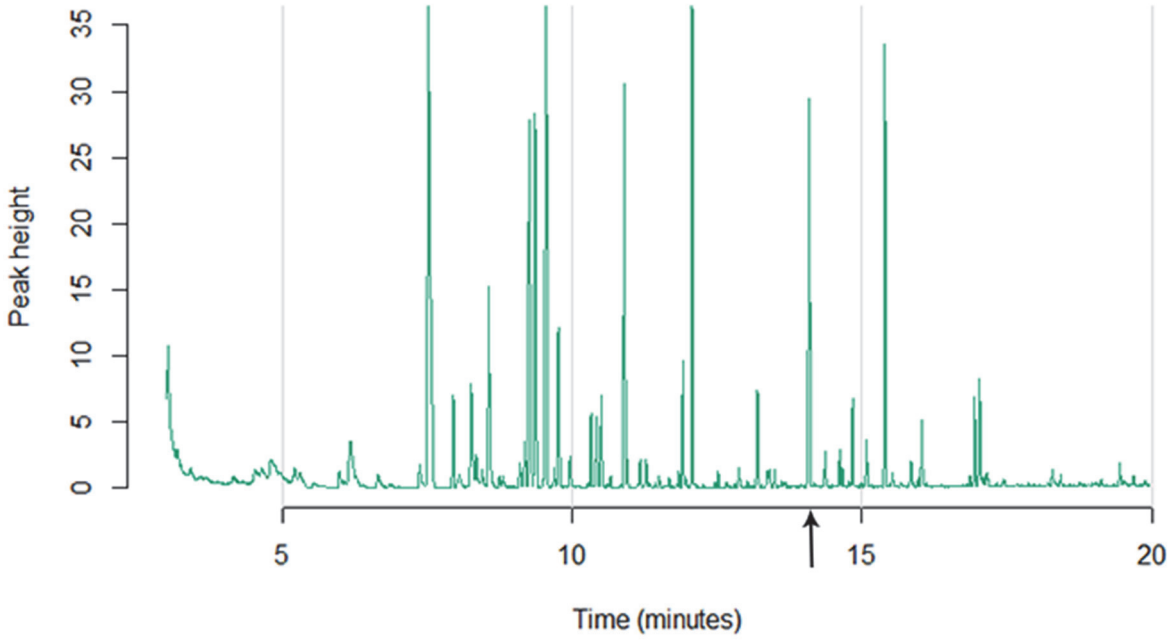


Figure 100 *Heracleum sphondylium* exemplar chromatogram, normalised to the IS (1-bromodecane peak @ 2.385 ppm at 14.15 minutes). The arrow shows the location of the IS.

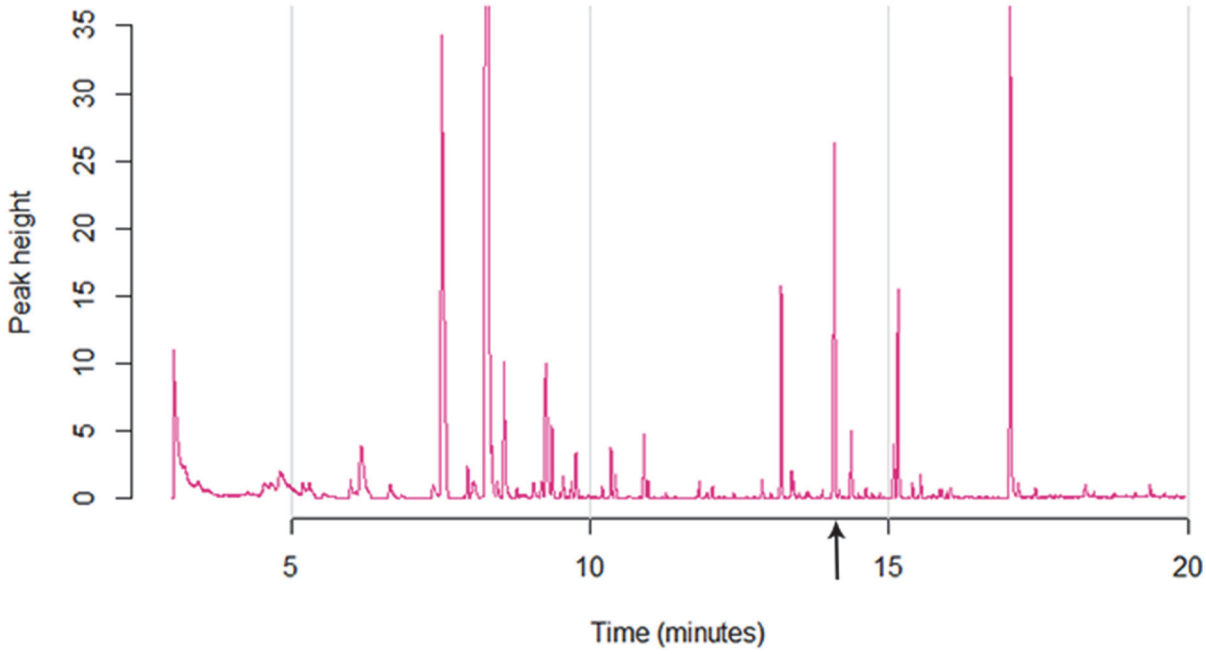


Figure 101 *Daucus carota* exemplar chromatogram, normalised to the IS (1-bromodecane peak @ 2.385 ppm at 14.15 minutes).

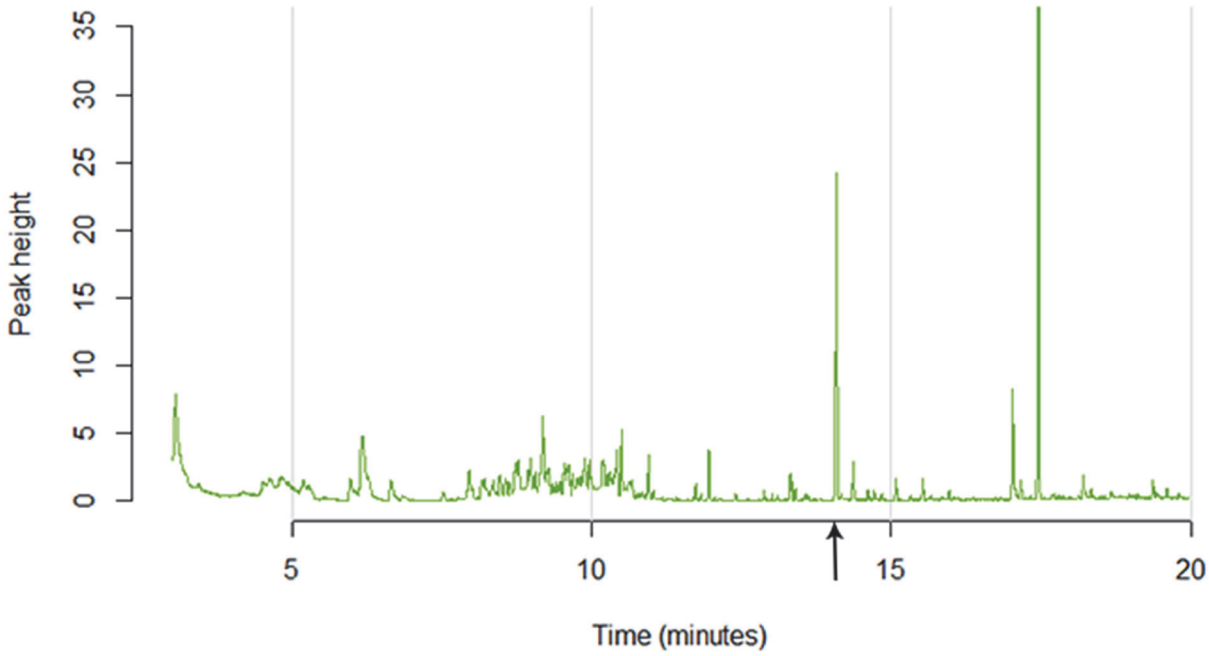


Figure 102 *Conopodium majus* exemplar chromatogram, normalised to the IS (1-bromodecane peak @ 2.385 ppm at 14.15 minutes). The arrow shows the location of the IS.

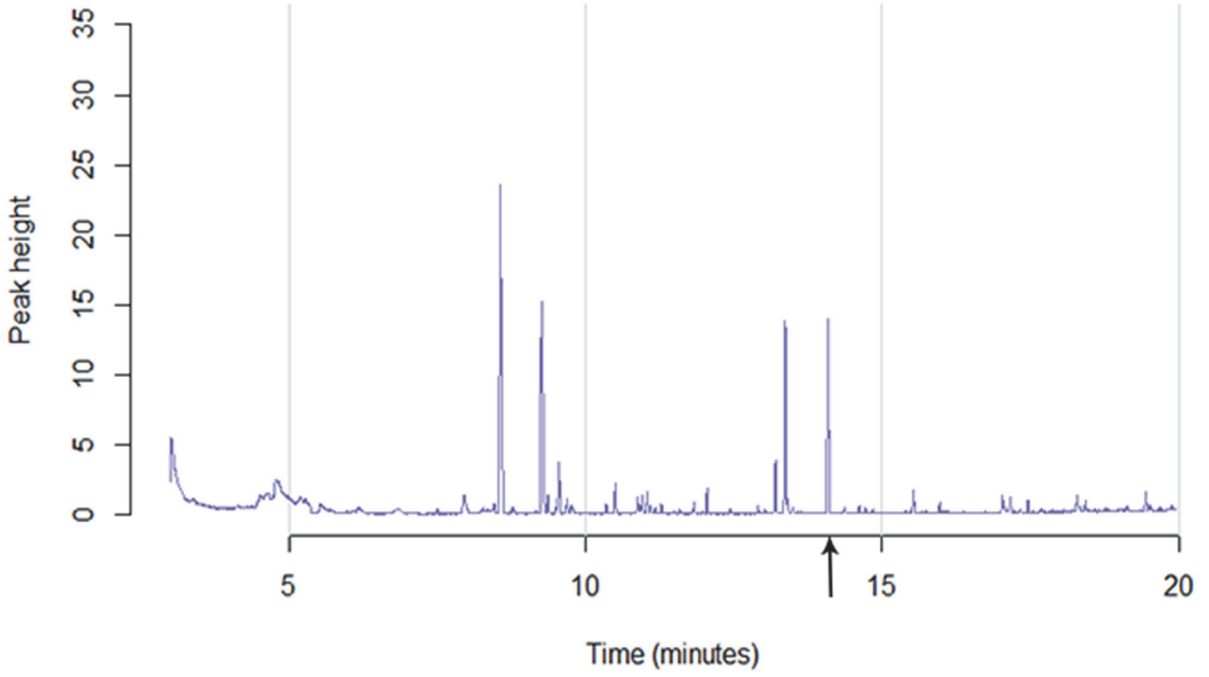


Figure 103 *Anthriscus sylvestris* exemplar chromatogram, normalised to the IS (1-bromodecane peak @ 2.385 ppm at 14.15 minutes) Here the IS peak (arrowed) reaches a lower height than in other chromatograms shown. This is due to the peak having a broad base in this instance, and as the quantification is based on the area under the peak, rather than peak height the height appears reduced.

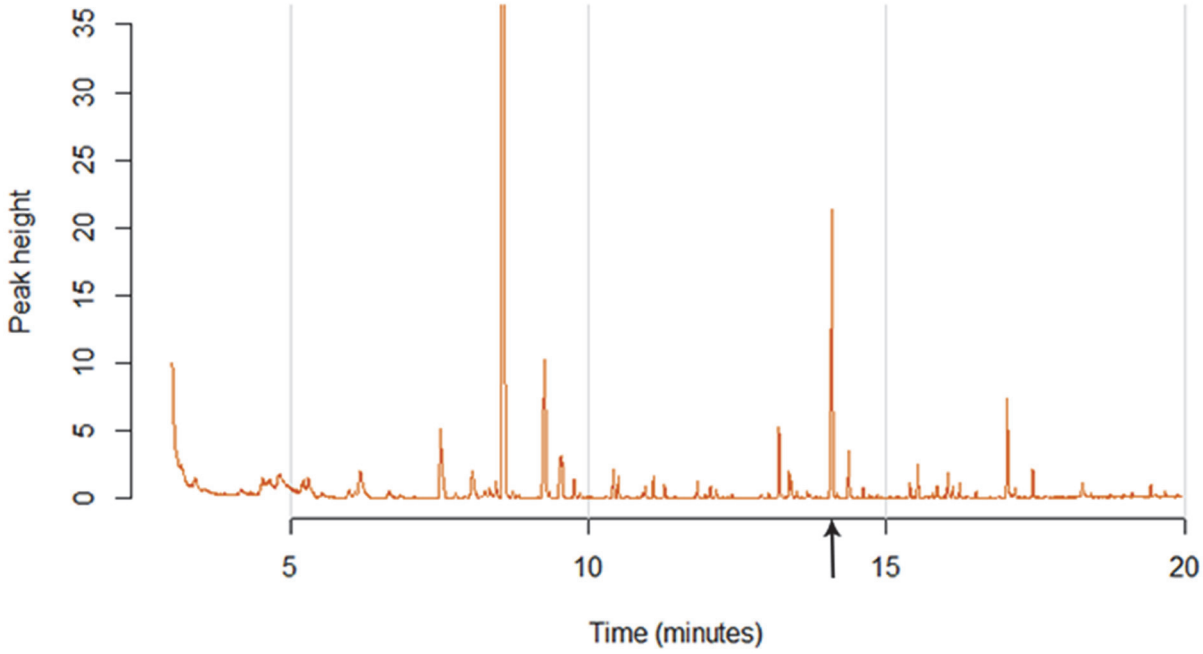


Figure 104 *Angelica sylvestris* exemplar chromatogram, normalised to the IS (1-bromodecane peak @ 2.385 ppm at 14.15 minutes). The arrow shows the location of the IS.

The results of HCA are shown in Figure 105. HCA used the 'complete' clustering method and 'Euclidean' distance to find similar clusters. In this analysis, each of the 25 chromatograms is, initially, treated as a separate cluster then the algorithm iteratively joins the two most similar clusters until all clusters have been joined to find the computed distances between the constituent clusters. The clustering algorithm was blind to *a priori* group membership, and so the displayed grouping is based upon the similarity in chromatograms only. Generally, the species are quite well clustered, and where species are split between different clusters, the distance (y-axis value) is not large. Where clustering by species is sometimes incomplete, it was suggestive of intra-species variation in VOC emissions. *C. majus* specimens were not well grouped by HCA and were distributed across the clusters; potentially, this is due to the very low emissions from this species causing its characteristics to be, relatively, indistinct within the analysis due to its lack of strong features. Another noteworthy clustering feature was the split of the *D. carota* specimens into two distinct groups, the distance between these groups is very large, and may indicate a genetic or other variation between individual plants in the field.

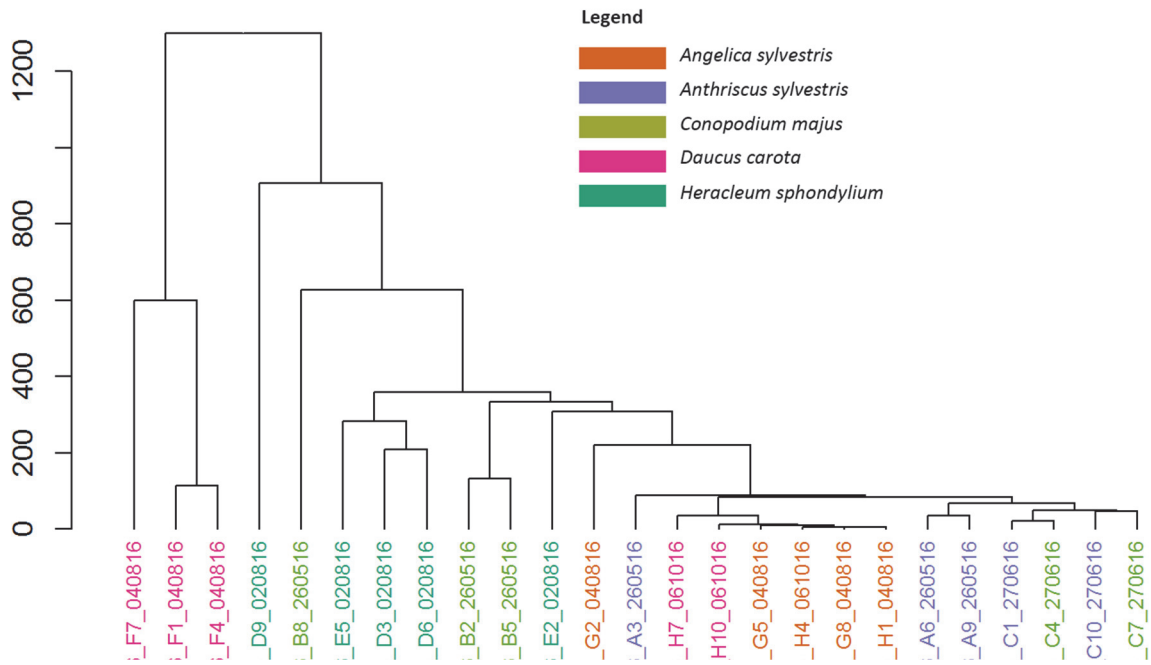


Figure 105 Results of HCA between the flower species when applied to the complete chromatogram. It shows some grouping by species, but not complete grouping. For example, four of the five *Ang. sylvestris* samples are tightly grouped with one slightly less close. Conversely, *C. majus* is not closely clustered, with separation at the second highest level for the B8 specimen. B2 and B5 are close together, but far from B8, C4 and C7 specimens.

5.1.2.4.4.2 Details of the MS identification of compounds

Included here are exemplar details from GCMS reports for the top 20 peaks, by peak area, for the time period 3 to 16 minutes. Many of the panel compounds do not feature in these top twenty lists reinforcing the need to conduct the larger scale 150 peak MS library searches employed in this study. The complete 150 peak data that correspond to the chromatograms shown in Figure 100 through Figure 104 are included in Appendix 8 through to Appendix 12. Subsequent analyses of these data focus on the compounds known to affect mosquito behaviour (Table 20**)** (Verhulst *et al.* 2011; Otienoburu *et al.* 2012; Nyasembe and Torto 2014; Pitts *et al.* 2014) rather than peak area based analyses of the whole headspace.

Table 21 Top 20 peaks from *Angelica sylvestris*, that eluted before 16 minutes, ordered by the area under the peak. Many of the target compounds do not appear in the top 20 peaks, reinforcing the need to utilise 150 peak GCMS reporting and subsequent analyses.

RT	Height	Area	Name_1	CAS_No_1	Match_1	Name_2	CAS_No_2	Match_2
8.588	308335552	10791911.0	β-Myrcene	123-35-3	904	β-Myrcene	123-35-3	896
3.042	34332016	3142177.2	Heptane	142-82-5	908	Heptane	142-82-5	904
9.317	34883192	1467954.4	Bicyclo[3.1.0]hex-2-ene, 4-methyl-1-(1-methylethyl)-	28634-89-1	927	Bicyclo[3.1.0]hexane, 4-methylene-1-(1-methylethyl)-	3387-41-5	857
4.846	5504035	977379.7	2-Pentene, 4,4-dimethyl-, (Z)-	762-63-0	769	3-Hexen-2-one	763-93-9	737
7.542	17394892	778868.6	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	855	1R-α-Pinene	7785-70-8	854
3.421	4701248	660369.1	Cyclohexane, methyl-	108-87-2	783	Diethylcyanamide	617-83-4	775
3.017	56182440	547938.4	Heptane	142-82-5	872	Heptane	142-82-5	865
13.183	17896196	458147.0	4-Hydroxy-2-methylacetophenone	875-59-2	845	4-Hydroxy-3-methylacetophenone	876-02-8	836
6.271	6168216	448270.8	p-Xylene	106-42-3	847	p-Xylene	106-42-3	846
4.554	4316444	391704.8	3-Hexanone	589-38-8	740	3-Hexanone	589-38-8	724
4.654	4089049	383010.8	2-Hexanone, 4-hydroxy-3-propyl-	62338-17-4	651	2-Hexanone	591-78-6	641
8.079	6975570	374331.5	Benzaldehyde	100-52-7	817	Benzaldehyde	100-52-7	807
5.312	4780042	363796.2	Heptane, 2,4-dimethyl-	2213-23-2	869	Hexane, 2,3,4-trimethyl-	921-47-1	853
9.579	10427734	304649.4	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	779	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	772
14.425	11866621	303390.9	Propanoic acid, 2-methyl-, 3-hydroxy-2,4,4-trimethylpentyl ester	74367-34-3	816	Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester	74367-31-0	797
5.233	4257425	289382.9	Cyclotrisiloxane, hexamethyl-	541-05-9	871	Cyclotrisiloxane, hexamethyl-	541-05-9	837
9.550	9801626	260140.2	Benzaldehyde, 2-hydroxy-	90-02-8	841	Benzaldehyde, 2-hydroxy-	90-02-8	832
15.600	8616859	202519.9	3-Isopropoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane		525	3-Butoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane	72439-84-0	517
3.600	1744205	197354.0	Hydroperoxide, hexyl	4312-76-9	590	1-Pentanol, 4-methyl-	626-89-1	589
10.463	7251467	183302.5	Decane	124-18-5	798	Hexane, 3,3-dimethyl-	563-16-6	797

Table 22 Top 20 peaks from *Anthriscus sylvestris*, that eluted before 16 minutes, ordered by the area under the peak. Many of the target compounds do not appear in the top 20 peaks, reinforcing the need to utilise 150 peak GCMS reporting and subsequent analyses.

RT	Height	Area	Name_1	CAS_No_1	Match_1	Name_2	CAS_No_2	Match_2
8.590	106502520	4146044.5	β-Myrcene	123-35-3	877	β-Myrcene	123-35-3	869
3.021	25055918	3076596.5	Heptane	142-82-5	842	Heptane	142-82-5	841
9.306	68286344	2477055.2	Limonene	138-86-3	887	Cyclohexene, 1-methyl-4-(1-methylethenyl)-, (S)-	5989-54-8	843
4.805	9548238	2228224.5	3-Penten-2-one, 4-methyl-	141-79-7	741	3-Penten-2-one, 4-methyl-	141-79-7	740
13.456	62354716	1678002.8	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	631	Cyclopropane, trimethyl(2-methyl-1-propenylidene)-	14803-30-6	628
3.389	4425264	814707.4	Diethylcyanamide	617-83-4	671	Diethylcyanamide	617-83-4	667
4.621	5274605	554528.1	Cyclohexane, 1,1'-(2-methyl-1,3-propanediyl)bis-	2883-08-1	450	2-Oxo-n-propyl-2-(veratrylidenehydrazino)acetamide	339241-37-1	447
9.595	16427211	514489.9	3-Carene	13466-78-9	830	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	828
5.207	4555402	454539.1	Cyclotrisiloxane, hexamethyl-	541-05-9	834	Cyclotrisiloxane, hexamethyl-	541-05-9	811
4.528	4259968	432008.2	3-Hexanone	589-38-8	800	3-Hexanone	589-38-8	785
5.299	4470925	407708.2	Heptane, 2,4-dimethyl-	2213-23-2	842	Hexane, 2,3,5-trimethyl-	1069-53-0	832
13.305	17134146	403353.0	Isobornyl acetate	125-12-2	807	Acetic acid, 1,7,7-trimethyl-bicyclo[2.2.1]hept-2-yl ester	92618-89-8	806
8.075	5848204	360814.5	Benzaldehyde	100-52-7	793	Benzaldehyde	100-52-7	783
5.504	3156976	293924.4	Formic acid, 1-methylethyl ester	625-55-8	722	Butanal, 3-hydroxy-	107-89-1	722
10.567	10051937	275774.2	Nonanal	124-19-6	830	Nonanal	124-19-6	821
6.936	1645726	203808.6	2,4,6-Trimethyl-1-nonene	55771-40-9	784	Pentane, 2,3,4-trimethyl-	565-75-3	713
12.133	8039638	191582.1	Decanal	112-31-2	891	Decanal	112-31-2	846
15.655	7355326	190413.8	3-Isopropoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane		630	1,1,1,3,5,5,7,7,7-Nonamethyl-3-(trimethylsiloxy)tetrasiloxane	38146-99-5	619
6.262	2150687	189994.7	p-Xylene	106-42-3	771	p-Xylene	106-42-3	769
11.094	6812706	188223.6	Benzyl nitrile	140-29-4	865	Benzyl nitrile	140-29-4	857

Table 23 Top 20 peaks from *Conopodium majus*, that eluted before 16 minutes, ordered by the area under the peak. Many of the target compounds do not appear in the top 20 peaks, reinforcing the need to utilise 150 peak GCMS reporting and subsequent analyses.

RT	Height	Area	Name_1	CAS_No_1	Match_1	Name_2	CAS_No_2	Match_2
3.067	38768772	5147640.0	Heptane	142-82-5	870	Heptane	142-82-5	870
14.143	118884808	2867399.0	Undecane, 1-bromo-	693-67-4	897	Decane, 1-bromo-	112-29-8	896
6.279	21186530	2102202.8	p-Xylene	106-42-3	876	p-Xylene	106-42-3	871
4.888	7480940	1853625.4	Diethylcyanamide	617-83-4	568	Cyclopropane, 1,1,2,2-tetramethyl-	4127-47-3	556
9.290	28191692	1498918.6	1-Hexanol, 2-ethyl-	104-76-7	778	1-Hexanol, 2-ethyl-	104-76-7	764
3.465	5411269	1090639.0	Cyclohexane, methyl-	108-87-2	649	2,4-Azetidinedione, 3,3-diethyl-1-methyl-	69315-91-9	639
4.675	6815676	700816.3	Pentanal, 2-methyl-	123-15-9	643	Pentanal, 2-methyl-	123-15-9	624
8.079	10432172	690936.1	Benzaldehyde	100-52-7	856	Benzaldehyde	100-52-7	840
10.542	24336838	684359.4	Nonanal	124-19-6	864	Nonanal	124-19-6	864
10.462	15308833	593397.1	Decane, 2,4-dimethyl-	2801-84-5	866	Undecane	1120-21-4	855
9.403	9024909	562871.8	Nonane, 3-methyl-	5911-04-6	760	Heptane, 2,2-dimethyl-	1071-26-7	739
4.574	5325346	512571.8	3-Hexanone	589-38-8	800	3-Hexanone	589-38-8	778
5.249	6611838	508457.4	Cyclotrisiloxane, hexamethyl-	541-05-9	841	Cyclotrisiloxane, hexamethyl-	541-05-9	813
10.219	9989887	503975.8	Undecane, 4,7-dimethyl-	17301-32-5	868	Undecane, 5-methyl-	1632-70-8	861
5.328	5210856	489077.3	Heptane, 2,4-dimethyl-	2213-23-2	846	Heptane, 2,4-dimethyl-	2213-23-2	832
9.666	11132638	466402.6	Decane, 2,5,9-trimethyl-	62108-22-9	835	Heptane, 3-ethyl-2-methyl-	14676-29-0	824
12.095	18644480	456356.1	Decanal	112-31-2	890	Decanal	112-31-2	886
6.727	7233508	454847.3	o-Xylene	95-47-6	874	p-Xylene	106-42-3	870
10.010	11171575	451318.1	Decane, 2,4,6-trimethyl-	62108-27-4	859	Undecane, 3,7-dimethyl-	17301-29-0	844
8.980	10922498	439906.2	Decane, 2,5,9-trimethyl-	62108-22-9	859	Octane, 3-methyl-	2216-33-3	858

Table 24 Top 20 peaks from *Daucus carota*, that eluted before 16 minutes, ordered by the area under the peak. Many of the target compounds do not appear in the top 20 peaks, reinforcing the need to utilise 150 peak GCMS reporting and subsequent analyses.

RT	Height	Area	Name_1	CAS_No_1	Match_1	Name_2	CAS_No_2	Match_2
8.289	771978624	30089376.0	Bicyclo[3.1.0]hexane, 4-methylene-1-(1-methylethyl)-	3387-41-5	932	β-Phellandrene	555-10-2	917
7.543	130602576	6476428.0	1R-α-Pinene	7785-70-8	927	Tricyclo[2.2.1.0(2,6)]heptane, 1,3,3-trimethyl-	488-97-1	917
3.050	33734860	3807942.8	Heptane	142-82-5	850	Heptane	142-82-5	834
15.215	59746288	1573149.2	Cyclohexene, 3-(1,5-dimethyl-4-hexenyl)-6-methylene-, [S-(R,*S*)]-	20307-83-9	864	1H-3a,7-Methanoazulene, octahydro-3,8,8-trimethyl-6-methylene-, [3R-(3α,3aβ,7β,8aα)]-		848
13.259	61110964	1528960.2	Bicyclo[2.2.1]heptan-2-ol, 1,7,7-trimethyl-, acetate, (1S-endo)-	5655-61-8	872	Acetic acid, 1,7,7-trimethyl-bicyclo[2.2.1]hept-2-yl ester	92618-89-8	870
9.294	37475056	1457341.0	Cyclohexene, 1-methyl-4-(1-methylethenyl)-, (S)-	5989-54-8	838	Limonene	138-86-3	827
8.586	38350504	1288304.6	β-Myrcene	123-35-3	888	β-Myrcene	123-35-3	886
4.822	6967380	1124963.4	3-Penten-2-one, 4-methyl-	141-79-7	813	3-Penten-2-one, 4-methyl-	141-79-7	794
6.270	13696787	1096061.1	p-Xylene	106-42-3	861	p-Xylene	106-42-3	854
3.431	4573488	560801.9	2,4-Azetidinedione, 3,3-diethyl-1-methyl-	69315-91-9	819	Cyclohexane, methyl-	108-87-2	779
9.394	20386110	549105.1	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	877	1R-α-Pinene	7785-70-8	848
14.423	19329732	494493.0	Propanoic acid, 2-methyl-, 3-hydroxy-2,4,4-trimethylpentyl ester	74367-34-3	837	Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester	74367-31-0	779
10.910	18228030	430244.1	Bicyclo[3.1.0]hex-2-ene, 4,4,6,6-tetramethyl-	19487-09-3	815	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	812
15.110	15510813	402567.9	1,3-Cyclohexadiene, 5-(1,5-dimethyl-4-hexenyl)-2-methyl-, [S-(R,*S*)]-	495-60-3	776	1H-3a,7-Methanoazulene, 2,3,4,7,8,8a-hexahydro-3,6,8,8-tetramethyl-, [3R-(3α,3aβ,7β,8aα)]-		764
9.796	12884567	396258.4	Cyclohexene, 4-methylene-1-(1-methylethyl)-	99-84-3	822	Bicyclo[3.1.0]hexan-2-ol, 2-methyl-5-(1-methylethyl)-	546-79-2	793
7.849	9251357	387190.6	Bicyclo[2.2.1]heptane, 7,7-dimethyl-2-methylene-	471-84-1	833	Camphene	79-92-5	809
10.458	14392633	383521.7	Octane, 3,5-dimethyl-	15869-93-9	701	Decane, 2,4,6-trimethyl-	62108-27-4	696
4.671	3750230	372053.2	Pentanal, 2,3-dimethyl-	32749-94-3	558	Pentanal, 2-methyl-	123-15-9	547
8.075	5043959	327711.9	Benzaldehyde	100-52-7	796	Benzaldehyde	100-52-7	789
4.562	3731922	297431.2	3-Hexanone	589-38-8	770	3-Hexanone	589-38-8	758

Table 25 Top 20 peaks from *Heracleum sphondylium*, that eluted before 16 minutes, ordered by the area under the peak. Many of the target compounds do not appear in the top 20 peaks, reinforcing the need to utilise 150 peak GCMS reporting and subsequent analyses.

RT	Height	Area	Name_1	CAS_No_1	Match_1	Name_2	CAS_No_2	Match_2
12.133	401357312	9623534.0	Acetic acid, octyl ester	112-14-1	909	Acetic acid, octyl ester	112-14-1	905
7.546	164453040	7928732.5	1R-α-Pinene	7785-70-8	941	1R-α-Pinene	7785-70-8	928
9.583	219240464	6576075.0	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	951	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	933
9.296	119464696	4027283.5	Limonene	138-86-3	920	D-Limonene	5989-27-5	878
9.396	121473832	3405012.8	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	931	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	927
15.454	144289952	3285783.2	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (E)-	18794-84-8	916	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (Z)-	28973-97-9	914
10.913	131490792	3253326.8	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	939	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	915
14.142	126904256	3029354.2	Dodecane, 1-bromo-	143-15-7	904	Decane, 1-bromo-	112-29-8	904
8.587	64895792	2161591.5	β-Myrcene	123-35-3	897	β-Myrcene	123-35-3	887
9.792	51867976	1475832.2	Bicyclo[3.1.0]hexan-2-ol, 2-methyl-5-(1-methylethyl)-	546-79-2	877	1,4-Cyclohexadiene, 1-methyl-4-(1-methylethyl)-	99-85-4	862
7.854	30132960	1354550.5	Camphene	79-92-5	910	Camphene	79-92-5	891
4.833	7775909	1336538.0	2-Pentene, 4,4-dimethyl-, (Z)-	762-63-0	736	3-Hexen-2-one	763-93-9	712
3.054	31102078	1304209.2	Heptane	142-82-5	847	Heptane	142-82-5	837
8.288	32426646	1158639.4	Bicyclo[3.1.0]hex-2-ene, 4-methyl-1-(1-methylethyl)-	28634-89-1	879	β-Phellandrene	555-10-2	835
6.279	14140980	1154420.8	p-Xylene	106-42-3	858	p-Xylene	106-42-3	857
11.938	41230332	985264.8	3-Octen-1-ol, acetate, (Z)-	69668-83-3	812	3-Octen-2-ol, (E)-	57648-55-2	806
10.538	29882360	864555.3	Nonanal	124-19-6	870	Nonanal	124-19-6	865
13.258	31687550	793801.4	Bornyl acetate	76-49-3	852	Isobornyl acetate	125-12-2	849
10.258	23916728	680321.3	trans-Linaloloxide		829	2-Furanmethanol, 5-ethenyltetrahydro-α,α,5-trimethyl-, trans-	34995-77-2	819
14.896	28915246	658234.9	Dodecanal	112-54-9	890	Dodecanal	112-54-9	879

5.1.2.4.4.3 Analysis of MS results related to the panel of target VOCs

Following analysis of the GC data, the MS data was utilised to enable the comparison to the table of target VOCs (Table 20) shown to influence male mosquito behaviour. When analysing the MS output reports the match score is a measure of how well matched an unknown compound is to the library spectrum. An excellent match would result in a value of 900 or greater, a good match would yield 800 – 900, a fair match would yield 700 – 800 and match scores of less than 600 would be a very poor match (Mikaia 2008). Preliminary analyses showed that match scores for alkane standards used in the retention curve creation and the 1-Bromodecane used as the internal standard received match scores between 800 and 900 in some samples. Therefore, only identifications where the match score was >= 800 were used in any part of these analyses, which should be considered a good match (Mikaia 2008; Stein 2012).

5.1.2.4.4.4 Comparison of target VOCs between sorbent sampling tubes

The sum of the number of different target VOCs was calculated for each candidate flower species for each of the three sorbent tubes involved in the DHS process. The breakthrough tubes showed no target VOCs, with the exception of two *D. carota* samples and three *Ang. sylvestris* samples which showed hexan-1-ol to be present. Pre-filter tubes showed a more diverse range of target VOCs, which was expected as these would be collected from the atmosphere around the DHS chamber as the air was drawn into the system by the vacuum unit. These VOCs, however, often had no corresponding peak in the capture tube and so it was accepted that the pre-filter was working as designed and preventing any target VOCs from sources other than the floral sample from reaching the capture tube. As confirmation of this assertion, in the negative only hexan-1-ol was identified on the capture tube, confirming that the pre-filter was efficient and that the chamber was not allowing ingress of contamination from any other source. With only isolated target VOC identifications on the breakthrough tubes, the lack of co-incidence of pre-filter and capture compounds and the evidence from the negative controls, it was accepted that the total VOCs collected on the capture tube could be considered the total emission from the specimen.

The five different Apiaceae species yielded different numbers of discrete target VOCs. *H. sphondylium* yielded the highest number with 13 different target compounds. *D. carota* and

Ant. sylvestris each emitted 9 different target compounds, *Ang. sylvestris* emitted 7 different target compounds and *C. majus* emitted the fewest discrete target compounds with 3. The collated data for the number of different target VOCs entrained on the capture tubes only for the five species, and the three negative controls are shown in Figure 106, the single VOC reported in the 3 negative controls was hexan-1-ol.

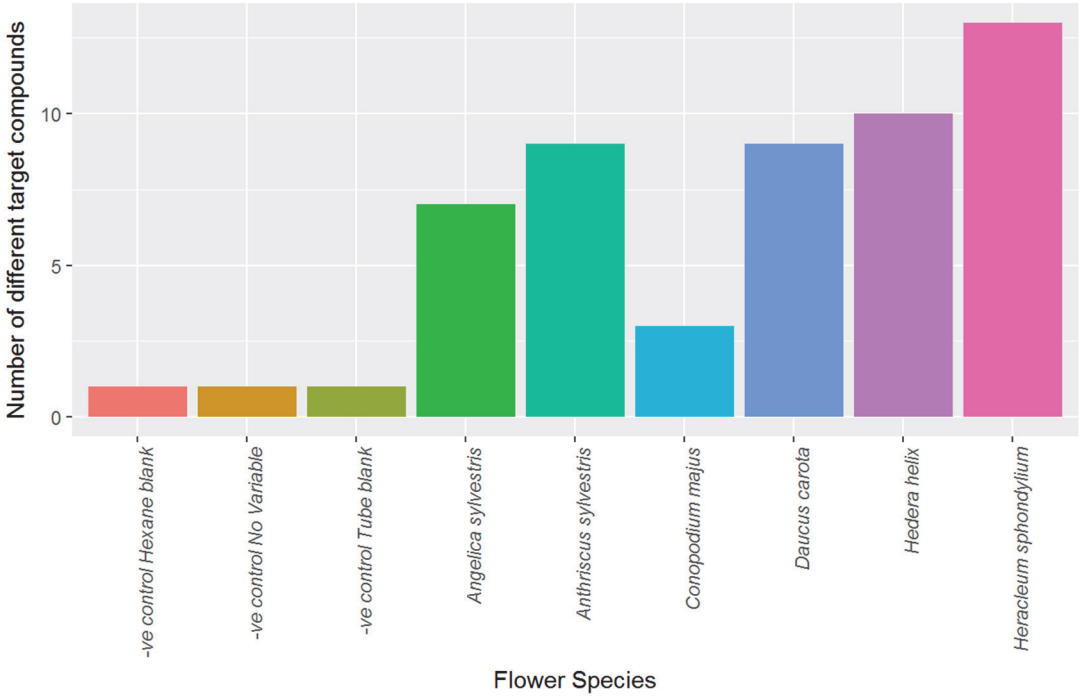


Figure 106 Number of target VOCs collected on the capture tube for the three types of negative control and the flower species. Only VOC identifications where TurboMass Match score was >= 800. The target VOC found in all negative controls was Hexan-1-ol and presumably a product from the Hexane solvent.

5.1.2.4.4.5 Comparison of intra- and inter-species chemical composition

Intraspecies variation in the number of different compounds and their recorded levels varied considerably. Figures 33 to 37 show the range of parts per million for each of the target compounds. *H. sphondylium* (Figure 107) had a very high emission of β-ocimene, with one outlier at almost 60 ppm and a mean of 10 ppm, relative to the other compounds, and the other floral species emissions of β-ocimene; it also had the highest emissions in general. At the other extreme, *C. majus* had a highest mean ppm of < 0.6 ppm for hexan-1-ol. It is quite likely that this compound is associated with the hexane being use as the solvent for desorption as it was often present in negative controls. Therefore the highest-level emission of target VOCs in C. majus is < 0.05 ppm from benzaldehyde (Figure 110). *Ant. sylvestris*

(Figure 109), *Ang. sylvestris* (Figure 108) and *D. carota* (Figure 111) all fall between the extremes of *H. sphondylium* and *C. majus*, but show considerable variation between species and also show considerable variation within species. As the only non-Apiaceae family species here, *H. helix*, the positive control, had a distinctly different target VOC emission pattern, with higher levels of (E)-β-Farnesene and Furanol than any of the Apiaceae species.

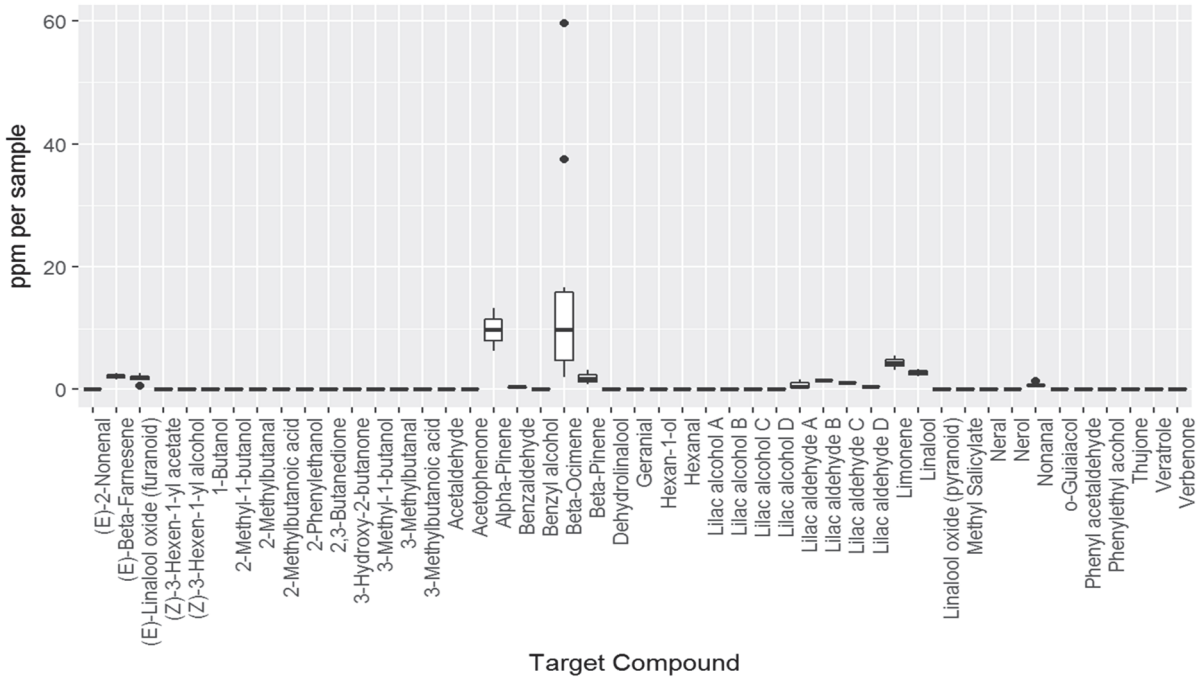


Figure 107 Ppm per VOC in samples from *H. sphondylium*, where TurboMass Match >= 800. *H.* sphondylium showed higher target VOC emissions and had more different target VOCs present than any of the other floral species. Y-axis is scaled to the data range. Dots are outliers (> 2 sd from the median).

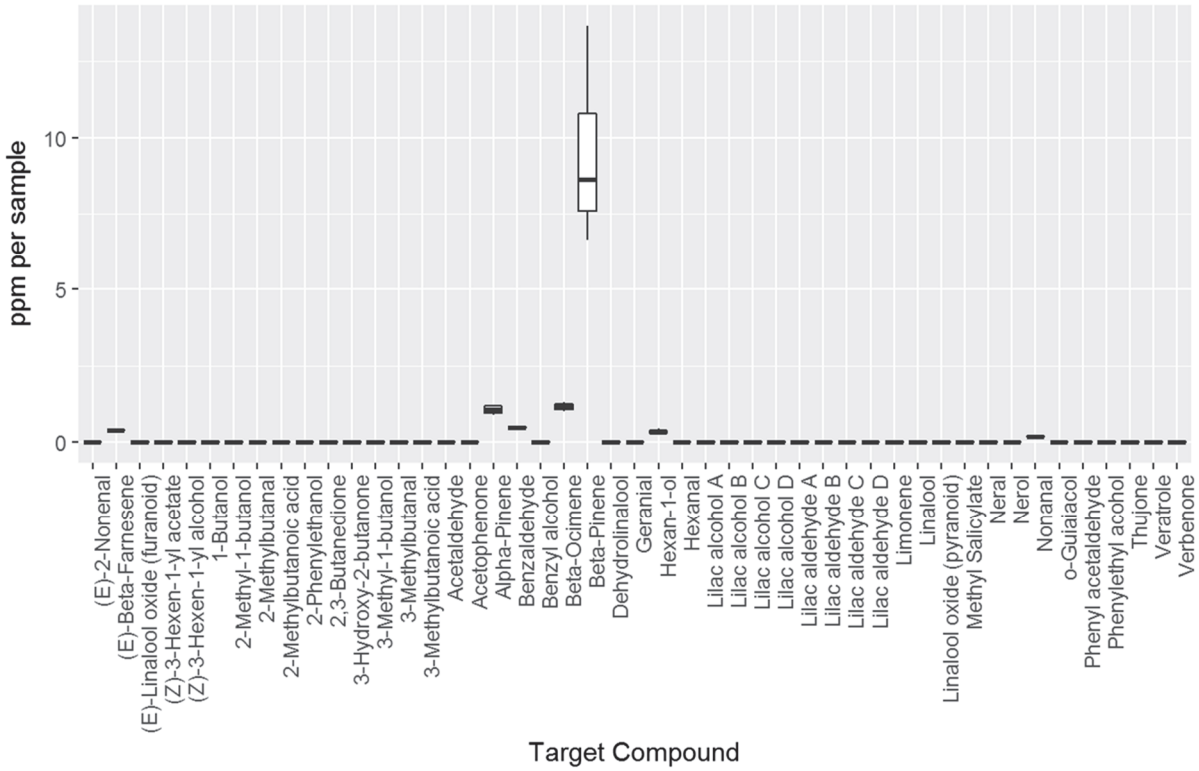


Figure 108 Ppm per VOC in samples from *Ang. sylvestris*, where TurboMass Match >= 800. Y-axis is scaled to the data range.

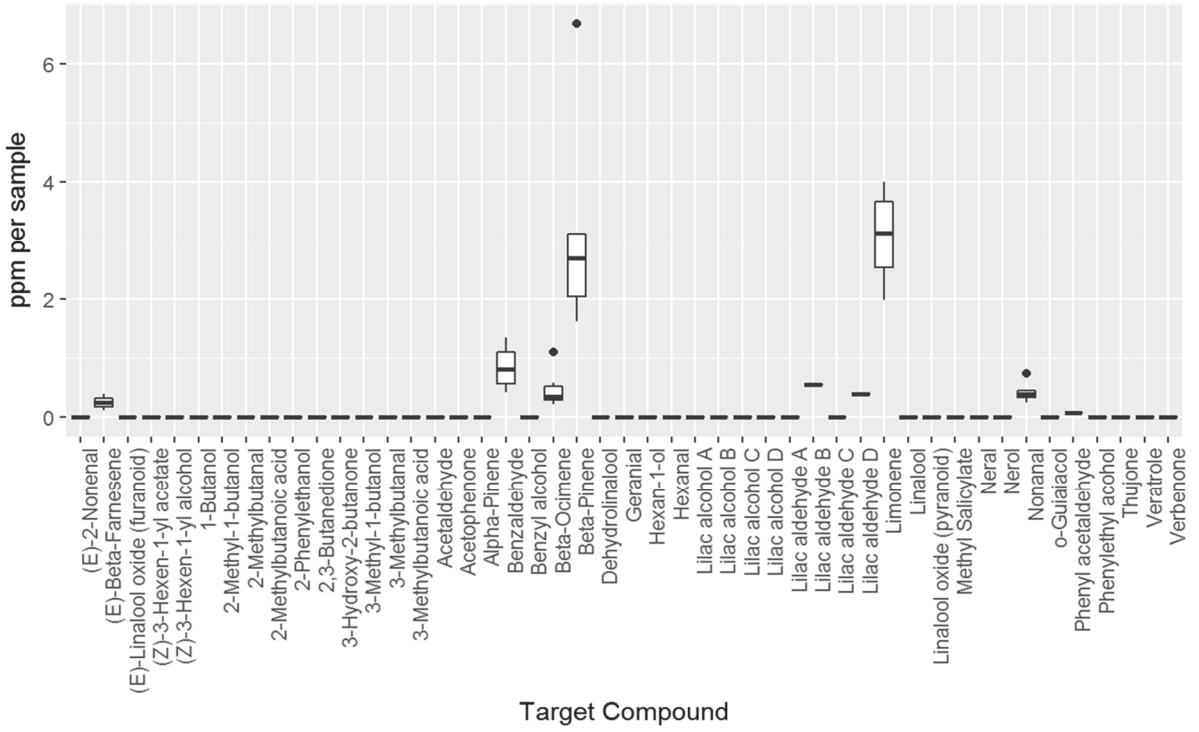


Figure 109 Ppm per VOC in samples from *Ant. sylvestris*, where TurboMass Match >= 800. Y-axis is scaled to the data range. Dots are outliers (> 2 sd from the median).

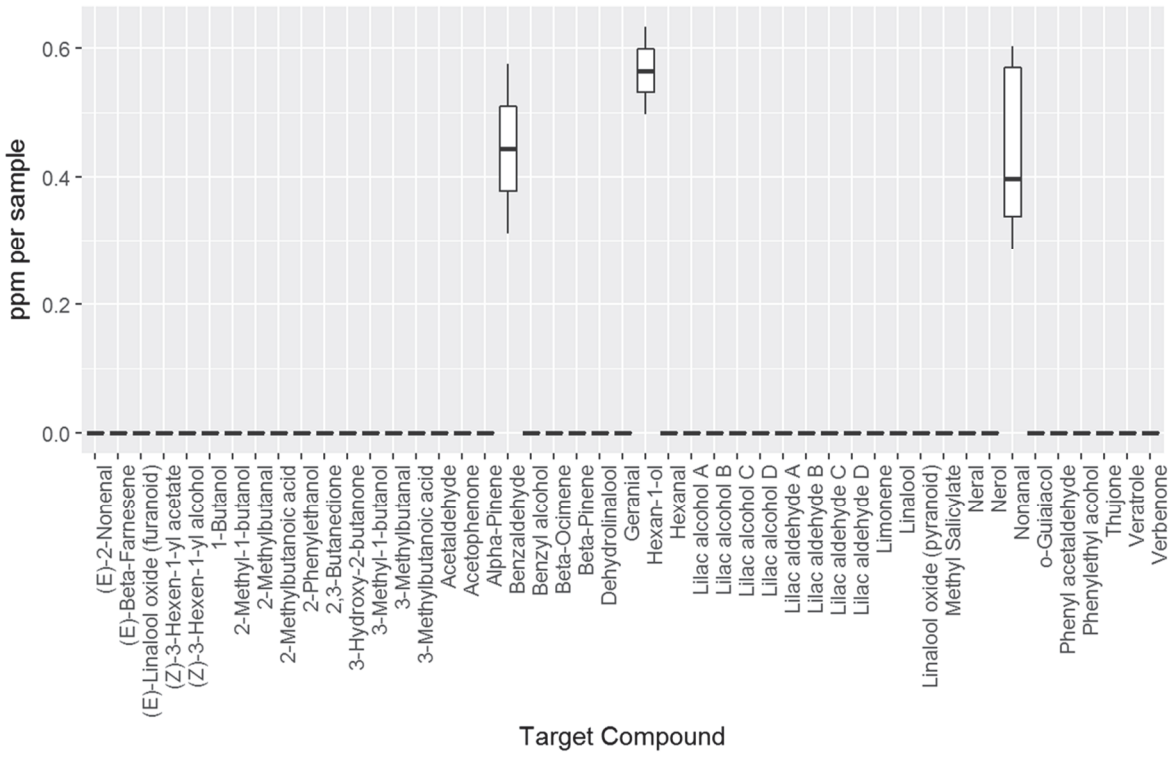


Figure 110 Ppm per VOC in samples from *C. majus*, where TurboMass Match >= 800. Y-axis is scaled to the data range.

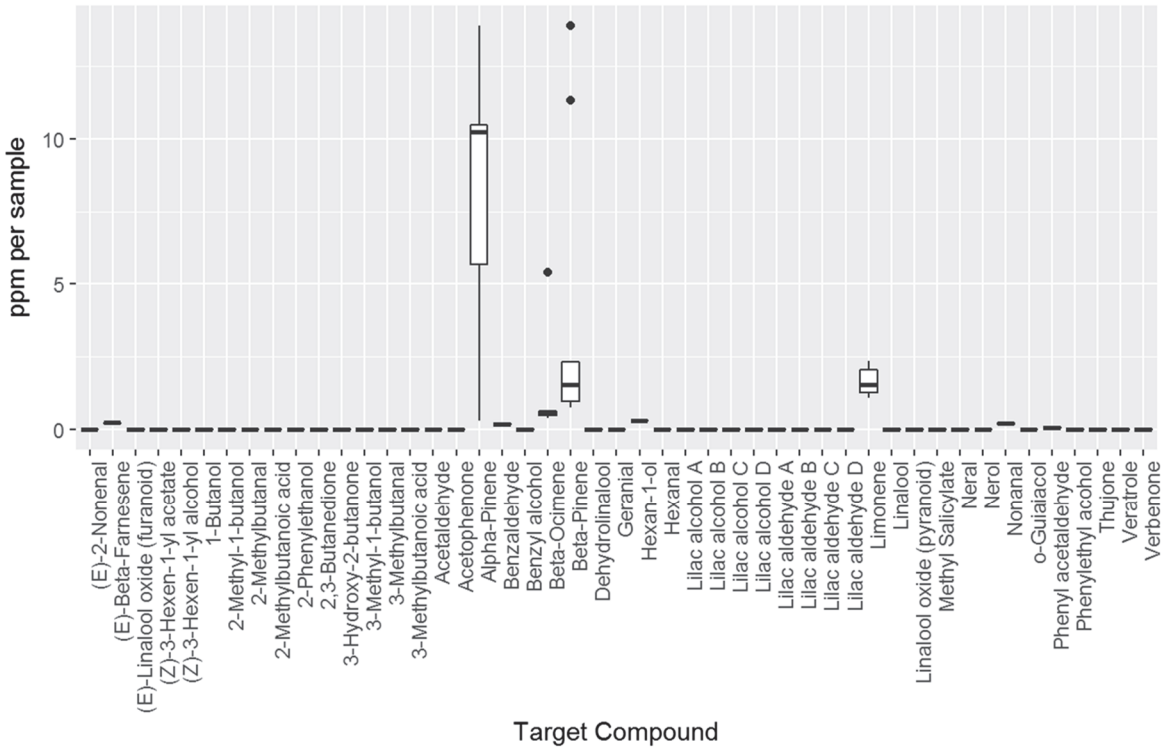


Figure 111 Ppm per VOC in samples from *D. carota*, where TurboMass Match >= 800. Y-axis is scaled to the data range. Dots are outliers (> 2 sd from the median).

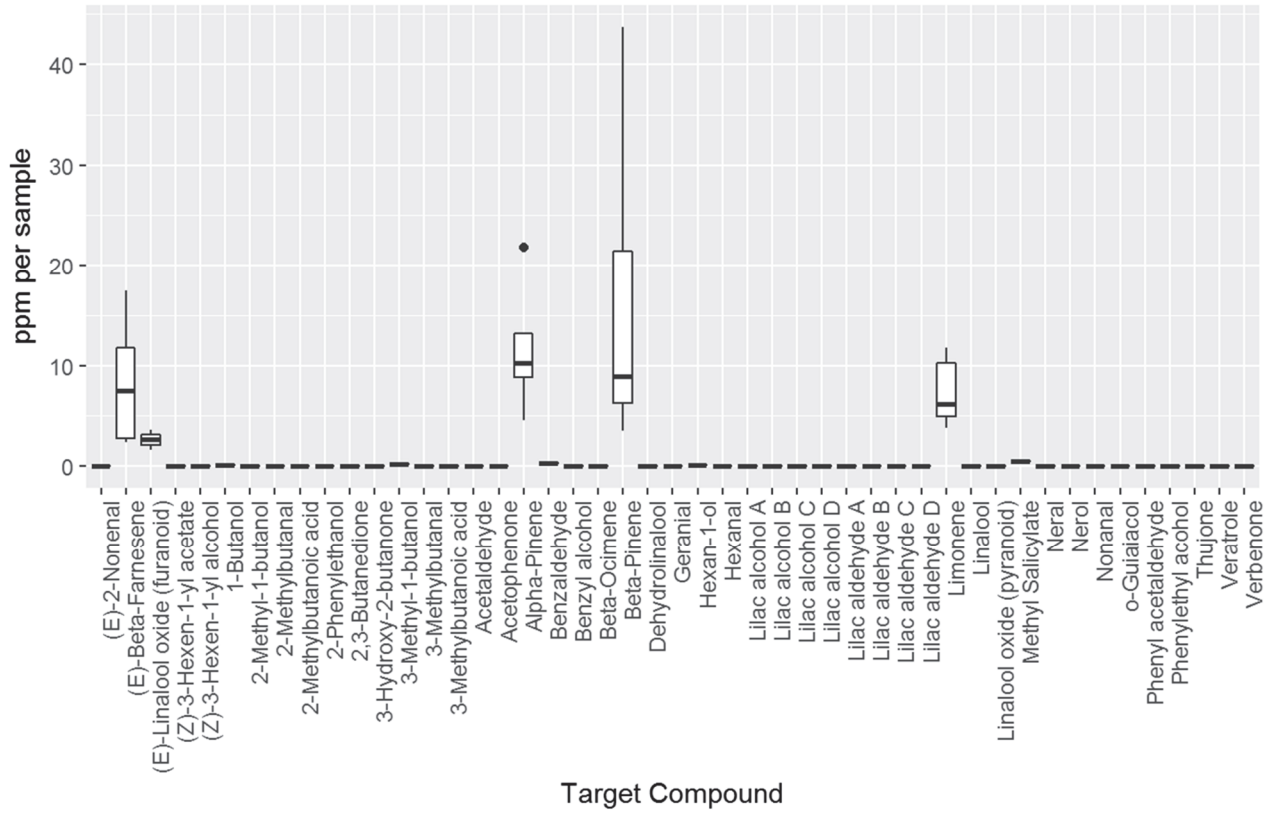


Figure 112 Ppm per VOC in samples from *H. helix*, where TurboMass Match >= 800. The only non-Apiaceae featured in these analyses has a different pattern of VOCs and included levels of (E) – B-Farnesene and Furanol not seen in any of the Apiaceae. Y-axis is scaled to the data range. Dots are outliers (> 2 sd from the median).

5.1.2.4.4.6 Odour signature by species

Having determined that there was a substantial variation in the emissions of target VOCs between the floral species under investigation, it was deemed appropriate to display these differences in an easy to compare, visual, method. Figure 113 shows the six different plant species under analysis alongside one another in an array of graphs of the median emission parts per million per sample for each target VOC from the panel being tested (Table 20). Each plot has comparable y-axes, and the order of the target VOCs on the x-axis is identical, allowing the comparison of the pattern of median to be easily carried out. This visualisation's purpose is to communicate that although these species are closely related, excluding *H. helix*, and yet they still have very different floral 'signatures'. As has already been stated olfaction is the most important sensory modality for insects (Yin *et al.* 2015), and is highly selective (Leal 2013; Montell and Zwiebel 2016), there can be no doubt that if GCMS can discern the differences between these odorant profiles, then male mosquitoes can.

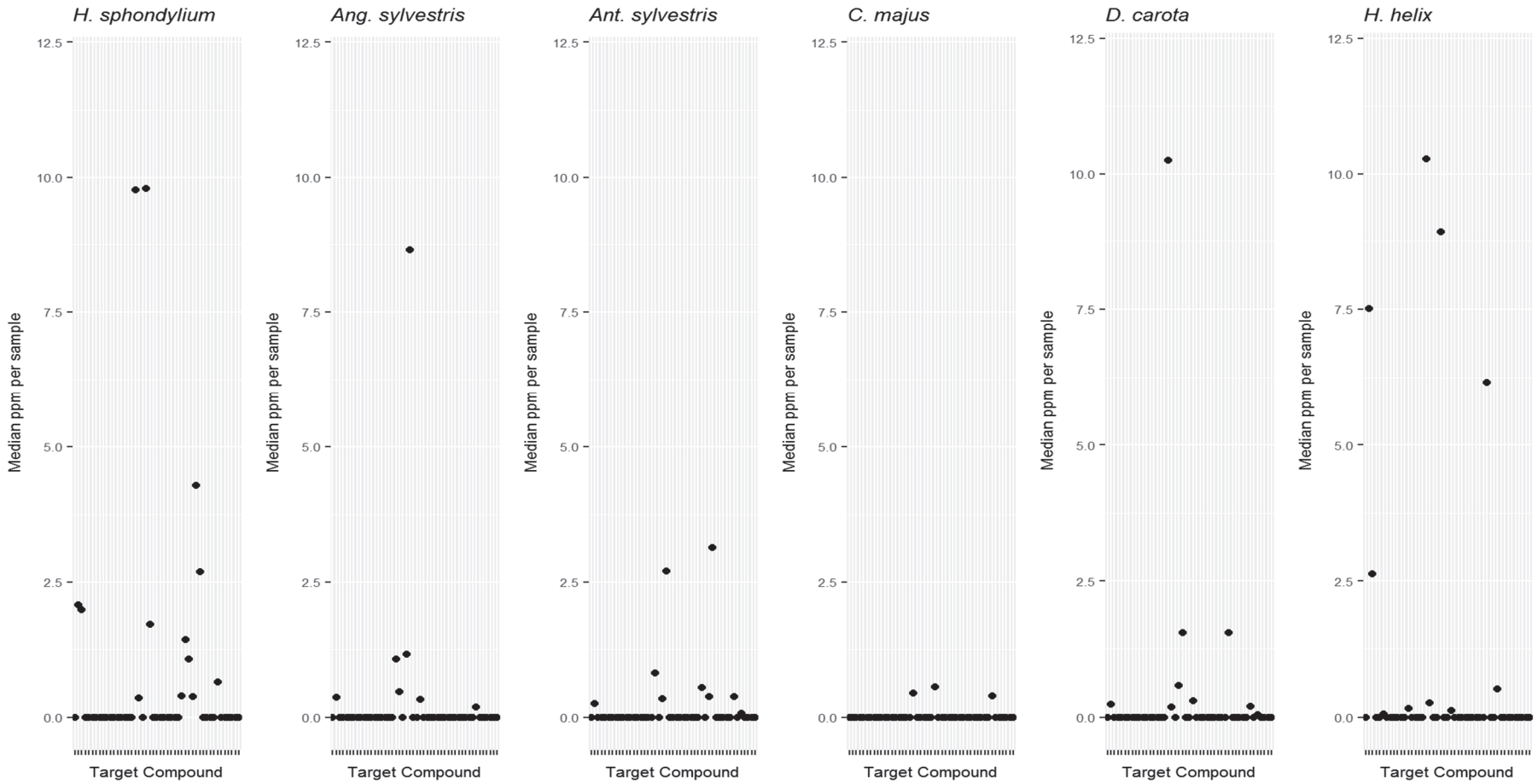


Figure 113 Floral "fingerprints" of the target VOCs. The x-axis tick marks are the different VOCs, in the same order on each graph. There are clear visual differences between the species in terms of the median ppm of VOC emissions. Supporting the assertion that this group of species have sufficient difference between them such that mosquito behaviour might be reasonably expected to be different towards them.

5.1.2.4.4.7 Discussion

These chemical analyses represent the first characterisation of the floral headspace of *Heracleum sphondylium, Daucus carota* and *Angelica sylvestris*. There have been analyses of essential oils and extracts from these Apiaceae species (Nissinen *et al.* 2005; Özek *et al.* 2005, 2008; Wedge *et al.* 2009), and of the headspace associated with their leaves (Nissinen *et al.* 2005) and roots (Yahyaa *et al.* 2015) analysis of the VOCs which are emitted without the application of extraction techniques, such as steam distillation, has been very limited to date. *Anthriscus sylvestris* was included in the floral headspace analysis of species of Apiaceae, collected in southern Sweden, conducted by (Borg-Karlson *et al.* 1993). Tollsten and Øvstedal (1994) demonstrated that within three Norwegian population of *Conopodium majus* the floral headspace profiles were sufficiently different to be able to identify the source of the flowers, suggesting geographic variation within these species and supporting the rationale of collecting geographically relevant samples when considering behavioural interactions.

This analysis was also the first semi-automatic, approach to the matching of the botanically derived chemical compounds known to be detectable by mosquitoes (Otienoburu *et al.* 2012; Nyasembe and Torto 2014) and those demonstrated to influence male mosquito behaviour (Pitts *et al.* 2014) to a data set of multiple GCMS samples; and the first matching, automated or otherwise, to headspace samples from Apiaceae collected in the UK. Mosquitoes can detect VOCs at very low concentrations; indeed it has been demonstrated that concentrations of 0.00001 µg of nonanal provoke a physiological response in *Culex* mosquitoes (Syed and Leal 2009) and in the context of the wider thesis, it cannot be assumed that the largest peaks are the most likely drivers of behaviour. Because the MS qualitative reporting was set to return 150 peak reports, it was ensured that even very small peaks within samples would be identified. A robust and useful data handling methodology was developed that enabled recording and analysis of 150 peaks per extraction. Therefore, this comprehensive reporting was employed. This reporting allied with the custom R script which handled the data entry interface with the Microsoft Access database ensured that no data entry/human errors could cause loss of fidelity in the reporting.

Amongst the compounds identified by MS were a number of semiochemicals identified as behaviour modifying in other organisms. Compounds such as camphene (CAS No. 79-92-5, RT

7.850), β-myrcene (CAS No. 123-35-3, RT 8.59), o-xylene (CAS No. 95-47-6, RT 6.725) and α-phellandrene (CAS No. 99-83-2, RT 8.89), were identified from the samples of at least three of the five plant species assayed, and are known to influence behaviour, either on their own, or more typically in the form of blend of chemicals (El-Sayed 2014). Other chemicals were present across all species. For example, ethylbenzene (CAS No. 100-41-4) eluted at 6.09 minutes, was common to all five plant species tested, and has been demonstrated to be an attractant to other dipterans, including the olive fruit fly, *Olea europaea* (Scarpati *et al.* 1993). P-xylene (CAS No. 106-42-3) eluting at 6.27 minutes and decanal (CAS No. 112-31-2) eluting at 12.10 minutes were also present across all species. P-xylene was found to be moderately attractive to fruit flies (Scarpati *et al.* 1993), but not attractive to pickleworm moths (*Diaphania nitidalis*) in isolation but was part of the floral profile of attractive plants (Peterson *et al.* 1994). Decanal has been shown to be attractive as a constituent of a blend of synthetic chemicals to lepidopterans (Cha *et al.* 2008) and dipterans (Padmaja *et al.* 2010); it was also attractive on its own to male black banded oak borer beetles, the females of which showed neither attraction nor antennal response in electroantennographic assays (Fürstenau *et al.* 2012).

Chemical analysis of the closely related Apiaceae species used in this study has shown that whilst the visual differences within the flowers were few, the ratio of chemical emissions are strikingly different. When these chemical signatures were visualised, by plotting the median ppm per VOC per sample (Figure 113), the different pattern of abundance of the VOCs between species was evident. The median was chosen as the measure of central tendency for these plots as it conveys the midpoint of observations without being skewed by extreme values. Among the panel of target compounds investigated in this thesis, much of the variance between species was associated with α-pinene, β-pinene, β-ocimene, limonene and (E)-β-farnesene (Figure 107 to Figure 112) These inter-species differences were anticipated and concurred with other studies which have analysed the differences between similarly closely related flower species. (Borg-Karlson *et al.* 1993; Suchet *et al.* 2011). In their study of DHS collections from Swedish Apiaceae, Borg-Karlson *et al.* (1993) showed similar variance between species, which originated, primarily, in the proportional differences between α-pinene, β-pinene, cis and trans isomers of β-ocimene, limonene, sabinene

(Bicyclo[3.1.0]hexane, 4-methylene-1-(1-methylethyl)- in MS reports in this thesis) and myrcene.

The use of solvent desorbed DHS as the sampling methodology proved to be an excellent choice. It proved to be sufficiently sensitive to collect samples of compounds that occurred at very low concentrations, and supported the use of 150 peak reports very well. It did have the drawback of requiring a relatively time intensive desorption and reconditioning process when compared to thermal desorption methods for instance. However, it did have the significant advantage that the process of desorption and GC injection were unpaired, meaning that it was nondestructive to the sample; therefore should a repeat of the GC analysis, or analysis using a different method or column, be required, the DHS sample remained available obviating the need to re-aquire the sampled from the source, which in the case of wild flowers may not be available until the next flowering season.

The behavioural effects of VOCs on mosquito behaviour are not limited to attraction. VOC emissions from plants may also act as repellents (Ballantyne and Willmer 2012; Champakaew *et al.* 2015). VOCs emitted by whole, undamaged, plants can offer sufficient repellency that they might be considered as part of mosquito vector control initiatives (Mng'ong'o *et al.* 2011). *Lantana camara* L. was applied to the problem of mosquitoes entering houses, and showed a reduction in the number of *An. gambiae s.s.* and *An. funestus s.s.* inside houses with these plants when used a physical and chemical barrier (Mng'ong'o *et al.* 2011). Plant derived compounds have been demonstrated to have repellent effects (Nerio *et al.* 2010) and often unrelated plant species have been demonstrated to share similar repellent constituents such as eucalyptol, *d*-pinene, geraniol, citronellal, camphor and linalool (Curtis *et al.* 1990). Indeed, the essential oils of members of the Lamiaceae, Myrtaceae, Poaceae and Pinaceae are commonly used as insect repellents (Maia and Moore 2011). Due to the composition of the panel of target VOCs used in this study, it is likely that some of the compounds on the list may act as repellents to mosquitoes. Therefore, when considering the chemical analysis results it is unlikely to follow that more matched target panel compounds, or the higher their concentration, will equate to higher attraction; rather it will be the nature of the blend of VOCs that is likely to drive the attraction or repulsion relative to the flowers (Smallegange *et al.* 2005, 2009).

The apparent height difference of the IS peak in the chromatograms is thought to be due to the difference in the shape of the IS peak in some of the samples, and a factor of possible variation in the sensitivity of the GC during a given analysis. The normalisation process that was utilised is based on a small time region, and therefore multiple data points that describe the peak (Vu *et al.* 2011), rather than just the highest point of it. Therefore, if the IS on a chromatogram has a broader base, the normalised plot should be expected to return a shorter peak. Excised florets were used throughout the chemical analysis study. Excising the florets rather than sampling the headspace of florets still attached the plant, may cause physiological changes to the florets (Wade and Wratten 2007) and therefore the VOC emissions from them. The emission of herbivory-related VOCs is a known phenomenon (Kessler and Baldwin 2000, 2002; Engelberth *et al.* 2004) and VOCs which are targets for this study have been observed among those indicated as herbivory-related in other flower species (Maia and Moore 2011). It was hypothesised, in other flower species, that there was a difference in the amount of nectar produced by excised flowers (Lee and Heimpel 2008) which might lead, therefore, to a reduction of VOCs related to nectar presence. Despite these potential drawbacks, using excised florets facilitated the use of a smaller DHS chamber in the laboratory, which allowed the control of most environmental variables. Critically, the use of excised florets during the chemical analysis matched the presentation of variables within the subsequent behavioural assays (5.3) Therefore the VOC emissions from the samples should be similar to those offered to the mosquitoes in those assays, enabling inferences between chemical analysis and behavioural findings. As part of further work related to the analysis of Apiaceae volatiles, the use of laboratory based DHS collections of whole plants, and non-excised inflorescences, potentially using variations of the nylon bag encapsulation method trialled in the field but using sorbent tubes rather than SPME. It would be particularly interesting to determine the emission source of VOCs and would facilitate the identification of those compounds specifically related to herbivory/damage.

In conclusion, the differences in the target VOCs emitted by these Apiaceae were found to be supportive of the further hypothesis that there will be sufficient inter-species variation to elicit different behavioural responses in sugar meal seeking mosquitoes, and therefore, these Apiaceae species were accepted as the floral variables for the mosquito behavioural assays.

5.2 Designing and building a dual choice olfactometer

The design of the olfactometer used for these studies was strongly influenced by those already used successfully in mosquito behavioural research. There has been a progressive development of the olfactometer over time with different researchers making changes to the design of the apparatus to suit their own needs. Early examples include those designed by Gouck and Schreck (1965) to investigate mosquito attractants, that they later further improved (Schreck *et al.* 1967), and Klowden's and Lea's (1978) olfactometer used to assess host seeking behaviour. These designs influenced the dual-port airflow olfactometer used by Hancock and Foster (1993). Further changes to the air delivery systems, led to the bench top filtering and moistening employed by Pates *et al.* (2001), and the addition of warming plates to the air preparation stage as seen in Verhulst *et al.* (2009). It is from this development history that the olfactometer design process in this chapter begins.

5.2.1 Design process

To minimise costs and ensure the final design was fit for purpose the olfactometer was designed and tested digitally using computer-aided design software (CAD). All component parts were accurately sized and assembled *in-silico* before being visualised in three dimensions. This allowed the virtual inspection of the flight arena, trapping and variable chambers; facilitating improvements to these components without requiring the materials to be purchased and parts constructed to test their suitability.

The dual choice olfactometer was designed initially using Autodesk Inventor Professional 2014, before being migrated to Autodesk Inventor Professional 2015 upon that product's release. Digital testing, using computational fluid dynamics (CFD), of the various designs of the dual choice olfactometer was carried out using Autodesk Simulation CFD 360. Software licensing was granted for all Autodesk products by their Student Licensing Programme.

The first digital prototype used a large cross-sectional area, with the intention being that it would facilitate the natural flight of mosquitoes and reduce interaction with the arena walls. However, once the first complete digital prototype was prepared, and the airflow parameters were calculated it became apparent that the flight chamber was too large resulting in substantial regions of dead space (Figure 114). To effectively distribute olfactory stimuli

through the arena at a rate suited to mosquito flight would place too high an air flow demand on the compressor. The increased air flow speeds in the release chamber, due to its much smaller cross-sectional area, would potentially become such that the upwind flight response would be triggered and the system become part wind tunnel and part flight arena, which was not the intention of the device. The main chamber was therefore resized to a smaller cross-sectional area, remodelled and retested. This improved matters, but still showed several areas of dead air with minimal air flow in the main arena, this could result in mosquitoes finding themselves in areas of the flight arena where they are not receiving the applied olfactory stimuli at all.

A further reduction of the cross-sectional area of the flight arena showed, after *in-silico* testing, a decrease in this dead space problem (Figure 115), whilst still offering sufficient flight space for the mosquitoes to move unimpeded in the process of reaching the decision ports at the far end of the chamber. Analysis of the airflow characteristics of the final design suggested that the chamber was sufficiently long for air flows to mix and for plumes to stabilise within the flight arena, facilitating mosquito's location of the source of the odour (Figure 116). Final airflow through the main flight arena was calculated at 0.01 m s^{-1} when the air flow delivery was regulated at 10 l min^{-1}.

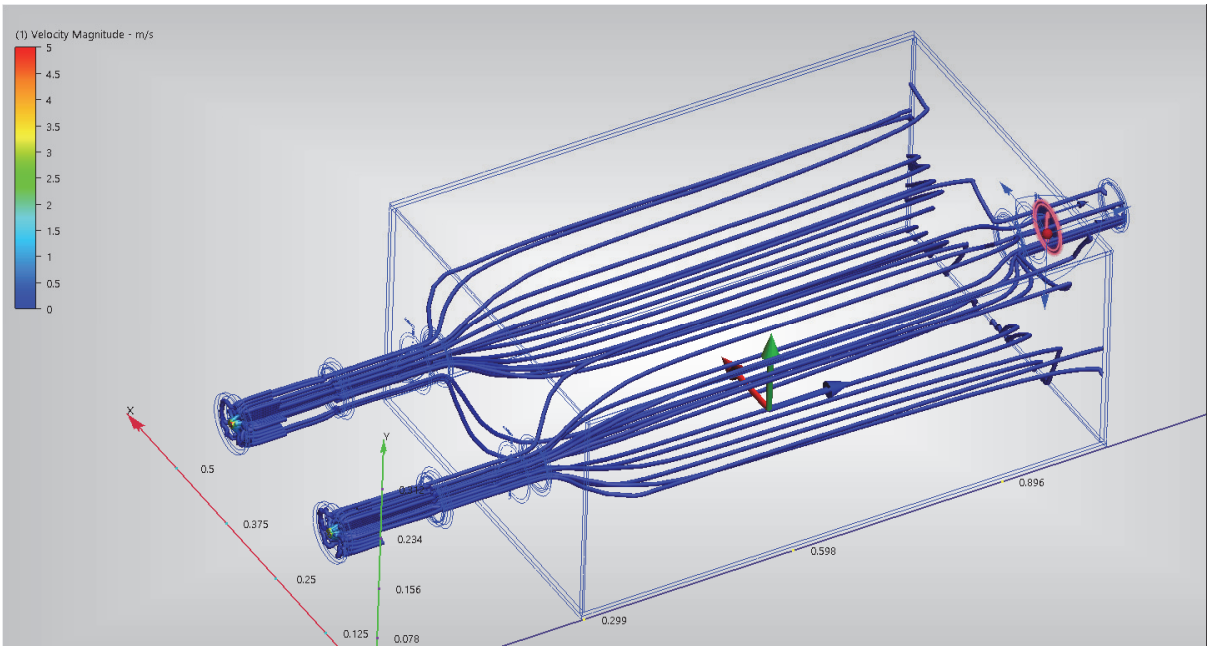


Figure 114 Particle trace of moist air flowing through the first design prototype of the dual choice olfactometer. This flow was suggestive of significant areas of potential 'dead space' within the arena, and was consequently rejected.

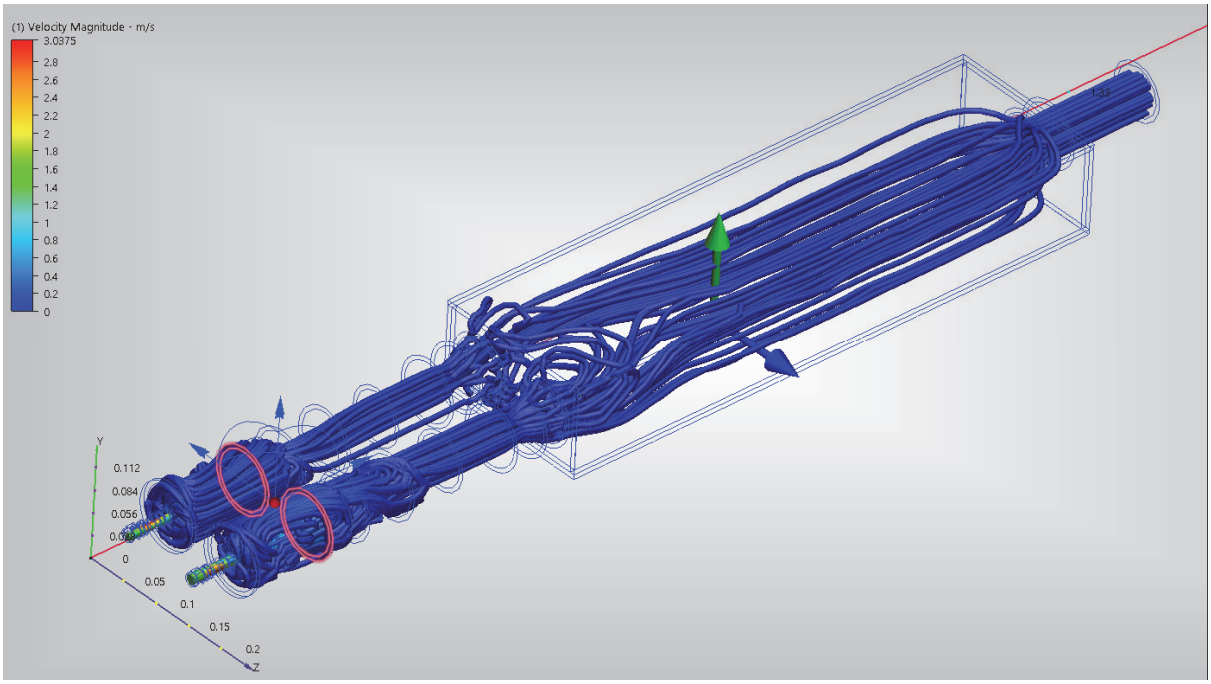


Figure 115 Particle trace of the moist air flow in the final design prototype of the dual choice olfactometer, shows a much reduced dead space for the main air flow.

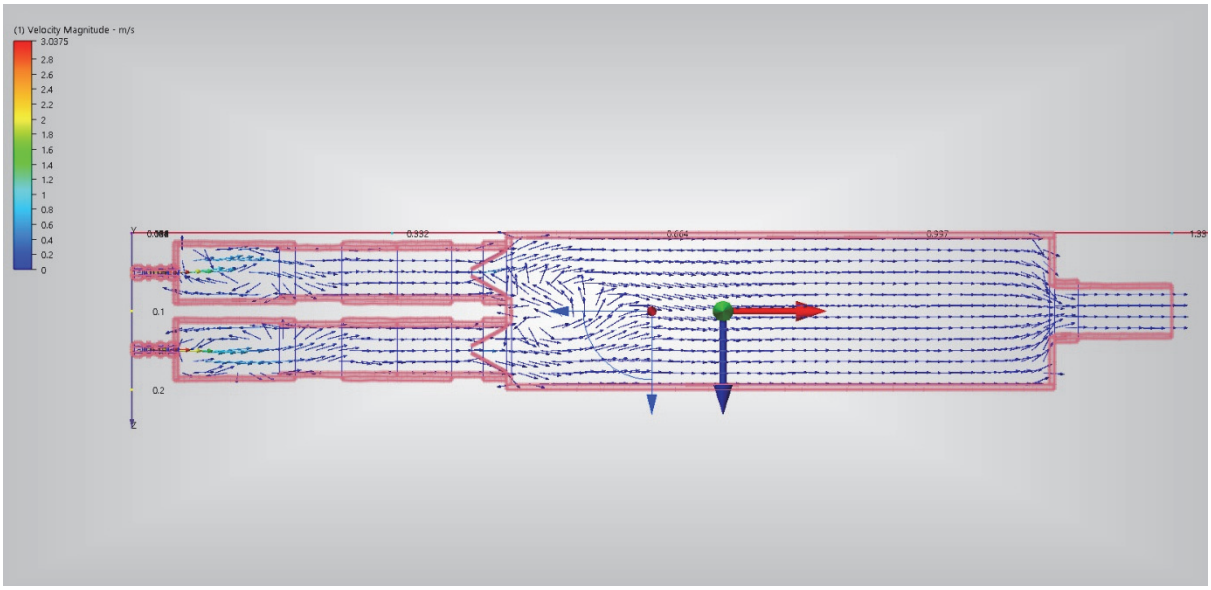


Figure 116 Airflow visualisation within the olfactometer, showing areas of mixing and subsequent laminar flow. Based on 10 l min^{-1} air supply the airflow rate was 0.01 metres per second.

5.2.1.1 Materials and dimensions

The labelled diagram of external dimension and parts (Figure 118), details the major characteristics of this dual choice olfactometer design. The main arena is built from 6 mm thick, transparent, colourless, acrylic sheet, chemically welded along each seam using EMA Plastic Weld, applied using capillary action guaranteeing a complete seal. The top sheet of the main arena was attached using stainless steel screws. The release chamber consisted of a removable, transparent, colourless, acrylic cylinder (150 mm in length, 70 mm outside diameter (OD), 5 mm wall thickness (WT)), closed at one end by a fine stainless-steel mesh, and with a rotating mesh gate at the other end allowing the release of the mosquitoes at the start of each assay. The trapping chambers were located at the opposite end of the flight arena from the release chamber, were of similar construction to the release chamber except for the rotating gate being replaced by a cone with 10 mm aperture (Figure 117), preventing mosquitoes from leaving the variable specific trap once they had entered it. Immediately upwind of the trapping chambers were the variable chambers. These were constructed using 80 mm OD, 5 mm WT, transparent, colourless, acrylic cylinders, and were closed at one end by an acrylic disc with a bulkhead fitting for an 8 mm OD nylon airline. The other end was left open such that following variable loading the chamber could be fitted over the end of the

trapping chamber. The release and trapping chambers were attached to the main arena by means of 80 mm OD, 5 mm WT, collars which were permanently fixed to the arena, facilitating a tight push fit coupling between parts.

The olfactometer was designed to be cleaned using detergent and water between each assay, rinsed with de-ionised water and dried with paper towels. To facilitate this cleaning, where possible all parts were designed to be removable without requiring tools. Only the lid of the main arena required a screwdriver to remove the 4 retaining screws. The permanently attached end caps and stainless steel meshes required the use of a bottle brush to clean thoroughly.

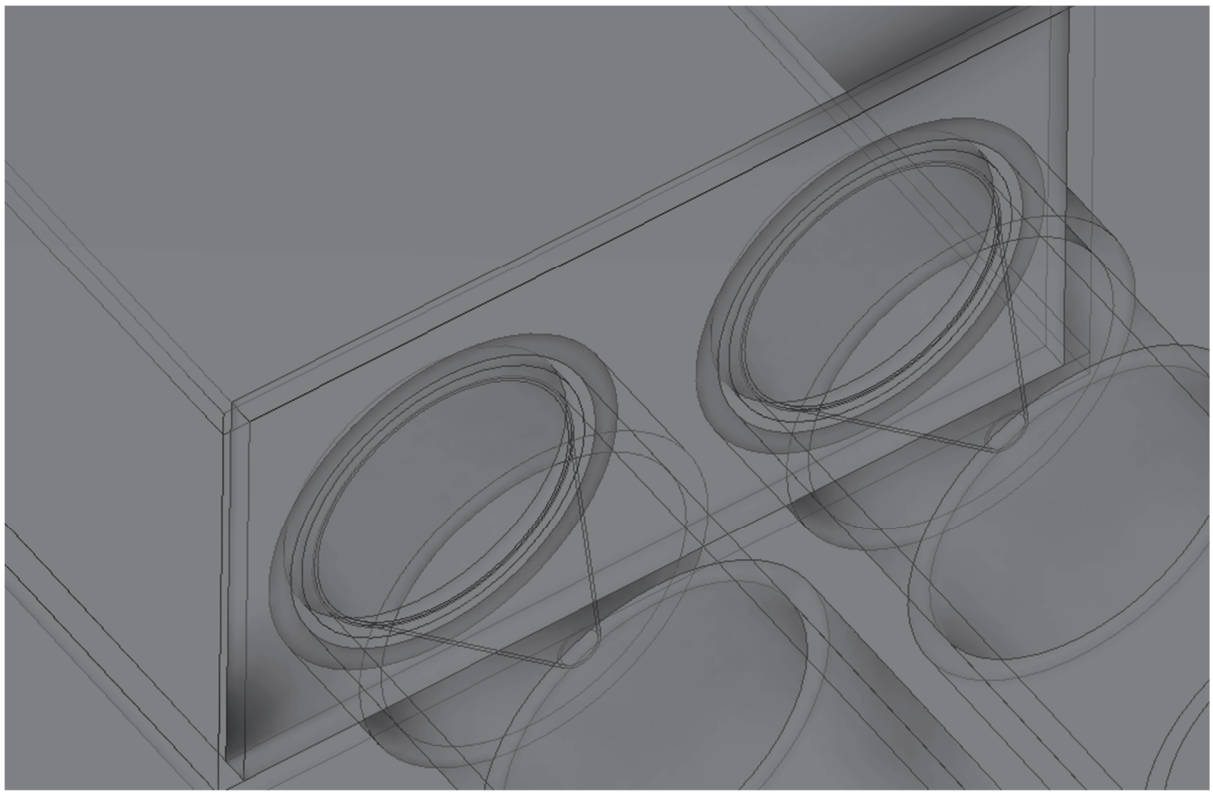


Figure 117 Detail view of the anti-return funnels in place. These were fitted to allow overnight assays to take place, whilst minimising the possibility of mosquitoes changing their 'decision' by moving between chambers.

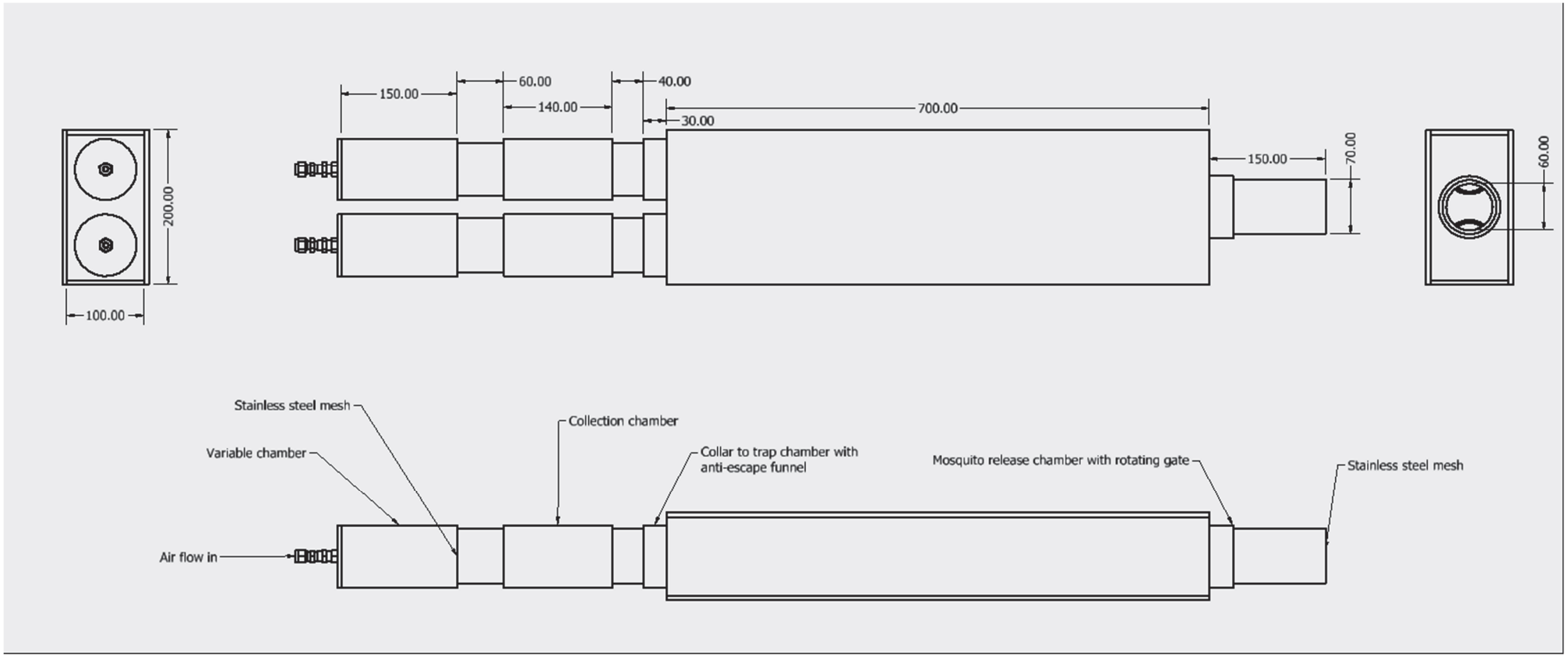


Figure 118 Two-dimensional orthographic projection diagram of the part dimensions (mm) and labels.

5.2.1.2 Air supply

Supply of clean air was accomplished using an oil-free compressor (Bambi HT24). This air was filtered using an activated carbon filter (Parker ACS010CGMX) prior to passing through a flowmeter (Cole-Parmer WZ-68560-20) to regulate the airflow. Air flow was set at 10 l min^{-1}. The air was then passed through a 250 ml Drechsel bottle filled with 150 ml of distilled water, which was warmed by standing on a laboratory heat plate maintained at 50 °C to control the temperature and humidity.

Air supplied as above resulted in RH 80 % (+/- 5%) and 25 °C (+/- 3 °C). These fluctuations were mostly thought to originate from fluctuation within the room within which the assays were conducted, and by the progressive reduction in water level in the Drechsel bottle.

5.2.1.3 Controlling for visual stimuli

Controlling the mosquitoes' access to visual stimuli was considered to be imperative if assay results are to be considered based on olfactory-driven behaviours. These stimuli might originate within and without the olfactometer. Within the olfactometer, very fine stainless steel mesh was used to separate the variables from the trapping chambers, which were already visually obscured by the non-return cones fitted to the entry of the trap chamber. This aperture size of this mesh was sufficiently small to prevent visual discernment of variables, even under close examination.

Visual stimuli within the room being used for the assays were obscured using blackout material which was draped over the whole olfactometer array for the duration of assays.

5.2.1.4 Testing for positional bias

Once built, the olfactometers were set up and run overnight, with no variables loaded so that the results could be inspected for positional bias. No significant positional bias was observed in either olfactometer; olfactometer No. 1, $\chi^2 = 2.134$, df = 4, p= 0.71, olfactometer No. 2, $\chi^2 = 3.628$, df = 4, p= 0.459. Each olfactometer was tested 5 times, and the mean number of mosquitoes per assay = 17 (+/- 7).

The mosquitoes used for these assays consisted of daily batches of unfed male *Culex pipiens s.s.*, reared on from eggs collected locally (Eastings 384742, Northings 348565), were allowed to mature for 24 - 36 hours post eclosion and given *ad-libitum* access to water.

5.2.2 Discussion

The dual choice olfactometer design discussed and built here proved to be fit for the purpose of studying the olfactory behaviour of mosquitoes. The transparent acrylic construction made the recording of results very easy. It also ensured that safe recollection of specimen mosquitoes was a simple matter, using an electric aspirator with a long rubber hose. The clear design also made inspection of the condition of the unit straightforward, promoting confidence in its continued use.

The design was sufficiently robust to be used for an extended period of time and the constructed units tolerated daily assays, including, daily deconstruction and cleaning without detriment to performance. The acrylic construction did restrict the use of some common laboratory cleaning products, due to the potential of alcohols attacking the material, although others have reported the use of alcohol to clean acrylic and plexiglass (Geier and Boeckh 1999; Otienoburu *et al.* 2012) spot testing with 70 % ethanol on the material used here showed damage and used was stopped. Cleaning was therefore conducted with mild dish soap, before thorough rinsing using copious distilled water and drying with disposable paper towels.

One potential limitation of the design of the variable chambers was that their internal diameter limited the physical size of stimuli which could be loaded into them. For this study, it caused no problems as the variable chambers were designed with the expected variables in mind. Should studies require physically larger variables, whole plants for example, then simply leaving the built-in chambers empty and adding external, inline, variable chambers would allow any size of odour source to be used.

Designing the olfactometer using Autodesk CAD software, whilst it felt somewhat time-consuming, was an excellent approach as it was highly supportive of a considered approach to the design process. The additional time spent digitally designing the olfactometer was probably rather minimal compared to the time cost of sourcing materials for and building and subsequently testing physical prototypes. Indeed, once the olfactometer components were

digitally constructed, it was very quick to generate new digital prototypes. One significant advantage of having carried out the *in-silico* testing of the flight arena designs was the financial savings made possible by digital prototyping. Because of this cost-saving, a second dual choice olfactometer was able to be constructed within the original budget, doubling the potential data collection rate for studies using them. All the parts of the olfactometer were selected to fit the aim of the project and resulted in a pair of highly capable and robust olfactometers being designed and built.

5.3 Behavioural Assays

5.3.1 Aims

The aim of this chapter was to explore the olfactory sugar feeding behaviour of male *Culex pipiens s.s.* and *Culex torrentium* mosquitoes in relation to semiochemical emissions from common flowers of the Apiaceae family.

Through experimentation, the preference towards these potential sugar sources of each mosquito species will be ascertained. This will allow the identification of the most preferred flower species for each mosquito species and the comparison of the behaviour of these two sibling mosquitoes. Data produced by this study need to be suitable for use in further analyses to allow significance testing and to inform further work relating to the investigation of these behavioural data against the data from the chemical analysis of volatile organic chemicals emitted by Apiaceae.

5.3.2 Method

5.3.2.1 Mosquitoes - rearing

Candidate mosquitoes were all wild-type *Culex pipiens s.s.* or *Culex torrentium* males reared from field collected eggs, to ensure that mosquitoes used in these experiments express behaviour representative of the wild population. Culex *spp.* egg rafts were collected from a 10 L bucket, placed in a suburban back garden (Eastings 384742 Northings 348565) and half filled with hay-infused water (Hazard *et al.* 1967). Additional egg collections were made from outdoor pools at a local aquatic plant retailer (Eastings 383469 Northings 354591).

Rearing of larvae through to adult emergence was conducted as previously described (3.2.1.7). Upon emergence of the first male specimens, species identification was carried out on no fewer than 3 individuals, and the now identified batch of siblings transferred to one of several species-specific mosquito cages. Newly emerged males from these species-specific cages were collected each day and stored in holding cages without food, but with ad-libitum access to water, for use in behavioural assays on the next day. Males utilised in assays were between 24 and 48 hrs old and unfed at the start of each assay. Males were only used in an assay once as learning has been seen to occur in mosquitoes (McCall and Eaton 2001; Kaur *et*

al. 2003; Vinauger *et al.* 2014), and so, following assays, used mosquitoes were moved to culturing cages with access to food and females as part of a separate culture project.

5.3.2.2 Plants – Selection, collection and preparation

Mature, whole, floral umbels were collected from the field not more than 24 hrs prior to assays and taken into the laboratory. The umbels were then placed in a beaker of tap water and stored in the laboratory, with access to sunlight, but under no other special conditions.

All flower collections were made from the Staffordshire University estate and nature reserve (Eastings 38839 Northings 34539). Nitrile gloves were worn throughout floral collection and preparation, to avoid contamination.

5.3.2.3 Dual choice assays

All assays were conducted using the dual-choice olfactometers which were designed and built for this study; these were specifically designed to restrict the stimuli received by the mosquitoes to those of an olfactory nature.

Prior to assays, all olfactometer parts were cleaned using deionised water and mild detergent, then rinsed thoroughly using distilled water before drying using disposable paper towels. Nitrile gloves were worn at all times during the cleaning and setup of the olfactometer and during collection, handling and preparation of floral variables, to avoid contamination by human odours.

For each assay, two 7 ml roll top glass vials were partially filled with deionised water and topped with aluminium foil. For the negative control vial, a hole was pierced through the foil lid (~ 5 mm) to allow evaporation to occur as in the variable preparation. For the variable, the stem of an excised section of the flowering umbel was passed through the aluminium foil lid, ensuring that the cut section remained submerged. Excised sections used were of similar size for each assay, regardless of species.

The variables and controls were then loaded into the stimuli chambers such that there was no contact between the variable and the stainless-steel mesh which separated it from the trapping chambers (Figure 119). Although no positional bias was detected in testing, the

position of the variables, their placement in the front or back chamber, was alternated between each assay.

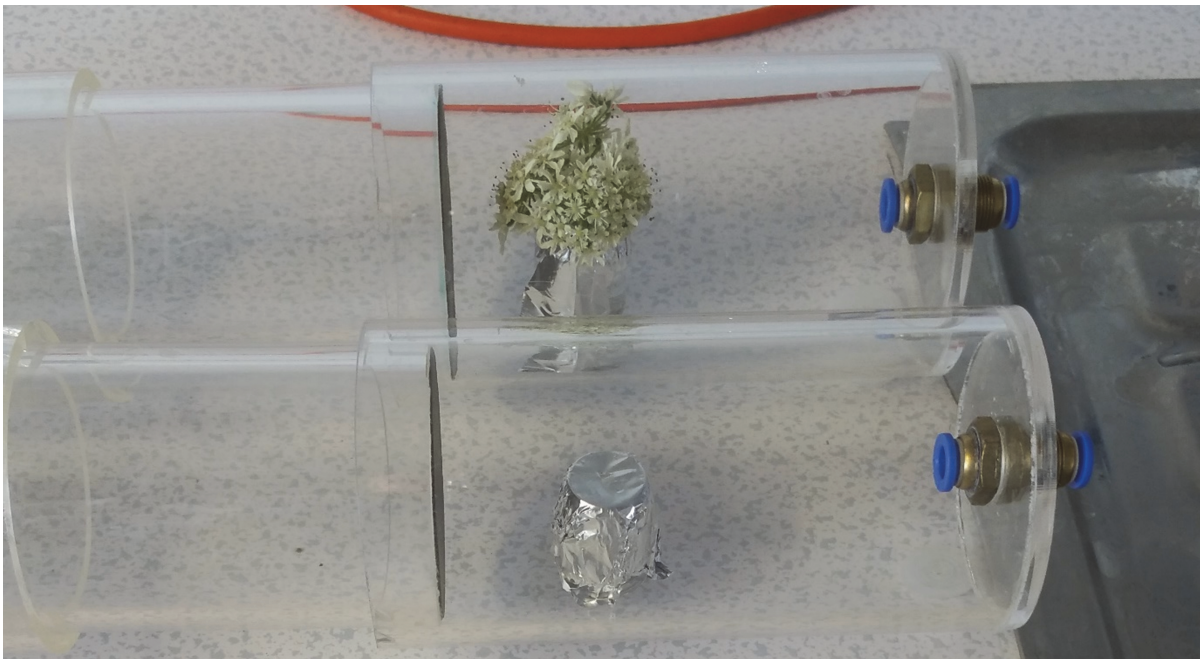


Figure 119 Example of variable and control configuration within the dual choice olfactometer. The top chamber is typical of the variable presentation and the bottom chamber shows a negative control.

Airflow was delivered by an oil free air compressor and passed through an activated carbon filter. The air was humidified by bubbling through warmed distilled water contained in a Drechsel bottle (250 ml) on a heater pad. Airflow was restricted to 10 l min^{-1} by a Cole Palmer inline flow valve, before being split into the two inlets of the dual choice olfactometer. This airflow resulted in a 0.01 m s^{-1} air speed in the main flight arena. Thirty minutes prior to starting an assay the heater pad was turned on, at 50 °C, under the Drechsel bottle, and airflow commenced. This allowed the conditions inside the olfactometer to stabilise. Internal conditions were 25 °C +/- 3 °C and 80 % +/- 5% relative humidity (RH).

All assays used only naïve mosquitoes, these being those that had not been utilised for behavioural studies previously and had no prior exposure to the olfactometer. This was important as mosquitoes have been demonstrated to have the capacity for learning which

could bias behaviour (McCall and Eaton 2001; Jhumur *et al.* 2006). In previous studies that have used olfactometers for behavioural analysis there has been a broad range in the number of mosquitoes used per assay. Some studies have used relatively low numbers with as few as five per assay (Jhumur *et al.* 2006), 10 per assay (Nyasembe *et al.* 2012) through to 25 – 30 mosquitoes (Logan *et al.* 2008), whilst others have used higher numbers of mosquitoes up to 200 – 300 mosquitoes per assay (Otienoburu *et al.* 2012). Here, due to the relatively low volume olfactometer design approximately 25 mosquitoes would be used for each assay.

Unfed, naïve, male mosquitoes (n = ~25 per assay) were collected from the relevant daily holding cages and introduced to the release chamber of the olfactometer. These mosquitoes were left to acclimatise for 15 minutes. Throughout the acclimation process, and the assaying process itself, the dual-choice olfactometers were covered with a blackout sheet to avoid interaction of visual cues in the room. Each assay was started at 17:00 and run overnight to 09:30 the next morning. At the end of the assay the location of all mosquitoes was recorded, with 4 possible locations, release chamber, main arena and two choice chambers (Figure 120). Those in the release chamber were considered as non-activated, those in the main arena as activated but having not made a choice, those in the variable chamber as having made a choice for the stimuli and those in the control chamber as having made a choice for the control.

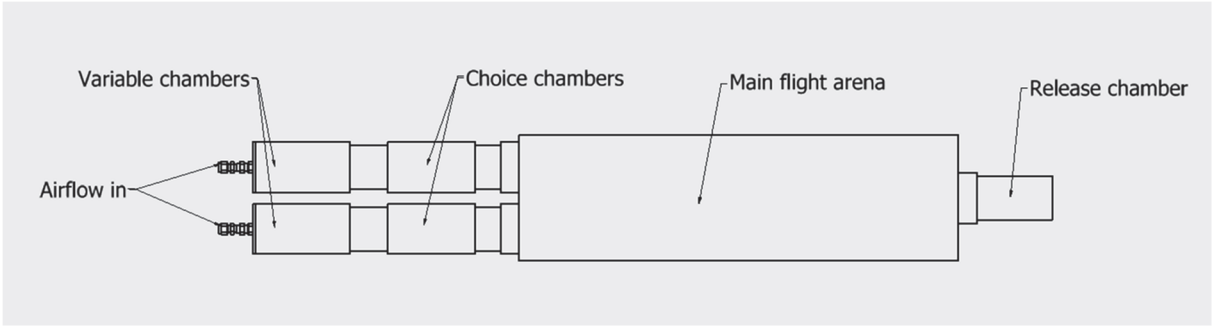


Figure 120 Plan view of the main zones of the olfactometer. At the end of each assay the number of mosquitoes in each choice chamber, the main arena and the release chamber will be counted separately.

Following each iteration, the olfactometer was disassembled, washed, rinsed and dried before reassembly for subsequent assays. On occasions where there were several days where assays would not be performed, such as during periods of insufficient numbers of mosquitoes

meeting the experimental requirements, the olfactometer would be cleaned again on the day of the next assay.

5.3.2.4 Quantification of preference

The preference index (PI) for each assay was calculated. Several versions of preference indices have been used to describe choices by groups of organisms in olfactometers. The PI is a measure of the relative attractiveness of a variable, and an often used version of the formula is (McBride *et al.* 2014; Wang *et al.* 2016):

PI = (Xv - Xc)/ (Xv + Xc)

Where:

PI = Preference index

Xv = number of mosquitoes choosing variable

Xc = number of mosquitoes choosing control

The PI calculation produces a result between +1 and -1 with 0 being no preference, i.e. 50 % of the mosquitoes chose each option, and +1 or -1 being an absolute preference for one variable or the other.

There is a weakness inherent in this measure of preference, however, as it does not consider the proportion of mosquitoes which did not respond to the stimuli. Two alternative preference index variants were devised and discussed, the Activated Preference Index (API), and the Expressed Preference Index (EPI).

In the API, a further modification was applied to the PI to account for the power of stimuli to cause the mosquitos to activate and move from their release chamber. The PI was multiplied by the activation percentage (the percentage of mosquitoes not found in the release chamber at the end of the assay). This weighted the results such that those eliciting the greater

activation percentage had, the largest influence on the overall measure of preference. Importantly, it prevented cases with very low activation rates having too high an impact on the preference index.

Therefore, the formula applied would be:

API = ((Xv - Xc)/ (Xv + Xc)) * A

Where:

API = Activation and Preference index

Xv = number of mosquitoes choosing variable

Xc = number of mosquitoes choosing control

Xr = number of mosquitoes in release chamber

Xm = number of mosquitoes in main arena

A = Activation percentage = (Xv + Xc + Xm) / ((Xv + Xc + Xm + Xr)) * 100

The API calculation produces a result between +100 and -100 with 0 being no preference, i.e. 50 % of the mosquitoes chose each option, and +100 or -100 being an absolute preference for one variable or the other.

The EPI is a logical continuation of the calculation which should be applied to the PI in high duration assays, such as these which were carried out overnight. It is reasonable to expect that the mosquitoes might move to find a preferable resting location, even if they are not directly activated by the stimuli presented to them. Therefore, the equation should be adjusted such that it is weighted by the proportion of those which expressed a preference,

i.e. those which are found in one of the two choice chambers, rather than simply having moved from the release location, giving:

EPI = ((Xv - Xc)/ (Xv + Xc)) * E

Where:

EPI = Expressed Preference index

Xv = number of mosquitoes choosing variable

Xc = number of mosquitoes choosing control

Xr = number of mosquitoes in release chamber

Xm = number of mosquitoes in main arena

E = Expression Percentage = (Xv + Xc) / ((Xv + Xc + Xm + Xr)) * 100

This is more conservative than PI and API in the values it returns, particularly when these results are being considered in the context of the 'best' results being those which indicate a variable which is highly attractive to mosquitoes and might be suitable as a lure.

Like the API, the EPI calculation produces a result between +100 and -100 with 0 being no preference, i.e. 50 % of the mosquitoes chose each option, and +100 or -100 being an absolute preference for one variable or the other, but the values returned are now reduced where mosquitoes were not seen to express a preference.

Within the results, the activation percentages and expression percentages are reported to show the influence arising from these measures as applied to the PI. Further analyses were carried out using the EPI results as this measure returns the most conservative results when considering the effective attractiveness of stimuli.

5.3.2.5 Statistical testing

Pearson's Chi-squared, with simulated p-value testing was carried out to test for significance in the raw count data for each valid combination from the possible contingencies. The simulated p-value option was used as in some assays there was a low number of mosquitoes found in the choice chambers, e.g. 8 in the variable and 1 in the control giving an expected value of 4.5, this expected value is less than 5, which is considered the cut-off for the reliability of Chi-squared.

Testing for difference of preference between floral stimuli offered, and, separately, between the preferences observed in each mosquito species, was carried out using ANOVA. Prior to carrying out the ANOVA test, each combination was tested for normality, using the Shapiro-Wilk normality test, and for homoscedasticity using the Bartlett test for homogeneity of variances.

Testing for a significant difference in the expression percentage between contingencies used the Shapiro-Wilk test and Bartlett test as above but resulted in non-normal and heteroscedastic data requiring the utilisation of the Kruskal-Wallis rank sum test.

5.3.3 Results

Due to insufficient temporal overlap between the availability of *Culex torrentium* and *C. majus*, assays were not possible; similarly, only two iterations of *Culex pipiens s.s.* against *Ant. sylvestris* were conducted. Therefore, only their activation/expression values are returned in Figure 121 and Figure 122; they were not included in any subsequent analyses.

5.3.3.1 Mosquito activation rate in assays

Mosquito activation rates were high throughout the assays with mean activation > 87.5 % for all flower species (Figure 121). Activation rate was also high for the double negative control assays used for tests of positional bias (Figure 121). There were no significant differences in the activation rates seen for the various floral stimuli, nor was there any significant difference between the mosquito species' activation rates to similar stimuli.

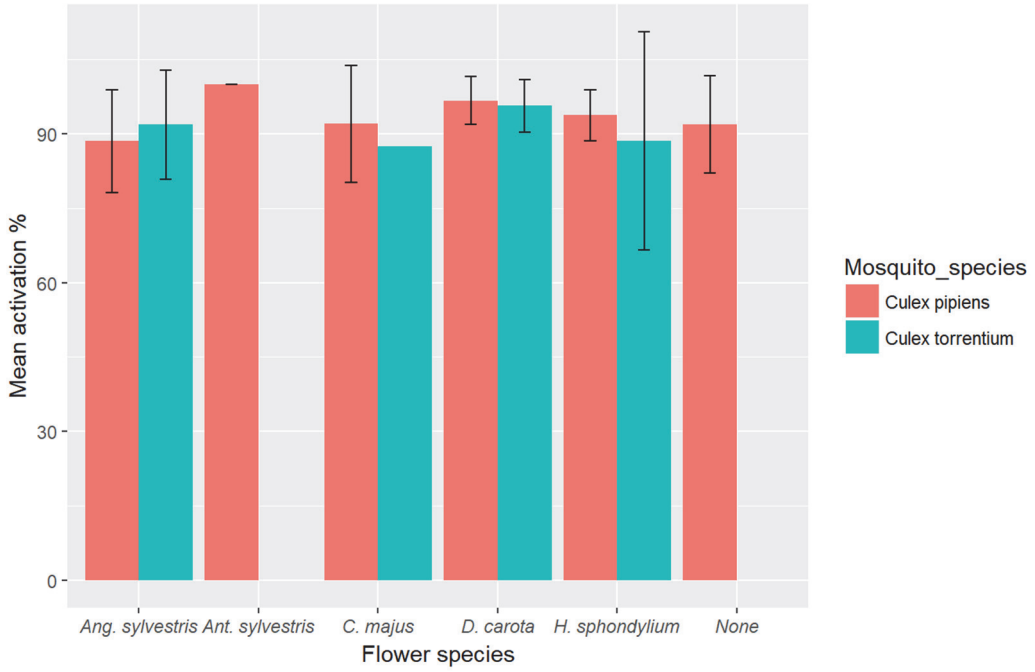


Figure 121 Mean activation percentage of each mosquito species when assayed against the named flower species. *Ant. sylvestris* was unavailable for assay against *Culex torrentium*. There was no significant difference in mean activation between assays using floral volatiles and those using negative controls only.

Using the more conservative measure of mosquito response, the expression percentage, values were lower and, superficially, more varied between species (Figure 122). This variation was shown not to be statistically significant using a Kruskall-Wallis rank sum test. *Culex pipiens s.s.* KW $\chi^2 = 8.7075$, df = 4. p = 0.06884, *Culex torrentium* KW $\chi^2 = 0.54113$, df = 2. p = 0.7629, *Culex pipiens s.s.* and *Culex torrentium* combined data KW $\chi^2 = 6.3779$, df = 4. p = 0.1726.

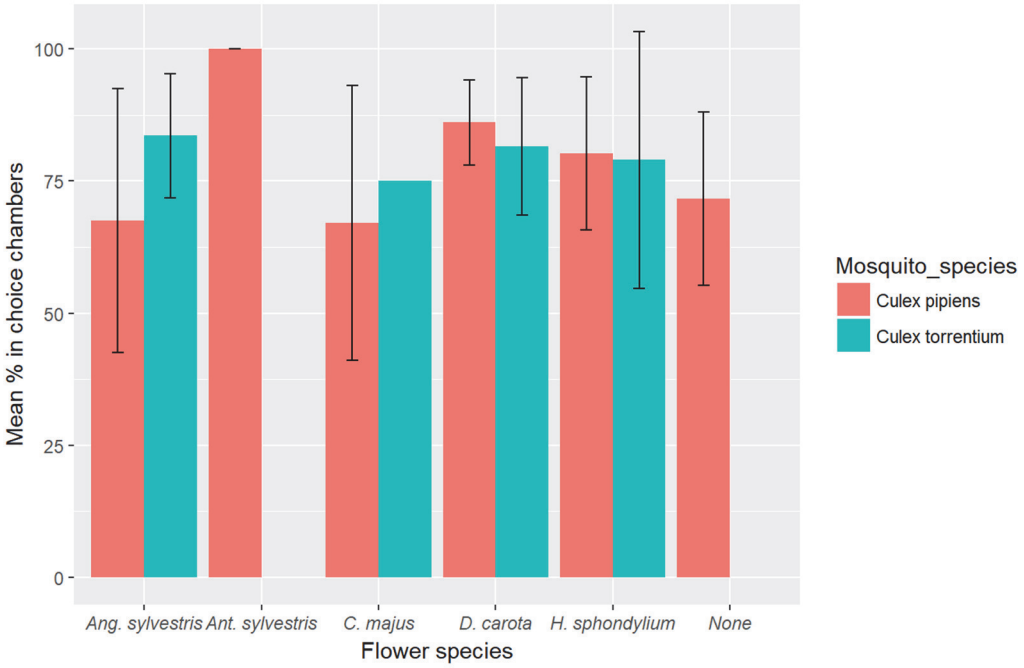


Figure 122 Mean percentage of mosquitoes found in choice chambers of each mosquito species when assayed against the named flower species. *Ant. sylvestris* was unavailable for assay against *Culex torrentium*. There was no significant difference in mean No. in choice chambers between assays using floral volatiles and those using negative controls only.

5.3.3.2 Preference of mosquitoes to floral stimuli

Culex pipiens s.s. and *Culex torrentium* each showed a significant preference for each floral variable offered when tested separately using Pearson's Chi-squared test ($p < 0.01$ for all floral species). Figure 123 and Figure 124 show the EPI scores for the combinations with sufficient iterations for analyses.

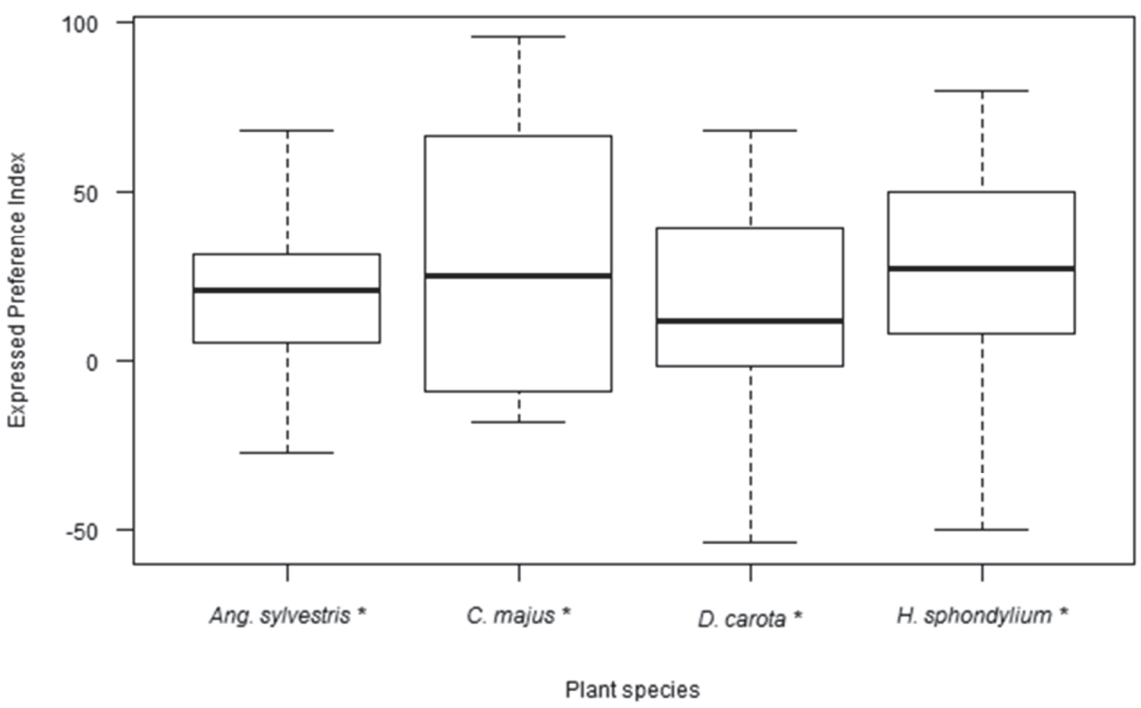


Figure 123 Expressed Preference Index values of *Culex pipiens s.s.* to selected floral odours. * denotes a statistically significant preference for the floral variable over negative control (χ^2 $p < 0.01$).

Figure 124 Activated Preference Index values of *Culex torrentium* to selected floral odours. * denotes a statistically significant preference for the floral variable over the negative control (χ^2 $p < 0.01$).

5.3.3.3 Interspecies comparison

Results were tested for variance of the EPI between the different floral variables assayed and between the mosquito species. EPI data were tested for normality and homoscedasticity in three groups prior to ANOVA testing: *Culex pipiens s.s.* Vs. floral species' EPIs, *Culex torrentium* Vs. floral species' EPIs and finally both mosquitoes collated Vs. floral species' EPIs. All combinations were normally distributed (Shapiro-Wilk test, p-values > 0.05) and homoscedastic (Bartlett test, p-values > 0.05), and therefore met the assumptions for ANOVA testing. Table 26 shows the results of ANOVA testing; all p-values were > 0.05 demonstrating that there was no significant difference between the EPIs of the floral volatiles assayed and that the observed differences shown in Figure 123 and Figure 124 are not significant.

Table 26 ANOVA test results of the EPIs of floral variables to the names mosquito species. $p < 0.05$ would denote a significant of expressed preference of the mosquito species towards one or more of the floral species assayed.

	P-values from ANOVA of EPIs to different floral variables
Culex pipiens s.s.	0.77
Culex torrentium	0.21
Culex pipiens s.s. and *Culex torrentium*	0.31

Having ascertained that there was no significant difference between EPIs of the various floral variables, in *Culex pipiens s.s.* or *Culex torrentium*, the EPIs of the two mosquito species were tested against each plant to ascertain whether there was any difference between mosquito species. These data were tested using Shapiro-Wilk and Bartlett tests, as before, and were shown to meet the requirements for ANOVA. Results of the ANOVA testing (Table 27) showed no significant difference in EPI between *Culex pipiens s.s.* and *Culex torrentium* in relation to any of the individual floral variables.

Table 27 ANOVA test results of the EPIs from *Culex pipiens s.s.* and *Culex torrentium* when assayed against similar flower species. P< 0.05 would denote a significant difference. *C. majus* is n/a as there were insufficient *Culex torrentium* assays against this species for meaningful analysis.

	P- values from ANOVA of EPI difference of *Culex pipiens s.s.* and *Culex torrentium*
Ang. sylvestris	0.36
C. majus	n/a
D. carota	0.51
H. sphondylium	0.67

5.3.3.4 Species collation

Due to the lack of significant differences between EPIs expressed by *Culex pipiens s.s.* and *Culex torrentium* towards similar floral variables, data for the two mosquito species were collated to visualise the variation shown in the data (Figure 125). Within the collated data all floral variables had statistically significant, positive, EPIs, although there were no significant differences between the EPIs of the different floral variables.

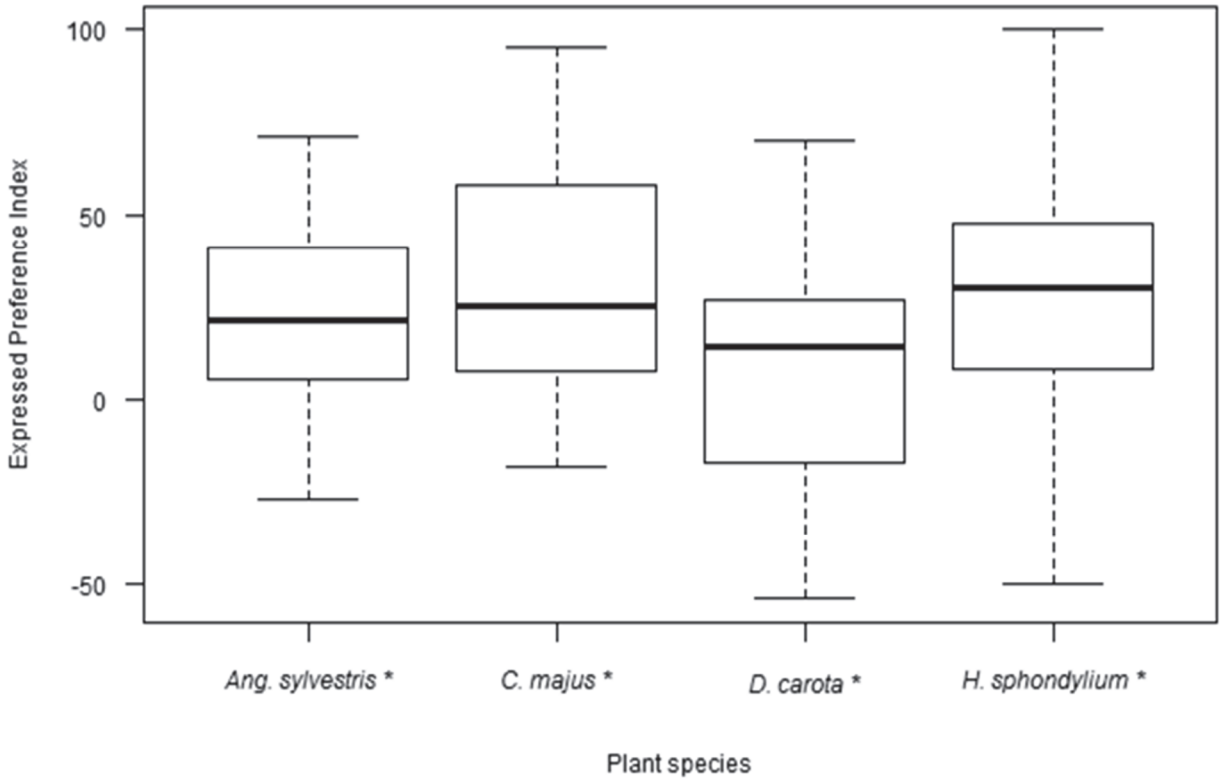


Figure 125 Expressed Preference Index results of *Culex pipiens s.s.* and *Culex torrentium* combined to selected floral odours. * denotes statistically significant results (χ^2 $p < 0.01$).

5.3.4 Discussion

This chapter described the first laboratory behavioural assays of male British *Culex pipiens s.s.* and *Culex torrentium* sources of floral nectar. It demonstrated that all of the members of the Apiaceae tested, *Heracleum sphondylium*, *Angelica sylvestris*, *Anthriscus sylvestris*, *Conopodium majus* and *Daucus carota*, were attractive to both mosquito species, despite large inter-specific differences in the ratio of floral VOCs emitted by these flowers. Both *Culex pipiens s.s.* and *Culex torrentium* species showed similar attraction to sugar sources in all cases.

The significant preference expressed towards the plant species used in these assays, each with demonstrably different VOC signatures (5.1.2.4), is strong evidence of olfactory-mediated behavioural responses being elicited by the flower species presented within this study. It highlights the ability of mosquitoes to use olfaction for long range location of sugar sources. The lack of significant differences between the EPIs to the different floral species is suggestive of multiple compounds being attractive to these mosquitoes; indeed, as these flowers species and mosquitoes share the same geographic range it would provide a distinct survival advantage for these mosquitoes to be able to identify them as sugar sources available for exploitation. Alternatively, there might be a ubiquitous chemical, not listed among those chemicals in the panel of target panel analysed in chapter 5 being chosen that explains the similar attraction; this does seem unlikely however especially since for the most part, phytochemical VOCs appear to be attractive constituents of a blend rather than in isolation (Peterson *et al.* 1994; Cha *et al.* 2008; Padmaja *et al.* 2010).

The different floral odorants emitted by various plant species can have a role in niche separation and reproductive isolation in insects (Byers *et al.* 2014), however, the lack of significant behavioural difference between *Culex pipiens s.s.* and *Culex torrentium* offers no evidence in support of sugar feeding, in relation to these floral variables, forming an axis for niche separation. Successful sugar feeding has been shown to improve reproductive fitness by providing sufficient energy for male swarming (Yuval *et al.* 1994). In arid areas of the world where flowering plants may be rare for some of the active period for mosquitoes, sugar has been seen to be a potentially limiting factor (Müller and Schlein 2005, 2006; Junnila *et al.* 2010) and the effect of experimentally removing a plentiful, preferred, sugar source, the

invasive shrub *Prosopis juliflora* caused a dramatic decrease in the mosquito population (Müller *et al.* 2017). However, in the UK during the late spring, summer and autumn mosquito active period, flowers are rarely likely to be a limiting factor, and in suburban settings with the additional sugar sources provided by planted gardens, it is even less likely. Also, the sharing of habitats and the identical morphology of their feeding structures/mouthparts has allowed development of similar preference in both mosquito species to these common plants. It is also highly likely for these species that floral preferences and the genes which code for odorant binding proteins and olfactory receptors (Leal 2013; Montell and Zwiebel 2016) predate speciation. Consequently, if as is likely, sugar sources are not a limiting factor (Foster 1995a) causing competition between these species, then no further behavioural change would be expected to occur (Wcislo 1989). It must also be considered that although the Apiaceae were selected for inclusion in this study based on field observations around nectar availability, seasonal occurrence and shared habitats with the mosquitoes, their floral emissions might only be mildly attractive to mosquitoes. There may be other flora which would be much preferred by mosquitoes and which might elicit differences in response still await discovery through assays against additional flower species. The expansion of behavioural assays to include more plant species is one of the key areas for further work in this area. With regard to the speciation of *Culex pipiens s.s.* and *Culex torrentium*, in the context of shared affinity to sugar sources described here, their sympatry as described in chapter 5 and in the literature (Service 1968; Jupp 1979), perhaps the most convincing hypothesis is that reproductive isolation between genetic variants was induced by cytoplasmic incompatibility caused by the infection of *Culex pipiens* by *Wolbachia pipientis* strains (Werblow *et al.* 2014). Thus far Wolbachia has not been isolated from *Culex torrentium*, but is widely distributed through *Culex pipiens* populations (Khrabrova *et al.* 2009).

The first measure of preference (PI) utilised, seemed, initially, to be adequate for these data, and was in line with approaches utilised by others in olfactory-mediated behavioural studies using dual choice olfactometers (Afify *et al.* 2014; Vinauger *et al.* 2014), some of whom chose to multiply the PI by 100 to return results on a -100 to +100 scale (Nyasembe *et al.* 2012). However, it does not factor in the mosquitoes' activation rate within the index itself, thereby requiring the activation percentages to be considered separately. The solution applied here

was to multiply the PI with the expressed percentage to generate the Expressed Preference Index (EPI), effectively weighting the PI with the power of the stimuli to provoke movement into the choice chambers, rather than remain in the release chamber or main arena. In lengthy assays of this type, the EPI is a more conservative measure than PI and should be considered as a desirable modification to avoid assays with very low activation rates but high preference indices having an overly large impact on the experimental outcome. Other measures of preference might be possible to use with these data, such as Ivlev's Electivity (Ivlev 1964) or Chesson's α (Chesson 1978), however these indices of preference are unnecessarily complex for these assays as they factor in aspects more related to predation – such as the effect of prey being removed from the available pool of prey available to be selected – which did not occur in this experimental model.

Using only wild mosquitoes and plant materials was challenging and increased the difficulty in carrying out assays related to flowers where there is little temporal overlap between flowering period and mosquito occurrence. For example, within the data, results are reported for the behaviour of *Culex pipiens s.s.* when presented with *Anthriscus sylvestris*, and *Culex torrentium* and *Conopodium majus*; in both of these cases the flowering season is relatively early in the year reducing the overlapping window of occurrence between the species within which experimentation could be carried out. However, in terms of attempting to assess the naturally occurring interspecies interactions and possible variance within species, it was deemed that this was a hardship worth persevering with, as opposed to using laboratory cultured mosquitoes with their necessarily reduced gene pool and potentially atypical behaviour (Huettel 1976) and cultured greenhouse plants.

The use of wild mosquitoes and the collection of daily batches of males lead to some differences in the number of mosquitoes used in each assay per day, with a mean number of 24 (SD = 7.26) and a total of 2449 mosquitoes were used for 104 assays. The daily variation was judged to be acceptable for the study as the daily number used remained within the range commonly seen by the author in mosquito olfactory-mediated behaviour assays (Jhumur *et al.* 2006; Logan *et al.* 2008; Nyasembe *et al.* 2012; Cator *et al.* 2013). Higher importance was placed on carrying out the assays using wild mosquitoes, despite their

inconsistency of supply, rather than cultured mosquitoes where supply could be more controlled.

The use of excised flower sections, rather than whole plants, offered significant advantages in terms of low space requirements, particularly when dealing with tall plants such as *H. sphondylium*. As an approach, it might also be considered reductionist in terms of not having soils present in the variable setting. However, the excision process should be expected to liberate compounds associated with herbivory (Lohman *et al.* 1996). These herbivory induced (HI) emissions have been demonstrated to attract predators to herbivore damaged plants (Kessler and Baldwin 2000, 2002). Therefore, there may be compounds available to the mosquitoes in the behavioural assays which would not be emitted from an undamaged, whole plant. Therefore caution must be exercised when comparing results between studies which have taken the opposite approach when handling plant specimens (Wade and Wratten 2007). However, as the botanical specimens used in behavioural assays and in the chemical analysis (5.1.2.4) were prepared in the same manner, any compounds so liberated would be available to both parts of the study.

5.4 Evaluation of synthetic blends

5.4.1 Introduction

The concept of applying multiple lures to traps for sampling and control is not new, and forms the basis of many effective control measures (Müller, Beier, *et al.* 2010; Beier *et al.* 2012; Harrup *et al.* 2012; Cameron and Lorenz 2013; Lühken *et al.* 2014), and the study of plant derived candidate chemicals as repellents and lures is an area of study receiving much attention. For example Ballantyne and Willmer (2012) and Champakaew *et al.* (2015) investigated the repellent properties of botanical VOCs, significant effort however, has been increasingly applied to attraction of VOCs (Gouagna *et al.* 2010; Otienoburu *et al.* 2012; Nikbakhtzadeh *et al.* 2014; von Oppen *et al.* 2015; Scott-Fiorenzano *et al.* 2017).

Having ascertained, in the previous chapter, that male *Culex pipiens s.s.* and *Culex torrentium* were attracted to the VOCs emitted by several different Apiaceae species in olfactometer based behavioural assays, logically there were two potential avenues to continue the investigation, each with different goals. The first would be to focus on the chemicals that are known to be detectable by mosquitoes (Otienoburu *et al.* 2012; Nyasembe and Torto 2014; Pitts *et al.* 2014) to test whether similar behaviour would be observed in relation to synthetic blends based on the headspace fraction of these flower species. The second would be to use electroantennography (Bjostad 1988; Olsson and Hansson 2013) to investigate whether there were other VOCs present in the emissions from the plant species that have yet to be determined as detectable by male mosquitoes (Pitts *et al.* 2014). The route chosen was to focus on those chemicals which have already been identified as detectable/behaviour modifying and to test synthetic blends of some of those chemicals against male British *Culex* mosquitoes, and also trial the blends in the field to test their efficacy as additional baits for CDC light traps.

This chapter describes the use of two approaches to assess the potential of these synthetic chemical blends; the first being behavioural assays similar to those conducted using excised plant sections, and the second taking the form of a field trial using light traps augmented with these test blends.

5.4.2 Methods

5.4.2.1 Preparation of the synthetic blends

Two synthetic blends were trialled, both in the laboratory in behavioural assays using a dual choice olfactometer and in the field using light traps augmented with the candidate synthetic blends.

Having already carried out behavioural assays of excised florets of several different Apiaceae and found them to be attractive to both *Culex pipiens s.s.* and *Culex torrentium* a determination needed to be made as to which chemicals should be used for these trials. Originally, the behavioural outcomes were designed to be modelled against the chemical composition of the floral species utilised in the assays. However, because there was no significant inter-species preference difference, i.e. there was insufficient evidence to statistically state that observed preference differences were not solely the result of chance, this method of dimension reduction was lost to the analysis, effectively preventing this top-down approach from being used. Also, it was beyond the scope of this study, in terms of allocated time, to allow individual chemicals to be tested to facilitate a bottom-up approach to designing a blend. Therefore, an intermediate position needed to be taken. By comparing the chemical signatures between the flower species used in both behavioural assays and chemical analysis and with reference to the literature (Verhulst *et al.* 2011; Nyasembe *et al.* 2012; Otienoburu *et al.* 2012; Byers *et al.* 2014) a synthetic blend was designed for testing. The literature did provide many other compounds which would be worthy candidates for inclusion in these tests. However, the analysis applied was limited to compounds from the panel of compounds known to be detectable by male mosquitoes and that were found in the flower species included during the chemical analyses (5.1.2.4).

The compounds incorporated included (+)-limonene, (1S)-(-)-α-pinene, β-ocimene, (1S)-(-)-β-pinene, hexanal, nonanal, trans-2-nonenal and benzaldehyde. The ratio of these compounds was derived from the median ratio to the internal standard (1-bromodecane) from the chemical analyses carried out previously (Figure 126). Two different blends were created one with the full complement of synthetic chemicals and one with the two forms of pinene removed to determine whether their presence influenced the attractiveness of the blend. α-Pinene been shown to be an attractant for various insect Orders, including dipterans (Scarpati

et al. 1993), hymenopterans (Sato and Maeto 2006; Costello *et al.* 2008) and coleopterans (Chénier and Philogène 1989; Sweeney *et al.* 2004). B-Pinene has also been shown to be attractive across a range of insect orders, Diptera (Rudinsky *et al.* 1971), Coleoptera (Chénier and Philogène 1989) and Hymenoptera (Pettersson *et al.* 2000) for example. Information relating to the sources and measures of the chemicals in the blends can be seen in Table 28.

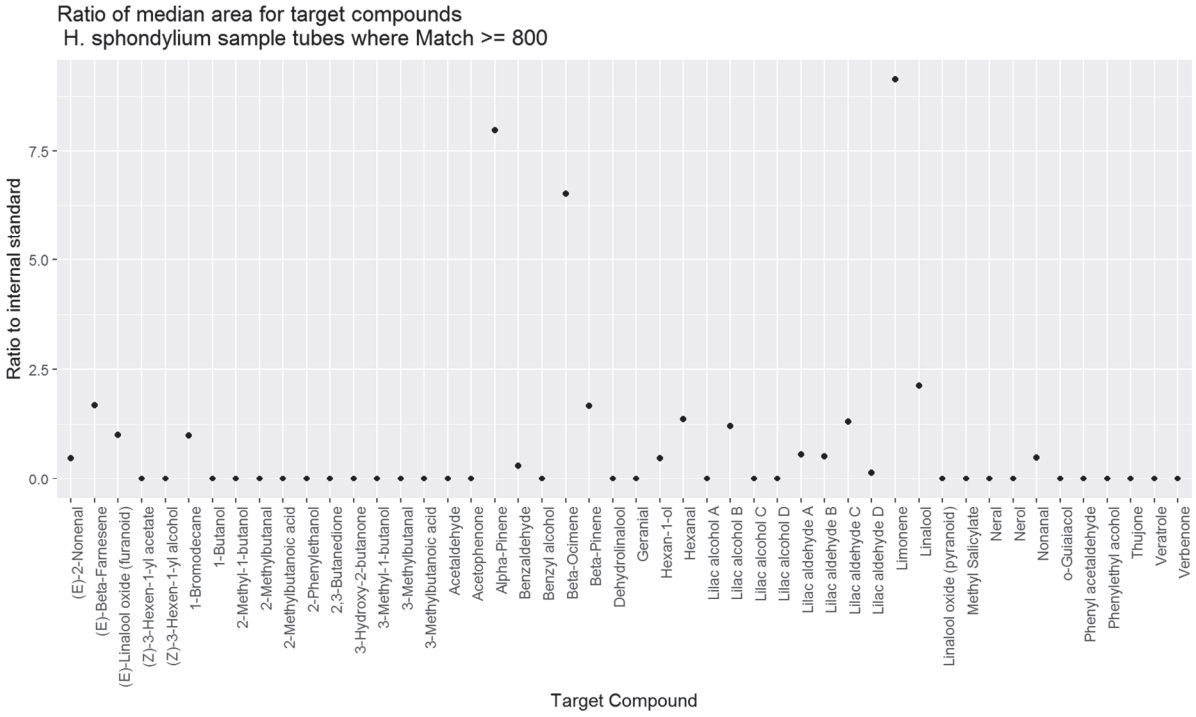


Figure 126 Ratio of the median area under the peak for the named chemicals to the internal standard (1-bromodecane). These ratios were the ratios chosen for the compounds used in the synthetic blends.

Table 28 Sourcing and details related to the chemicals used in the synthetic blend testing. The final two columns show the volumes of each compound included in the named blend stock mixture.

Chemical name	**Purity %**	**Manufacturer**	**Lot No.**	**Blend one volume (µl)**	**Blend two volume (µl)**
(+)-Limonene	96 %	Acros Organics	A0361193	950	950
(1S)-(-)-α-Pinene	98 %	Acros Organics	A0340129	0	816
β-Ocimene	>= 90 % as part of a mix of isomers	Sigma-Aldrich	MKBW4748V	720	720
(1S)-(-)-β-Pinene	98 %	Acros Organics	A0355854	0	173
Hexanal	96 %	Acros Organics	A0364536	146	146
Nonanal	97 %	Alfa Aeser	10179106	51	51
Trans-2-nonenal	97 %	Alfa Aeser	10197454	48	48
Benzaldehyde	99.5 %	Acros Organics	A0361752	30.1	30.1

Stock blends were prepared on a daily basis for both field and laboratory based applications and used for all assays and trials on that day. For the delivery mechanism, an approach modified from Cilek *et al.* (2003) was employed: 4 ml glass headspace vials, with plastic caps and polytetrafluoroethylene (PTFE) septa, were used as delivery mechanisms for the synthetic blend (Figure 127). The use of pipe-cleaners as wicks for the delivery of VOCs has been successfully used in a number of studys (Hayes *et al.* 1993; Kline 1994; Cilek 1999; Cilek and Hallmon 2008), due to its practicality and ease of standardising the delivery rate by adjusting the length of the exposed wick. The same approach was applied in the field and laboratory except for the length of the exposed wick being shortened for the laboratory assays. The wick consisted of a length of pipe cleaner piercing the septum, one end of which was submerged in the liquid contained in the vial and the other protruding into the space surrounding the vial. For the field testing, the exposed wick was 20 mm long, and in the laboratory assays, the exposed wick was 10 mm long to reduce the VOC emission rate. Synthetic blend treatment

preparations consisted of 2000 µl of distilled water and 200 µl of the prepared stock blend. Because these blends were non-miscible with water and less dense than it, the blend existed as a floating layer mid-way up the delivery vial; this ensured that the the wick was always submerged and that the synthetic blend could be emitted throughout the assay. Negative controls consisted of 2200 µl of distilled water presented in the same manner.

Figure 127 Example of the delivery mechanism for synthetic blends. New 4 ml glass vials with plastic caps and PTFE septa were used to hold liquids. New pipe cleaners were used as a wick. In the field, the exposed wick length was 20 mm; in the laboratory, a 10 mm exposed wick length was used.

5.4.2.2 Field evaluation using light-trapping

Field evaluation of the synthetic blends was conducted to investigate whether the inclusion of either of these specific synthetic blends as an additional lure, augmenting the light trap's lamp, would result in increased collection of mosquitoes compared with the lamp alone. Four identical CDC 512 mini light traps equipped with "air-actuated gates systems" and LCS-2 photo switches (Hock 2016) were used for the field trial and each powered by 6-volt rechargeable batteries, which supplied sufficient power for three nights trapping before recharging was required. The traps' LCS-2 photo switches were calibrated before deployment such that under similar fading light conditions they all switched the traps' lamps and fans on within a range of two minutes; ensuring standardisation of sampling effort.

For each trapping session, two traps were augmented with the synthetic blend treatment, and two with the negative control. A randomised 2x2 Latin square design was implemented to determine the starting positions and rotation of treatments, thereby controlling for any location specific effects on sampling (Perry *et al.* 1980). Variables and controls were prepared as above and mounted in the trapping chamber of each of the four light traps. To avoid contamination between treatments, a new piece of White Tack (Bostik) was used to secure the vial in an upright position in the centre of the mesh that forms the exhaust of the trap (Figure 128) at the start of each trapping sessions. By locating the lure inside the trapping chamber, at the centre of the trap's exhaust mesh, the volatile emissions from the treatment would be actively transported by the air flow from the trap's fan, leading to efficient distribution of the odour plume. The position of the lure on the inside of the trap lead to a configuration more similar to that used in the Biogents Sentinel and Mosquito Magnet traps where the generated airflow is utilised to drive the delivery of the lure (Biogents, 2011; Mosquito Magnet, 2013). By locating the lure inside the trapping chamber, it was sheltered from the wind thereby controlling for different rates of evaporation from the wick that could be associated with the vial being located externally on or adjacent to the trap. In other studies, lures used in conjunction with CDC mini light traps have been arranged in several ways (Reisen *et al.* 2000). Ritchie and Kline (1995) hung a CO_2 lure adjacent to the trap and positioned a vial of octenol next to the trap's entrance. Similarly, Van Essen *et al.* (1994) placed their lure adjacent to trap entry point. Obenauer *et al.* (2013) hung their chemical lure of the outside of the trap such that it was adjacent to the mid-point of the collection sock, approximately equidistant from the trap entry and the exhaust. It is possible that the different volatile delivery method may cause a difference to its efficacy, and so the location of lures should be considered when comparing results between studies. Variables and controls were transported to the field as prepared vials with intact septa. Each septum was carefully pierced using a mounted needle to create a pilot hole; then the pre-cut pipe cleaner wicks were pushed through the hole ensuring the wick was tightly fixed in position with 20 mm exposed to the external atmosphere.

Figure 128 Placement of the variable/control within the trapping chamber of the CDC 512 mini light trap. A new piece of White Tack (UHU) was used to secure the vial in a vertical position in the centre of the mesh of the chamber at the start of each session.

Trapping was conducted at Wybunbury Moss NNR, under kind permission of Natural England. A survey of the Wybunbury Moss larval and adult mosquito populations has been carried out previously in this thesis, and so it was possible to select an area at the north of the site which was seen to have a relatively high adult mosquito catch rate (Figure 129). The surrounding flora was similar at all four trap locations and consisted of alder carr woodland; all four traps were located within the woodland approximately equidistant from the nearest border (Figure 130) to try and minimise the edge effect on any particular trap (Ries *et al.* 2004). The immediate surrounding for each trap was similar in terms of vegetation but differed at ground level. There were no visible sources of sugar within the vicinity of the trap locations, the nearest being garden plants adjacent to the housing directly north of the traps. Traps LT01 and LT02 were placed adjacent to semi-permanent pools of water (Figure 131). These pools

typically had available water but had been seen to dry up briefly in the height of summer 2015. Throughout this trapping study, in autumn 2016, the pools were filled. The ground under trap LT03 and trap LT04 was considerably drier than LT01 and LT02 and had no standing water adjacent; it did still have a thick moss layer which released water when trodden on (Figure 132). Differences between the trap locations were addressed in the experimental design by rotating the variables and controls after every three nights, ensuring that all treatment combinations occurred across the four traps.

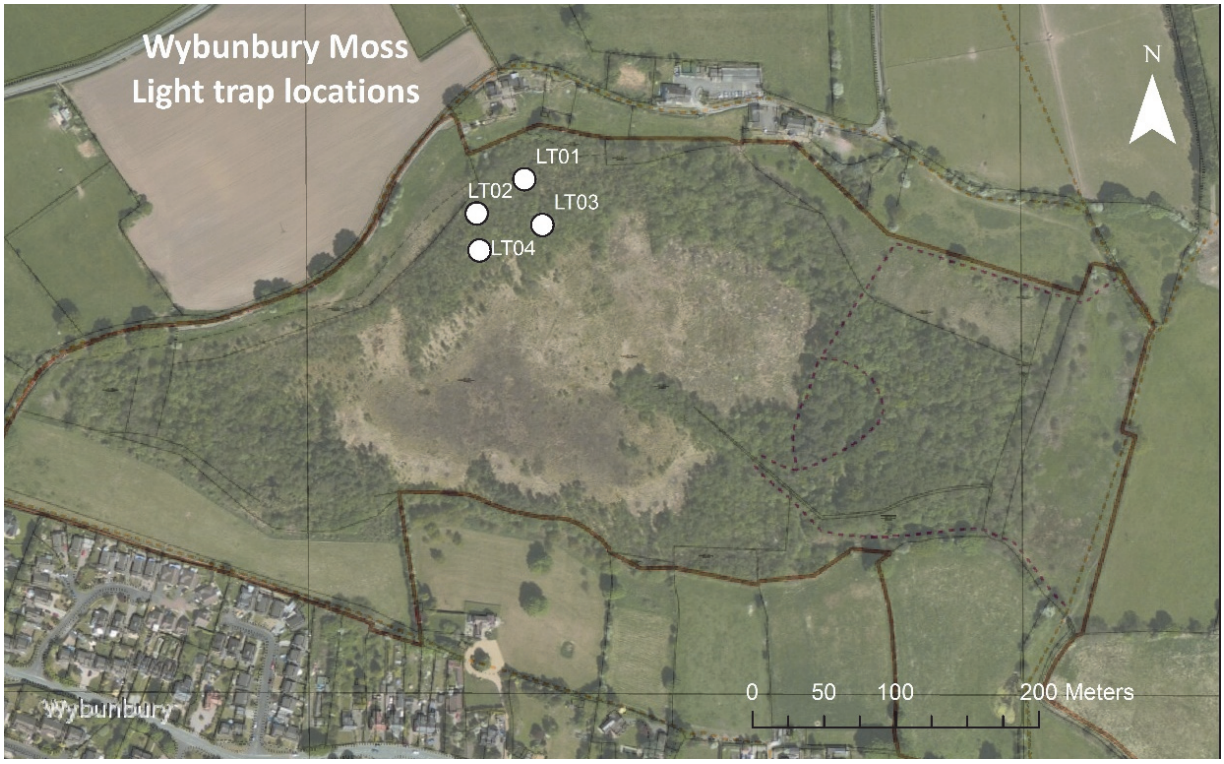


Figure 129 CDC 512 mini light traps were placed at Wybunbury Moss. The traps were located in an area previously identified as being a relative hotspot for the collection of adult mosquitoes using these traps.

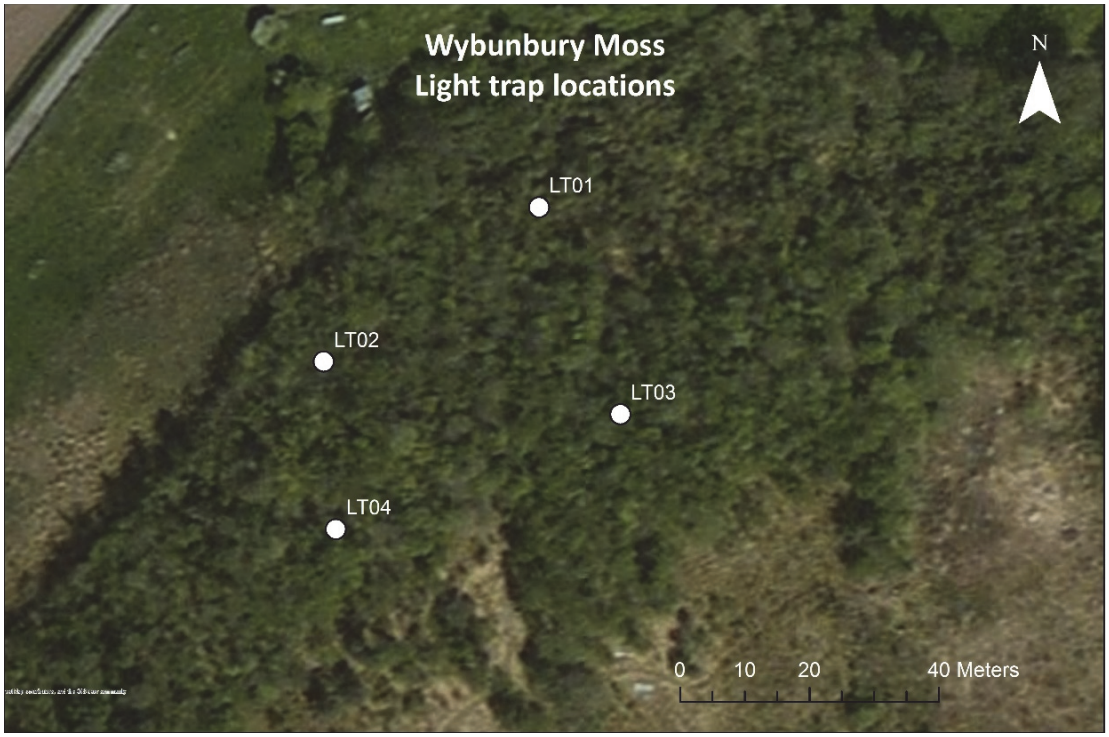


Figure 130 Area chosen for the placement of the light traps consisted of mixed alder and birch woodland, in peat soils. LT01 and LT02 were immediately adjacent to shallow pools which were observed to contain juvenile mosquitoes. LT03 and LT04 were on wet ground, but without obvious pooling of water.

Figure 131 Shallow pools, known to be inhabited by juvenile mosquitoes, were adjacent to LT01 and LT02. Here the proximity of the water to LT02 is apparent. LT01 was similarly closely associated with water.

Figure 132 LT03 and LT04 were set over wet ground. Although pools of water were absent, the ground was sufficiently wet that water would temporarily pool in footprints before the moss sprung back to fill the depression.

Each trapping event consisted of two replications of one trap being a negative control and one utilising the experimental treatment. Initial trap conditions were generated randomly and subsequently cycled through all permutations. After three nights of trapping with any given configuration, the batteries were collected and recharged overnight, before returning to the field for the next night's trapping. The experimental treatment and negative control configuration was moved to the next arrangment for the next three repetitions following each battery recharge. Twenty-nine separate trapping events were conducted, during autumn 2016, generating 116 trapping nights of data for analysis.

Trap manipulation took place between 16:00 and 18:00 each night and involved the collection of the previous night's samples and the replacement of variables/controls with new ones. Samples were collected with an electric aspirator, where possible live Lepidoptera were counted and released in the field, the remaining specimens were transferred to labelled vials prior to storage at -20 °C until processing. Samples were processed in the laboratory following the conclusion of all trapping. Processing consisted of counting the number of non-target non-Lepidopterans, the number of Lepidopterans (collated with the in-field count) and the number, species and sex of the mosquitoes collected per trap per night.

The count data for mosquitoes and non-target insects was collated and tested using a Chi-squared test to determine whether capture rates between treatments offered evidence of differential trapping rates.

5.4.2.3 Laboratory evaluation using behavioural assays

Behavioural assays of synthetic blends were conducted using the same method as that applied to the flower species assays, barring that only *Culex pipiens s.s.* were used for these assays and that the variables being delivered were the synthetic blends and controls.

Analysis of the results of the behavioural assays was broadly the same as that applied to the floral variables. The only difference between the analyses was the application of Wilcoxon rank sum testing, rather than Kruskal-Wallis, as for this analysis only two groups would be tested against one another.

5.4.3 Results

5.4.3.1 Results of field evaluation using light-trapping

The field trial of the two blends resulted in the collection of four different species of mosquitoes, these being *Culiseta annulata, Culex pipiens s.s., Culex pipiens* s.l. *and Anopheles claviger*. A total of eighteen individuals were collected in Blend 1 traps, two in Blend 2 traps and eighteen in negative control traps. Collection data (Figure 133) were tested for normality using the Shapiro-Wilk test and found to be non-normally distributed: Blend 1 treatment $W = 0.644$, $p < 0.01$ and negative control $W = 0.496$, $p < 0.01$. Blend 2 treatment $W = 0.250$, $p < 0.01$, $p < 0.01$ and negative control $W = 0.168$, $p < 0.01$. Therefore, the Wilcoxon rank sum test was applied to each treatment and its control. There was insufficient evidence to reject the null hypothesis that each treatment and control will have equal medians. Blend 1 vs control $W = 296$, $p = 0.8531$ ($n = 24$) and Blend 2 vs control $W = 593$, $p = 0.6049$ ($n = 35$).

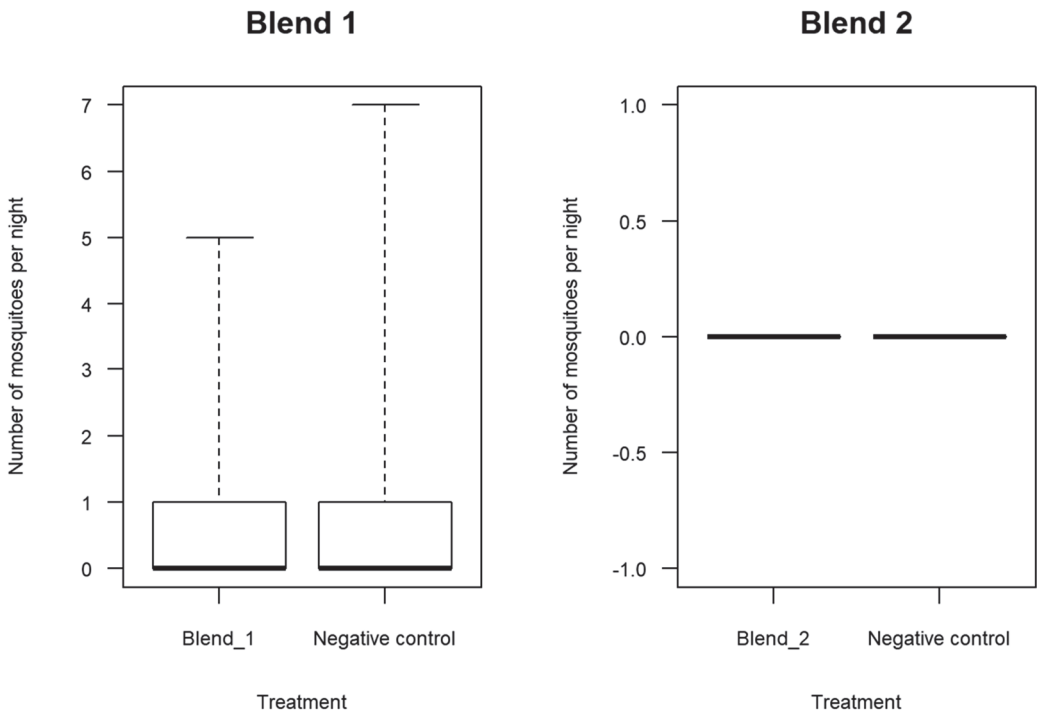


Figure 133 Number of mosquitoes caught per night during the two non-contemporaneous sampling tranches. Blend 2 catch numbers were either 0 or 1 and therefore have insufficient integer range for plotting as a box plot. Differences between treatments and controls were not found to be significant using Wilcoxon rank sum test (Blend 1: W = 296, p = 0.8531 (n= 24), Blend 2: W = 593, p = 0.6049 (n= 35)).

For context, the number of non-target insects collected during the trapping effort was also recorded and are reported here. These treatment data (Figure 134) were also tested for significance against their controls, and insufficient evidence was shown in either case to reject the null hypothesis. Blend 1 vs control W = 294.5, p = 0.9012, Blend 2 vs control W = 648, p = 0.3854.

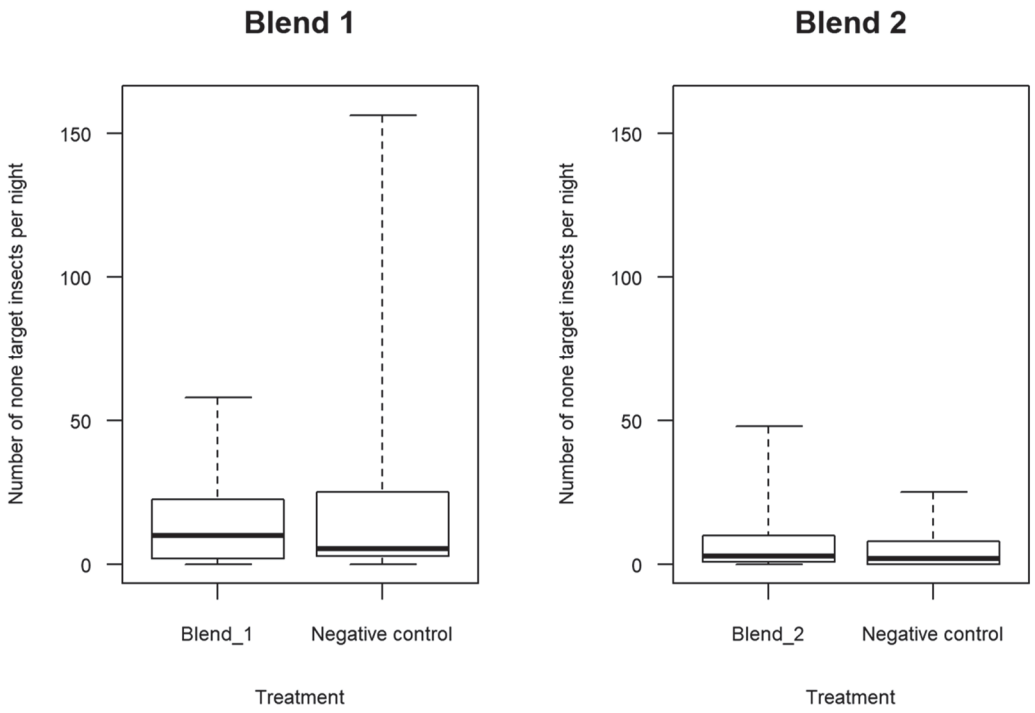


Figure 134 Daily catch rate of non-target insects are shown here as a comparison to the mosquito catch rate. These highlight that the traps were functioning properly throughout the sampling, and show a significant reduction in the number of insects collected between the sampling tranches. When the negative controls were compared between sampling periods, Wilcoxon rank sum testing revealed that the collection rates were significantly different (W = 550, p = 0.0125) (Blend 1 n = 24, Blend 2 n = 35).

Blend 1 and Blend 2 mosquito catch data were not tested directly against each other statistically. This was because the tests were not conducted contemporaneously and the overall catch rate changed dramatically between sampling tranches; this was exemplified by the significantly reduced non-target insect catch rate in the negative controls between the two tranches of sampling (Wilcoxon rank sum test, W = 550, p = 0.0125). Whilst it would be possible to transform the Blend 1 and Blend 2 data by the ratio of the negative controls for each treatment the lack of significance of each blend against controls rendered this step redundant.

5.4.3.2 Results of laboratory evaluation using behavioural assays

Pearson's Chi-squared testing of the number of mosquitoes in choice chambers when presented with each blend showed no significance between blend 1 and the negative control $\chi^2 = 11.983$, p = 0.088, or blend 2 and the negative control $\chi^2 = 4.5833$, p = 0.239. Due to the small number of repeats (Blend 1 n = 8, and Blend 2 n = 4), the Pearson's Chi-squared test was

calculated using p-value computed for a Monte Carlo test, based on 2000 replicates, by adding the *simulate.p.value = TRUE* switch to the R code for the calculation.

Figure 135 shows the distribution of the Expressed Preference Index (EPI), as introduced in the Behavioural Analysis chapter. These EPI data, when tested using the Wilcoxon rank sum test, did not show a significant difference between the two synthetic blends under trial (W = 17, p = 0.932).

The calculation for EPI is repeated here:

EPI = ((Xv - Xc)/ (Xv + Xc)) * E

Where:

EPI = Expressed Preference index

Xv = number of mosquitoes choosing variable

Xc = number of mosquitoes choosing control

Xr = number of mosquitoes in release chamber

Xm = number of mosquitoes in main arena

E = Expression Percentage = (Xv + Xc) / ((Xv + Xc + Xm + Xr)) * 100

When the EPI results for the synthetic blends (Figure 135) were compared against those of the naturally occurring VOCs emitted during assays of *Culex pipiens s.s.* (Figure 136), the reduction of attractiveness shown towards the synthetic blends appeared very apparent, although when the aggregate data, i.e. all flower species and both synthetic blends, were tested using the Kruskal-Wallis rank sum test, differences between the groups were found not to be significant KW χ^2 = 3.238, df = 5, p = 0.663.

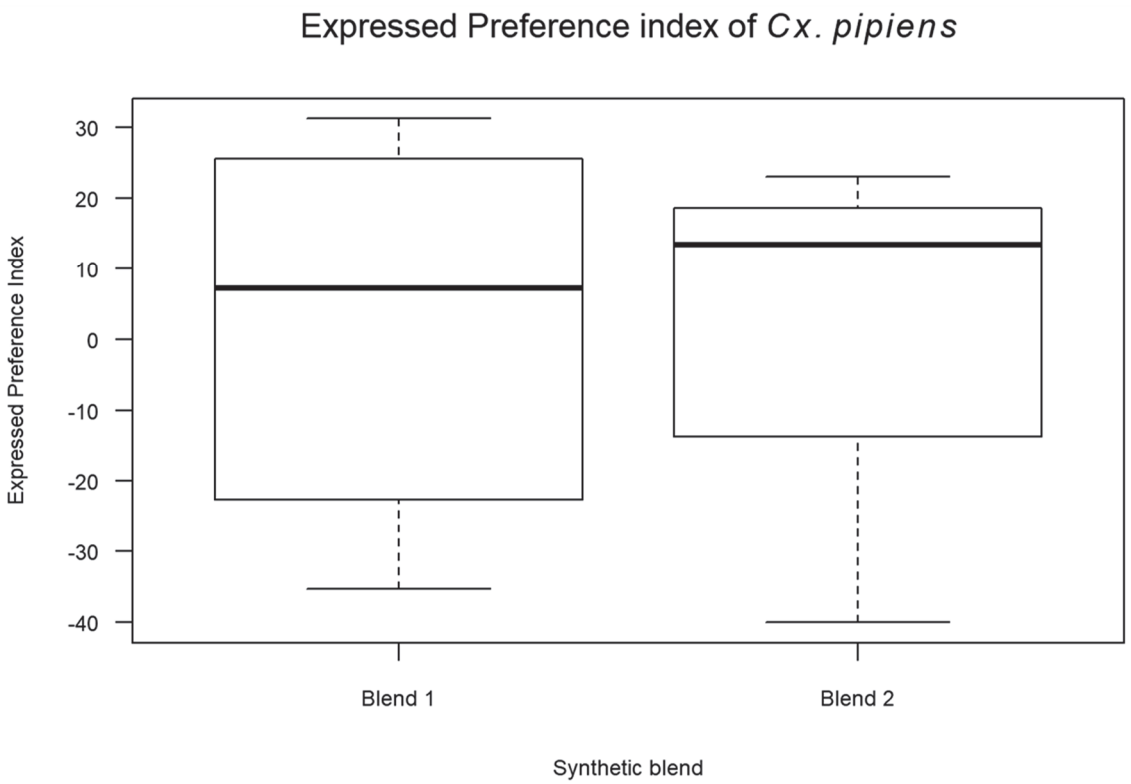


Figure 135 Expressed Preference Index (EPI) of *Culex pipiens s.s.* mosquitoes towards the two synthetic blends tested. There was no significant difference between Blends 1 and 2 when tested using the Wilcoxon rank sum test W = 17 p = 0.932

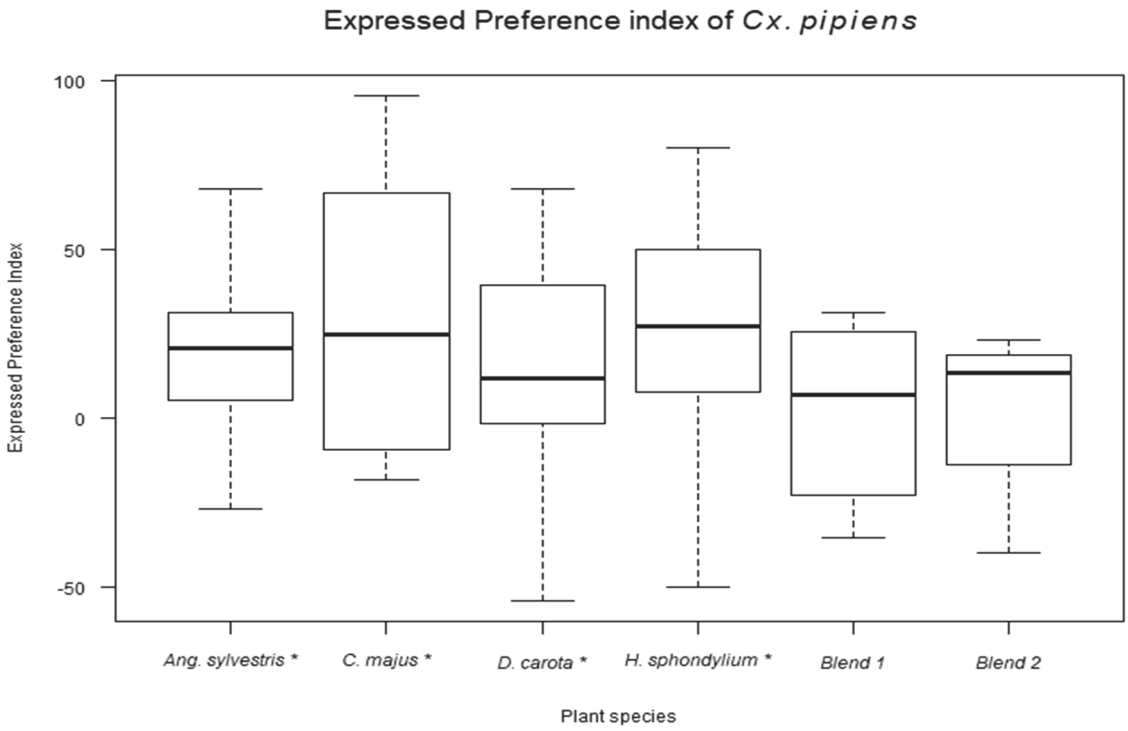


Figure 136 Expressed Preference Index of *Culex pipiens s.s.* to selected floral odours. These data were generated in the behavioural assays against excised floret of the species named here. * denote a statistically significant preference for the variable over negative control (χ^2 $p < 0.01$).

5.4.4 Discussion

Laboratory behaviour assays showed no significant difference in attractiveness, as reported by calculated EPI, between the two blends, and the count data between the variable and the negative control was also found to be insignificant. The results of the field testing showed no significant change in the number of mosquitoes collected in light traps augmented with either of the synthetic blends tested compared to the negative controls. Similarly, there was no significant change in the number of non-target insects. Due to the low catch rate of adult mosquitoes observed during trapping using CDC light traps at Wybunbury Moss both during this field test of volatile blends and the previous survey work (3.2.2.6)**,** this study is likely to be somewhat underpowered and would be likely to only detect very strong preference towards the lure. Considering the intended use for this lure as part of mosquito control, any response less than a strong effect would be insufficient for purpose. Given the results of the laboratory testing of these lures there is no evidence of any such large preference for the lure.

The rationale for testing these blends late in the year was based on the observation that light trapping of UK *Culex* mosquitoes can be more efficient during September and October as mosquitoes become more likely to enter buildings for diapause (Silver 2008c). It was also hoped that the pre-diapause sugar feeding drive (Bowen 1992a), and the observed reduction in the number of plants in flower at the test site, Wybunbury Moss, might combine to create an ideal situation for trapping using a sugar associated attractant. The substantial reduction in overall collection rate observed between the blend 1 and blend 2 trapping tranches was a factor of the weather, specifically a reduction in overnight temperatures, rather than any change in the method or in relation to the blend being tested.

The testing of synthetic chemical blends was conducted concurrently in the laboratory and the field to facilitate the employment of both approaches during 2016. The non-significant results of the field experiments, when considered alongside the synthetic blend behavioural assays provided corroborating evidence that the two synthetic blends were less attractive than the total odorant emissions of the flowers. This supported the hypothesis that mosquito phytochemical related behaviour is dependent upon the ratio of a number of VOCs (Bruce *et al.* 2005).

The methods of testing applied here were shown to be fit for purpose and could be used to test the efficacy of different combinations of floral derivative VOCs for the purpose of creating lures suitable for the augmentation of control and surveying tools currently available. Through augmentation using phytochemical, sugar associated, VOCs, there is the potential to improve the effectiveness of any control or surveying method that relies upon the attraction of mosquitoes to a surveying device or control measure. These VOCs also have the advantage of being attractive to both sexes as opposed to host, or oviposition derived odours which attract females predominately (Piñero and Dorn 2007; Pitts *et al.* 2014; Montell and Zwiebel 2016). Therefore, applications such as attractive toxic sugar baits (ATSBs) (Müller, Beier, *et al.* 2010; Revay *et al.* 2014; Ding *et al.* 2016; Fikrig *et al.* 2017) or delivery of selective entomopathogenic fungi (Shah and Pell 2003) or bacteria such as the trans-infection of Wolbachia from one species to another to induce cytoplasmic incompatibility (Laven 1967; Werren *et al.* 2008; Iturbe-ormaetxe *et al.* 2011), could benefit and display enhanced performance with optimised phytochemical lures.

When considering the methods applied and potential changes for further study, the decision to use wild mosquitoes during the laboratory behavioural assays caused some resource shortage towards the end of the experimentation. Wild mosquitoes, reared from field collected eggs, were used throughout the behavioural assays to maintain consistency with the previously conducted behavioural assays using natural, excised, florets. Towards the end of the study availability of new specimens became the limiting factor as wild females stopped oviposition and focussed on preparing for overwintering; this availability effectively capped the number of possible repetitions. More repeats would, eventually, enable the declaration that the results were significant and not chance. However, if the preference were to remain similar to that shown in the data so far, the magnitude of the attraction would be too low for either of the blends tested to be a viable candidate for field applications.

A limitation of this study was that the availability of authentic standard chemicals was limited and so it is possible that one or more essential ingredients were missing from the blends tested. Further studies should be carried out on these compounds and others from the panel of target VOCs shown to be involved in mosquito olfactory mediated behaviour and should take the form of laboratory-based subtractive assays. Subtractive assays are those where components of a blend are removed one at a time, and the reduced blend then assayed against the complete one to determine the effect of removal (Otienoburu *et al.* 2012). Also, in future work, testing different VOC emission concentrations in the behavioural assays is also worth investigation as dose dependent differences in attraction to floral odorants has been demonstrated (Hao *et al.* 2013) as has dose-dependent detection in EAG studies (Beyaert *et al.* 2010). Once a blend which exhibits a significantly increased attraction rate, of sufficient magnitude, perhaps EPI > 80 % then a trial of the blend in the field should be carried out. It would help these further studies to unshackle the synthetic chemical blend development process from being constrained by the characteristics of the Apiaceae analysed here as there are several compounds which do not feature prominently in the headspace of these flowers which have been seen to be attractive to mosquitoes. For example, throughout the synthetic blend development process, the relative lack of phenylacetaldehyde in the chemical analysis of Apiaceae was especially interesting, where it was only found in *Anthriscus sylvestris* at < 0.1 ppm. Phenylacetaldehyde has been shown to be attractive to mosquitoes and was one of only three compounds required for the most attractive blend found by Otienoburu *et al.*

(2012) in their subtractive blend optimisation paper. In other flower species, this has been found to be amongst the most prominent compounds within the headspace (Jhumur *et al.* 2007, 2008) and therefore might be considered to be one of the primary attractants for any blend. At the other end of the spectrum of attractiveness, some of the compounds included in these synthetic blends, limonene, for example, have been referred to as insecticides (Hebeish *et al.* 2008). Evidence exists of limonene acting as a repellent when part of an essential oil blend (Nerio *et al.* 2010), although limonene alone was not always seen to be repellent to mosquitoes (Dogan and Rossignol 1999). VOC's activity may be different depending on the blend of the other emitted odours (Smallegange *et al.* 2005, 2009), they may also have different activity depending upon the mosquitoes' current life stage, for example, a compound associated strongly with sugar sources might be unattractive to a female mosquito requiring a blood meal (Gary and Foster 2006), where the same odour pre-mating might be an attractant.

The challenge and work involved in developing a floral based attractant should not be underestimated, and lures which have been demonstrated to be attractive in laboratory assays (Jhumur *et al.* 2007; von Oppen *et al.* 2015) might not be effective in larger semi-field assays and field deployments as demonstrated by Fikrig *et al.* (2017). Perhaps part of the issue when testing preference in the laboratory is that the techniques have developed to a point where even small attraction differences can be statistically tested and found to be significant. Preference assays are always carefully controlled to remove non-candidate stimuli to maximise confidence in the result. However, the magnitude of the attraction is often lost in the pursuit of a significant result. The likelihood that a mosquito would choose to interact with compounds that show significant attraction may be overstated when these compounds are moved into a field setting with the plethora of other stimuli. This the justification behind the proposal of the expressed preference index. To achieve a high EPI within the low airspeed method used here, a large proportion of mosquitoes must activate and make their way to the selection chamber associated with the variable; any that choose the control or do not make a choice, i.e. those that remain in the main flight arena or the release chamber, reduce the EPI. This is a deliberately conservative measure of preference designed to prevent overstatement of attraction, and high EPI values communicate both a strong preference for and high motivation to move toward the preferred stimuli.

6 General Discussion

The success of the thesis to address the overall aims of the research is reflected upon in this chapter and is conducted as an overview as chapter specific content was discussed in each chapter.

New knowledge has been generated regarding the mosquito occurrence within the study area. North Staffordshire and the surrounding area previously held very few mosquito records, and as such any attempt at suggesting which mosquitoes were present would be subject to many assumptions based on other research from other locales. Some of the studies in the literature do not appear to translocate well and even recent, robust, works such as Townroe and Callaghan (2014) suggest a more apparent spatial separation between *Culex pipiens s.s.* and *Culex torrentium* than the level of sympatry observed in this thesis. Whilst mosquitoes were recovered from all sites surveyed, and occasionally in high numbers, particularly as juveniles, generally collection rates were fairly low across most sites. Therefore, despite the confirmed vectoral capacity of *Culex pipiens* and *Culex torrentium* the opportunity for transmission of disease to and between humans is probably limited. However, through climate change, increased urbanisation and the increased use of rainwater capture and storage devices (water butts) there is the potential to see increased mosquito density, due to a longer breeding season, greater urban heat island effect and more breeding sites (Townroe and Callaghan 2014).

A new procedure for the discrimination between female *Culex pipiens s.s.* and *Culex torrentium* was also developed, tested and optimised. Part of the process involved the assessment of aspects of the wing morphometric approach which was so successful in Germany for Börstler *et al.*, (2014), however these could not be validated for the locally caught specimens; indeed, within the first few wings measured, the wing vein r2+3/r3 ratio was seen to overlap, where they found no such overlap across all their specimens, effectively ruling out bivariate approaches to this identification. Only through significant optimisation and the development of a new coded solution based on the R software (R Core Team 2015) packages 'adegenet' (Jombart 2008) and 'Geomorph' (Adams and Otárola-Castillo 2013) was

it possible to create an approach which discriminated multivariate wing data to give results with high accuracy (> 84 %) and complete precision (100 %) when tested against novel specimens. The final multivariate approach, using photographs of slide mounted wings, was tested and found to be robust. As this stands currently, the proposed wing morphometric approach is a useful tool for discrimination between female *Culex pipiens s.s.* and *Culex torrentium*, and as the training dataset will expand through use over time, is expected to generate increasingly high accuracy results over time. That the Börstler *et al.* (2014) specimens were so much more separated by the r2+3/r3 wing vein ratio than the specimens collected in North Staffordshire and its environs, is interesting when considered as a descriptive difference between geographically distinct populations; trying to generate causal argument however is very challenging. There is no evidence of hybridisation between *Culex pipiens s.s.* and *Culex torrentium* in the field, so there is no gene flow between these sympatric siblings. However, because of the apparent stability of the unreliability of wing vein ratios for the discrimination between these species in the UK (Service 1968) it is most likely that the genetics of the expressed phenotypes are relatively unchanged suggesting founder effects (Barton and Charlesworth 1984) during range expansion of either species causing an initial difference which remains unaltered.

Comparative behavioural studies between *Culex pipiens s.s.* and *Culex torrentium* showed that there was no significant difference between the species' attraction to candidate flowers dual choice olfactometer assays. Both species showed attraction to flowers of the five species from the Apiaceae family, and although there was some visible behavioural variation between flower species in the data, these variations were found to be non-significant. Assays took place in a purpose designed and built olfactometer and the preference was assessed using a new, novel, single measure, the Expressed Preference Index (EPI) which was designed to take into account the location of all mosquitoes involved in each assay, and is particularly suited to longer behavioural assays where 'activation' should not be considered to be a valid metric. This resulted in a highly conservative preference index which only generates high behavioural values when a large proportion of the mosquitoes in an assay actively express a preference for a variable.

Following chemical analysis of the volatile organic chemicals (VOCs) emitted by the flowers involved in the behavioural assays, candidate blends were tested in the field for efficacy as lures to increase trap effectiveness, and in the laboratory to generate EPI values for *Culex pipiens*. The blends tested appeared to be less attractive than their whole floret counterparts, although the reduction was not found to be significant. In the field, they did not cause a significant increase in mosquito catch rates either. The lack of significant attraction to the synthetic blend in field and laboratory trials was somewhat disappointing. However, when considered within the context of the number of possible candidate VOCs, and the resultant contingency table of permutations available within the tested flora, it would be unlikely to devise a powerful attractant at the first attempt. The development of a lure of this type is a lengthy process requiring significant optimisation of chemical composition, dose and delivery mechanisms, and therefore null results as those seen for the synthetic blends used here are important pieces of information which feed into the design and optimisation process. However, despite the null result found here, the method of testing was determined to be suitable for testing the efficacy of a candidate lure.

Assuming that a strongly attractive floral derived lure for male mosquitoes can be developed through further research, strategies for its application would need to be devised. The swarming ecology and behaviour demonstrated by many of the key vector species, including the major African malaria vectors (Sawadogo *et al.* 2017) and many *Culex* mosquitoes with important vectoral roles (Knab 1906; Reisen *et al.* 1986; Service 1994; Fonseca *et al.* 2004) can potentially be exploited to target these accumulations of males. It has been demonstrated that by targeting male swarms the local mosquito population can be reduced significantly, by as much as 80% (Sawadogo *et al.* 2017). In their study, swarms were targeted by the application of aerosolised insecticides directly by an operator. Sawadogo *et al.* (2017) described their study as the first of its kind and so their approach will need to be repeated by others, and in other locations, to ascertain whether the male focused control measures are truly effective. Even with this caveat however, it does suggest that male-centric control opportunities are currently being underexploited (Diabate and Tripet 2015). Further questions relate to whether the Sawadogo *et al.* (2017) method could be changed to support the use of trapping and killing, or the application of attractive sugar baits, rather than operator delivered insecticides. In such control applications, traps should be baited with the

best combination of attractants that can be brought to bear on the problem, within the control budget. Therefore, in all likelihood the type of olfactory attractant discussed here would be allied with carefully selected trap type (Hoel *et al.* 2009; Lühken *et al.* 2014), colouration (Burkett and Butler 2005) and additional lures which may include auditory stimuli (Diabate and Tripet 2015) to deliver the highest possible lure and kill rate from these swarms.

The data gathered in the preference studies here support the notion of mosquitoes being generalist feeders in relation to sugars (Foster 1995b). However, other studies have demonstrated some strong preferences towards certain flower species (Müller and Schlein 2006; Otienoburu *et al.* 2012; Nikbakhtzadeh *et al.* 2014; Chen and Kearney 2015) and strong aversion (Mng'ong'o *et al.* 2011) from others. A potential limitation of lures based on mosquito's drive to sugar feed when applied to control is that there are often many other sources of sugar available to mosquitoes. Manipulation of the local flora through the removal of highly attractive flower species and the addition of aversive species (Mng'ong'o *et al.* 2011) could be used to develop a push-pull system (Takken and Koenraadt 2013); where the trap represents the most attractive sugar source adjacent to the swarming location. This would lead to maximum efficiency for trapping using these types of lure. Ideally these attractants and strategies would form a part of the arsenal of tools available to integrated vector management programmes. This type of approach is supported by the WHO (2017) who called for a "comprehensive approach to vector control" to deliver the effective reduction of disease in their Global Vector Control 2017-2030 report.

6.1 Conclusion

In view of the evidence presented in this thesis, I am confident to suggest that continued efforts are needed to supplement the mosquito species record in the British Isles; much of the existing data are old, and due to changes in climate, land use and human activity, we must refresh the data to avoid being caught out by changes to distributions. Within the region studied I have collected significant data, at a high spatial resolution, for addition to the area's species record, and developed site-specific, detailed, information about mosquito occurrence at the SSSI and NNR Wybunbury Moss.

Species discrimination between female *Culex pipiens s.s.* and *Culex torrentium* continues to defy simple, univariate, morphological techniques. However, using the tool that I have developed the discrimination can be made with high accuracy using geometric morphometric analysis of the wing venation. Hopefully, through the adoption of this technique by other researchers, extra resolution can be added to *Culex torrentium* distribution data in particular, due to increased confidence in identification for entomologists that do not have access to molecular methods.

Culex pipiens s.s. and *Culex torrentium* showed olfactory mediated preference towards preference VOC emissions from Apiaceae flowers. Synthetic chemical blends did not elicit the same behavioural response, and were found to lack efficacy as baits to augment trapping efforts. However, there is potential using the approach utilised, and the apparatus designed and built herein to develop and optimise such a lure.

Other contributions to the field from this thesis, aside from those specific to the aims, include the programmatical data handling tools developed. Those associated with the extraction and storage of data from high peak number mass spectrometry reports, and their subsequent analysis against a panel of compounds of interest, were particularly powerful. They have the advantage of being easy to reapply to stored data in light of either new MS reports or the addition of new compounds to the target compound panel. This approach is also very easy to repurpose to other studies, requiring only superficial adjustments to the code to use with any comparative chemical analysis.

6.2 Further work

Within the context of the current rapidly changing distributions of vectors and vector-borne disease, and the likely acceleration of this phenomena over time due to climate change, land use change and increased human population and movement, it has never been more important to continue to press forward with efforts to strengthen our understanding of disease vectors. The continuation of mosquito distribution and ecological studies in the region are vital to generate temporally relevant data to bolster the currently poor data provision. As these data become more comprehensive, then confidence in computer modelling of vector ecology becomes incrementally more robust. Similarly, good quality data, allied with modern

analysis methods will help respond to any future vector-borne disease outbreak in the British Isles. Reflecting on the fieldwork conducted in this thesis, a continuation of preimaginal surveying using traditional dipping methods, allied with Mosquito Magnet™ sampling of the adult populations would be most likely to yield the strongest and most comprehensive results, as both approaches collected the broadest range of species compared to other methods that were employed.

With regard to the wing morphometric identification method, additional specimens should be added to the database from across the country to further enhance the robustness of the training dataset. There is also an opportunity to combine the wing morphometric based approach used here with artificial neural network analysis (Lorenz *et al.* 2015) and compare their results. Both approaches have been demonstrated to be powerful tools for discrimination, and so if identifications were based upon agreement between the two approaches then high confidence in those identifications should be possible. It is possible to combine both methods in one software tool, using R software for example, therefore requiring no further data collection. There is also potential to develop a coded software solution to create comprehensive system that can identify to species from samples across multiple genera. This would be quite data intensive and require a large reference data set, but this could be collated should demand require a tool of this type.

Now that a benchmark of olfactory-mediated floral attraction has been established for *Culex pipiens s.s.* and *Culex torrentium*, subsequent efforts to find attractive floral VOCs will be easier and should continue with the addition of new flower species. EPI for new species can quickly be compared to the Apiaceae data facilitating the shortlisting of species for further analysis. These new species should include non-native species and cultivars, as although this thesis was deliberately constrained to species found in the region under study, further work could be unshackled from this restriction. Behavioural assays against synthetic blends should continue, and utilise both individual compounds and subtractive assays (Otienoburu *et al.* 2012; Qiao *et al.* 2012) to determine blend efficacy and help optimise the attraction before testing the blends in field settings. Future studies should be assisted by the use of electroantennographic detection to allow the confirmation of physiological responses to the

offered stimuli, reducing the reliance on the literature, and enabling the detection of new interactions between mosquitoes and VOCs which would be missed otherwise.

7 References

Acuff, V.R. (1976) 'Trap biases influencing mosquito collections.', *Mosquito News*, 36(2), 51–53.

Adams, D.C., Otárola-Castillo, E. (2013) 'Geomorph: an R Package for the Collection and Analysis of Geometric Morphometric Shape Data', *Methods in Ecology and Evolution*, 4(4), 393–399.

Adams, D.C., Rohlf, F.J., Slice, D.E. (2004) 'Geometric morphometrics: Ten years of progress following the "revolution"', *Italian Journal of Zoology*, 71(1), 5–16.

Afify, A., Horlacher, B., Roller, J., Galizia, C.G. (2014) 'Different repellents for *Aedes aegypti* against blood-feeding and oviposition', *PLoS ONE*, 9(7), 1–7.

Ahmed, N., Dawson, M., Smith, C., Wood, E. (2007) *Biology of Disease*, Laboratory investigation, Taylor & Francis Group: Abingdon.

Allan, S.A., Bernier, U.R., Kline, D.L. (2005) 'Evaluation of oviposition substrates and organic infusions on collection of Culex in Florida.', *Journal of the American Mosquito Control Association*, 21(3), 268–273.

Allan, S.A., Day, J.F., Edman, J.D. (1987) 'Visual ecology of biting flies.', *Annual review of entomology*, 32, 297–316.

Alphey, L. (2009) 'Natural and engineered mosquito immunity.', *Journal of biology*, 8(4), 40.

Alphey, L. (2014) 'Genetic control of mosquitoes.', *Annual Review of Entomology*, 59, 205–224.

Alphey, L., McKemey, A., Nimmo, D., Neira Oviedo, M., Lacroix, R., Matzen, K., Beech, C. (2013) 'Genetic control of *Aedes* mosquitoes', *Pathogens and Global Health*, 107(4), 170–179.

Alto, B.W., Muturi, E.J., Lampman, R.L. (2012) 'Effects of nutrition and density in *Culex pipiens*', *Medical and Veterinary Entomology*, 26(4), 396–406.

Andersson, I.H., Jaenson, T.G.T. (1987) 'Nectar feeding by mosquitoes in Sweden, with special reference to *Culex pipiens* and *Cx torrentium*.', *Medical and veterinary entomology*, 1(1), 59–64.

Anon (2017) Hamamatsu NanoZoomer Range [online], available: http://www.hamamatsu.com/eu/en/product/category/5002/5007/index.html [accessed 10 May 2017].

Anonymous (2017) Loynton Moss [online], available: http://www.staffs-wildlife.org.uk/reserves/loynton-moss [accessed 16 Jan 2017].

Anonymous (n.d.) How Hot Spot Analysis (Getis-Ord Gi*) Works [online], available: http://desktop.arcgis.com/en/arcmap/10.3/tools/spatial-statistics-toolbox/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm.

Ashraf, U., Ye, J., Ruan, X., Wan, S., Zhu, B., Cao, S. (2015) 'Usutu virus: An emerging flavivirus in Europe', *Viruses*, 7(1), 219–238.

Babushok, V.I., Linstrom, P.J., Zenkevich, I.G. (2011) 'Retention Indices for Frequently Reported Compounds of Plant Essential Oils', *Journal of Physical and Chemical Reference Data*, 40(4), 043101.

Bahnck, C.M., Fonseca, D.M. (2006) 'Rapid assay to identify the two genetic forms of *Culex (Culex) pipiens* L. (Diptera: Culicidae) and hybrid populations', *The American Journal of Tropical Medicine and Hygiene*, 75(2), 251–5.

Bailey, C.L., Eldridge, B.F., Hayes, D.E., Watts, D.M., Tammariello, R.F., Dalrymple, J.M. (1978) 'Isolation of St. Louis encephalitis virus from overwintering *Culex pipiens* mosquitoes', *Science*, 199(4335), 1346–1349.

Baker, T.C. (1995) 'Electroantennographic and coupled gas chromatographic-electroantennographic responses of the mediterranean fruit fly, *Ceratitis capitata*, to male-produces volatiles and mango odor', *Journal of Chemical Ecology*, 21(11).

Bale, J.S., Hayward, S.A.L. (2010) 'Insect overwintering in a changing climate', *Journal of Experimental Biology*, 213(6), 980–994.

Balenghien, T., Vazeille, M., Grandadam, M., Schaffner, F., Zeller, H., Reiter, P., Sabatier, P., Fouque, F., Bicout, D.J. (2008) 'Vector competence of some French *Culex* and *Aedes* mosquitoes for West Nile virus.', *Vector borne and zoonotic diseases (Larchmont, N.Y.)*, 8(5), 589–595.

Ballantyne, G., Willmer, P. (2012) 'Nectar Theft and Floral Ant-Repellence: A Link between Nectar Volume and Ant-Repellent Traits?', *PLoS ONE*, 7(8), 1–10.

Ballhorn, D.J., Kautz, S., Heil, M. (2013) 'Distance and sex determine host plant choice by herbivorous beetles.', *PloS one*, 8(2), e55602.

Barton, N.H., Charlesworth, B. (1984) 'Genetic revolutions, founder effect, and speciation', *Ann. Rev. Ecol. Syst.*, 15, 133–164.

Batzer, D.P. (1993) 'Technique for surveying larval populations of *Coquillettidia perturbans*.', *Journal of the American Mosquito Control Association*, 9(3), 349–51.

Baum, J.A., Roberts, J.K. (2014) *Progress Towards RNAi-Mediated Insect Pest Management*, 1st ed, Advances in Insect Physiology, Elsevier Ltd.

Becker, N., Jöst, A., Weitzel, T. (2012) 'The *Culex pipiens* complex in Europe', *Journal of the American Mosquito Control Association*, 28(suppl 4), 53–67.

Becker, N., Petrić, D., Zgomba, M., Boase, C., Madon, M., Dahl, C., Kaiser, A. (2010a) *Mosquitoes and Their Control*, 2nd ed, Rangeland Ecology & Management, Springer.

Becker, N., Petrić, D., Zgomba, M., Boase, C., Madon, M., Dahl, C., Kaiser, A. (2010b) 'Chapter 7: Key to Male Mosquitoes', in *Mosquitoes and Their Control*, Springer, 113–134.

Beier, J.C. (2002) 'Vector Incrimination and Entomological Inoculation Rates', in Doolan, D.L., ed., *Malaria Methods and Protocols*, Humana Press, 3–11.

Beier, J.C., Müller, G.C., Gu, W., Arheart, K.L., Schlein, Y. (2012) 'Attractive toxic sugar bait (ATSB) methods decimate populations of *Anopheles* malaria vectors in arid environments regardless of the local availability of favoured sugar-source blossoms', *Malaria Journal*, 11(1), 31.

Benedict, M.Q., Levine, R.S., Hawley, W.A., Lounibos, L.P. (2007) 'Spread of the Tiger: Global Risk of Invasion by the Mosquito *Aedes albopictus*', *Growth (Lakeland)*, 7(1), 76–85.

Bentley, M.D., Day, J.F. (1989) 'Chemical ecology and behavioral aspects of mosquito oviposition', *Annual review of entomology*, 34, 401–421.

Bernier, U.R., Kline, D.L., Barnard, D.R., Schreck, C.E., Yost, R.A. (2000) 'Analysis of Human Skin Emanations by Gas Identification of Volatile Compounds That Are Candidate Attractants for the Yellow Fever Mosquito (*Aedes aegypti*)', *Analytic Chemistry*, 72(4), 747–756.

Bessell, P.R., Robinson, R.A., Golding, N., Searle, K.R., Handel, I.G., Boden, L.A., Purse, B. V., Bronsvoort, B.M. de C. (2014) 'Quantifying the Risk of Introduction of West Nile Virus into Great Britain by Migrating Passerine Birds', *Transboundary and Emerging Diseases*.

Beyaert, I., Wäschke, N., Scholz, A., Varama, M., Reinecke, A., Hilker, M. (2010) 'Relevance of resource-indicating key volatiles and habitat odour for insect orientation', *Animal Behaviour*, 79(5), 1077–1086.

Bhatt, S., Gething, P.W., Brady, O.J., Messina, J.P., Farlow, A.W., Moyes, C.L., Drake, J.M., Brownstein, J.S., Hoen, A.G., Sankoh, O., Myers, M.F., George, D.B., Jaenson, T.G.T. (2013) 'The global distribution and burden of dengue', *Nature*, 496(7446), 504–507.

Bhatt, S., Weiss, D.J., Cameron, E., Bisanzio, D., Mappin, B., Dalrymple, U. (2016) 'The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015', *Nature*, 526(7572), 207–211.

Bidlingmayer, W.L. (1967) 'A comparison of trapping methods for adult mosquitoes: species response and environmental influence', *Journal of Medical Entomology*, 4(2), 200–220.

Bidlingmayer, W.L. (1975) 'Mosquito Flight Paths in Relation to Environment - Effect of Vertical and Horizontal Visual Barriers', *Annals of the Entomological Society of America*, 68, 51–57.

Biessmann, H., Andronopoulou, E., Biessmann, M.R., Douris, V., Dimitratos, S.D., Eliopoulos,

E., Guerin, P.M., Iatrou, K., Justice, R.W., Kröber, T., Marinotti, O., Tsitoura, P., Woods, D.F., Walter, M.F. (2010) 'The *Anopheles gambiae* odorant binding protein 1 (AgamOBP1) mediates indole recognition in the antennae of female mosquitoes', *PLoS ONE*, 5(3).

Biogents (2011) 'Instruction Manual for the Biogents Sentinel', 1–13.

Birdsall, K., Zimmerman, E., Teeter, K., Gibson, G. (2000) 'Genetic variation for the positioning of wing veins in *Drosophila melanogaster*', *Evolution and Development*, 2(1), 16–24.

Birkett, M.A., Bruce, T.J.A., Martin, J.L., Smart, L.E., Oakley, J., Wadhams, L.J. (2004) 'Responses of female orange wheat blossom midge, *Sitodiplosis mosellana*, to wheat panicle volatiles', *Journal of Chemical Ecology*, 30(7), 1319–1328.

Bjostad, L.B. (1988) 'Insect electroantennogram responses to semiochemicals recorded with an inexpensive personal computer', *Physiological Entomology*, 13(2), 139–145.

Bjostad, L.B., Roelofs, W.L. (1980) 'An inexpensive electronic device for measuring electroantennogram responses to sex pheromone components with a voltmeter', *Physiological entomology*, 5(4), 309–314.

Blagrove, M.S.C., Sherlock, K., Chapman, G.E., Impoinvil, D.E., McCall, P.J., Medlock, J.M., Lycett, G., Solomon, T., Baylis, M. (2016) 'Evaluation of the vector competence of a native UK mosquito *Ochlerotatus detritus* (*Aedes detritus*) for dengue, chikungunya and West Nile viruses', *Parasites & Vectors*, 9(452), 1–9.

Bohbot, J.D., Durand, N.F., Vinyard, B.T., Dickens, J.C. (2013) 'Functional development of the octenol response in Aedes aegypti', *Frontiers in Physiology*, 4(39), 1–8.

Bohbot, J.D., Jones, P.L., Wang, G., Pitts, R.J., Pask, G.M., Zwiebel, L.J. (2011) 'Conservation of Indole Responsive Odorant Receptors in Mosquitoes Reveals an Ancient Olfactory Trait', *Chemical Senses*, 39, 149–160.

Bohnenstengel, S.I., Evans, S., Clark, P.A., Belcher, S.E. (2011) 'Simulations of the London urban heat island', *Quarterly Journal of the Royal Meteorological Society*, 137(659), 1625–1640.

Bonilauri, P., Bellini, R., Calzolari, M., Angelini, R., Venturi, L., Fallacara, F., Cordioli, P., Angelini, P., Venturelli, C., Merialdi, G., Dottori, M. (2008) 'Chikungunya Virus in Aedes albopictus, Italy', *Emerging Infectious Diseases*, 14(5), 852–853.

Borg-Karlson, A.-K., Valterová, I., Nilsson, L.A. (1993) 'Volatile compounds from flowers of six species in the family Apiaceae: Bouquets for different pollinators?', *Phytochemistry*, 35(1), 111–119.

Börstler, J., Jöst, H., Garms, R., Krüger, A., Tannich, E., Becker, N., Schmidt-Chanasit, J., Lühken, R. (2016) 'Host-feeding patterns of mosquito species in Germany', *Parasites & Vectors*, 9(1), 318.

Börstler, J., Lühken, R., Rudolf, M., Steinke, S., Melaun, C., Becker, S., Garms, R., Krüger, A. (2014) 'The use of morphometric wing characters to discriminate female *Culex pipiens* and *Culex torrentium*', *Journal of Vector Ecology*, 39(June), 204–212.

Bowen, M.F. (1990) 'Post-diapause sensory responsiveness in *Culex pipiens*', *Journal of Insect Physiology*, 36(12), 923–929.

Bowen, M.F. (1992a) 'Patterns of sugar feeding in diapausing and nondiapausing *Culex pipiens* (Diptera: Culicidae) females.', *Journal of Medical Entomology*, 29(5), 843–849.

Bowen, M.F. (1992b) 'Terpene-sensitive receptors in female *Culex pipiens* mosquitoes: Electrophysiology and behaviour', *Journal of Insect Physiology*, 38(10), 759–764.

Brisco, K.K., Cornel, A.J., Lee, Y., Mouatcho, J., Braack, L. (2016) 'Comparing efficacy of a sweep net and a dip method for collection of mosquito larvae in large bodies of water in South Africa', *F1000Research*, 5(May), 713.

Bruce, T.J. a, Wadhams, L.J., Woodcock, C.M. (2005) 'Insect host location: a volatile situation.', *Trends in Plant Science*, 10(6), 269–74.

Brugman, V.A., Hernández-Triana, L.M., Prosser, S.W.J., Weland, C., Westcott, D.G., Fooks, A.R., Johnson, N. (2015) 'Molecular species identification, host preference and detection of myxoma virus in the *Anopheles maculipennis* complex (Diptera: Culicidae) in southern England, UK', *Parasites & Vectors*, 8(421).

Buckley, A., Dawson, A., Gould, E.A. (2006) 'Detection of seroconversion to West Nile virus, Usutu virus and Sindbis virus in UK sentinel chickens', *Virology Journal*, 3(1), 1–6.

Buckley, A., Dawson, A., Moss, S.R., Hinsley, S.A., Bellamy, P.E., Gould, E.A. (2003) 'Serological evidence of West Nile virus, Usutu virus and Sindbis virus infection of birds in the UK', *Journal of General Virology*, 84(10), 2807–2817.

Bueno-Marí, R., Jiménez-Peydró, R. (2011) '*Anopheles plumbeus* Stephens, 1828: a neglected malaria vector in Europe', *Malaria Reports*, 1.

Burkett, D. a., Butler, J.F. (2005) 'Laboratory Evaluation of Colored Light As an Attractant for Female *Aedes aegypti, Aedes albopictus, Anopheles quadrimaculatus*, and Culex nigripalpus</i>', *Florida Entomologist*, 88(4), 383–389.

Burla, H., Brito da Cunha, A., Cordeiro, A.R., Dobzhansky, T., Malogolowkin, C., Pavan, C. (1949) 'The *Willistoni* Group of Sibling Species of *Drosophila*', *Evolution*, 3(4), 300–314.

Byers, K.J.R.P., Bradshaw, H.D., Riffell, J.A. (2014) 'Three floral volatiles contribute to differential pollinator attraction in monkeyflowers (*Mimulus*).', *The Journal of experimental biology*, 217(Pt 4), 614–23.

Calistri, P., Giovannini, A., Hubálek, Z., Ionescu, A., Monaco, F., Savini, G., Lelli, R. (2010) 'Epidemiology of West Nile in Europe and in the Mediterranean Basin', *The Open*

Virology Journal, 4, 29–37.

Cameron, M.M., Lorenz, L.M. (2013) *Biological and Environmental Control of Disease Vectors*, CABI: Wallingford.

Campbell, L.P., Luther, C., Peterson, a T. (2015) 'Climate Change Influences on Global Vector Distributions for Dengue and Chikungunya Viruses'.

Van Cann, J., Virgilio, M., Jordaens, K., De Meye, M. (2015) 'Wing morphometrics as a possible tool for the diagnosis of the *Ceratitis fasciventris, C. Anonae, C. rosa* complex (Diptera, Tephritidae)', *ZooKeys*, 2015(540), 489–506.

Canon (2018) Canon CanoScan LiDE 110 - CanoScan Flatbed Scanners - Canon UK [online], available: https://www.canon.co.uk/scanners/flatbed-scanners/canoscan_lide_110/ [accessed 27 Mar 2018].

Carey, A.F., Wang, G., Su, C.-Y., Zwiebel, L.J., Carlson, J.R. (2010) 'Odorant reception in the malaria mosquito *Anopheles gambiae*.', *Nature*, 464(7285), 66–71.

Carrieri, M., Angelini, P., Venturelli, C., Maccagnani, B., Bellini, R. (2011) '*Aedes albopictus* (Diptera: Culicidae) Population Size Survey in the 2007 Chikungunya Outbreak Area in Italy. I. Characterization of Breeding Sites and Evaluation of Sampling Methodologies', *Journal of Medical Entomology*, 48(6), 1214–1225.

Cator, L.J., George, J., Blanford, S., Murdock, C.C., Baker, T.C., Read, A.F., Thomas, M.B. (2013) '"Manipulation" without the parasite: altered feeding behaviour of mosquitoes is not dependent on infection with malaria parasites.', *Proceedings. Biological sciences / The Royal Society*, 280(1763), 20130711.

CDC (2017) West Nile Virus Activity by State - United States | West Nile Virus | CDC [online], available: https://www.cdc.gov/westnile/statsmaps/preliminarymapsdata/activitystatedate.html [accessed 6 Jul 2017].

Cha, D.H., Nojima, S., Hesler, S.P., Zhang, A., Linn, C.E., Roelofs, W.L., Loeb, G.M. (2008) 'Identification and field evaluation of grape shoot volatiles attractive to female grape berry moth (*Paralobesia viteana*)', *Journal of Chemical Ecology*, 34(9), 1180–1189.

Chamberlain, R.W., Sudia, W.D. (1961) 'Mechanism of transmission of viruses by mosquitoes.', *Annual review of entomology*, 6, 371–390.

Champakaew, D., Junkum, A., Chaithong, U., Jitpakdi, A., Riyong, D., Sanghong, R., Intirach, J., Muangmoon, R., Chansang, A., Tuetun, B., Pitasawat, B. (2015) 'Angelica sinensis (Umbelliferae) with proven repellent properties against *Aedes aegypti*, the primary dengue fever vector in Thailand', *Parasitology Research*.

Chandler, P.J. (2017) 'An Update of the 1998 Checklist of Diptera of the British Isles [updated 16 January 2017]', *Dipterists Digest*, (January), 1–430.

Chen, Z., Kearney, C.M. (2015) 'Nectar protein content and attractiveness to *Aedes aegypti* and *Culex pipiens* in plants with nectar/insect associations', *Acta Tropica*, 1–8.

Chénier, J.V.R., Philogène, B.J.R. (1989) 'Field responses of certain forest Coleoptera to conifer monoterpenes and ethanol', *Journal of Chemical Ecology*, 15(6), 1729–1745.

Chesson, J. (1978) 'Measuring preference in selective predation', *Ecology*, 59(2), 211–215.

Christe, R. de O., Wilke, A.B.B., Vidal, P.O., Marrelli, M.T. (2016) 'Wing sexual dimorphism in *Aedes fluviatilis* (Diptera: Culicidae)', *Infection, Genetics and Evolution*, 45, 434–436.

Christensen, T.A., Hildebrand, J.G. (2002) 'Pheromonal and host-odor processing in the insect antennal lobe: how different?', *Current opinion in neurobiology*, 12(4), 393–9.

Cilek, J.E. (1999) 'Evaluation of various substances to increase adult *Stomoxys calcitrans* (Diptera: Muscidae) collections on alsynite cylinder traps in north Florida.', *Journal of medical entomology*, 36(5), 605–9.

Cilek, J.E., Hallmon, C.F. (2008) 'Controlled Release Chamber for Dispensing Aromatic Substances', *Journal of the American Mosquito Control Association*, 24(4), 594–597.

Cilek, J.E., Kline, D.L., Hallmon, C.F. (2003) 'Evaluation of a novel removal trap system to reduce biting midge (Diptera: Ceratopogonidae) populations in Florida backyards.', *Journal of vector ecology : journal of the Society for Vector Ecology*, 28(1), 23–30.

Clarkson, M., Setzkorn, C. (2011) 'The domestic mosquitoes of the Neston area of Cheshire, UK', *European Mosquito Bulletin*, 29, 122–128.

Clements, A. (1999) *Biology of Mosquitoes, Volume 2: Sensory Reception and Behaviour*.

Cole-Parmer (2018) Cole-Parmer Microscope w/Built-In Camera, 4/10/40/100x, 3 Megapixels from Cole-Parmer United Kingdom [online], available: https://www.coleparmer.co.uk/i/cole-parmer-microscope-w-built-in-camera-4-10-40-100x-3-megapixels/4892345 [accessed 27 Mar 2018].

Comstock, J.H., Needham, J.G. (1898) 'The Wings of Insects . Chapter III . The Specialization of Wings by Reduction', *The American Naturalist*, 32(376), 231–257.

Cook, J.I., Majeed, S., Ignell, R., Pickett, J. a., Birkett, M. a, Logan, J.G. (2011) 'Enantiomeric selectivity in behavioural and electrophysiological responses of *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes.', *Bulletin of entomological research*, 101(5), 541–50.

Cooperband, M.F., McElfresh, J.S., Millar, J.G., Cardé, R.T. (2008) 'Attraction of female *Culex quinquefasciatus* Say (Diptera: Culicidae) to odors from chicken feces', *Journal of Insect Physiology*, 54(7), 1184–1192.

Costa, M., Mateus, R.P., Moura, M.O. (2015) 'Constant fluctuating asymmetry but not directional asymmetry along the geographic distribution of *Drosophila antonietae*

(Diptera, Drosophilidae)', *Revista Brasileira de Entomologia*, 59(4), 337–342.

Costello, S.L., Negrón, J.F., Jacobi, W.R. (2008) 'Traps and attractants for wood-boring insects in ponderosa pine stands in the Black Hills, South Dakota.', *Journal of economic entomology*, 101(2), 409–420.

Coy, M.R., Sanscrainte, N.D., Chalaire, K.C., Inberg, A., Maayan, I., Glick, E., Paldi, N., Becnel, J.J. (2012) 'Gene silencing in adult *Aedes aegypti* mosquitoes through oral delivery of double-stranded RNA', *Journal of Applied Entomology*, 136(10), 741–748.

Crans, W.J. (2004) 'A classification system for mosquito life cycles: life cycle types for mosquitoes of the northeastern United States.', *Journal of Vector Ecology*, 29(1), 1–10.

Cranston, P., Ramsdale, C.D., Snow, K.R., White, G. (1987) *Keys to the Adults, Male Hypopygia, Fourth-Instar Larvae and Pupae of the British Mosquitoes (Culicidae) with Notes on Their Ecology and Medical Importance.*, Freshwater Biological Association: Cumbria.

Curtis, C.F., Lines, J.D., Baolin, L., Renz, A. (1990) 'Natural and Synthetic Repellents', in Curtis, C.F., ed., *Appropriate Technology in Vector Control*, CRC Press: Boca Raton, 75–92.

D'Alessandro, M., Brunner, V., von Mérey, G., Turlings, T.C.J. (2009) 'Strong attraction of the parasitoid *Cotesia marginiventris* towards minor volatile compounds of maize', *Journal of Chemical Ecology*, 35(9), 999–1008.

Dahl, C. (1988) 'Taxonomic studies on *Culex pipiens* and *C. torrentium*', in Service, M.W., ed., *Biosystematics of Haematophagous Insects*, Clarendon PRess: Oxford, 149–175.

Danabalan, R. (2010) *Mosquitoes of Southern England and Northern Wales: Identification, Ecology and Host Selection*.

Danabalan, R., Monaghan, M.T., Ponsonby, D.J., Linton, Y.-M. (2014) 'Occurrence and host preferences of *Anopheles maculipennis* group mosquitoes in England and Wales.', *Medical and veterinary entomology*, 28(2), 169–78.

Danabalan, R., Ponsonby, D.J., Linton, Y.-M. (2012) 'A critical assessment of available molecular identification tools for determining the status of *Culex pipiens* s.l. in the United Kingdom', *The American Mosquito Control Association*, 28(4), 68–74.

Davis, E.E., Sokolove, P.G. (1976) 'Lactic acid sensitive receptors on the antennae of the mosquito, *Aedes aegypti*', *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 105(4), 43–54.

Debat, V., Bégin, M., Legout, H., David, J.R. (2003) 'Allometric and nonallometric components of *Drosophila* wing shape respond differently to developmental temperature', *Evolution*, 57(12), 2773–2784.

Deblauwe, I., Sohier, C., Schaffner, F., Rakotoarivony, L.M., Coosemans, M. (2014) 'Implementation of surveillance of invasive mosquitoes in Belgium according to the ECDC

guidelines.', *Parasites & vectors*, 7(1), 201.

Demeestere, K., Dewulf, J., De Witte, B., Van Langenhove, H. (2007) 'Sample preparation for the analysis of volatile organic compounds in air and water matrices', *Journal of Chromatography A*, 1153(1–2), 130–144.

Dennison, N.J., Jupatanakul, N., Dimopoulos, G. (2014) 'The mosquito microbiota influences vector competence for human pathogens', *Current Opinion in Insect Science*, 1(3), 6–13.

Derraik, J.G.B. (2005) 'Mosquitoes breeding in container habitats in urban and peri-urban areas in the Auckland Region, New Zealand', *Entomotropica*, 20(2), 89–93.

Diabaté, A., Dabiré, R.K., Heidenberger, K., Crawford, J., Lamp, W.O., Culler, L.E., Lehmann, T. (2008) 'Evidence for divergent selection between the molecular forms of *Anopheles gambiae*: role of predation', *BMC Evolutionary Biology*, 8(1), 5.

Diabate, A., Tripet, F. (2015) 'Targeting male mosquito mating behaviour for malaria control', *Parasites and Vectors*, 8(1), 1–13.

Ding, Y., Hu, Y., Yu, B.-T., Mo, X.C., Mo, J.C. (2016) 'Laboratory evaluation of differential attraction of *Culex pipiens pallens* to fruit-based sugar baits', *Acta Tropica*, 163, 20–25.

Dobson, M.J. (1998) 'Bitter- sweet solutions for malaria: exploring natural remedies from the past', *Parassitologia*, (40), 69–81.

Dogan, E.B., Rossignol, P.A. (1999) 'An olfactometer for discriminating between attraction, inhibition, and repellency in mosquitoes (Diptera: Culicidae).', *Journal of medical entomology*, 36(6), 788–93.

Van Den Dool, H., Kratz, P.D. (1963) 'A generalization of the retention index system including linear temperature programmed gas-liquid partition chromatography', *Journal of Chromatography*, (3), 463–471.

Dujardin, J.-P. (2016) 'CLIC (http://mome-clic.com/)'.

Dujardin, J.-P. Al, Kaba, D., Henry, A.B. (2010) 'The exchangeability of shape.', *BMC research notes*, 3(1), 266.

Dujardin, J.P. (2011) 'Modern Morphometrics of medically important insects', in Tibayrenc, M., ed., *Genetics and Evolution of Infectious Diseases*, Elsevier: Amsterdam, 473–501.

ECDC (2012) *Guidelines for the Surveillance of Invasive Mosquitoes in Europe*, Euro surveillance : bulletin européen sur les maladies transmissibles = European communicable disease bulletin.

ECDC (2014) *Guidelines for the Surveillance of Native Mosquitoes in Europe*, Stockholm.

El-Sayed, A.M. (2014) The Pherobase: Database of Pheremones and Semiochemicals [online],

available: http://www.pherobase.com.

Engelberth, J., Alborn, H.T., Schmelz, E.A., Tumlinson, J.H. (2004) 'Airborne signals prime plants against insect herbivore attack.', *Proceedings of the National Academy of Sciences of the United States of America*, 101(6), 1781–5.

ESRI (2014) 'ArcGIS'.

Van Essen, P.H.., Kemme, J.A., Ritchie, S.A., Kay, B.H. (1994) 'Differential responses of *Aedes* and *Culex* mosquitoes to octenol or light in combination with carbon dioxide in Queensland, Australia', *Medical and Veterinary Entomology*, 8(1), 63–67.

Ewing, D.A., Cobbold, C.A., Purse, B. V., Nunn, M.A., White, S.M. (2016) 'Modelling the effect of temperature on the seasonal population dynamics of temperate mosquitoes', *Journal of Theoretical Biology*.

Farajollahi, A., Fonseca, D.M., Kramer, L.D., Marm Kilpatrick, A. (2011) '"Bird biting" mosquitoes and human disease: a review of the role of *Culex pipiens* complex mosquitoes in epidemiology.', *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, 11(7), 1577–85.

Fay, R.W., Eliason, D.A. (1966) 'A Preferred Oviposition Site as a Surveillance Method for *Aedes aegypti*', *Mosquito News*, 26(4), 531–535.

Fedorova, M. V., Shaikevich, E. V (2007) 'Morphological and molecular-genetic distinctions between adult mosquitoes *Culex torrentium* Martini and *C. pipiens* Linnaeus (Diptera, Culicidae) from Moscow Province', *Entomological Review*, 87(2), 127–135.

Fikrig, K., Johnson, B.J., Fish, D., Ritchie, S.A. (2017) 'Assessment of synthetic floral-based attractants and sugar baits to capture male and female *Aedes aegypti* (Diptera: Culicidae)', *Parasites & Vectors*, 1–9.

Fillinger, U., Sombroek, H., Majambere, S., van Loon, E., Takken, W., Lindsay, S.W. (2009) 'Identifying the most productive breeding sites for malaria mosquitoes in The Gambia.', *Malaria journal*, 8, 62.

Fischer-Scientific (2018) Fisher Scientific - Laboratory Supplies, Equipments, Chemical Manufacturing, MSDS Sheets [online], available: https://www.fishersci.co.uk/gb/en/home.html [accessed 27 Mar 2018].

Fonseca, D.M., Keyghobadi, N., Malcolm, C.A., Mehmet, C., Schaffer, F., Mogi, M., Fleischer, R.C., Wilkerson, R.C. (2004) 'Emerging Vectors in the *Culex pipiens* Complex.', *Science*, 303, 1535–1538.

Fonseca, D.M., Smith, J.L., Kim, H.C., Mogi, M. (2009) 'Population genetics of the mosquito *Culex pipiens pallens* reveals sex-linked asymmetric introgression by *Culex quinquefasciatus*', *Infection, Genetics and Evolution*, 9(6), 1197–1203.

Foster, W.A. (1995a) 'Mosquito sugar feeding and reproductive energetics', *Annual Review of Entomology*, 40, 443–474.

Foster, W.A. (1995b) 'Mosquito sugar feeding and reproductive energetics.', *Annual Review of Entomology*, 40, 443–474.

Francoy, T.M., Silva, R.A.., Menezes, C., Impera riz-Fonseca, V.L. (2009) 'Gender identification of five genera of stingless bees (Apidae , *Meliponini*) based on wing morphology', *Genetics and Molecular Research*, 8(1), 207–214.

Francoy, T.M., Wittmann, D., Drauschke, M., Muller, S., Steinhage, V., Bezerra-Laure, M.A.F., Jong, D. De, Gonçalves, L.S. (2008) 'Identification of Africanized honey bees through wing morphometrics : two fast and efficient procedures', *Science*, 39(5), 488–494.

Fürstenau, B., Rosell, G., Guerrero, A., Quero, C. (2012) 'Electrophysiological and behavioral responses of the black-banded oak borer, *Coroebus florentinus*, to conspecific and host-plant volatiles', *Journal of Chemical Ecology*, 38(4), 378–388.

Gadawski, R.M., Smith, S.M. (1992) 'Nectar sources and age structure in a population of *Aedes provocans* (Diptera: Culicidae)', *J Med Entomol*, 29(5), 879–886.

Gaffigan, T., Pecor, J.E. (1997) WRBU: Collecting, Rearing, Mounting and Shipping Mosquitoes [online], available: http://www.wrbu.org/about/techniques.html [accessed 20 Jul 2014].

Garcia-bellido, A., Celis, J.F. De (1992) 'Developmental genetics of the venation pattern of *Drosphila*', *Annual review of genetics*, 26(0), 277–304.

García Rojo, M., Bueno García, G., Carbajo Vicente, M. (2003) 'Critical comparison of 31 commercially available digital slide systems in pathology', *Techniques in Pathology*, 14(4), 285–305.

Gary, R.E., Foster, W.A. (2006) 'Diel timing and frequency of sugar feeding in the mosquito *Anopheles gambiae*, depending on sex, gonotrophic state and resource availability', *Medical and Veterinary Entomology*, 20(3), 308–316.

Gavrielides, M.A., Conway, C., O'Flaherty, N., Gallas, B.D., Hewitt, S.M. (2014) 'Observer performance in the use of digital and optical microscopy for the interpretation of tissue-based biomarkers', *Analytical Cellular Pathology*, 2014(2), 157308.

Geier, M., Boeckh, J. (1999) 'A new Y-tube olfactometer for mosquitoes to measure the attractiveness of host odours', *Entomologia Experimentalis et Applicata*, 92, 9–19.

Geier, M., Bosch, O., Boeckh, J. (1999) 'Influence of odour plume structure on upwind flight of mosquitoes towards hosts', *The Journal of experimental biology*, 202 (Pt 12, 1639–48.

Gerberg, E.J. (1970) *Manual for Mosquito Rearing and Experimental Techniques.*, American Mosquito Control Association: Selma.

Getis, A., Ord, J.K. (1992) 'The Analysis of Spatial Association by use of distance statistics', *Geographical Analysis*, 24(3), 189–206.

Ghaninia, M., Hansson, B.S., Ignell, R. (2007) 'The antennal lobe of the African malaria mosquito, *Anopheles gambiae* - innervation and three-dimensional reconstruction', *Arthropod Structure and Development*, 36(1), 23–39.

Ghosh, S.K.B., Hunter, W.B., Park, A.L., Gundersen-Rindal, D.E. (2017) 'Double strand RNA delivery system for plant-sap-feeding insects', *Plos One*, 12(2), e0171861.

Gibson, G., Torr, S.J. (1999) 'Visual and olfactory responses of haematophagous Diptera to host stimuli', *Medical and Veterinary Entomology*, 13(1), 2–23.

Golding, N. (2013) *Mapping and Understanding the Distributions of Potential Vector Mosquitoes in the UK : New Methods and Applications Mapping and Understanding the Distributions of Potential Vector Mosquitoes in the UK : New Methods and Applications*.

Golding, N., Nunn, M.A., Medlock, J.M., Purse, B. V., Vaux, A.G.C., Schäfer, S.M. (2012) 'West Nile virus vector *Culex modestus* established in southern England.', *Parasites & vectors*, 5(1), 32.

Golding, N., Nunn, M.A., Purse, B. V. (2015) 'Identifying biotic interactions which drive the spatial distribution of a mosquito community', *Parasites & Vectors*, 8(367).

Gouagna, L.-C., Poueme, R.S., Dabiré, K.R., Ouédraogo, J.-B., Fontenille, D., Simard, F. (2010) 'Patterns of sugar feeding and host plant preferences in adult males of *An. gambiae* (Diptera: Culicidae).', *Journal of Vector Ecology : Journal of the Society for Vector Ecology*, 35(2), 267–76.

Gouck, H.K., Schreck, C.E. (1965) 'An Olfactometer for Use in the Study of Mosquito Attractants', *Journal of Economic Entomology*, 58(3), 589–590.

Gouinguené, S., Pickett, J. a., Wadhams, L.J., Birkett, M. a., Turlings, T.C.J. (2005) 'Antennal Electrophysiological Responses of Three Parasitic Wasps to Caterpillar-Induced Volatiles from Maize (*Zea mays mays*), Cotton (*Gossypium herbaceum*), and Cowpea (*Vigna unguiculata*)', *Journal of Chemical Ecology*, 31(5), 1023–1038.

Grimstad, P., DeFoliart, G. (1974) 'Nectar sources of Wisconsin mosquitoes', *Journal of Medical Entomology*, 11(3), 331–341.

Grottola, A., Marcacci, M., Tagliazucchi, S., Gennari, W., Di Gennaro, A., Orsini, M., Monaco, F., Marchegiano, P., Marini, V., Meacci, M., Rumpianesi, F., Lorusso, A., Pecorari, M., Savini, G. (2017) 'Usutu virus infections in humans: a retrospective analysis in the municipality of Modena, Italy', *Clinical Microbiology and Infection*, 23(1), 33–37.

Hall, M.J.R., MacLeod, N., Wardhana, A.H.H. (2014) 'Use of wing morphometrics to identify populations of the Old World screwworm fly, *Chrysomya bezziana* (Diptera: Calliphoridae): A preliminary study of the utility of museum specimens', *Acta Tropica*,

138, S49–S55.

Hamer, G.L., Kelly, P.H., Focks, D. a, Goldberg, T.L., Walkers, E.D. (2011) 'Evaluation of a novel emergence trap to study *Culex* mosquitoes in urban catch basins.', *Journal of American Mosquito Control Association*, 27(2), 142–7.

Hamer, G.L., Kitron, U.D., Brawn, J.D., Loss, S.R., Ruiz, M.O., Goldberg, T.L., Walker, E.D. (2008) '*Culex pipiens* (Diptera: Culicidae): a bridge vector of West Nile virus to humans.', *Journal of medical entomology*, 45(1), 125–8.

Hancock, R.G., Foster, W.A. (1993) 'Effects of preblood-meal sugar on sugar seeking and upwind flight by gravid and parous Aedes aegypti (Diptera: Culicidae).', *Journal of Medical Entomology*, 30(2), 353–9.

Van Handel, E. (1972) 'The Detection of Nectar in Mosquitoes', *Mosquito News*, 32, 458.

Van Handel, E. (1984) 'Metabolism of Nutrients in the adult mosquito', *Mosquito News*, 44(4), 573–579.

Hanson, B.A. (2014) 'ChemoSpec : An R Package for Chemometric Analysis of Spectroscopic Data and Chromatograms (Package Version 4.4.17)', 1–41.

Hao, H., Sun, J., Dai, J. (2013) 'Dose-dependent behavioral response of the mosquito *Aedes albopictus* to floral odorous compounds', *Journal of Insect Science*, 13(127), 1–8.

Hapairai, L.K., Joseph, H., Sang, M. a C., Melrose, W., Ritchie, S.A., Burkot, T.R., Sinkins, S.P., Bossin, H.C. (2013) 'Field evaluation of selected traps and lures for monitoring the filarial and arbovirus vector, *Aedes polynesiensis* (Diptera: Culicidae), in French Polynesia.', *Journal of Medical Entomology*, 50(4), 731–9.

Haramis, L.D., Foster, W.A. (1990) 'Gonotrophic cycle duration, population age structure, and onset of sugar feeding and insemination of *Aedes triseriatus* (Diptera: Culicidae)', *Journal of Medical Entomology*, 27(4), 421–428.

Harbach, R.E. (2007) 'The Culicidae (Diptera): a review of taxonomy, classification and phylogeny', *Zootaxa*, 638, 591–638.

Harbach, R.E. (2012) 'Culex pipiens: species versus species complex taxonomic history and perspective.', *Journal of the American Mosquito Control Association*, 28(4 Suppl), 10–23.

Harbach, R.E., Howard, T.M. (2007) 'Index of currently recognized mosquito species (Diptera: Culicidae)', *European Mosquito Bulletin*, 23(August), 1–66.

Harrup, L.E., Logan, J.G., Cook, J.I., Golding, N., Birkett, M.A., Pickett, J.A., Sanders, C., Barber, J., Rogers, D.J., Mellor, P.S., Purse, B. V., Carpenter, S. (2012) 'Collection of *Culicoides* (Diptera: Ceratopogonidae) using CO2 and enantiomers of 1-octen-3-ol in the United Kingdom.', *Journal of Medical Entomology*, 49(1), 112–21.

Hartberg, W.K. (1971) 'Observations on the mating behavior of *Aedes aegypti* in nature', *Bulletin of the World Health Organization*, 45(6), 847–850.

Hassell, M.P., Southwood, T.R.E. (1978) 'Foraging strategies of insects', *Annual Review of Ecology and Systematics*, 9(1978), 75–98.

Hawley, W.A., Pumpuni, C.B., Brady, R.H., Craig, G.B. (1989) 'Overwintering Survival of *Aedes albopictus* (Diptera: Culicidae) Eggs in Indiana', *Journal of Medical Entomology*, 26(2), 122–129.

Hayes, R.O., Doane, O.W., Sakolski, G., Berrick, S. (1993) 'Evaluation of attractants in traps for greenhead fly (Diptera:Tabanidae) collections on a cape cod, Massachusetts, salt marsh', *Journal of American Mosquito Association*, 9(4), 436–440.

Hazard, E.I., Mayer, M.S., Savage, K.E. (1967) 'Attraction and oviposition stimulation of gravid female mosquitoes by bacteria isolated from hay infusions', *Mosquito News*, 27(2), 133–136.

Hebeish, A., Fouda, M.M.G., Hamdy, I.A., EL-Sawy, S.M., Abdel-Mohdy, F.A. (2008) 'Preparation of durable insect repellent cotton fabric: Limonene as insecticide', *Carbohydrate Polymers*, 74(2), 268–273.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R. (2003) 'Biological identifications through DNA barcodes', *Proceedings of the Royal Society B: Biological Sciences*, 270(1512), 313–321.

Hebert, P.D.N., Ratnasingham, S., de Waard, J.R. (2003) 'Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species', *Proceedings of the Royal Society B: Biological Sciences*, 270(Suppl_1), S96–S99.

Heil, M., Silva Bueno, J.C. (2007) 'Within-plant signaling by volatiles leads to induction and priming of an indirect plant defense in nature.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), 5467–5472.

Hernández-Triana, L.M., Brugman, V.A., Prosser, S.W.J., Weland, C., Nikolova, N., Thorne, L., De Marco, M.F., Fooks, A.R., Johnson, N. (2017) 'Molecular approaches for blood meal analysis and species identification of mosquitoes (Insecta: Diptera: Culicidae) in rural locations in southern England, United Kingdom', *Zootaxa*, 4250(1), 67–76.

Herrera-Varela, M., Lindh, J.M., Lindsay, S.W., Fillinger, U. (2014) 'Habitat discrimination by gravid *Anopheles gambiae* sensu lato - a push-pull system.', *Malaria journal*, 13(1), 133.

Herrmann, A. (2010) *The Chemistry and Biology of Volatiles*, John Wiley & Sons, Ltd: Chichester, UK.

Hesson, J.C., Ignell, R., Hill, S.R., Östman, Ö., Lundström, J.O. (2015) 'Trapping biases of *Culex torrentium* and *Culex pipiens* revealed by comparison of captures in CDC traps, ovitraps, and gravid traps', *Journal of Vector Ecology*, 40(1), 158–163.

Hesson, J.C., Rettich, F., Merdić, E., Vignjević, G., Ostman, O., Schäfer, M., Schaffner, F., Foussadier, R., Besnard, G., Medlock, J.M., Scholte, E.-J., Lundström, J.O. (2014) 'The arbovirus vector *Culex torrentium* is more prevalent than *Culex pipiens* in northern and central Europe.', *Medical and veterinary entomology*, 28(2), 179–86.

Hesson, J.C., Verner-Carlsson, J., Larsson, A., Ahmed, R., Lundkvist, Å., Lundström, J.O. (2015) '*Culex torrentium* Mosquito Role as Major Enzootic Vector Defined by Rate of Sindbis Virus Infection, Sweden, 2009', *Emerging infectious diseases*, 21(5).

Hickling, R., Roy, D.B., Hill, J.K., Thomas, C.D. (2005) 'A northward shift of range margins in British Odonata', *Global Change Biology*, 11(3), 502–506.

Hock, J.W. (2016) CDC Miniature Light Trap [online], available: http://johnwhock.com/products/mosquito-sandfly-traps/cdc-miniature-light-trap/ [accessed 15 Jan 2016].

Hoel, D.F., Allan, S.A., Kline, D.L. (2009) 'Evaluation of Six Mosquito Traps for Collection of *Aedes albopictus* and Associated Mosquito Species in a Suburban Setting in North Central Florida', *Journal of the American Mosquito Control Association*, 25(1), 47–57.

Hopkins, G.W., Freckleton, R.P. (2002) 'Declines in the numbers of amateur and professional taxonomists: Implications for conservation', *Animal Conservation*, 5(3), 245–249.

Hosken, D.J., Blanckenhorn, W.U., Ward, P.I. (2000) 'Developmental stability in yellow dung flies (*Scathophaga stercoraria*): fluctuating asymmetry , heterozygosity and environmental stress', *Science*, 13, 919–926.

Hubálek, Z. (2008) 'Mosquito-borne viruses in Europe', *Parasitology Research*, 103(SUPPL. 1), S29-43.

Hudson, R. (2000) 'Odor and odorant: a terminogical clarification', *Chemical Senses*, 25, 693.

Huettel, M.D. (1976) 'Monitoring the quality of laboratory-reared insects: a biological and behavioural perspective', *Environmental Entomology*, 5(5), 807–814.

Hughes, D.T., Pelletier, J., Luetje, C.W., Leal, W.S. (2010) 'Odorant Receptor from the Southern House Mosquito Narrowly Tuned to the Oviposition Attractant Skatole', *Journal of Chemical Ecology*, 36, 797–800.

Hughes, D.T., Wang, G., Zwiebel, L.J., Luetje, C.W. (2014) 'A Determinant of Odorant Specificity Is Located at the Extracellular Loop 2-Transmembrane Domain 4 Interface of an Anopheles gambiae Odorant Receptor Subunit', *Chemical Senses*, 39(9), 761–769.

Hutchinson, R.A. (2004) *Mosquito Borne Diseases in England : Past, Present and Future Risks, with Special Reference to Malaria in the Kent Marshes*.

Hutchinson, R.A., West, P.A., Lindsay, S.W. (2007) 'Suitability of two carbon dioxide-baited traps for mosquito surveillance in the United Kingdom.', *Bulletin of Entomological*

Research, 97(6), 591–7.

Ishida, Y., Chen, A., Tsuruda, J., Cornel, A., Debboun, M., Leal, W.S. (2004) 'Intriguing olfactory proteins from the yellow fever mosquito, *Aedes aegypti*', *Naturwissenschaften*, 91(9), 426–431.

Ishida, Y., Leal, W.S. (2005) 'Rapid inactivation of a moth pheromone.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 14075–14079.

Ishida, Y., Leal, W.S. (2008) 'Chiral discrimination of the Japanese beetle sex pheromone and a behavioral antagonist by a pheromone-degrading enzyme', *Proc Natl Acad Sci U S A*, 105(26), 9076–9080.

Iturbe-ormaetxe, I., Walker, T., Neill, S.L.O. (2011) 'Wolbachia and the biological control of mosquito borne disease', *Nature Publishing Group*, 12(6), 508–518.

Ivlev, V.S. (1964) 'Experimental ecology of the feeding of fishes'.

Jaenson, T.G.T. (1987) 'Overwintering of *Culex* mosquitoes and their potential as reservoirs of human pathogens', *Medical and Veterinary Entomology*, 1, 151–156.

Jaenson, T.G.T. (1988) 'Diel activity patterns of blood-seeking anthropophilic mosquitoes in central Sweden', *Medical and Veterinary Entomology*, 2(2), 177–187.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013) *Springer Texts in Statistics An Introduction to Statistical Learning*, Springer: New York.

Jhumur, U.S., Dötterl, S., Jürgens, A. (2006) 'Naive and conditioned responses of *Culex pipiens pipiens* biotype *molestus* (Diptera: Culicidae) to flower odors.', *Journal of Medical Entomology*, 43(6), 1164–70.

Jhumur, U.S., Dötterl, S., Jürgens, A. (2007) 'Electrophysiological and behavioural responses of mosquitoes to volatiles of *Silene otites* (Caryophyllaceae)', *Arthropod-Plant Interactions*, 1(4), 245–254.

Jhumur, U.S., Dötterl, S., Jürgens, A. (2008) 'Floral Odors of *Silene otites*: Their Variability and Attractiveness to Mosquitoes', *Journal of Chemical Ecology*, 34(1), 14–25.

Jombart, T. (2008) 'Adegenet: a R package for the multivariate analysis of genetic markers', *Bioinformatics*, 24(11), 1403–1405.

Jombart, T. (2012) 'A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 1 . 3-4', *Rvignette*, 1–37.

Jombart, T., Devillard, S. (2010) 'Discriminant analysis of principal components: a new method for the analysis of genetically structured populations', *BMC Genetics*, 11(94).

Junnila, A., Müller, G.C., Schlein, Y. (2010) 'Species identification of plant tissues from the gut

of *An. sergentii* by DNA analysis', *Acta Tropica*, 115(3), 227–233.

Jupp, P.G. (1979) ' *Culex (Culex) pipiens* and *Culex (Cx.) torrentium* (Diptera: Culicidae) in England: Notes on their Taxonomy and Biology', *Mosquito Systematics*, 11(2), 121–126.

Kampen, H., Medlock, J.M., Vaux, A.G.C., Koenraadt, C.J.M., van Vliet, A.J.H., Bartumeus, F., Oltra, A., Sousa, C. a, Chouin, S., Werner, D. (2015) 'Approaches to passive mosquito surveillance in the EU', *Parasites & Vectors*, 8(1), 9.

Kaur, J.S., Lai, Y.L., Giger, a D. (2003) 'Learning and memory in the mosquito *Aedes aegypti* shown by conditioning against oviposition deterrence.', *Medical and veterinary entomology*, 17(4), 457–60.

Keele Climate [online] (2017) *Met Office*, available: http://www.metoffice.gov.uk/public/weather/climate/gcqmkqe11 [accessed 1 Feb 2017].

Kelly, D.W. (2001) 'Why are some people bitten more than others?', *Trends in Parasitology*, 17(12), 578–581.

Kenney, J.L., Romo, H., Duggal, N.K., Tzeng, W., Kristen, L., Brault, A.C., Savage, H.M. (2017) 'Transmission Incompetence of *Culex quinquefasciatus* and *Culex pipiens pipiens* from North America for Zika Virus', *American Journal of Tropical Medicine and Hygiene*, 96(5), 1235–1240.

Kessler, A., Baldwin, I.T. (2000) 'Defensive Function of Herbivore-Induced Plant Volatile Emissions in Nature', *Plant Cell Physiol. Proc. Natl. Acad. Sci. U.S.A. Science J. Zurzycki K. Nozue et al. Proc. Natl. Acad. Sci. U.S.A. Plant Cell*, 41(12), 415–50.

Kessler, A., Baldwin, I.T. (2002) 'Plant Responses to Insect Herbivory: The Emerging Molecular Analysis', *Annual Review of Plant Biology*, 53(1), 299–328.

Khatchikian, C.E., Dennehy, J.J., Vitek, C.J., Livdahl, T.P. (2010) 'Environmental effects on bet hedging in *Aedes* mosquito egg hatch', *Evolutionary Ecology*, 24(5), 1159–1169.

Khrabrova, N. V, Bukhanskaya, E.D., Sibataev, A.K., Volkova, T. V (2009) 'The distribution of strains of endosymbiotic bacteria *Wolbachia pipientis* in natural populations of *Culex pipiens* mosquitoes (Diptera: Culicidae).', *European Mosquito Bulletin*, 27(27), 18–22.

Kiarie-Makara, M.W., Ngumbi, P.M., Lee, D.-K. (2016) 'Effects of Temperature on the Growth and Development of *Culex pipiens* Complex Mosquitoes (Diptera: Culicidae)', *IOSR Journal of Pharmacy and Biological Sciences Ver. II*, 10(January), 2319–7676.

Kiflawi, M., Blaustein, L., Mangel, M. (2003) 'Oviposition habitat selection by the mosquito *Culiseta longiareolata* in response to risk of predation and conspecific larval density', *Ecological Entomology*, 28(2), 168–173.

Kilpatrick, A.M., Kramer, L.D., Jones, M.J., Marra, P.P., Daszak, P. (2006) 'West Nile virus

epidemics in North America are driven by shifts in mosquito feeding behavior.', *PLoS biology*, 4(4), 0606–0610.

Kline, D.L. (1994) 'Olfactory attractants for mosquito surveillance and control: 1-Octen-3-Ol', *Journal of the American Mosquito Control Association*, 10(2), 280–287.

Klowden, M.J. (1999) 'The check is in the male: male mosquitoes affect female physiology and behavior.', *Journal of the American Mosquito Control Association*, 15(2), 213–220.

Klowden, M.J. (2006) 'Switchover to the mated state by spermathecal activation in female *Anopheles gambiae* mosquitoes', *Journal of Insect Physiology*, 52(7), 679–684.

Klowden, M.J., Lea, A.O. (1978) 'Blood mean size as a factor affecting continued host-seeking by *Aedes aegypti* (L.)', *American Journal of Tropical Medicine and Hygiene*, 27, 827–831.

Knab, F. (1906) 'The swarming of *Culex pipiens*', *Psyche: A Journal of Entomology*, 123–134.

Knudsen, J.T., Knudsen, J.T., Eriksson, R., Eriksson, R., Gershenzon, J., Gershenzon, J., Ståhl, B., Ståhl, B. (2006) 'Diversity and Distribution of Floral Scent', *The Botanical Review*, 72(1), 1–120.

Knudsen, J.T., Tollsten, L., Bergstrom, L.G. (1993) 'Floral Scents-a checklist of volatile compounds by head-space techniques', *Phytochemistry*, 33(2), 253–280.

Kolar, C.S., Lodge, D.M. (2001) 'Progress in invasion biology : predicting invaders', *Trends in Ecology & Evolution*, 16(4), 199–204.

Krebs, B.L., Anderson, T.K., Goldberg, T.L., Hamer, G.L., Kitron, U.D., Newman, C.M., Ruiz, M.O., Walker, E.D., Brawn, J.D., Krebs, B.L. (2014) 'Host group formation decreases exposure to vector-borne disease : a field experiment in a " hotspot " of West Nile virus transmission', *Proceedings of the Royal Society B*, 281.

Lalkhen, A.G., McCluskey, A. (2008) 'Clinical tests: Sensitivity and specificity', *Continuing Education in Anaesthesia, Critical Care and Pain*, 8(6), 221–223.

Lang, W.D. (1920) *A Handbook of British Mosquitoes.*, A Handbook of British Mosquitoes., British Museum (Natural History): London.

Larsson, M.C., Stensmyr, M.C., Bice, S.B., Hansson, B.S. (2003) 'Attractiveness of fruit and flower odorants detected by olfactory receptor neurons in the fruit chafer *Pachnoda marginata*.', *Journal of Chemical Ecology*, 29(5), 1253–68.

Laughlin, J.D., Ha, T.S., Jones, D.N.M., Smith, D.P. (2008) 'Activation of Pheromone-Sensitive Neurons Is Mediated by Conformational Activation of Pheromone-Binding Protein', *Cell*, 133(7), 1255–1265.

Laurence, B.R., Pickett, J.A. (1982) 'erythro-6-Acetoxy-5-hexadecanolide, the major component of a mosquitoe oviposition attractant pheromone', *Journal of the Chemical*

Society, Chemical Communications, (1), 59–60.

Laven, H. (1967) 'Eradication of *Culex pipiens fatigans* through cytoplasmic incompatibility', *Nature*, 216, 383–384.

Lazar, I. (2010) 'Gel Analyzer 2010: Freeware 1D gel electrophoresis image analysis software.'

Leal, W.S. (2012) 'An Acute Olfactory System Is Essential for Reproduction—the Raison d'être for Adult Insects 1', *Proceedings of the american philosophical society*, 156(3), 295–302.

Leal, W.S. (2013) 'Odorant Reception in Insects: Roles of Receptors, Binding Proteins, and Degrading Enzymes', *Annual Review of Entomology*, 58(1), 373–391.

Leal, W.S., Choo, Y., Xu, P., da Silva, C.S.B., Ueira-Vieira, C. (2013) 'Differential expression of olfactory genes in the southern house mosquito and insights into unique odorant receptor gene isoforms.', *Proceedings of the National Academy of Sciences of the United States of America*, 110(46), 18704–18709.

Lee, J.C., Heimpel, G.E. (2008) 'Effect of floral nectar, water, and feeding frequency on *Cotesia glomerata* longevity', *BioControl*, 53(2), 289–294.

Lees, R.S., Knols, B.G.J., Bellini, R., Benedict, M.Q., Bheecarry, A., Bossin, H.C., Chadee, D.D., Charlwood, J., Dabiré, R.K., Djogbenou, L., Egyir-Yawson, A., Gato, R., Gouagna, L.-C., Hassan, M.M., Khan, S.A., Koekemoer, L.L., Lemperiere, G., Manoukis, N.C., Mozuraitis, R., Pitts, R.J., Simard, F., Gilles, J.R.L. (2014) 'Review: Improving our knowledge of male mosquito biology in relation to genetic control programmes', *Acta Tropica*, 132(1), S2–S11.

Leggewie, M., Badusche, M., Rudolf, M., Jansen, S., Jessica, B., Krumkamp, R., Huber, K., Kr, A., Tannich, E., Becker, S.C. (2016) '*Culex pipiens* and *Culex torrentium* populations from Central Europe are susceptible to West Nile virus infection', *One Health*, 2, 88–94.

Lenhart, A.E., Walle, M., Cedillo, H., Kroeger, A. (2005) 'Building a better ovitrap for detecting *Aedes aegypti* oviposition', *Acta Tropica*, 96(1), 56–59.

Lens, L., Dongen, S. v., Wilder, C.M., Brooks, T.M., Matthysen, E. (1999) 'Fluctuating asymmetry increases with habitat disturbance in seven bird species of a fragmented afrotropical forest', *Proceedings of the Royal Society B: Biological Sciences*, 266(1425), 1241–1246.

Leonard, A.S., Masek, P. (2014) 'Multisensory integration of colors and scents: Insights from bees and flowers', *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 200(6), 463–474.

Lindsay, S.W., Hole, D.G., Hutchinson, R.A., Richards, S.A., Willis, S.G. (2010) 'Assessing the future threat from vivax malaria in the United Kingdom using two markedly different modelling approaches.', *Malaria Journal*, 9, 70.

Liu, C., Pitts, R.J., Bohbot, J.D., Jones, P.L., Wang, G., Laurence, J. (2010) 'Distinct Olfactory Signaling Mechanisms in the Malaria Vector Mosquito *Anopheles gambiae*', *PLoS Biology*, 8(8).

Logan, J.G. (2008) 'Why Do Mosquitoes "Choose" to Bite Some People More Than Others?', *Outlooks on Pest Management*, 19(6), 280–283.

Logan, J.G., Birkett, M.A. (2007) 'Semiochemicals for biting fly control: their identification and exploitation', *Pest Management Science*, 63, 647–657.

Logan, J.G., Birkett, M.A., Clark, S.J., Powers, S., Seal, N.J., Wadhams, L.J., Mordue, A.J., Pickett, J.A. (2008) 'Identification of human-derived volatile chemicals that interfere with attraction of *Aedes aegypti* mosquitoes', *Journal of Chemical Ecology*, 34(3), 308–322.

Logan, J.G., Pickett, J.A., Cameron, M.M. (2013) 'Chapter 6. Vector control using semiochemicals', in Cameron, M.M. and Lorenz, L.M., eds., *Biological and Environmental Control of Disease Vectors*, CABI, 95–116.

Lohman, D.., Zangerl, A.R., Berenbaum, M.R. (1996) 'Impact of Floral Herbivory by Parsnip Webworm (Oecophoridae: *Depressaria pastinacella* Duponchel) on Pollination and Fitness of Wild Parsnip (Apiaceae: *Pastinaca sativa* L.)', *The American Midland Naturalist*, 136(2), 407–412.

Van Loon, J.J.A., Smallegange, R.C., Bukovinszkiné-Kiss, G., Jacobs, F., De Rijk, M., Mukabana, W.R., Verhulst, N.O., Menger, D.J., Takken, W. (2015) 'Mosquito Attraction: Crucial Role of Carbon Dioxide in Formulation of a Five-Component Blend of Human-Derived Volatiles', *Journal of Chemical Ecology*.

Loor, K.A., DeFoliart, G.R. (1969) 'An oviposition trap for detecting the presence of *Aedes trisriatus* (Say)', *Mosquito News*, 29(3), 487–488.

Lorenz, C., Almeida, F., Almeida-Lopes, F., Louise, C., Pereira, S.N., Petersen, V., Vidal, P.O., Virginio, F., Suesdek, L. (2017) 'Geometric morphometrics in mosquitoes: What has been measured?', *Infection, Genetics and Evolution*, 54, 205–215.

Lorenz, C., Marques, T.C., Sallum, M.A.M., Suesdek, L. (2012) 'Morphometrical diagnosis of the malaria vectors *Anopheles cruzii*, *An. homunculus* and *An. bellator*', *Parasites & Vectors*, 5(1), 257.

Lorenz, C., Sergio, A., Suesdek, L. (2015) 'Artificial Neural Network applied as a methodology of mosquito species identification', *Acta Tropica*, 152, 165–169.

Lounibos, L.P., Escher, R.L. (1983) 'Seasonality and sampling of *Coquillettidia perturbans* (Diptera: Culicidae) in south Florida', *Environmental Entomology*, 12(4), 1087–1093.

Lovett-Doust, J., J, L., Doust, L. (1982) 'Life-history patterns in british umbelliferae: a review', *Botanical Journal of the Linnean Society*, 85, 179–194.

Lucero, M., Estell, R., Tellez, M., Fredrickson, E. (2009) 'A retention index calculator simplifies identification of plant volatile organic compounds', *Phytochemical Analysis*, 20(5), 378–384.

Lühken, R., Pfitzner, W.P., Börstler, J., Garms, R., Huber, K., Schork, N., Steinke, S., Kiel, E., Becker, N., Tannich, E., Krüger, A. (2014) 'Field evaluation of four widely used mosquito traps in Central Europe.', *Parasites & vectors*.

Lühken, R., Steinke, S., Leggewie, M., Tannich, E., Krüger, A., Becker, S., Kiel, E. (2015) 'Physico-Chemical Characteristics of *Culex pipiens* sensu lato and *Culex torrentium* (Diptera: Culicidae) Breeding Sites in Germany: Table 1.', *Journal of Medical Entomology*, tjv070.

Lundström, J.O., Niklasson, B., Francy, D.B. (1990) 'Swedish *Culex torrentium* and *Cx. pipiens* (Diptera: Culicidae) as experimental vectors of Ockelbo virus', *J Med Entomol*, 27(4), 561–563.

Magnarelli, L.A. (1977) 'Nectar feeding by *Aedes sollicitans* and its relation to gonotrophic activity', *Environmental Entomology*, 6, 237–242.

Magurano, F., Remoli, M.E., Baggieri, M., Fortuna, C., Marchi, A., Fiorentini, C., Bucci, P., Benedetti, E., Ciufolini, M.G., Rizzo, C., Piga, S., Salcuni, P., Rezza, G., Nicoletti, L. (2012) 'Circulation of West Nile virus lineage 1 and 2 during an outbreak in Italy', *Clinical Microbiology and Infection*, 18(12), E545–E547.

Maia, M.F., Moore, S.J. (2011) 'Plant-based insect repellents: a review of their efficacy, development and testing.', *Malaria journal*, 10 Suppl 1(Suppl 1), S11.

Manda, H., Gouagna, L.-C., Foster, W.A., Jackson, R.R., Beier, J.C., Githure, J.I., Hassanali, A. (2007) 'Effect of discriminative plant-sugar feeding on the survival and fecundity of *Anopheles gambiae*.', *Malaria journal*, 6, 113.

Marcondes, C.B., Borges, P.S.S. (2000) 'Distinction of Males of the *Lutzomyia intermedia* (Lutz & Neiva, 1912) Species Complex by Ratios between Dimensions and by an Artificial Neural Network (Diptera: Psychodidae, Phlebotominae)', *Memorias do Instituto Oswaldo Cruz*, 95(5), 685–688.

Martin, J.P., Beyerlein, A., Dacks, A.M., Reisenman, C.E., Riffell, J.A., Lei, H., Hildebrand, J.G. (2011) 'The neurobiology of insect olfaction: sensory processing in a comparative context.', *Progress in neurobiology*, 95(3), 427–47.

Mauer, D.J., Rowley, W.A. (1999) 'Attraction of *Culex pipiens pipiens* (Diptera: Culicidae) to flower volatiles.', *Journal of Medical Entomology*, 36(4), 503–7.

McBride, C.S., Baier, F., Omondi, A.B., Spitzer, S.A., Lutomiah, J., Sang, R., Ignell, R., Vosshall, L.B. (2014) 'Evolution of mosquito preference for humans linked to an odorant receptor', *Nature*, 515(7526), 222–227.

McCall, P.J., Eaton, G. (2001) 'Olfactory memory in the mosquito *Culex quinquefasciatus*',

Medical and Veterinary Entomology, 15, 197–203.

McCallum, E.J., Cunningham, J.P., Lucker, J., Zalucki, M.P., De Voss, J.J., Botella, J.R. (2011) 'Increased plant volatile production affects oviposition, but not larval development, in the moth *Helicoverpa armigera*', *Journal of Experimental Biology*, 214(21), 3672–3677.

Medlock, J.M., Avenell, D., Barrass, I., Leach, S.A. (2006) 'Analysis of the potential for survival and seasonal activity of *Aedes albopictus* (Diptera: Culicidae) in the United Kingdom.', *Journal of Vector Ecology*, 31(2), 292–304.

Medlock, J.M., Hansford, K.M. (2012) 'Mosquito nuisance and control in the UK–A questionnaire-based survey of local authorities', *European Mosquito Bulletin*, 30(February), 15–29.

Medlock, J.M., Leach, S.A. (2012) 'Effects of climate change on vector-borne diseases', in Vardoulakis, S. and Heaviside, C., eds., *Health Effects of Climate Change in the UK 2012*, Health Protection Agency: London, 159–199.

Medlock, J.M., Leach, S.A. (2015) 'Effect of climate change on vector-borne disease risk in the UK', *The Lancet Infectious Diseases*, 3099(15).

Medlock, J.M., Snow, K.R., Leach, S.A. (2005) 'Potential transmission of West Nile virus in the British Isles: an ecological review of candidate mosquito bridge vectors.', *Medical and Veterinary Entomology*, 19(1), 2–21.

Medlock, J.M., Snow, K.R., Leach, S.A. (2007) 'Possible ecology and epidemiology of medically important mosquito-borne arboviruses in Great Britain.', *Epidemiology and Infection*, 135(3), 466–82.

Medlock, J.M., Vaux, A.G.C. (2009) '*Aedes* (*Aedes*) *geminus* Peus (Diptera , Culicidae) – an addition to the British mosquito fauna', *Dipterists Digest*, 16(1818), 147–150.

Medlock, J.M., Vaux, A.G.C. (2011) 'Assessing the possible implications of wetland expansion and management on mosquitoes in Britain', *European Mosquito Bulletin*, 29(March), 38–65.

Medlock, J.M., Vaux, A.G.C. (2014) 'Colonization of a newly constructed urban wetland by mosquitoes in England: implications for nuisance and vector species.', *Journal of vector ecology : journal of the Society for Vector Ecology*, 39(2), 249–60.

Medlock, J.M., Vaux, A.G.C. (2015a) 'Impacts of the creation, expansion and management of English wetlands on mosquito presence and abundance – developing strategies for future disease mitigation', *Parasites & Vectors*, 8(1).

Medlock, J.M., Vaux, A.G.C. (2015b) 'Seasonal dynamics and habitat specificity of mosquitoes in an English wetland : implications for UK wetland management and restoration', *Journal of Vector Ecology*, 40(1), 90–106.

Medlock, J.M., Vaux, A.G.C. (2016) Vector-Borne Diseases in the 21st Century | Public Health Matters [online], *Public Health England*, available: https://publichealthmatters.blog.gov.uk/2016/01/08/mosquito-borne-diseases-in-the-21st-century/ [accessed 7 Jul 2017].

Medlock, J.M., Vaux, A.G.C., Cull, B., Schaffner, F., Gillingham, E., Pfluger, V., Leach, S.A. (2017) 'Detection of the invasive mosquito species *Aedes albopictus* in southern England', *The Lancet Infectious Diseases*, 17(2), 140.

Meixner, M.., Pinto, M.A., Bouga, M., Kryger, P., Ivanova, E., Fuchs, S. (2013) 'Standard methods for characterising subspecies and ecotypes of *Apis mellifera*', *Journal of apicultural research*, 52(4), 1–28.

Merritt, R.W., Dadd, R.H., Walker, E.D. (1992) 'Feeding behavior, natural food, and nutritional relationships of larval mosquitoes.', *Annual review of entomology*, 37(128), 349–76.

Mikaia, A. (2008) 'NIST Standard Reference Database 1A', 1–49.

Mng'ong'o, F.C., Sambali, J.J., Sabas, E., Rubanga, J., Magoma, J., Ntamatungiro, A.J., Turner, E.L., Nyogea, D., Ensink, J.H.J., Moore, S.J. (2011) 'Repellent plants provide affordable natural screening to prevent mosquito house entry in tropical rural settings-results from a pilot efficacy study', *PLoS ONE*, 6(10).

Mohrig, W. (1969) *Die Culiciden Deutschlands: Untersuchungen Zur Taxonomie, Biologie Und Ökologie Der Einheimischen Stechmücken*, G. Fischer.

Mokany, A., Shine, R. (2003) 'Oviposition site selection by mosquitoes is affected by cues from conspecific larvae and anuran tadpoles', *Austral Ecology*, 28(1), 33–37.

Montell, C., Zwiebel, L.J. (2016) 'Mosquito Sensory Systems', in Raikhel, A.S., ed., *Advances in Insect Physiology*, Academic Press: London, 293–328.

Morris, M.G. (1987) 'Changing attitudes to nature conservation: the entomological perspective', *Biological Journal of the Linnean Society*, (32), 213–223.

Mosquito_Magnet (2013) 'Mosquito Magnet, Executive - Operation Manual'.

Müller, G.C., Beier, J.C., Traore, S.F., Toure, M.B., Traore, M.M., Bah, S., Doumbia, S., Schlein, Y. (2010) 'Successful field trial of attractive toxic sugar bait (ATSB) plant-spraying methods against malaria vectors in the *Anopheles gambiae* complex in Mali, West Africa.', *Malaria journal*, 9(210), 1–7.

Müller, G.C., Junnila, A., Schlein, Y. (2010) 'Effective control of adult *Culex pipiens* by spraying an attractive toxic sugar bait solution in the vegetation near larval habitats.', *Journal of medical entomology*, 47(1), 63–66.

Müller, G.C., Junnila, A., Traore, M.M., Traore, S.F., Doumbia, S., Sissoko, F., Dembele, S.M., Schlein, Y., Arheart, K.L., Revay, E.E., Kravchenko, V.D., Witt, A., Beier, J.C. (2017) 'The

invasive shrub *Prosopis juliflora* enhances the malaria parasite transmission capacity of Anopheles mosquitoes: a habitat manipulation experiment', *Malaria Journal*, 16(1), 237.

Müller, G.C., Schlein, Y. (2005) 'Plant tissues: The frugal diet of mosquitoes in adverse conditions', *Medical and Veterinary Entomology*, 19(4), 413–422.

Müller, G.C., Schlein, Y. (2006) 'Sugar questing mosquitoes in arid areas gather on scarce blossoms that can be used for control.', *International journal for parasitology*, 36(10–11), 1077–80.

Müller, G.C., Xue, R.-D., Schlein, Y. (2011) 'Differential attraction of *Aedes albopictus* in the field to flowers, fruits and honeydew.', *Acta tropica*, 118(1), 45–9.

Muturi, E.J., Mwangangi, J., Shililu, J., Muriu, S., Jacob, B., Mbogo, C.M., John, G., Novak, R. (2007) 'Evaluation of four sampling techniques for surveillance of *Culex quinquefasciatus* (Diptera: Culicidae) and other mosquitoes in African rice agroecosystems.', *Journal of Medical Entomology*, 44(3), 503–508.

Myung, D., Jais, A., He, L., Chang, R.T. (2014) 'Simple, Low-Cost Smartphone Adapter for Rapid, High Quality Ocular Anterior Segment Imaging: A Photo Diary', *Journal of Mobile Technology in Medicine*, 3(1), 2–8.

Najar-Rodriguez, A.J., Galizia, C.G., Stierle, J., Dorn, S. (2010) 'Behavioral and neurophysiological responses of an insect to changing ratios of constituents in host plant-derived volatile mixtures.', *The Journal of experimental biology*, 213(Pt 19), 3388–3397.

Nasci, R.S., Savage, H.M., White, D.J. (2001) 'West Nile virus in overwintering *Culex* mosquitoes, New York City, 2000.', *Emerging infectious …*, 7(4), 742–744.

Nayar, J.K., Van Handel, E. (1971) 'The fuel for sustained mosquito flight', *Journal of Insect Physiology*, 17(3), 471–481.

NBN (2017) Home - NBN Atlas [online], available: https://nbnatlas.org/ [accessed 11 Jul 2017].

Nerio, L.S., Olivero-Verbel, J., Stashenko, E. (2010) 'Repellent activity of essential oils: A review', *Bioresource Technology*, 101(1), 372–378.

Nielsen, E.T., Nielsen, A.T. (1953) 'Field Observations on the Habits of *Aedes Taeniorhynchus*', *Ecology*, 34(1), 141–156.

Niemirski, R., Zych, M. (2011) 'Fly pollination of dichogamous *Angelica sylvestris* (Apiaceae): How (functionally) specialized can a (morphologically) generalized plant be?', *Plant Systematics and Evolution*, 294(3), 147–158.

Nikbakhtzadeh, M.R., Terbot, J.W., Otienoburu, P.E., Foster, W.A. (2014) 'Olfactory Basis of Floral Preference of the Malaria Vector *Anopheles gambiae* (Diptera: Culicidae) Among

Common African Plants', *Journal of Vector Ecology*, 39(2), 372–383.

Nishida, R. (2014) 'Chemical ecology of insect–plant interactions: ecological significance of plant secondary metabolites', *Bioscience, Biotechnology and Biochemistry*, 78(1), 1–13.

Nissinen, A., Ibrahim, M., Kainulainen, P., Tiilikkala, K., Holopaine, J.K. (2005) 'Influence of carrot psyllid (Trioza apicalis) feeding or exogenous limonene or methyl jasmonate treatment on composition of carrot (*Daucus carota*) leaf essential oil and headspace volatiles', *Journal of Agricultural and Food Chemistry*, 53(22), 8631–8638.

Nsoesie, E.O., Kraemer, M.U., Golding, N., Pigott, D.M., Brady, O.J., Moyes, C.L. (2016) 'Global distribution and environmental suitability for chikungunya virus , 1952 to 2015', *Euro Surveillance*, 21(20).

Nyasembe, V.O., Teal, P.E. a, Mukabana, W.R., Tumlinson, J.H., Torto, B. (2012) 'Behavioural response of the malaria vector *Anopheles gambiae* to host plant volatiles and synthetic blends.', *Parasites & Vectors*, 5(1), 234.

Nyasembe, V.O., Torto, B. (2014) 'Volatile phytochemicals as mosquito semiochemicals', *Phytochemistry Letters*, 8(1), 196–201.

Obenauer, P.J., Abdel-Dayem, M.S., Stoops, C. a, Villinski, J.T., Tageldin, R., Fahmy, N.T., Diclaro, J.W., Bolay, F. (2013) 'Field responses of *Anopheles gambiae* complex (Diptera: Culicidae) in Liberia using yeast-generated carbon dioxide and synthetic lure-baited light traps.', *Journal of Medical Entomology*, 50(4), 863–70.

Odom, A.L., Bell, C. V. (2011) 'Distinguishing among declarative, descriptive and causal questions to guide field investigations and student assessment', *Journal of Biological Education*, 45(4), 222–228.

Oleksa, A., Tofilski, A. (2015) 'Wing geometric morphometrics and microsatellite analysis provide similar discrimination of honey bee subspecies', *Apidologie*, 46, 49–60.

Olsen, S.R., Bhandawat, V., Wilson, R.I. (2007) 'Excitatory Interactions between Olfactory Processing Channels in the *Drosophila* Antennal Lobe', *Neuron*, 54(1), 89–103.

Olsson, S.B., Hansson, B.S. (2013) 'Electroantennogram and Single Sensillum Recording in Insect Antennae', in Touhara, K., ed., *Methods in Molecular Biology*, Methods in Molecular Biology, Humana Press: Totowa, NJ, 157–177.

Omrani, S.-M., Vatandoost, H., Oshaghi, M.A., Shokri, F., Guerin, P.M., Yaghoobi Ershadi, M.R., Rassi, Y., Tirgari, S. (2010) 'Fabrication of an olfactometer for mosquito behavioural studies.', *Journal of Vector Borne Diseases*, 47(1), 17–25.

Onyeka, J.O.A. (1982) 'The taxonomic value of prealar scales in the identification of *Culex pipiens* and *Culex torrentium* (Diptera Culicidae)', *Mosquito Systematics*, 14(1), 41–46.

Onyeka, J.O.A. (1983) 'Studies on the natural predators of *Culex pipiens* L. and *Culex*

torrentium Martini (Diptera, Culicidae) in England', *Bulletin of Entomological Research*, 73(02), 185–194.

von Oppen, S., Masuh, H., Licastro, S., Zerba, E., Gonzalez-Audino, P. (2015) 'A floral-derived attractant for *Aedes aegypti* mosquitoes', *Entomologia Experimentalis et Applicata*, (2014), 1–9.

Osorio, Costa, H., Ze-Ze, Libia, Alves, Joao, M. (2012) 'Host-Feeding Patterns of *Culex pipiens* and Other Potential Mosquito Vectors (Diptera: Culicidae) of West Nile Virus (Flaviviridae) Collected in Portugal', *Journal of Medical Entomology*, 49(3), 717–721.

Otienoburu, P.E. (2011) *Plant Semiochemicals as Mosquito Attractants*.

Otienoburu, P.E., Ebrahimi, B., Phelan, P.L., Foster, W.A. (2012) 'Analysis and optimization of a synthetic milkweed floral attractant for mosquitoes.', *Journal of Chemical Ecology*, 38(7), 873–81.

Özek, T., Özek, G., Baser, K.H.C., Duran, A. (2005) 'Comparison of the Essential Oils of Three Endemic Turkish *Heracleum* Species Obtained by Different Isolation Techniques', *Journal of Essential Oil Research*, 17(October 2012), 605–610.

Özek, T., Özek, G., Can Başer, K.H., Duran, A., Sagiroglu, M. (2008) 'Composition of the Essential Oils of *Angelica sylvestris* L. var. *sylvestris* Isolated from the Fruits by Different Isolation Techniques.', *Journal of Essential Oil Research*, 20(5), 408–411.

Packer, M.J., Corbet, P.S. (1989) 'Size variation and reproductive success of female *Aedes punctor* (Diptera: Culicidae)', *Ecological Entomology*, 14(3), 297–309.

Padmaja, P.G., Woodcock, C.M., Bruce, T.J.A. (2010) 'Electrophysiological and Behavioral Responses of Sorghum Shoot Fly, *Atherigona soccata*, to Sorghum Volatiles', *Journal of Chemical Ecology*, 36(12), 1346–1353.

Palmer, A.R. (1994) 'Fluctuating asymmetry analyses: A primer', in Markow, T.A., ed., *Developmental Instability: Its Origins and Evolutionary Implications*, Kluwer: Dordrecht, 335–364.

Palmer, A.R., Strobeck, C. (2003) 'Fluctuating asymmetry analyses revisited', in Polak, M., ed., *Developmental Instability (DI): Causes and Consequences*, Oxford University Press: Oxford, 279–319.

Papa, A., Xanthopoulou, K., Gewehr, S., Mourelatos, S. (2011) 'Detection of West Nile virus lineage 2 in mosquitoes during a human outbreak in Greece', *Clinical Microbiology and Infection*, 17(8), 1176–1180.

Paré, P.W., Tumlinson, J.H. (1999) 'Plant volatiles as a defense against insect herbivores', *Plant Physiology*, 121, 325–331.

Parsons, P.A. (1990) 'Fluctuating Asymmetry: an Epigenetic Measure of Stress', *Biological*

Reviews, 65(2), 131–145.

Pates, H.V., Takken, W., Stuke, K., Curtis, C.F. (2001) 'Differential behaviour of Anopheles gambiae sensu stricto (Diptera: Culicidae) to human and cow odours in the laboratory', *Bulletin of Entomological Research*, 91(04), 289–296.

Pearson, D.L., Shetterly, J.A. (2006) 'How Do Published Field Guides Influence Interactions between Amateurs and Professionals in Entomology ?', *American Entomologist*, (1999), 246–252.

Pecorari, M., Longo, G., Gennari, W., Grottola, A., Sabbatini, A.M.T., Tagliazucchi, S. (2009) 'First human case of Usutu virus neuroinvasive infection, Italy, August-September 2009', *Euro Surveillance*, 14(50), 1–2.

Pelletier, J., Guidolin, A., Syed, Z., Cornel, A.J., Leal, W.S. (2010) 'Knockdown of a mosquito odorant-binding protein involved in the sensitive detection of oviposition attractants', *Journal of Chemical Ecology*, 36(3), 245–248.

Pelletier, J., Hughes, D.T., Luetje, C.W., Leal, W.S. (2010) 'An odorant receptor from the Southern house mosquito *Culex pipiens quinquefasciatus* sensitive to oviposition attractants', *PLoS ONE*, 5(4), 1–8.

Pelletier, J., Leal, W.S. (2009) 'Genome analysis and expression patterns of odorant-binding proteins from the Southern House mosquito *Culex pipiens quinquefasciatus*', *PLoS ONE*, 4(7).

Perry, J.N., Wall, C., Greenway, A.R. (1980) 'Latin Square designs in field experiments involving insect sex attractants', *Ecological Entomology*, 5(4), 385–396.

Peterson, J.K., Horvat, R.J., Elsey, K.D. (1994) 'Squash leaf glandular trichome volatiles: identification and influence on behaviour of female pickleworm moth [*Diaphania nitidalis* (Stoll.)] (Lepidoptera: Pyralidae)', *Journal of Chemical Ecology*, 20(8), 2099–2109.

Pettersson, E.M., Sullivan, B.T., Anderson, P., Berisford, C.W., Birgersson, G. (2000) 'Odor perception in the bark beetle parasitoid *Roptrocerus xylophagorum* exposed to host associated volatiles', *Journal of Chemical Ecology*, 26(11), 2507–2525.

PHE (2014) 'Investigations: Investigations of Red Rash', *UK Standards for Microbiology*, G7(2.1), 1–21.

Piñero, J.C., Dorn, S. (2007) 'Synergism between aromatic compounds and green leaf volatiles derived from the host plant underlies female attraction in the oriental fruit moth', *Entomologia Experimentalis et Applicata*, 125(2), 185–194.

Piper, R.W., Compton, S.G. (2013) 'How limited is dispersal in the rare beetle, Cryptocephalus decemmaculatus (Chrysomelidae, Cryptocephalinae)?', *Journal of Insect Conservation*, 17(6), 1229–1235.

Pitts, R.J., Mozuraitis, R., Gauvin-Bialecki, A., Lempérière, G. (2014) 'The roles of kairomones, synomones and pheromones in the chemically-mediated behaviour of male mosquitoes', *Acta Tropica*, 132(1), 26–34.

Poland., J., Clement., E.J. (2009) *The Vegetative Key to the British Flora : A New Approach to Naming British Vascular Plants Based on Vegetative Characters*, John Poland.

Posey, K.H., Barnard, D.R., Schreck, C.E. (1998) 'Triple cage olfactometer for evaluating mosquito (Diptera: Culicidae) attraction responses.', *Journal of medical entomology*, 35(3), 330–4.

Qiao, H.L., Lu, P.F., Chen, J., Ma, W.S., Qin, R.M., Li, X.M. (2012) 'Antennal and behavioural responses of *Heortia vitessoides* females to host plant volatiles of *Aquilaria sinensis*', *Entomologia Experimentalis et Applicata*, 143(3), 269–279.

Qiu, Y.T., van Loon, J.J.A. (2010) 'Olfactory physiology of blood-feeding vector mosquitoes', in *Olfaction in Vector-Host Interactions*, Wageningen Academic Publishers, 39–61.

Qiu, Y.T., Smallegange, R.C., Smid, H., Loon, J.J.A. Van, Galimard, A., Posthumus, M.A., Beek, T. Van, Takken, W. (2004) 'GC-EAG analysis of human odours that attract the malaria mosquito *Anopheles gambiae* sensu stricto', *Proceedings of the Netherlands Entomological Meeting*, 15, 59–64.

Qualls, W. a., Müller, G.C., Revay, E.E., Allan, S. a., Arheart, K.L., Beier, J.C., Smith, M.L., Scott, J.M., Kravchenko, V.D., Hausmann, A., Yefremova, Z. a., Xue, R. De (2014) 'Evaluation of attractive toxic sugar bait (ATSB)-Barrier for control of vector and nuisance mosquitoes and its effect on non-target organisms in sub-tropical environments in Florida', *Acta Tropica*, 131(1), 104–110.

R Core Team (2015) 'R: A Language and Environment for Statistical Computing'.

R Development Core Team, R. (2011) *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, R Foundation for Statistical Computing, R Foundation for Statistical Computing.

Raguso, R.A., Pellmyr, O. (1998) 'Dynamic headspace analysis of floral volatiles: a comparison of methods', *Oikos*, 81(2), 238–254.

Ramsdale, C.D., Gunn, N. (2005) 'History of and prospects for mosquito-borne disease in Britain', *Journal of the European Mosquito Control Association*, 20, 15–31.

Ramsdale, C.D., Snow, K.R. (1995) *Mosquito Control in Britain*, University of East London: Dagenham.

Refsnider, J.M., Janzen, F.J. (2010) 'Putting Eggs in One Basket: Ecological and Evolutionary Hypotheses for Variation in Oviposition-Site Choice', *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 39–57.

Reinert, J.F., Harbach, R.E., Kitching, I.A.N.J. (2006) 'Phylogeny and classification of *Finlaya* and allied taxa (Diptera : Culicidae : Aedini) based on morphological data from all life stages', *Zoological Journal of the Linnean Society*, 148, 1–101.

Reinert, J.F., Harbach, R.E., Kitching, I.J. (2009) 'Phylogeny and classification of tribe Aedini (Diptera: Culicidae)', *Zoological Journal of the Linnean Society*, 157(4), 700–794.

Reisen, W.K., Meyer, R.P., Cummings, R.F., Delgado, O. (2000) 'Effects of trap design and CO2 presentation on the measurement of adult mosquito abundance using Centers for Disease Control-style miniature light traps.', *Journal of the American Mosquito Control Association*, 16(June), 13–18.

Reisen, W.K., Meyer, R.P., Milby, M.M. (1986) 'Patterns of fructose feeding by *Culex tarsalis* (Diptera: Culicidae)', *Journal of Medical Entomology*, 23(4), 366–373.

Reiter, P. (2001) 'Climate Change and Mosquito-Borne Disease', *Environmental Health Prespective*, 109(July 2000), 141–161.

Revay, E.E., Müller, G.C., Qualls, W. a., Kline, D.L., Naranjo, D.P., Arheart, K.L., Kravchenko, V.D., Yefremova, Z.A., Hausmann, A., Beier, J.C., Schlein, Y., Xue, R. De (2014) 'Control of *Aedes albopictus* with attractive toxic sugar baits (ATSB) and potential impact on non-target organisms in St. Augustine, Florida', *Parasitology Research*, 113(1), 73–79.

Rezza, G., Nicoletti, L., Romi, R., Finarelli, A.C., Panning, M., Cordioli, P., Fortuna, C., Boros, S., Magurano, F., Silvi, G., Angelini, P., Dottori, M., Ciufolini, M.G., Majori, G.C., Cassone, A. (2007) 'Infection with chikungunya virus in Italy: an outbreak in a temperate region', *Lancet*, 370, 1840–46.

Richards, S.L., Anderson, S.L., Yost, S. a (2012) 'Effects of blood meal source on the reproduction of *Culex pipiens quinquefasciatus* (Diptera: Culicidae).', *Journal of vector ecology : journal of the Society for Vector Ecology*, 37(1), 1–7.

Ries, L., Fletcher, R.J., Battin, J., Sisk, T.D. (2004) 'Ecological Responses to Habitat Edges : Mechanisms , Models , and Variability Explained', *Annual Reviews in Ecology, Evolution and Systematic*, 35(2004), 491–522.

Riffell, J.A., Shlizerman, E., Sanders, E., Abrell, L., Medina, B., Hinterwirth, A.J., Kutz, J.N. (2014) 'Flower discrimination by pollinators in a dynamic chemical environment', *Science*, 344(6191), 1515–1518.

Ripley, B., Lapsley, M. (2016) 'RODBC: ODBC Database Access'.

Ritchie, S.A., Kline, D.L. (1995) 'Comparison of CDC and EVS Light Traps Baited with Carbon Dioxide and Octenol for Trapping Mosquitoes in Brisbane, Queensland (Diptera: Culicidae)', *Australian Journal of Entomology*, 34(3), 215–218.

Ritchie, S.A., Zborowski, P., Banks, D., Walsh, I., Davis, J. (2008) 'Efficacy of Novel Updraft Traps for Collection of Mosquitoes in Cairns, Australia', *Journal of the American Mosquito*

Control Association, 24(4), 520–527.

Rizzoli, A., Bolzoni, L., Chadwick, E. a, Capelli, G., Montarsi, F., Grisenti, M., de la Puente, J.M., Muñoz, J., Figuerola, J., Soriguer, R., Anfora, G., Di Luca, M., Rosà, R. (2015) 'Understanding West Nile virus ecology in Europe: *Culex pipiens* host feeding preference in a hotspot of virus emergence', *Parasites & Vectors*, 8(1).

Robich, R.M., Denlinger, D.L. (2005) 'Diapause in the mosquito *Culex pipiens* evokes a metabolic switch from blood feeding to sugar gluttony.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(44), 15912–15917.

Rohlf, F.J. (1993) 'Relative warps analysis and an example of its application to mosquito wings', *Contributions to morphometrics*, 131–159.

Rohlf, F.J. (1999) 'Shape Statistics: Procrustes Superimpositions and Tangent Spaces', *Journal of Classification*.

Rohlf, F.J., Archie, J.W. (1984) 'A Comparison of Fourier Methods for the Description of Wing Shape in Mosquitoes (Diptera: Culicidae)', *Systematic Zoology*, 33(3), 302–317.

Rohlf, F.J., Marcus, L.F. (1993) 'A revolution in Morphometrics', *Tree*, 8(4), 129–132.

Romi, R. (2001) '*Aedes albopictus* in Italia: Un problema sanitario sottovalutato', *Annali dell'Istituto Superiore di Sanita*, 37(2), 241–247.

Romi, R., Pontuale, G., Ciufolini, M.G., Fiorentini, G., Marchi, A., Nicoletti, L., Cocchi, M., Tamburro, A. (2004) 'Potential vectors of West Nile Virus following an equine disease outbreak in Italy', *Medical and Veterinary Entomology*, 18(1), 14–19.

Rose, F. (2006) *The Wild Flower Key (Revised Edition) : How to Identify Wild Plants, Trees and Shrubs in Britain and Ireland*, 2nd ed, F. Warne.

Rousseau, C.L. (2016) Best Universal Lenses For Android Phones | Android Central [online], *Android Central*, available: https://www.androidcentral.com/best-universal-lenses-android-phones [accessed 14 Jul 2017].

RStudio Team (2015) 'RStudio: Integrated Development for R.'

Rudinsky, J.A., Novak, V., Svihra, P. (1971) 'Attraction of the bark beetle *Ips typographus* L. to terpenes and male -produced pheromone', *Zeitschrift für Angewandte Entomologie*, 67, 179–188.

Rudolf, M., Czajka, C., Börstler, J., Melaun, C., Jöst, H., von Thien, H., Badusche, M., Becker, N., Schmidt-Chanasit, J., Krüger, A., Tannich, E., Becker, S. (2013) 'First nationwide surveillance of *Culex pipiens* complex and *Culex torrentium* mosquitoes demonstrated the presence of *Culex pipiens* biotype *pipiens/molestus* hybrids in Germany.', *PloS one*, 8(9), e71832.

Rund, S.S.C., O'Donnell, A.J., Gentile, J.E., Reece, S.E. (2016) 'Daily rhythms in mosquitoes and their consequences for malaria transmission', *Insects*, 7(2), 1–20.

Samy, A.M., Thomas, S.M., Abd, A., Wahed, E., Cohoon, K.P., Peterson, A.T. (2016) 'Mapping the global geographic potential of Zika virus spread', 111(September), 559–560.

Sanders, C.J., Gubbins, S., Mellor, P.S., Barber, J., Golding, N., Harrup, L.E., Carpenter, S.T. (2012) 'Investigation of diel activity of Culicoides biting midges (Diptera: Ceratopogonidae) in the United Kingdom by using a vehicle-mounted trap.', *Journal of Medical Entomology*, 49(3), 757–65.

Sant'Ana, D.C. ristina, Sa, I.L. uizi R. de, Sallum, M.A.M. (2014) 'Effectiveness of Mosquito Magnet trap in rural areas in the southeastern tropical Atlantic Forest', *Memorias do Instituto Oswaldo Cruz*, 109(8), 1021–1029.

Sato, S., Maeto, K. (2006) 'Attraction of female Japanese horntail *Urocerus japonicus* (Hymenoptera : Siricidae) to alpha-pinene', *Applied Entomology and Zoology*, 41(2), 317–323.

Sawadogo, S.P., Niang, A., Bilgo, E., Millogo, A., Maïga, H., Dabire, R.K., Tripet, F., Diabaté, A. (2017) 'Targeting male mosquito swarms to control malaria vector density', *PLoS ONE*, 12(3), 1–11.

Scarpati, M.L., Lo Scalzo, R., Vita, G. (1993) '*Olea europaea* volatiles attractive and repellent to the olive fruit fly', *Journal of Chemical Ecology*, 19(4), 881–891.

Schiestl, F.P., Dötterl, S. (2012) 'The evolution of floral scent and olfactory preferences in pollinators: Coevolution or pre-existing bias?', *Evolution*, 66(7), 2042–2055.

Schlein, Y., Müller, G.C. (2008) 'An approach to mosquito control: using the dominant attraction of flowering *Tamarix jordanis* trees against *Culex pipiens*.', *Journal of medical entomology*, 45(3), 384–390.

Schreck, C.E., Gouck, H.K., Smith, N. (1967) 'An improved olfactometer for use in studying mosquito attractants and repellents', *Journal of Economic Entomology*, 60(4), 1188–1190.

Scott-Fiorenzano, J.M., Fulcher, A.P., Seeger, K.E., Allan, S.A., Kline, D.L., Koehler, P.G., Müller, G.C., Xue, R.-D. (2017) 'Evaluations of dual attractant toxic sugar baits for surveillance and control of *Aedes aegypti* and *Aedes albopictus* in Florida', *Parasites & Vectors*, 10(1), 1–9.

Service, M.W. (1968) 'The taxonomy and biology of two sympatric sibling species of *Culex, C. pipiens* and *C. torrentium* (Diptera, Culicidae)', *Journal of the Zoological Society of London*, 156, 313–323.

Service, M.W. (1971a) 'Feeding behaviour and host preferences of British mosquitoes.', *Bulletin of entomological research*, 60(4), 653–61.

Service, M.W. (1971b) 'The daytime distribution of mosquitoes resting in vegetation.', *Journal of medical entomology*, 8(3), 271.

Service, M.W. (1994) 'Male swarming of the mosquito *Culex(Culex) torrentium* in England', *Medical and Veterinary Entomology*, 8, 95–98.

Seyoum, A., Pålsson, K., Kung'a, S., Kabiru, E.W., Lwande, W., Killeen, G.F., Hassanali, A., Knols, B.G.J. (2002) 'Traditional use of mosquito-repellent plants in western Kenya and their evaluation in semi-field experimental huts against *Anopheles gambiae*: Ethnobotanical studies and application by thermal expulsion and direct burning', *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 96(3), 225–231.

Shah, P.A., Pell, J.K. (2003) 'Entomopathogenic fungi as biological control agents', *Applied Microbiology and Biotechnology*, 61, 413–423.

Shaikevich, E. V (2007) 'PCR-RFLP of the COI gene reliably differentiates *Cx. pipiens*, *Cx. pipiens* f. *molestus* and *Cx. torrentium* of the Pipiens Complex', *European Mosquito Bulletin*, 22, 25–30.

Shute, P.G. (1951) '*Culex molestus*', *Transactions of the Royal Entomological Society of London*, 102(7), 380–382.

Silver, J.B. (2008a) *Mosquito Ecology - Field Sampling Methods*, 3rd ed, Springer: New York.

Silver, J.B. (2008b) 'Sampling the Larval Population', in *Mosquito Ecology: Field Sampling Methods*, Springer Netherlands.

Silver, J.B. (2008c) 'Sampling Adults with Light-Traps', in *Mosquito Ecology: Field Sampling Methods*, Springer: New York, 845–946.

Sirot, L.K., Poulson, R.L., Mckenna, M.C., Girnary, H., Mariana, F., Harrington, L.C. (2009) 'Control of Female Feeding and Reproduction', *Insect Biochemistry*, 38(2), 176–189.

Slaff, M., Haefner, J.D., Parsons, R.E., Wilson, F. (1984) 'A Modified Pyramidal Emergence Trap for Collecting Mosquitoes', *Journal of the American Mosquito Control Association*, 44(2), 197–199.

Slatkin, M. (1987) 'Gene Flow and the Geographic Structure of Natural Populations', *Science*, 236(4803), 787–792.

Smallegange, R.C., Knols, B.G.J., Takken, W. (2010) 'Effectiveness of synthetic versus natural human volatiles as attractants for *Anopheles gambiae* (Diptera: Culicidae) *sensu stricto*.', *Journal of medical entomology*, 47(3), 338–344.

Smallegange, R.C., Qiu, Y.T., Bukovinszkiné-Kiss, G., Van Loon, J.J.A., Takken, W. (2009) 'The effect of aliphatic carboxylic acids on olfaction-based host-seeking of the malaria mosquito *Anopheles gambiae sensu stricto*', *Journal of Chemical Ecology*, 35(8), 933–943.

Smallegange, R.C., Qiu, Y.T., Van Loon, J.J.A., Takken, W. (2005) 'Synergism between ammonia, lactic acid and carboxylic acids as kairomones in the host-seeking behaviour of the malaria mosquito *Anopheles gambiae sensu stricto* (Diptera: Culicidae)', *Chemical Senses*, 30(2), 145–152.

Smith, J.J.L., Fonseca, D.M. (2004) 'Rapid assays for identification of members of the *Culex (Culex) pipiens* complex, their hybrids, and other sibling species (Diptera: Culicidae)', *The American journal of tropical medicine and …*, 70(4), 339–345.

Smithburn, K.C., Hughes, T.P., Burke, A.W., Paul, J.H. (1940) 'A Neurotropic Virus Isolated from the Blood of a Native of Uganda1', *The American Journal of Tropical Medicine and Hygiene*, s1-20(4), 471–492.

Snow, K.R. (1986) 'Mosquito nuisance and control in UK.', *Environmental Health*, 94(10), 265–268.

Snow, K.R. (1990) *Mosquitoes*, The Richmond Publishing Co. Ltd.: Slough.

Snow, K.R., Medlock, J.M. (2006) 'The potential impact of climate change on the distribution and prevalence of mosquitoes in Britain', *European Mosquito Bulletin*, 21, 1–10.

Snow, K.R., Medlock, J.M. (2008) 'The mosquitoes of Epping Forest, Essex, UK', *European Mosquito Bulletin*, 26(December), 9–17.

Spielman, A., Wong, J. (1973) 'Environmental Control of Ovarian Diapause in *Culex pipiens*', *Annals of the Entomological Society of America*, 66(4), 905–908.

Stegmann, M.B., Gomez, D.D. (2002) 'A Brief Introduction to Statistical Shape Analysis', (March), 1–15.

Stein, S. (2012) 'Mass spectral reference libraries: An ever-expanding resource for chemical identification', *Analytical Chemistry*, 84(17), 7274–7282.

Steinbrecht, R.A. (1997) 'Pore Structures a Review in Insect Olfactory Sensilla : of Data and Concepts', *International Journal of Insect Morphology and Embryology*, 26(3/4), 229–245.

Stengl, M., Funk, N.W. (2013) 'The role of the coreceptor Orco in insect olfactory transduction', *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 199(11), 897–909.

Stewart-Jones, A., Poppy, G.M. (2006) 'Comparison of glass vessels and plastic bags for enclosing living plant parts for headspace analysis', *Journal of Chemical Ecology*, 32(4), 845–864.

Sturaro, A., Parvoli, G., Doretti, L. (1992) 'Artifacts produced by Porapak Q sorbent tubes on solvent desorption', *Chromatographia*, 33(1–2), 53–57.

Suchet, C., Dormont, L., Schatz, B., Giurfa, M., Simon, V., Raynaud, C., Chave, J. (2011) 'Floral scent variation in two *Antirrhinum majus* subspecies influences the choice of naïve bumblebees', *Behavioral Ecology and Sociobiology*, 65(5), 1015–1027.

Sulaiman, S. (1982) *The Ecology of* Aedes Cantans *(Meigen) and Biology of* Culex Pipiens *L. in Hibernation Sites in Northern England.*

Sulaiman, S. (1983) 'The difference in the instar composition of immature mosquitoes sampled by three sampling techniques', *Japanese Journal of Medical Science and Biology*, 36, 281–288.

Sullivan, G. a., Liu, C., Syed, Z. (2014) 'Oviposition signals and their neuroethological correlates in the *Culex pipiens* complex', *Infection, Genetics and Evolution*, (October).

Sullivan, K., Kloess, J., Qian, C., Bell, D., Hay, A., Lin, Y.P., Gu, Y. (2012) 'High throughput virus plaque quantitation using a flatbed scanner', *Journal of Virological Methods*, 179(1), 81–89.

Sumruayphol, S., Apiwathnasorn, C., Ruangsittichai, J., Sriwichai, P., Attrapadung, S., Samung, Y., Dujardin, J.P. (2016) 'DNA barcoding and wing morphometrics to distinguish three *Aedes* vectors in Thailand', *Acta Tropica*, 159, 1–10.

Sutcliffe, J.F. (1987) 'Distance orientation of biting flies to their hosts', *Insect Science and its Applications*, 8(4/5/6), 611–616.

Sweeney, J., Groot, P. De, Macdonald, L., Smith, S., Kenis, M., Gutowski, J.M., Sweeney, J.O.N., Groot, P.D.E., Cocquempot, C. (2004) 'Host Volatile Attractants and Traps for Detection of *Tetropium fuscum* (F .), *Tetropium castaneum* L ., and Other Longhorned Beetles (Coleoptera : Cerambycidae)', *Environmental entomology*, 33(4), 844–854.

Syed, Z., Leal, W.S. (2009) 'Acute olfactory response of *Culex* mosquitoes to a human- and bird-derived attractant', *Proceedings of the National Academy of Sciences USA*, 106(44), 18803–18808.

Takken, W., Dekker, T., Wijnholds, Y.G.I. (1997) 'Odor-Mediated Flight Behavior of *Anopheles gambiae* Giles Sensu Stricto and *An . stephensi* Liston in Response to CO2 , Acetone , and 1-Octen-3-ol (Diptera : Culicidae)', *Journal of Insect Behavior*, 10(3), 395–407.

Takken, W., Knols, B.G.J. (1999) 'Odor-mediated behavior of Afrotropical malaria mosquitoes.', *Annual Review of Entomology*, 44, 131–57.

Takken, W., Knols, B.G.J. (2007) *Emerging Pests and Vector-Borne Diseases in Europe*, Wageningen Academic Publishers: Wageningen.

Takken, W., Koenraadt, C.J.M. (2013) *Ecology of Parasite-Vector Interactions*, Wageningen Academic Publishers: Wageningen.

Takken, W., Verhulst, N.O. (2013) 'Host Preferences of Blood-Feeding Mosquitoes', *Annual*

Review of Entomology, (September), 433–453.

Tasin, M., Bäckman, A.C., Bengtsson, M., Ioriatti, C., Witzgall, P. (2006) 'Essential host plant cues in the grapevine moth', *Naturwissenschaften*, 93(3), 141–144.

Taylor, L.H., Latham, S.M., Woolhouse, M.E.J. (2001) 'Risk factors for human disease emergence'.

Tholl, D., Boland, W., Hansel, A., Loreto, F., Röse, U.S.R., Schnitzler, J.-P. (2006) 'Practical approaches to plant volatile analysis.', *The Plant journal : for cell and molecular biology*, 45(4), 540–60.

Thullen, J.S., Sartoris, J.J., Walton, W.E. (2002) 'Effects of vegetation management in constructed wetland treatment cells on water quality and mosquito production', *Ecological Engineering*, 18(4), 441–457.

Tollsten, L., Knudsen, J.T., Bergstrom, L.G. (1994) 'Floral scent in generalistic *Angelica* (Apiaceae) - An adaptive character?', *Biochemical Systematics and Ecology*, 22(2), 161–169.

Tollsten, L., Øvstedal, D.O. (1994) 'Differentiation in floral scent chemistry among populations of *Conopodium majus* (Apiaceae)', *Nordic Journal of Botany*, 14(4), 361–368.

Torto, B., Carroll, M.J., Duehl, A., Fombong, A.T., Gozansky, T.K., Nazzi, F., Soroker, V., Teal, P.E.A. (2013) 'Standard methods for chemical ecology research in Apis mellifera', *Journal of Apicultural Research*, 52(4), 1–34.

Townroe, S., Callaghan, A. (2014) 'British container breeding mosquitoes: the impact of urbanisation and climate change on community composition and phenology.', *PloS one*, 9(4), e95325.

Trpis, M., Häusermann, W., Craig, Jr., G.B. (1995) 'Estimates of population size, dispersal, and longevity of domestic *Aedes aegypti aegypti* (Diptera: Culicidae) by mark-release-recapture in the village of Shauri Moyo in eastern Kenya', *Journal of Medical Entomology*, 32(1), 27–33.

Tsai, T.F., Popovici, F., Cernescu, C., Campbell, G.L., Nedelcu, N.I. (1998) 'West Nile encephalitis epidemic in southeastern Romania', *Lancet*, 352(9130), 767–771.

Turell, M.J., Dohm, D.J., Sardelis, M.R., O'guinn, M.L., Andreadis, T.G., Blow, J.A. (2005) 'An Update on the Potential of North American Mosquitoes (Diptera: Culicidae) to Transmit West Nile Virus', *Journal of Medical Entomology*, 42(1), 57–62.

Turell, M.J., Sardelis, M.R., Jones, J.W., Watts, D.M., Fernandez, R., Carabajal, F., Pecor, J.E., Klein, T.A. (2008) 'Seasonal distribution, biology, and human attraction patterns of mosquitoes (Diptera: Culicidae) in a rural village and adjacent forrested site near Iquitos, Peru', *Journal of Medical Entomology*, 45(6), 1165–1172.

Vaidyanathan, R., Scott, T.W. (2007) 'Geographic Variation in Vector Competence for West Nile Virus in the *Culex pipiens* (Diptera: Culicidae) Complex in California', *Vector-Borne and Zoonotic Diseases*, 7(2), 193–198.

Vardoulakis, S., Heaviside, C. (2012) *Health Effects of Climate Change in the UK 2012*, London.

Vaux, A.G.C., Gibson, G., Hernández-Triana, L.M., Cheke, R. a, McCracken, F., Jeffries, C.L., Horton, D.L., Springate, S., Johnson, N., Fooks, A.R., Leach, S.A., Medlock, J.M. (2015) 'Enhanced West Nile virus surveillance in the North Kent marshes, UK', *Parasites & Vectors*, 8(1).

Vaux, A.G.C., Medlock, J.M. (2015) 'Current status of invasive mosquito surveillance in the UK.', *Parasites & vectors*, 8(JUNE), 351.

Verhulst, N.O., Beijleveld, H., Knols, B.G.J., Takken, W., Schraa, G., Bouwmeester, H.J., Smallegange, R.C. (2009) 'Cultured skin microbiota attracts malaria mosquitoes.', *Malaria journal*, 8, 302.

Verhulst, N.O., Mbadi, P.A., Kiss, G.B., Mukabana, W.R., van Loon, J.J.A., Takken, W., Smallegange, R.C. (2011) 'Improvement of a synthetic lure for *Anopheles gambiae* using compounds produced by human skin microbiota.', *Malaria journal*, 10(28), 1–9.

Versteirt, V., Boyer, S., Damiens, D., De Clercq, E.M., Dekoninck, W., Ducheyne, E., Grootaert, P., Garros, C., Hance, T., Hendrickx, G., Coosemans, M., Van Bortel, W. (2013) 'Nationwide inventory of mosquito biodiversity (Diptera: Culicidae) in Belgium, Europe', *Bulletin of Entomological Research*, 103(2), 193–203.

Vinauger, C., Lutz, E.K., Riffell, J.A. (2014) 'Olfactory learning and memory in the disease vector mosquito *Aedes aegypti*', *The Journal of Experimental Biology*, (217), 2321–2330.

Vinogradova, E.B. (2003) 'Ecophysiological and morphological variations in mosquitoes of the *Culex pipiens* complex (Diptera: Culicidae)', *Acta Societatis Zoologicae Bohemicae*, 67, 41–50.

Vinogradova, E.B., Shaikevich, E. V (2007) 'Morphometric , physiological and molecular characteristics of underground populations of the urban mosquito *Culex pipiens* Linnaeus f . *molestus* Forskål (Diptera : Culicidae) from several areas of Russia', *European Mosquito Bulletin*, 22, 17–24.

Vogels, C.B., Göertz, G.P., Pijlman, G.P., Koenraadt, C.J.M. (2017) 'Vector competence of northern and southern European Culex pipiens pipiens mosquitoes for West Nile virus across a gradient of temperatures', *Medical and Veterinary Entomology*, 31(4), 358–364.

Vosshall, L.B., Hansson, B.S. (2011) 'A unified nomenclature system for the insect olfactory coreceptor', *Chemical Senses*, 36(6), 497–498.

Vu, T.N., Laukens, K. (2013) 'Getting Your Peaks in Line: A Review of Alignment Methods for NMR Spectral Data', *Metabolites*, 3(2), 259–276.

Vu, T.N., Valkenborg, D., Smets, K., Verwaest, K. a, Dommisse, R., Lemière, F., Verschoren, A., Goethals, B., Laukens, K. (2011) 'An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data', *BMC Bioinformatics*, 12(1), 405.

Wade, M.R., Wratten, S.D. (2007) 'Excised or intact inflorescences? Methodological effects on parasitoid wasp longevity', *Biological Control*, 40(3), 347–354.

Walker, A., Crosby, T. (1988) *The Preparation and Curation of Insects*, New revise. ed, Science Information Publishing Centre: Wellington.

Walsh, P.S., Metzger, D.A., Higuchi, R. (1991) 'Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material', *Biotechniques*, 10(4), 506–513.

Wang, B., Liu, Y., He, K., Wang, G. (2016) 'Comparison of research methods for functional characterization of insect olfactory receptors', *Scientific Reports*, 6(1), 32806.

Wcislo, W.T. (1989) 'Behavioral Environments and Evolutionary Change', *Annual Review of Ecology and Systematics*, 20, 137–169.

Wedge, D.E., Klun, J. a., Tabanca, N., Demirci, B., Ozek, T., Baser, K.H.C., Liu, Z., Zhang, S., Cantrell, C.L., Zhang, J. (2009) 'Bioactivity-guided fractionation and GC/MS fingerprinting of *Angelica sinensis* and *Angelica archangelica* root components for antifungal and mosquito deterrent activity', *Journal of Agricultural and Food Chemistry*, 57(2), 464–470.

Werblow, A., Klimpel, S., Bolius, S., Dorresteijn, A.W.C., Sauer, J., Melaun, C. (2014) 'Population Structure and Distribution Patterns of the Sibling Mosquito Species *Culex pipiens* and *Culex torrentium* (Diptera: Culicidae) Reveal Different Evolutionary Paths.', *PloS one*, 9(7), e102158.

Werren, J.H., Baldo, L., Clark, M.E. (2008) 'Wolbachia: master manipulators of invertebrate biology', *Nature Reviews Microbiology*, 6(10), 741–751.

Wessnitzer, J., Webb, B. (2006) 'Multimodal sensory integration in insects—towards insect brain control architectures', *Bioinspiration & Biomimetics*, 1(3), 63–75.

WHO (2016) Mosquito-Borne Diseases [online], *WHO*, available: http://www.who.int/neglected_diseases/vector_ecology/mosquito-borne-diseases/en/ [accessed 6 Jul 2017].

WHO (2017) *Global Vector Control Response: 2017-2030*, Geneva.

Whyard, S., Erdelyan, C., Partridge, A.L., Singh, A.D., Beebe, N.W., Capina, R. (2015) 'Silencing the buzz: a new approach to population suppression of mosquitoes by feeding larvae double-stranded RNAs', *Parasites & Vectors*, 8(1), 96.

Wickham, H., Francois, R. (2016) 'dplyr: A Grammar of Data Manipulation'.

Wilke, A.B.B., De Oliveira Christe, R., Multini, L.C., Vidal, P.O., Wilk-Da-silva, R., De Carvalho, G.C., Marrelli, M.T. (2016) 'Morphometric wing characters as a tool for mosquito identification', *PLoS ONE*, 11(8), 1–12.

Wilkerson, R.C., Linton, Y.-M., Fonseca, D.M., Schultz, T.R., Price, D.C., Strickman, D.A. (2015) 'Making mosquito taxonomy useful: A stable classification of tribe Aedini that balances utility with current knowledge of evolutionary relationships', *PLoS ONE*, 10(7), 1–26.

Woolfenden, E. (2010a) 'Sorbent-based sampling methods for volatile and semi-volatile organic compounds in air. Part 1: Sorbent-based air monitoring options', *Journal of Chromatography A*, 1217(16), 2674–2684.

Woolfenden, E. (2010b) 'Sorbent-based sampling methods for volatile and semi-volatile organic compounds in air. Part 2. Sorbent selection and other aspects of optimizing air monitoring methods', *Journal of Chromatography A*, 1217(16), 2685–2694.

Wuriyanghan, H., Rosa, C., Falk, B.W. (2011) 'Oral delivery of double-stranded RNAs and siRNAs induces RNAi effects in the potato/tomato psyllid, Bactericerca cockerelli', *PLoS ONE*, 6(11).

Yahyaa, M., Tholl, D., Cormier, G., Jensen, R., Simon, P.W., Ibdah, M. (2015) 'Identification and Characterization of Terpene Synthases Potentially Involved in the Formation of Volatile Terpenes in Carrot (*Daucus carota* L.) Roots', *Journal of Agricultural and Food Chemistry*, 63(19), 4870–4878.

Yates, M.G. (1979) 'The biology of the tree-hole breeding mosquito *Aedes geniculatus* (Olivier) (Diptera: Culicidae) in southern England', *Bulletin of Entomological Research*, 69(04), 611–628.

Ye, J., Coulouris, G., Zaratskaya, I., Cutcutache, I., Rozen, S., Madden, T. (2012) 'Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction.', *BMC Bioinformatics*, 13, 134.

Yee, W.L., Foster, W.A. (1992) 'Diel Sugar-Feeding and Host-Seeking Rhythms in Mosquitoes (Diptera : Culicidae) Under Laboratory Conditions', *Journal of Medical Entomology*, 29(5), 784–791.

Yee, W.L., Foster, W.A., Howe, M.J., Hancock, R.G. (1992) 'Simultaneous Field Comparison of Evening Temporal Distributions of Nectar and Blood Feeding By *Aedes vexans* and *Ae trivittatus* (Diptera, Culicidae) in Ohio', *Journal of Medical Entomology*, 29(2), 356–360.

Yin, J., Choo, Y.-M., Duan, H., Leal, W.S. (2015) 'Selectivity of odorant-binding proteins from the southern house mosquito tested against physiologically relevant ligands', *Frontiers in Physiology*, 6(February), 1–8.

Yuval, B., Holliday-Hanson, M.L., Washino, R.K. (1994) 'Energy budget of swarming male mosquitoes', *Ecological Entomology*, 19(1), 74–78.

Yuval, B., Wekesa, J.W., Washino, R.K. (1993) 'Effect of body size on swarming behaviour and mating success of male *Anopheles freeborni* (Diptera: Culicidae)', *Journal of Insect Behavior*, 6(3), 155–158.

Zelditch, M.L., Swiderski, D.L., Sheets, H.D. (2004) *Geometric Morphometrics for Biologists: A Primer*, 2nd ed, Elsevier (USA): Oxford.

Zuur, A.F., Ieno, E.N., Elphick, C.S. (2010) 'A protocol for data exploration to avoid common statistical problems', *Methods in Ecology and Evolution*, 1(1), 3–14.

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M. (2009) *Mixed Effects Models and Extensions in Ecology with R*.

Zwiebel, L.J., Takken, W. (2004) 'Olfactory regulation of Mosquito - host interactions', *Insect Biochemistry and Molecular Biology*, 34(7), 645–652.

Zych, M. (2007) 'On flower visitors and true pollinators: The case of protandrous *Heracleum sphondylium* L. (Apiaceae)', *Plant Systematics and Evolution*, 263(3–4), 159–179.

8 Appendices

Appendix 1 Primer blast reports

Primer blast reports for primers from (Rudolf *et al.* 2013).

>DQ470148.1 *Culex pipiens* pipiens strain 258c microsatellite CQ11 sequence

product length = 163
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22
Reverse primer 1 CGTCCTCAAACATCCAGACA 20
Template 165 146

>DQ470147.1 *Culex pipiens* pipiens strain 258b microsatellite CQ11 sequence

product length = 163
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22
Reverse primer 1 CGTCCTCAAACATCCAGACA 20
Template 165 146

>DQ470146.1 *Culex pipiens* pipiens strain 258a microsatellite CQ11 sequence

product length = 163
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22
Reverse primer 1 CGTCCTCAAACATCCAGACA 20
Template 165 146

>DQ470145.1 *Culex pipiens* pipiens strain 260b microsatellite CQ11 sequence

product length = 165
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22
Reverse primer 1 CGTCCTCAAACATCCAGACA 20
Template 167 148

>DQ470144.1 *Culex pipiens* pipiens strain 260a microsatellite CQ11 sequence

product length = 165
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22
Reverse primer 1 CGTCCTCAAACATCCAGACA 20
Template 167 148

>DQ470143.1 *Culex pipiens* pipiens strain 262 microsatellite CQ11 sequence

product length = 167
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22
Reverse primer 1 CGTCCTCAAACATCCAGACA 20
Template 169 150

>DQ470142.1 *Culex pipiens* pipiens strain 264b microsatellite CQ11 sequence

product length = 169
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22
Reverse primer 1 CGTCCTCAAACATCCAGACA 20
Template 171 152

>DQ470141.1 *Culex pipiens* pipiens strain 264a microsatellite CQ11 sequence

product length = 169
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22
Reverse primer 1 CGTCCTCAAACATCCAGACA 20
Template 171 152

>DQ470140.1 *Culex pipiens* pipiens strain 266 microsatellite CQ11 sequence

product length = 171
Forward primer 1 GCGGCCAAATATTGAGACTT 20
Template 3 22

Reverse primer 1 CGTCCTCAAACATCCAGACA 20

Template 173 154

>AY497525.1 *Culex torrentium* acetylcholinesterase (Ace2) gene, partial cds

product length = 116
Forward primer 1 GACACAGGACGACAGAAA 18
Template 86 103
Reverse primer 1 GCCTACGCAACTACTAAA 18
Template 201 184

Appendix 2 An example of the code for optimisation and analysis of the number of landmarks required for best operation of the geometric morphometric species discrimination model.

```
---
title: "Female *Culex* wing morphometrics 18 points both wings"
author: "Rich Halfpenny"
date: "14 December 2016"
output:
html_document: default
word_document:
fig_height: 5
fig_width: 7
---

##Declaration of changed variables in this document

```{r, this contains all of variables that need to be edited for different approaches}
#edit these variables as needed

description <- "Female Culex wing morphometrics both wings 18 points"
num_lms <- 18


whichsex <- "Female"



#sql for the training data

qry1 <- "SELECT Wing_morphometrics_LM.ID1, Wing_morphometrics_LM.ID,
    Wing_morphometrics_LM.COMMENT, Wing_morphometrics_LM.VAR, Wing_morphometrics_LM.x1,
    Wing_morphometrics_LM.y1, Wing_morphometrics_LM.x2, Wing_morphometrics_LM.y2,
    Wing_morphometrics_LM.x3, Wing_morphometrics_LM.y3, Wing_morphometrics_LM.x4,
    Wing_morphometrics_LM.y4, Wing_morphometrics_LM.x5, Wing_morphometrics_LM.y5,
    Wing_morphometrics_LM.x6, Wing_morphometrics_LM.y6, Wing_morphometrics_LM.x7,
    Wing_morphometrics_LM.y7, Wing_morphometrics_LM.x8, Wing_morphometrics_LM.y8,
    Wing_morphometrics_LM.x9, Wing_morphometrics_LM.y9, Wing_morphometrics_LM.x10,
    Wing_morphometrics_LM.y10, Wing_morphometrics_LM.x11, Wing_morphometrics_LM.y11,
    Wing_morphometrics_LM.x12, Wing_morphometrics_LM.y12, Wing_morphometrics_LM.x13,
    Wing_morphometrics_LM.y13, Wing_morphometrics_LM.x14, Wing_morphometrics_LM.y14,
    Wing_morphometrics_LM.x15, Wing_morphometrics_LM.y15, Wing_morphometrics_LM.x16,
    Wing_morphometrics_LM.y16, Wing_morphometrics_LM.x17, Wing_morphometrics_LM.y17,
    Wing_morphometrics_LM.x18, Wing_morphometrics_LM.y18
FROM Wing_morphometrics_LM;
"
#sql for the known unknowns

```
```

```
qry2 <- "SELECT Wing_morphometrics_LM_FM.ID1, Wing_morphometrics_LM_FM.ID,
    Wing_morphometrics_LM_FM.COMMENT, Wing_morphometrics_LM_FM.VAR,
    Wing_morphometrics_LM_FM.x1, Wing_morphometrics_LM_FM.y1,
    Wing_morphometrics_LM_FM.x2, Wing_morphometrics_LM_FM.y2,
    Wing_morphometrics_LM_FM.x3, Wing_morphometrics_LM_FM.y3,
    Wing_morphometrics_LM_FM.x4, Wing_morphometrics_LM_FM.y4,
    Wing_morphometrics_LM_FM.x5, Wing_morphometrics_LM_FM.y5,
    Wing_morphometrics_LM_FM.x6, Wing_morphometrics_LM_FM.y6,
    Wing_morphometrics_LM_FM.x7, Wing_morphometrics_LM_FM.y7,
    Wing_morphometrics_LM_FM.x8, Wing_morphometrics_LM_FM.y8,
    Wing_morphometrics_LM_FM.x9, Wing_morphometrics_LM_FM.y9,
    Wing_morphometrics_LM_FM.x10, Wing_morphometrics_LM_FM.y10,
    Wing_morphometrics_LM_FM.x11, Wing_morphometrics_LM_FM.y11,
    Wing_morphometrics_LM_FM.x12, Wing_morphometrics_LM_FM.y12,
    Wing_morphometrics_LM_FM.x13, Wing_morphometrics_LM_FM.y13,
    Wing_morphometrics_LM_FM.x14, Wing_morphometrics_LM_FM.y14,
    Wing_morphometrics_LM_FM.x15, Wing_morphometrics_LM_FM.y15,
    Wing_morphometrics_LM_FM.x16, Wing_morphometrics_LM_FM.y16,
    Wing_morphometrics_LM_FM.x17, Wing_morphometrics_LM_FM.y17,
    Wing_morphometrics_LM_FM.x18, Wing_morphometrics_LM_FM.y18
FROM Wing_morphometrics_LM_FM;"

```

```{r packages, echo=FALSE}
suppressMessages(library(geomorph))
suppressMessages(library(RODBC))
suppressMessages(library(dplyr))
suppressMessages(library(adegenet))
```

```{r get data, echo=FALSE}

con <- odbcConnect("MDnew")


fromdb1 <- sqlQuery(con, qry1, stringsAsFactors = FALSE)
fromdb1[, (5:40)] <- fromdb1[, (5:40)] * -1 # this is just to flip the data over so the wings are not upside down in the visualisation
fromdb <- fromdb1 # this is just a fudge for those sections of the data which do not have  criteria to match and so don't use the fromdb1 source
#how to collect the classifier data in a separate array - geomorph does not expect to see these in the landmark data

    classifiers <- fromdb1[, 2:4]
    species <- classifiers[, 1]
    sex <- classifiers [, 2]
    whichwing <- classifiers[, 3]
```

```
61
62
63 #remove the classifiers from the raw data a create a landmark file 64
65 coords <- as.matrix(fromdb1[, -(1:4)])
66
67
68 coords <- arrayspecs(coords, num_lms, 2)
69
70
71
72 ```
73
74 ##Perform Generalised Procrustes Analysis with Partial Procrustes Superimposition 75
76 ```{r adegenet get data and remove outliers from females, echo=FALSE}
77
78
79 fromdb <- filter(fromdb1, COMMENT == whichsex )
80
81
82
83 classifiers <- fromdb[, 2:4]
84
85     species <- classifiers[, 1]
86     sex <- classifiers [, 2]
87     whichwing <- classifiers[, 3]
88
89
90 coords <- as.matrix(fromdb[, -(1:4)])
91
92 coords <- arrayspecs(coords, num_lms, 2) #Convert landmark data to a two dimensional array
93
94 y <- gpagen(coords, PrinAxes = TRUE, print.progress = FALSE) #Conduct GPA
95
96     summary(y)
97     ```
98
99      ## Plot of the landmark data after GPA process. Black spots are the mean position for each landmark, grey
        points are individual landmarks
100     ```{r, echo=FALSE}
101
102 plot(y)
103
104 ```
105
106 #Plot of the Procrustes distance from the Mean. Red values indicate potential outliers to the data ( > upper quartile)
```

```{r, echo=FALSE}

plotOutliers(y$coords, groups = NULL)


outliers <- plotOutliers(y$coords)

```

###Largest outlier deformation

    ```{r, echo=FALSE}
    plotRefToTarget(mshape(y$coords), y$coords[,, outliers[1]], method = "vector", label = T)
    ```

2nd Largest outlier deformation

    ```{r, echo=FALSE}
    plotRefToTarget(mshape(y$coords), y$coords[,, outliers[2]], method = "vector", label = T)
    ```

3rd Largest outlier deformation

    ```{r, echo=FALSE}
    plotRefToTarget(mshape(y$coords), y$coords[,, outliers[3]], method = "vector", label = T)
    ```

4th Largest outlier deformation

    ```{r, echo=FALSE}
    plotRefToTarget(mshape(y$coords), y$coords[,, outliers[4]], method = "vector", label = T)
    ```

###Outliers can be removed

```{r}

coords <- as.matrix(fromdb[-c(113, 46), -(1:4)])

classifiers <- fromdb[-c(113, 46), 2:4]

```

```{r, echo=FALSE}


#fromdb <- fromdb[-c(113, 46), ]


    species <- classifiers[, 1]
    sex <- classifiers [, 2]
    whichwing <- classifiers[, 3]

coords <- arrayspecs(coords, num_lms, 2)

y <- gpagen(coords, PrinAxes = TRUE, print.progress = FALSE)
coords <- two.d.array(y$coords)

```

#Test the assumption that the landmark data for *Culex pipiens* and *Culex torrentium* are actually different

```{r quantify between types, echo=FALSE}

set.seed(123)  # set the starting conditions for randomisation
morphol.disparity(y$coords ~ 1, groups = ~ species, iter = 9999, print.progress =
                                  FALSE) # estimate morphological disparity and perform pairwise comparison among groups


```

```{r visualisation, echo=FALSE}

    species <- as.factor(species)
    col.gp <- rainbow(length(levels(species)))
    names(col.gp) <- levels(species)
    col.gp <- col.gp[match(species, names(col.gp))]
    ```

##Visualistion of the separation of these species on the first 2 Principal Components

```{r, echo=FALSE}

PCA <- plotTangentSpace(y$coords, groups = col.gp, legend = TRUE)
```

```

##The weighting the principal components

Although there are some components that are more influential it is obvious that there are number of potentially influential components.

```{r, echo=FALSE}

pvar <- (PCA$sdev^2) / (sum(PCA$sdev^2))
names(pvar) <- seq(1:length(pvar))
barplot(pvar, xlab = "Principal Components", ylab = "% Variance", main = "Principal components weighting")

```

```{r plot an individual reference to a target specimen, echo=FALSE}
ref <- mshape(y$coords) # this gets the mean shape data for all the coordinates
```

###Reference plot of a *Culex pipiens s.s.* wing compared to the group mean for both species

```{r, echo=FALSE}

plotRefToTarget(ref, y$coords[,, 1], mag = 1, method = "TPS")  # this is useful for examining the difference in the shape of an individual - mag = 3 makes the magnitude of the difference 3 times greater for ease of inference.

```

###Reference plot of a *Culex torrentium* wing compared to the group mean for both species

```{r, echo=FALSE}

plotRefToTarget(ref, y$coords[,, 195], mag = 1, method = "TPS")  # this is useful for examining the difference in the shape of an individual - mag = 3 makes the magnitude of the difference 3 times greater for ease of inference.

```

##Selection and optimisation of PCs for species discrimination

Having demonstrated that there is a significant difference between these species, the optimisation of the method needs to take place. Using too many PCs can create a

model which can fit the training data very well, but may be unstable, using too few

PCs can result in the discarding of major discriminatory power.

First, an iterative process using Cross Validation of DAPC will be used to find the number of PCs which best match the training data set.

Second, an approach using alpha score optimisation will be used to measure the ratio of successful group assignment to the training data and values obtained through random discrimination. It is the proportion of successful reassignment corrected for the number of retained PCs. Multiple DAPC analyses are carried out using random group; alpha scores are calculated for each group as are average alpha scores. (Jombart & Collins: A tutorial for Discriminant Analysis of Principal Components
```

(DAPC) using *adegenet* 2.0.2 (2016))

Third, a number of PCs are simply chosen following the inspection of the loadings data and the PC scree graph.

```{r adegenet DAPC females, echo=FALSE}

set.seed(123)

system.time( xvalfemales <- xvalDapc(coords, species, n.pca.max = 100, n.da = NULL,
training.set = 0.9, result = "overall",
center = FALSE, scale = FALSE,
n.pca = 2:30, n.rep = 1000, xval.plot = TRUE, parallel = "snow", ncpus
= 4))

```

Result of the longform results from the Cross Validation DAPC

```{r, echo=FALSE}

xvalfemales[-1]

```

Box and whisker plot of the confidence of correct assignment by number of PCs

```{r, echo=FALSE}
boxplot(xvalfemales$`Cross-Validation Results`$success ~ xvalfemales$`Cross-Validation Results`$n.pca, xlab = " Number of PCA components", main = "Confidence in correct species assignment to PCA number", ylab = "Probability
of correct assignment")
```

Plot of the confidence of individuals to species, with DAPC calculated on the number of PC from xval DAPC method

```{r, echo=FALSE}

dapc1 <- dapc(coords, grp = species, n.pca = as.numeric(xvalfemales$`Number of PCs Achieving Highest Mean Success`), n.da = 1)

compoplot(dapc1, lab = "", xlab = "Individuals")

```

Scatter plot of the single discriminant function between species, showing the separation of these species.

```{r, echo=FALSE}
```

```
                scatter.dapc(dapc1, legend = TRUE, posi.leg = "topright", scree.pca = TRUE,
                posi.pca = "topleft", cleg = 0.75, xax = 1, yax = 2, ratio.pca = 0.2)
                ```
```

## Assignment heatmap of the species identification. Blue crosses indicate actual identification. White = 0 % confidence, Red = 100 %.

```{r, echo= FALSE}

assignplot(dapc1)

```

Loading plot of the individual input variables (the transformed (by GPA) x and y coordinates which make a landmark), allowing the identification of the actual variables which are most important in the discrimination of species.

```{r, echo= FALSE}

loadingplot(dapc1$var.contr, main = "Loading plot of the influence of named variables")


```

Detailed visualisation of the individuals that are not assigned to the correct species with >90 % confidence

```{r lets look at those that donot fit a group well female wings, echo=FALSE}

    temp <- 0
    temp <- which(apply(dapc1$posterior, 1, function(e) all(e < 0.9)))
    #Specimens which are not at least 90% for correct group:
    temp <- rbind(temp, 0)

        suppressWarnings(
        if (temp >= 1) {
        compoplot(dapc1, subset = temp, posi = "bottomright", txt.leg = c("Cx. pipiens",
        "Cx.torrentium"), ncol = 2)
        } else {
        print("All specimens at least 90% correct for group")
        })

rm(temp)

```

The method above gives a very high percentage of fit for the training data, with an almost perfect fit. - This does not necessarily mean that it will be the best

predictive model as it might be unstable
```

```
# Import known unknowns, using SQL (qry2) to test the model's predictive power
```{r collect the data for unknowns, echo=FALSE}


fromdb2 <- sqlQuery(con, qry2, stringsAsFactors = FALSE)

fromdb2 <- filter(fromdb2, COMMENT == whichsex )

    fromdb2[, (5:40)] <- fromdb2[, (5:40)] * -1 # this is just to flip the data over so the wings are not upside down in the visualisation
    unknown <- fromdb2 # this is just a fudge for those sections of the data which do not have  criteria to match and so don't use the fromdb1 source
    #how to collect the classifier data in a separate array - geomorph does not expect to see these in the landmark data

    uclassifiers <- fromdb2[, 2:4]
    uspecies <- uclassifiers[, 1]
    usex <- uclassifiers [, 2]
    uwhichwing <- uclassifiers[, 3]


#remove the classifiers from the raw data a create a landmark file
ucoords <- as.matrix(fromdb2[, -(1:4)])
ucoords <- arrayspecs(ucoords, num_lms, 2)
y <- gpagen(ucoords, PrinAxes = TRUE, print.progress = FALSE)
ucoords <- two.d.array(y$coords)

```

Scatter plot of the single discriminant function between species, showing the separation of these species, overlaid with the new known unknown individuals (circles)

```{r predict species, echo=FALSE}

specify <- predict(dapc1, ucoords)

    scatter(dapc1, scree.da = TRUE)
    points(specify$ind.scores, rep(0.1, 77))   # the last number here needs to match the number of "unknowns being plotted"
    ```

Assignment heatmap of the "known unknowns" species identification. White = 0 % confidence, Red = 100 %.

    ```{r, echo= FALSE}
    assignplot(dapc1, new.pred = specify, only.grp = "unknown" )
    title("78 known unknowns - first 41 pipiens, 2nd 37 torrentium")
    ```
```

```
## Plot of the confidence of individual "known unknowns" identification to species, with DAPC calculated on the number of PC from xval DAPC method

    ```{r, echo= FALSE}
    compoplot(dapc1, new.pred = specify, ncol = 2, only.grp = "unknown")
    title("78 known unknowns - first 41 pipiens, 2nd 37 torrentium")

```

```{r, round the identification confidence and begin data tidy up, echo=FALSE}
assigned_ids <- specify$posterior


assignedround <- round(assigned_ids, digits = 2)

colnames(assignedround)[colnames(assignedround) == "Cx. pipiens"] <- "Identification"
# round the data and make into english

    assignedround <- assignedround[, -2]
    assignedround[assignedround > .50] <- "*Culex pipiens*"
    assignedround[assignedround == .50] <- "No Choice"
    assignedround[assignedround < .50] <- "*Culex torrentium*"


```

Table of the Predicted species and the actual species for each known unknown individual

```{r, compile the correct id list compare and report, echo=FALSE}
#correct id list from the database prior identification


uspecies <- gsub(pattern = "\\d : UnkPipiens$", replacement = "*Culex pipiens*", uspecies, ignore.case = F)

    uspecies <- gsub(pattern = "\\d : UnkTorrentium$", replacement = "*Culex torrentium*", uspecies, ignore.case = F)
    uspecies <- gsub(pattern = "\\d", replacement = "", uspecies, ignore.case = F)


#join the 2 objects

compare <- cbind(assignedround, uspecies)


```
```

```
colnames(compare) <- c("Predicted id", "Actual id")

#compare the data sets and declare whether they are correct or not
    compare <- as.data.frame(compare)
    compare[, "Correct_id"] <- NA

    #fix for factor levels not matching
    compare[, 2] <- factor(compare[, 2], levels = levels(compare[, 1]))
            for (i in seq(along = compare[, 1])){
            ifelse(compare[i, 1] == compare[i, 2], Correct <- TRUE, Correct <- FALSE )
            Correct <- as.vector(Correct)
            compare[i, 3] <- Correct
            }

compare


```

Correct identifications of "known unknowns"

```{r, echo= FALSE}

correct_id <- plyr::count(compare$Correct_id)

    correct_id
    ```

percentage correct identifcation precentage rate

    ```{r, echo=FALSE}
    accuracy_percent <- round(correct_id[2, 2] / (correct_id[1, 2] + correct_id[2, 2] )
    * 100, digits = 2)

accuracy_percent

output <- as.data.frame(cbind(description, accuracy_percent))

```

This section saves the results to an Access database to allow the investigation of the method's precision

```{r send results to Access, echo= FALSE}

        #Send it to access
        # save to a new table called Wingmorph_data, if the table already exists then add it to the end of the table that already exists.
        sqlSave(con, output, tablename = "Wingmorph_data", rownames = FALSE, append = TRUE, safer = TRUE, fast = FALSE)
```

```

##using the alpha score to decide which PCs to keep

Optimise the number of PCs based on the alpha score using optim.a.score function. Here the calculated dapc1 object is used as it has already been calculated using the recommended number of PCs from the cross-validation method. Remember, the cross validation makes the best model for predicting the training data and so tends to retain lots of PCs

```{r, alpha score section}

dapc1 <- dapc(coords, grp = species, n.pca = 30, n.da = 1)

# run optimisation

ascoreresult <- optim.a.score(dapc1, n.pca = 1:30) # optim.a.score is still in development, and so might not be great - however, it is being used to indicate  starting point which would subsequently be tested empirically.

# redo dapc this time using the number of pcs calculated using optim.a.score

dapc1 <- dapc(coords, grp = species, n.pca = as.numeric(ascoreresult$best), n.da = 1)

```

Plot of the confidence of individual identification to species, with DAPC calculated on the number of PC from Alpha score optimisation method

```{r, echo=FALSE}
#visualise these results on the fit of the training data

compoplot(dapc1, lab = "", xlab = "Individuals")

```

Scatter plot of the single discriminant function between species, showing the separation of these species.

```{r, echo=FALSE}

scatter.dapc(dapc1, legend = TRUE, posi.leg = "topright", scree.pca = TRUE,
posi.pca = "topleft", cleg = 0.75, xax = 1, yax = 2, ratio.pca = 0.2)

```

Assignment heatmap of the species identification. Blue crosses indicate actual identification. White = 0 % confidence, Red = 100 %.
```

```{r, echo=FALSE}

assignplot(dapc1)

```

```{r, echo= FALSE}

loadingplot(dapc1$var.contr, main = "Loading plot of the influence of named variables")

```

Detailed visualisation of the individuals that are not assigned to the correct species with >90 % confidence

```{r lets look at those that donot fit a group well female wings alpha score section, echo=FALSE}
temp <- 0
temp <- which(apply(dapc1$posterior,1, function(e) all(e < 0.9)))
#Specimens which are not at least 90% for correct group:
temp <- rbind(temp, 0)

    suppressWarnings(
    if (temp >= 1) {
    compoplot(dapc1, subset = temp, posi = "bottomright", txt.leg = c("Cx. pipiens",
    "Cx.torrentium"), ncol = 2)
    } else {
    print("All specimens at least 90% correct for group")
    })

rm(temp)

```

As can be seen, this model fits the data less well than the higher PC model, but still describes the data quite well. How does it perform on the known unknowns as a predictor?

```{r predict species alpha score section, echo=FALSE}

specify <- predict(dapc1, ucoords)

```

Scatter plot of the single discriminant function between species, showing the separation of these species, overlaid with the new "known unknown" individuals

(circles)

```{r, echo= FALSE}
scatter(dapc1, scree.da = TRUE)
points(specify$ind.scores, rep(0.1, 77))   # the last number here needs to match the number of "unknowns being plotted"
```

Assignment heatmap of the "known unknowns" species identification. White = 0 % confidence, Red = 100 %.

```{r, echo= FALSE}

assignplot(dapc1, new.pred = specify, only.grp = "unknown" )
title("78 known unknowns - first 41 pipiens, 2nd 37 torrentium")

```

Plot of the confidence of individual "known unknowns" identification to species, with DAPC calculated on the number of PC from alpha score DAPC method

```{r, echo= FALSE}
compoplot(dapc1, new.pred = specify, ncol = 2, only.grp = "unknown")
title("78 known unknowns - first 41 pipiens, 2nd 37 torrentium")

```

```{r, round the identification confidence and begin data tidy up alpha score section, echo=FALSE}

assigned_ids <- specify$posterior


assignedround <- round(assigned_ids, digits = 2)

colnames(assignedround)[colnames(assignedround) == "Cx. pipiens"] <- "Identification"
# round the data and make into english

assignedround <- assignedround[, -2]

assignedround[assignedround > .50] <- "*Culex pipiens*"
assignedround[assignedround < .50] <- "*Culex torrentium*"


```

Table of the Predicted species and the actual species for each known unknown individual

```{r, compile the correct id list compare and report alpha score section, echo=FALSE}
#correct id list from the database prior identification


uspecies <- gsub(pattern = "\\d : UnkPipiens$", replacement = "*Culex pipiens*", uspecies, ignore.case = F)

    uspecies <- gsub(pattern = "\\d : UnkTorrentium$", replacement = "*Culex torrentium*", uspecies, ignore.case = F)
    uspecies <- gsub(pattern = "\\d", replacement = "", uspecies, ignore.case = F)




#join the 2 objects

compare <- cbind(assignedround, uspecies)


colnames(compare) <- c("Predicted id", "Actual id")

#compare the data sets and declare whether they are correct or not
    compare <- as.data.frame(compare)
    compare[, "Correct_id"] <- NA


    #fix for factor levels not matching
    compare[, 2] <- factor(compare[, 2], levels = levels(compare[, 1]))


            for (i in seq(along = compare[, 1])){
            ifelse(compare[i, 1] == compare[i, 2], Correct <- TRUE, Correct <- FALSE )
            Correct <- as.vector(Correct)
            compare[i, 3] <- Correct
            }

    compare
    ```

Correct identifications of "known unknowns"

    ```{r, echo= FALSE}
    correct_id <- plyr::count(compare$Correct_id)

    correct_id
    ```

percentage correct identification precentage rate

```{r, echo= FALSE}
```

```
accuracy_percent <- round(correct_id[2, 2] / (correct_id[1, 2] + correct_id[2, 2] )
    * 100, digits = 2)

accuracy_percent


output <- as.data.frame(cbind(description, accuracy_percent))



```

    ```{r send results to Access alpha score section}
    ## This section saves the results to an Access database to allow the investigation of the method's precision

        #Send it to access
        # save to a new table called Wingmorph_data, if the table already exists then add it to the end of the table that already exists.
        sqlSave(con, output, tablename = "Wingmorph_data_alpha_optimised", rownames = FALSE, append = TRUE, safer = TRUE, fast = FALSE)


```

using a user defined number of PCs (n = 9)

Plot of the confidence of individual identification to species, with DAPC calculated on the number of 9 PCs

```{r, 9 PCs section}

dapc1 <- dapc(coords, grp = species, n.pca = 9, n.da = 1)
#visualise these results on the fit of the training data

compoplot(dapc1, lab = "", xlab = "Individuals")

```

Scatter plot of the single discriminant function between species, showing the separation of these species.

```{r, echo= FALSE}

            scatter.dapc(dapc1, legend = TRUE, posi.leg = "topright", scree.pca = TRUE,
            posi.pca = "topleft", cleg = 0.75, xax = 1, yax = 2, ratio.pca = 0.2)
            ```

Assignment heatmap of the species identification. Blue crosses indicate actual identification. White = 0 % confidence, Red = 100 %.

```{r, echo=FALSE}

assignplot(dapc1)

```

```{r, echo= FALSE}

loadingplot(dapc1$var.contr, main = "Loading plot of the influence of named variables")

```

Detailed visualisation of the individuals that are not assigned to the correct species with >90 % confidence

```{r lets look at those that donot fit a group well female wings 9 PCs section, echo=FALSE}
temp <- 0
temp <- which(apply(dapc1$posterior, 1, function(e) all(e < 0.9)))
#Specimens which are not at least 90% for correct group:
temp <- rbind(temp, 0)

    suppressWarnings(
    if (temp >= 1) {
    compoplot(dapc1, subset = temp, posi = "bottomright", txt.leg = c("Cx. pipiens",
    "Cx.torrentium"), ncol = 2)
    } else {
    print("All specimens at least 90% correct for group")
    })

rm(temp)

```

As can be seen, this model fits the data less well than the higher PC model, but still describes the data quite well, and is better than the very small number of pcs identified using aplha score. How does it perform on the known unknowns as a predictor?

```{r predict species 9 PCs section, echo=FALSE}

specify <- predict(dapc1, ucoords)
```

Scatter plot of the single discriminant function between species, showing the separation of these species, overlaid with the new "known unknown" individuals (circles)

```{r, echo= FALSE}
scatter(dapc1, scree.da = TRUE)
points(specify$ind.scores, rep(0.1, 77))   # the last number here needs to match the number of "unknowns being plotted"
```

Assignment heatmap of the "known unknowns" species identification. White = 0 % confidence, Red = 100 %.

```{r, echo= FALSE}
assignplot(dapc1, new.pred = specify, only.grp = "unknown" )
title("38 known unknowns - first 20 pipiens, 2nd 18 torrentium")
```

Plot of the confidence of individual "known unknowns" identification to species, with DAPC calculated on the number of 9 PCs

```{r, echo= FALSE}
compoplot(dapc1, new.pred = specify, ncol = 2, only.grp = "unknown")
title("38 known unknowns - first 20 pipiens, 2nd 18 torrentium")
```

```{r, round the identification confidence and begin data tidy up 9 PCs section, echo=FALSE}

assigned_ids <- specify$posterior


assignedround <- round(assigned_ids, digits = 2)

colnames(assignedround)[colnames(assignedround) == "Cx. pipiens"] <- "Identification"
# round the data and make into english

assignedround <- assignedround[, -2]

assignedround[assignedround > .50] <- "*Culex pipiens*"
assignedround[assignedround < .50] <- "*Culex torrentium*"



```

Correct identifications of "known unknowns"
Table of the Predicted species and the actual species for each known unknown individual

```{r, compile the correct id list compare and report 9 PCs section, echo=FALSE}
#correct id list from the database prior identification


uspecies <- gsub(pattern = "\\d : UnkPipiens$", replacement = "*Culex pipiens*", uspecies, ignore.case = F)

    uspecies <- gsub(pattern = "\\d : UnkTorrentium$", replacement = "*Culex torrentium*", uspecies, ignore.case = F)
    uspecies <- gsub(pattern = "\\d", replacement = "", uspecies, ignore.case = F)




#join the 2 objects

compare <- cbind(assignedround, uspecies)


colnames(compare) <- c("Predicted id", "Actual id")

#compare the data sets and declare whether they are correct or not
    compare <- as.data.frame(compare)
    compare[, "Correct_id"] <- NA


    #fix for factor levels not matching
    compare[, 2] <- factor(compare[, 2], levels = levels(compare[, 1]))
            for (i in seq(along = compare[, 1])){
            ifelse(compare[i, 1] == compare[i, 2], Correct <- TRUE, Correct <- FALSE )
            Correct <- as.vector(Correct)
            compare[i, 3] <- Correct
            }

    compare
    ```


Correct identifications of "known unknowns"

    ```{r, echo= FALSE}
    correct_id <- plyr::count(compare$Correct_id)

correct_id

```


percentage correct identification precentage rate

```{r, echo= FALSE}
```

```
accuracy_percent <- round(correct_id[2, 2] / (correct_id[1, 2] + correct_id[2, 2] )
    * 100, digits = 2)

accuracy_percent


output <- as.data.frame(cbind(description, accuracy_percent))



```

    ```{r send results to Access 9 PCs section}
    ## This section saves the results to an Access database to allow the investigation of the method's precision

        #Send it to access
        # save to a new table called Wingmorph_data, if the table already exists then add it to the end of the table that already exists.
        sqlSave(con, output, tablename = "Wingmorph_data_9_pcs", rownames = FALSE, append = TRUE, safer = TRUE, fast = FALSE)
        odbcClose(con)
        ```

Appendix 3 Identification tool, verbose return with code visible (normally hidden).

Wing Morphometric discrimination between *Culex pipiens s.s. Culex torrentium*

Rich Halfpenny

05 January 2017

```
suppressMessages(library(geomorph))
suppressMessages(library(RODBC))
suppressMessages(library(dplyr))
suppressMessages(library(adegenet))
suppressMessages(library(knitr))

##Edit this region only

verbose <- TRUE  #Select "TRUE" for the full output, including descriptive
 statistics
#for the training data. Select "FALSE" for identification of unknowns only
.

con <- odbcConnect("MDnew") # Replace MDnew with different ODBC identifier
 as required

remoutliers <- -c(22, 21) # change these values in response to the outlier
 plot for
#the training data- Use Verbose <- TRUE to see this plot

# SQL query to import the training data
qry1 <- "SELECT Wing_morphometrics_LM_combined.ID1, Wing_morphometrics_LM_
combined.ID, Wing_morphometrics_LM_combined.COMMENT, Wing_morphometrics_LM
_combined.VAR, Wing_morphometrics_LM_combined.x2, Wing_morphometrics_LM_co
mbined.y2, Wing_morphometrics_LM_combined.x3, Wing_morphometrics_LM_combin
ed.y3, Wing_morphometrics_LM_combined.x7, Wing_morphometrics_LM_combined.y
7, Wing_morphometrics_LM_combined.x8, Wing_morphometrics_LM_combined.y8, W
ing_morphometrics_LM_combined.x9, Wing_morphometrics_LM_combined.y9, Wing_
morphometrics_LM_combined.x17, Wing_morphometrics_LM_combined.y17, Wing_mo
rphometrics_LM_combined.x18, Wing_morphometrics_LM_combined.y18
FROM Wing_morphometrics_LM_combined;
"
#SQL query for the specimen/s to be identified

qry2 <- "SELECT Wing_morphometrics_LM_FM.ID1, Wing_morphometrics_LM_FM.ima
ge, Wing_morphometrics_LM_FM.COMMENT, Wing_morphometrics_LM_FM.VAR,  Wing_
morphometrics_LM_FM.x2, Wing_morphometrics_LM_FM.y2, Wing_morphometrics_LM
_FM.x3, Wing_morphometrics_LM_FM.y3,  Wing_morphometrics_LM_FM.x7, Wing_mo
rphometrics_LM_FM.y7, Wing_morphometrics_LM_FM.x8, Wing_morphometrics_LM_F
M.y8, Wing_morphometrics_LM_FM.x9, Wing_morphometrics_LM_FM.y9,   Wing_mor
phometrics_LM_FM.x17, Wing_morphometrics_LM_FM.y17, Wing_morphometrics_LM_
FM.x18, Wing_morphometrics_LM_FM.y18
FROM Wing_morphometrics_LM_FM;"
```

```
##Editable region end

#Main code

fromdb1 <- sqlQuery(con, qry1, stringsAsFactors = FALSE)
fromdb1[, (5:18)] <- fromdb1[, (5:18)] * -1 # this is just to flip the dat
a over so the wings are not upside down in the visualisation
fromdb <- fromdb1 # this is just a fudge for those sections of the data wh
ich do not have  criteria to match and so don't use the fromdb1 source
#how to collect the classifier data in a separate array - geomorph does no
t expect to see these in the landmark data

classifiers <- fromdb1[, 2:4]
species <- classifiers[, 1]
sex <- classifiers [, 2]
whichwing <- classifiers[, 3]


#remove the classifiers from the raw data a create a landmark file

coords <- as.matrix(fromdb1[, -(1:4)])

coords <- arrayspecs(coords, 7, 2)

fromdb <- filter(fromdb1, COMMENT == "Female" )

classifiers <- fromdb[, 2:4]

species <- classifiers[, 1]
sex <- classifiers [, 2]
whichwing <- classifiers[, 3]


coords <- as.matrix(fromdb[, -(1:4)])

coords <- arrayspecs(coords, 7, 2)

y <- gpagen(coords, PrinAxes = TRUE, print.progress = FALSE)

if (verbose == TRUE) {
  summary(y)

  plot(y)

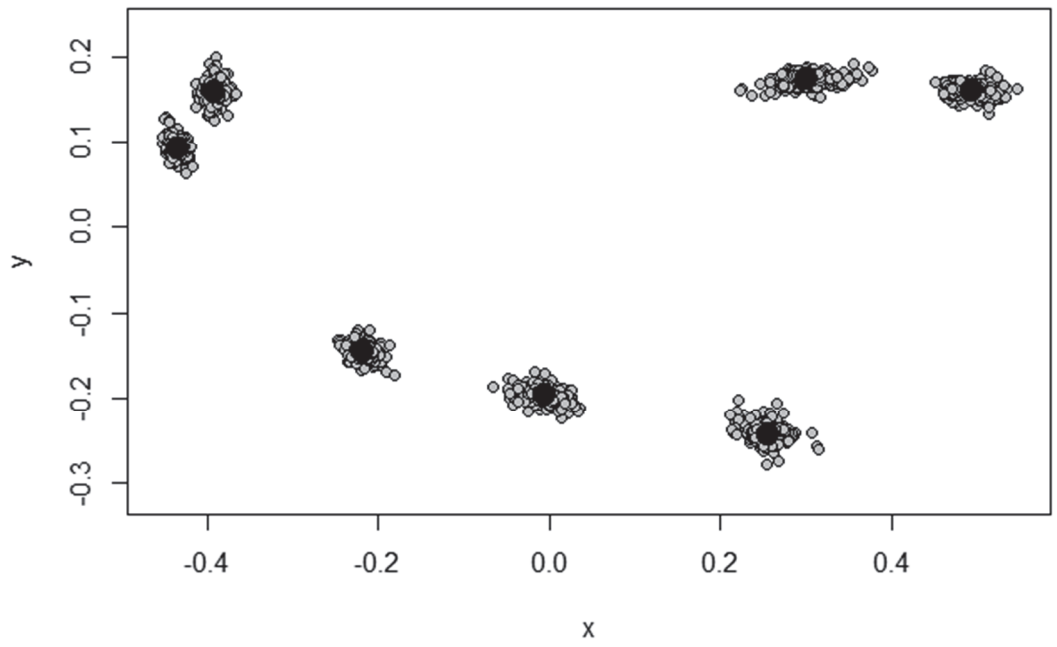
  plotOutliers(y$coords, groups = NULL)

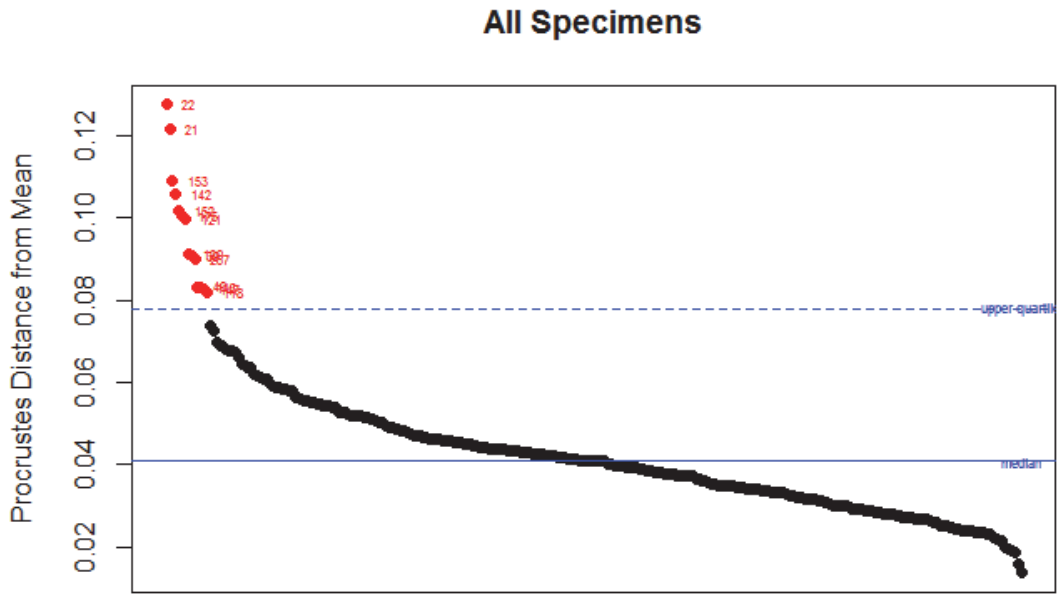
  outliers <- plotOutliers(y$coords)

}

##
## Call:
## gpagen(A = coords, PrinAxes = TRUE, print.progress = FALSE)
##
```

```
##
##
## Generalized Procrustes Analysis
## with Partial Procrustes Superimposition
##
## 7 fixed landmarks
## 0 semilandmarks (sliders)
## 2-dimensional landmarks
## 2 GPA iterations to converge
##
##
## Consensus (mean) Configuration
##
##                  X           Y
## [1,] -0.392109153  0.15806714
## [2,] -0.435419231  0.09291591
## [3,] -0.217991450 -0.14541685
## [4,] -0.006359123 -0.19660302
## [5,]  0.255769473 -0.24290203
## [6,]  0.494276613  0.16015223
## [7,]  0.301832871  0.17378662
```






```
coords <- as.matrix(fromdb[remoutliers, -(1:4)])
classifiers <- fromdb[remoutliers, 2:4]
fromdb <- fromdb[remoutliers, ]

######

species <- classifiers[, 1]
```

```
sex <- classifiers [, 2]
whichwing <- classifiers[, 3]

coords <- arrayspecs(coords, 7, 2) # re add this if outliers are being rem
oved

y <- gpagen(coords, PrinAxes = TRUE, print.progress = FALSE)

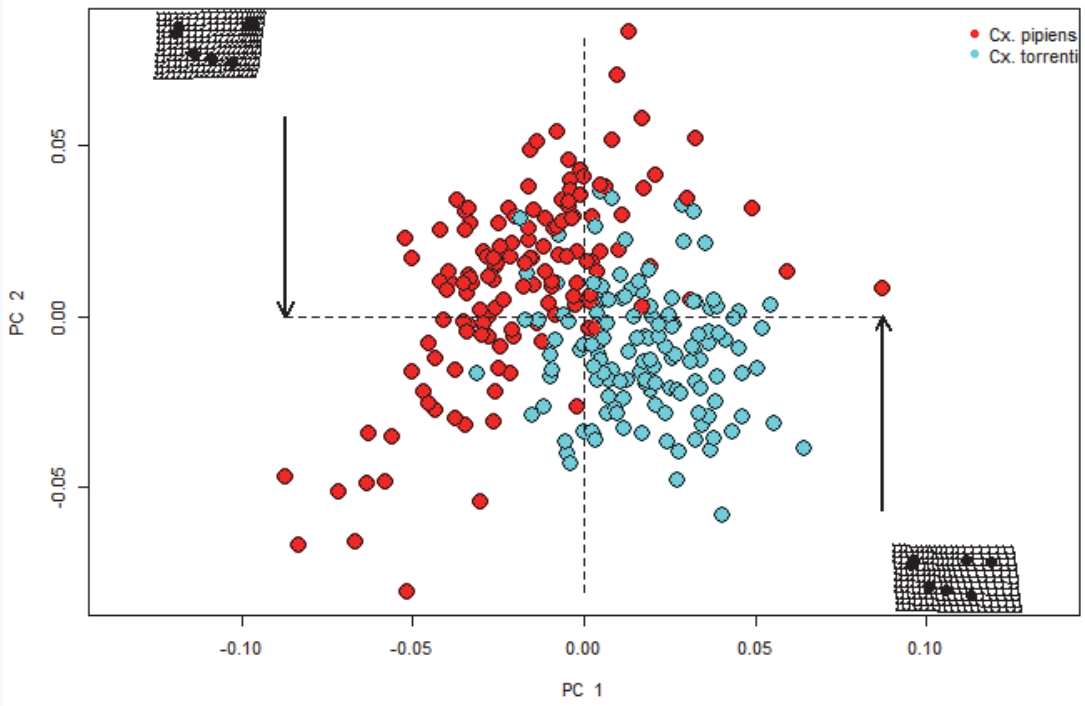
coords <- two.d.array(y$coords)

if (verbose == TRUE){
  set.seed(123)
  morphol.disparity(y$coords ~ 1, groups = ~ species, iter = 9999, print.p
rogress = FALSE)

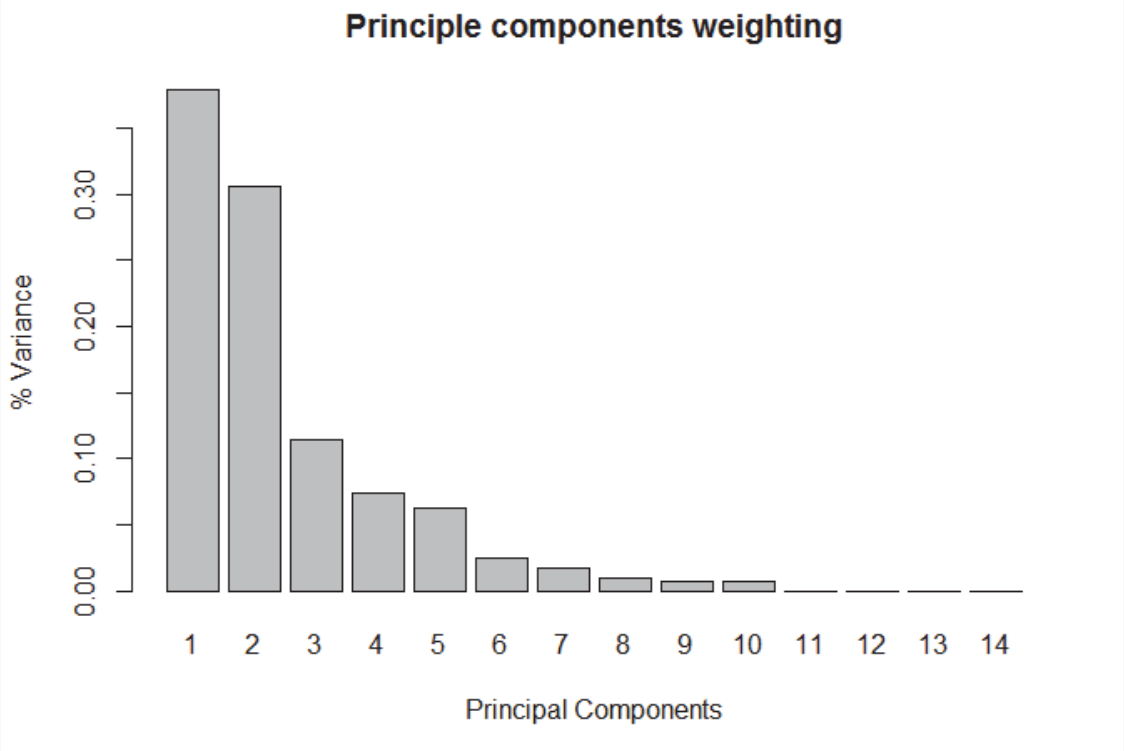

species <- as.factor(species)
col.gp <- rainbow(length(levels(species)))
names(col.gp) <- levels(species)
col.gp <- col.gp[match(species, names(col.gp))]

PCA <- plotTangentSpace(y$coords, groups = col.gp, legend = TRUE)

pvar <- (PCA$sdev^2) / (sum(PCA$sdev^2))
names(pvar) <- seq(1:length(pvar))
barplot(pvar, xlab = "Principal Components", ylab = "% Variance", main  =
"Principal components weighting")
```



}




```
dapc1 <- dapc(coords, grp = species, n.pca = 7, n.da = 1)

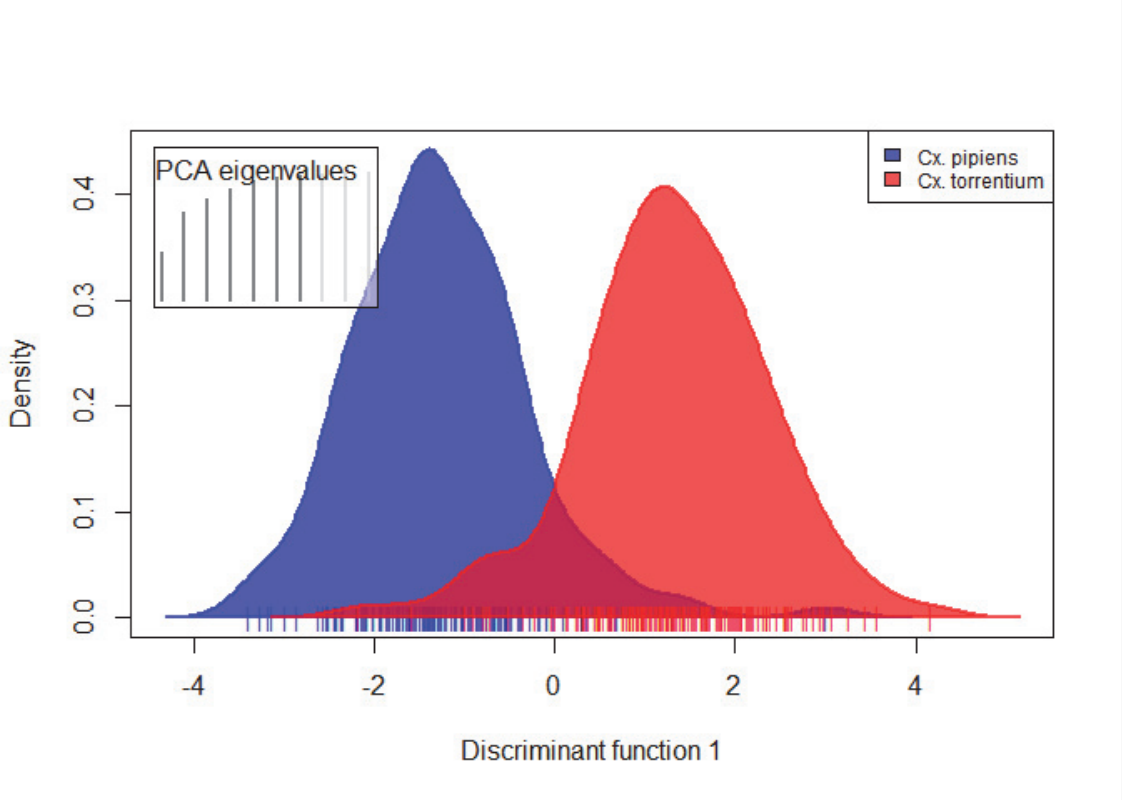
#visualise these results on the fit of the training data

if (verbose == TRUE){
scatter.dapc(dapc1, legend = TRUE, posi.leg = "topright", scree.pca = TRUE
```

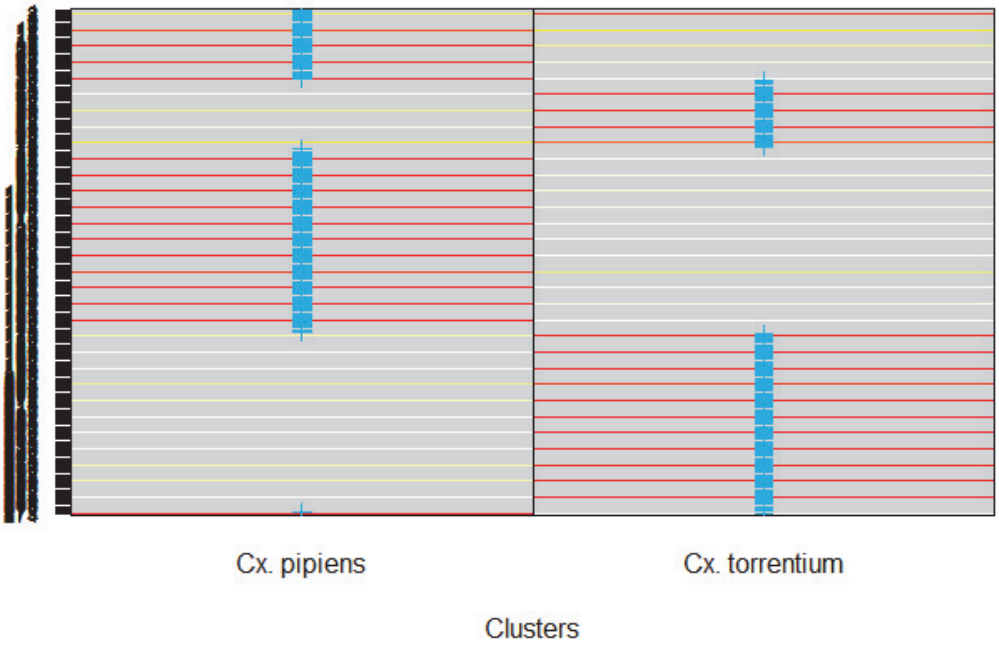
```
,
             posi.pca = "topleft", cleg = 0.75, xax = 1, yax = 2, ratio.pc
a = 0.2)

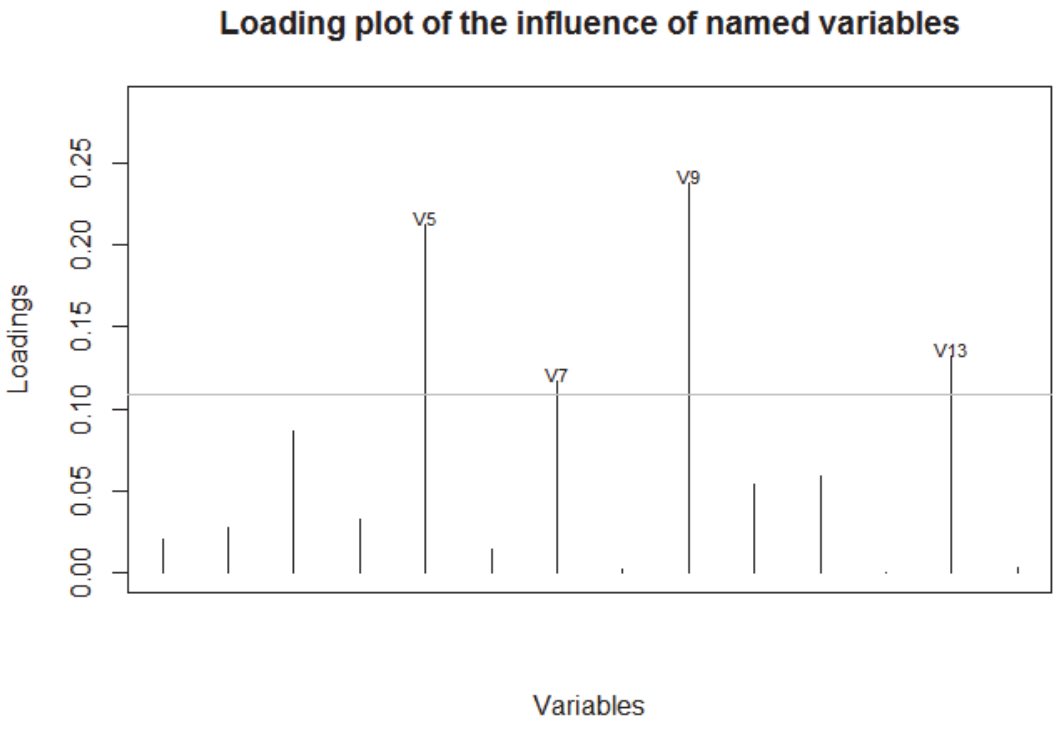
assignplot(dapc1)

loadingplot(dapc1$var.contr, main = "Loading plot of the influence of name
d variables")
```




```
}
```






```
#Prediction section

#Import the data for the specimens which need to be identified
fromdb2 <- sqlQuery(con, qry2, stringsAsFactors = FALSE)
```

```
fromdb2 <- filter(fromdb2 )

fromdb2[, (5:18)] <- fromdb2[, (5:18)] * -1 # this is just to flip the dat
a over so the wings are not upside down in the visualisation
unknown <- fromdb2


uclassifiers <- fromdb2[, 2:4]
uniqueid <- uclassifiers[, 1]

#remove the classifiers from the raw data a create a landmark file

ucoords <- as.matrix(fromdb2[, -(1:4)])

#count how many unknowns have been added
no_specimens <- nrow(ucoords)

#bind training data and unknowns so that GPA and DAPC can be carried on wh
ole data
totcoords <- rbind(coords, ucoords)
totcoords <- arrayspecs(totcoords, 7, 2)
suppressWarnings(y <- gpagen(totcoords, PrinAxes = TRUE, print.progress =
FALSE))
totcoords <- two.d.array(y$coords)
nrowstrain <- nrow(totcoords) - no_specimens
dapc1 <- dapc(head(totcoords, n = nrowstrain), grp = species, n.pca = 7, n
.da = 1)

#separate the unknowns again so they can be used for prediction
ucoords <- tail(totcoords, n = no_specimens)


specify <- predict(dapc1, ucoords)

assigned_ids <- specify$posterior


assignedround <- round(assigned_ids, digits = 2)

colnames(assignedround)[colnames(assignedround) == "Cx. pipiens"] <- "Iden
tification"

# round the data and make into english

assignedround <- assignedround[, -2]

assignedround[assignedround > .50] <- "Culex pipiens"

assignedround[assignedround == .50] <- "No Choice"

assignedround[assignedround < .50] <- "Culex torrentium"

Specimen_identified_as <- assignedround

Specimen_reference <- uniqueid
```

```
idlist <- cbind(Specimen_reference, Specimen_identified_as)
```

Appendix 4 Example of identification result output from the

```
kable(idlist)
```

Specimen_reference	Specimen_identified_as
10263_PIP_F_L.tif	*Culex pipiens*
10263_PIP_F_R.tif	*Culex pipiens*
10264_PIP_F_L.tif	*Culex torrentium*
10264_PIP_F_R.tif	*Culex pipiens*
10265_PIP_F_L.tif	*Culex pipiens*
10265_PIP_F_R.tif	*Culex pipiens*
10266_PIP_F_L.tif	*Culex pipiens*
10266_PIP_F_R.tif	*Culex pipiens*
10267_PIP_F_L.tif	*Culex pipiens*
10267_PIP_F_R.tif	*Culex pipiens*
10268_PIP_F_L.tif	*Culex pipiens*
10268_PIP_F_R.tif	*Culex pipiens*
10269_PIP_F_L.tif	*Culex pipiens*
10269_PIP_F_R.tif	*Culex pipiens*
10270_PIP_F_L.tif	*Culex pipiens*
10270_PIP_F_R.tif	*Culex pipiens*
10271_PIP_F_L.tif	*Culex pipiens*
10271_PIP_F_R.tif	*Culex pipiens*
10272_PIP_F_L.tif	*Culex pipiens*
10272_PIP_F_R.tif	*Culex pipiens*
10273_PIP_F_L.tif	*Culex torrentium*
10273_PIP_F_R.tif	*Culex torrentium*
10275_PIP_F_L.tif	*Culex pipiens*
10275_PIP_F_R.tif	*Culex pipiens*
10276_PIP_F_L.tif	*Culex pipiens*
10276_PIP_F_R.tif	*Culex pipiens*
10277_PIP_F_L.tif	*Culex pipiens*
10277_PIP_F_R.tif	*Culex pipiens*
10278_PIP_F_L.tif	*Culex pipiens*
10278_PIP_F_R.tif	*Culex pipiens*
10279_PIP_F_L.tif	*Culex pipiens*
10279_PIP_F_R.tif	*Culex pipiens*
10280_PIP_F_L.tif	*Culex pipiens*
10280_PIP_F_R.tif	*Culex pipiens*
10281_PIP_F_L.tif	*Culex pipiens*
10281_PIP_F_R.tif	*Culex pipiens*
10282_PIP_F_L.tif	*Culex torrentium*
10282_PIP_F_R.tif	*Culex pipiens*
10283_PIP_F_L.tif	*Culex pipiens*

10283_PIP_F_R.tif	*Culex pipiens*
10284_TORR_F_L.tif	*Culex torrentium*
10284_TORR_F_R.tif	*Culex torrentium*
10285_TORR_F_L.tif	*Culex torrentium*
10285_TORR_F_R.tif	*Culex torrentium*
10286_TORR_F_L.tif	*Culex torrentium*
10287_TORR_F_L.tif	*Culex torrentium*
10287_TORR_F_R.tif	*Culex torrentium*
10288_TORR_F_R.tif	*Culex torrentium*
10289_TORR_F_L.tif	*Culex torrentium*
10289_TORR_F_R.tif	*Culex torrentium*
10290_TORR_F_L.tif	*Culex pipiens*
10290_TORR_F_R.tif	*Culex torrentium*
10291_TORR_F_L.tif	*Culex torrentium*
10291_TORR_F_R.tif	*Culex torrentium*
10292_TORR_F_L.tif	*Culex torrentium*
10292_TORR_F_R.tif	*Culex torrentium*
10293_TORR_F_L.tif	*Culex torrentium*
10293_TORR_F_R.tif	*Culex torrentium*
10294_TORR_F_L.tif	*Culex pipiens*
10294_TORR_F_R.tif	*Culex pipiens*
10295_TORR_F_L.tif	*Culex torrentium*
10295_TORR_F_R.tif	*Culex pipiens*
10296_TORR_F_L.tif	*Culex torrentium*
10296_TORR_F_R.tif	*Culex torrentium*
10297_TORR_F_L.tif	*Culex torrentium*
10297_TORR_F_R.tif	*Culex torrentium*
10298_TORR_F_L.tif	*Culex torrentium*
10299_TORR_F_L.tif	*Culex torrentium*
10299_TORR_F_R.tif	*Culex torrentium*
10300_TORR_F_L.tif	*Culex torrentium*
10300_TORR_F_R.tif	*Culex torrentium*
10301_TORR_F_L.tif	*Culex torrentium*
10301_TORR_F_R.tif	*Culex torrentium*
10302_TORR_F_L.tif	*Culex torrentium*
10302_TORR_F_R.tif	*Culex torrentium*
10303_TORR_F_L.tif	*Culex torrentium*
10303_TORR_F_R.tif	*Culex torrentium*

Appendix 5 An example photograph of a *Culex pipiens* s.l. mosquitoes captured using a clip-on macro lens and a smart phone (Samsung Galaxy S3).

Appendix 6 R code to import MS data to Access database

```
library(RODBC)   # this library allows the connection to the access .accdb
library (dplyr)

setwd("~/PHD/R/phd/GC")    # This is your working directory - in this case it needs to be where the .csv files are placed before import

files = dir(pattern="_150.CSV")    # this .CSV is case sensitive - I assume that TurboMass will always output with capital file extension, can add logic so that it is not case sensitive if needed.
files2 = dir(pattern="_lib.CSV")

for (i in 1:length(files)) {      # this loops through the following section as many times as the number of files found

 db = odbcConnect("MDnew")

 strs <- readLines(files[i])


 #get all content from CSV page table content
 dats <- read.csv(text=strs,           # read from an R object rather than a file
             skip=5  )

 dats <- dats[grep("[[:digit:]]", dats$X), ]      # might seem odd to do this in 2 steps but it does need to be this way
 dats <- dats[grep("[[:digit:]]", dats$X.6), ]
 dats <- select(dats, X, X.1, X.2, X.3, X.4, X.5, X.6)


 dats2 <- do.call("rbind", lapply(files[i], function(x) {
  dats2 <- read.csv(x, header=FALSE,row.names=NULL, nrows = length(strs) - 236)
  dats2$fileName <- tools::file_path_sans_ext(basename(x))
  dats2
 }))

#  dats2 <- do.call("rbind", lapply(files, function(x) {                        #temp edit for a manual workthrough
#    dats2 <- read.csv(x, header=FALSE,row.names=NULL, nrows = length(strs) - 236)
#    dats2$fileName <- tools::file_path_sans_ext(basename(x))
#    dats2
#  }))


 dats2 <- subset(dats2, !duplicated(fileName))

 atrs <- readLines(files2[i])
 # strs <- readLines("H_sphondylium_5_capture_080915_091115_024828.CSV")

 #get all content from CSV page table content
 dat <- read.csv(text=atrs, sep=",",           # read from an R object rather than a file
             skip=5  )
 dat <- select(dat, X, X.1, X.2, X.3, X.4, X.5)

 dat <- dat[grep("[[:digit:]]", dat$X.2), ]      # might seem odd to do this in 2 steps but it does need to be this way
 dat <- filter(dat, !grepl("Vial",X))
 dat <- filter(dat, !grepl("Inst",X))
 dat <- filter(dat, !grepl("Print",X))
 dat <- filter(dat, !grepl("Page",X))

 dat1 <- filter(dat, grepl("1",X))
 dat1 <- select(dat1,X.1, X.2, X.3, X.4, X.5)
 names(dat1)[1] <- "Lib_1"
 names(dat1)[2] <- "Match_1"
 names(dat1)[3] <- "R.Match_1"
 names(dat1)[4] <- "Name_1"
 names(dat1)[5] <- "CAS_No_1"

 dat2 <- filter(dat, grepl("2",X))
 dat2 <- select(dat2,X.1, X.2, X.3, X.4, X.5)
 names(dat2)[1] <- "Lib_2"
 names(dat2)[2] <- "Match_2"
 names(dat2)[3] <- "R.Match_2"
 names(dat2)[4] <- "Name_2"
 names(dat2)[5] <- "CAS_No_2"
```

```
 dat3 <- filter(dat, grepl("3",X))
 dat3 <- select(dat3,X.1, X.2, X.3, X.4, X.5)
 names(dat3)[1] <- "Lib_3"
 names(dat3)[2] <- "Match_3"
 names(dat3)[3] <- "R.Match_3"
 names(dat3)[4] <- "Name_3"
 names(dat3)[5] <- "CAS_No_3"

 dat4 <- filter(dat, grepl("4",X))
 dat4 <- select(dat4,X.1, X.2, X.3, X.4, X.5)
 names(dat4)[1] <- "Lib_4"
 names(dat4)[2] <- "Match_4"
 names(dat4)[3] <- "R.Match_4"
 names(dat4)[4] <- "Name_4"
 names(dat4)[5] <- "CAS_No_4"

 dat5 <- filter(dat, grepl("5",X))
 dat5 <- select(dat5,X.1, X.2, X.3, X.4, X.5)
 names(dat5)[1] <- "Lib_5"
 names(dat5)[2] <- "Match_5"
 names(dat5)[3] <- "R.Match_5"
 names(dat5)[4] <- "Name_5"
 names(dat5)[5] <- "CAS_No_5"

 datall <- cbind(dat1, dat2, dat3, dat4, dat5)



 todb <- cbind(dats2["fileName"],dats, datall)
 #give the columns sensible names

 names(todb)[2] <- "PeakNumber"
 names(todb)[3] <- "RT"
 names(todb)[4] <- "Scan"
 names(todb)[5] <- "Height"
 names(todb)[6] <- "Area"
 names(todb)[7] <- "AreaPercent"
 names(todb)[8] <- "NormPercent"


 #Send it to access
 # save to a new table called MS Outputs, if the table already exists then add it to the end of the table that already exists.
 sqlSave(db, todb, tablename = "Hedera_helix_gcms_IS", rownames = FALSE, append = TRUE, safer = TRUE, fast = FALSE)
 odbcClose(db)
}

#move the .CSV files so you can't accidently import them twice - Edit the folder path in to="" to your receiving folder

file.copy(from=files, to="C:/Users/Rich/Documents/PHD/R/phd/GC/IMPORTED/",
       overwrite = FALSE, recursive = FALSE,
       copy.mode = TRUE)
file.copy(from=files2, to="C:/Users/Rich/Documents/PHD/R/phd/GC/IMPORTED/",
       overwrite = FALSE, recursive = FALSE,
       copy.mode = TRUE)
# delete the original .csv files after moving

file.remove(files)
file.remove(files2)
```

Appendix 7 R Markdown code to compare target VOCs to MS outputs - final

```
---
title: "STEP 2 IS 2017 Compare CAS to targets"
author: "Rich Halfpenny"
date: "24 March 2017"
output:
  word_document: default
  html_document: default
---
```{r, get data, echo=FALSE}
## Packages used
suppressMessages(library(RODBC))
suppressMessages(library(dplyr))
suppressMessages(library(knitr))
suppressMessages(library(ggplot2))
suppressMessages(library(reshape2))
## Get the data from the database
con = odbcConnect("MDnew")
qry1 = "SELECT GC_target_compounds.Compound_name, GC_target_compounds.CAS_registry_number, GC_target_compounds.Average_RI_for_DB5_column FROM GC_target_compounds;"
qry2 = "SELECT MS_outputs_150.fileName, MS_outputs_150.PeakNumber, MS_outputs_150.RT, MS_outputs_150.Scan, MS_outputs_150.Height, MS_outputs_150.Area, MS_outputs_150.AreaPercent, MS_outputs_150.NormPercent, MS_outputs_150.Lib_1, MS_outputs_150.Match_1, MS_outputs_150.RMatch_1, MS_outputs_150.Name_1, MS_outputs_150.CAS_No_1, MS_outputs_150.Lib_2, MS_outputs_150.Match_2, MS_outputs_150.RMatch_2, MS_outputs_150.Name_2, MS_outputs_150.CAS_No_2, MS_outputs_150.Lib_3, MS_outputs_150.Match_3, MS_outputs_150.RMatch_3, MS_outputs_150.Name_3, MS_outputs_150.CAS_No_3, MS_outputs_150.Lib_4, MS_outputs_150.Match_4, MS_outputs_150.RMatch_4, MS_outputs_150.Name_4, MS_outputs_150.CAS_No_4, MS_outputs_150.Lib_5, MS_outputs_150.Match_5, MS_outputs_150.RMatch_5, MS_outputs_150.Name_5, MS_outputs_150.CAS_No_5
FROM MS_outputs_150;"
qry3 = "SELECT Hedera_helix_gcms_IS.fileName, Hedera_helix_gcms_IS.PeakNumber, Hedera_helix_gcms_IS.RT, Hedera_helix_gcms_IS.Scan, Hedera_helix_gcms_IS.Height,Hedera_helix_gcms_IS.Area, Hedera_helix_gcms_IS.AreaPercent, Hedera_helix_gcms_IS.NormPercent, Hedera_helix_gcms_IS.Lib_1, Hedera_helix_gcms_IS.Match_1, Hedera_helix_gcms_IS.RMatch_1, Hedera_helix_gcms_IS.Name_1, Hedera_helix_gcms_IS.CAS_No_1, Hedera_helix_gcms_IS.Lib_2, Hedera_helix_gcms_IS.Match_2, Hedera_helix_gcms_IS.RMatch_2, Hedera_helix_gcms_IS.Name_2, Hedera_helix_gcms_IS.CAS_No_2, Hedera_helix_gcms_IS.Lib_3, Hedera_helix_gcms_IS.Match_3, Hedera_helix_gcms_IS.RMatch_3, Hedera_helix_gcms_IS.Name_3, Hedera_helix_gcms_IS.CAS_No_3, Hedera_helix_gcms_IS.Lib_4, Hedera_helix_gcms_IS.Match_4, Hedera_helix_gcms_IS.RMatch_4, Hedera_helix_gcms_IS.Name_4, Hedera_helix_gcms_IS.CAS_No_4, Hedera_helix_gcms_IS.Lib_5, Hedera_helix_gcms_IS.Match_5, Hedera_helix_gcms_IS.RMatch_5, Hedera_helix_gcms_IS.Name_5, Hedera_helix_gcms_IS.CAS_No_5
FROM Hedera_helix_gcms_IS;"
targets = sqlQuery(con, qry1)
targets = arrange(targets, Compound_name)
CASlist <- targets$CAS_registry_number
ISCAS <- "112-29-8"    # CAS number for Internal standard in this case 1-Bromodecane
gcdata = sqlQuery(con, qry2)
gcdata2 = sqlQuery(con, qry3)
gcdata = suppressWarnings(bind_rows(gcdata, gcdata2))  #this generates 11 warnings - can't see anything wrong with the data and function is fine, so suppressing the warnings
#set the names of the species to objects for later use
hs <- "Heracleum sphondylium"
ap <- "Angelica sylvestris"
hh <- "Hedera helix"
dc <- "Daucus carota carota"
anp <- "Anthriscus sylvestris"
cm <- "Conopodium majus"

#THE CODE BELOW filters SO THAT IT ONLY retains rows in the target compound list and MATCH VALUE => 800
foundcompounds <- gcdata %>%
  filter(CAS_No_1 %in% CASlist & Match_1 >= 800| CAS_No_2 %in% CASlist & Match_2 >= 800| CAS_No_3 %in% CASlist & Match_3 >= 800| CAS_No_4 %in% CASlist & Match_4 >= 800| CAS_No_5 %in% CASlist & Match_5 >= 800)
# code to replace the name_1 value with the name of the target compound found rather than that of the first of 5 chemicals proposed for the match.
foundtemp <- merge(x = foundcompounds, targets, by.x = "CAS_No_5", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp2 <- merge(x = foundcompounds, targets, by.x = "CAS_No_4", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp3 <- merge(x = foundcompounds, targets, by.x = "CAS_No_3", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp4 <- merge(x = foundcompounds, targets, by.x = "CAS_No_2", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp5 <- merge(x = foundcompounds, targets, by.x = "CAS_No_1", by.y = "CAS_registry_number", all.x = FALSE)

#this union creates duplicates... but the duplicates are based on the fact that in some cases a peak is identified as 2 different compounds in the target list - e.g. myrcene and pinene or the Lilac alcohol group being represented multiple times.

foundtemp6 <- union(foundtemp, foundtemp2)
foundtemp6 <- union(foundtemp6, foundtemp3)
```
```

```
foundtemp6 <- union(foundtemp6, foundtemp4)
foundtemp6 <- union(foundtemp6, foundtemp5)

#NEXT STEP REMOVE THE DUPLICATES LEAVING THE BEST MATCH BEHIND

foundtemp7 <- subset(foundtemp6, !duplicated(Height))

foundcompounds <- foundtemp7 %>%
  select(-Name_1) %>%
  mutate(Found_target_compound = Compound_name)

#rename the object that will contain the internal standard data for later use

foundcompoundsIS <- foundcompounds

# now strip the internal standard from the data  so that it does not appear in the graphs and tables of data for the flowers - only want it to appear in the later stuff to do with the ratio of the compounds

foundcompounds <- foundcompounds %>%
  filter(CAS_No_1 != "112-29-8" ) %>%
  filter(CAS_No_2 != "112-29-8") %>%
  filter(CAS_No_3 != "112-29-8") %>%
  filter(CAS_No_4 != "112-29-8") %>%
  filter(CAS_No_5 != "112-29-8")

#THE CODE BELOW filters SO THAT IT ONLY retains rows in the target compound list and MATCH VALUE => 900
foundcompounds900 <- gcdata %>%
  filter(CAS_No_1 %in% CASlist & Match_1 >= 900| CAS_No_2 %in% CASlist & Match_2 >= 900| CAS_No_3 %in% CASlist & Match_3 >= 900|
CAS_No_4 %in% CASlist & Match_4 >= 900| CAS_No_5 %in% CASlist & Match_5 >= 900)

# code to replace the name_1 value with the name of the target compound found rather than that of the first of 5 chemicals proposed for the match.

foundtemp <- merge(x = foundcompounds900, targets, by.x = "CAS_No_5", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp2 <- merge(x = foundcompounds900, targets, by.x = "CAS_No_4", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp3 <- merge(x = foundcompounds900, targets, by.x = "CAS_No_3", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp4 <- merge(x = foundcompounds900, targets, by.x = "CAS_No_2", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp5 <- merge(x = foundcompounds900, targets, by.x = "CAS_No_1", by.y = "CAS_registry_number", all.x = FALSE)

#this union creates duplicates... but the duplicates are based on the fact that in some cases a peak is identified as 2 different compounds in the target list - e.g. myrcene and pinene or the Lilac alcohol group being represented multiple times.

foundtemp6 <- union(foundtemp, foundtemp2)
foundtemp6 <- union(foundtemp6, foundtemp3)
foundtemp6 <- union(foundtemp6, foundtemp4)
foundtemp6 <- union(foundtemp6, foundtemp5)

#NEXT STEP REMOVE THE DUPLICATES LEAVING THE BEST MATCH BEHIND
foundtemp7 <- subset(foundtemp6, !duplicated(Height))
foundcompounds900 <- foundtemp7 %>%
  select(-Name_1) %>%
  mutate(Found_target_compound = Compound_name)
#THE CODE BELOW filters SO THAT IT ONLY retains rows in the target compound list and MATCH VALUE => 950
foundcompounds950 <- gcdata %>%
  filter(CAS_No_1 %in% CASlist & Match_1 >= 950| CAS_No_2 %in% CASlist & Match_2 >= 950| CAS_No_3 %in% CASlist & Match_3 >= 950|
CAS_No_4 %in% CASlist & Match_4 >= 950| CAS_No_5 %in% CASlist & Match_5 >= 950)
# code to replace the name_1 value with the name of the target compound found rather than that of the first of 5 chemicals proposed for the match.

foundtemp <- merge(x = foundcompounds950, targets, by.x = "CAS_No_5", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp2 <- merge(x = foundcompounds950, targets, by.x = "CAS_No_4", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp3 <- merge(x = foundcompounds950, targets, by.x = "CAS_No_3", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp4 <- merge(x = foundcompounds950, targets, by.x = "CAS_No_2", by.y = "CAS_registry_number", all.x = FALSE)
foundtemp5 <- merge(x = foundcompounds950, targets, by.x = "CAS_No_1", by.y = "CAS_registry_number", all.x = FALSE)

#this union creates duplicates... but the duplicates are based on the fact that in some cases a peak is identified as 2 different compounds in the target list - e.g. myrcene and pinene or the Lilac alcohol group being represented multiple times.

foundtemp6 <- union(foundtemp, foundtemp2)
foundtemp6 <- union(foundtemp6, foundtemp3)
foundtemp6 <- union(foundtemp6, foundtemp4)
```

```
foundtemp6 <- union(foundtemp6, foundtemp5)

#NEXT STEP REMOVE THE DUPLICATES LEAVING THE BEST MATCH BEHIND

foundtemp7 <- subset(foundtemp6, !duplicated(Height))

foundcompounds950 <- foundtemp7 %>%
 select(-Name_1) %>%
 mutate(Found_target_compound = Compound_name)
```
```{r mulitplot function, echo=FALSE}

multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
 require(grid)

 # Make a list from the ... arguments and plotlist
 plots <- c(list(...), plotlist)

 numPlots = length(plots)

 # If layout is NULL, then use 'cols' to determine layout
 if (is.null(layout)) {
  # Make the panel
  # ncol: Number of columns of plots
  # nrow: Number of rows needed, calculated from # of cols
  layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                   ncol = cols, nrow = ceiling(numPlots/cols))
 }

 if (numPlots==1) {
  print(plots[[1]])

 } else {
  # Set up the page
  grid.newpage()
  pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

  # Make each plot, in the correct location
  for (i in 1:numPlots) {
   # Get the i,j matrix positions of the regions that contain this subplot
   matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

   print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                   layout.pos.col = matchidx$col))
  }
 }
}
```

Interpretation of the Match_x and R.Match_x values (where x is 1 to 5 depending on possible identification compound)
Flower species specific data
###collate data by species and tube type - i.e. pre filter, capture, breakthrough
Negative control - Hexane blank - Hexane only


```{r, echo=FALSE}
#orignal table code (changed to a function)- used later for the >=900 match graph leave in addition to the shorter version below

# speciesfilterfunction short - df = data source i = filename search string
speciesfilterlong <- function (df, i) {
 df %>%
  filter(grepl(i,fileName, ignore.case = TRUE)) %>%
  select(fileName, RT, Height, Area,
Found_target_compound,CAS_No_1,Match_1,CAS_No_2,Match_2,CAS_No_3,Match_3,CAS_No_4,Match_4,CAS_No_5,Match_5) %>%
  arrange(Found_target_compound)
}

hexane_blank <- speciesfilterlong(foundcompounds,"Hexane_blank")

# speciesfilterfunction short - df = data source i = filename search string
speciesfilter <- function (df, i) {
```

```
  df%>%
    filter(grepl(i,fileName, ignore.case = TRUE)) %>%
    select(Found_target_compound,fileName, RT, Height, Area) %>%
    arrange(Found_target_compound)
}
hexane_blanks <- speciesfilter(foundcompounds,"Hexane_blank")
# make a list of compounds in no variable and subtract _ need to this data off the GC still.

#kable(n_var_pre, caption = "Target compounds found in the No Variable negative control Pre Filter results")
#kable(n_var_bre, caption = "Target compounds found in the No Variable negative control Breakthrough results")
kable(hexane_blanks, caption = "Target compounds found in the hexane blank results")
```
Negative control - Solvent desorption using hexane from the Sorbent tubes only.
Should only be whatever is found on the PorapakQ or in the Hexane solvent
```{r, echo=FALSE}
tube_blank <- speciesfilterlong(foundcompounds,"Tube_blank")
tube_blanks <- speciesfilter(foundcompounds,"Tube_blank")
# make a list of compounds in no variable and subtract _ need to this data off the GC still.
#kable(n_var_pre, caption = "Target compounds found in the No Variable negative control Pre Filter results")
#kable(n_var_bre, caption = "Target compounds found in the No Variable negative control Breakthrough results")
kable(tube_blanks, caption = "Target compounds found in the Tube blank results")
```
-ve control - No variable loaded but full 20hr sampling carried out
```{r, echo=FALSE}
 speciestubefilterlong <- function (df, i, x) {
   df %>%
    filter(grepl(i,fileName, ignore.case = TRUE)) %>%
    filter(grepl(x, fileName))%>%
    select(fileName, RT, Height, Area,
Found_target_compound,CAS_No_1,Match_1,CAS_No_2,Match_2,CAS_No_3,Match_3,CAS_No_4,Match_4,CAS_No_5,Match_5) %>%
    arrange(Found_target_compound)
 }
n_var_pre <- speciestubefilterlong(foundcompounds, "No_variable", "pre_fil")
n_var_bre <- speciestubefilterlong(foundcompounds, "No_variable", "break")
n_var_cap <- speciestubefilterlong(foundcompounds, "No_variable", "capture")
 speciestubefilter <- function (df, i, x) {
   df %>%
    filter(grepl(i,fileName, ignore.case = TRUE)) %>%
    filter(grepl(x, fileName))%>%
    select(Found_target_compound,fileName, RT, Height, Area) %>%
    arrange(Found_target_compound)
 }

n_var_pres <- speciestubefilter(foundcompounds, "No_variable", "pre_fil")
n_var_bres <- speciestubefilter(foundcompounds, "No_variable", "break")
n_var_caps <- speciestubefilter(foundcompounds, "No_variable", "capture")
# make a list of compounds in no variable and subtract _ need to this data off the GC still.
kable(n_var_pres, caption = "Target compounds found in the No Variable negative control Pre Filter results")
kable(n_var_bres, caption = "Target compounds found in the No Variable negative control Breakthrough results")
kable(n_var_caps, caption = "Target compounds found in the No Variable negative control capture results")
```
##H_sphondylium
```{r kable, echo=FALSE}
foundcompoundstemp <- foundcompounds %>%
  filter(!grepl("H_sphondylium_1_",fileName))
h_sphon_pre <- speciestubefilterlong(foundcompoundstemp, "H_sphondylium_1", "pre_fil")
h_sphon_bre <- speciestubefilterlong(foundcompoundstemp, "H_sphondylium_1", "break")
h_sphon_cap <- speciestubefilterlong(foundcompoundstemp, "H_sphondylium_1", "capture")
h_sphon_pres <- speciestubefilter(foundcompoundstemp, "H_sphondylium_1", "pre_fil")
h_sphon_bres <- speciestubefilter(foundcompoundstemp, "H_sphondylium_1", "break")
h_sphon_caps <- speciestubefilter(foundcompoundstemp, "H_sphondylium_1", "capture")
# make a list of compounds in no variable and subtract _ need to this data off the GC still.
 kable(h_sphon_pres, caption = "Target compounds found in the Heracleum sphondylium Pre Filter results")
kable(h_sphon_bres, caption = "Target compounds found in the Heracleum sphondylium Breakthrough results")
kable(h_sphon_caps, caption = "Target compounds found in the Heracleum sphondylium Capture results")
```
Angelica_sylvestris
```{r, echo=FALSE}
foundcompoundstemp <- foundcompounds %>%
  filter(!grepl("Angelica_sylvestris_1_",fileName))
Angelica_sylvestris_pre <- speciestubefilterlong(foundcompoundstemp, "Angelica_sylvestris_1", "pre_fil")
```

```
Angelica_sylvestris_bre <- speciestubefilterlong(foundcompoundstemp, "Angelica_sylvestris_1", "break")
Angelica_sylvestris_cap <- speciestubefilterlong(foundcompoundstemp, "Angelica_sylvestris_1", "capture")
Angelica_sylvestris_pres <- speciestubefilter(foundcompoundstemp, "Angelica_sylvestris_1", "pre_fil")
Angelica_sylvestris_bres <- speciestubefilter(foundcompoundstemp, "Angelica_sylvestris_1", "break")
Angelica_sylvestris_caps <- speciestubefilter(foundcompoundstemp, "Angelica_sylvestris_1", "capture")
kable(Angelica_sylvestris_pres, caption = "Target compounds found in the Angelica sylvestris Pre Filter results")
kable(Angelica_sylvestris_bres, caption = "Target compounds found in the Angelica sylvestris Breakthrough results")
kable(Angelica_sylvestris_caps, caption = "Target compounds found in the Angelica sylvestris Capture results")
```
H_helix

```{r, echo=FALSE}
h_helix_pre <- speciestubefilterlong(foundcompounds, "H_helix", "pre_fil")
h_helix_bre <- speciestubefilterlong(foundcompounds, "H_helix", "break")
h_helix_cap <- speciestubefilterlong(foundcompounds, "H_helix", "capture")
h_helix_pres <- speciestubefilter(foundcompounds, "H_helix", "pre_fil")
h_helix_bres <- speciestubefilter(foundcompounds, "H_helix", "break")
h_helix_caps <- speciestubefilter(foundcompounds, "H_helix", "capture")
kable(h_helix_pres, caption = "Target compounds found in the Hedera helix Pre Filter results")
kable(h_helix_bres, caption = "Target compounds found in the Hedera helix Breakthrough results")
kable(h_helix_caps, caption = "Target compounds found in the Hedera helix Capture results")
```
Anthriscus_sylvestris
```{r, echo=FALSE}
Anthriscus_sylvestris_pre <- speciestubefilterlong(foundcompounds, "Anthriscus_sylvestris", "pre_fil")
Anthriscus_sylvestris_bre <- speciestubefilterlong(foundcompounds, "Anthriscus_sylvestris", "break")
Anthriscus_sylvestris_cap <- speciestubefilterlong(foundcompounds, "Anthriscus_sylvestris", "capture")
Anthriscus_sylvestris_pres <- speciestubefilter(foundcompounds, "Anthriscus_sylvestris", "pre_fil")
Anthriscus_sylvestris_bres <- speciestubefilter(foundcompounds, "Anthriscus_sylvestris", "break")
Anthriscus_sylvestris_caps <- speciestubefilter(foundcompounds, "Anthriscus_sylvestris", "capture")
kable(Anthriscus_sylvestris_pres, caption = "Target compounds found in the Anthriscus sylvestris Pre Filter results")
kable(Anthriscus_sylvestris_bres, caption = "Target compounds found in the Anthriscus sylvestris Breakthrough results")
kable(Anthriscus_sylvestris_caps, caption = "Target compounds found in the Anthriscus sylvestris Capture results")
```
C_majus

```{r, echo=FALSE}
c_majus_pre <- speciestubefilterlong(foundcompounds, "c_majus", "pre_fil")
c_majus_bre <- speciestubefilterlong(foundcompounds, "c_majus", "break")
c_majus_cap <- speciestubefilterlong(foundcompounds, "c_majus", "capture")
c_majus_pres <- speciestubefilter(foundcompounds, "c_majus", "pre_fil")
c_majus_bres <- speciestubefilter(foundcompounds, "c_majus", "break")
c_majus_caps <- speciestubefilter(foundcompounds, "c_majus", "capture")
kable(c_majus_pres, caption = "Target compounds found in the Conopodium majus Pre Filter results")
kable(c_majus_bres, caption = "Target compounds found in the Conopodium majus Breakthrough results")
kable(c_majus_caps, caption = "Target compounds found in the Conopodium majus Capture results")
```
Daucus carota

```{r, echo=FALSE}
d_carota_pre <- speciestubefilterlong(foundcompounds, "d_carota", "pre_fil")
d_carota_bre <- speciestubefilterlong(foundcompounds, "d_carota", "break")
d_carota_cap <- speciestubefilterlong(foundcompounds, "d_carota", "capture")
d_carota_pres <- speciestubefilter(foundcompounds, "d_carota", "pre_fil")
d_carota_bres <- speciestubefilter(foundcompounds, "d_carota", "break")
d_carota_caps <- speciestubefilter(foundcompounds, "d_carota", "capture")
kable(d_carota_pres, caption = "Target compounds found in the Daucus carota Pre Filter results")
kable(d_carota_bres, caption = "Target compounds found in the Daucus carota Breakthrough results")
kable(d_carota_caps, caption = "Target compounds found in the Daucus carota Capture results")
```
Comparison of the number of target compounds found in each flower species
Graph of the number of target compounds by flower species with Match >= 800
```{r, generate number of different compounds per species, echo=FALSE}
#create an empty matrix to receive data
distinct_compounds <- matrix(ncol=0, nrow=0)
# enter the data into the matrix
distinct_compounds$flower_species <- c("-ve control Hexane blank", "-ve control Tube blank", "-ve control No Variable", "Heracleum sphondylium", "Angelica sylvestris", "Hedera helix", "Anthriscus sylvestris", "Conopodium majus", "Daucus carota")
distinct_compounds$distinct_compounds <- c(nrow(distinct(hexane_blank, Found_target_compound)),nrow(distinct(tube_blank, Found_target_compound)),nrow(distinct(n_var_cap, Found_target_compound)),nrow(distinct(h_sphon_cap, Found_target_compound)), nrow(distinct(Angelica_sylvestris_cap, Found_target_compound)), nrow(distinct(h_helix_cap,
```

```
Found_target_compound)),nrow(distinct(Anthriscus_sylvestris_cap, Found_target_compound)),nrow(distinct(c_majus_cap, Found_target_compound)), nrow(distinct(d_carota_cap, Found_target_compound)))
#change to data frame - makes it easy to print as a table if needed
distinct_compounds <- data.frame(distinct_compounds)
kable(distinct_compounds, caption = "The number of distinct compounds found in \n each experimental group")
```
```{r,echo=FALSE, fig.height= 5, fig.width= 7}
p <- ggplot(distinct_compounds, aes(fill=flower_species, y=distinct_compounds, x=flower_species))
p + geom_bar(position="dodge", stat="identity") +  xlab("Flower Species") + ylab("Number of different target compounds") + # Set axis labels
 ggtitle("Inter-specific comparison of the\n number of distinct target compounds") + # Set title
 theme(axis.text.x = element_text(face = "italic", angle = 90, hjust = 1)) + theme(legend.position = "none")
 ```
Graph of the number of target compounds by flower species with Match >= 900
```{r, generate number of different compounds per species 900, echo=FALSE}
hexane_blank900 <- hexane_blank %>%
 filter(Match_1 >=900)
tube_blank900 <- tube_blank %>%
 filter(Match_1 >=900)
n_var_cap900 <- n_var_cap %>%
 filter(Match_1 >=900)
h_sphon_cap900 <- h_sphon_cap %>%
 filter(Match_1 >=900)
Angelica_sylvestris_cap900 <- Angelica_sylvestris_cap %>%
 filter(Match_1 >=900)
h_helix_cap900 <- h_helix_cap %>%
 filter(Match_1 >=900)
Anthriscus_sylvestris_cap900 <- Anthriscus_sylvestris_cap %>%
 filter(Match_1 >=900)
c_majus_cap900 <- c_majus_cap %>%
 filter(Match_1 >=900)
d_carota_cap900 <- d_carota_cap %>%
 filter(Match_1 >=900)
#create an empty matrix to receive data
distinct_compounds900 <- matrix(ncol=0, nrow=0)
# enter the data into the matrix
distinct_compounds900$flower_species <- c("-ve control Hexane blank", "-ve control Tube blank", "-ve control No Variable", "Heracleum sphondylium", "Angelica sylvestris","Hedera_helix", "Anthriscus sylvestris", "Conopodium majus", "Daucus carota")
distinct_compounds900$distinct_compounds <- c(nrow(distinct(hexane_blank900, Found_target_compound)),nrow(distinct(tube_blank900, Found_target_compound)),nrow(distinct(n_var_cap900, Found_target_compound)),nrow(distinct(h_sphon_cap900, Found_target_compound)), nrow(distinct(Angelica_sylvestris_cap900, Found_target_compound)), nrow(distinct(h_helix_cap900, Found_target_compound)), nrow(distinct(Anthriscus_sylvestris_cap900, Found_target_compound)), nrow(distinct(c_majus_cap900, Found_target_compound)), nrow(distinct(d_carota_cap900, Found_target_compound)))
#change to data frame - makes it easy to print as a table if needed
distinct_compounds900 <- data.frame(distinct_compounds900)
kable(distinct_compounds900, caption = "The number of distinct compounds found in \n each experimental group where Match >= 900")
```

```{r,echo=FALSE, fig.height= 5, fig.width= 7}
p <- ggplot(distinct_compounds900, aes(fill=flower_species, y=distinct_compounds, x=flower_species))
p + geom_bar(position="dodge", stat="identity") + xlab("Flower Species") + ylab("Number of different target compounds") + # Set axis labels
 ggtitle("Inter-specific comparison of the number of\n distinct target compounds where Match >= 900 ") + # Set title
 theme(axis.text.x = element_text(face = "italic", angle = 90, hjust = 1)) + theme(legend.position = "none")
 ```
Graph of the number of unique target compounds by flower species
Variation in the VOCs found in Heracleum sphondylium by sample tube - Check that capture tube has most peaks
This uses the Match >= 800 data to
```{r H sphondylium, echo=FALSE}
#create an empty matrix to receive data
distinct_compounds_h_sphon <- matrix(ncol=0, nrow=0)
# enter the data into the matrix
distinct_compounds_h_sphon$sample_type <- c("pre filter tube", "capture tube", "breakthrough tube")
distinct_compounds_h_sphon$distinct_compounds <- c(nrow(distinct(h_sphon_pre, Found_target_compound)),nrow(distinct(h_sphon_cap, Found_target_compound)),nrow(distinct(h_sphon_bre, Found_target_compound)))
#change to data frame - makes it easy to print as a table if needed
distinct_compounds_h_sphon <- data.frame(distinct_compounds_h_sphon)
#kable(distinct_compounds_h_sphon, caption = "The number of distinct compounds found in \n H. sphondylium DHS sample tubes")
```
```{r,echo=FALSE, fig.height= 5, fig.width= 7}
```

```
p <- ggplot(distinct_compounds_h_sphon, aes(fill=sample_type, y=distinct_compounds, x=sample_type))
p + geom_bar(position="dodge", stat="identity") + scale_y_continuous(breaks = round(seq(min(distinct_compounds_h_sphon$distinct_compounds), max(distinct_compounds_h_sphon$distinct_compounds), by = 1),1)) + xlab("Sample Tube") + ylab("Number of different target compounds") + # Set axis labels
 ggtitle("The number of distinct compounds found in \n H. sphondylium sample tubes where Match >= 800 ") +  # Set title
 theme(axis.text.x = element_text(angle = 90, hjust = 1)) + theme(legend.position = "none")
 ```
Variation in the VOCs found in Angelica sylvestris by sample tube - Check that capture tube has most peaks
This uses the Match >= 800 data to
```{r A podagraria, echo=FALSE}
#create an empty matrix to receive data
distinct_compounds_Angelica_sylvestris <- matrix(ncol=0, nrow=0)
# enter the data into the matrix
distinct_compounds_Angelica_sylvestris$sample_type <- c("pre filter tube", "capture tube", "breakthrough tube")
distinct_compounds_Angelica_sylvestris$distinct_compounds <- c(nrow(distinct(Angelica_sylvestris_pre, Found_target_compound)),nrow(distinct(Angelica_sylvestris_cap, Found_target_compound)),nrow(distinct(Angelica_sylvestris_bre, Found_target_compound)))
#change to data frame - makes it easy to print as a table if needed
distinct_compounds_Angelica_sylvestris <- data.frame(distinct_compounds_Angelica_sylvestris)
#kable(distinct_compounds_Angelica_sylvestris, caption = "The number of distinct compounds found in \n Angelica sylvestris DHS sample tubes")
```
```{r,echo=FALSE, fig.height= 5, fig.width= 7}
p <- ggplot(distinct_compounds_Angelica_sylvestris, aes(fill=sample_type, y=distinct_compounds, x=sample_type))
p + geom_bar(position="dodge", stat="identity") + scale_y_continuous(breaks = round(seq(min(distinct_compounds_Angelica_sylvestris$distinct_compounds), max(distinct_compounds_Angelica_sylvestris$distinct_compounds), by = 1),1)) + xlab("Sample Tube") + ylab("Number of different target compounds") + # Set axis labels
 ggtitle("The number of distinct compounds found in \n Angelica sylvestris sample tubes where Match >= 800 ") + # Set title
 theme(axis.text.x = element_text(angle = 90, hjust = 1)) + theme(legend.position = "none")
 ```
Variation in the VOCs found in Hedera helix by sample tube - Check that capture tube has most peaks
This uses the Match >= 800 data to
```{r H helix, echo=FALSE}
#create an empty matrix to receive data
distinct_compounds_h_helix <- matrix(ncol=0, nrow=0)
# enter the data into the matrix
distinct_compounds_h_helix$sample_type <- c("pre filter tube", "capture tube", "breakthrough tube")
distinct_compounds_h_helix$distinct_compounds <- c(nrow(distinct(h_helix_pre, Found_target_compound)),nrow(distinct(h_helix_cap, Found_target_compound)),nrow(distinct(h_helix_bre, Found_target_compound)))
#change to data frame - makes it easy to print as a table if needed
distinct_compounds_h_helix <- data.frame(distinct_compounds_h_helix)
#kable(distinct_compounds_h_helix, caption = "The number of distinct compounds found in \n H. helix DHS sample tubes")

```
```{r,echo=FALSE, fig.height= 5, fig.width= 7}

p <- ggplot(distinct_compounds_h_helix, aes(fill=sample_type, y=distinct_compounds, x=sample_type))
p + geom_bar(position="dodge", stat="identity") + scale_y_continuous(breaks = round(seq(min(distinct_compounds_h_helix$distinct_compounds), max(distinct_compounds_h_helix$distinct_compounds), by = 1),1)) + xlab("Sample Tube") + ylab("Number of different target compounds") + # Set axis labels
 ggtitle("The number of distinct compounds found in \n H. helix sample tubes where Match >= 800 ") + # Set title
 theme(axis.text.x = element_text(angle = 90, hjust = 1)) + theme(legend.position = "none")
 ```

Variation in the VOCs found in Anthriscus sylvestris by sample tube - Check that capture tube has most peaks
This uses the Match >= 800 data to
```{r A sylvestris, echo=FALSE}
#create an empty matrix to receive data
distinct_compounds_Anthriscus_sylvestris <- matrix(ncol=0, nrow=0)
# enter the data into the matrix
distinct_compounds_Anthriscus_sylvestris$sample_type <- c("pre filter tube", "capture tube", "breakthrough tube")
distinct_compounds_Anthriscus_sylvestris$distinct_compounds <- c(nrow(distinct(Anthriscus_sylvestris_pre, Found_target_compound)),nrow(distinct(Anthriscus_sylvestris_cap, Found_target_compound)),nrow(distinct(Anthriscus_sylvestris_bre, Found_target_compound)))
#change to data frame - makes it easy to print as a table if needed
distinct_compounds_Anthriscus_sylvestris <- data.frame(distinct_compounds_Anthriscus_sylvestris)
#kable(distinct_compounds_Anthriscus_sylvestris, caption = "The number of distinct compounds found in \n Anthriscus sylvestris DHS sample tubes")
```
```{r,echo=FALSE, fig.height= 5, fig.width= 7}
```

```
p <- ggplot(distinct_compounds_Anthriscus_sylvestris, aes(fill=sample_type, y=distinct_compounds, x=sample_type))
p + geom_bar(position="dodge", stat="identity") + scale_y_continuous(breaks = round(seq(min(distinct_compounds_Anthriscus_sylvestris$distinct_compounds),
max(distinct_compounds_Anthriscus_sylvestris$distinct_compounds), by = 1),1)) + xlab("Sample Tube") + ylab("Number of different target compounds") + # Set axis labels
 ggtitle("The number of distinct compounds found in \n Anthriscus sylvestris sample tubes where Match >= 800 ") + # Set title
 theme(axis.text.x = element_text(angle = 90, hjust = 1)) + theme(legend.position = "none")
```

Variation in the VOCs found in Conopodium majus by sample tube - Check that capture tube has most peaks

This uses the Match >= 800 data to

```
{r C majus, echo=FALSE}

#create an empty matrix to receive data
distinct_compounds_c_majus <- matrix(ncol=0, nrow=0)

# enter the data into the matrix
distinct_compounds_c_majus$sample_type <- c("pre filter tube", "capture tube", "breakthrough tube")

distinct_compounds_c_majus$distinct_compounds <- c(nrow(distinct(c_majus_pre, Found_target_compound)),nrow(distinct(c_majus_cap, Found_target_compound)),nrow(distinct(c_majus_bre, Found_target_compound)))

#change to data frame - makes it easy to print as a table if needed
distinct_compounds_c_majus <- data.frame(distinct_compounds_c_majus)

#kable(distinct_compounds_c_majus, caption = "The number of distinct compounds found in \n C. majus DHS sample tubes")
```

```
{r,echo=FALSE, fig.height= 5, fig.width= 7}


p <- ggplot(distinct_compounds_c_majus, aes(fill=sample_type, y=distinct_compounds, x=sample_type))
p + geom_bar(position="dodge", stat="identity") + scale_y_continuous(breaks = round(seq(min(distinct_compounds_c_majus$distinct_compounds), max(distinct_compounds_c_majus$distinct_compounds), by = 1),1)) + xlab("Sample Tube") + ylab("Number of different target compounds") + # Set axis labels
 ggtitle("The number of distinct compounds found in \n C. majus sample tubes where Match >= 800 ") + # Set title
 theme(axis.text.x = element_text(angle = 90, hjust = 1)) + theme(legend.position = "none")
```

Variation in the VOCs found in Daucus carota by sample tube - Check that capture tube has most peaks

This uses the Match >= 800 data to

```
{r D carota, echo=FALSE}

#create an empty matrix to receive data
distinct_compounds_d_carota <- matrix(ncol=0, nrow=0)

# enter the data into the matrix
distinct_compounds_d_carota$sample_type <- c("pre filter tube", "capture tube", "breakthrough tube")

distinct_compounds_d_carota$distinct_compounds <- c(nrow(distinct(d_carota_pre, Found_target_compound)),nrow(distinct(d_carota_cap, Found_target_compound)),nrow(distinct(d_carota_bre, Found_target_compound)))

#change to data frame - makes it easy to print as a table if needed
distinct_compounds_d_carota <- data.frame(distinct_compounds_d_carota)

#kable(distinct_compounds_d_carota, caption = "The number of distinct compounds found in \n C. majus DHS sample tubes")
```

```
{r,echo=FALSE, fig.height= 5, fig.width= 7}
```

```
p <- ggplot(distinct_compounds_d_carota, aes(fill=sample_type, y=distinct_compounds, x=sample_type))
p + geom_bar(position="dodge", stat="identity") + scale_y_continuous(breaks = round(seq(min(distinct_compounds_d_carota$distinct_compounds), max(distinct_compounds_d_carota$distinct_compounds), by = 1),1))
+ xlab("Sample Tube") + ylab("Number of different target compounds") + # Set axis labels
 ggtitle("The number of distinct compounds found in \n D. carota sample tubes where Match >= 800 ") + # Set title
 theme(axis.text.x = element_text(angle = 90, hjust = 1)) + theme(legend.position = "none")
```

```

## Inspection of the target compounds found in the capture tubes of each species.

These data show whether there is any intra-species variation in the VOCs sampled.

## Heracleum sphondylium

```{r, echo=FALSE}

# do not include the samples 1 to 5 or we have too many to compare to other plants - just use the 2016 data
#h_sphon_cap1 <- h_sphon_caps %>%
#  filter(grepl("lium_1_",fileName)) %>%
#  arrange(Found_target_compound)

#h_sphon_cap2 <- h_sphon_caps %>%
#  filter(grepl("lium_2_",fileName)) %>%
#  arrange(Found_target_compound)

#h_sphon_cap3 <- h_sphon_caps %>%
#  filter(grepl("lium_3_",fileName)) %>%
#  arrange(Found_target_compound)

#h_sphon_cap4 <- h_sphon_caps %>%
#  filter(grepl("lium_4_",fileName)) %>%
#  arrange(Found_target_compound)

#h_sphon_cap5 <- h_sphon_caps %>%
#  filter(grepl("lium_5_",fileName)) %>%
#  arrange(Found_target_compound)

h_sphon_cap11 <- h_sphon_caps %>%
  filter(grepl("lium_11_",fileName)) %>%
  arrange(Found_target_compound)

h_sphon_cap12 <- h_sphon_caps %>%
  filter(grepl("lium_12_",fileName)) %>%
  arrange(Found_target_compound)

h_sphon_cap13 <- h_sphon_caps %>%
  filter(grepl("lium_13_",fileName)) %>%
  arrange(Found_target_compound)

h_sphon_cap14 <- h_sphon_caps %>%
  filter(grepl("lium_14_",fileName)) %>%
  arrange(Found_target_compound)

h_sphon_cap15 <- h_sphon_caps %>%
  filter(grepl("lium_15_",fileName)) %>%
  arrange(Found_target_compound)

h_sphon_cap16 <- h_sphon_caps %>%
  filter(grepl("lium_16_",fileName)) %>%
  arrange(Found_target_compound)

#kable(h_sphon_cap1, caption = "Target compounds found in H_sphondyium capture tube 1")
#kable(h_sphon_cap2, caption = "Target compounds found in H_sphondyium capture tube 2")
#kable(h_sphon_cap3, caption = "Target compounds found in H_sphondyium capture tube 3")
#kable(h_sphon_cap4, caption = "Target compounds found in H_sphondyium capture tube 4")
#kable(h_sphon_cap5, caption = "Target compounds found in H_sphondyium capture tube 5")
kable(h_sphon_cap11, caption = "Target compounds found in H_sphondyium capture tube 11")
```
```

```
kable(h_sphon_cap12, caption = "Target compounds found in H_sphondyium capture tube 12")
kable(h_sphon_cap13, caption = "Target compounds found in H_sphondyium capture tube 13")
kable(h_sphon_cap14, caption = "Target compounds found in H_sphondyium capture tube 14")
kable(h_sphon_cap15, caption = "Target compounds found in H_sphondyium capture tube 15")
kable(h_sphon_cap16, caption = "Target compounds found in H_sphondyium capture tube 16")
```

Angelica sylvestris

```{r, echo=FALSE}
# do not include the samples 1 to 5 or we have too many to compare to other plants - just use the 2016 data

#Angelica_sylvestris_cap1 <- Angelica_sylvestris_caps %>%
#  filter(grepl("Angelica_sylvestris_1_",fileName)) %>%
#  arrange(Found_target_compound)

#Angelica_sylvestris_cap2 <- Angelica_sylvestris_caps %>%
#  filter(grepl("Angelica_sylvestris_2_",fileName)) %>%
#  arrange(Found_target_compound)

#Angelica_sylvestris_cap3 <- Angelica_sylvestris_caps %>%
#  filter(grepl("Angelica_sylvestris_3_",fileName)) %>%
#  arrange(Found_target_compound)

#Angelica_sylvestris_cap4 <- Angelica_sylvestris_caps %>%
#  filter(grepl("Angelica_sylvestris_4_",fileName)) %>%
#  arrange(Found_target_compound)

#Angelica_sylvestris_cap5 <- Angelica_sylvestris_caps %>%
#  filter(grepl("Angelica_sylvestris_5_",fileName)) %>%
#  arrange(Found_target_compound)

Angelica_sylvestris_cap6 <- Angelica_sylvestris_caps %>%
  filter(grepl("Angelica_sylvestris_6_",fileName)) %>%
  arrange(Found_target_compound)

Angelica_sylvestris_cap11 <- Angelica_sylvestris_caps %>%
  filter(grepl("Angelica_sylvestris_11_",fileName)) %>%
  arrange(Found_target_compound)

Angelica_sylvestris_cap12 <- Angelica_sylvestris_caps %>%
  filter(grepl("Angelica_sylvestris_12_",fileName)) %>%
  arrange(Found_target_compound)

Angelica_sylvestris_cap13 <- Angelica_sylvestris_caps %>%
  filter(grepl("Angelica_sylvestris_13_",fileName)) %>%
  arrange(Found_target_compound)

Angelica_sylvestris_cap14 <- Angelica_sylvestris_caps %>%
  filter(grepl("Angelica_sylvestris_14_",fileName)) %>%
  arrange(Found_target_compound)

#kable(Angelica_sylvestris_cap1, caption = "Target compounds found in Angelica sylvestris capture tube 1")
#kable(Angelica_sylvestris_cap2, caption = "Target compounds found in Angelica sylvestris capture tube 2")
#kable(Angelica_sylvestris_cap3, caption = "Target compounds found in Angelica sylvestris capture tube 3")
#kable(Angelica_sylvestris_cap4, caption = "Target compounds found in Angelica sylvestris capture tube 4")
#kable(Angelica_sylvestris_cap5, caption = "Target compounds found in Angelica sylvestris capture tube 5")
kable(Angelica_sylvestris_cap6, caption = "Target compounds found in Angelica sylvestris capture tube 6")
kable(Angelica_sylvestris_cap11, caption = "Target compounds found in Angelica sylvestris capture tube 11")
kable(Angelica_sylvestris_cap12, caption = "Target compounds found in Angelica sylvestris capture tube 12")
kable(Angelica_sylvestris_cap13, caption = "Target compounds found in Angelica sylvestris capture tube 13")
kable(Angelica_sylvestris_cap14, caption = "Target compounds found in Angelica sylvestris capture tube 14")
```

Hedera helix

```{r, echo=FALSE}
```

```
h_helix_cap1 <- h_helix_caps %>%
  filter(grepl("helix_1_",fileName)) %>%
  arrange(Found_target_compound)

h_helix_cap2 <- h_helix_caps %>%
  filter(grepl("h_helix_2_",fileName)) %>%
  arrange(Found_target_compound)

h_helix_cap3 <- h_helix_caps %>%
  filter(grepl("h_helix_3",fileName)) %>%
  arrange(Found_target_compound)

h_helix_cap4 <- h_helix_caps %>%
  filter(grepl("h_helix_4",fileName)) %>%
  arrange(Found_target_compound)

h_helix_cap5 <- h_helix_caps %>%
  filter(grepl("h_helix_5",fileName)) %>%
  arrange(Found_target_compound)


kable(h_helix_cap1, caption = "Target compounds found in H. helix capture tube 1")
kable(h_helix_cap2, caption = "Target compounds found in H. helix capture tube 2")
kable(h_helix_cap3, caption = "Target compounds found in H. helix capture tube 3")
kable(h_helix_cap4, caption = "Target compounds found in H. helix capture tube 4")
kable(h_helix_cap5, caption = "Target compounds found in H. helix capture tube 5")
```

```
## Anthriscus sylvestris
```

```{r, echo=FALSE}
Anthriscus_sylvestris_cap1 <- Anthriscus_sylvestris_caps %>%
  filter(grepl("Anthriscus_sylvestris_1_",fileName)) %>%
  arrange(Found_target_compound)

Anthriscus_sylvestris_cap2 <- Anthriscus_sylvestris_caps %>%
  filter(grepl("Anthriscus_sylvestris_2_",fileName)) %>%
  arrange(Found_target_compound)

Anthriscus_sylvestris_cap3 <- Anthriscus_sylvestris_caps %>%
  filter(grepl("Anthriscus_sylvestris_3",fileName)) %>%
  arrange(Found_target_compound)

Anthriscus_sylvestris_cap4 <- Anthriscus_sylvestris_caps %>%
  filter(grepl("Anthriscus_sylvestris_4",fileName)) %>%
  arrange(Found_target_compound)

Anthriscus_sylvestris_cap5 <- Anthriscus_sylvestris_caps %>%
  filter(grepl("Anthriscus_sylvestris_5",fileName)) %>%
  arrange(Found_target_compound)


kable(Anthriscus_sylvestris_cap1, caption = "Target compounds found in Anthriscus sylvestris capture tube 1")
kable(Anthriscus_sylvestris_cap2, caption = "Target compounds found in Anthriscus sylvestris capture tube 2")
kable(Anthriscus_sylvestris_cap3, caption = "Target compounds found in Anthriscus sylvestris capture tube 3")
kable(Anthriscus_sylvestris_cap4, caption = "Target compounds found in Anthriscus sylvestris capture tube 4")
kable(Anthriscus_sylvestris_cap5, caption = "Target compounds found in Anthriscus sylvestris capture tube 5")
```

```
## Conopodium majus
```

```{r, echo=FALSE}
```

```
c_majus_cap1 <- c_majus_caps %>%
  filter(grepl("c_majus_1_",fileName)) %>%
  arrange(Found_target_compound)

c_majus_cap2 <- c_majus_caps %>%
  filter(grepl("c_majus_2_",fileName)) %>%
  arrange(Found_target_compound)

c_majus_cap3 <- c_majus_caps %>%
  filter(grepl("c_majus_3",fileName)) %>%
  arrange(Found_target_compound)

c_majus_cap4 <- c_majus_caps %>%
  filter(grepl("c_majus_4",fileName)) %>%
  arrange(Found_target_compound)

c_majus_cap5 <- c_majus_caps %>%
  filter(grepl("c_majus_5",fileName)) %>%
  arrange(Found_target_compound)


kable(c_majus_cap1, caption = "Target compounds found in C. majus capture tube 1")
kable(c_majus_cap2, caption = "Target compounds found in C. majus capture tube 2")
kable(c_majus_cap3, caption = "Target compounds found in C. majus capture tube 3")
kable(c_majus_cap4, caption = "Target compounds found in C. majus capture tube 4")
kable(c_majus_cap5, caption = "Target compounds found in C. majus capture tube 5")
```

```
## Daucus carota
```

```{r, echo=FALSE}

d_carota_cap1 <- d_carota_caps %>%
  filter(grepl("d_carota_1_",fileName)) %>%
  arrange(Found_target_compound)

d_carota_cap2 <- d_carota_caps %>%
  filter(grepl("d_carota_2_",fileName)) %>%
  arrange(Found_target_compound)

d_carota_cap3 <- d_carota_caps %>%
  filter(grepl("d_carota_3",fileName)) %>%
  arrange(Found_target_compound)

d_carota_cap4 <- d_carota_caps %>%
  filter(grepl("d_carota_4",fileName)) %>%
  arrange(Found_target_compound)

d_carota_cap5 <- d_carota_caps %>%
  filter(grepl("d_carota_5",fileName)) %>%
  arrange(Found_target_compound)


kable(d_carota_cap1, caption = "Target compounds found in Daucus carota capture tube 1")
kable(d_carota_cap2, caption = "Target compounds found in Daucus carota capture tube 2")
kable(d_carota_cap3, caption = "Target compounds found in Daucus carota capture tube 3")
kable(d_carota_cap4, caption = "Target compounds found in Daucus carota capture tube 4")
kable(d_carota_cap5, caption = "Target compounds found in Daucus carota capture tube 5")
```

```
# Semi-quantitative section - ppm determination (Hedera helix only at this time)
```

Test the calculation of the ppm of target compounds in the Hedera helix samples. We have a known ppm of 1-Bromodecane added to the Hedera helix samples as an internal standard.

```{r ppm of 1-Bromodecane}

#Parts per million calculator


# ppm = 1g/m3 = 1mg/L = 1??g/mL

#convert the mass to milligrams

#e.g. A solution has a concentration of 1.25g/L.
#What is its concentration in ppm?

#Convert the mass in grams to a mass in milligrams:
#  1.25g = 1.25 x 1000mg = 1250mg

#Re-write the concentration in mg/L = 1250mg/L = 1250ppm

#edit the below to match solution

#some info for 1-Bromodecane used

# CAS 112-29-8
# Purity 98%
# Molecular weight 221.18
# Linear formula CH3(CH2)9Br
# Density 1.066g/ml at 25 C(lit.)

#ppm of 1st internal standard solution

purity <- .98   # convert from percentage purity to decimal

solutedensity <- 1.066  #g/ml

initialsolutevol <-  2.5e-03 #ml volume in ml of 1-bromo decane stock added

initialsolventvol <- 2.5 #ml  volume in ml of hexane as solvent


solutionvol <- initialsolutevol + initialsolventvol # ml

rawsolutemassmg <-  ((solutedensity*initialsolutevol)*purity)*1000  #mass of solute in mg

firstppm <- (1000/solutionvol)*rawsolutemassmg

# ppm for serial dilution

dilutionfactor <- 1/5 # ratio of dilution  as a fraction i.e. 1:10 would be 1/10

secondppm <- firstppm*dilutionfactor

# 2 ul of this solution is added to each sample. Each pre-evaporation sample is ~2.5 ml. 3 ml hexane is used for solvent desorption but some is lost due to evaporation during the process. However, because the 1-bromodecane is significantly less volatile than the solvent or any of the target compounds being collected, the solvent volume of interest is the post evaporation volume of 0.175 ml

ISppm <- secondppm * (0.002/0.175)

# no further calculation needs to be applied to the internal standard as all processes which may transform the ppm of the internal standard are applied to the IS from now on. The ppm of in the pre-evaporation sample is directly proportional to the target compounds found.

ppm_in_air <- ISppm * (0.175/ (355*60*20))   # this is based on 95% pump speed and capture in fume cupboard in R121. See Notebook 5 P25 N12. Mean ml/min was 355. This is multiplied to get ml in 20 hrs


```

The ppm of 1-Bromodecane added to each sample, pre-evaporation, is `r ISppm`. The actual volume of 1-Bromodecane added is 2 ul of `r ISppm` ppm.

It follows that, therefore, the ppm of any other compound in the sample is the ppm of 1-Bromodecane multiplied by the ratio of their integrated peak to that of 1-Bromodecane.(Chemistry and biology of volatiles Hermann 2010) (Response factor of sample/Response factor of standard=RRF)

The tables below show the ppm of the target compounds found in each Hedera helix. The ppm concentrated into the solution by the DHS process are shown, as is the ppm in air, i.e. the ppm in the solution divided by the volumn of air drawn through the DHS entrainment arena

```{r target compound ppm determination, echo= FALSE}

#THE CODE BELOW filters SO THAT IT ONLY retains IS and MATCH VALUE => 800
foundIS <- gcdata %>%
 filter(CAS_No_1 %in% ISCAS & Match_1 >= 800 | CAS_No_2 %in% ISCAS & Match_2 >= 800 | CAS_No_3 %in% ISCAS & Match_3 >= 800 | CAS_No_4 %in% ISCAS & Match_4 >= 800 | CAS_No_5 %in% ISCAS & Match_5 >= 800)

IS <- foundIS %>%
 filter(grepl("capture",fileName, ignore.case = TRUE)) %>%
 select(fileName, RT, Height, Area) %>%
 filter(!grepl("tube_blank", fileName, ignore.case = TRUE)) %>%
 arrange(fileName)

kable(IS, caption = "1-Bromodecane (IS) peak Areas in capture results")

# collect the IS level in each capture sample

h_helix_1_capture_IS <- IS %>%
 filter(grepl("h_helix_1_capture",fileName, ignore.case = TRUE))

h_helix_2_capture_IS <- IS %>%
 filter(grepl("h_helix_2_capture",fileName, ignore.case = TRUE))

h_helix_3_capture_IS <- IS %>%
 filter(grepl("h_helix_3_capture",fileName, ignore.case = TRUE))

h_helix_4_capture_IS <- IS %>%
 filter(grepl("h_helix_4_capture",fileName, ignore.case = TRUE))

h_helix_5_capture_IS <- IS %>%
 filter(grepl("h_helix_5_capture",fileName, ignore.case = TRUE))

# Response factor of sample/Response factor of standard=RRF

#Capture tube 1

 h_helix_1_cap <- foundcompounds %>%
 filter(grepl("h_helix_1",fileName, ignore.case = TRUE)) %>%
 filter(grepl("capture", fileName, ignore.case = TRUE))%>%
 select(Found_target_compound,fileName, RT, Height, Area) %>%
  mutate(response_factor = Area / h_helix_1_capture_IS$Area) %>%
  mutate(ppm_in_sample = response_factor*ISppm) %>%
  mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
 arrange(Found_target_compound)


#Capture tube 2

 h_helix_2_cap <- foundcompounds %>%
 filter(grepl("h_helix_2",fileName, ignore.case = TRUE)) %>%
 filter(grepl("capture", fileName, ignore.case = TRUE))%>%
 select(Found_target_compound,fileName, RT, Height, Area) %>%
  mutate(response_factor = Area / h_helix_2_capture_IS$Area) %>%
  mutate(ppm_in_sample = response_factor*ISppm) %>%
  mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
 arrange(Found_target_compound)


#Capture tube 3

 h_helix_3_cap <- foundcompounds %>%
 filter(grepl("h_helix_3",fileName, ignore.case = TRUE)) %>%
 filter(grepl("capture", fileName, ignore.case = TRUE))%>%
 select(Found_target_compound,fileName, RT, Height, Area) %>%
  mutate(response_factor = Area / h_helix_3_capture_IS$Area) %>%
  mutate(ppm_in_sample = response_factor*ISppm) %>%
  mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
```

```
 arrange(Found_target_compound)


#Capture tube 4

 h_helix_4_cap <- foundcompounds %>%
  filter(grepl("h_helix_4",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / h_helix_4_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)


#Capture tube 5

 h_helix_5_cap <- foundcompounds %>%
  filter(grepl("h_helix_5",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / h_helix_5_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)



kable(h_helix_1_cap, digits = c(0,0,3,0,1,5,10,10), caption = "Target compounds' response factor and derived ppm, found in Hedera helix sample 1 Capture results")
kable(h_helix_2_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Hedera helix sample 2 Capture results")
kable(h_helix_3_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Hedera helix sample 3 Capture results")
kable(h_helix_4_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Hedera helix sample 4 Capture results")
kable(h_helix_5_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Hedera helix sample 5 Capture results")

##########################################################################################################
```

```{r}
# collect the IS level in each Anthriscus sylvestris capture sample

Anthriscus_sylvestris_1_capture_IS <- IS %>%
  filter(grepl("Anthriscus_sylvestris_1_capture",fileName, ignore.case = TRUE))

Anthriscus_sylvestris_2_capture_IS <- IS %>%
  filter(grepl("Anthriscus_sylvestris_2_capture",fileName, ignore.case = TRUE))

Anthriscus_sylvestris_3_capture_IS <- IS %>%
  filter(grepl("Anthriscus_sylvestris_3_capture",fileName, ignore.case = TRUE))

Anthriscus_sylvestris_4_capture_IS <- IS %>%
  filter(grepl("Anthriscus_sylvestris_4_capture",fileName, ignore.case = TRUE))

Anthriscus_sylvestris_5_capture_IS <- IS %>%
  filter(grepl("Anthriscus_sylvestris_5_capture",fileName, ignore.case = TRUE))

# Response factor of sample/Response factor of standard=RRF

# Anthriscus sylvestris Capture tube 1

 Anthriscus_sylvestris_1_cap <- foundcompounds %>%
  filter(grepl("Anthriscus_sylvestris_1",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Anthriscus_sylvestris_1_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
```

```
  mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
 arrange(Found_target_compound)


#Anthriscus sylvestris Capture tube 2

 Anthriscus_sylvestris_2_cap <- foundcompounds %>%
  filter(grepl("Anthriscus_sylvestris_2",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Anthriscus_sylvestris_2_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)


#Anthriscus sylvestris Capture tube 3

 Anthriscus_sylvestris_3_cap <- foundcompounds %>%
  filter(grepl("Anthriscus_sylvestris_3",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Anthriscus_sylvestris_3_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)

#Anthriscus sylvestris Capture tube 4

 Anthriscus_sylvestris_4_cap <- foundcompounds %>%
  filter(grepl("Anthriscus_sylvestris_4",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Anthriscus_sylvestris_4_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)


#Anthriscus sylvestris Capture tube 5

 Anthriscus_sylvestris_5_cap <- foundcompounds %>%
  filter(grepl("Anthriscus_sylvestris_5",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Anthriscus_sylvestris_5_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)




kable(Anthriscus_sylvestris_1_cap, digits = c(0,0,3,0,1,5,10,10), caption = "Target compounds' response factor and derived ppm, found in Anthriscus sylvestris sample 1 Capture results")
kable(Anthriscus_sylvestris_2_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Anthriscus sylvestris sample 2 Capture results")
kable(Anthriscus_sylvestris_3_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Anthriscus sylvestris sample 3 Capture results")
kable(Anthriscus_sylvestris_4_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Anthriscus sylvestris sample 4 Capture results")
kable(Anthriscus_sylvestris_5_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Anthriscus sylvestris sample 5 Capture results")


##############################################################################################################
```


```{r}
# collect the IS level in each Conopodium majus capture sample
```

```
c_majus_1_capture_IS <- IS %>%
  filter(grepl("c_majus_1_capture",fileName, ignore.case = TRUE))

c_majus_2_capture_IS <- IS %>%
  filter(grepl("c_majus_2_capture",fileName, ignore.case = TRUE))

c_majus_3_capture_IS <- IS %>%
  filter(grepl("c_majus_3_capture",fileName, ignore.case = TRUE))

c_majus_4_capture_IS <- IS %>%
  filter(grepl("c_majus_4_capture",fileName, ignore.case = TRUE))

c_majus_5_capture_IS <- IS %>%
  filter(grepl("c_majus_5_capture",fileName, ignore.case = TRUE))


# Response factor of sample/Response factor of standard=RRF

# Conopodium majus Capture tube 1

 c_majus_1_cap <- foundcompounds %>%
  filter(grepl("c_majus_1",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / c_majus_1_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)

#Conopodium majus Capture tube 2

 c_majus_2_cap <- foundcompounds %>%
  filter(grepl("c_majus_2",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / c_majus_2_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)


#Conopodium majus Capture tube 3

 c_majus_3_cap <- foundcompounds %>%
  filter(grepl("c_majus_3",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / c_majus_3_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)


#Conopodium majus Capture tube 4

 c_majus_4_cap <- foundcompounds %>%
  filter(grepl("c_majus_4",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / c_majus_4_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)


#Conopodium majus Capture tube 5

 c_majus_5_cap <- foundcompounds %>%
  filter(grepl("c_majus_5",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
```

```
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / c_majus_5_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)


kable(c_majus_1_cap, digits = c(0,0,3,0,1,5,10,10), caption = "Target compounds' response factor and derived ppm, found in Conopodium majus sample 1 Capture results")
kable(c_majus_2_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Conopodium majus sample 2 Capture results")
kable(c_majus_3_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Conopodium majus sample 3 Capture results")
kable(c_majus_4_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Conopodium majus sample 4 Capture results")
kable(c_majus_5_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Conopodium majus sample 5 Capture results")
```
```{r}
# collect the IS level in each Heracleum sphondlyium capture sample
h_sphon_11_capture_IS <- IS %>%
  filter(grepl("h_sphondylium_11_capture",fileName, ignore.case = TRUE))
h_sphon_12_capture_IS <- IS %>%
  filter(grepl("h_sphondylium_12_capture",fileName, ignore.case = TRUE))
h_sphon_13_capture_IS <- IS %>%
  filter(grepl("h_sphondylium_13_capture",fileName, ignore.case = TRUE))
h_sphon_14_capture_IS <- IS %>%
 filter(grepl("h_sphondylium_14_capture",fileName, ignore.case = TRUE))
h_sphon_15_capture_IS <- IS %>%
  filter(grepl("h_sphondylium_15_capture",fileName, ignore.case = TRUE))
h_sphon_16_capture_IS <- IS %>%
  filter(grepl("h_sphondylium_16_capture",fileName, ignore.case = TRUE))

# Response factor of sample/Response factor of standard=RRF
# Heracleum sphondylium Capture tube 11
h_sphon_11_cap <- foundcompounds %>%
  filter(grepl("h_sphondylium_11",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / h_sphon_11_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
#Heracleum sphondylium Capture tube 12
h_sphon_12_cap <- foundcompounds %>%
  filter(grepl("h_sphondylium_12",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / h_sphon_12_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
 #Heracleum sphondylium Capture tube 13
h_sphon_13_cap <- foundcompounds %>%
  filter(grepl("h_sphondylium_13",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / h_sphon_13_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
#Heracleum sphondylium Capture tube 14
 h_sphon_14_cap <- foundcompounds %>%
  filter(grepl("h_sphondylium_14",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / h_sphon_14_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
#Heracleum sphondylium Capture tube 15
```

```
h_sphon_15_cap <- foundcompounds %>%
  filter(grepl("h_sphondylium_15",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / h_sphon_15_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)

 #Heracleum sphondylium Capture tube 16
 h_sphon_16_cap <- foundcompounds %>%
  filter(grepl("h_sphondylium_16",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / h_sphon_16_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)


kable(h_sphon_11_cap, digits = c(0,0,3,0,1,5,10,10), caption = "Target compounds' response factor and derived ppm, found in Heracleum sphondylium sample 11 Capture results")
kable(h_sphon_12_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Heracleum sphondylium sample 12 Capture results")
kable(h_sphon_13_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Heracleum sphondylium sample 13 Capture results")
kable(h_sphon_14_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Heracleum sphondylium sample 14 Capture results")
kable(h_sphon_15_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Heracleum sphondylium sample 15 Capture results")
kable(h_sphon_16_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Heracleum sphondylium sample 16 Capture results")
```
```{r}
# collect the IS level in each Angelica sylvestris capture sample

Angelica_sylvestris_11_capture_IS <- IS %>%
  filter(grepl("Angelica_sylvestris_11_capture",fileName, ignore.case = TRUE))
Angelica_sylvestris_12_capture_IS <- IS %>%
  filter(grepl("Angelica_sylvestris_12_capture",fileName, ignore.case = TRUE))
Angelica_sylvestris_13_capture_IS <- IS %>%
  filter(grepl("Angelica_sylvestris_13_capture",fileName, ignore.case = TRUE))
Angelica_sylvestris_14_capture_IS <- IS %>%
  filter(grepl("Angelica_sylvestris_14_capture",fileName, ignore.case = TRUE))
Angelica_sylvestris_15_capture_IS <- IS %>%
  filter(grepl("Angelica_sylvestris_15_capture",fileName, ignore.case = TRUE))
Angelica_sylvestris_16_capture_IS <- IS %>%
  filter(grepl("Angelica_sylvestris_16_capture",fileName, ignore.case = TRUE))

# Response factor of sample/Response factor of standard=RRF
# Angelica sylvestris Capture tube 11

 Angelica_sylvestris_11_cap <- foundcompounds %>%
  filter(grepl("Angelica_sylvestris_11",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Angelica_sylvestris_11_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)

#Angelica sylvestris Capture tube 12
 Angelica_sylvestris_12_cap <- foundcompounds %>%
  filter(grepl("Angelica_sylvestris_12",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Angelica_sylvestris_12_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
```

```
#Angelica sylvestris Capture tube 13  No Target compounds in this tube
Angelica_sylvestris_13_cap <- foundcompounds %>%
  filter(grepl("Angelica_sylvestris_13",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Angelica_sylvestris_13_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
#Angelica sylvestris Capture tube 14
Angelica_sylvestris_14_cap <- foundcompounds %>%
  filter(grepl("Angelica_sylvestris_14",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Angelica_sylvestris_14_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
#Angelica sylvestris Capture tube 15
Angelica_sylvestris_15_cap <- foundcompounds %>%
  filter(grepl("Angelica_sylvestris_15",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / Angelica_sylvestris_15_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)

 #Angelica sylvestris Capture tube 16
# Angelica_sylvestris_16_cap <- foundcompounds %>%
#  filter(grepl("Angelica_sylvestris_16",fileName, ignore.case = TRUE)) %>%
#  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
#  select(Found_target_compound,fileName, RT, Height, Area) %>%
#  mutate(response_factor = Area / Angelica_sylvestris_16_capture_IS$Area) %>%
#   mutate(ppm_in_sample = response_factor*ISppm) %>%
#   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
#  arrange(Found_target_compound)


kable(Angelica_sylvestris_11_cap, digits = c(0,0,3,0,1,5,10,10), caption = "Target compounds' response factor and derived ppm, found in Angelica sylvestris sample 11 Capture results")
kable(Angelica_sylvestris_12_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Angelica sylvestris sample 12 Capture results")
kable(Angelica_sylvestris_13_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Angelica sylvestris sample 13 Capture results")
kable(Angelica_sylvestris_14_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Angelica sylvestris sample 14 Capture results")
kable(Angelica_sylvestris_15_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Angelica sylvestris sample 15 Capture results")
#kable(Angelica_sylvestris_16_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Angelica sylvestris sample 16 Capture results")
```
```{r}
# collect the IS level in each Daucus carota capture sample
d_carota_1_capture_IS <- IS %>%
  filter(grepl("d_carota_1_capture",fileName, ignore.case = TRUE))
d_carota_2_capture_IS <- IS %>%
  filter(grepl("d_carota_2_capture",fileName, ignore.case = TRUE))
d_carota_3_capture_IS <- IS %>%
  filter(grepl("d_carota_3_capture",fileName, ignore.case = TRUE))
d_carota_4_capture_IS <- IS %>%
  filter(grepl("d_carota_4_capture",fileName, ignore.case = TRUE))
d_carota_5_capture_IS <- IS %>%
  filter(grepl("d_carota_5_capture",fileName, ignore.case = TRUE))


# Response factor of sample/Response factor of standard=RRF
# Daucus carota Capture tube 1
d_carota_1_cap <- foundcompounds %>%
  filter(grepl("d_carota_1",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
```

```
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / d_carota_1_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
#Daucus carota Capture tube 2
 d_carota_2_cap <- foundcompounds %>%
  filter(grepl("d_carota_2",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / d_carota_2_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
 #Daucus carota Capture tube 3
d_carota_3_cap <- foundcompounds %>%
  filter(grepl("d_carota_3",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / d_carota_3_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
#Daucus carota Capture tube 4
d_carota_4_cap <- foundcompounds %>%
  filter(grepl("d_carota_4",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / d_carota_4_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)
#Daucus carota Capture tube 5
 d_carota_5_cap <- foundcompounds %>%
  filter(grepl("d_carota_5",fileName, ignore.case = TRUE)) %>%
  filter(grepl("capture", fileName, ignore.case = TRUE))%>%
  select(Found_target_compound,fileName, RT, Height, Area) %>%
   mutate(response_factor = Area / d_carota_5_capture_IS$Area) %>%
   mutate(ppm_in_sample = response_factor*ISppm) %>%
   mutate(ppm_in_atmosphere = response_factor*ppm_in_air) %>%
  arrange(Found_target_compound)

kable(d_carota_1_cap, digits = c(0,0,3,0,1,5,10,10), caption = "Target compounds' response factor and derived ppm, found in Daucus carota
sample 1 Capture results")
kable(d_carota_2_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Daucus carota
sample 2 Capture results")
kable(d_carota_3_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Daucus carota
sample 3 Capture results")
kable(d_carota_4_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Daucus carota
sample 4 Capture results")
kable(d_carota_5_cap, digits = c(0,0,3,0,1,5,10,10),caption = "Target compounds' response factor and derived ppm, found in Daucus carota
sample 5 Capture results")
```
graphs of ppm

```{r}

# make an object based on the target data, minus DEEt and DDT
template <- targets %>%
  filter( !grepl("DEET", Compound_name) ) %>%
  filter( !grepl("DDT", Compound_name) ) %>%
  select(Compound_name) %>%
  rename(Found_target_compound = Compound_name)
# Function to group the by compounds and bind to the template
sampspeciesppm <- function (i) {
  test <- i %>%
   group_by(Found_target_compound)
   test <- full_join(template, test, by= "Found_target_compound")
  test <- test %>%
    replace(is.na(.), 0)
}
```

```
# Function to group the by compounds, calculate the mean and bind to the template
sampspeciesmedianppm <- function(i) {
 test <- i %>%
  group_by(Found_target_compound) %>%
 summarise(avg = median(ppm_in_sample))
 test <- full_join(template, test, by= "Found_target_compound")
  test <- test %>%
  replace(is.na(.), 0)
}



# plot point graphs for the mean data
d_carota_ppms <- rbind(d_carota_1_cap, d_carota_2_cap, d_carota_3_cap, d_carota_4_cap, d_carota_5_cap)
c_majus_ppms <- rbind(c_majus_1_cap, c_majus_2_cap, c_majus_3_cap, c_majus_4_cap, c_majus_5_cap)
h_helix_ppms <- rbind(h_helix_1_cap, h_helix_2_cap, h_helix_3_cap, h_helix_4_cap, h_helix_5_cap)
h_sphon_ppms <- rbind(h_sphon_11_cap, h_sphon_12_cap, h_sphon_13_cap, h_sphon_14_cap, h_sphon_15_cap, h_sphon_16_cap)
Angelica_sylvestris_ppms <- rbind(Angelica_sylvestris_11_cap, Angelica_sylvestris_12_cap, Angelica_sylvestris_13_cap,
Angelica_sylvestris_14_cap, Angelica_sylvestris_15_cap)
Anthriscus_sylvestris_ppms <- rbind(Anthriscus_sylvestris_1_cap, Anthriscus_sylvestris_2_cap, Anthriscus_sylvestris_3_cap,
Anthriscus_sylvestris_4_cap, Anthriscus_sylvestris_5_cap)

#run both functions over the data saving to objects
d_carota_coll_ppms <- sampspeciesppm(d_carota_ppms )
d_carota_median_ppms <- sampspeciesmedianppm(d_carota_ppms )
c_majus_coll_ppms <- sampspeciesppm(c_majus_ppms )
c_majus_median_ppms <- sampspeciesmedianppm(c_majus_ppms )
h_helix_coll_ppms <- sampspeciesppm(h_helix_ppms )
h_helix_median_ppms <- sampspeciesmedianppm(h_helix_ppms )
h_sphon_coll_ppms <- sampspeciesppm(h_sphon_ppms )
h_sphon_median_ppms <- sampspeciesmedianppm(h_sphon_ppms )
Angelica_sylvestris_coll_ppms <- sampspeciesppm(Angelica_sylvestris_ppms )
Angelica_sylvestris_median_ppms <- sampspeciesmedianppm(Angelica_sylvestris_ppms )
Anthriscus_sylvestris_coll_ppms <- sampspeciesppm(Anthriscus_sylvestris_ppms )
Anthriscus_sylvestris_median_ppms <- sampspeciesmedianppm(Anthriscus_sylvestris_ppms )

#remove the redundant internal standard from the graphs, leaving behind only those compounds found in the headspace
d_carota_coll_ppms <- d_carota_coll_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
d_carota_median_ppms <- d_carota_median_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
c_majus_coll_ppms <- c_majus_coll_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
c_majus_median_ppms <- c_majus_median_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
h_helix_coll_ppms <- h_helix_coll_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
h_helix_median_ppms <- h_helix_median_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
h_sphon_coll_ppms <- h_sphon_coll_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
h_sphon_median_ppms <- h_sphon_median_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
Angelica_sylvestris_coll_ppms <- Angelica_sylvestris_coll_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
Angelica_sylvestris_coll_ppms <- Angelica_sylvestris_coll_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
Anthriscus_sylvestris_coll_ppms <- Anthriscus_sylvestris_coll_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
Anthriscus_sylvestris_median_ppms <- Anthriscus_sylvestris_median_ppms %>%
 filter(Found_target_compound != "1-Bromodecane")
#plot boxplot of the data for collated function
ggplot(d_carota_coll_ppms, aes(Found_target_compound, ppm_in_sample)) +
 geom_boxplot() + xlab("Target Compound") + ylab("ppm per sample") +
 ggtitle("Parts per million for target compounds \n Daucus carota sample tubes where Match >= 800 ") +
 theme(axis.text.x = element_text(angle = 90, hjust = 1))

ggplot(c_majus_coll_ppms, aes(Found_target_compound, ppm_in_sample)) +
 geom_boxplot() + xlab("Target Compound") + ylab("ppm per sample") +
```

```
  ggtitle("Parts per million for target compounds \n C. majus sample tubes where Match >= 800 ") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

ggplot(h_helix_coll_ppms, aes(Found_target_compound, ppm_in_sample)) +
  geom_boxplot() + xlab("Target Compound") + ylab("ppm per sample") +
  ggtitle("Parts per million for target compounds \n H. helix sample tubes where Match >= 800 ") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

ggplot(h_sphon_coll_ppms, aes(Found_target_compound, ppm_in_sample)) +
  geom_boxplot() + xlab("Target Compound") + ylab("ppm per sample") +
  ggtitle("Parts per million for target compounds \n H. sphondylium sample tubes where Match >= 800 ") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

ggplot(Angelica_sylvestris_coll_ppms, aes(Found_target_compound, ppm_in_sample)) +
  geom_boxplot() + xlab("Target Compound") + ylab("ppm per sample") +
  ggtitle("Parts per million for target compounds \n Ang. sylvestris sample tubes where Match >= 800 ") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

ggplot(Anthriscus_sylvestris_coll_ppms, aes(Found_target_compound, ppm_in_sample)) +
  geom_boxplot() + xlab("Target Compound") + ylab("ppm per sample") +
  ggtitle("Parts per million for target compounds \n Ant. sylvestris sample tubes where Match >= 800 ") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))

#then need to collate the x_x_cap objects for each species to a collated object - need the mulitplot function too
```
```{r, fig.width=10, fig.height=6}
p1y <- ggplot(h_sphon_median_ppms, aes( y=avg, x=Found_target_compound))  + ylim(0, 12) +
 geom_point( stat="identity") + xlab("Target Compound") + ylab("Median ppm per sample") + # Set axis labels
  ggtitle("H. sphondylium") +
  theme(text = element_text(size=8),axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(axis.text.x=element_blank(), plot.title = element_text(face = "italic"))

p2y <- ggplot(h_helix_median_ppms, aes( y=avg, x=Found_target_compound))  + ylim(0, 12) +
 geom_point(stat="identity") + xlab("Target Compound") + ylab("Median ppm per sample") + # Set axis labels
  ggtitle("H. helix") + # Set title
  theme(text = element_text(size=8),axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(axis.text.x=element_blank(), plot.title = element_text(face = "italic"))

p3y <- ggplot(Angelica_sylvestris_median_ppms, aes( y=avg, x=Found_target_compound))  + ylim(0, 12) +
 geom_point( stat="identity") + xlab("Target Compound") + ylab("Median ppm per sample") + # Set axis labels
  ggtitle("Ang. sylvestris") + # Set title
  theme(text = element_text(size=8),axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(axis.text.x=element_blank(), plot.title = element_text(face = "italic"))

p4y <- ggplot(Anthriscus_sylvestris_median_ppms, aes( y=avg, x=Found_target_compound))  +  ylim(0, 12) +
 geom_point( stat="identity") + xlab("Target Compound") + ylab("Median ppm per sample") + # Set axis labels
  ggtitle("Ant. sylvestris") + # Set title
  theme(text = element_text(size=8),axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(axis.text.x=element_blank(), plot.title = element_text(face = "italic"))

p5y <- ggplot(c_majus_median_ppms, aes( y=avg, x=Found_target_compound))  + ylim(0, 12) +
 geom_point(stat="identity") + xlab("Target Compound") + ylab("Median ppm per sample") + # Set axis labels
  ggtitle("C. majus") + # Set title
  theme(text = element_text(size=8),axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(axis.text.x=element_blank(), plot.title = element_text(face = "italic"))

p6y <- ggplot(d_carota_median_ppms, aes( y=avg, x=Found_target_compound))  + ylim(0, 12) +
 geom_point( stat="identity") + xlab("Target Compound") + ylab("Median ppm per sample") + # Set axis labels
  ggtitle("D. carota") + # Set title
  theme(text = element_text(size=8),axis.text.x = element_text(angle = 90, hjust = 1)) +
  theme(axis.text.x=element_blank(), plot.title = element_text(face = "italic"))

multiplot(p1y, p2y, p3y, p4y, p5y, p6y,  cols=6)
```

Appendix 8 Top 150 peak MS report for *Angelica sylvestris*, limited to first 20 minutes of GCMS method, ordered by area under the peak.

RT	Height	Area	Name_1	CAS_No_1	Match1	Name_2	CAS_No_2	Match2
8.588	308335552	10791911.0	β-Myrcene	123-35-3	904	β-Myrcene	123-35-3	896
3.042	34332016	3142177.2	Heptane	142-82-5	908	Heptane	142-82-5	904
9.317	34883192	1467954.4	Bicyclo[3.1.0]hex-2-ene, 4-methyl-1-(1-methylethyl)-	28634-89-1	927	Bicyclo[3.1.0]hexane, 4-methylene-1-(1-methylethyl)-	3387-41-5	857
4.846	5504035	977379.7	2-Pentene, 4,4-dimethyl-, (Z)-	762-63-0	769	3-Hexen-2-one	763-93-9	737
7.542	17394892	778868.6	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	855	1R-α-Pinene	7785-70-8	854
3.421	4701248	660369.1	Cyclohexane, methyl-	108-87-2	783	Diethylcyanamide	617-83-4	775
17.104	25194546	643408.8	Propanoic acid, 2-methyl-, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester		872	2,2,4-Trimethyl-1,3-pentanediol diisobutyrate	6846-50-0	828
3.017	56182440	547938.4	Heptane	142-82-5	872	Heptane	142-82-5	865
13.183	17896196	458147.0	4-Hydroxy-2-methylacetophenone	875-59-2	845	4-Hydroxy-3-methylacetophenone	876-02-8	836
6.271	6168216	448270.8	p-Xylene	106-42-3	847	p-Xylene	106-42-3	846
4.554	4316444	391704.8	3-Hexanone	589-38-8	740	3-Hexanone	589-38-8	724
4.654	4089049	383010.8	2-Hexanone, 4-hydroxy-3-propyl-	62338-17-4	651	2-Hexanone	591-78-6	641
8.079	6975570	374331.5	Benzaldehyde	100-52-7	817	Benzaldehyde	100-52-7	807
5.312	4780042	363796.2	Heptane, 2,4-dimethyl-	2213-23-2	869	Hexane, 2,3,4-trimethyl-	921-47-1	853
9.579	10427734	304649.4	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	779	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	772
14.425	11866621	303390.9	Propanoic acid, 2-methyl-, 3-hydroxy-2,4,4-trimethylpentyl ester	74367-34-3	816	Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester	74367-31-0	797
5.233	4257425	289382.9	Cyclotrisiloxane, hexamethyl-	541-05-9	871	Cyclotrisiloxane, hexamethyl-	541-05-9	837
9.550	9801626	260140.2	Benzaldehyde, 2-hydroxy-	90-02-8	841	Benzaldehyde, 2-hydroxy-	90-02-8	832
15.600	8616859	202519.9	3-Isopropoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane		525	3-Butoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane	72439-84-0	517
3.600	1744205	197354.0	Hydroperoxide, hexyl	4312-76-9	590	1-Pentanol, 4-methyl-	626-89-1	589
10.463	7251467	183302.5	Decane	124-18-5	798	Hexane, 3,3-dimethyl-	563-16-6	797
17.533	7362122	168580.7	1,6-Cyclodecadiene, 1-methyl-5-methylene-8-(1-methylethyl)-, [s-(E,E)]-		468	Isoledene		468
13.421	6636166	157781.9	Bis(trimethylsilyl) 2,2-difluoro-1-(trifluoromethyl)ethenylphosphonate	110410-49-6	607	Cyclohexasiloxane, dodecamethyl-	540-97-6	595
16.108	6355993	147242.4	1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-, (Z,E)-	26560-14-5	781	1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-, (Z,E)-	26560-14-5	692
10.542	5698395	145630.6	Nonanal	124-19-6	781	Nonanal	124-19-6	767

11.121	5562106	135334.8	Lilac aldehyde A	53447-45-3	780	Lilac aldehyde B	53447-46-4	775
9.721	4849725	125949.5	Undecane, 4,7-dimethyl-	17301-32-5	815	Undecane, 5,7-dimethyl-	17312-83-3	812
8.479	3990652	114620.1	Cyclotetrasiloxane, octamethyl-	556-67-2	787	Cyclotetrasiloxane, octamethyl-	556-67-2	784
18.375	3475349	111912.4	2-Bromononane	2216-35-5	785	Tetradecane, 1-iodo-	19218-94-1	781
4.162	1586189	108508.2	Benzonitrile, m-phenethyl-	34176-91-5	652	Bicyclo[3.2.0]hepta-2,6-diene	2422-86-8	629
6.725	1819105	105623.0	o-Xylene	95-47-6	803	p-Xylene	106-42-3	790
13.458	3951030	102193.2	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	762	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	740
11.875	4175122	91309.7	1-Octene, 6-methyl-	13151-10-5	845	3-Octene, 2,6-dimethyl-	6874-28-8	817
11.258	3338428	87922.9	Lilac aldehyde A	53447-45-3	723	Lilac aldehyde B	53447-46-4	718
5.517	1051924	84946.4	4-Penten-2-ol	625-31-0	680	Formic acid, 1-methylethyl ester	625-55-8	678
6.088	1867842	84735.2	Ethylbenzene	100-41-4	765	Ethylbenzene	100-41-4	755
15.458	3631289	80461.1	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (E)-	18794-84-8	736	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (E)-	18794-84-8	701
19.454	3383634	80417.8	1-Iodo-2-methylnonane		753	Octane, 2,7-dimethyl-	1072-16-8	750
16.300	3837521	79245.2	Bicyclo[4.1.0]-3-heptene, 2-isopropenyl-5-isopropyl-7,7-dimethyl-		712	(+)-Epi-bicyclosesquiphellandrene	54324-03-7	710
16.200	2919172	77969.2	Bicyclo[3.1.1]heptane, 6-methyl-2-methylene-6-(4-methyl-3-pentenyl)-, [1R-(1α,5α,6β)]-		710	1,3-Cyclohexadiene, 5-(1,5-dimethyl-4-hexenyl)-2-methyl-, [S-(R,S)]-	495-60-3	690
5.083	1886549	76727.2	Cyclotrisiloxane, hexamethyl-	541-05-9	772	Trimethyl[4-(1,1,3,3,-tetramethylbutyl)phenoxy]silane	78721-87-6	760
15.917	2758232	75569.8	1,6-Cyclodecadiene, 1-methyl-5-methylene-8-(1-methylethyl)-, [s-(E,E)]-		633	Bicyclo[4.1.0]-3-heptene, 2-isopropenyl-5-isopropyl-7,7-dimethyl-		622
5.042	2113440	73415.1	Cyclotrisiloxane, hexamethyl-	541-05-9	744	Cyclotrisiloxane, hexamethyl-	541-05-9	720
10.988	2896710	72438.4	Cyclopentasiloxane, decamethyl-	541-02-6	720	Cyclopentasiloxane, decamethyl-	541-02-6	699
12.096	2904312	69934.8	Decanal	112-31-2	825	Decanal	112-31-2	769
14.671	2630980	54345.3	1-Decene, 8-methyl-	61142-79-8	792	3-Undecene, 9-methyl-, (Z)-	74630-50-5	779
8.371	1910201	52145.9	Bicyclo[3.1.0]hexane, 4-methylene-1-(1-methylethyl)-	3387-41-5	551	β-Myrcene	123-35-3	537
12.192	2097400	49797.5	Lilac alcohol B	33081-35-5	717	Lilac alcohol A	33081-34-4	715
8.283	1556181	49046.8	Ethyl (1-adamantylamino)carbothioylcarbamate	36997-89-4	485	Oxazolid-2-one, 3-acetyl-2',4-diphenyl-		420
17.238	2207982	45501.0	1-Iodoundecane	4282-44-4	758	Tetradecane, 1-iodo-	19218-94-1	754
6.917	713873	44408.0	3,4-Hexanedione, 2,2,5-trimethyl-	20633-03-8	784	Acetyl valeryl	96-04-8	778
8.779	1609387	41545.6	Decane, 6-ethyl-2-methyl-	62108-21-8	755	Octane, 2-bromo-	557-35-7	750

5.146	1637558	41087.2	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	792	Trimethyl[4-(1,1,3,3,-tetramethylbutyl)phenoxy]silane	78721-87-6	758
7.850	1383649	40685.4	Camphene	79-92-5	666	Camphene	79-92-5	656
13.546	1711873	40633.9	(S)-3,4-Dimethylpentanol		734	Hydroperoxide, hexyl	4312-76-9	728
19.692	1701614	39278.3	Amyl Nitrite	110-46-3	643	Isopropyl Myristate	110-27-0	628
13.688	1495646	38395.4	Furan, tetrahydro-2,5-dimethyl-	1003-38-9	800	2(3H)-Furanone, dihydro-5-methyl-	108-29-2	798
9.400	1613544	34428.0	3,5-Methanocyclopentapyrazole, 3,3a,4,5,6,6a-hexahydro-3a,4,4-trimethyl-		762	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	759
16.575	1554290	31042.6	Bicyclo[3.1.1]heptane, 6,6-dimethyl-2-methylene-, (1S)-	18172-67-3	704	1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-, (Z,E)-	26560-14-5	701
19.517	738315	28990.5	Nonane, 1-iodo-	4282-42-2	730	Tetradecane, 1-iodo-	19218-94-1	719
6.383	887232	28136.4	2-Propenoic acid, 2-methyl-, ethenyl ester	4245-37-8	513	3-Pentyn-2-ol	27301-54-8	503
9.813	1206178	27786.4	Pentane, 2,3,3-trimethyl-	560-21-4	779	Pentane, 2,3,3-trimethyl-	560-21-4	775
18.508	1194252	26622.4	Benzoic acid, 2-ethylhexyl ester	5444-75-7	709	1-Butanol, 3-methyl-, benzoate	94-46-2	682
19.208	1234427	26402.8	Butane, 2,3-dimethoxy-2-methyl-	74421-00-4	543	3-Methoxy-3-methylbutanol	56539-66-3	495
13.075	1225653	25982.3	Nonane, 1-iodo-	4282-42-2	742	Tetradecane, 1-iodo-	19218-94-1	701
10.696	541479	25914.5	Hexane, 2,2,3,3-tetramethyl-	13475-81-5	728	Octane, 3,4,5,6-tetramethyl-	62185-21-1	727
3.721	692564	25668.8	2-Propenoic acid, 2-methyl-, 1-methylbutyl ester	94159-12-3	514	1,3-Butadien-1-ol, acetate	1515-76-0	453
15.846	1238115	25406.0	Benzene, 1-(1,5-dimethyl-4-hexenyl)-4-methyl-	644-30-4	657	Benzene, 1-(1,5-dimethyl-4-hexenyl)-4-methyl-	644-30-4	651
6.175	999269	24982.2	Hexane, 2,3,4-trimethyl-	921-47-1	856	Hexane, 2,3,4-trimethyl-	921-47-1	827
19.883	815984	24415.1	1-Iodo-2-methylnonane		695	2-Bromononane	2216-35-5	686
12.463	921827	23668.0	Benzothiazole	95-16-9	680	Benzothiazole	95-16-9	677
14.233	1053255	23147.9	Butane, 1-propoxy-	3073-92-5	673	Propylene Carbonate	108-32-7	660
13.725	791374	22958.8	Heptane, 2,5,5-trimethyl-	1189-99-7	681	Octane, 2,6,6-trimethyl-	54166-32-4	659
12.929	1063226	22233.3	Ethanone, 1-(4-ethylphenyl)-	937-30-4	707	Ethanone, 1-(4-ethylphenyl)-	937-30-4	707
16.038	926780	21589.2	1-Iodo-2-methylnonane		732	Nonane, 3,7-dimethyl-	17302-32-8	719
18.846	701726	20560.1	Nonane, 1-iodo-	4282-42-2	704	Hexane, 3,3-dimethyl-	563-16-6	688
10.325	517827	20282.1	1,6-Heptadien-4-ol, 4-propyl-	52939-61-4	653	Cyclopropanemethanol, α-butyl-	4379-16-2	642
13.238	759837	19347.3	4-Ethylbenzoic acid, tridec-2-ynyl ester		781	1-Propanone, 2-chloro-1-(2,4-dimethylphenyl)-2-methyl-	54965-53-6	765
14.908	812262	18781.5	2-Hexen-1-ol, (E)-	928-95-0	691	2-Hexen-1-ol, (E)-	928-95-0	681
17.750	376958	18434.4	Decane, 1-chloro-	1002-69-3	616	Dodecane, 1-chloro-	112-52-7	614

8.888	672450	18145.2	α-Phellandrene	99-83-2	730	1,4-Methano-1H-Cyclopropa[d]pyridazine, 4,4a,5,5a-tetrahydro-6,6-dimethyl-, (1α,4α,4aα,5aα)-		720
10.913	955728	17600.7	1,3-Cyclohexadiene, 1,5,5,6-tetramethyl-	514-94-3	749	Cyclopentene, 1,2,3,3-tetramethyl-4-methylene-		733
14.775	931113	16897.2	Tetradecane, 1-iodo-	19218-94-1	726	3-Hexanone, 2,4-dimethyl-	18641-70-8	717
4.321	730150	16059.8	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		818	5-Methyl-2-trimethylsilyloxy-acetophenone	97389-69-0	616
8.842	589954	15493.3	1-Pentanol, 3-methyl-	589-35-5	670	1-Pentene, 4-methyl-	691-37-2	657
12.004	707196	14109.4	Butane, 2,2-dimethyl-	75-83-2	749	3-Hexanone, 2,2-dimethyl-	5405-79-8	740
16.938	421463	12601.3	2,4-Dinitrophenyl crotonate	69817-88-5	721	2-Propenoic acid, 2-methyl-, octyl ester	2157-01-9	707
19.058	523344	12364.1	Acetic acid, trifluoro-, 2,2-dimethylpropyl ester	7556-79-8	632	1-Nonene, 4,6,8-trimethyl-	54410-98-9	582
17.954	404616	12122.2	1-Nonene, 4,6,8-trimethyl-	54410-98-9	696	Nitric acid, nonyl ester	20633-13-0	693
17.396	452322	11935.7	1-Undecyn-4-ol	22127-86-2	701	1-Dodecyn-4-ol	74646-36-9	698
12.292	423327	11754.7	Adamantane, 1,3-dimethyl-	702-79-4	668	Tricyclo[4.3.1.1(3,8)]undecane, 1-chloro-	27011-46-7	654
18.288	395203	11570.7	1-Octanol, 2-nitro-	2882-67-9	655	1-Undecyn-4-ol	22127-86-2	651
15.221	491276	11412.2	1-Butanone, 2-chloro-3-methyl-1-[4-(1-methylethyl)phenyl]-	55955-90-3	555	Benzene, (1-butylhexyl)-	4537-11-5	534
11.479	377020	11405.4	Lilac aldehyde C	53447-47-5	519	Lilac aldehyde A	53447-45-3	515
7.963	484489	11373.4	4-Pentyn-2-ol	2117-11-5	672	2-Pentanol, 3-chloro-4-methyl-, (R,*S*)-(ñ)-	74685-48-6	659
4.254	673571	11366.3	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		837	Cyclotrisiloxane, hexamethyl-	541-05-9	680
18.671	438839	11346.7	Benzaldehyde, 3-benzyloxy-2-fluoro-4-methoxy-		558	N,O-Dicarbobenzyloxy-L-tyrosine	57228-29-2	558
4.338	606470	11208.7	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	671	Cyclotrisiloxane, hexamethyl-	541-05-9	665
10.850	453282	11167.1	2-(3-Methylbuta-1,3-dienyl)cyclohexanone		682	1-Methyl-3-ethyladamantane	1687-34-9	681
19.104	334717	11044.5	Propane, 1-chloro-2,2-dimethyl-	753-89-9	483	Propane, 1-chloro-2,2-dimethyl-	753-89-9	477
19.804	290370	10706.0	1,10-Dichlorodecane	2162-98-3	529	1,10-Dichlorodecane	2162-98-3	522
18.733	430714	10567.9	Benzofenac methyl ester	63007-66-9	577	Benzene, 1-methyl-4-[(phenylmethyl)sulfonyl]-	5395-20-0	557
7.779	390156	10230.0	DL-3,4-Dimethyl-3,4-hexanediol	32388-94-6	563	Acetic acid, 1,1-dimethylethyl ester	540-88-5	554
16.446	361319	9695.6	4-Heptanone, 3-methyl-	15726-15-5	634	3-Hexanone, 2,5-dimethyl-	1888-57-9	631
4.279	562958	9652.6	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		886	5-Methyl-2-trimethylsilyloxy-acetophenone	97389-69-0	800
3.900	371992	9537.7	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		810	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	687
19.729	399303	9487.8	Benzofenac methyl ester	63007-66-9	646	Benzaldehyde, 3-benzyloxy-2-fluoro-4-methoxy-		639

15.733	338455	9355.4	1-Hexene, 3,3,5-trimethyl-	13427-43-5	725	Isooctanol	26952-21-6	721
16.388	236284	9319.7	2-Propenoic acid, 2-methyl-, ethenyl ester	4245-37-8	700	Vinyl crotonate	14861-06-4	677
15.817	421474	9101.9	Copaene	3856-25-5	643	Bicyclo[4.1.0]-3-heptene, 2-isopropenyl-5-isopropyl-7,7-dimethyl-		637
18.996	269681	8726.7	Pentane, 2,3,4-trimethyl-	565-75-3	618	Pentane, 2,3,4-trimethyl-	565-75-3	570
16.804	183924	8432.0	Pentane, 2-bromo-4-methyl-	30310-22-6	551	Pentane, 2,3,3,4-tetramethyl-	16747-38-9	537
7.163	262704	7908.1	Vinyl crotonate	14861-06-4	632	2,4-Dinitrophenyl crotonate	69817-88-5	631
10.008	387284	7678.9	1-Pentanol, 2-ethyl-	27522-11-8	723	Isooctanol	26952-21-6	720
17.200	501834	7380.1	Tricyclo[3.1.0.0(2,4)]hexane, 3,6-diethyl-3,6-dimethyl-, trans-		650	1,3-Cyclopentadiene, 1,3-bis(1-methylethyl)-	123278-27-3	631
9.213	300857	7164.5	2-Ethyl-6-phenyl-1,3,4-thiadiazolo(3,2-a)(1,3,5)-triazine-5,7-dione	69378-04-7	738	1-Pentanone, 1-(4-methylphenyl)-	1671-77-8	716
15.117	404439	7124.2	Bicyclo[4.1.0]-3-heptene, 2-isopropenyl-5-isopropyl-7,7-dimethyl-		618	α-Cubebene	17699-14-8	611
3.783	475251	6872.3	Urea, N-[5-(ethylsulfonyl)-1,3,4-thiadiazol-2-yl]-N,N'-dimethyl-	30043-49-3	488	Acetic acid, [4-(1,1-dimethylethyl)phenoxy]-, methyl ester	88530-52-3	475
19.642	220019	6853.5	1-Dodecyn-4-ol	74646-36-9	583	Methyl 2-hydroxydodecanoate	51067-85-7	554
11.842	311802	6828.1	2,4,5-Trihydroxypyrimidine	496-76-4	699	1-[2-Pyridyl]-2,2-dimethyl-2-morpholino ethanol		697
18.046	252749	6677.2	3-Bromooctane	999-64-4	610	1-Hexene, 3,5,5-trimethyl-	4316-65-8	599
13.633	237095	6555.3	S-(Neopentyloxythiocarbonyl)thiohydroxylamine	35659-84-8	532	Nitric acid, heptyl ester	20633-12-9	489
5.679	257783	6475.8	Cyclobutanol	2919-23-5	437	Cyclopropyl carbinol	2516-33-8	409
11.417	206046	6396.7	2,4,6-Trimethyl-1-nonene	55771-40-9	669	1-Hexene, 3,3,5-trimethyl-	13427-43-5	641
19.146	261875	6256.1	1-Penten-3-ol	616-25-1	653	Pentane, 1,3-epoxy-4-methyl-	15045-60-0	621
18.875	357699	6210.4	Hexane, 3-chloro-3-methyl-	43197-78-0	400	2-Acetyl-6-methyl-6,7-dihydro-9H-5-oxa-9-azabenzocyclohepten-8-one		343
4.358	513163	6173.1	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		747	Cyclotrisiloxane, hexamethyl-	541-05-9	685
10.079	219497	5900.6	Heptane, 3-bromo-	1974-05-6	531	Heptane, 3-(chloromethyl)-	123-04-6	529
15.142	277522	5899.8	Trifluoromethanesulfonic anhydride	358-23-6	593	Vinyl crotonate	14861-06-4	580
16.892	155732	5724.8	Hexane, 2,2,5-trimethyl-	3522-94-9	536	Heptane, 2,2-dimethyl-	1071-26-7	520
14.963	232333	5646.6	4-Octene, 2,3,7-trimethyl-, [S-(E)]-	52763-13-0	705	4-Nonene, 3-methyl-, (Z)-	63830-69-3	694
11.313	229597	5615.7	Heptafluorobutyric acid, 2-methylpropyl ester	49669-71-8	677	Trifluoroacetic acid, 2-methylpropyl ester	17355-83-8	652
10.254	174464	5454.2	1-Propanesulfonyl chloride	10147-36-1	580	Propane, 2-nitro-	79-46-9	571
14.829	227707	5222.9	4-Octene, 2,3,6,7-tetramethyl-	63830-66-0	629	1-Dodecyn-4-ol	74646-36-9	626

13.325	215245	5143.7	Silane, trimethylphenyl-	768-32-1	646	o-Methoxybenzoic acid, 5-fluoro-2-nitrophenyl ester		638
15.263	215032	5133.7	5,6-Decadien-3-yne, 5,7-diethyl-	61227-89-2	639	2-Oxo-3-[3-methylbenzoyl]propanoic acid, ethyl ester	5814-37-9	626
13.954	220295	5027.6	2-Oxetanone, 4-methyl-, (ñ)-	36536-46-6	647	Acetic acid, hydrazide	1068-57-1	641
4.396	280007	4906.7	Cyclotrisiloxane, hexamethyl-	541-05-9	697	Cyclotrisiloxane, hexamethyl-	541-05-9	605
19.392	207200	4887.6	Pentanal	110-62-3	544	Pentanal	110-62-3	537
8.217	187770	4711.3	1-Pentene, 4,4-dimethyl-1,3-diphenyl-1-(trimethylsilyloxy)-		613	1-Heptene, 1,3-diphenyl-1-(trimethylsilyloxy)-		612
9.917	226317	4606.6	Ethanone, 2-(formyloxy)-1-phenyl-	55153-12-3	821	6-Benzamido-4-benzoyl-1,2,4-triazine-3,5(2H,4H)-dione		814
15.392	226891	4588.0	4-Hepten-2-one, (E)-	36678-43-0	661	Ethanone, 1-cyclopropyl-	765-43-5	652
19.942	221377	4502.0	1,9-Dichlorononane	821-99-8	504	Phenylacetic acid, 3-methylbut-2-enyl ester	151091-17-7	502
13.850	238646	4373.2	Butane, 2,2-dimethyl-	75-83-2	600	Butane, 2,2-dimethyl-	75-83-2	583
12.663	169043	4287.7	Acetic acid, 4-acetoxy-tricyclo[4.3.1.0(3,8)]dec-10-yl ester		412	1-(2,4-Dimethylphenyl)ethanol	99500-87-5	408
19.296	181389	4265.5	1-Bromo-7-(tetrahydro-2-pyranyloxy)heptane	10160-25-5	376	Furan, tetrahydro-2,4-dimethyl-, trans-	39168-02-0	339
3.921	209027	4152.5	Trimethyl[4-(1,1,3,3,-tetramethylbutyl)phenoxy]silane	78721-87-6	697	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		688
14.579	214823	4063.8	Phenol, 2-(5-isoxazolyl)-	61348-47-8	647	4-Butylbenzoic acid, tridec-2-ynyl ester		551
5.725	213119	4056.6	Phosphine, tris(trifluoromethyl)-	432-04-2	610	Trifluoromethanesulfonic anhydride	358-23-6	579

Appendix 9 Top 150 peak MS report for *Anthriscus sylvestris*, limited to first 20 minutes of GCMS method, ordered by area under the peak.

RT	Height	Area	Name_1	CAS_No_1	Match1	Name_2	CAS_No_2	Match2
8.590	106502520	4146044.5	β-Myrcene	123-35-3	877	β-Myrcene	123-35-3	869
3.021	25055918	3076596.5	Heptane	142-82-5	842	Heptane	142-82-5	841
9.306	68286344	2477055.2	Limonene	138-86-3	887	Cyclohexene, 1-methyl-4-(1-methylethenyl)-, (S)-	5989-54-8	843
4.805	9548238	2228224.5	3-Penten-2-one, 4-methyl-	141-79-7	741	3-Penten-2-one, 4-methyl-	141-79-7	740
13.456	62354716	1678002.8	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	631	Cyclopropane, trimethyl(2-methyl-1-propenylidene)-	14803-30-6	628
3.389	4425264	814707.4	Diethylcyanamide	617-83-4	671	Diethylcyanamide	617-83-4	667
4.621	5274605	554528.1	Cyclohexane, 1,1'-(2-methyl-1,3-propanediyl)bis-	2883-08-1	450	2-Oxo-n-propyl-2-(veratrylidenehydrazino)acetamide	339241-37-1	447

9.595	16427211	514489.9	3-Carene	13466-78-9	830	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	828
5.207	4555402	454539.1	Cyclotrisiloxane, hexamethyl-	541-05-9	834	Cyclotrisiloxane, hexamethyl-	541-05-9	811
4.528	4259968	432008.2	3-Hexanone	589-38-8	800	3-Hexanone	589-38-8	785
5.299	4470925	407708.2	Heptane, 2,4-dimethyl-	2213-23-2	842	Hexane, 2,3,5-trimethyl-	1069-53-0	832
13.305	17134146	403353.0	Isobornyl acetate	125-12-2	807	Acetic acid, 1,7,7-trimethyl-bicyclo[2.2.1]hept-2-yl ester	92618-89-8	806
8.075	5848204	360814.5	Benzaldehyde	100-52-7	793	Benzaldehyde	100-52-7	783
5.504	3156976	293924.4	Formic acid, 1-methylethyl ester	625-55-8	722	Butanal, 3-hydroxy-	107-89-1	722
10.567	10051937	275774.2	Nonanal	124-19-6	830	Nonanal	124-19-6	821
17.175	5329910	208692.3	4,4-Dimethyl-1-hexene	1647-08-1	755	Pentanoic acid, 2,2,4-trimethyl-3-carboxyisopropyl, isobutyl ester		737
6.936	1645726	203808.6	2,4,6-Trimethyl-1-nonene	55771-40-9	784	Pentane, 2,3,4-trimethyl-	565-75-3	713
12.133	8039638	191582.1	Decanal	112-31-2	891	Decanal	112-31-2	846
15.655	7355326	190413.8	3-Isopropoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane		630	1,1,1,3,5,5,7,7,7-Nonamethyl-3-(trimethylsiloxy)tetrasiloxane	38146-99-5	619
6.262	2150687	189994.7	p-Xylene	106-42-3	771	p-Xylene	106-42-3	769
11.094	6812706	188223.6	Benzyl nitrile	140-29-4	865	Benzyl nitrile	140-29-4	857
18.443	5363376	175275.3	Hexadecane	544-76-3	776	Dodecane, 1-iodo-	4292-19-7	773
8.486	3203332	155479.1	Cyclotetrasiloxane, octamethyl-	556-67-2	833	Cyclotetrasiloxane, octamethyl-	556-67-2	817
19.528	6355568	154956.4	Decane, 3,8-dimethyl-	17312-55-9	783	Hexadecane	544-76-3	781
9.407	5462084	138449.0	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	815	3,5-Methanocyclopentapyrazole, 3,3a,4,5,6,6a-hexahydro-3a,4,4-trimethyl-		802
11.015	5423816	136784.3	Cyclopentasiloxane, decamethyl-	541-02-6	715	Cyclopentasiloxane, decamethyl-	541-02-6	676
9.733	4510782	120763.5	Octane, 2,4,6-trimethyl-	62016-37-9	836	Decane, 2,4-dimethyl-	2801-84-5	831
17.304	4916890	111963.5	Octane, 2,7-dimethyl-	1072-16-8	790	Dodecane, 3-methyl-	17312-57-1	783
13.498	4351969	106694.7	Cyclobutanone, 2,3-dimethyl-, trans-	1942-42-3	537	Formic acid, butyl ester	592-84-7	532
8.301	1673056	106565.9	1-Heptene, 1,3-diphenyl-1-(trimethylsilyloxy)-		678	Cyclotetrasiloxane, octamethyl-	556-67-2	659
10.939	4783878	104654.7	Bicyclo[3.1.0]hex-2-ene, 4,4,6,6-tetramethyl-	19487-09-3	731	2,4,6-Octatriene, 3,4-dimethyl-	57396-75-5	729
5.655	1307803	99583.7	Pyridine, 2-methyl-	109-06-8	707	Aniline	62-53-3	703
17.602	4148196	91692.7	Cyclooctasiloxane, hexadecamethyl-	556-68-3	647	Silane, [[4-[1,2-bis[(trimethylsilyl)oxy]ethyl]-1,2-phenylene]bis(oxy)]bis[trimethyl-		640
8.377	1229439	87204.7	1-Pentene, 4,4-dimethyl-1,3-diphenyl-1-(trimethylsilyloxy)-		667	Cyclotetrasiloxane, octamethyl-	556-67-2	661
11.907	3784420	85190.1	1-Octene, 3,7-dimethyl-	4984-01-4	837	1-Octene, 3,7-dimethyl-	4984-01-4	836

16.098	3632136	84832.6	1-Iodo-2-methylnonane		820	Dodecane, 2,6,10-trimethyl-	3891-98-3	814
9.809	2330566	83375.3	α-Phellandrene	99-83-2	738	1,4-Cyclohexadiene, 1-methyl-4-(1-methylethyl)-	99-85-4	735
11.283	3051802	80820.2	Lilac aldehyde A	53447-45-3	755	Lilac aldehyde B	53447-46-4	754
18.582	3233737	80144.2	Benzoic acid, 2-ethylhexyl ester	5444-75-7	771	2-Octyl benzoate	6938-51-8	699
10.479	2911336	80134.1	Decane, 2,5,9-trimethyl-	62108-22-9	807	Decane, 2,4-dimethyl-	2801-84-5	790
8.854	2097037	72420.8	Octanal	124-13-0	767	Octanal	124-13-0	761
6.166	1021782	66882.6	Octane, 4-methyl-	2216-34-4	786	Hexane, 2,3,4-trimethyl-	921-47-1	772
3.624	1642837	66725.9	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		747	Cyclotrisiloxane, hexamethyl-	541-05-9	624
7.543	1510037	64257.9	Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-	2867-05-2	759	Bicyclo[3.1.0]hexane, 4-methyl-1-(1-methylethyl)-, didehydro deriv.	58037-87-9	758
4.160	1077523	59075.0	Benzonitrile, m-phenethyl-	34176-91-5	569	Benzimidazole, 2-benzylsulfonyl-	100872-42-2	545
12.970	2734788	58404.8	m-Ethylacetophenone	22699-70-3	802	Ethanone, 1-(4-ethylphenyl)-	937-30-4	800
3.678	1018461	53497.3	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		592	Silane, 1,4-phenylenebis[trimethyl-	13183-70-5	573
19.285	1697329	53433.2	Ethanedioic acid, bis(trimethylsilyl) ester	18294-04-7	517	Mercaptoacetic acid, bis(trimethylsilyl)-	6398-62-5	513
11.153	2188783	50364.3	Lilac aldehyde C	53447-47-5	713	Lilac aldehyde A	53447-45-3	706
17.840	1125030	49913.2	Decane, 2,5,9-trimethyl-	62108-22-9	718	Heptane, 2,3,5-trimethyl-	20278-85-7	700
19.599	1369462	49893.4	Hexane, 3,3-dimethyl-	563-16-6	749	Nonane, 3,7-dimethyl-	17302-32-8	744
9.553	2212252	48390.0	Benzeneacetaldehyde	122-78-1	849	Benzeneacetaldehyde	122-78-1	849
14.721	2216653	47883.3	1-Heptanol, 6-methyl-	1653-40-3	813	Isooctanol	26952-21-6	811
11.216	1897278	47102.4	2,6,6-Trimethyl-2-cyclohexene-1,4-dione	1125-21-9	722	2,6,6-Trimethyl-2-cyclohexene-1,4-dione	1125-21-9	709
14.474	1719258	46837.4	3-Hexanone, 2,5-dimethyl-4-nitro-	59906-54-6	716	Pentanoic acid, 2,2,4-trimethyl-3-hydroxy-, isobutyl ester	244074-78-0	706
13.590	1868662	46536.5	(S)-3,4-Dimethylpentanol		742	Hydroperoxide, hexyl	4312-76-9	724
6.048	719388	45675.3	Carbamic acid, N-(3-oxo-4-isoxazolidinyl)-, benzyl ester	28832-02-2	656	Sulfoxide, methyl phenethyl	7714-32-1	647
18.917	1104179	45515.4	1-Iodo-2-methylnonane		757	Eicosane	112-95-8	740
14.825	1937640	44305.4	Decane, 6-ethyl-2-methyl-	62108-21-8	774	Octane, 3-ethyl-2,7-dimethyl-	62183-55-5	771
11.597	1675741	41666.0	1,4-Cyclohexanedione, 2,2,6-trimethyl-	20547-99-3	764	1-Decene, 2-methyl-	13151-27-4	681
19.771	1431870	40522.0	Valeric acid, 2-pentadecyl ester		647	Valeric acid, 4-tridecyl ester		626
12.497	1423124	39557.2	Benzothiazole	95-16-9	744	Benzothiazole	95-16-9	743
14.951	1518394	36436.1	2-Decen-1-ol	22104-80-9	778	2-Decen-1-ol, (E)-	18409-18-2	755
19.189	910965	34313.4	1,3-Dioxolan-4-one, 2-(1,1-dimethylethyl)-5-(phenylmethyl)-5-(2-propenyl)-, (2S-cis)-		525	Propanehydrazide, 2-benzyl-2-hydroxy-3-phenyl-N2-ethylideno-		491

8.791	1132364	32735.5	Hexadecane	544-76-3	774	Octane, 2,7-dimethyl-	1072-16-8	766
17.464	1334637	32311.9	Octadecane, 1-(ethenyloxy)-	930-02-9	724	1-Nonadecanol	1454-84-8	714
9.214	1060484	28934.1	Benzene, 1-methyl-2-(1-methylethyl)-	527-84-4	791	Benzene, 1-methyl-4-(1-methylethyl)-	99-87-6	772
18.025	1057153	27820.7	Tetradecane, 1-iodo-	19218-94-1	739	Heptadecane, 2,6-dimethyl-	54105-67-8	725
6.702	574216	27693.4	Ethylbenzene	100-41-4	685	Ethylbenzene	100-41-4	684
11.509	1007332	27206.1	1,6,10-Dodecatrien-3-ol, 3,7,11-trimethyl-, [S-(Z)]-	142-50-7	686	Lilac aldehyde A	53447-45-3	649
13.117	1245924	25755.4	Decane, 2,9-dimethyl-	1002-17-1	724	Decane, 2,4-dimethyl-	2801-84-5	720
10.726	413165	22156.9	Hexane, 4-ethyl-2,2-dimethyl-	52896-99-8	713	Heptane, 2,3-dimethyl-	3074-71-3	711
18.808	799127	21541.4	Benzofenac methyl ester	63007-66-9	569	2,4-Difluorobenzene, 1-benzyloxy-	152434-86-1	562
18.624	879548	21113.1	Hexadecanal	629-80-1	588	Cyclopentanol	96-41-3	564
13.733	876679	20726.8	Butane, 1-chloro-	109-69-3	777	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	763
18.360	795116	20393.3	Methyl 2-hydroxydodecanoate	51067-85-7	678	4-Tetradecanol	1653-33-4	669
15.868	777442	20234.6	2,5-Pyrrolidinedione, 3-ethyl-1,3-dimethyl-	13861-99-9	652	3,4-Dihydroxy-5-aminopyridazine	4655-95-2	625
19.134	816460	20173.8	Decane, 6-ethyl-2-methyl-	62108-21-8	727	2-Bromononane	2216-35-5	722
7.405	436421	18492.0	Cyclopropane, 1-propenyl-	4663-21-2	679	2H-Pyran, 2,5-diethenyltetrahydro-	25724-33-8	592
18.753	855808	18475.7	Heneicosane, 11-phenyl-	6703-80-6	568	1,10-Dichlorodecane	2162-98-3	564
3.754	834427	18333.3	Cyclotrisiloxane, hexamethyl-	541-05-9	535	1,4-Benzenediol, 2,6-bis(1,1-dimethylethyl)-	2444-28-2	439
15.512	625019	17262.4	Trifluoroacetyl-lavandulol	28673-24-7	687	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (E)-	18794-84-8	660
15.793	642962	16638.3	4-Trifluoroacetoxyoctane	116465-17-9	730	Isooctanol	26952-21-6	706
11.869	688284	16510.5	1-[2-Pyridyl]-2,2-dimethyl-2-morpholino ethanol		731	Etidocaine	36637-18-0	722
16.195	691488	15620.6	Phenol, 2,4-bis(1,1-dimethylethyl)-	96-76-4	636	Phenol, 3,5-bis(1,1-dimethylethyl)-	1138-52-9	631
16.869	604580	15610.1	Decane, 6-ethyl-2-methyl-	62108-21-8	736	Undecane, 5-ethyl-	17453-94-0	726
19.218	787184	15600.1	Decane, 3-methyl-	13151-34-3	654	Heptane, 4-ethyl-2,2,6,6-tetramethyl-	62108-31-0	646
12.041	713533	15196.0	Decane, 2,5,9-trimethyl-	62108-22-9	714	1H-Tetrazol-5-amine	4418-61-5	710
16.249	653064	14954.0	1-Hexene, 4,4-diethyl-		694	Acetic acid, trifluoro-, 2,2-dimethylpropyl ester	7556-79-8	652
10.047	460948	14754.3	Tetrazolo[1,5-b]1,2,4-triazine, 5,6,7,8-tetrahydro-6,7-dimethyl-		661	Hexane, 1-fluoro-	373-14-8	643
18.201	653838	14334.1	4-Hepten-3-one, 5-methyl-	1447-26-3	637	Cyclohexane, 1-bromo-4-methyl-	6294-40-2	626
18.988	461819	14047.8	Pentane, 3-ethyl-2,4-dimethyl-	1068-87-7	615	Pentane, 2,3,3,4-tetramethyl-	16747-38-9	613
10.357	491592	13780.4	1-Nonene, 4,6,8-trimethyl-	54410-98-9	678	Octane, 3-ethyl-2,7-dimethyl-	62183-55-5	676
19.453	512547	13459.6	Heptane, 2,5-dimethyl-	2216-30-0	622	Hexane, 2,5-dimethyl-	592-13-2	621
19.871	581807	13240.2	Decane, 1-chloro-	1002-69-3	682	Dodecane, 1-chloro-	112-52-7	671

19.800	468370	13209.8	β-l-Rhamnopyranoside, phenyl-2,3-O-ethylboranediyl-4-O-benzyl-		564	3-Benzoyl-8-oxo-6-azabicyclo[3.2.1]octan-6,7-dicarboxylic acid, dibenzyl ester		519
3.846	477138	13022.1	Cyclotrisiloxane, hexamethyl-	541-05-9	626	Cyclotrisiloxane, hexamethyl-	541-05-9	582
18.113	521601	12552.6	Undecane, 3-methyl-	1002-43-3	649	Heptane, 2,2,4-trimethyl-	14720-74-2	645
4.311	625537	11735.7	Cyclotrisiloxane, hexamethyl-	541-05-9	688	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	661
13.770	466127	11635.8	1H-Tetrazol-5-amine	4418-61-5	656	3-Hexanone, 2,5-dimethyl-	1888-57-9	642
12.242	553286	11483.9	1-Heptene, 6-methyl-	5026-76-6	761	Cyclopropane, 1,1-dimethyl-2-pentyl-	62167-97-9	751
13.670	512222	11374.7	Phthalic acid, monoethyl ester	2306-33-4	601	Phthalamic acid	88-97-1	592
9.013	249062	11188.6	1-Pentanol, 2-ethyl-4-methyl-	106-67-2	579	Borinic acid, diethyl-	4426-31-7	572
16.513	430528	11069.7	Octane, 3,4,5,6-tetramethyl-	62185-21-1	679	Hexadecane	544-76-3	667
17.681	471369	11055.5	Sulfoxide, methyl phenethyl	7714-32-1	587	Benzofenac methyl ester	63007-66-9	587
16.174	472134	11046.4	Propanoic acid, 2,2-dimethyl-, 2,6-bis(1,1-dimethylethyl)-4-methylphenyl ester		438	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl-, methylcarbamate	1918-11-2	436
12.702	409712	10934.8	Benzenepropanol, α-methyl-, acetate	10415-88-0	599	1-Pent-3-ynylcyclopenta-1,3-diene		576
10.269	452964	10931.7	Benzenemethanol, α,α-dimethyl-	617-94-7	687	Benzenemethanol, α,α-dimethyl-	617-94-7	662
19.063	393783	10803.3	1H-Tetrazol-5-amine	4418-61-5	575	Dodecane, 4-methyl-	6117-97-1	538
13.238	503777	10394.4	(-)-trans-Pinocarvyl acetate		651	(-)-Myrtenyl acetate		610
14.281	327042	10233.0	3-Buten-2-ol	598-32-3	628	2-Butanone, 1-(acetyloxy)-	1575-57-1	596
12.326	376432	9938.5	1-Methyl-3-ethyladamantane	1687-34-9	727	Tricyclo[4.3.1.1(3,8)]undecane, 1-chloro-	27011-46-7	674
7.158	255288	9790.3	Sulfurous acid, dipropyl ester	623-98-3	547	Propanoic acid, 2-methyl-	79-31-2	546
19.721	409517	9649.7	Octadecane, 1-(ethenyloxy)-	930-02-9	633	Decane, 1-(ethenyloxy)-	765-05-9	626
15.596	416876	9528.4	Butane, 1-chloro-3,3-dimethyl-	2855-08-5	587	3-Hexanone, 2,4-dimethyl-	18641-70-8	576
11.681	386306	8746.5	7-Octene-1,2-diol	85866-02-0	500	5-Hepten-2-ol, 6-methyl-	1569-60-4	487
19.901	444972	8590.7	Cyclopentane, 1,1'-ethylidenebis-	4413-21-2	658	1,1'-Bicycloheptyl	23183-11-1	632
10.098	230456	8431.0	Glycidol	556-52-5	555	Cyclopropyl carbinol	2516-33-8	555
10.872	388051	8355.0	1,2-Benzenedicarboxylic acid, dipentyl ester	131-18-0	617	1,2-Benzenedicarboxylic acid, diheptyl ester	3648-21-3	607
14.662	285207	8190.3	Thiolane-3,3,4,4-tetracarbonitrile, 2,5-di-tert-butyl-		643	Valeraldehyde, 4,4-dimethyl-2-methylene-	5375-28-0	627
4.361	672070	8155.2	Cyclotrisiloxane, hexamethyl-	541-05-9	793	Cyclotrisiloxane, hexamethyl-	541-05-9	726
15.449	277504	8069.0	4-Hepten-2-one, (E)-	36678-43-0	631	4-Penten-2-one, 4-methyl-	3744-02-3	611
17.895	273720	7888.7	Glycine, N-(N-L-alanylglycyl)-	3146-40-5	506	Piperazine	110-85-0	485
13.900	209441	7804.8	2-Hexen-1-ol, (E)-	928-95-0	576	2-Penten-1-ol, (E)-	1576-96-1	575
4.206	550369	7745.8	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	690	Cyclotrisiloxane, hexamethyl-	541-05-9	686

15.943	380373	7583.9	Heptane, 2,2,3,3,5,6,6-heptamethyl-	7225-67-4	756	Hexadecane	544-76-3	732
9.692	303758	7487.2	Butane, 1-chloro-3,3-dimethyl-	2855-08-5	664	1-Bromo-3,3-dimethylbutane	1647-23-0	634
17.962	250458	7338.0	Decane, 2,3,7-trimethyl-	62238-13-5	595	Heptane, 2,2,3,3,5,6,6-heptamethyl-	7225-67-4	593
18.301	240438	7269.0	4-Cyclopropylcarbonyloxytridecane		520	1-Pentanol, 4-methyl-2-propyl-	54004-41-0	514
19.637	396680	7083.9	2-Acetyl-1-pyrroline	85213-22-5	492	2,4,6,8-Tetraazabicyclo[3.3.0]octan-3-one, 7-nitroimino-	223690-54-8	464
14.884	289131	6913.2	Acetic acid, trifluoro-, 2,2-dimethylpropyl ester	7556-79-8	578	(2S,3S)-(-)-3-Propyloxiranemethanol	89321-71-1	564
18.691	227280	6706.7	Vinyl lauryl ether	765-14-0	496	Trichloroacetic acid, tetradecyl ester	74339-52-9	485
3.934	387296	6268.2	1,1,1,3,3,3-Hexafluoro-2-[(1-isopropoxycarbonyl)ethoxyimino] propane	74387-56-7	465	2,4-Dimethoxycinnamic acid	6972-61-8	428
11.350	226835	6041.0	Butane, 1-chloro-3,3-dimethyl-	2855-08-5	589	1-Hexene, 3,5,5-trimethyl-	4316-65-8	585
16.957	272229	5916.1	Undecane, 6-methyl-	17302-33-9	690	Dodecane, 6-methyl-	6044-71-9	686
17.384	238897	5789.1	Acetic acid, trifluoro-, 2,2-dimethylpropyl ester	7556-79-8	578	3-Hexanone, 2,2-dimethyl-	5405-79-8	568
4.273	276500	5713.9	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		781	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	670
16.794	173486	5643.0	Nonadecane, 1-chloro-	62016-76-6	561	Undecane, 5-cyclohexyl-	13151-80-9	545
15.273	264495	5566.1	Ethanone, 1-[4-(1-methylethyl)phenyl]-	645-13-6	591	Ethanone, 1,1'-(1,4-phenylene)bis-	1009-61-6	588
9.989	225663	5565.4	1-Pentanol, 2-ethyl-4-methyl-	106-67-2	572	1-Hexene, 5,5-dimethyl-	7116-86-1	570
16.722	241492	5353.2	3,4-Hexanedione, 2,2,5-trimethyl-	20633-03-8	620	3-Hexanone	589-38-8	602
15.018	220366	5203.6	Acetic acid, trifluoro-, 3,7-dimethyloctyl ester	28745-07-5	624	2,6-Dimethyl-6-trifluoroacetoxyoctane	61986-67-2	624
12.878	173574	5101.0	1H-Indene, 1-hexadecyl-2,3-dihydro-	55334-29-7	514	Benzenepropanamine, α-methyl-	22374-89-6	509
16.429	166218	5079.1	Ethanol, 2-(2-propynyloxy)-	3973-18-0	532	1-Dodecyn-4-ol	74646-36-9	530
4.235	263412	5047.0	1-Benzopyrylium, 2-phenyl-	14051-53-7	439	1-Butanol, 2,3-dimethyl-	19550-30-2	417
9.934	191846	5028.0	Benzoyl bromide	618-32-6	711	Ethanone, 2,2-dihydroxy-1-phenyl-	1075-06-5	710
4.043	439976	4909.7	Trimethyl(4-tert.-butylphenoxy)silane	25237-79-0	559	2,4-Cyclohexadien-1-one, 3,5-bis(1,1-dimethylethyl)-4-hydroxy-	54965-43-4	549
7.238	224143	4769.5	Tert.-butylaminoacrylonitryl	77376-84-2	311	Cyclopentanol	96-41-3	294
17.011	184587	4731.0	3-Penten-1-ol, 4-methyl-	763-89-3	467	2,2-Dimethylpropionic acid, cyclopentyl ester		467

Appendix 10 Top 150 peak MS report for *Conopodium majus*, limited to first 20 minutes of GCMS method, ordered by area under the peak.

RT	Height	Area	Name_1	CAS_No_1	Match1	Name_2	CAS_No_2	Match2
17.514	239202752	5905569.5	Apiol	523-80-8	863	Apiol	523-80-8	822
3.067	38768772	5147640.0	Heptane	142-82-5	870	Heptane	142-82-5	870
14.143	118884808	2867399.0	Undecane, 1-bromo-	693-67-4	897	Decane, 1-bromo-	112-29-8	896
6.279	21186530	2102202.8	p-Xylene	106-42-3	876	p-Xylene	106-42-3	871
4.888	7480940	1853625.4	Diethylcyanamide	617-83-4	568	Cyclopropane, 1,1,2,2-tetramethyl-	4127-47-3	556
9.290	28191692	1498918.6	1-Hexanol, 2-ethyl-	104-76-7	778	1-Hexanol, 2-ethyl-	104-76-7	764
3.465	5411269	1090639.0	Cyclohexane, methyl-	108-87-2	649	2,4-Azetidinedione, 3,3-diethyl-1-methyl-	69315-91-9	639
17.103	39137376	1061691.8	Propanoic acid, 2-methyl-, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester		874	2,2,4-Trimethyl-1,3-pentanediol diisobutyrate	6846-50-0	826
4.675	6815676	700816.3	Pentanal, 2-methyl-	123-15-9	643	Pentanal, 2-methyl-	123-15-9	624
8.079	10432172	690936.1	Benzaldehyde	100-52-7	856	Benzaldehyde	100-52-7	840
10.542	24336838	684359.4	Nonanal	124-19-6	864	Nonanal	124-19-6	864
10.462	15308833	593397.1	Decane, 2,4-dimethyl-	2801-84-5	866	Undecane	1120-21-4	855
9.403	9024909	562871.8	Nonane, 3-methyl-	5911-04-6	760	Heptane, 2,2-dimethyl-	1071-26-7	739
4.574	5325346	512571.8	3-Hexanone	589-38-8	800	3-Hexanone	589-38-8	778
5.249	6611838	508457.4	Cyclotrisiloxane, hexamethyl-	541-05-9	841	Cyclotrisiloxane, hexamethyl-	541-05-9	813
10.219	9989887	503975.8	Undecane, 4,7-dimethyl-	17301-32-5	868	Undecane, 5-methyl-	1632-70-8	861
5.328	5210856	489077.3	Heptane, 2,4-dimethyl-	2213-23-2	846	Heptane, 2,4-dimethyl-	2213-23-2	832
9.666	11132638	466402.6	Decane, 2,5,9-trimethyl-	62108-22-9	835	Heptane, 3-ethyl-2-methyl-	14676-29-0	824
12.095	18644480	456356.1	Decanal	112-31-2	890	Decanal	112-31-2	886
6.727	7233508	454847.3	o-Xylene	95-47-6	874	p-Xylene	106-42-3	870
10.010	11171575	451318.1	Decane, 2,4,6-trimethyl-	62108-27-4	859	Undecane, 3,7-dimethyl-	17301-29-0	844
8.980	10922498	439906.2	Decane, 2,5,9-trimethyl-	62108-22-9	859	Octane, 3-methyl-	2216-33-3	858
6.090	6951844	436341.5	Ethylbenzene	100-41-4	842	Ethylbenzene	100-41-4	836
10.990	15583901	427665.8	Cyclopentasiloxane, decamethyl-	541-02-6	829	Cyclopentasiloxane, decamethyl-	541-02-6	777
8.682	6958224	411356.9	Undecane, 4,7-dimethyl-	17301-32-5	754	Undecane	1120-21-4	730
8.938	9947740	396126.9	Octane, 2,5,6-trimethyl-	62016-14-2	715	Undecane, 6-methyl-	17302-33-9	692
14.428	14046640	373978.0	Propanoic acid, 2-methyl-, 3-hydroxy-2,4,4-trimethylpentyl ester	74367-34-3	775	Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester	74367-31-0	750
9.101	11135269	373115.2	Undecane	1120-21-4	814	Octane, 3,5-dimethyl-	15869-93-9	806

10.085	10146656	369701.0	Hexane, 2,2,3,3-tetramethyl-	13475-81-5	832	Hexane, 2,4-dimethyl-	589-43-5	817
8.268	6379809	346087.7	2,2,7,7-Tetramethyloctane	1071-31-4	708	Octane, 2,2,6-trimethyl-	62016-28-8	693
9.750	10704183	317107.4	Butane, 2,2,3,3-tetramethyl-	594-82-1	778	Butane, 2,2,3,3-tetramethyl-	594-82-1	762
9.721	10732332	308409.6	Heptane, 2,4-dimethyl-	2213-23-2	851	Decane, 2,4-dimethyl-	2801-84-5	843
10.701	6225369	282795.1	Heptane, 4-ethyl-	2216-32-2	791	Hexane, 3-ethyl-2,5-dimethyl-	52897-04-8	784
8.331	6559336	275176.3	Decane, 2,4-dimethyl-	2801-84-5	776	3-Ethyl-3-methylheptane	17302-01-1	771
10.642	5673273	269887.4	Nonane, 3-methyl-	5911-04-6	851	Decane, 2,5,9-trimethyl-	62108-22-9	841
13.418	9623927	262973.6	Cyclohexasiloxane, dodecamethyl-	540-97-6	608	Cyclohexasiloxane, dodecamethyl-	540-97-6	602
8.473	5245829	261800.2	Cyclotetrasiloxane, octamethyl-	556-67-2	773	Cyclotetrasiloxane, octamethyl-	556-67-2	747
18.372	8128229	260361.8	Octane, 2,7-dimethyl-	1072-16-8	795	Decane, 6-ethyl-2-methyl-	62108-21-8	792
8.774	5937077	258456.8	Octane, 3,3-dimethyl-	4110-44-5	847	Heptane, 2,5,5-trimethyl-	1189-99-7	835
9.813	7448684	211385.2	Decane, 2-methyl-	6975-98-0	857	Undecane	1120-21-4	853
9.047	6904864	209584.8	Octane, 3,6-dimethyl-	15869-94-0	845	Octane, 2,3,7-trimethyl-	62016-34-6	833
15.592	7765004	204572.8	3-Isopropoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane		656	3-Butoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane	72439-84-0	630
15.144	7637076	184255.1	Bicyclo[7.2.0]undec-4-ene, 4,11,11-trimethyl-8-methylene-,[1R-(1R,*4Z,9S*)]-		782	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (E)-	18794-84-8	774
19.453	6869680	183397.2	Dodecane, 3-methyl-	17312-57-1	829	1-Iodo-2-methylundecane	73105-67-6	814
9.888	4172104	182767.3	Heptane, 2,2,4,6,6-pentamethyl-	13475-82-6	814	Heptane, 2,2,4,6,6-pentamethyl-	13475-82-6	810
9.164	6624856	171711.6	Pentane, 2,2,3,4-tetramethyl-	1186-53-4	884	Pentane, 2,2,3,4-tetramethyl-	1186-53-4	846
13.456	6095946	165123.9	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	722	1-Butanol	71-36-3	697
7.552	3278326	155374.9	Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-	2867-05-2	785	Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-	2867-05-2	769
17.237	6333548	152477.3	Undecane, 3-methyl-	1002-43-3	813	1-Iodo-2-methylundecane	73105-67-6	811
4.185	1642566	144582.7	Bicyclo[3.2.0]hepta-2,6-diene	2422-86-8	710	Toluene	108-88-3	678
9.926	4331736	142496.5	Heptane, 2,4,6-trimethyl-	2613-61-8	714	Decane, 2,5,6-trimethyl-	62108-23-0	705
11.877	6093542	138868.7	1-Octene, 6-methyl-	13151-10-5	844	1-Octene, 3,7-dimethyl-	4984-01-4	841
8.837	4322892	134160.8	Octane, 1,1'-oxybis-	629-82-3	771	Octanal	124-13-0	753
6.919	1624738	130567.1	Hexane, 2,4-dimethyl-	589-43-5	779	Hexane, 2,4-dimethyl-	589-43-5	770
9.595	3825900	115449.9	Nonane, 3,7-dimethyl-	17302-32-8	706	Octane, 2,6,6-trimethyl-	54166-32-4	701
11.065	3432983	103756.2	Benzene, 1-isocyano-3-methyl-	20600-54-8	832	Benzyl nitrile	140-29-4	811
10.341	4704398	101218.7	Undecane	1120-21-4	863	Undecane	1120-21-4	819
10.877	1939175	98318.8	Decane, 2,5,9-trimethyl-	62108-22-9	773	Undecane, 3,8-dimethyl-	17301-30-3	762
19.515	2160222	95494.2	Dodecane, 2,6,10-trimethyl-	3891-98-3	791	Octane, 2,3,3-trimethyl-	62016-30-2	790

13.544	3917256	95265.6	Decanal	112-31-2	756	Undecanal	112-44-7	752
10.303	4382942	91420.4	Heptane, 2,3,6-trimethyl-	4032-93-3	832	Heptane, 3-ethyl-	15869-80-4	829
9.537	3722704	88710.0	4,5-Dihydro-2(1H)-pentalenone		860	Benzeneacetaldehyde	122-78-1	795
18.841	2387902	88537.8	1-Iodoundecane	4282-44-4	775	Nonane, 1-iodo-	4282-42-2	765
14.670	3711036	87517.5	1-Undecene, 9-methyl-	74630-41-4	814	2-Decene, 6-methyl-, (Z)-	74630-31-2	810
19.687	3372592	87393.0	Isopropyl Myristate	110-27-0	726	Isopropyl Myristate	110-27-0	695
18.510	3455692	85034.0	Benzoic acid, 2-ethylhexyl ester	5444-75-7	777	2-Octyl benzoate	6938-51-8	714
10.600	2744698	83465.1	Hexane, 2,3,3-trimethyl-	16747-28-7	708	2,2'-Bi-2H-pyran, octahydro-	16282-29-4	706
12.933	3740359	82390.0	m-Ethylacetophenone	22699-70-3	772	Ethanone, 1-(4-ethylphenyl)-	937-30-4	759
16.040	3395184	80806.7	Octane, 2,7-dimethyl-	1072-16-8	815	Decane, 6-ethyl-2-methyl-	62108-21-8	809
14.775	3503866	80497.8	Undecane, 4,7-dimethyl-	17301-32-5	800	Decane, 6-ethyl-2-methyl-	62108-21-8	794
17.773	1659208	75858.8	Octane, 3,4,5,6-tetramethyl-	62185-21-1	737	Tridecane, 6-methyl-	13287-21-3	724
9.482	3015155	74578.9	1-Octene, 4-methyl-	13151-12-7	786	Hexane, 3-ethyl-2-methyl-	16789-46-1	775
5.542	821361	73809.9	2-Hexanol, (R)-	26549-24-6	750	2-Hexanol, (S)-	52019-78-0	746
8.599	2210752	73031.8	Pentane, 2,2,3,4-tetramethyl-	1186-53-4	812	Decane, 2,5,9-trimethyl-	62108-22-9	811
12.460	2544509	72744.2	Benzothiazole	95-16-9	794	Benzothiazole	95-16-9	781
13.691	2278269	66060.0	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	800	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	750
13.728	1800277	58968.4	Decane, 2,4-dimethyl-	2801-84-5	732	Decane, 6-ethyl-2-methyl-	62108-21-8	730
14.235	2247402	57555.4	Ethanol, 2-(2-butoxyethoxy)-, acetate	124-17-4	708	1,2-Butanediol, 3,3-dimethyl-	59562-82-2	698
14.905	2435045	55120.8	2-Decen-1-ol	22104-80-9	789	Decanal	112-31-2	786
12.003	2561420	54187.7	Decane, 6-ethyl-2-methyl-	62108-21-8	798	Decane, 2,5,9-trimethyl-	62108-22-9	794
19.884	1784126	54018.4	Oxirane, [(tetradecyloxy)methyl]-	38954-75-5	716	1-Octanol, 2,7-dimethyl-	15250-22-3	708
15.734	1458960	49756.9	3-Undecene, 2-methyl-, (Z)-	74630-48-1	797	1-Undecene, 10-methyl-	22370-55-4	793
13.075	2233948	49200.7	Decane, 2,9-dimethyl-	1002-17-1	746	Decane, 2,4-dimethyl-	2801-84-5	745
7.950	969916	46192.9	Pyruvic acid, butyl ester	20279-44-1	767	Heptane, 2,5-dimethyl-	2216-30-0	760
3.770	1103868	45961.9	Propanedioic acid, oxo-, bis(2-methylpropyl) ester	92778-43-3	532	2-Penten-1-ol, (Z)-	1576-95-0	531
16.454	1122091	42746.3	Tetradecane, 1-iodo-	19218-94-1	705	Decane, 6-ethyl-2-methyl-	62108-21-8	698
19.206	1449109	42631.8	Mercaptoacetic acid, bis(trimethylsilyl)-	6398-62-5	510	Ethanedioic acid, bis(trimethylsilyl) ester	18294-04-7	506
10.780	1807253	41953.8	Nonane, 3,7-dimethyl-	17302-32-8	785	Decane, 3,7-dimethyl-	17312-54-8	771
19.097	1172927	41856.8	Hexane, 1,6-dicyclohexyl-	1610-23-7	611	Cyclohexane, (1-methylethyl)-	696-29-7	592
17.954	1427689	41293.5	Octane, 2-bromo-	557-35-7	719	Hexadecane	544-76-3	716
10.374	2216351	40940.4	Ether, tert-butyl isopropylidenecyclopropyl	24524-56-9	629	Heptane, 2,2,3,5-tetramethyl-	61868-42-6	612

15.395	1493108	40508.1	1,7-Nonadien-4-ol, 4,8-dimethyl-	17920-92-2	720	5,9-Undecadien-2-one, 6,10-dimethyl-, (Z)-	3879-26-3	701
13.238	1571089	39305.5	Ethanone, 1-(4-ethylphenyl)-	937-30-4	762	Ethanone, 1-(4-ethylphenyl)-	937-30-4	741
19.055	1382767	38538.3	Tridecane, 1-iodo-	35599-77-0	755	1-Nonene, 4,6,8-trimethyl-	54410-98-9	750
17.397	1326848	35917.6	Z-2-Dodecenol	69064-36-4	729	Oxirane, [(tetradecyloxy)methyl]-	38954-75-5	726
5.697	457865	32846.5	2,4,6-Trimethyl-1-nonene	55771-40-9	654	1-Hexanol, 3-methyl-	13231-81-7	654
4.361	676239	32628.1	2-Methyl-3-phenyl-pyrrolo(2,3-b)pyrazine	56015-25-9	519	2-Phenyl-3-methyl-pyrrolo(2,3-b)pyrazine	56015-26-0	497
18.674	1177483	31333.3	1-Chloroundecane	2473-03-2	568	Decane, 1-chloro-	1002-69-3	563
10.822	1447950	30780.3	Undecane, 2,4-dimethyl-	17312-80-0	667	Octane, 3,4,5,6-tetramethyl-	62185-21-1	652
12.288	781232	30770.6	Adamantane, 1,3-dimethyl-	702-79-4	712	1-Methyl-3-ethyladamantane	1687-34-9	698
15.914	942488	30459.7	Bicyclo[4.1.0]-3-heptene, 2-isopropenyl-5-isopropyl-7,7-dimethyl-		674	1-Ethyl-3-vinyl-adamantane		665
19.817	696092	29927.9	Dodecane, 1-chloro-	112-52-7	654	Hexadecane, 1,16-dichloro-	7735-39-9	649
8.536	1043485	28338.2	Hexane, 2,4-dimethyl-	589-43-5	805	Hexane, 2,4-dimethyl-	589-43-5	793
18.041	1033845	27234.0	Decane, 2,5,9-trimethyl-	62108-22-9	631	Octadecane, 6-methyl-	10544-96-4	626
11.836	1020139	26676.2	Naphthalene	91-20-3	744	Naphthalene	91-20-3	734
11.312	764007	26607.7	Decane	124-18-5	770	3,4-Hexanedione, 2,2,5-trimethyl-	20633-03-8	764
18.732	1047413	26202.7	Hexane, 1-chloro-	544-10-5	589	Benzeneacetic acid, 3-tetradecyl ester		589
18.293	983495	26113.8	4-Tetradecanol	1653-33-4	625	4-Dodecanol	10203-32-4	623
19.143	874929	25978.6	1-Iodo-2-methylundecane	73105-67-6	736	Decane, 2,5,9-trimethyl-	62108-22-9	732
12.204	984079	25767.3	2-Decene, 9-methyl-, (Z)-	74630-24-3	785	4-Undecene, 10-methyl-, (E)-	74630-60-7	782
16.806	968729	24899.4	3-Hexanone, 2,4-dimethyl-	18641-70-8	681	Pentane, 2,3,3-trimethyl-	560-21-4	659
18.984	873223	24439.5	1-Octanol, 2-butyl-	3913-02-8	759	1-Decanol, 2-ethyl-	21078-65-9	732
11.459	594761	22589.8	3,4-Hexanedione, 2,2,5-trimethyl-	20633-03-8	689	Hexane, 2,5-dimethyl-	592-13-2	670
14.704	995443	21542.2	Cyclohexane, 1-ethenyl-1-methyl-2,4-bis(1-methylethenyl)-	110823-68-2	617	1,3-Bis-(2-cyclopropyl,2-methylcyclopropyl)-but-2-en-1-one		614
11.186	753505	21338.2	1-Aza-4-oxabicyco[4.1.0]heptan-5-one, 2,2-dimethyl-	94637-30-6	568	Cyclopentane, (1-methylethyl)-	3875-51-2	560
19.725	812716	20476.9	Benzoic acid, 2-fluoro-3-hydroxy-, benzyl ester		668	3-Benzylsulfanyl-3-fluoro-2-trifluoromethyl-acrylonitrile		625
4.290	1036500	19234.3	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	764	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		718
14.817	527364	18413.4	3-Methyl-2-hexene	17618-77-8	638	3-Methyl-3-hexene	3404-65-7	634
16.186	893566	18299.3	Octadecane, 1-(ethenyloxy)-	930-02-9	673	Oxirane, (3,3-dimethylbutyl)-	53907-77-0	658
18.552	775308	17817.9	3-Decyn-2-ol	69668-93-5	674	Pentanoic acid, 10-undecenyl ester		651

4.265	1077526	17802.9	Cyclotrisiloxane, hexamethyl-	541-05-9	797	Cyclotrisiloxane, hexamethyl-	541-05-9	740
19.381	536379	17702.2	2-Ethyl-1-dodecanol	19780-33-7	598	1-Dodecanol, 2-methyl-, (S)-	57289-26-6	598
11.417	683220	17362.2	1-Pentanol, 2-ethyl-4-methyl-	106-67-2	701	2,4,6-Trimethyl-1-nonene	55771-40-9	697
12.665	718268	17091.0	1-Pent-3-ynylcyclopenta-1,3-diene		577	4-(1-Hydroxyethyl)benzaldehyde	80463-21-4	556
11.555	806633	17079.1	Heptane, 2,3,5-trimethyl-	20278-85-7	737	1-Butanol, 3,3-dimethyl-	624-95-3	733
17.623	717551	16451.0	Benzeneacetic acid, 4-tridecyl ester		579	Benzene, (3,3-dimethyldecyl)-	55134-09-3	577
15.546	670999	16378.3	Butane, 1-chloro-3,3-dimethyl-	2855-08-5	549	6-Chloro-2,2,9,9-tetramethyl-3,7-decadiyn-5-ol		514
15.215	562997	15971.4	Ethanone, 1,1'-(1,4-phenylene)bis-	1009-61-6	622	1(3H)-Isobenzofuranone, 3,3-dimethyl-	1689-09-4	614
3.904	604321	15387.2	Pentane, 3-ethyl-	617-78-7	566	1-Benzopyrylium, 2-phenyl-	14051-53-7	557
3.971	372888	15290.4	Cyclotrisiloxane, hexamethyl-	541-05-9	672	Cyclotrisiloxane, hexamethyl-	541-05-9	640
16.111	751639	15142.4	Butylated Hydroxytoluene	128-37-0	702	Phenol, 2,6-bis(1,1-dimethylethyl)-4-methyl-, methylcarbamate	1918-11-2	683
13.954	448852	15102.3	1-Propanesulfonyl chloride	10147-36-1	673	Propane, 2-nitro-	79-46-9	652
5.739	366059	14952.3	Aniline	62-53-3	512	2-Pyridineacetic acid	13115-43-0	503
16.136	794971	14577.9	2-Imino-6-mercapto-4,4-dimethyl-1,2,3,4-tetrahydro-pyridine-3,5-dicarbonitrile		576	Phenol, 2-(1,1-dimethylethyl)-4-(1,1,3,3-tetramethylbutyl)-	5806-73-5	555
17.824	642075	14496.4	1-Tridecyn-4-ol	74646-37-0	717	1-Dodecyn-4-ol	74646-36-9	710
16.341	499554	14365.6	1,3-Benzodioxole, 4-methoxy-6-(2-propenyl)-	607-91-0	619	1,3-Benzodioxole, 4-methoxy-6-(2-propenyl)-	607-91-0	607
17.309	447634	13953.3	Isooctane, (ethenyloxy)-	37769-62-3	639	Hydroxylamine, O-decyl-	29812-79-1	621
12.970	596502	13913.6	1-Hexene, 3,5,5-trimethyl-	4316-65-8	685	Decane, 1-chloro-	1002-69-3	683
16.890	661892	13771.5	3-Hexanone, 2,2-dimethyl-	5405-79-8	734	Butane, 2,2-dimethyl-	75-83-2	727
19.277	348736	13679.6	Octadecane, 1-(ethenyloxy)-	930-02-9	511	Heptadecane, 9-hexyl-	55124-79-3	496
13.326	624165	13629.4	1-Penten-3-ol, 4-methyl-	4798-45-2	605	3-Pentanol, 2-chloro-4-methyl-, (R,*S*)-(ñ)-	74685-65-7	602
19.645	536472	13621.1	Octadecane, 1-(ethenyloxy)-	930-02-9	652	1-Pentanol, 2-ethyl-4-methyl-	106-67-2	651
3.934	657905	13575.2	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	660	2-Chloroaniline-5-sulfonic acid	98-36-2	622
13.624	573886	13533.9	Acetate, 2-hydroxy-2-(3-chloro-4,5-dihydro-5-isoxazolyl)-, ethyl ester		545	Indan-1,2,3-trione	938-24-9	515
16.576	390437	13481.9	1-Iodo-2-methylnonane		653	Heptane, 2,2,3,3,5,6,6-heptamethyl-	7225-67-4	650
13.850	407602	13208.1	Octane, 3-ethyl-2,7-dimethyl-	62183-55-5	671	Octane, 2,4,6-trimethyl-	62016-37-9	671
16.685	469942	13089.8	1-Nonene, 4,6,8-trimethyl-	54410-98-9	709	Hexadecane	544-76-3	666
4.110	681487	12733.3	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	624	Trimethyl(4-tert.-butylphenoxy)silane	25237-79-0	582
17.887	390039	11647.4	Hexane, 3,3-dimethyl-	563-16-6	742	3-Hexanone, 2,2-dimethyl-	5405-79-8	733

13.134	502247	11598.8	1-Hexene, 3,5,5-trimethyl-	4316-65-8	707	1-Hexene, 3,5,5-trimethyl-	4316-65-8	704

Appendix 11 Top 150 peak MS report for *Daucus carota*, limited to first 20 minutes of GCMS method, ordered by area under the peak.

RT	Height	Area	Name_1	CAS_No_1	Match1	Name_2	CAS_No_2	Match2
8.289	771978624	30089376.0	Bicyclo[3.1.0]hexane, 4-methylene-1-(1-methylethyl)-	3387-41-5	932	β-Phellandrene	555-10-2	917
7.543	130602576	6476428.0	1R-α-Pinene	7785-70-8	927	Tricyclo[2.2.1.0(2,6)]heptane, 1,3,3-trimethyl-	488-97-1	917
17.103	181271616	4613334.5	Propanoic acid, 2-methyl-, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester		892	2,2,4-Trimethyl-1,3-pentanediol diisobutyrate	6846-50-0	845
3.050	33734860	3807942.8	Heptane	142-82-5	850	Heptane	142-82-5	834
15.215	59746288	1573149.2	Cyclohexene, 3-(1,5-dimethyl-4-hexenyl)-6-methylene-, [S-(R,*S*)]-	20307-83-9	864	1H-3a,7-Methanoazulene, octahydro-3,8,8-trimethyl-6-methylene-, [3R-(3α,3aβ,7β,8aα)]-		848
13.259	61110964	1528960.2	Bicyclo[2.2.1]heptan-2-ol, 1,7,7-trimethyl-, acetate, (1S-endo)-	5655-61-8	872	Acetic acid, 1,7,7-trimethyl-bicyclo[2.2.1]hept-2-yl ester	92618-89-8	870
9.294	37475056	1457341.0	Cyclohexene, 1-methyl-4-(1-methylethenyl)-, (S)-	5989-54-8	838	Limonene	138-86-3	827
8.586	38350504	1288304.6	β-Myrcene	123-35-3	888	β-Myrcene	123-35-3	886
4.822	6967380	1124963.4	3-Penten-2-one, 4-methyl-	141-79-7	813	3-Penten-2-one, 4-methyl-	141-79-7	794
6.270	13696787	1096061.1	p-Xylene	106-42-3	861	p-Xylene	106-42-3	854
3.431	4573488	560801.9	2,4-Azetidinedione, 3,3-diethyl-1-methyl-	69315-91-9	819	Cyclohexane, methyl-	108-87-2	779
9.394	20386110	549105.1	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	877	1R-α-Pinene	7785-70-8	848
14.423	19329732	494493.0	Propanoic acid, 2-methyl-, 3-hydroxy-2,4,4-trimethylpentyl ester	74367-34-3	837	Propanoic acid, 2-methyl-, 2-ethyl-3-hydroxyhexyl ester	74367-31-0	779
10.910	18228030	430244.1	Bicyclo[3.1.0]hex-2-ene, 4,4,6,6-tetramethyl-	19487-09-3	815	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	812
15.110	15510813	402567.9	1,3-Cyclohexadiene, 5-(1,5-dimethyl-4-hexenyl)-2-methyl-, [S-(R,*S*)]-	495-60-3	776	1H-3a,7-Methanoazulene, 2,3,4,7,8,8a-hexahydro-3,6,8,8-tetramethyl-, [3R-(3α,3aβ,7β,8aα)]-		764
9.796	12884567	396258.4	Cyclohexene, 4-methylene-1-(1-methylethyl)-	99-84-3	822	Bicyclo[3.1.0]hexan-2-ol, 2-methyl-5-(1-methylethyl)-	546-79-2	793
7.849	9251357	387190.6	Bicyclo[2.2.1]heptane, 7,7-dimethyl-2-methylene-	471-84-1	833	Camphene	79-92-5	809
10.458	14392633	383521.7	Octane, 3,5-dimethyl-	15869-93-9	701	Decane, 2,4,6-trimethyl-	62108-27-4	696
4.671	3750230	372053.2	Pentanal, 2,3-dimethyl-	32749-94-3	558	Pentanal, 2-methyl-	123-15-9	547
8.075	5043959	327711.9	Benzaldehyde	100-52-7	796	Benzaldehyde	100-52-7	789

4.562	3731922	297431.2	3-Hexanone	589-38-8	770	3-Hexanone	589-38-8	758
3.017	42961652	288011.5	Heptane	142-82-5	838	Hexane, 3-methyl-	589-34-4	821
6.082	4745038	253178.7	Ethylbenzene	100-41-4	834	Ethylbenzene	100-41-4	832
6.718	4028622	239208.0	p-Xylene	106-42-3	848	o-Xylene	95-47-6	848
5.316	3670982	234346.7	Heptane, 2,4-dimethyl-	2213-23-2	838	Heptane, 2,4-dimethyl-	2213-23-2	825
3.586	2021647	228434.6	1-Pentanol, 2,3-dimethyl-	10143-23-4	605	Di-n-hexyl-diselenide	52056-08-3	593
5.236	3647212	217916.9	Cyclotrisiloxane, hexamethyl-	541-05-9	856	Cyclotrisiloxane, hexamethyl-	541-05-9	817
7.418	3529357	197911.1	α-Phellandrene	99-83-2	757	α-Phellandrene	99-83-2	757
9.579	6318585	184105.4	1,3,7-Octatriene, 3,7-dimethyl-	502-99-8	734	3-Carene	13466-78-9	731
13.414	7508516	182443.3	Bis(trimethylsilyl) 2,2-difluoro-1-(trifluoromethyl)ethenylphosphonate	110410-49-6	600	Cyclohexasiloxane, dodecamethyl-	540-97-6	596
9.223	4801297	182332.9	Benzene, 1-methyl-3-(1-methylethyl)-	535-77-3	771	Benzene, 1-methyl-2-(1-methylethyl)-	527-84-4	771
9.089	4500516	182050.5	Cyclohexene, 3-methyl-6-(1-methylethylidene)-	586-63-0	779	Cyclohexene, 4-methylene-1-(1-methylethyl)-	99-84-3	715
10.537	6853828	172254.2	Nonanal	124-19-6	804	Nonanal	124-19-6	801
15.592	6805464	167479.1	3-Isopropoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane		603	3-Butoxy-1,1,1,7,7,7-hexamethyl-3,5,5-tris(trimethylsiloxy)tetrasiloxane	72439-84-0	588
9.721	4547644	128671.7	Undecane, 4,7-dimethyl-	17301-32-5	767	Heptane, 2,4-dimethyl-	2213-23-2	766
8.477	4243915	124730.3	Cyclotetrasiloxane, octamethyl-	556-67-2	790	Cyclotetrasiloxane, octamethyl-	556-67-2	778
13.452	4643000	121897.0	Furan, tetrahydro-2,5-dimethyl-	1003-38-9	699	Furan, tetrahydro-2,5-dimethyl-	1003-38-9	661
18.372	3492366	117565.9	Hexadecane	544-76-3	794	Eicosane	112-95-8	774
12.924	5351468	113686.5	m-Ethylacetophenone	22699-70-3	774	Ethanone, 1-(2,5-dimethylphenyl)-	2142-73-6	764
10.986	4520434	110923.9	Cyclopentasiloxane, decamethyl-	541-02-6	815	Cyclopentasiloxane, decamethyl-	541-02-6	791
11.873	4712418	98817.8	Cyclopropane, 1-ethyl-2-heptyl-	74663-86-8	835	1-Octene, 6-methyl-	13151-10-5	832
5.517	1135231	91174.6	Formic acid, 1-methylethyl ester	625-55-8	726	4-Amino-1-butanol	13325-10-5	684
15.454	4088414	90549.4	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (E)-	18794-84-8	743	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (Z)-	28973-97-9	706
10.248	3286073	88723.0	Cyclohexene, 3-methyl-6-(1-methylethylidene)-	586-63-0	769	(+)-4-Carene	29050-33-7	693
17.233	3924637	86716.0	Octane, 2,7-dimethyl-	1072-16-8	746	1-Iodo-2-methylundecane	73105-67-6	741
19.448	3501144	86609.8	Decane, 2,9-dimethyl-	1002-17-1	760	Tetradecane, 1-iodo-	19218-94-1	752
12.091	3331811	83173.5	Decanal	112-31-2	829	2-Decen-1-ol, (E)-	18409-18-2	781
17.526	2760139	79233.0	Silane, [[4-[1,2-bis[(trimethylsilyl)oxy]ethyl]-1,2-phenylene]bis(oxy)]bis[trimethyl-		643	Silane, [[4-[1,2-bis[(trimethylsilyl)oxy]ethyl]-1,2-phenylene]bis(oxy)]bis[trimethyl-		632
8.779	2726246	74891.1	Undecane, 4,7-dimethyl-	17301-32-5	771	Tridecane	629-50-5	767

14.666	2745048	70445.5	1-Decene, 8-methyl-	61142-79-8	824	1-Octene, 6-methyl-	13151-10-5	814
16.094	2782948	68469.3	1,3,6-Heptatriene, 2,5,5-trimethyl-	29548-02-5	664	1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-, (Z,E)-	26560-14-5	649
14.231	2437907	59063.3	Ethanol, 2-(2-butoxyethoxy)-, acetate	124-17-4	779	Ethanol, 2-(2-butoxyethoxy)-, acetate	124-17-4	752
4.177	1169962	56015.3	Toluene	108-88-3	702	Bicyclo[3.2.0]hepta-2,6-diene	2422-86-8	701
6.915	881812	48830.9	Octane	111-65-9	790	Octane	111-65-9	782
15.952	2349068	47806.3	Benzene, 1,1'-ethylidenebis-	612-00-0	755	Benzene, 1,1'-ethylidenebis-	612-00-0	725
15.910	2174658	45211.8	1-Ethyl-3-vinyl-adamantane		675	1,6-Cyclodecadiene, 1-methyl-5-methylene-8-(1-methylethyl)-, [s-(E,E)]-		663
13.938	2229171	43115.1	Cyclohexene, 3-methyl-6-(1-methylethylidene)-	586-63-0	738	1,3,6-Heptatriene, 2,5,5-trimethyl-	29548-02-5	700
13.544	1635343	42635.1	(S)-3,4-Dimethylpentanol		779	Hydroperoxide, hexyl	4312-76-9	725
16.036	1957388	40232.2	Tetradecane, 1-iodo-	19218-94-1	775	Nonane, 1-iodo-	4282-42-2	774
6.174	1096254	38388.6	Hexane, 2,3,4-trimethyl-	921-47-1	807	Hexane, 2,3,4-trimethyl-	921-47-1	803
12.451	1308317	38012.6	1H-Pyrazolo[3,4-d]pyrimidin-4-amine	2380-63-4	714	Benzothiazole	95-16-9	702
10.680	511524	35523.7	Heptane, 3,4,5-trimethyl-	20278-89-1	732	Hexane, 2,3,3-trimethyl-	16747-28-7	719
18.506	1614278	35073.3	Benzoic acid, 2-ethylhexyl ester	5444-75-7	698	1-Butanol, 3-methyl-, benzoate	94-46-2	667
4.206	918682	33004.9	Pentane, 2-methyl-	107-83-5	684	(Aminomethyl)cyclopropane	2516-47-4	647
13.071	1512167	32246.1	Octane, 3,4,5,6-tetramethyl-	62185-21-1	727	Nonane, 1-iodo-	4282-42-2	725
11.312	1535425	31616.3	Ethanone, 1-(2-methylphenyl)-	577-16-2	644	Benzene, 1-ethenyl-4-methoxy-	637-69-4	639
18.841	1078632	31517.5	3-Hexanone, 4-ethyl-	6137-12-8	683	Tetradecane, 1-iodo-	19218-94-1	682
13.682	1463866	31488.8	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	799	Furan, tetrahydro-2,5-dimethyl-	1003-38-9	799
3.733	586440	31072.2	1-Propene, 1-(2-propenyloxy)-, (E)-	61142-13-0	612	2-Chloro-2-methylhexane	4398-65-6	594
19.201	1228097	30411.5	2-(2',4',4',6',6',8',8'-Heptamethyltetrasiloxan-2'-yloxy)-2,4,4,6,6,8,8,10,10-nonamethylcyclopentasiloxane		481	1,3-Dioxolane	646-06-0	473
11.999	1438923	30408.9	3,4-Hexanedione, 2,2,5-trimethyl-	20633-03-8	718	Octane, 2,7-dimethyl-	1072-16-8	715
13.728	1238034	30029.3	2-Furanmethanol, tetrahydro-	97-99-4	659	3-Heptanone, 2,4-dimethyl-	18641-71-9	639
9.532	1319336	29283.9	Benzaldehyde, 4-benzyloxy-3-methoxy-2-nitro-	2450-27-3	837	Benzene, 2-methoxy-1-(2-nitroethenyl)-3-(phenylmethoxy)-	74810-83-6	825
8.892	980579	28892.9	1,4-Methano-1H-Cyclopropa[d]pyridazine, 4,4a,5,5a-tetrahydro-6,6-dimethyl-, (1α,4α,4aα,5aα)-		745	α-Phellandrene	99-83-2	714
8.925	826308	27720.9	Bicyclo[3.1.0]hex-2-ene, 4-methyl-1-(1-methylethyl)-	28634-89-1	714	Cyclohexene, 4-methylene-1-(1-methylethyl)-	99-84-3	702
17.773	734122	27512.6	Heptane, 2,2,3,3,5,6,6-heptamethyl-	7225-67-4	651	Decane, 2,5,9-trimethyl-	62108-22-9	646
14.901	1263634	26828.4	Decanal	112-31-2	780	2-Decen-1-ol	22104-80-9	759
19.515	749863	26249.3	Pentane, 2,3,3-trimethyl-	560-21-4	724	3-Ethyl-3-methylheptane	17302-01-1	716

14.771	1198328	25771.0	4-Heptanone, 3-methyl-	15726-15-5	800	3,4-Hexanedione, 2,2,5-trimethyl-	20633-03-8	792
14.545	1456436	25589.8	α-Cubebene	17699-14-8	658	α-Cubebene	17699-14-8	652
19.687	1053882	25166.2	1-Octanol, 2-nitro-	2882-67-9	619	Methyl 2-hydroxydecanoate	71271-24-4	588
10.006	832285	24204.8	1-Undecyn-4-ol	22127-86-2	644	1,7-Nonadien-4-ol, 4,8-dimethyl-	17920-92-2	638
4.327	502005	20975.1	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	733	Trimethyl[4-(1,1,3,3,-tetramethylbutyl)phenoxy]silane	78721-87-6	694
19.884	637316	20662.0	Isooctane, (ethenyloxy)-	37769-62-3	679	Nonane, 1-iodo-	4282-42-2	669
5.701	392024	19973.8	Cyclopropanecarboxylic acid,3-methylbutyl ester		667	1-Hexanol, 4-methyl-	818-49-5	667
19.105	466092	17325.4	1-Ethyl-3,5-di-n-propyladamantane		505	5-Methyl-2-(N-methylbenzylamino)-2-thiazoline	77158-40-8	493
9.160	491291	16989.3	Octane, 2,3,7-trimethyl-	62016-34-6	640	Octane, 2-bromo-	557-35-7	636
16.371	458334	16600.6	Butane, 1,3-dichloro-2-methyl-	23010-07-3	477	5-Hepten-1-yne, 6-methyl	22842-10-0	467
15.805	882033	16563.9	1H-Cyclopenta[1,3]cyclopropa[1,2]benzene, octahydro-7-methyl-3-methylene-4-(1-methylethyl)-, [3aS-(3aα,3bβ,4β,7α,7aS*)]-		640	1,6-Cyclodecadiene, 1-methyl-5-methylene-8-(1-methylethyl)-, [s-(E,E)]-		613
8.842	613744	16297.1	Pentane, 3-methylene-	760-21-4	677	2-Pentene, 3-methyl-, (Z)-	922-62-3	676
5.630	421816	16263.1	Urea, N-cyclohexyl-N'-phenyl-	886-59-9	612	Imidazole, 2-cyano-	31722-49-3	582
12.656	578667	15934.4	Tricyclo[4.2.1.0(2,5)]non-3-en-9-endo-ol, 9-exo-ethyl-, endo-		567	4-(1-Hydroxyethyl)benzaldehyde	80463-21-4	563
19.381	561992	15384.1	1-Dodecyn-4-ol	74646-36-9	709	3-Trifluoroacetoxy-6-ethyldecane	116436-59-0	698
16.442	569806	15260.9	1-Iodoundecane	4282-44-4	724	Decane, 2,4-dimethyl-	2801-84-5	719
17.953	722657	15098.2	Octane, 2-bromo-	557-35-7	744	Nitric acid, nonyl ester	20633-13-0	706
17.870	598308	13866.8	Xanthene, 9,9-dimethyl-	19814-75-6	634	Xanthene, 9,9-dimethyl-	19814-75-6	629
10.328	430939	13795.2	Nitric acid, nonyl ester	20633-13-0	665	1-Nonene, 4,6,8-trimethyl-	54410-98-9	654
15.722	566530	13340.1	3-Undecene, 2-methyl-, (Z)-	74630-48-1	748	4-Trifluoroacetoxyoctane	116465-17-9	743
19.055	593157	12985.9	1-Iodo-2-methylnonane		669	1-Iodo-2-methylundecane	73105-67-6	667
18.289	453133	12660.2	4-Dodecanol	10203-32-4	623	4-Dodecanol	10203-32-4	603
12.288	503694	12625.4	Ethaneperoxoic acid, 1-cyano-1-[2-(2-phenyl-1,3-dioxolan-2-yl)ethyl]pentyl ester		720	1,3-Dioxolane-2-heptanenitrile, α-methyl-ë-oxo-2-phenyl-	58422-90-5	683
15.261	629653	12570.8	1,3,5-Cycloheptatriene, 2,4-diethyl-7,7-dimethyl-		613	1,3,5-Cycloheptatriene, 2,5-diethyl-7,7-dimethyl-		607
12.200	516486	12278.2	1-Hexene, 3,3,5-trimethyl-	13427-43-5	789	5-Undecene, 5-methyl-, (Z)-	57024-93-8	772
11.836	615451	12192.3	2,4,5-Trihydroxypyrimidine	496-76-4	712	1-[2-Pyridyl]-2,2-dimethyl-2-morpholino ethanol		703
16.350	537480	12101.9	Naphthalene, 1,2,3,5,6,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, (1S-cis)-		600	Naphthalene, 1,2,3,5,6,8a-hexahydro-4,7-dimethyl-1-(1-methylethyl)-, (1S-cis)-		597

16.563	352035	11630.4	Propanoic acid, 2-methyl-, 3,7-dimethyl-2,6-octadienyl ester, (E)-	2345-26-8	540	2-Bromononane	2216-35-5	536
7.954	470287	11576.3	Ethanone, 1-(3-methyloxiranyl)-	17257-79-3	683	Propanoic acid, 2-oxo-	127-17-3	675
11.471	337226	11316.6	3-Cyclohexene-1-ethanol, α-ethenyl-α,3-dimethyl-6-(1-methylethylidene)-		570	1,3,6,10-Dodecatetraene, 3,7,11-trimethyl-, (Z,E)-	26560-14-5	515
3.900	399681	11294.9	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		706	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	689
10.144	344134	10752.8	Butane, 1-chloro-3,3-dimethyl-	2855-08-5	648	1-Dodecyn-4-ol	74646-36-9	621
9.913	470447	10267.7	Ethanone, 2-(formyloxy)-1-phenyl-	55153-12-3	833	Phenacylidene diacetate	5062-30-6	832
4.311	602953	10132.1	2-Furanmethanol, tetrahydro-	97-99-4	515	2-Methyl-pentanoic acid [4-(2-methyl-pentanoylsulfamoyl)-phenyl]-amide		509
11.115	292774	9935.8	1,6-Octadien-3-ol, 3,7-dimethyl-, propanoate	144-39-8	566	1-Dodecyn-4-ol	74646-36-9	511
16.136	493975	9847.8	3,4-Dimethyl-2-(3-methyl-butyryl)-benzoic acid, methyl ester	71940-29-9	625	2-Imino-6-mercapto-4,4-dimethyl-1,2,3,4-tetrahydro-pyridine-3,5-dicarbonitrile		599
10.077	402192	9407.2	1,3-Dioxan-4-one, 2-(1,1-dimethylethyl)-6-methyl-	113505-75-2	567	Propane, 1-chloro-3-diethylboryloxy-2,2-dimethyl-		536
18.674	391383	9343.0	Benzoic acid, 2-fluoro-3-hydroxy-, benzyl ester		575	Benzofenac methyl ester	63007-66-9	572
10.843	454462	9117.5	2-(3-Methylbuta-1,3-dienyl)cyclohexanone		644	1-Methyl-3-ethyladamantane	1687-34-9	633
17.392	438120	8853.4	1-Octanol, 2-nitro-	2882-67-9	638	1-Hexene, 5,5-dimethyl-	7116-86-1	635
18.992	329603	8553.9	1-Nonene, 4,6,8-trimethyl-	54410-98-9	652	Nitric acid, nonyl ester	20633-13-0	640
11.249	352366	8523.3	1-Octanol, 2-nitro-	2882-67-9	559	4-Octene, 2,3,6-trimethyl-	63830-65-9	551
19.147	335481	8522.0	Decane, 2,5,9-trimethyl-	62108-22-9	757	Decane, 6-ethyl-2-methyl-	62108-21-8	756
16.802	324861	8438.5	Pentane, 1,3-epoxy-4-methyl-	15045-60-0	662	Hexadecane	544-76-3	645
16.182	359007	8115.9	1-Octanol, 2-nitro-	2882-67-9	661	1-Undecyn-4-ol	22127-86-2	637
19.809	315369	8101.3	1-Pentene, 3,3,4-trimethyl-5-phenyl-		519	Benzene, (1-ethyloctyl)-	4621-36-7	517
15.537	326866	8070.5	6-Chloro-2,2,9,9-tetramethyl-3,7-decadiyn-5-ol		562	Propanoic acid, 2,2-dimethyl-, 2,6-bis(1,1-dimethylethyl)-4-methylphenyl ester		552
18.037	374256	8057.5	Butane, 1-chloro-3,3-dimethyl-	2855-08-5	669	Butane, 2,3-dimethyl-2,3-dinitro-	3964-18-9	645
4.080	251860	7881.9	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		824	5-Methyl-2-trimethylsilyloxy-acetophenone	97389-69-0	723
18.724	337053	7841.3	β-l-Rhamnopyranoside, phenyl-2,3-O-ethylboranediyl-4-O-benzyl-		557	Benzofenac methyl ester	63007-66-9	545
12.799	378308	7642.6	Propanoic acid, 2-methyl-, 2-phenylethyl ester	103-48-0	755	β-Phenylethyl butyrate	103-52-6	747
9.666	373771	7588.7	3-Hexanone, 2,4-dimethyl-	18641-70-8	652	Decane, 2,5,9-trimethyl-	62108-22-9	647
11.970	455837	7496.5	5,7-Octadien-2-ol, 2,6-dimethyl-	5986-38-9	660	7-Octen-2-ol, 2-methyl-6-methylene-	543-39-5	653

4.265	495740	7289.6	4-Methyl-2,4-bis(4'-trimethylsilyloxyphenyl)pentene-1		847	Cyclotrisiloxane, hexamethyl-	541-05-9	756
15.391	304694	7083.8	Ethanone, 1-cyclopropyl-	765-43-5	625	Ethanone, 1-cyclopropyl-	765-43-5	615
16.685	217767	7000.4	Vinyl butyrate	123-20-6	661	2H-Pyran-2,3-diol, tetrahydro-, diacetate, trans-	3021-94-1	655
15.839	427421	6943.2	Bicyclo[3.1.1]hept-2-ene, 2,6-dimethyl-6-(4-methyl-3-pentenyl)-	17699-05-7	652	1,3-Cyclohexadiene, 5-(1,5-dimethyl-4-hexenyl)-2-methyl-, [S-(R,*S*)]-	495-60-3	643
17.610	280079	6529.2	Benzene, [(methylsulfinyl)methyl]-	824-86-2	610	Benzofenac methyl ester	63007-66-9	610
18.875	472995	6510.9	5-Hepten-3-one, 5-ethyl-4-methyl-	74764-56-0	491	1,3-Benzodioxol-2-one, hexahydro-, trans-	20192-66-9	465
14.964	288399	6405.6	4-Pentenal, 2,2-dimethyl-	5497-67-6	601	4-Pentenal, 2-methyl-	5187-71-3	589
13.854	241664	6172.5	Octane, 2-methyl-	3221-61-2	593	1-Trifluoroacetoxy-2-methylpentane	155089-96-6	587
14.825	283768	6129.3	3-Hexene, 2,5-dimethyl-, (E)-	692-70-6	689	4-Trifluoroacetoxyoctane	116465-17-9	687
4.026	281408	6123.6	2-t-Butyl-5-(dimethoxy-phosphoryl)-3-methyl-4-oxoimidazolidine-1-carboxylic acid, t-butyl ester		505	6H-1,3-Oxazin-6-one, 4-acetoxy-2-bromomethyl-		365
11.429	218528	6064.2	4-Methylcyclohexanol acetate	22597-23-5	577	3-Hepten-1-ol	10606-47-0	566
12.832	186718	6010.0	1-Adamantanecarboxylic acid, 3-phenyl-2-propenyl ester		440	1,5,6,7-Tetrahydro-4-indolone	13754-86-4	439
3.875	256849	5799.2	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	507	2-Methylpentyl formate		476
7.782	275772	5793.3	Di-n-propyl ether	111-43-3	682	Di-n-propyl ether	111-43-3	645
14.587	287242	5670.8	Vinylbital	2430-49-1	698	5,16:8,13-Diethenodibenzo[a,g]cyclododecane, 6,7,14,15-tetrahydro	17341-02-5	630
18.134	217993	5585.3	Dodecane, 1-chloro-	112-52-7	524	Decane, 1-chloro-	1002-69-3	520
17.819	235688	5517.0	Nitric acid, nonyl ester	20633-13-0	687	Nitric acid, decyl ester	2050-78-4	663
8.976	196533	5450.5	2,2,4-Trimethyl-3-pentanone	5857-36-3	553	Phenol, m-(1-methyl-3-propyl-3-pyrrolidinyl)-	428-37-5	536

Appendix 12 Top 150 peak MS report for *Heracleum sphondylium*, limited to first 20 minutes of GCMS method, ordered by area under the peak.

RT	Height	Area	Name_1	CAS_No_1	Match1	Name_2	CAS_No_2	Match2
12.133	401357312	9623534.0	Acetic acid, octyl ester	112-14-1	909	Acetic acid, octyl ester	112-14-1	905
7.546	164453040	7928732.5	1R-α-Pinene	7785-70-8	941	1R-α-Pinene	7785-70-8	928
9.583	219240464	6576075.0	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	951	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	933
9.296	119464696	4027283.5	Limonene	138-86-3	920	D-Limonene	5989-27-5	878
9.396	121473832	3405012.8	1,3,6-Octatriene, 3,7-dimethyl-, (E)-	3779-61-1	931	1,3,6-Octatriene, 3,7-dimethyl-, (Z)-	3338-55-4	927

15.454	144289952	3285783.2	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (E)-	18794-84-8	916	1,6,10-Dodecatriene, 7,11-dimethyl-3-methylene-, (Z)-	28973-97-9	914
10.913	131490792	3253326.8	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	939	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	915
14.142	126904256	3029354.2	Dodecane, 1-bromo-	143-15-7	904	Decane, 1-bromo-	112-29-8	904
8.587	64895792	2161591.5	β-Myrcene	123-35-3	897	β-Myrcene	123-35-3	887
9.792	51867976	1475832.2	Bicyclo[3.1.0]hexan-2-ol, 2-methyl-5-(1-methylethyl)-	546-79-2	877	1,4-Cyclohexadiene, 1-methyl-4-(1-methylethyl)-	99-85-4	862
7.854	30132960	1354550.5	Camphene	79-92-5	910	Camphene	79-92-5	891
4.833	7775909	1336538.0	2-Pentene, 4,4-dimethyl-, (Z)-	762-63-0	736	3-Hexen-2-one	763-93-9	712
3.054	31102078	1304209.2	Heptane	142-82-5	847	Heptane	142-82-5	837
8.288	32426646	1158639.4	Bicyclo[3.1.0]hex-2-ene, 4-methyl-1-(1-methylethyl)-	28634-89-1	879	β-Phellandrene	555-10-2	835
6.279	14140980	1154420.8	p-Xylene	106-42-3	858	p-Xylene	106-42-3	857
11.938	41230332	985264.8	3-Octen-1-ol, acetate, (Z)-	69668-83-3	812	3-Octen-2-ol, (E)-	57648-55-2	806
17.100	34687888	909825.1	Propanoic acid, 2-methyl-, 1-(1,1-dimethylethyl)-2-methyl-1,3-propanediyl ester		895	2,2,4-Trimethyl-1,3-pentanediol diisobutyrate	6846-50-0	825
10.538	29882360	864555.3	Nonanal	124-19-6	870	Nonanal	124-19-6	865
13.258	31687550	793801.4	Bornyl acetate	76-49-3	852	Isobornyl acetate	125-12-2	849
10.258	23916728	680321.3	trans-Linaloloxide		829	2-Furanmethanol, 5-ethenyltetrahydro-α,α,5-trimethyl-, trans-	34995-77-2	819
17.013	29117984	670717.4	Hexanoic acid, octyl ester	4887-30-3	849	Hexanoic acid, 9-decen-1-yl ester	180252-09-9	801
14.896	28915246	658234.9	Dodecanal	112-54-9	890	Dodecanal	112-54-9	879
10.458	22358918	580132.8	1,6-Octadien-3-ol, 3,7-dimethyl-, propanoate	144-39-8	695	1,6-Octadien-3-ol, 3,7-dimethyl-	78-70-6	692
16.100	21760886	494208.1	α-Farnesene	502-61-4	800	α-Farnesene	502-61-4	779
4.658	5401630	474232.8	1-Propen-2-ol, formate	32978-00-0	600	Pentanal, 2,3-dimethyl-	32749-94-3	558
9.208	14503679	467454.4	1,3,8-p-Menthatriene	21195-59-5	840	Benzene, 1-methyl-4-(1-methylethyl)-	99-87-6	777
4.554	4715120	391548.0	3-Hexanone	589-38-8	774	3-Hexanone	589-38-8	769
15.179	15012979	332704.0	Propanoic acid, 2,2-dimethyl-, octyl ester	27751-88-8	825	Butanoic acid, 2-methyl-, octyl ester	29811-50-5	791
10.004	10160286	326720.0	Cyclopropane, pentyl-	2511-91-3	773	1-Octanol	111-87-5	764
5.317	4168394	324694.2	Hexane, 2,3,5-trimethyl-	1069-53-0	802	Undecane, 2,4-dimethyl-	17312-80-0	795
15.913	8371123	315565.4	1,6-Cyclodecadiene, 1-methyl-5-methylene-8-(1-methylethyl)-, [s-(E,E)]-		711	1-Ethyl-3-vinyl-adamantane		702
5.238	5116076	303026.9	Cyclotrisiloxane, hexamethyl-	541-05-9	854	Cyclotrisiloxane, hexamethyl-	541-05-9	817
14.421	11544446	292202.0	Butanoic acid, butyl ester	109-21-7	786	Propanoic acid, 2-methyl-, 3-hydroxy-2,4,4-trimethylpentyl ester	74367-34-3	786

7.408	7025064	291493.9	Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-	2867-05-2	816	Bicyclo[3.1.0]hex-2-ene, 2-methyl-5-(1-methylethyl)-	2867-05-2	794
14.613	11776112	277388.1	Propanoic acid, 2-methyl-, octyl ester	109-15-9	822	Butanoic acid, octyl ester	110-39-4	817
9.129	7879642	264147.0	1-Methoxycycloheptatriene	27324-85-2	818	Benzene, 1-methoxy-4-methyl-	104-93-8	814
8.075	4343627	261225.7	Benzaldehyde	100-52-7	752	Benzaldehyde	100-52-7	748
11.117	8882651	253760.1	Lilac aldehyde C	53447-47-5	741	Lilac aldehyde B	53447-46-4	734
6.725	4434094	251588.1	o-Xylene	95-47-6	833	p-Xylene	106-42-3	828
4.171	2665883	236010.5	Bicyclo[3.2.0]hepta-2,6-diene	2422-86-8	808	1,3,5-Cycloheptatriene	544-25-2	775
11.250	8786061	235237.7	Lilac aldehyde C	53447-47-5	801	Lilac aldehyde B	53447-46-4	800
6.088	4552611	229255.5	Ethylbenzene	100-41-4	851	Ethylbenzene	100-41-4	841
8.371	8187808	218950.1	Cyclohexene, 4-methylene-1-(1-methylethyl)-	99-84-3	752	Bicyclo[3.1.1]heptane, 6,6-dimethyl-2-methylene-, (1S)-	18172-67-3	751
12.092	10090580	204491.7	Decanal	112-31-2	853	Decanal	112-31-2	825
10.988	8415760	183474.2	Cyclopentasiloxane, decamethyl-	541-02-6	870	Cyclopentasiloxane, decamethyl-	541-02-6	837
18.371	5120590	175070.4	Octane, 2,3,3-trimethyl-	62016-30-2	761	Hexadecane	544-76-3	753
19.450	7189442	169310.2	Decane, 6-ethyl-2-methyl-	62108-21-8	818	2-Bromo dodecane	13187-99-0	808
12.929	6411080	167844.6	m-Ethylacetophenone	22699-70-3	756	Ethanone, 1-(2,5-dimethylphenyl)-	2142-73-6	753
3.450	3185958	164256.5	2,4-Azetidinedione, 3,3-diethyl-1-methyl-	69315-91-9	729	Cyclohexane, methyl-	108-87-2	715
13.450	5345192	157980.5	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	767	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	751
13.413	5166556	143321.8	Bis(trimethylsilyl) 2,2-difluoro-1-(trifluoromethyl)ethenylphosphonate	110410-49-6	616	Cyclohexasiloxane, dodecamethyl-	540-97-6	586
9.717	5919174	138914.5	Undecane, 4,7-dimethyl-	17301-32-5	823	Heptane, 2,4-dimethyl-	2213-23-2	816
8.479	5043051	138667.0	Cyclotetrasiloxane, octamethyl-	556-67-2	780	Cyclotetrasiloxane, octamethyl-	556-67-2	763
14.663	6110902	136625.8	2-n-Butylacrolein	1070-66-2	638	3-Methylpenta-1,3-diene-5-ol, (E)-	1572-08-3	614
13.542	5653552	129874.9	Oxirane, decyl-	2855-19-8	763	Decanal	112-31-2	762
3.208	3252793	126861.2	Furan, tetrahydro-2,5-dimethyl-, trans-(ñ)-	38484-59-2	846	Furan, tetrahydro-2,5-dimethyl-, cis-	2144-41-4	817
11.875	5508188	124779.1	5-Methyl-1-heptanol	7212-53-5	827	1-Octene, 3,7-dimethyl-	4984-01-4	820
15.592	4544320	119516.5	Trisiloxane, 1,1,1,5,5,5-hexamethyl-3,3-bis[(trimethylsilyl)oxy]-	3555-47-3	541	Pentasiloxane, dodecamethyl-	141-63-9	531
5.529	1536600	116093.4	Formic acid, 1-methylethyl ester	625-55-8	703	Formic acid, 1-methylethyl ester	625-55-8	694
12.504	5063126	112622.8	Butanoic acid, 3-methyl-, hexyl ester	10032-13-0	813	Butanoic acid, 2-methyl-, hexyl ester	10032-15-2	800
3.708	1341916	106303.3	1,2,4-Benzenetricarboxylic acid, 1,2-dimethyl ester	54699-35-3	537	Methyl vinyl ketone	78-94-4	534
3.629	1362266	105261.1	Hydroperoxide, hexyl	4312-76-9	563	Heptanol	53535-33-4	551
8.842	3130878	98813.8	Octanal	124-13-0	778	Octanal	124-13-0	766
8.779	3281019	94197.4	Decane, 2,5,9-trimethyl-	62108-22-9	803	Decane, 6-ethyl-2-methyl-	62108-21-8	783

17.233	4169574	91232.3	Hexadecane	544-76-3	796	Dodecane, 3-methyl-	17312-57-1	778
17.508	2379580	83902.4	Apiol	523-80-8	704	Apiol	523-80-8	672
18.500	3666238	81766.9	Benzoic acid, 2-ethylhexyl ester	5444-75-7	748	2-Octyl benzoate	6938-51-8	703
11.642	3120308	80622.6	1,4-Pentadiene, 2,3,3-trimethyl-	756-02-5	815	1,4-Pentadiene, 2,3,3-trimethyl-	756-02-5	797
11.475	3419412	80225.0	Lilac aldehyde C	53447-47-5	642	Lilac aldehyde A	53447-45-3	638
16.938	3359002	78663.5	(E)-2-Butenoic acid, 2-(methylenecyclopropyl)prop-2-yl ester		706	1,5-Heptadiene, 2,6-dimethyl-	6709-39-3	689
10.758	3342244	76381.3	3-Octanol, acetate	4864-61-3	729	Acetic acid, octyl ester	112-14-1	672
19.683	3157570	72514.2	Isopropyl Myristate	110-27-0	681	Isopropyl Myristate	110-27-0	660
16.033	2790214	67611.4	Decane, 2,5,9-trimethyl-	62108-22-9	794	Hexadecane	544-76-3	786
12.000	2283291	63123.0	Decane, 2,4-dimethyl-	2801-84-5	773	3-Hexanone, 2,4-dimethyl-	18641-70-8	772
13.688	2008008	56811.9	Oxirane, 2-methyl-3-(1-methylethyl)-	1192-31-0	775	Furan, tetrahydro-2,5-dimethyl-	1003-38-9	721
17.196	2988118	56738.8	Tricyclo[3.1.0.0(2,4)]hexane, 3,6-diethyl-3,6-dimethyl-, trans-		652	1,3-Cyclopentadiene, 1,3-bis(1-methylethyl)-	123278-27-3	647
19.204	1993723	52145.2	3-Dodecanol, 3,7,11-trimethyl-	7278-65-1	605	3-Pentanol, 2,4-dimethyl-	600-36-2	602
17.142	2368030	50601.2	(S)-(+)-6-Methyl-1-octanol	110453-78-6	777	1-Octene, 6-methyl-	13151-10-5	728
19.883	1451883	50196.5	Tridecane, 1-iodo-	35599-77-0	727	1-Nonene, 4,6,8-trimethyl-	54410-98-9	725
18.837	1150102	49663.0	Tetradecane, 1-iodo-	19218-94-1	754	1-Iodoundecane	4282-44-4	735
19.517	1282987	49187.7	Heptane, 2,5,5-trimethyl-	1189-99-7	749	Undecane, 3,3-dimethyl-	17312-65-1	733
15.113	1562182	46686.7	1H-Cyclopenta[1,3]cyclopropa[1,2]benzene, octahydro-7-methyl-3-methylene-4-(1-methylethyl)-, [3aS-(3aα,3bβ,4β,7α,7aS*)]-		674	1,6-Cyclodecadiene, 1-methyl-5-methylene-8-(1-methylethyl)-, [s-(E,E)]-		653
14.771	1834521	46399.0	Tetradecane, 1-iodo-	19218-94-1	760	Decane, 1-iodo-	2050-77-3	757
12.663	1773903	45327.4	1-(2,4-Dimethylphenyl)ethanol	99500-87-5	560	1-Pent-3-ynylcyclopenta-1,3-diene		554
13.071	1917955	43123.6	Tetradecane, 1-iodo-	19218-94-1	723	Nonane, 1-iodo-	4282-42-2	722
12.454	1657998	41911.8	1H-Pyrazolo[3,4-d]pyrimidin-4-amine	2380-63-4	695	Benzothiazole	95-16-9	673
17.767	891076	38040.3	Tridecane, 6-methyl-	13287-21-3	721	Undecane, 3-methyl-	1002-43-3	718
14.229	1420477	36349.8	1,2-Butanediol, 3,3-dimethyl-	59562-82-2	689	1,2-Butanediol, 3,3-dimethyl-	59562-82-2	682
6.958	722388	35861.4	Butane, 2-cyclopropyl-	5750-02-7	593	Furan, 2,5-dihydro-	1708-29-8	561
13.754	1435189	34845.5	Cyclohexane, chloro-	542-18-7	745	Cyclohexane, chloro-	542-18-7	742
15.729	1308172	34330.2	2-Octene, 2,6-dimethyl-	4057-42-5	777	1-Decene, 8-methyl-	61142-79-8	762
6.921	1096816	32603.9	Octane, 2,4,6-trimethyl-	62016-37-9	776	Nonane	111-84-2	768
11.413	1197056	31969.2	1-Heptene, 5-methyl-	13151-04-7	777	1-Decene, 8-methyl-	61142-79-8	775
11.308	1048442	31861.3	1-Hexene, 3,3,5-trimethyl-	13427-43-5	716	Cyclopentane, nitro-	2562-38-1	710

16.883	947624	31597.2	Oxirane, [(tetradecyloxy)methyl]-	38954-75-5	746	Nitric acid, nonyl ester	20633-13-0	735
17.950	1020078	31456.5	Octane, 2-bromo-	557-35-7	754	Heptane, 3-ethyl-5-methyl-	52896-90-9	738
10.700	602162	31087.0	Heptane, 2,3-dimethyl-	3074-71-3	673	Pentane, 3-ethyl-2,3-dimethyl-	16747-33-4	655
12.196	1045439	30273.0	1-Decene, 8-methyl-	61142-79-8	677	3-Undecene, 9-methyl-, (E)-	74630-54-9	675
17.392	1335327	30207.5	Z-2-Dodecenol	69064-36-4	726	Decanal	112-31-2	721
13.717	1169075	30007.2	Decane, 6-ethyl-2-methyl-	62108-21-8	686	Octane, 3,4,5,6-tetramethyl-	62185-21-1	670
4.333	937427	29346.3	2-Furanol, tetrahydro-2-methyl-	7326-46-7	672	2-Furanmethanol, tetrahydro-	97-99-4	664
16.179	1302465	29039.8	1-Octanol, 2-nitro-	2882-67-9	704	Oxirane, [(tetradecyloxy)methyl]-	38954-75-5	697
18.288	1302641	27293.0	3-Dodecanol, 3,7,11-trimethyl-	7278-65-1	501	4-Dodecanol	10203-32-4	493
5.713	496073	27194.0	Hexane, 2,3-dimethyl-	584-94-1	478	Butane, 1-chloro-3-methyl-	107-84-6	474
19.100	934309	26686.4	1,2,3-Undecanetriol	100537-15-3	506	1-Dodecyn-4-ol	74646-36-9	500
19.133	919312	26433.1	Heptane, 3-ethyl-5-methyl-	52896-90-9	723	Hexane, 3,3-dimethyl-	563-16-6	701
15.033	1266732	26188.2	1,6-Octadiene, (E)-	19036-81-8	663	3-Octen-1-ol, (Z)-	20125-84-2	660
10.204	1341565	24948.2	1,4-Cyclohexadiene, 3,3,6,6-tetramethyl-	2223-54-3	718	2,4,6-Octatriene, 2,6-dimethyl-, (E,Z)-	7216-56-0	714
5.650	537924	23075.8	1H-Pyrazole-4-carbonitrile	31108-57-3	645	1H-Pyrazole-4-carbonitrile	31108-57-3	619
11.837	1051526	22780.4	Naphthalene	91-20-3	723	Naphthalene	91-20-3	710
14.012	836640	22507.8	Heptane, 3-ethyl-5-methyl-	52896-90-9	705	Pentane, 2,3,4-trimethyl-	565-75-3	699
10.317	684697	22254.6	1-Octene, 6-methyl-	13151-10-5	710	Cyclopentane, 1,2,4-trimethyl-	2815-58-9	702
6.179	729781	20059.4	Hexane, 2,3,4-trimethyl-	921-47-1	801	Octane, 4-methyl-	2216-34-4	797
12.292	771013	19753.9	Adamantane, 1,3-dimethyl-	702-79-4	677	1-Methyl-3-ethyladamantane	1687-34-9	667
15.254	939711	19365.1	(+)-Epi-bicyclosesquiphellandrene	54324-03-7	636	2-Phenylacetamide, N-(1-phenyl-2-propyl)-		629
3.904	533999	19091.6	Pentane, 3-ethyl-	617-78-7	694	Hexane, 2,3-dimethyl-	584-94-1	693
19.054	842024	18398.6	Octane, 3,4,5,6-tetramethyl-	62185-21-1	722	1-Nonene, 4,6,8-trimethyl-	54410-98-9	716
5.142	826107	18268.9	Trimethyl[4-(1,1,3,3,-tetramethylbutyl)phenoxy]silane	78721-87-6	752	Cyclotrisiloxane, hexamethyl-	541-05-9	747
16.442	559090	16643.0	3-Hexanone, 2,5-dimethyl-	1888-57-9	713	3,5-Dimethyl-4-octanone	7335-17-3	683
6.488	447701	16374.2	Methyl vinyl ketone	78-94-4	721	Methyl vinyl ketone	78-94-4	713
12.833	902532	16334.6	Benzenemethanol, α,α,4-trimethyl-	1197-01-9	589	Benzenemethanol, α,α,4-trimethyl-	1197-01-9	569
15.538	716945	15831.2	6-Chloro-2,2,9,9-tetramethyl-3,7-decadiyn-5-ol		564	Decane, 1-chloro-	1002-69-3	534
16.800	590511	15606.1	Octane, 2-bromo-	557-35-7	648	2-Bromo-6-methylheptane	4730-24-9	645
18.038	655628	14633.8	2-Bromononane	2216-35-5	639	Hexadecanoic acid, 2-oxo-, methyl ester	55836-30-1	635
17.825	499340	13846.9	2-Hexene, 4,4,5-trimethyl-	55702-61-9	688	2-Heptene, 5-ethyl-2,4-dimethyl-	74421-06-0	667

19.808	463219	13701.5	m-Toluylic acid, 2-ethylhexyl ester	16397-66-3	567	p-Toluylic acid, 2-ethylhexyl ester	16397-65-2	557
19.367	403883	12908.5	1-Dodecyn-4-ol	74646-36-9	699	Methyl 2-hydroxydecanoate	71271-24-4	684
18.675	592157	12667.1	Octane, 1-chloro-	111-85-3	609	Benzoic acid, 2-fluoro-3-hydroxy-, benzyl ester		598
9.867	409748	12425.9	Oxirane, (3,3-dimethylbutyl)-	53907-77-0	706	1-Hexene, 4,4-diethyl-		683
15.762	604277	12407.9	Ethanone, 1-[4-(1-hydroxy-1-methylethyl)phenyl]-	54549-72-3	571	Benzyl alcohol, α,α-dimethyl-p-isopropyl-	3445-42-9	536
11.587	478958	12311.5	1-Pentene, 3,4-dimethyl-	7385-78-6	699	1-Pentene, 2,4-dimethyl-	2213-32-3	695
18.133	359306	12243.2	4-Trifluoroacetoxyoctane	116465-17-9	645	1-Dodecyn-4-ol	74646-36-9	643
13.321	571324	12152.7	1-Penten-3-ol, 4-methyl-	4798-45-2	593	Cyclobutanemethanol, α-methyl-	7515-29-9	573
13.879	459528	12143.6	2,2-Dimethylpropionic acid, cyclopentyl ester		596	6-Hepten-3-one, 4-methyl-	26118-97-8	580
7.950	423864	11936.5	2-Carbomethoxyaziridine	5950-34-5	574	2-Carbomethoxyaziridine	5950-34-5	574
12.708	557545	11899.1	5-Hexen-2-one, 5-methyl-	3240-09-3	586	Ethanone, 1-cyclopropyl-	765-43-5	582
10.846	618547	11820.7	1,2-Benzenedicarboxylic acid, dihexyl ester	84-75-3	617	1H-S-Triazolo[1,5-a]pyridin-4-ium, 2-hydroxy-1-methyl-, hydroxide, inner salt		611
18.721	510050	11796.4	Benzofenac methyl ester	63007-66-9	630	Phenol, 2-benzyloxy-3,6-difluoro-	152434-80-5	568
10.142	314631	11630.0	2-Propenoic acid, 2-methyl-, octyl ester	2157-01-9	633	4-Hexen-3-one, 2,2-dimethyl-	20971-19-1	617
8.996	301892	11516.7	Trifluoroacetic acid, 2-methylpropyl ester	17355-83-8	669	2-Propanone, 1,1,1-trifluoro-	421-50-1	645
16.725	266404	11291.0	Pentanenitrile, 4-methyl-	542-54-1	616	2,3-Epoxyhexanol	90528-63-5	613
18.175	357633	11214.6	4-Heptanol, 2,6-dimethyl-	108-82-7	417	2,4-Pentanedione, 1,1,1-trifluoro-	367-57-7	414
15.387	418747	10958.7	Ethanone, 1-cyclopropyl-	765-43-5	616	3-Butyn-2-ol, 2-methyl-	115-19-5	589
18.983	509651	10867.2	Heptafluorobutyric acid, 2-tetrahydrofurylmethyl ester		599	1-Nonene, 4,6,8-trimethyl-	54410-98-9	597
14.542	731459	10826.4	Bicyclo[4.1.0]-3-heptene, 2-isopropenyl-5-isopropyl-7,7-dimethyl-		636	Copaene	3856-25-5	613
13.625	487233	10715.5	2-Butanethiol, 2-methyl-	1679-09-0	547	Nitric acid, heptyl ester	20633-12-9	537
14.825	367205	10367.4	2-Heptene, 4-methyl-, (E)-	66225-17-0	676	1-Pentene, 2,3-dimethyl-	3404-72-6	671
13.958	305885	10317.8	1-Nonene, 4,6,8-trimethyl-	54410-98-9	636	Pentane, 2,3,3-trimethyl-	560-21-4	633
11.558	569688	10294.0	(S)-3,4-Dimethylpentanol		688	Nitric acid, nonyl ester	20633-13-0	684
13.854	398671	10073.4	3,4-Hexanedione, 2,2,5-trimethyl-	20633-03-8	725	Butane, 2,2-dimethyl-	75-83-2	708