Universidade Nova de Lisboa

Instituto de Higiene e Medicina Tropical

Genetic studies on the mosquito vector *Culex pipiens*.

Bruno Gomes da Silva

Licenciado em Biologia da Universidade do Porto

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Doutor no Ramo de Ciências Biomédicas, Especialidade em Parasitologia, realizada sob orientação científica do Prof. Dr. João Pinto.

Orientador: Prof. Dr. João Pinto
Unidade de Parasitologia Médica
Instituto de Higiene e Medicina Tropical

Co-orientadores: Prof. Dr. António P.G. Almeida
Unidade de Parasitologia Médica
Instituto de Higiene e Medicina Tropical

Prof. Dr. Martin J. Donnelly
Department of Vector Biology
Liverpool School of Tropical Medicine

Comissão Tutorial: Prof. Dr. Henrique Silveira
Unidade de Parasitologia Médica
Instituto de Higiene e Medicina Tropical

O trabalho foi financiado pela Fundação para a Ciência e Tecnologia, através da bolsa de doutoramento SFRH/BD/36410/2007 e dos projectos de investigação POCI/BIA-BDE/57650/2004 e PPCDT/BIA-BDE/57650/2004.

JANEIRO, 2013

Para o casal que me deu a fala,

E o tradutor que me ligou ao Mundo.

Acknowledgements

Firstly, I would like to acknowledge to my colleagues at IHMT and LSTM. My sincerely thanks to João Pinto, my supervisor, for his knowledge, enthusiasm and patience, which were welcome and essential in helping me to complete this project and thesis. Martin J. Donnelly, my co-supervisor, provided support and advice during the period I spent at Liverpool School of Tropical Medicine and I appreciate his enthusiasm in the development of this project. António P. G. Almeida, my co-supervisor, provided essential support in the field and in the ecological perspectives of the thesis.

Carla A. Sousa, Patricia Salgueiro and José L. Vicente were essential partners during the last four years to perform several tasks in the laboratory and field at IHMT during the last four years. Craig Wilding, David Weetman and Keith Steen supported the development of the laboratory and molecular analysis performed at LSTM. Other colleagues from the IHMT, Ana R. Côrte-Real, Eliane Arez, Ferdinando B. Freitas, Inês Vieira, Isabel Calderón, Joana Alves, Leonor Pinho, Maria T. Freitas, Ricardo Alves and Teresa L. Silva also provided valuable assistance in this project.

I would like to thank Harry Savage (Centers for Disease Control and Prevention, USA), John Vontas (University of Crete, Greece) and Marta Santa-Ana (University of Madeira, Portugal), for field work support and providing mosquito samples. Also, thanks to Deirdre Walshe and Teresa L. Silva for proofreading this thesis.

I would like to acknowledge the administrative support from the Departments of Entomology, Malaria, and Parasitology at IHMT and Vector Biology at LSTM. This study was funded by a PhD fellowship from the Fundação para a Ciência e Tecnologia/MCTES (SFRH/BD/36410/2007) and by projects of the Fundação para a Ciência e a Tecnologia, Portugal (POCI/BIA-BDE/57650/2004 and PPCDT/BIA-BDE/57650/2004).

Last but by no means least, thank you to all of my family and friends who helped me to grown as a person, especially my parents and my big brother. The moments that we spent together, from calm conversations to noisy moments, helped me to maintain my sanity over the past four years.

Resumo

As duas espécies do complexo *Culex pipiens* com maior distribuição geográfica, *Culex quinquefasciatus* e *Culex pipiens sensu stricto*, são importantes vectores de filárias e arbovírus. *Culex pipiens s.s.* apresenta categorias intra-específicas definidas por características ecológicas e fisiológicas, das quais as formas pipiens e molestus têm sido implicadas na transmissão do vírus da Febre do Nilo Ocidental na Europa e América do Norte.

Hibridação entre *Cx. quinquefasciatus* e *Cx. pipiens s.s.* foi documentada em algumas regiões geográficas onde ambas espécies coexistem simpatricamente. Este fenómeno também foi descrito entre as formas molestus e pipiens, em áreas de simpatria e quando existe contacto limitado em certas épocas do ano. No entanto, o impacto da hibridação na divergência genética entre as espécies ou formas está por clarificar. Além disso, a hibridação pode afectar características ecológicas/fisiológicas das espécies/formas, que podem influenciar a sua capacidade vectorial. Neste contexto, foram analisadas populações do complexo *Cx. pipiens* da Europa, EUA e da Macaronésia com objectivo de determinar níveis de diferenciação genética e taxas de hibridação entre os membros do complexo.

As amostras de mosquitos foram obtidas por diferentes métodos de colheita no terreno e a partir de colónias laboratoriais, entre 2005 e 2011. As análises genéticas realizadas foram baseadas em microssatélites e por polimorfismos de comprimento de fragmentos amplificados. Foram efectuadas comparações abordando questões específicas a diferentes níveis taxonómicos, que estão descritas nos cinco capítulos de resultados da tese.

A distribuição e níveis de hibridação entre *Cx. quinquefasciatus* e *Cx. pipiens s.s.* foram avaliados nas ilhas da Macaronésia, o que permitiu detectar híbridos (~40%) em duas ilhas do arquipélago de Cabo Verde. A distribuição destas espécies na região reflecte a biogeografia e aspectos históricos da colonização humana.

A coexistência em habitats de superfície das formas molestus e pipiens na região da Comporta (Portugal), foi demostrada pela combinação de análises fenotípicas e genéticas. As análises moleculares também sugerem a existência de um padrão de introgressão assimétrica, de molestus para pipiens. Estudos adicionais, sugerem uma

maior tendência da forma molestus para explorar habitats intradomiciliares/antropogénicos quando comparada com a forma pipiens. Em ambas as formas, mais de 90% das refeições sanguíneas foram realizadas em aves.

Foi ainda efectuada a primeira análise genómica focada na divergência entre os genomas das formas molestus e pipiens. Esta análise indicou uma baixa divergência entre os dois genomas (1,4%–3,1%), o que é consistente com um processo de especiação simpátrica com fluxo génico.

Finalmente, foram realizadas análises genéticas em amostras de *Cx. pipiens s.s.* colhidas na Grécia durante um surto de Febre do Nilo Ocidental, em 2010. Populações simpátricas de molestus e pipiens com introgressão assimétrica foram identificadas na região onde o surto ocorreu, enquanto uma população homogénea de molestus foi encontrada numa região sem transmissão do vírus.

Estes resultados evidenciam a importância da caracterização da variação genética e das relações evolutivas entre os membros do complexo *Cx. pipiens* para entender o seu potencial como vectores de doenças. Também abrem novas perspectivas para a investigação da ecologia e evolução deste complexo de espécies com importância médica.

PALAVRAS-CHAVE: *Culex pipiens*, *Culex quinquefasciatus*, molestus, genética populacional, hibridação, especiação com fluxo génico, vírus de Febre do Nilo.

Abstract

The two widespread species of the *Culex pipiens* complex, *Culex quinquefasciatus* and *Culex pipiens sensu stricto*, are major vectors of filarial worms and arboviruses. *Culex pipiens s.s.* is also divided into intraspecific categories defined by ecological and physiological traits. Of these, two forms, denoted pipiens and molestus, have been implicated in West Nile virus transmission in Europe and North America.

Inter-specific hybridisation between *Cx. quinquefasciatus* and *Cx. pipiens s.s.* has been documented in some geographic regions where both species occur sympatrically. Likewise, hybridisation between molestus and pipiens forms has been described in areas of sympatry or when the forms become in contact during certain times of the year. However, the impact of hybridisation on the extent of genetic divergence between species or forms remains uncertain. Moreover, hybridisation may affect ecological and physiological traits of the species/forms, which may influence their vectorial capacity. In this context, the degree of genetic differentiation and hybridisation between members of the *Cx. pipiens* complex was studied in populations from Europe, USA and Macaronesian islands.

Mosquito samples were obtained from field collections or laboratory colonies between 2005 and 2011. Genetic analyses were based on microsatellite genotypes and amplified fragment length polymorphisms. Comparisons were made at different taxonomic levels, addressing specific questions. These are described in the five results chapters of this thesis.

The distribution and hybridisation between *Cx. quinquefasciatus* and *Cx. pipiens s.s.* were assessed in Macaronesian islands. Hybrid rates ~40% were detected in two islands of the Cape Verde archipelago. The distribution of the species reflects both the islands' biogeography and historical aspects of human colonization.

A combination of phenotypic and genetic analyses conducted in Comporta (Portugal) revealed the co-occurrence of molestus and pipiens forms of *Cx. pipiens s.s.* in aboveground habitats. Moreover, a pattern of asymmetric introgression from molestus into pipiens was found. Subsequent molecular and ecological analyses carried out in the same region suggested that the molestus form has a higher tendency to

explore indoor/anthropogenic habitats, when compared with the sympatric pipiens form. In both forms, over 90% of blood meals were made on avian hosts.

The first genomic scan addressing levels of genome divergence between molestus and pipiens forms was implemented. Low levels of inter-form genomic divergence (1.4%–3.1%) were detected, consistent with a process of sympatric speciation with gene flow.

Finally, *Cx. pipiens s.s.* samples collected in Greece during a WNV outbreak in 2010 were genetically characterised. Sympatric molestus and pipiens populations with asymmetric introgression were detected in the region where the outbreak occurred, whereas a more genetically homogenous molestus population was found in a region with no WNV transmission.

These results highlight the importance of characterizing patterns of genetic variation and evolutionary relations among members of the *Cx. pipiens* complex as a requirement for understanding the potential of these species to act as disease vectors. They also open new perspectives for further research on the ecology and evolution of this species complex of medical importance.

KEYWORDS: *Culex pipiens, Culex quinquefasciatus,* molestus, population genetics, hybridisation, speciation with gene flow, West Nile virus.

Table of Contents

Abbreviations

6-FAM	6-carboxyfluorescein, fluorescent dye with an absorbance maximum of 492 nm and an emission maximum of 517 nm.
ace-2	acetylcholinesterase-2 gene
Ae.	*Aedes*
AFLP	amplified fragment length polymorphism
An.	*Anopheles*
A_R	allelic richness
bp	base pair
BWh	arid hot steppe climate
BxP	backcross *Cx. pipiens s.s.* (in chapter 2)
ca.	*circa*
CDC	Centers for Disease Control and Prevention
CO_2	carbon dioxide
CQ11FL	molecular assay in the 5' flanking region of the CQ11 microsatellite
Csa	temperate climate with dry and hot summers
Csb	temperate climate with dry and warm summers
Cx.	*Culex*
cyt b	*cytochrome b*
Dc	chord distance
DENV1	dengue virus serotype 1
DGS	Direcção-Geral de Saúde
d.f.	degrees of freedom
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide triphosphates
DV/D	morphological identification of males based on the length of the dorsal and ventral arms of the phallosome
ECDC	European Centre for Disease Prevention and Control
EDTA	ethylenediaminetetraacetic acid
e.g.	*exempli gratia*
ELISA	enzyme-linked immunosorbent assay
et al.	*et alli*
f	frequency
F1	first generation of hybrids or colony offspring
F2	hybrids of the second generation (cross from two hybrids individuals)
F_{IS}	inbreeding coefficient
F_{ST}	genetic differentiation according to Weir & Cockerham (1984)
i.e.	*id est*
IgG	immunoglobulin G
IHMT	Instituto de Higiene e Medicina Tropical

H'	Shannon index
H_e	expected heterozygosity,
HEX	hexachloro-fluorescein, fluorescent dye with an absorbance maximum of 535 nm and an emission maximum of 556 nm
HLC	human landing catches
IR	indoor resting collections
JEV	Japanese encephalitis virus
JEVs	Japanese encephalitis virus serogroup
K	cluster
L.	Linnaeus
LSTM	Liverpool School of Tropical Medicine
MDE	mutation drift equilibrium
$MgCl_2$	magnesium chloride
min	minutes
mtDNA	mitochondrial deoxyribonucleic acid
N	North
N	sample size
NCBI	National Center for Biotechnology Information
NED	fluorescent dye with an absorbance maximum of 546 nm and an emission maximum of 575 nm.
NGS	next generation sequencing
NJ	neighbor-joining
P	probability
P450s	a large and diverse group of enzymes that catalyze the oxidation of organic substances
${}_pA_R$	private allele richness
PCA	principal component analysis
PCR	polymerase chain reaction
PLP	proportion of polymorphic loci at the 5% level
PO	posterior odds
Pr(X|*K*)	probability of the data under each *K*
q_i	probability of ancestry
RH	relative humidity
R_{ST}	genetic differentiation according Slatkin (1995)
sec	seconds
s.l.	*sensu lacto*
SMM	stepwise mutation model
spp.	several species
s.s.	*sensu stricto*
Tq	posterior probability threshold
TPM	wo-phase model
U	Unit

UK	United Kingdom
USA	United States of America
VEEV	Venezuelan equine encephalitis virus
vs.	versus
W.	*Wolbachia*
W	West
WHO	World Health Organization
WNV	West Nile virus
χ^2	P-values of chi-square tests
α	nominal significance level of rejection of the null hypothesis
α	Parameter of the admixture model in the STUCTURE
ΔK	*ad hoc* approach to infer the most likely number of clusters in the sample by the STRUCTURE
λ	Parameter of the frequency correlation model in the STUCTURE

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 1.

General Introduction

Vector-borne diseases: an emerging challenge

Arthropods are involved in the transmission of pathogens to vertebrates being responsible for the dissemination of several infectious diseases which impact human populations. The phylum Arthropoda is the largest and most diverse animal group and occupies almost every habitat and major food chain of the planet. An important behavioural feature for an arthropod to become a vector of pathogens is blood feeding (hematophagy). This behaviour evolved independently in several arthropod taxa such as ticks, fleas, and dipterous insects (Grimaldi & Engel, 2005). The family Culicidae (Order: Diptera) is considered to be the most important arthropod taxa from a medical perspective due to their role in the transmission of vector-borne diseases such as malaria, yellow fever and dengue (Eldridge, 2005).

Malaria, caused by *Plasmodium spp*. parasites, is the vector-borne disease with the highest impact in terms of human deaths worldwide, with more than 650,000 deaths per year estimated in 2010 (WHO, 2011). Malaria control campaigns in the 1950s/1960s successfully eradicated the disease from the majority of the temperate regions. However, malaria control efforts have suffered several setbacks in the last decades, including the emergence of drug resistance in parasite populations and insecticide resistance in some of the major mosquito vectors (Vinayak *et al.,* 2010; Ranson *et al.,* 2011; Goldberg *et al.,* 2012; Khamsiriwatchara *et al.,* 2012; Tan *et al.,* 2012).

Over the past few decades, other vector-borne diseases have increased their impact on human populations, especially diseases caused by arthropod-borne viruses (arboviruses), expanding their original endemic areas into new/naïve territories. Dengue virus has been estimated to cause 50-100 million cases of dengue fever and 250,000 to 500,000 cases of dengue haemorrhagic fever worldwide every year (Rigau-Pérez *et al.,* 1998; WHO 2012a). It was first reported in the 18th century as a benign, nonfatal disease of visitors to the tropics. The first dengue pandemics occurred during World War II in southeast Asia and it became a global problem after the expansion of the virus and its main vector, *Aedes aegypti* (Linnaeus 1762), during the 1980s (Gubler & Clark, 1995).

West Nile virus (WNV) originally detected as a cause of encephalitis in Uganda in 1937, and confined to the eastern hemisphere, with outbreaks in Africa, Europe and Asia, until 1999, made its appearance that same year in New York (USA). West Nile virus rapidly spread throughout the USA, reaching the west coast in only three years (1999-2002). In the first decade in that continent, WNV caused ~360,000 illness cases of which over 1,300 were fatal (Kilpatrick, 2011).

In the summer of 2007, the first large outbreak of Chikungunya virus documented in a temperate climate and European country occurred in Italy. Molecular analysis confirmed 204 of the 281 suspected cases tested (Angelini *et al.*, 2007). A recently established mosquito population of *Aedes albopictus* Skuse 1894, introduced in the early 1990s, was the vector of the disease that may infected about 10% of the 3,968 inhabitants of two small towns in north-eastern Italy (Moro *et al.*, 2010; Poletti *et al.*, 2011; Medlock *et al.*, 2012).

More recently, a dengue outbreak occurred in Madeira Island (Portugal). The first infections were detected in the beginning of October 2012 and by the beginning of December, a total of 1993 cases had been reported (DGS, 2012). Transmission of DENV1 virus has been sustained by a recently introduced *Ae. aegypti* population that was first recorded in 2005 (Almeida *et al.*, 2007).

Several factors may contribute to the emergence and re-emergence of vector-borne diseases. The spread of resistance to drugs and insecticides is considered a major obstacle for the sustainability of control programmes (Ranson *et al.*, 2011). Climate/environmental changes, especially the increase of temperature and variation in rainfall partners, has been associated with expansion of vector populations and higher vector-borne incidence (McMichael *et al.*, 2006; Vora, 2008). The wide geographic distribution and high abundance of synanthropic vector species such as *Ae. aegypti* and *Culex quinquefasciatus* Say 1823 have been associated with human-mediated dispersal (human travelling) and the increase of urbanization (Subra, 1981; Lounibos, 2002; Gubler, 2011). Also, the increase of urbanization and deforestation has promoted a higher contact of humans with pathogens and their vectors previously restricted to wild habitats (Vora, 2008). This new paradigm is a cause of concern in public health systems of many temperate countries which have discontinued vector control programs for

almost 50 years. Such was the case of several European countries after obtaining the malaria eradication status by the World Health Organization, during the 1970s (WHO, 2006).

The complexity of host-parasite relations and vectorial systems

The life-cycle of vector-borne pathogens is complex. It involves the interaction of the pathogen with two very different host organisms, normally an arthropod vector and a vertebrate host.

Most arthropod species are not involved in transmission of pathogens since several traits are required in order to acquire vector competence. Intrinsic permissiveness to infection, replication/multiplication, and transmission of the pathogen (vector competence) is essential for an arthropod to act as a vector. In addition, ecological/physiological characteristics of the vector (*e.g.* host preference, longevity and biting cycle) and interactions with the ecosystem also influence its role as a vector in the natural life-cycle of the pathogen. Therefore, depending on the ecological setting an arthropod species with high vector competence may have a minor importance than another with lower vector competence. Finally, transmission of pathogens in nature is often sustained by more than one vector species that occur sympatrically.

Some of the better studied vectorial systems are those that sustain malaria transmission. There are five *Plasmodium* species that infect humans, which can be transmitted by females of *circa* 70 out of 480 *Anopheles* species. The importance of these species in the spread of malaria may vary throughout geographic regions, and primary vector species in one area may have little importance in another (Service, 1993a,b).

Dengue virus transmission may involve a sylvatic cycle among monkeys and arboreal canopy-dwelling *Aedes spp.* mosquitoes in Southeast Asia and West Africa. This cycle is restricted to the forest but accidental infection of humans and transient spillover to peri-urban areas may occur. However, dengue is also maintained in an urban cycle among humans by two urban mosquito species, *Ae. aegypti* and *Ae. albopictus*, in tropical and subtropical areas (Vasilakis *et al.*, 2011).

In addition to being transmitted by several vector species, pathogens can also infect more than one vertebrate taxon. Pathogens or diseases that normally circulate in non-human hosts but can occasionally infect humans are designated as zoonotic. Bubonic plague was probably the vector-borne zoonosis with highest impact in human history, with three large pandemics occurring between ancient and modern ages (Walløe, 2008). This disease is caused by a bacterial infection, *Yersinia pestis* Lehmann & Neumann 1896, which is maintained in nature by an enzootic cycle among small rodents and fleas. Human infections are acquired by the bite of infected fleas (Gage, 2005). Several vector-borne zoonosis caused by arboviruses such as Japanese encephalitis virus (JEV) and Venezuelan equine encephalitis virus (VEEV) also have a heavy toll on human health. The JEV circulate in an enzootic cycle among birds and mosquitoes, being transmitted mostly by mosquitoes of the genus *Culex*. Humans and domestic mammals are accidentally infected (Hayes, 2001). The VEEV is maintained in nature by an enzootic cycle in rodents or a secondary cycle in horses. The mosquito species involved in those cycles vary with the vertebrate hosts (Weaver & Barrett, 2004).

The ability of a vector to transmit a pathogen among different vertebrate species depends on its blood feeding preferences. Some vector species bite a wider range of vertebrate hosts while others are more specific. The feeding behaviour may also vary between populations of the same species since host availability and habitat characteristics may also influence this trait (Balenghien *et al.*, 2011). Species and populations with a wider range of hosts will be better zoonotic vectors than specialist species. A catholic feeding behaviour may also be promoted by gene flow between populations with different feeding preferences, of which, members of the *Culex pipiens* complex are one of the classic examples (Fonseca *et al.*, 2004).

The strategies used to control vector-borne diseases vary according to the type of pathogen, vectors and hosts involved, thus, the disease ecosystem. The characterisation of the vectorial systems at local and regional levels is, therefore, crucial for the design and implementation of sustainable control strategies and programmes, especially when control measures are focused on the vector populations in order to prevent or interrupt the transmission chain (Higgs & Beaty, 2004).

Population structure, speciation and hybridisation of vectors

The identification of the vector species involved in the transmission of vector-borne diseases has presented many challenges to medical entomologists. A classic example is the *"Anophelism without malaria"* paradigm. Following the discovery of malaria transmission by mosquitoes of the genus *Anopheles* (Grassi *et al.,* 1899 *fidé* Bruce-Chwatt, 1988; Gilles, 1993), malaria epidemiologists soon realised that the incidence of malaria cases in Europe at the beginning of the 20th century only partially overlapped with the distribution of the malaria vector *Anopheles maculipennis* Meigen 1818 *sensu lato* (Stepphens & Christophers 1902, *fidé* Fantini, 1994). This scenario was only fully understood when the notion of sibling/cryptic species emerged during the 1930s/1940s: *"We have defined as sibling species sympatric forms which are morphologically very similar or indistinguishable, but which possess specific biological characteristics and are reproductively isolated"* (Mayr, 1999). This concept supported the description of five sibling species related to *An. maculipennis* in Europe that showed differences in vector competence for malaria transmission (Bates, 1940). Today, many major insect vectors of various pathogens were found to belong to sibling species complexes. Among the many examples are the *Anopheles gambiae* complex, which includes *Anopheles gambiae* Giles 1902 *sensu stricto* and *Anopheles arabiensis* Patton 1905 as major Afrotropical malaria vectors; the *Culex pipiens* complex, that includes two ubiquitous vectors of arboviroses and filarial worms, *Culex pipiens* L. 1758 *s.s.* and *Cx. quinquefasciatus*; and the *Simulium damnosum* complex that gathers 12 cytoforms capable of transmitting the human onchocerciasis (river blindness) worm, *Onchocerca volvulus* Leuckart 1893, in West Africa (Smith & Fonseca, 2004; Post *et al.,* 2007, 2011; Kamali *et al.*, 2012).

The definition of species and the interspecific variation among vectors are not the only challenges in vector studies. Intraspecific variation among populations often occurs when the geographic distribution range of a species exceeds the capacity of individual dispersal, or when different selective pressures act in the different populations. In both scenarios, the evolutionary forces promoting genetic variation in local populations, coupled with restricted gene flow, may result in population differentiation across the species range. This may ultimately lead to sufficient phenotypic differentiation resulting in reproductive isolation and speciation. In this

context, incipient species are considered to be a group of genetically distinct populations that are in the process of becoming "true species" (*i.e.* reproductively isolated). The divergence among intraspecific populations may promote variation in the ability of transmitting pathogens.

Hybridisation is the inter-crossing of genetically distinct taxa leading to the production of viable offspring. Hybridisation was considered to be a rare phenomenon but several studies have shown a greater range of hybridisation in nature than previously predicted (Mallet, 2005). Hybridisation can lead to introgression, *i.e.* the invasion of foreign genetic material into a genome, which may have great evolutionary impact. The genetic exchange can vary among the entities involved (*e.g.* asymmetric, bidirectional), and across the genome of each entity. The gene flow may be overcome by divergent selection in specific genomic regions leading to adaptive divergence (Mallet, 2005; Nosil *et al.,* 2009). This process of heterogeneous genetic divergence is considered a major process of sympatric ecological speciation and has been described in several insect species (Machado *et al.,* 2002; Turner *et al.,* 2005; Egan *et al.,* 2008; Weetman *et al.,* 2012).

In medically important arthropods, the evolutionary relevance of hybridisation has also a public health dimension, as they affect dispersal of genes of interest such as those related with insecticide resistance or refractoriness to infection. As an example, the Afrotropical malaria vector *An. gambiae s.s.* comprises two molecular forms, denoted M and S, that are considered units of an on-going incipient speciation process (della Torre *et al.,* 2001). Genetic analyses have revealed remarkable differences between sympatric M and S populations in genes encoding for the complement-like Thioester-Containing Protein 1 with antiparasitic activity and in the frequency of insecticide knockdown resistance mutations (Santolamazza *et al.,* 2008; White *et al.,* 2011).

The genomic leap of population genetics

Genomics and molecular biology have enabled the discovery of several types of DNA markers that have been used in population genetic and evolutionary studies (Sunnucks, 2000). The information obtained from each type of marker varies with

characteristics such as abundance/distribution in the genome, degree of polymorphism and dominance. Different markers may also serve different research purposes and selection of the appropriate genetic marker depends on the type of biological/evolutionary question to be addressed.

Population genetic studies often rely on assessing variation at a relatively small number of loci, which may not be sufficient to fully understand the genetic factors that influence evolutionary processes in natural populations. Molecular techniques that collect high amounts of genetic data across the genome have been recently developed resulting in the emergence of population genomics as a scientific field. Population genomics allows tracing levels of diversity and differentiation across the whole genome by the identification of genome-wide averages and outlier regions for population genetics parameters (*e.g.* nucleotide diversity, fixation index). Therefore, population genomics provides a genome-wide view of the effect of both neutral processes and selective pressures acting on specific genomic regions (Hohenlohe *et al.*, 2012).

Microsatellites are among the most popular markers used by the scientific community. These DNA sequences are composed by tandemly repeated short motifs (Guichoux *et al.*, 2011). These markers are largely abundant in eukaryotic genomes, highly polymorphic due to elevated mutation rates, and codominant with respect to amplification and genotyping. Since their discovery, microsatellites have been used for a variety of purposes and the first population genomics studies have utilised these markers (Payseur *et al.*, 2002). However, there are limitations for analysing large numbers of microsatellites inherent to the laborious genotyping procedures, especially when compared to more recent methods based on single nucleotide polymorphisms.

Amplified fragment length polymorphism (AFLP) is a technique used to collect anonymous genetic variation among the genome and combines a global enzymatic restriction with subsequent amplification of random fragments. This technique provides a large number of loci across the genome that allows inference of genome-wide population divergence averages and detection of outlier loci. The markers obtained by this technique are traditionally dominant, with alleles being scored by the presence or absence of fragments of a particular size. The lower information obtained for each locus (*e.g.* no direct estimation of heterozygosity) is compensated by the large dataset

obtained, which allows a statistical power in population structure analysis similar to that of codominant markers (Campbell *et al.*, 2003).

More recently, new technologies based on DNA microarrays and next generation sequencing (NGS) are promoting significant advances in population genomic studies. In the case of NGS, these methods intend to collect complete genetic information for entire genome sequences of multiple individuals. However, the high cost associated to these techniques is still a drawback for its application in many population genetic and evolutionary studies (Hohenlohe *et al.*, 2012).

Population genetics has also benefited from breakthroughs in bioinformatics and computation science that have allowed the widespread application of simulation methods (*e.g.* Markov-chain Monte Carlo procedures) to statistical inference in biological and evolutionary studies (Hartl & Clark, 2007; Okasha, 2012). These novel statistical approaches can now be applied to large datasets such as those obtained by NGS methods, which has been a major advance for population genomics studies.

The *Culex pipiens* complex

Taxonomy and systematics

According to the classification of Grimaldi & Engel (2005) for higher taxa (*i.e.* above Family) and Knight & Stone 1977 for lower taxa, the systematic position of the genus *Culex* is:

Kingdom	Animalia
Phylum	Arthropoda
Subphylum	Mandibulata
Superclass	Panhexapoda
Epiclass	Hexapoda
Class	Insecta (Ectognatha)
Subclass	Dicondylia
Superorder	Panorpida
Order	Diptera
Suborder	Culicomorpha
Family	Culicidae
Subfamily	Culicinae
Tribe	Culicini
Genus	*Culex*

The genus *Culex* is divided into 26 subgenus and includes 768 formally recognized species. The species of the *Cx. pipiens* complex belong to the subgenus *Culex* (Harbach, 2011). The taxonomy of this complex has been under debate since the 1950s (Vinogradova, 2000). Consequently, the number of members and their classification varies according to the authors.

According to the mosquito systematic reference book *"A Catalog of the Mosquitoes of the World"* (Knight & Stone, 1977; Knight, 1978; Ward, 1984, 1992) the six entities of the complex include three species without intraspecific stratification: *Culex quinquefasciatus, Culex australicus* Dobrotworsky & Drummond 1953, and *Culex globocoxitus* Dobrotworsky 1953; and one species, *Culex pipiens s.s.*, which is subdivided into three intraspecific forms, denoted pallens (Coquillett 1898), molestus (Forskål 1775), and pipiens (Knight & Stone, 1977; Knight, 1978; Ward, 1984, 1992).

The status of pallens as an intraspecific form has been abandoned by most authors, who have upgraded this entity to a subspecies of *Cx. pipiens s.s.* (*Culex pipiens pallens*). However, the main debate surrounding *Cx. pipiens pallens* is its existence as an evolutionary entity. Previous studies regarding the male genitalia, length of the dorsal and ventral arms of the phallosome (DV/D ratio), classified the populations of *Cx. pipiens pallens* as hybrid populations between *Cx. pipiens s.s.* and *Cx. quinquefasciatus* (Bekku, 1956; Laven, 1967; Cornel *et al.,* 2003). This idea has been abandoned with molecular studies suggesting the existence of *Cx. pipiens pallens* as a single entity (Fonseca *et al.,* 2009).

The status of *Cx. quinquefasciatus* as a species has been questioned by authors that consider this taxon subspecies of *Culex pipiens* s.s (*Culex pipiens quinquefasciatus*) (Kothera *et al.,* 2009; Diaz-Badillo *et al.,* 2011; Atyame *et al.,* 2011). This alternative classification is based on the incomplete isolation between *Cx. quinquefasciatus* and *Culex pipiens s.s.* in the contact zone of North America (Barr, 1957; Kothera *et al.,* 2009). Hybridisation between these entities is not exclusive to North America as it has also been described in Argentina and in islands near the African continent (Urbanelli *et al.,* 1995; Humeres *et al.,* 1998; Alves *et al.,* 2010). However, this scenario of incomplete isolation has not been found in all sympatric populations of these taxa (*e.g.* East Africa) (Cornel *et al.,* 2003).

Culex pervigilans Von Bergroth 1889 (New Zealand), *Culex torrentium* Martini 1925 (Europe), and *Culex vagans* Wiedemann 1828 (central and eastern Asia) have been originally included in the complex based on their morphological similarity (classical sibling species) with the other complex species (Vinogradova, 2000). However, other studies, including phylogenetic analyses, have classified these species in a different subgroup (Trifilatus subgroup) (Miller *et al.,* 1996; Weitzel *et al.,* 2009; Harbach, 2011). Other *Culex* species outside the complex have shown morphological similarity with the complex members but the scientific community does not include them in the complex (*e.g. Culex restuans* Theobald 1901, *Culex nigripalpus* Theobald 1901 and *Culex salinarius* Coquillett 1904). The females of these species are considered to be undistinguishable from the *Cx. pipiens* complex and several molecular approaches have been developed in order to avoid misidentification (Farajollahi *et al.*, 2011). Therefore, at the present date, the *Cx. pipiens* complex is considered to include *Cx. australicus, Cx. globocoxitus, Cx. pipiens s.s.,* and *Cx. quinquefasciatus*.

The European intraspecific forms: pipiens and molestus

The taxonomic status of molestus is probably the most controversial of the *Cx. pipiens* complex entities. The elevation of molestus to species (*Culex molestus*) in the 1970s was based in pre-mating isolation mechanisms found in crossing experiments between molestus and *Cx. quinquefasciatus* (Miles, 1977, 1978). However, this claim was not supported by molecular studies performed by Miles & Paterson (1979) which showed a higher similarity between molestus and *Culex pipiens s.s.* than with *Cx. quinquefasciatus.* The return of molestus to an infraspecific classification level was promoted by the work of Harbach *et al.* (1984) describing a neotype for this entity with high genetic and morphological similarity to *Cx. pipiens s.s.* but displaying differences in behaviour and physiology. This intraspecific status launched the debate about the evolutionary origin of molestus populations that divided authors between two hypotheses: 1) the molestus form derives from the pipiens form through multiple local adaptations to anthropogenic underground habitats (Vinogradova, 2000), 2) the molestus form is an evolutionarily independent entity and the colonization of northern underground habitats was made by molestus populations from southern latitudes (Fonseca *et al.,* 2004). The last hypothesis has been reinforced by a common ancestry

among geographically distinct molestus populations determined by microsatellite based analyses (Fonseca *et al.*, 2004).

Ecology and Behaviour

The distribution of the members *Cx. pipiens* complex varies according to the geographic regions and reflects bioecological features of the taxa. *Culex australicus* and *Cx. globocoxitus* are endemic to Australia while *Cx. quinquefasciatus* and *Cx. pipiens s.s.* show ubiquitous distribution in tropical and temperate regions, respectively. Within *Cx. pipiens s.s.*, the pallens form is restricted to Southeast Asia while molestus and pipiens forms have a continental distribution in North America, Europe and most of the Asian continent (Table 1; Dobrotworsky, 1953; Vinogradova, 2000; Cornel *et al.*, 2003).

Table 1. Distribution and ecology/behaviour in the *Culex pipiens* complex.

Classification		**Distribution**	**Physiology/Ecology**			**Male genitalia**
Species	**form**		**ED**	**MB**	**LC**	
Culex australicus		Australia	An	Er	He	DV/D ratio 0.2-0.4
Culex globocoxitus		South Australia	An	St	Ho	Enlarged coxit
Culex pipiens s.s.	pallens form	Southeast Asia	An	Er	He	DV/D ratio 0.2-0.4
	pipiens form	All temperate regions except Australia (Holarctic)	An	Er	He	DV/D ratio <0.2 Dorsal arm blunt ended
	molestus form	All temperate regions	Au	St	Ho	
Culex quinquefasciatus		Cosmotropical	An	St	Ho	DV/D ratio >0.4 Dorsal arm sharp tipped

ED: egg development; MB: mating behaviour; LC: life cycle; An: anautogenous (requires a bloodmeal to develop eggs); Au: autogenous (egg development without a bloodmeal); Er: Eurigamous (requires open spaces to mate); St: stenogamous (able to mate in confined spaces); He: heterodynamic (performs diapause); Ho: homodynamic (unable to perform diapause). (Dobrotworsky, 1953; Vinogradova, 2000; Cornel *et al.*, 2003)

Most ecological studies have focused mainly on the two ubiquitous species of the complex. These studies were particularly important to distinguish pipiens and molestus forms since no significant morphological differences have been described.

Autogeny (*i.e.* the capacity of laying eggs without a blood meal) has been considered one of the main ecological differences between the two forms. The molestus form is the only entity in the complex considered to be autogenous as females of the other entities require a blood meal to lay eggs (Table 1). Differences in mating behaviour (stenogamy versus eurygamy) and life cycle (homodynamy versus heterodynamy) have also been found between molestus and pipiens forms. *Culex quinquefasciatus* and *Cx. australicus* share the same characteristics of molestus (stenogamous and homodynamic) while the pallens form and *Cx. globocoxitus* present similar behaviours to those of the pipiens form (Table 1; Dobrotworsky & Drummond, 1953; Dobrotworsky, 1953; Harbach *et al.*, 1984, 1985; Vinogradova, 2000).

The species of this complex are generally associated with humans. In this context, *Cx. quinquefasciatus* and *Cx. pipiens s.s.* are commonly referred to as the southern and northern house mosquito, respectively. However, the tolerance to urbanization varies among species. *Culex australicus, Cx. globocoxitus* and *Cx. pipiens* are considered common in countryside rural areas while *Cx. quinquefasciatus* is associated to highly urbanized regions (Dobrotworsky & Drummond, 1953; Dobrotworsky, 1953, Ribeiro *et al.*, 1980; Subra, 1981). The molestus form is an exception within *Cx. pipiens s.s.* since it is found mostly associated to urban areas and restricted to underground habitats such as sewers, subway tunnels and sellers in northern latitudes (Byrne & Nichols, 1999; Vinogradova, 2000; Huang *et al.*, 2008).

Taxa without the capacity of performing diapause (*i.e.* homodynamic) are likely to be less adapted to survive the cold winters of temperate regions. This characteristic may explain the confinement of the global distribution of *Cx. quinquefasciatus* to tropical and subtropical regions, and also of the molestus form to underground habitats at northern latitudes. The capacity of molestus to survive in underground habitats is also favoured by its ability to mate in confined spaces (stenogamy) and lay eggs without requiring a blood meal (autogeny).

The involvement of the entities of this complex in the transmission of arboviruses is dependent, among other factors, on host preferences. This trait varies among members of the complex between ornithophily (*i.e.* preference in biting birds) and mammophily (*i.e.* preference in biting mammals). Four of the entities of the

complex, namely *Cx. australicus*, *Cx. globocoxitus*, pallens and pipiens forms of *Cx. pipiens s.s.*, are associated with ornithophilic behaviour. Preference for biting mammals (and humans in particular) is normally attributed to the molestus form and to *Cx. quinquefasciatus*. However, ornithophilic populations of the latter species have been described in North America (Subra, 1981; Harbach *et al.*, 1984; Savage *et al.*, 2007; Molaei *et al.*, 2010).

Hybridisation between members of the *Cx. pipiens* complex with different host preferences may promote a catholic feeding behaviour, thus increasing the importance of host availability (Fonseca *et al.*, 2004; Balenghien *et al.*, 2011). A broader host range of these populations would increase the potential for bridge vector transmission since variations in the abundance of the traditional hosts (*e.g.* bird migrations) may promote a shift of the feeding preference to non-traditional hosts, as described in USA in the end of the summer (Kilpatrick *et al.*, 2006).

Notes on the Culex pipiens genome

Most mosquito genomes are organized in three pairs of chromosomes (2n=6). Species of the Anophelinae subfamiliy present two pairs of metacentric chromosomes of unequal size and one pair of heteromorphic sex chromosomes while species of the Toxorhynchitinae and Culicinae subfamilies have three pairs of homomorphic metacentric or slightly submetacentric chromosomes without sexual dimorphism. In *Cx. pipiens s.l.*, sex is determined by genes located in chromosome 1 (Vinogradova, 2000; Severson & Black IV, 2005).

The genome of *Cx. quinquefasciatus* was sequenced in 2010 (Arensburger *et al.*, 2010). A large fraction (29%) of the genome assembly was found to be composed by transposable elements and 18,883 protein-coding genes have been described. This number of protein-coding genes is larger than in the other two known mosquito genomes (15,419 genes in *Ae. aegypti* and 12,457 genes in *An. gambiae s.s.*), which may be explained by a substantially high number of expanded gene families observed (Arensburger *et al.*, 2010). Among the expanded families identified, were cytochrome P450s, glutathione-S-transferases and choline/carboxylesterases, involved in detoxification processes associated with insecticide resistance (Reddy *et al.*, 2012). These families may also be important in adaptation to polluted water at urban breeding

sites, characteristic of *Cx. quinquefasciatus*. Expanded gene families also included genes related to ecological and behavioural traits that differ among taxa of the complex, such as juvenile hormone genes and olfactory-receptor genes. Juvenile hormones have been associated with regulation of diapause and egg development, while olfactory-receptors may be involved in mating choice, host-seeking behaviour and host preference (Raikhel & Lea, 1991; Robich *et al.,* 2007; Nozawa & Nei, 2007; Arensburger *et al.,* 2010).

Diseases transmitted by the *Culex pipiens* complex

Females of the *Cx. pipiens* complex have been implicated as vectors of several infections with impact on both human and animal health. The majority of the infectious agents transmitted by the *Cx. pipiens* complex are arboviruses, such as the Japanese encephalitis virus serocomplex (Flaviviridae), Rift valley virus (Bunyaviridae), and viruses from the Togaviridae family (*i.e.* Western equine virus and Sindbis virus) (Solomon, 2004; Hubálek, 2008). The Japanese encephalitis virus serocomplex (JEVs) belongs to the genus Flavivirus. Five members of the serocomplex cause encephalitis in humans in tropical and temperate regions: Japanese encephalitis virus (JEV), Murray Valley encephalitis virus, St. Louis encephalitis virus, Usutu virus and West Nile virus (WNV) that include the Kunjin virus in Australia (Solomon, 2004; Rossi *et al.,* 2010). These viruses have enzootic cycles which alternate between birds and mosquitoes mainly from the genus *Culex.* The accidental transmission to mammals, including humans, often results into a “dead-end” infection.

The disease transmitted by *Cx. pipiens* complex mosquitoes with higher impact on human health is lymphatic filariasis. This parasitic infection affects over 120 million people in 80 countries throughout tropical and subtropical regions (WHO, 2012b). The disease is caused by three nematode species *Wuchereria bancrofti* Cobbold 1877, *Brugia malayi* Brug 1927 and *Brugia timori* Partono *et al.* 1977. These worms invade the lymph nodes and cause improper functioning of the lymphatic system, resulting in fluid collection and swelling (lymphedema). The decreased function of the lymph system by the parasite impairs the immune system of the patient reducing its capacity to respond to other infections. An increase in bacterial infections of the skin and lymph

system may lead to a hardening and thickening of the skin, which together with the underlying lymphedema and increase in volume of the limbs is called elephantiasis.

Finally, members of the *Cx. pipiens* complex have important roles in transmitting several veterinary diseases such as avian malaria and avian pox. These diseases have caused the decline of a number of endemic bird populations in oceanic islands leading to conservation concerns in Hawaii and other biodiversity hotspots (Fonseca *et al.,* 2000; Bataille *et al.,* 2009; Carlson *et al.,* 2011).

West Nile virus

Of the arboviral infections transmitted by mosquitoes, WNV has been the virus with highest impact in temperate regions of North America and Europe. It is the most widely distributed arthropod-borne flavivirus in the world, occurring in all continents except Antarctica (Briton, 2002). Although the burden of WNV on human and animal health may not be as severe as that imposed by dengue or yellow fever flaviviruses, it still is a significant and frequently underestimated infection. Most WNV infections of humans are asymptomatic (*ca.* 80%), while clinical cases translate into an acute self-limited illness (known as West Nile fever) accompanied by non-specific influenza-like symptoms (Zeller & Schuffenecker, 2004). However, severe forms of this disease do occur, usually in elderly or debilitated patients. In less than 1% of cases, the virus gains access to the central nervous system where it replicates, resulting in encephalitis and/or meningitis, acute flaccid-paralysis and respiratory failure. In such cases, those who survive may experience long-lasting or permanent neurological sequelae (Dauphin *et al.,* 2004).

As in the case of the other JEVs viruses, the WNV natural cycle involves enzootic transmission between mosquitoes and avian species. Therefore, transmission of WNV to humans requires previous infection of the mosquito with WNV acquired from a blood meal on an infected bird, while humans act as dead-end hosts as viraemia is not sufficient for these to become infective/reservoir hosts (Figure 1).

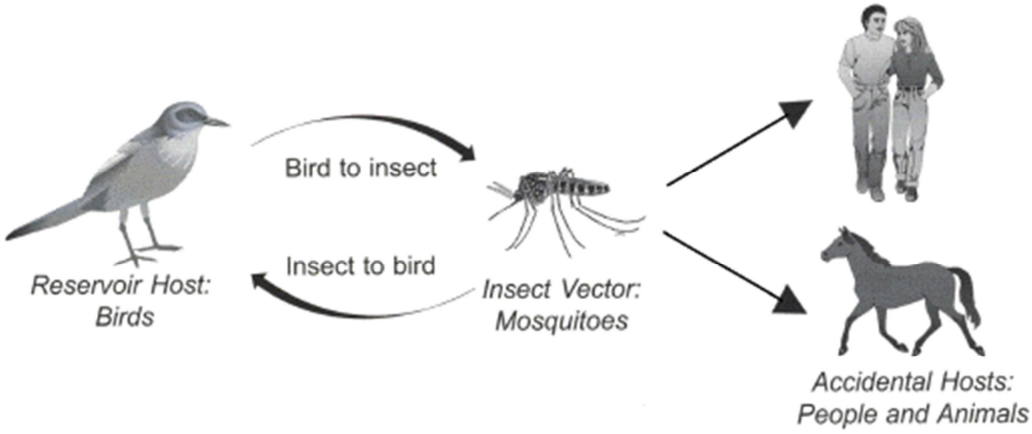

Figure 1. West Nile virus transmission cycle (Butte County Mosquito & Vector Control District 2012).

Species from different genera of mosquitoes (Diptera: Culicidae) have been found to be infected with WNV (Higgs *et al.*, 2004). In Europe, three species are considered the main vectors of WNV: *Culex pipiens s.s.*, *Culex modestus* Ficalbi 1889 and *Coquillettidia richiardii* Ficalbi 1889 (Higgs *et al.*, 2004; Hubálek, 2008). A fourth mosquito species, *Culex perexiguus* Theobald 1903, was recently found to play an important role in the transmission of WNV in the Iberian Peninsula (Muñoz *et al.*, 2012). Interestingly, the first detection of this virus in Portugal was in *Anopheles atroparvus* Van Thiel 1927 (Filipe 1972).

Knowledge of the infection rates in the local avian fauna and contributing ecological factors such as migratory routes and nidification areas/seasons are important to establish the risk of WNV outbreaks (Komar *et al.*, 2003; Rappole & Hubálek, 2003). Bird migration journeys normally follow a north-south axis, linking breeding regions (Arctic and temperate) to non-breeding regions (temperate and tropical). Eight well-established migration routes, called flyways, have been described (Si *et al.*, 2009; Birdlife International, 2010; Boere & Piersma, 2012). The two flyways involving the European continent are the East Atlantic flyway, that includes sites from the east of Canada, central Siberia, western Europe and western Africa (Birdlife International, 2010) (Figure 2); and the Mediterranean/Black Sea flyway (Figure 2), which includes sites from western Siberia, eastern and central Europe, the Middle East and Africa (Birdlife International, 2010).

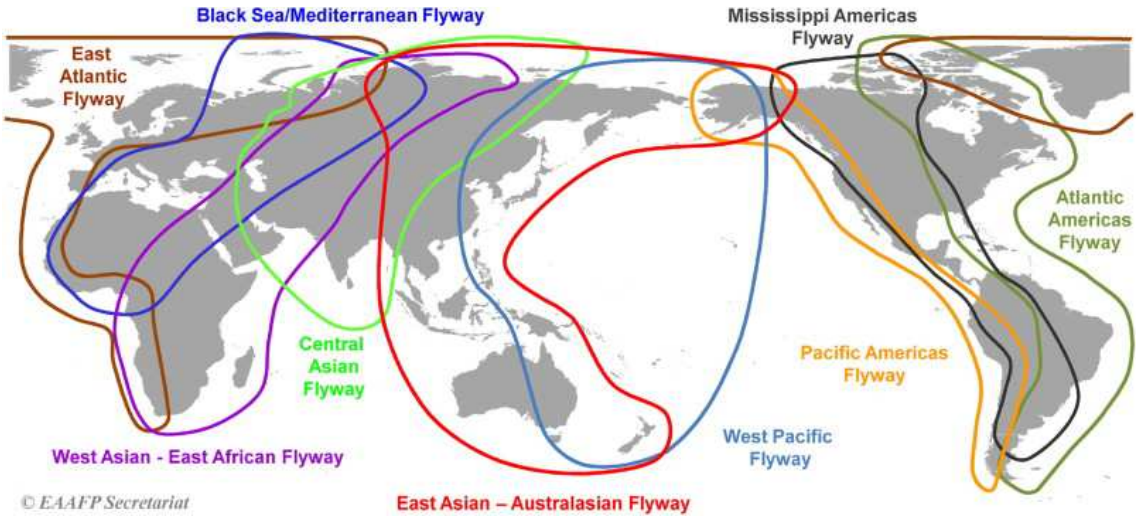

Figure 2. Main waterbird (including shorebirds) flyways of the world (Boere & Piersma, 2012).

Different avian groups also display differences in the capacity to build sufficient viraemia for infecting mosquitoes. Bird species belonging to orders such as Passeriformes, Charadriiformes, Pelecaniformes and some raptors (Falconiformes and Strigiformes) are known to present higher reservoir competences than other bird groups (Komar *et al.*, 2003; Wheeler *et al.*, 2009).

In 1996, WNV became a public health concern in Europe when an outbreak of WNV occurred in Romania, with over 500 cases. This was followed by other smaller outbreaks in Italy, France, Israel, Tunisia, Morocco and Russia (Tsai *et al.*, 1998; Zeller & Schuffenecker, 2004). West Nile virus activity in Europe is usually transient and mostly seasonal (associated with the summer months, which are coincident with the peak periods of mosquito populations), giving rise to occasional outbreaks. This observation supports the hypothesis that WNV is repeatedly introduced into Europe from Africa via migratory birds (May *et al.*, 2011). The WNV episodes in North America have been the most severe in temperate regions, with ~360,000 WNV human cases reported in the last decade, of which 1,308 were fatal (Kilpatrick, 2011).

In Portugal, the incidence of WNV infections in humans remains relatively unknown. Serological surveys of WNV-specific neutralizing antibodies carried out from the 1960s in humans and horses suggested a low transmission level of the virus (Filipe *et al.*, 1972; Filipe & Andrade, 1990; Formosinho *et al.*, 2006; Barros *et al.*, 2011).

Until the present day, West Nile fever has only been confirmed in two linked cases. These were two Irish bird-watching tourists that visited Ria Formosa, a natural reserve and bird sanctuary in Algarve, southern Portugal (Connell *et al.*, 2004). The virus was subsequently isolated from *Cx. pipiens s.s.* and *Cx. univittatus s.l.* mosquitoes collected in the putative infection area of the former (Esteves *et al.*, 2005).

Thesis outline and objectives

The central objective of this thesis was to investigate the evolutionary and genetic relations among different taxa of the *Cx. pipiens* complex and to assess how these relations may impact the transmission of vector-borne diseases sustained by these vectors, the WNV in particular. The major research interests focused on accurately determining levels of genetic differentiation, rates of hybridisation and to detect the occurrence of genomic regions under divergent selection.

The thesis is divided into seven chapters. The current chapter, General Introduction, intends to provide the reader with the background information required to appreciate the research questions addressed and to critically interpret the results obtained and conclusions drawn. Therefore, basic information and general concepts are given for key topics such as the increasing importance of emergent vector-borne diseases, the complexity of vector systems and the importance of bio-ecological, population genetics and evolutionary studies of mosquito vector populations. Following this introduction, five results chapters are presented in the form of scientific articles[1].

Chapter 2 focuses on the interaction between the sibling species *Cx. pipiens s.s.* and *Cx. quinquefasciatus,* in the islands of Macaronesia. This is the first molecular study addressing levels of hybridisation between these two sibling species in the western African biogeographic region, where the presence of the Sahara desert imposes a natural barrier between them. Furthermore, the peculiar history of human colonization of these islands coupled with their biogeography provided an opportunity to address issues of island colonization by insect vector species.

[1] Two published, two submitted, and one in preparation.

Chapter 3 is devoted to the intraspecific heterogeneity within *Cx. pipiens s.s.* in a southern European continental area (Comporta, Portugal). Using a combination of phenotypic and genetic data, the basis for the sympatric occurrence of molestus and pipiens forms in surface habitats is established and the levels of hybridisation and introgression between them are quantified.

The finding of molestus and pipiens sympatric populations in the region of Comporta (Portugal) and the occurrence of important levels of gene flow between these two forms led to subsequent investigations aimed at establishing the role of genetic introgression in key behavioural traits, such as host preferences and biting behaviour, for WNV transmission. These results are described in Chapter 4.

In Chapter 5, a different methodological approach, AFLP analysis, was implemented to genetically characterise *Cx. pipiens s.s.* samples from different geographic origins from Europe and North America. This is the first genomic scan that specifically addresses levels of genome divergence between molestus and pipiens forms. This is, therefore, to the best of our knowledge, a first attempt to determine genomic regions under divergent selection that may contain genes putatively associated with reproductive isolation between these two forms.

In the last results chapter (Chapter 6) the microsatellite-based approach developed in Chapter 3, which allowed discriminating molestus and pipiens genetic backgrounds, was used to characterise *Cx. pipiens s.s.* populations collected in another southern European continental region (Greece), in the midst of an active WNV transmission outbreak that occurred in 2010.

The final chapter (chapter 7) aims to integrate the main findings in a discussion about how evolutionary and genetic relations among the different taxa analysed, may affect their role as vectors of disease. In addition, it provides future perspectives for further research to clarify the evolution of the members of the *Cx. pipiens* complex and its impact on human health.

References

Almeida, A.P.G., Gonçalves, Y.M., Novo, M.T., Sousa, C.A., Melim, M. & Grácio, A.J.S. (2007) Vector monitoring of *Aedes aegypti* in the autonomous region of Madeira, Portugal. *Eurosurveillance.* 12 (11), E071115.6.

Alves, J., Gomes, B., Rodrigues, R., Silva, J., Arez, A. P., Pinto, J. & Sousa, C. A. (2010) Mosquito fauna on the Cape Verde islands (West Africa): An update on species distribution and a new finding. *Journal of Vector Ecology*. 35 (2), 307–312.

Angelini, R., Finarelli, A.C., Angelini, P., Po, C., Petropulacos, K., Silvi, G., Macini, P., Fortuna, C., Venturi, G., Magurano, F., Fiorentini, C., Marchi, A., Benedetti, E., Bucci, P., Boros, S., Romi, R., Majori, G., Ciufolini, M.G., Nicoletti, L., Rezza, G. & Cassone, A. (2007) Chikungunya in north-eastern Italy: a summing up of the outbreak. *Eurosurveillance*. 12 (11), E071122.2.

Arensburger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E., Camara, F., Campbell, C.L., Campbell, K.S., Casola, C., Castro, M.T., Chandramouliswaran, I., Chapman, S.B., Christley, S., Costas, J., Eisenstadt, E., Feschotte, C., Fraser-Liggett, C., Guigo, R., Haas, B., Hammond, M., Hansson, B.S. Hemingway, J., Hill, S.R., Howarth, C., Ignell, R., Kennedy, R.C., Kodira, C.D., Lobo, N.F., Mao, C., Mayhew, G., Michel, K., Mori, A., Liu, N., Naveira, H., Nene, V., Nguyen, N., Pearson, M.D., Pritham, E.J., Puiu, D., Qi, Y., Ranson, H., Ribeiro, J.M.C., Roberston, H.M., Severson, D.W., Shumway, M., Stanke, M., Strausberg, R.L., Sun, C., Sutton, G., Tu, Z.J., Tubio, J.M.C., Unger, M.F., Vanlandingham, D.L., Vilella, A.J., White, O., White, J.R., Wondji, C.S., Wortman, J., Zdobnov, E.M., Birren, B., Christensen, B.M., Collins, F.H., Cornel, A., Dimopoulos, G., Hannick, L.I., Higgs, S., Lanzaro, G.C., Lawson, D., Lee, N.H., Muskavitch, M.A.T., Raikhel, A.S. & Atkinson, P.W. (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science.* 330 (6000), 86–88.

Atyame, C. M., Pasteur, N., Dumas, E., Tortosa, P., Tantely, M. L., Pocquet, N., Licciardi, S., Bheecarry, A., Zumbo, B., Weill, M. & Duron, O. (2011) Cytoplasmic incompatibility as a means of controlling *Culex pipiens quinquefasciatus* mosquito in the islands of the south-western Indian Ocean. *PLoS Neglected Tropical Diseases*. 5 (12), e1440.

Balenghien, T., Fouque, F., Sabatier, P. & Bicout, D.J. (2011) Theoretical formulation for mosquito host-feeding patterns: Application to a West Nile virus focus of southern France. *Journal of Medical Entomology*. 48 (5), 1076–1090.

Barr, A.R. (1957) The distribution of *Culex p. pipiens* and *C. p. quinquefasciatus* in North America. *The American journal of Tropical Medicine and Hygiene*. 6 (1), 153–165.

Barros, S. C., Ramos, F., Fagulha, T., Duarte, M., Henriques, M., Luís, T. & Fevereiro, M. (2011) Serological evidence of West Nile virus circulation in Portugal. *Veterinary Microbiology*. 152 (3-4), 407–410.

Bataille, A., Cunningham, A.A., Cedeño, V., Cruz, M., Eastwood, G., Fonseca, D.M., Causton, C.E., Azuero, R., Loayza, J., Martinez, J.D.C. & Goodman, S.J. (2009) Evidence for regular ongoing introductions of mosquito disease vectors into the Galápagos Islands. *Proceedings. Biological sciences / The Royal Society*. 276 (1674), 3769–3775.

Bates, M. (1940) The nomenclature and taxonomic status of the mosquitoes of the *Anopheles maculipennis* complex. *Annals of the Entomological Society of America*. 33 (2), 343–356.

Bekku, H. (1956) Studies on the *Culex pipiens* group of Japan. *Nagasaki Medical Journal*. 31 (11), 956–966.

Birdlife International (2010) *The flyways concept can help coordinate global efforts to conserve migratory birds*. [Online]. 2010. Available from: http://www.birdlife.org/datazone/sowb/casestudy/20 [Accessed: 23 October 2012].

Boere, G.C. & Piersma, T. (2012) Flyway protection and the predicament of our migrant birds: A critical look at international conservation policies and the Dutch Wadden Sea. *Ocean & Coastal Management*. 68, 157–168.

Brinton, M.A. (2002) The molecular biology of West Nile virus: a new invader of the western hemisphere. *Annual review of Microbiology*. 56, 371–402.

Bruce-Chwatt, L.J. (1988) History of malaria from prehistory to eradication. In: W.H. Wernsdorfer & I. McGregor (eds.). *Malaria. Principles of Malariology Volume 1*. London, Churchill Livinstone Inc. pp. 1–59.

Butte County Mosquito & Vector Control District (2012) *Viruses (West Nile virus, WNV) - How do people get WNV?*. [Online]. 2012. Available from: http://www.bcmvcd.com/viruses.html [Accessed: 23 October 2012].

Byrne, K. & Nichols, R.A. (1999) *Culex pipiens* in London Underground tunnels: differentiation between surface and subterranean populations. *Heredity*. 82 (1), 7–15.

Campbell, D., Duchesne, P. & Bernatchez, L. (2003) AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Molecular Ecology*. 12 (7), 1979–1991.

Carlson, J.S., Martínez-Gómez, J.E., Cornel, A., Loiseau, C. & Sehgal, R.N.M. (2011) Implications of *Plasmodium* parasite infected mosquitoes on an insular avifauna: the case of Socorro Island, México. *Journal of Vector Ecology*. 36 (1), 213–220.

Connell, J., McKeown, P., Garvey, P., Cotter, S., Conway, A., Darina O'Flanagan, P. O'Herlihy, B., Morgan, D., Nicoll, A. & Lloyd, G. (2004) Two linked cases of West Nile virus (WNV) acquired by Irish tourists in the Algarve, Portugal. *Eurosurveillance*. 8 (32), pii=2517.

Cornel, A.J., Mcabee, R.D., Rasgon, J., Stanich, M.A., Scott, T.W. & Coetzee, M. (2003) Differences in extent of genetic introgression between sympatric *Culex pipiens* and *Culex quinquefasciatus* (Diptera: Culicidae) in California and South Africa. *Journal of Medical Entomology*. 40 (1), 36–51.

Dauphin, G., Zientara, S., Zeller, H. & Murgue, B. (2004) West Nile: worldwide current situation in animals and humans. *Comparative Immunology, Microbiology and Infectious Diseases*. 27 (5), 343–355.

Diaz-Badillo, A., Bolling, B.G., Perez-Ramirez, G., Moore, C.G., Martinez-Munoz, J.P., Padilla-Viveros, A.A., Camacho-Nuez, M., Diaz-Perez, A., Beaty, B.J. & Munoz, M.L. (2011) The distribution of potential West Nile virus vectors, *Culex pipiens pipiens* and *Culex pipiens quinquefasciatus* (Diptera: Culicidae), in Mexico City. *Parasites & Vectors*. 4 (1), 70.

Dobkotworsky, N.V. (1953) The *Culex pipiens* group in south-eastern Australia I. *Proceedings of The Linnean Society of New South Wales*. 77 (3-4), 357–360.

Dobrotworsky, N.V. & Drummond, F.H. (1953) The *Culex pipiens* group in south-eastern Australia II. *Proceedings of The Linnean Society of New South Wales*. 77 (5-6), 131–146.

Egan, S.P., Nosil, P., Funk, D.J. & Pfennig, D. (2008) Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* Leaf Beetles. *Evolution*. 62 (5), 1162–1181.

Eldridge, B.F. (2004) Mosquitoes, the Culicidae. In: W.C. Marquardt (ed.). *Biology of Disease Vectors.* 2nd ed. Elsevier Academic Press. pp. 95–112.

Esteves, A., Almeida, A.P.G., Galão, R.P., Parreira, R.,Piedade, J., Rodrigues, J. C., Sousa, C.A. & Novo, M.T. (2005) West Nile virus in southern Portugal, 2004. *Vector-Borne and Zoonotic Diseases*. 5 (4), 410–413.

Direcção Geral da Saúde (DGS) (2012) *Dengue na Região Autónoma da Madeira - atualização 05/12/2012.* [Online] Available from: http://www.dgs.pt/?cn=683368347243AAAAAAAAAAAA [Accessed: 10 December 2012].

Fantini, B. (1994) Anophelism without malaria: an ecological and epidemiological puzzle. *Parassitologia*. 36 (1-2), 83–106.

Farajollahi, A., Fonseca, D.M., Kramer, L.D. & Kilpatrick, A.M. (2011) “Bird biting” mosquitoes and human disease: a review of the role of *Culex pipiens* complex mosquitoes in epidemiology. *Infection, Genetics and Evolution.* 11 (7), 1577–1585.

Filipe, A.R. (1972) Isolation in Portugal of West Nile virus from *Anopheles maculipennis* mosquitoes. *Acta Virologica*. 16 (4), 361.

Filipe, A.R. & de Andrade, H.R. (1990) Arboviruses in the Iberian Peninsula. *Acta Virologica*. 34 (6), 582–591.

Fonseca, D.M., Lapointe, D.A. & Fleischer, R.C. (2000) Bottlenecks and multiple introductions: population genetics of the vector of avian malaria in Hawaii. *Molecular Ecology*. 9 (11), 1803–1814.

Fonseca, D.M., Keyghobadi, N., Malcolm, C.A., Mehmet, C., Schaffner, F., Mogi, M., Fleischer, R.C. & Wilkerson, R.C. (2004) Emerging vectors in the *Culex pipiens* complex. *Science*. 303 (5663), 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.

Formosinho, P., Santos-Silva, M.M., Santos, A., Melo, P., Encarnação, V., Santos, N., Nunes, T., Agrícola, R. & Portas, M. (2006). O vírus West Nile em Portugal – estudos de vigilância epidemiológica. *Revista Portuguesa de Ciências Veterinárias*. 101 (557-558), 61–68.

Gage, K.L. (2004) Fleas, the Siphonaptera. In: W.C. Marquardt (ed.). *Biology of Disease Vectors.* 2nd ed. Elsevier Academic Press. pp. 77 – 92.

Gilles, H.M. (1993) The malaria parasites. In: H.M. Gilles & D.A. Warrel (eds.). *Bruce-Chwatt's Essential Malariology*. 3rd ed. Kent, Edward Arnold. pp. 124–163.

Goldberg, D.E., Siliciano, R.F. & Jacobs, W.R. (2012) Outwitting evolution: fighting drug-resistant TB, malaria, and HIV. *Cell*. 148 (6), 1271–1283.

Grimaldi, D. & Engel, M.S. (2005) *Evolution of the Insects (Cambridge Evolution Series)*. Cambridge University Press.

Gubler, D.J. & Clark, G.G. (1995) Dengue/dengue hemorrhagic fever: the emergence of a global health problem. *Emerging Infectious Diseases*. 1 (2), 55–57.

Gubler, D.J. (2011) Dengue, urbanization and globalization: The unholy trinity of the 21st Century. *Tropical Medicine and Health*. 39 (4 Suppl), 3–11.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F. & Petit, R.J. (2011) Current trends in microsatellite genotyping. *Molecular Ecology resources*. 11 (4), 591–611.

Hayes, C.G. (2001) West Nile virus: Uganda, 1937, to New York City, 1999. *Annals of the New York Academy of Sciences*. 951 (1), 25–37.

Harbach, R.E., Harrison, B.A. & Gad, A.M. (1984) *Culex (Culex) molestus* Forskål (Diptera, Culicidae) - neotype designation, description, variation, and taxonomic status. *Proceedings of the Entomological Society of Washington*. 86 (3), 521–542.

Harbach, R.E., Dahl, C. & White, G.B. (1985) *Culex (Culex) pipiens* Linnaeus (Diptera, Culicidae) - Concepts, type designations, and description. *Proceedings of the Entomological Society of Washington*. 87 (1), 1–24.

Harbach, R.E. (2011) Classification within the cosmopolitan genus *Culex* (Diptera: Culicidae): the foundation for molecular systematics and phylogenetic research. *Acta Tropica*. 120 (1-2), 1–14.

Hartl, D.L. & Clark, A.G. (2007) *Principles of Population Genetics.* 4th ed. Sunderland, MA, Sinauer Associates, Inc.

Higgs, S. & Beaty, B.J. (2004) Natural Cycles of Vector-Borne Pathogens. In: W.C. Marquardt (ed.). *Biology of Disease Vectors*. 2nd ed. Elsevier Academic Press. pp. 167–186.

Higgs, S., Snow, K. & Gould, E.A. (2004) The potential for West Nile virus to establish outside of its natural range: a consideration of potential mosquito vectors in the United Kingdom. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 98 (2), 82–87.

Hohenlohe, P.A., Catchen, J. & Cresko, W.A. (2012) Population genomic analysis of model and nonmodel organisms using sequenced RAD tags. *Methods in molecular biology (Clifton, N.J.)*. pp. 235–260.

Huang, S., Molaei, G. & Andreadis, T.G. (2008) Genetic insights into the population structure of *Culex pipiens* (Diptera: Culicidae) in the northeastern United States by using microsatellite analysis. *American Journal of Tropical Medicine and Hygiene*. 79 (4), 518–527.

Hubálek, Z. (2008) Mosquito-borne viruses in Europe. *Parasitology Research*. 10 (1), 329–343.

Humeres, S.G., Almirón, W.R., Sabattini, M.S. & Gardenal, C.N. (1998) Estimation of genetic divergence and gene flow between *Culex pipiens* and *Culex quinquefasciatus* (Diptera: Culicidae) in Argentina. *Memórias do Instituto Oswaldo Cruz*. 93 (1), 57–62.

Kamali, M., Xia, A., Tu, Z. & Sharakhov, I.V. (2012) A New chromosomal phylogeny supports the repeated origin of vectorial capacity in Malaria mosquitoes of the *Anopheles gambiae* complex. *PLoS Pathogens*. 8 (10), e1002960.

Khamsiriwatchara, A., Sudathip, P., Sawang, S., Vijakadge, S., Potithavoranan, T., Sangvichean, A., Satimai, W., Delacollette, C., Singhasivanon, P., Lawpoolsri, S. & Kaewkungwal, J. (2012) Artemisinin resistance containment project in Thailand. (I): Implementation of electronic-based malaria information system for early case detection and individual case management in provinces along the Thai-Cambodian border. *Malaria Journal*. 11 (1), 247.

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), e82.

Kilpatrick, A.M. (2011) Globalization, Land Use, and the Invasion of West Nile virus. *Science*. 334 (6054), 323–327.

Knight, K.L. & Stone, A. (1977) *A catalog of the mosquitoes of the world (Diptera: Culicidae)*. 2nd ed. College Park, Maryland, USA, Thomas Say Foundation, Entomological Society of America.

Knight, K.L. (1978) *Supplement to A Catalog of the mosquitoes of the World*. College Park, Maryland, Thomas Say Foundation, Entomological Society of America.

Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R. & Bunning, M. (2003) Experimental infection of north American birds with the New York 1999 strain of West Nile virus. *Emerging Infectious Diseases*. 9 (3), 311–322.

Kothera, L., Zimmerman, E.M., Richards, C.M. & Savage, H.M. (2009) Microsatellite characterization of subspecies and their hybrids in *Culex pipiens* complex (Diptera: Culicidae) mosquitoes along a North-South transect in the central United States. *Journal of Medical Entomology*. 46 (2), 236–248.

Laven, H. (1967) Speciation and evolution in *Culex pipiens*. In: J.W. Wright & R. Pal (eds.). *Genetics of Insect Vectors of Disease*. Amsterdam, Elsevier Academic Press. pp. 251–275.

Lounibos, L.P. (2002) Invasions by insect vectors of human disease. *Annual Review of Entomology*. 47 (1), 233–266.

Machado, C.A., Kliman, R.M., Markert, J.A. & Hey, J. (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*. 19 (4), 472–488.

Mallet, J. (2005) Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*. 20 (5), 229–237.

May, F.J., Davis, C.T., Tesh, R.B. & Barrett, A.D.T. (2011) Phylogeography of West Nile virus: from the cradle of evolution in Africa to Eurasia, Australia, and the Americas. *Journal of Virology*. 85 (6), 2964–2974.

Mayr, E. (1999) *Systematics and the Origin of Species from the Viewpoint of a Zoologist*. Harvard University Press.

McMichael, A.J., Woodruff, R.E. & Hales, S. (2006) Climate change and human health: present and future risks. *Lancet*. 367 (9513), 859–869.

Medlock, J. M., Hansford, K. M., Schaffner, F., Versteirt, V., Hendrickx, G., Zeller, H. & Van Bortel, W. (2012) A review of the invasive mosquitoes in Europe: ecology, public health risks, and control options. *Vector Borne and Zoonotic Diseases.* 12 (6), 435–447.

Miles, S.J. (1977) Laboratory evidence for mate recognition behaviour in a member of the *Culex pipiens* complex (Diptera: Culicidae). *Australian Journal of Zoology*. 25 (3), 491–498.

Miles, S.J. (1978) Assortative mating between *Culex fatigans* and *C. molestus* (Diptera, Culicidae) under simulated field conditions. *Australian Journal of Entomology*. 16 (4), 389–392.

Miles, S.J. & Paterson, H.E. (1979) Protein variation and systematics in the *Culex pipiens* group of species. *Mosquito Systematics*. 11 (3), 187–202.

Miller, B.R., Crabtree, M.B. & Savage, H.M. (1996) Phylogeny of fourteen *Culex* mosquito species, including the *Culex pipiens* complex, inferred from the internal transcribed spacers of ribosomal DNA. *Insect Molecular Biology*. 5 (2), 93–107.

Molaei, G., Cummings, R.F., Su, T., Armstrong, P.M., Williams, G.A., Cheng, M.-L., Webb, J.P. & Andreadis, T.G. (2010) Vector-host interactions governing epidemiology of West Nile virus in southern California. *The American Journal of Tropical Medicine and Hygiene*. 83 (6), 1269–1282.

Muñoz, J., Ruiz, S., Soriguer, R., Alcaide, M., Viana, D. S., Roiz, D., Vázquez, A. & Figuerola, J. (2012) Feeding patterns of potential West Nile virus vectors in south-west Spain. *PloS ONE*. 7 (6), e39549.

Moro, M.L., Gagliotti, C., Silvi, G., Angelini, R., Sambri, V., Rezza, G., Massimiliani, E., Mattivi, A., Grilli, E., Finarelli, A.C., Spataro, N., Pierro, A.M., Seyler, T. & Macini, P. (2010) Chikungunya virus in north-eastern Italy: a seroprevalence survey. *The American Journal of Tropical Medicine and Hygiene*. 82 (3), 508–511.

Nosil, P., Harmon, L.J. & Seehausen, O. (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*. 24 (3), 145–156.

Nozawa, M. & Nei, M. (2007) Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. *Proceedings of the National Academy of Sciences of the United States of America*. 104 (17), 7122–7127.

Okasha, S. (2012) Population Genetics. In: E.N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy*. Fall 2012. [Online]. Available from: http://plato.stanford.edu/archives/fall2012/entries/population-genetics/ [Accessed: 5 November 2012].

Payseur, B.A., Cutter, A.D. & Nachman, M.W. (2002) Searching for evidence of positive selection in the Human genome using patterns of microsatellite variability. *Molecular Biology and Evolution*. 19 (7), 1143–1153.

Poletti, P., Messeri, G., Ajelli, M., Vallorani, R., Rizzo, C. & Merler, S. (2011) Transmission potential of chikungunya virus and control measures: the case of Italy. *PloS ONE*. 6 (5), e18860.

Post, R.J., Mustapha, M. & Krueger, A. (2007) Taxonomy and inventory of the cytospecies and cytotypes of the *Simulium damnosum* complex (Diptera: Simuliidae) in relation to onchocerciasis. *Tropical Medicine & International Health.* 12 (11), 1342–1353.

Post, R.J., Onyenwe, E., Somiari, S.A.E., Mafuyai, H.B., Crainey, J.L. & Ubachukwu, P.O. (2011) A guide to the *Simulium damnosum* complex (Diptera: Simuliidae) in Nigeria, with a cytotaxonomic key for the identification of the sibling species. *Annals of Tropical Medicine and Parasitology*. 105 (4), 277–297.

Raikhel, A.S. & Lea, A.O. (1991) Control of follicular epithelium development and vitelline envelope formation in the mosquito; role of juvenile hormone and 20-hydroxyecdysone. *Tissue and Cell*. 23 (4), 577–591.

Ranson, H., N'guessan, R., Lines, J., Moiroux, N., Nkuni, Z. & Corbel, V. (2011) Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? *Trends in Parasitology*. 27 (2), 91–98.

Rappole, J.H. & Hubálek, Z. (2003) Migratory birds and West Nile virus. *Journal of Applied Microbiology*. 94, 47–58.

Reddy, B.P.N., Labbé, P. & Corbel, V. (2012) *Culex* genome is not just another genome for comparative genomics. *Parasites & Vectors*. 5 (1), 63.

Ribeiro, H., Ramos, H.C., Capela, R.A. & Pires, C.A. (1980) *Os Mosquitos de Cabo Verde (Diptera: Culicidae) – Sistemática, distribuição, bioecologia e importância Médica*. Lisboa, Junta de Investigações Científica do Ultramar.

Rigau-Pérez, J.G., Clark, G.G., Gubler, D.J., Reiter, P., Sanders, E.J. & Vorndam, A.V. (1998) Dengue and dengue hemorrhagic fever. *Lancet*. 352 (9132), 971–977.

Robich, R.M., Rinehart, J.P., Kitchen, L.J. & Denlinger, D.L. (2007) Diapause-specific gene expression in the northern house mosquito, *Culex pipiens* L., identified by suppressive subtractive hybridization. *Journal of Insect Physiology*. 53 (3), 235–245.

Rossi, *S.L.*, Ross, T.M. & Evans, J.D. (2010) West Nile virus. *Clinics in Laboratory Medicine*. 30 (1), 47–65.

Santolamazza, F., Calzetta, M., Etang, J., Barrese, E., Dia, I., Caccone, A., Donnelly, M.J., Petrarca, V., Simard, F., Pinto, J. & della Torre, A. (2008) Distribution of knock-down resistance mutations in *Anopheles gambiae* molecular forms in west and west-central Africa. *Malaria Journal*. 7 (1), 74.

Savage, H.M., Aggarwal, D., Apperson, C.S., Katholi, C.R., Gordon, E., Hassan, H.K., Anderson, M., Charnetzky, D., McMillen, L., Unnasch, E.A. & Unnasch, T.R. (2007) Host choice and West Nile virus infection rates in blood-fed Mosquitoes, including members of the *Culex pipiens* complex, from Memphis and Shelby County, Tennessee, 2002–2003. *Vector-Borne and Zoonotic Diseases*. 7 (3), 365–386.

Service, M.W. (1993a) Mosquitoes (Culicidae). In: R.P. Lane & R.W. Crosskey (eds.). *Medical Insects and Arachnids*. 1st ed. London, Chapman & Hall. pp. 120–240.

Service, M.W. (1993b) The *Anopheles* vector. In: H.M. Gilles & D.A. Warrel (eds.). *Bruce-Chwatt's Essential Malariology*. 3rd ed. Kent, Edward Arnold. pp. 97–123.

Severson, D.W. & Black IV, W.C. (2005) Genome Evolution in Mosquitoes. In: W.C. Marquardt (ed.). *Biology of Disease Vectors*. 2nd ed. Elsevier Academic Press. pp. 449–464.

Si, Y., Skidmore, A. K., Wang, T., Boer, W. F. de, Debba, P., Toxopeus, A. G., Li, L. & Prins, H.H.T. (2009) Spatio-temporal dynamics of global H5N1 outbreaks match bird migration patterns. *Geospatial Health*. 4 (1), 65–78.

Smith, 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 Hygiene*. 70 (4), 339–345.

Solomon, T. (2004) Flavivirus encephalitis. *The New England Journal of Medicine*. 351 (4), 370–378.

Subra, R. (1981) Biology and control of *Culex pipiens quinquefasciatus* Say, 1823 (Diptera, Culicidae) with special reference to Africa. *International Journal of Tropical Insect Science*. 1 (4), 319–338.

Sunnucks, P. (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution*. 15 (5), 199–203.

Tan, W.-L., Wang, Z.-M., Li, C.-X., Chu, H.-L., Xu, Y., Dong, Y.-D., Wang, Z., Chen, D.-Y. Liu, H., Liu, D.-P., Liu, N., Sun, J. & Zhao, T. (2012) First report on co-occurrence knockdown resistance mutations and susceptibility to beta-cypermethrin in *Anopheles sinensis* from Jiangsu Province, China. *PloS ONE*. 7 (1), e29242.

della Torre, A., Fanello, C., Akogbeto, M., Dossou-yovo, J., Favia, G., Petrarca, V. & Coluzzi, M. (2001) Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa. *Insect Molecular Biology*. 10 (1), 9–18.

Tsai, T., Popovici, F., Cernescu, C., Campbell, G. & Nedelcu, N. (1998) West Nile encephalitis epidemic in southeastern Romania. *The Lancet*. 352 (9130), 767–771.

Turner, T.L., Hahn, M.W. & Nuzhdin, S.V. (2005) Genomic islands of speciation in *Anopheles gambiae. PLoS Biology*. 3 (9), e285.

Urbanelli, S., Silvestrini, F., Sabatinelli, G., Raveloarifera, F., Petrarca, V. & Bullini, L. (1995) Characterization of the *Culex pipiens* complex (Diptera: Culicidae) in Madagascar. *Journal of Medical Entomology*. 32 (6), 778-786.

Vasilakis, N., Cardosa, J., Hanley, K.A., Holmes, E.C. & Weaver, S.C. (2011) Fever from the forest: prospects for the continued emergence of sylvatic dengue virus and its impact on public health. *Nature reviews Microbiology*. 9 (7), 532–541.

Vinayak, S., Alam, M.T., Mixson-Hayden, T., McCollum, A.M., Sem, R., Shah, N.K., Lim, P., Muth, S., Rogers, W.O., Fandeur, T., Barnwell, J.W., Escalante, A.A., Wongsrichanalai, C., Ariey, F., Meshnick, S.R. & Udhayakumar, V. (2010) Origin and evolution of sulfadoxine resistant *Plasmodium falciparum*. *PLoS Pathogens*. 6 (3), e1000830.

Vinogradova, E.B. (2000) *Culex pipiens pipiens mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control*. S.I. Golovatch (ed.). Sofia, Pensoft.

Vora, N. (2008) Impact of anthropogenic environmental alterations on vector-borne diseases. *Medscape Journal of Medicine*. 10 (10), 238.

Walløe, L. (2008) Medieval and modern bubonic plague: some clinical continuities. *Medical History. Supplement*. 27, 59–73.

Ward, R.A. (1984) Second supplement to "A Catalog of the mosquitoes of the World". *Mosquito Systematics*. 16 (3), 227–270.

Ward, R.A. (1992) Third supplement to "A Catalog of the mosquitoes of the World". *Mosquito Systematics*. 24 (3), 177–230.

Weaver, S.C. & Barrett, A.D.T. (2004) Transmission cycles, host range, evolution and emergence of arboviral disease. *Nature reviews Microbiology*. 2 (10), 789–801.

Weetman, D., Wilding, C.S., Steen, K., Pinto, J. & Donnelly, M.J. (2012) Gene flow dependent genomic divergence between *Anopheles gambiae* M and S forms. *Molecular Biology and Evolution*. 29 (1), 279–291.

Weitzel, T., Collado, A., Jöst, A., Pietsch, K., Storch, V. & Becker, N. (2009) Genetic differentiation of populations within the *Culex pipiens* complex and phylogeny of related species. *Journal of the American Mosquito Control Association*. 25 (1), 6–17.

Wheeler, S. S., Barker, C. M., Fang, Y., Veronica Armijos, M., Carroll, B. D., Husted, S., Johnson, W. O. & Reisen, W.K. (2009) Differential impact of West Nile virus on California birds. *The Condor*. 111 (1), 1–20.

White, B. J., Lawniczak, M. K. N., Cheng, C., Coulibaly, M. B., Wilson, M. D., Sagnon, N., Costantini, C., Simard, F., Christophides, G.K. & Besansky, N.J. (2011) Adaptive divergence between incipient species of *Anopheles gambiae* increases resistance to *Plasmodium*. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (1), 244–249.

World Health Organization (WHO) - Regional Office for Europe (2006) *Regional strategy: From Malaria control to elimination in the WHO European region, 2006-2015*. [Online]. p.50. Available from: http://www.euro.who.int/Document/E88840.pdf. [Accessed: 8 December 2012].

World Health Organization (WHO) (2011) *World Malaria Report 2011*. [Online]. p.248. Available from: http://www.who.int/malaria/world_malaria_report_2011/en/ [Accessed: 29 November 2012].

World Health Organization (WHO) (2012a) *WHO - Dengue and severe dengue*. [Online]. 2012. Available from: http://www.who.int/mediacentre/factsheets/fs117/en/index.html [Accessed: 23 October 2012].

World Health Organization (WHO) (2012b) *WHO - Lymphatic filariasis*. [Online]. 2012. Available from: http://www.who.int/lymphatic_filariasis/en/ [Accessed: 23 October 2012].

Zeller, H.G. & Schuffenecker, I. (2004) West Nile virus: an overview of its spread in Europe and the Mediterranean basin in contrast to its spread in the Americas. *European Journal of Clinical Microbiology & Infectious Diseases*. 23 (3), 147–156.

.

Chapter 2.

Hybridisation and population structure of the *Culex pipiens* complex in the islands of Macaronesia

Published as:

Gomes, B., Alves, J., Sousa, C.A., Santa-Ana, M., Vieira, I., Silva, T.L., Almeida, A.P.G., Donnelly, M.J. & Pinto, J. (2012) Hybridization and population structure of the *Culex pipiens* complex in the islands of Macaronesia. *Ecology and Evolution*. 2 (8), 1889–1902.

Abstract

The *Culex pipiens* complex includes two widespread mosquito vector species, *Cx. pipiens* and *Cx. quinquefasciatus*. The distribution of these species varies in latitude, with the former being present in temperate regions and the latter in tropical and subtropical regions. However, their distribution range overlaps in certain areas and interspecific hybridisation has been documented. Genetic introgression between these species may have epidemiological repercussions for West Nile virus (WNV) transmission. Bayesian clustering analysis based on multilocus genotypes of 12 microsatellites was used to determine levels of hybridisation between these two species in Macaronesian islands, the only contact zone described in West Africa. The distribution of the two species reflects both the islands' biogeography and historical aspects of human colonization. Madeira Island displayed a homogenous population of *Cx. pipiens*, whereas Cape Verde showed a more intriguing scenario with extensive hybridisation. In the islands of Brava and Santiago, only *Cx. quinquefasciatus* was found, while in Fogo and Maio high hybrid rates (~40%) between the two species were detected. Within the admixed populations, second-generation hybrids (~50%) were identified suggesting a lack of isolation mechanisms. The observed levels of hybridisation may locally potentiate the transmission to humans of zoonotic arboviruses such as WNV.

Introduction

The biological diversity of islands with recent volcanic origin and high isolation from mainland is a result of the colonizers ability to break the isolation and survive the island's environmental conditions. The highly stochastic nature of colonizing events means that only a very limited number of taxa may be present in each archipelago (Gillespie & Roderick, 2002). For example, in Hawaii, only 15% of the known insect families were observed (Howarth & Ramsay, 1991), and a similar scenario occurs in the Macaronesian region (Juan *et al.*, 2000; Gillespie & Roderick, 2002). This region is formed by four archipelagos of volcanic islands located in the northern hemisphere of the Atlantic Ocean: Azores, Canary Islands, Cape Verde, and Madeira. Isolation and low colonization rates in these islands promote divergence by adaptive radiation,

leading to a higher proportion of neoendemic species than in regions with lower levels of genetic isolation (Gillespie & Roderick, 2002). In Macaronesia, there are several examples of adaptive radiations in vertebrate species such as lizards (Gallotiinae, Gekkonidae, and Scincidae; Carranza *et al.,* 2001, 2002; Carranza & Arnold 2006; Cox *et al.,* 2010) and invertebrates such as beetles (*Calathus*, *Meladema*, *Pimelia*, *Tarphius*), butterflies (*Gonepteryx*), and spiders (*Pholcus*; Brunton & Hurst 1998; Emerson *et al.*, 2000a,b; Contreras-Diaz *et al.,* 2003; Ribera *et al.,* 2003; Dimitrov *et al.,* 2008). However, rates of island endemism appear to be lower for mosquitoes (Diptera: Culicidae). Of the 11 mosquito species/subspecies found in the Canary Islands, only two are endemic for Macaronesia and these are shared with Madeira (Capela, 1982; Báez & Oromí, 2010). This contrasts with the nearly 50% endemism rate among terrestrial invertebrate species in Canary Islands (Juan *et al.,* 2000). A similarly low proportion of endemic mosquitoes is observed in other volcanic islands such as Cape Verde and Hawaii (Shroyer, 1981; Alves *et al.,* 2010). The reason for the relative paucity of adaptive radiation in island mosquito populations is that they are very recent colonizers often as a result of multiple human-mediated introductions (Fonseca *et al.,* 2000; Lounibos, 2002; Bataille *et al.,* 2009).

Invasions of certain mosquito species can have a negative impact in vertebrates and humans due to their ability to serve as transmission vectors of diseases (Lounibos, 2002; Delatte *et al.,* 2011). A remarkable example was the decline of native bird' populations in Hawaii associated with avian malaria and avian pox virus transmitted by the introduced mosquito vector *Culex quinquefasciatus* Say, 1823 (Fonseca *et al.,* 2000; Lapointe *et al.,* 2012).

The *Culex pipiens* complex (Figure 1) comprises mosquito vectors responsible for the transmission of lymphatic filariasis and neurotropic arboviruses from the Japanese encephalitis serogroup including the West Nile virus (WNV) to humans (Smith & Fonseca, 2004; Solomon, 2004). The nominal species of the complex, *Culex pipiens* Linnaeus 1758 *sensu stricto* (hereafter termed *Cx. pipiens*) and *Cx. quinquefasciatus* are the most common and widespread species. The former is found primarily in temperate zones, whereas the latter occurs in tropical and subtropical zones. *Cx. pipiens* has a greater ecological range with populations found from the low subarctic of Siberia and Scandinavian countries to the semidesert regions of the

Maghreb (Vinogradova, 2000). *Cx. quinquefasciatus* is confined to warmer tropical and subtropical regions with a higher degree of humidity (Subra, 1981; Fonseca *et al.*, 2006). However, it is possible to find regions where both species coexist sympatrically and where hybrids of the two species have been observed (Urbanelli *et al.*, 1995; Humeres *et al.*, 1998; Kothera *et al.*, 2009; Alves *et al.*, 2010).

Figure 1. Mosquito (female) of the *Culex pipiens* complex.

Photograph with a digital camera SC30 (OLYMPUS, Tokyo, Japan) under a stereomicroscope OLYMPUS SZ61 (12× magnification).

In North America, morphological identification of males based on the length of the dorsal and ventral arms of the phallosome, namely the DV/D ratio, revealed the presence of only *Cx. pipiens* at latitudes above 39ºN while *Cx. quinquefasciatus* was the only species found at latitudes below 36ºN (Barr, 1957). In the areas between 36ºN and 39ºN, a hybrid zone between the two species has been described (Barr, 1957; Savage *et al.*, 2008). Females are morphologically indistinguishable, and several molecular methods have been described to identify these sibling species (Farajollahi *et al.*, 2011). Of these, the PCR assay based on species-specific polymorphisms in the intron-2 of the acetylcholinesterase-2 gene (*ace*-2) has been one of the most widely used (Smith & Fonseca, 2004). Allozyme studies confirmed the latitudinal cline between the two species (Cornel *et al.*, 2003). A recent microsatellite-based study extended the geographic limits of "Barr's hybrid zone" suggesting a wider area between 30ºN and 40ºN (Kothera *et al.*, 2009).

In contrast with the American continent, isolation between *Cx. pipiens* from Mediterranean Europe and *Cx. quinquefasciatus* from the northern hemisphere of Africa

appears to be absolute. The most plausible explanation for this isolation is the presence of the Sahara desert. This inhospitable region lying between 15ºN and 33ºN acts as a barrier to gene flow not only for insects but also for other organisms (Douady *et al.*, 2003; Kodandaramaiah & Wahlberg, 2007). An exception is likely to be found in the Macaronesian region. Madeira, the Canary Islands, and Cape Verde locate within the latitudinal interval of the Saharan desert. In spite of the influence of the Saharan winds, the islands that compose these archipelagos have quite varying climates, ranging from temperate with dry summers (Madeira: Csb, Köppen Classification System) to arid with hot temperatures (Cape Verde: BWh; Peel *et al.*, 2007). Importantly, many of these islands display environmental conditions for sustaining mosquito populations.

Populations of *Cx. pipiens* have been identified in the four archipelagos, and *Cx. quinquefasciatus* has been recorded in Cape Verde (Capela, 1982; Alves *et al.*, 2010; Báez & Oromí, 2010; Vieira *et al.*, 2010). The observation of intermediate DV/D values for the male genitalia of some specimens from Cape Verde suggested the presence of hybrids (Ribeiro *et al.*, 1980). Similarly, in a recent update on the mosquito fauna of Cape Verde, molecular identification based on the *ace*-2 marker suggested hybrid frequencies of 39–67% in two islands of the archipelago (Alves *et al.*, 2010). However, the extent of hybridisation and genetic introgression between *Cx. pipiens* and *Cx. quinquefasciatus* in these islands is still largely unknown.

There are several examples of species expansion mediated by human activity that have broken the geographic isolation between sibling species of insects and other organisms (Pinto *et al.*, 2005; Steeves *et al.*, 2010). The lack of other isolation mechanisms between these species may allow introgression leading to species assimilation or erosion of species boundaries. There is evidence supporting that two isolation mechanisms between *Cx. pipiens* and *Cx. quinquefasciatus* are likely to occur: (1) prezygotic isolation may result from differences in species distribution and in mating behavior; and (2) intrinsic postzygotic may result from cytoplasmic incompatibility that creates unviable hybrids (Vinogradova, 2000; Cornel *et al.*, 2003). A role of extrinsic postzygotic mechanisms linked to hybrid fitness (McBride & Singer, 2010) in the isolation of the *Cx. pipiens* complex members remains unclear.

Both species also display important bio-ecological differences. *Culex quinquefasciatus* is generally considered a more synanthropic urban mosquito compared with a more rural *Cx. pipiens* (Ribeiro *et al.*, 1980; Subra, 1981). In Brazil and in Africa, *Cx. quinquefasciatus* displays a strong preference for mammals (including humans; Subra, 1981; Muturi *et al.*, 2008; Lorosa *et al.*, 2010). In North America, there are differences in host preference among populations of *Cx. quinquefasciatus*, with some preferring mammals (Zinser *et al.*, 2004; Molaei *et al.*, 2007) and others birds (Savage *et al.*, 2007; Molaei *et al.*, 2010). *Culex pipiens* preferentially feeds upon birds (Kilpatrick *et al.*, 2006, 2007; Molaei *et al.*, 2006). Hybridisation between members of the *Cx. pipiens* complex with different host preferences may promote a more opportunistic feeding behavior increasing the importance of the host availability (Fonseca *et al.*, 2004; Balenghien *et al.*, 2011). Consequently, this population with more catholic feeding behavior would have an increased potential as a bridge vector between bird and humans for the transmission of WNV (Molaei *et al.*, 2007; Savage *et al.*, 2007; Kilpatrick, 2011).

In this study, Bayesian model-based methods were applied to multilocus microsatellite genotypes to infer the genetic structure of the *Cx. pipiens* complex in Madeira and in four islands of Cape Verde. The aims were (1) to determine the degree of genetic differentiation among island populations; (2) to measure rates of hybridisation and genetic introgression between the sibling species; and (3) to infer about the colonization process and their impacts in Macaronesian region.

Material and methods

Mosquito collections

Indoor resting collections of adult mosquitoes using mechanical aspirators were carried out in four localities of Madeira Island in September 2006 and in June 2007 (Figure 2). Given the very low adult mosquito densities found in Cape Verde (Pinto *et al.*, 1999), collections of immature culicids were undertaken using dippers and pipettes, between November and December 2007, in four islands of Cape Verde: Brava, Fogo, Santiago, and Maio (Figure 2). Immature collections were made on a wide range of breeding sites such as ponds, pools, swamps, pits, water tanks, and septic tanks.

Information on the localities in which *Cx. pipiens s.l.* larvae were sampled in Cape Verde is shown in Table S1 (Supporting information; see also Alves *et al.*, 2010). Collected larvae were transported to a laboratory in Praia (Santiago Island) and reared to adulthood. Adult mosquitoes were killed by freezing and identified to species/complex using morphological keys (Ribeiro & Ramos, 1995, 1999). Samples were stored over silica gel until DNA extraction.

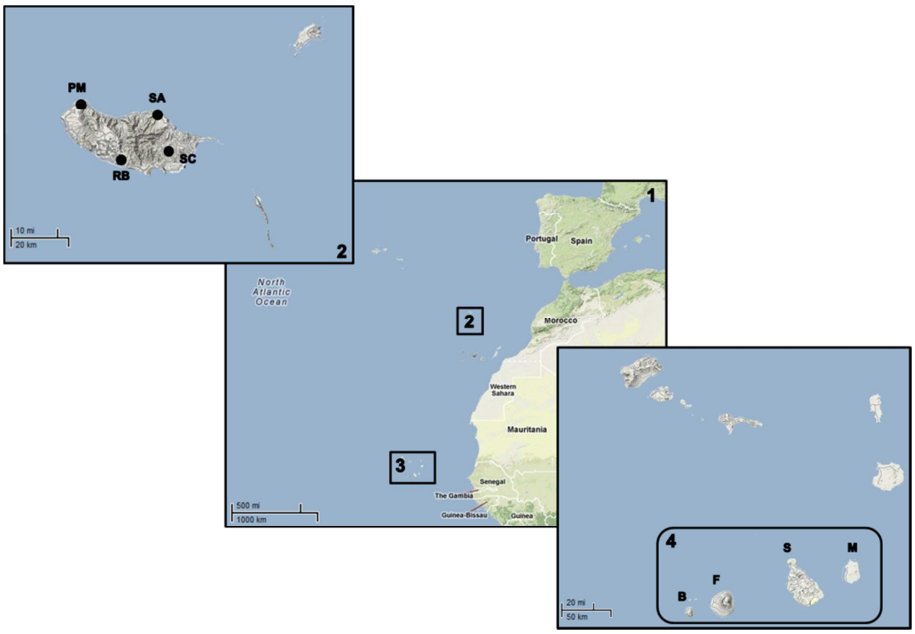

Figure 2. Maps of the North Atlantic region showing the localities/islands sampled.

(1) North Atlantic region including West Africa and Mediterranean region; (2) Madeira archipelago; PM: Porto Moniz; RB: Ribeira Brava; SC: Santa Cruz; SA, Santana; (3) Cape Verde archipelago; (4) Islands sampled in Cape Verde; B: Brava; F: Fogo; S: Santiago; M: Maio. Images collected in Google Maps – ©2012 Google (http://maps.google.com/).

Molecular analyses

DNA extraction from individual female mosquitoes was performed using the method of Collins *et al.* (1987). Each specimen was identified to species by a multiplex PCR assay targeting species-specific polymorphisms in intron-2 of the *ace*-2 gene using primers specific for *Cx. pipiens*, *Cx. quinquefasciatus*, and *Cx. torrentium* (Smith & Fonseca, 2004).

Twelve microsatellites (Fonseca *et al.*, 1998; Keyghobadi *et al.*, 2004; Smith *et al.*, 2005) were genotyped (see Table S2). For each specimen, each locus was amplified separately in a 20 µL PCR reaction that contained 1X GoTaq® Flexi Buffer (Promega,

Madison, Wisconsin), 2.5 mM $MgCl_2$, 0.25 mM dNTPs, 0.20 mg/mL Bovine Serum Albumin, 0.20 μm of each primer, and 0.5 U GoTaq® Flexi DNA polymerase (Promega). For each locus, one of the primers was fluorescently labeled (NED, HEX our 6-FAM; Applied Biosystems, Carlsbad, California). Thermocycling conditions included an initial denaturation step of 5 min at 96ºC, followed by 30 cycles each of 96ºC for 30 sec, annealing at 52ºC-56ºC (locus-dependent) for 30 sec and 72ºC for 30 sec. After a final extension step of 5 min at 72ºC, reactions were stopped at 4ºC.

Amplified products were separated by capillary electrophoresis in a genetic analyzer ABI3730 (Applied Biosystems) at Yale DNA Analysis Facility (USA). Fragment sizes and genotypes were scored using the software GeneMarker 1.4. (Softgenetics, State College, Pennsylvania).

Data analysis

Genetic diversity at each microsatellite locus was characterized by estimates of expected heterozygosity (Nei, 1987) and inbreeding coefficient (F_{IS}). Significance of F_{IS} values was assessed by randomization tests. These analyses were performed using FSTAT v. 2.9.3.2. (Goudet, 1995). Estimates of allele richness (A_R), a measure of allele diversity adjusted for the lowest sample size, were obtained by the statistical rarefaction approach implemented in HP-RARE (Kalinowski, 2005). Departures from Hardy–Weinberg proportions were tested by exact tests available in ARLEQUIN v.3.11 (Excoffier *et al.*, 2005). The same software was used to perform exact tests of linkage equilibrium between pairs of loci based on the expectation-maximization approach described by Slatkin & Excoffier (1996). The software Micro-Checker 2.2.3. (van Oosterhout *et al.*, 2004) was used to test for the presence of null alleles (99% confidence interval) at each locus/sample.

Bayesian clustering analysis as implemented by STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) was used to infer population substructure/ancestry from the data set without prior information of sampling groups under the conditions of admixture (α allowed to vary between 0 and 10), and allele frequencies correlated among populations (λ was set at 1, default value). Ten independent runs with 10^5 iterations and replications were performed for each value of K (K = 1–10 clusters). The inference of the number of genetic clusters (K) in the Bayesian method implemented by STRUCTURE is not

straightforward, and it is normally performed by *ad hoc* approaches: an estimation of ln[Pr(X|*K*)], described in the original publication (Pritchard *et al.,* 2000) and the ΔK statistic (Evanno *et al.,* 2005). We used a combination of these approaches with a sequential procedure in which data were analyzed at three levels: (1) all samples, (2) each archipelago, and (3) each island. Following the suggestions of Vähä & Primmer (2006), individual genetic assignment to clusters was based on a minimum posterior probability threshold (*Tq*) of 0.90. Individuals displaying $0.1 \leq q_i \leq 0.90$ were considered of admixed ancestry. The information from the outputs of each *K* (10 runs) was aligned by the Greedy method implemented in CLUMPP (Jakobsson & Rosenberg, 2007).

The Bayesian method implemented by NEWHYBRIDS 1.1. (Anderson & Thompson, 2002) was used to assign individuals into six classes: two pure (parental *Cx. pipiens* and *Cx. quinquefasciatus*) and four hybrid (F1, F2, and backcrosses with the parental populations). The approach of uniform priors was used because it reduces the influence of low-frequency alleles thus which may result from sampling and genotyping errors in closely related populations. Results were based on the average of five independent runs of 10^5 iterations. Following the suggestions of Anderson & Thompson (2002), individual genetic assignment to classes was based on a minimum posterior probability threshold (*Tq*) of 0.50.

A neighbor-joining (NJ) tree based on Cavalli-Sforza & Edwards (1967) chord distance (Dc) was used to represent the relationships among genetic clusters and geographic samples. Individuals with an admixed genetic background (*i.e.* with a probability of assignment not attributable to any of the purebred or hybrid clusters) were excluded from this analysis. A consensus tree was obtained by bootstrapping (1000 replicates) distance values over loci. Calculations were performed with the program Populations 1.2.30 (Langella, 1999). The software Treeview (Page, 1996) was used to visualize the tree.

Whenever multiple testing was performed, the nominal significance level of rejection of the null hypothesis ($\alpha = 0.05$) was corrected by the sequential Bonferroni procedure (Holm, 1979).

Results

ace-2 molecular identification

A total of 374 females (Madeira: 190 and Cape Verde: 184, distributed as follows, Brava: 31, Fogo, 36, Santiago: 54, Maio: 63) were analyzed by the molecular assay *ace*-2 (Smith & Fonseca, 2004; Table 1). Of these, 203 were identified as *Cx. pipiens* and were collected in Madeira (*N* = 190) and in Maio (*N* = 13). *Culex quinquefasciatus* was found in the four islands of Cape Verde (*N* = 115), and it was the only member of the complex present in the collections from Brava (*N* = 31) and Santiago (*N* = 54). Fifty-six mosquitoes displayed a heterozygous pattern for *ace*-2 and were collected in Fogo (*N* = 14) and Maio (*N* = 42). The island of Maio was the only island where the two species and putative hybrids were found in sympatry.

Table 1. Molecular identification of *Culex pipiens* complex species based on the molecular assay in the *ace-2.*

	N	**Localities**							
		Madeira				**Cape Verde**			
		PM	**RB**	**SC**	**SA**	**B**	**F**	**S**	**M**
Cx. pipiens	203	66 (100.0)	39 (100.0)	34 (100.0)	51 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	13 (20.6)
Hybrids	56	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	14 (38.9)	0 (0.0)	42 (66.7)
Cx. quinquefasciatus	115	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	31 (100.0)	22 (61.1)	54 (100.0)	8 (12.7)

N: number of individuals; PM: Porto Moniz; RB: Ribeira Brava; SC: Santa Cruz; SA: Santana; B: Brava Island; F: Fogo Island; S: Santiago Island; M: Maio Island. Values in parenthesis refer to the relative genotypic frequencies (in percentage) within each locality.

Clustering analysis

Genetic diversity estimates for the 12 microsatellites in whole sample (*N* = 374) and subsamples determined by *ace*-2 identification and geographic location are shown in Table S3. Loci CQ26 and CQ41 exhibited heterozygote deficits in all subsamples from Madeira, possibly reflecting locus-specific effects, such as null alleles or selection. The analysis performed by the Micro-Checker software confirmed the possibility of null alleles at loci CQ26 and CQ41 in samples from Madeira island (see Table S3). These loci were therefore excluded from Bayesian assignment and genetic differentiation analyses.

The Bayesian analysis implemented in STRUCTURE and the two *ad hoc* approaches to define the number of clusters revealed a homogeneous population in Madeira (K = 1) and the intriguing scenario of Cape Verde with three possible subdivisions (K = 2, K = 3, or K = 4; Figure 3, see Figure S1, S2). The sequential procedure under the three levels of organization (whole sample, archipelago, and island) highlighted a further subdivision within the islands of Maio and Fogo providing support for K = 3 in the archipelago of Cape Verde and consequently a K = 4 for the whole sample (Figure 3, see Figure S3).

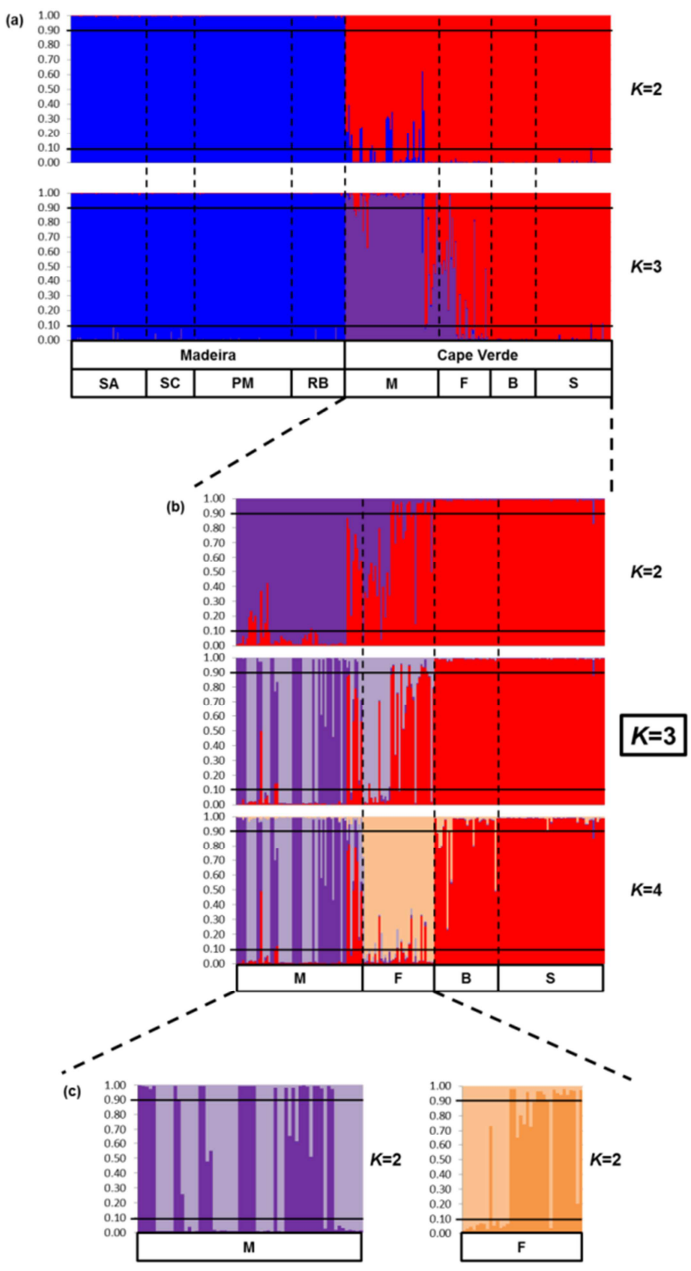

Figure 3. Bayesian cluster analysis conducted by STRUCTURE at three different levels.

(a) all samples, (b) Cape Verde samples, (c) Maio and Fogo islands. *K*: number of clusters. Columns correspond to the multilocus genotype of each individual, partitioned in different colors representing the probability of ancestry (q_i) to each cluster. Individuals were ordered according to their geographic information. Lines indicate the qi threshold used to determine admixed individuals (see 'Materials and Methods').

The combination of the Bayesian clustering results with the *ace*-2 identification clarified the separation of sampled mosquitoes into four different clusters (Table 2): Cluster 1 (C1) grouped all the 190 *Cx. pipiens* from Madeira, while the other three clusters were restricted to Cape Verde. Cluster 4 (C4) was the most abundant in the archipelago with 91 specimens from three islands (Brava, Fogo and Santiago), all identified as *Cx. quinquefasciatus* by *ace*-2 PCR. Cluster 2 (C2) was the smallest cluster with 25 specimens from Maio Island being classified as *Cx. pipiens* or hybrid by *ace*-2 PCR. Cluster 3 (C3) includes individuals from Fogo and Maio Islands, and the majority (87.2%) of the specimens were identified as hybrids by *ace*-2 PCR. Twenty-nine specimens, the majority of which from Maio and Fogo, were not assigned to any of the four clusters and were thus considered admixed.

Table 2. *ace-2* PCR species composition and relative distribution per locality/island of each genetic cluster revealed by STRUCTURE.

	N	*ace-2*			Localities							
		P	H	Q	Madeira				Cape Verde			
					PM	RB	SC	SA	B	F	S	M
Cluster 1	190	190 (100.0)	0 (0.0)	0 (0.0)	66 (34.7)	39 (20.5)	34 (17.8)	51 (26.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Cluster 2	25	9 (36.0)	16 (64.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	25 (100.0)
Cluster 3	39	1 (2.6)	34 (87.2)	4 (10.2)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	14 (35.9)	0 (0.0)	25 (64.1)
Cluster 4	91	0 (0.0)	0 (0.0)	91 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	31 (34.1)	7 (7.7)	53 (58.2)	0 (0.0)
Admixed	29	3 (10.3)	6 (20.7)	20 (69.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	15 (51.6)	1 (3.4)	13 (44.8)

N: number of individuals; P: *Culex pipiens* by *ace-2* identification; Q: *Culex quinquefasciatus* by *ace-2* identification; H: hybrids between *Culex pipiens* and *Culex quinquefasciatus* by *ace-2* identification; PM: Porto Moniz; RB: Ribeira Brava; SC: Santa Cruz; SA: Santana; B: Brava Island; F: Fogo Island; S: Santiago Island; M: Maio Island. Values in parenthesis refer to the frequencies (in percentage) within each cluster.

The analysis with NEWHYBRIDS confirmed the homogeneity of the Madeira population (C1). In Cape Verde, all the samples from C4 were classified as pure *Cx. quinquefasciatus*, while the majority of the individuals of C2 (96.0%) were classified as pure *Cx. pipiens* and one individual was classified as a backcross with *Cx. pipiens* (BxP). The majority of individuals of C3 (87.1%) were classified as hybrids. Of these,

10 (nine in Fogo, one in Maio) were classified as F2 hybrids and nine individuals from Maio were backcrosses with *Cx. pipiens* (BxP; Table 3).

Table 3. Frequencies purebred and hybrid individuals detected by NEWHYBRIDS in each of the ancestry clusters revealed by STRUCTURE.

	N	NEWHYBRIDS							
		P	Q	H	H				
					F1	F2	BxP	BxQ	H'
Cluster 1	190	189 (99.4)	0 (0.0)	1 (0.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)
Cluster 2	25	24 (96.0)	0 (0.0)	1 (4.0)	0 (0.0)	0 (0.0)	1 (4.0)	0 (0.0)	0 (0.0)
Cluster 3	39	4 (10.3)	1 (2.6)	34 (87.1)	0 (0.0)	10 (25.6)	9 (23.1)	0 (0.0)	15 (38.4)
Cluster 4	91	0 (0.0)	91 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Admixed	29	3 (10.4)	13 (44.8)	13 (44.8)	0 (0.0)	8 (27.6)	0 (0.0)	0 (0.0)	5 (17.2)

N: number of individuals; P: pure *Culex pipiens*; Q: pure *Culex quinquefasciatus*; H: hybrids between the pure groups (*Culex pipiens* and *Culex quinquefasciatus*); F1: hybrid first generation; F2: hybrids second generation; BxP: backcross *Culex pipiens*; BxQ: backcross *Culex quinquefasciatus*: H′: hybrids defined by the sum of assignment probabilities for all hybrid classes. Values in parenthesis refer to the frequencies (in percentage) within each cluster.

The Dc-based NJ tree was consistent with the presence of the four clusters identified in the analysis performed by STRUCTURE (Figure 4). *Culex pipiens* samples from Madeira (C1) and Maio (C2) displayed a high genetic distance but were still grouped in a common cluster separated from the remaining samples. Samples from cluster C3, composed mainly by hybrid individuals, displayed an intermediate position in the topology of the tree. *Culex quinquefasciatus* samples from cluster C4 shared the same cluster, but it was possible to observe significant divergence between the populations of the three islands (Brava, Fogo, and Santiago).

Discussion

In this study, the distribution and levels of hybridisation between *Cx. pipiens* and *Cx. quinquefasciatus* were found to differ among islands of the Macaronesian region. Madeira showed a genetically homogenous *Cx. pipiens* population. In Cape Verde, it was possible to identify monospecific populations of *Cx. quinquefasciatus* in Brava and

Santiago, while admixed populations between both species were observed in Maio and Fogo. The species diagnostic *ace*-2 PCR was effective in the identification of each species in the allopatric populations of Madeira, Brava, and Santiago. However, in sympatric populations with interspecific admixture such as those of Maio and Fogo, repeated introgression and recombination lead to a disruption of the linkage between the diagnostic alleles and the respective genetic backgrounds of each species. Under these conditions, a more cautious interpretation of the results obtained by a single diagnostic marker such as the *ace*-2 is needed for the correct identification of each species and hybrids (McAbee *et al.*, 2008; Fonseca *et al.*, 2009).

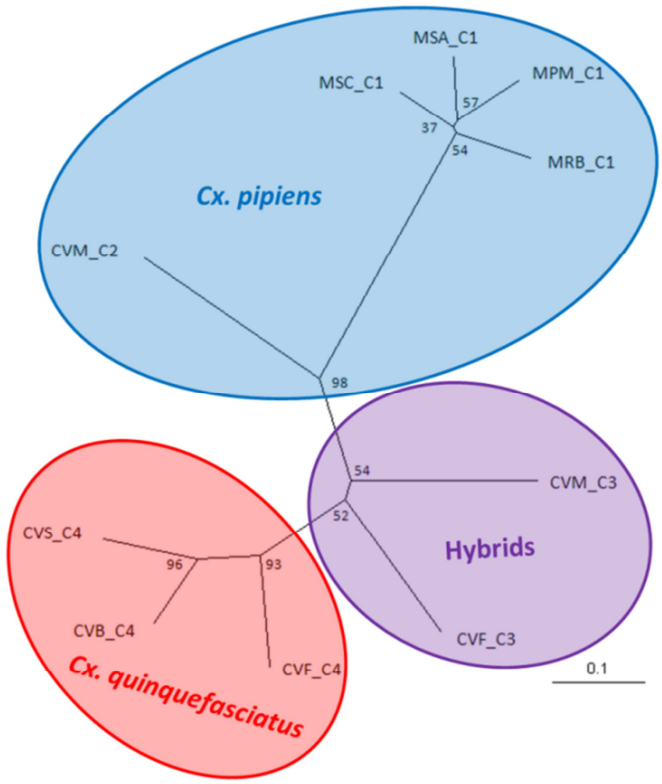

Figure 4. Phylogenetic tree of the *Culex pipiens* complex in Madeira and Cape Verde.

CVM: Maio; CVF: Fogo; CVB: Brava; CVS: Santiago; MSC: Santa Cruz (Madeira); MAS: Santana (Madeira); MPM: Porto Moniz (Madeira); MRB: Ribeira Brava (Madeira); C1: cluster 1; C2: cluster 2; C3: cluster 3; and C4: cluster 4.

The presence of a monospecific population of *Cx. pipiens* in Madeira agrees with previous reports (Capela, 1981; Fonseca *et al.*, 2004). This volcanic island locates in the temperate zone of the North Atlantic, 700 km off the coast of Morocco and 850 km from continental Portugal. In both countries, only *Cx. pipiens* has been identified (Trari *et al.*, 2002; Almeida *et al.*, 2008). Since its discovery in the 15th century, Madeira has been an important port-of-call in the Atlantic so that the introduction of *Cx.*

pipiens could have resulted from human-mediated passive dispersal (Lounibos, 2002). The Mediterranean temperate climate of this island should also be compatible with the establishment of *Cx. pipiens*. The absence of *Cx. quinquefasciatus* may reflect a lower tolerance of this species to more temperate climates with lower temperatures during winter. The possibility of this vector having never been introduced into this island is probably less likely in spite of the *ca.* 2000 km distance between Madeira Island and the sub-Saharan African coast. Migration by human-mediated dispersal in *Cx. pipiens* complex includes long-distance introductions that spread organophosphate insecticides resistance genes between continents and established invasive populations in isolated archipelagos such as Hawaii (Chevillon *et al.*, 1999; Fonseca *et al.*, 2006).

The distribution of the two members of the *Cx. pipiens* complex in Cape Verde is more intricate and reflects both bio-geographic features and historical aspects of the human peopling of the islands. The apparent predominance of *Cx. quinquefasciatus* on the archipelago agrees with the bio-geographic context of the islands, which lie in the tropical zone of the North Atlantic. The most likely origin of this species would be the West African continent. However, Fonseca *et al.* (2006) in a worldwide genetic survey of *Cx. quinquefasciatus* provided evidence for a recent introduction of *Cx. quinquefasciatus* in West Africa from the New World. Given the geographic intermediate location and the strategic importance of Cape Verde in maritime routes, one cannot exclude the possibility of introduction of mosquitoes of New World origin. The occurrence of *Cx. pipiens* most likely reflects the historical relationship of the archipelago with the Mediterranean region of the European continent. The islands were discovered by Portuguese sailors in the 15th century, and the subsequent peopling was made by migrants of both European and African origin. The Portuguese traders used the archipelago as a port-of-call for ship provisioning during travels between Europe and the African continent and also as a Senegambian slave outpost for the Atlantic (Brehm *et al.*, 2002). The considerable movement of ships between the islands and both continents could have provided the opportunity for the introduction of both mosquito species. It remains to be determined whether the present mosquito populations in Cape Verde islands result from multiple introduction events of one or both species. The analysis of mainland samples and of other molecular markers (*e.g.* mtDNA) would be required for this purpose (Hardouin *et al.*, 2010).

In Maio and Fogo, a considerable number of individuals were assigned as hybrids and yet one of the parental species was absent from the samples (*Cx. quinquefasciatus* in Maio; *Cx. pipiens* in Fogo). A possible explanation for this apparent contradiction could be insufficient sampling of the least abundant species. Factors that may have contributed for an insufficient sampling were the collection method used (immature captures) coupled with the low number of breeding sites positive for *Cx. pipiens s.l.* larvae. In Cape Verde, very low adult mosquito densities preclude the use of collection methods targeting adult mosquitoes for sampling sufficient numbers of individuals (Ribeiro *et al.,* 1980; Pinto *et al.,* 1999). However, other explanations may be proposed for these observations. Maio is the driest of the islands sampled and a lower density or virtual absence of a stable *Cx. quinquefasciatus* population agrees with its lower tolerance to aridity. Fogo has a very steep topography marked by the presence of a volcanic cone. A similar scenario to that of Madagascar, where *Cx. quinquefasciatus* predominated in the lowland urban areas and *Cx. pipiens* were found at altitudes above 1300 m (Urbanelli *et al.,* 1995), may occur in this island. Insufficient sampling could also explain the apparent absence of *Cx. pipiens* in Brava and Santiago, although in these cases there was no evidence of admixture. While the absence of this species agrees with previous surveys in Brava, the same does not hold for Santiago. In this island, *Cx. pipiens* prevailed over *Cx. quinquefasciatus* in larval collections performed in the late 1970s (Ribeiro *et al.,* 1980). This apparent inversion in the relative abundance of both species may be associated with an increase in urbanization of this island over the past recent years. Such an increase in urbanization could confer a greater adaptive advantage for *Cx. quinquefasciatus* over *Cx. pipiens*.

High hybridisation rates between *Cx. pipiens* and *Cx. quinquefasciatus* were detected in two islands of Cape Verde. These rates (Fogo: 39%; Maio: 40%) are comparable with those recorded in the hybrid zone of North America (Savage *et al.,* 2008; Kothera *et al.,* 2009) and contrast with the pattern of sympatry without hybridisation observed in southeast Africa (Cornel *et al.,* 2003). The lack of hybridisation in southeast Africa was justified by the presence of *Wolbachia pipientis* only in *Cx. quinquefasciatus*, whereas in North America, both species are infected with the same strain (Cornel *et al.,* 2003; Rasgon & Scott, 2003). *Wolbachia pipientis* infection can induce sterility by cytoplasmic incompatibility (an intrinsic postzygotic

isolation mechanism) between infected males and uninfected females or females infected by incompatible strains (Atyame *et al.*, 2011). In West Africa and in the Mediterranean region, both *Cx. pipiens* and *Cx. quinquefasciatus* populations share closely related strains of *W. pipientis* (Atyame *et al.*, 2011). Assuming a putative origin of both species from those regions, the introduction of mosquito populations possessing similar strains of *W. pipientis* (or no infection) may explain the high levels of hybridisation in Cape Verde. Molecular analysis of *W. pipientis* in these mosquito populations would help clarifying this hypothesis.

Isolation between close species can be promoted by several mechanisms that may act in simultaneous. The lack or incomplete action of prezygotic (*e.g.* mating behavior) and intrinsic postzygotic (*e.g.* cytoplasmic incompatibility) mechanisms allows hybridisation creating first generation (F1) hybrids. However, an extrinsic postzygotic mechanism such as hybrid sterility or hybrid low fitness can restrict gene flow to one generation avoiding introgression (Bono & Markow, 2009; McBride & Singer, 2010; Muñoz *et al.*, 2010). The analysis performed by NEWHYBRIDS in Cape Verde samples showed the presence of ~50% of second generation hybrids (25.6% F2 and 23.1% BxP; Table 3) within the hybrid cluster. The repeated hybridisation and backcrossing with *Cx. pipiens* indicate mating success of F1 individuals (males and females) suggesting a low effect of extrinsic postzygotic isolation mechanisms between the *Cx. pipiens* and *Cx. quinquefasciatus* in Cape Verde.

Macaronesia is a passage route and breeding region for migratory birds (Garcia-del-Rey, 2011). These birds may potentially introduce parasites and viruses that are known to be transmitted by the *Cx. pipiens* complex such as *Plasmodium relictum* (avian malaria), avian pox virus, Usutu virus and WNV. The introduction of these pathogens may place the endemic bird populations in danger (Kilpatrick, 2011; Savini *et al.*, 2011; Lapointe *et al.*, 2012). Furthermore, the high levels of hybridisation between *Cx. pipiens* and *Cx. quinquefasciatus* may promote a more opportunistic feeding behavior increasing the chance for the accidental transmission of WNV to humans. A serologic survey in 1980s showed 40% positive cases of WNV in Cape Verdean children (Vieira, 1985). However, it has been recognized that WNV serologic surveys of the last century had considerable false positives due to cross-reactivity with other flavivirus (Tardei *et al.*, 2000). Even so, the possibility of disease outbreaks

should not be neglected given the outcome of the introduction of WNV to the western hemisphere (Kilpatrick, 2011), or the more recent dengue epidemic in Cape Verde in 2009 (Franco *et al.,* 2010), highlighting the receptivity of a territory once a suitable vector is present.

Acknowledgments

We are grateful to the inhabitants of Madeira and Cape Verde for their collaboration in this study. We acknowledge the logistic support in Cape Verde given by the General Department of Health (J. Pereira), the delegates and technicians of the Health Care Units of Praia, St. Cruz, Tarrafal, S. Miguel, Órgãos, Rª Grande de Santiago and St. Catarina, in Santiago island,, and of Maio, Fogo and Brava islands. We thank the support of Regina Rodrigues (National Malaria Program) and João Silva (CMDT/IHMT) in the field collections. This study was funded by Fundação para a Ciência e a Tecnologia, Portugal (POCI/BIA-BDE/57650/2004 and PPCDT/BIA-BDE/57650/2004; PPCDT/SAU-ESP/55110/2004). Joana Alves and Bruno Gomes were funded by a PhD fellowship of Fundação para a Ciência e Tecnologia/MCTES (SFRH/BD/153451/2005, SFRH/BD/36410/2007).

References

Almeida, A.P.G., Galão, R.P., Sousa, C.A., Novo, M.T., Parreira, R., Pinto, J., Piedade, J. & Esteves, A. (2008) Potential mosquito vectors of arboviruses in Portugal: species, distribution, abundance and West Nile infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 102 (8), 823–832

Alves, J., Gomes, B., Rodrigues, R., Silva, J., Arez, A.P., Pinto, J. & Sousa, C. A. (2010) Mosquito fauna on the Cape Verde islands (West Africa): An update on species distribution and a new finding. *Journal of Vector Ecology*. 35 (2), 307–312.

Anderson, E.C. & Thompson, E.A. (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 160 (3), 1217–1229.

Atyame, C.M., Delsuc, F., Pasteur, N., Weill, M. & Duron, O. (2011) Diversification of *Wolbachia* endosymbiont in the *Culex pipiens* mosquito. *Molecular Biology and Evolution*. 28 (10), 2761–2772.

Balenghien, T., Fouque, F., Sabatier, P. & Bicout, D.J. (2011) Theoretical formulation for mosquito host-feeding patterns: application to a West Nile virus focus of Southern France. *Journal of Medical Entomology*. 48 (5), 1076–1090.

Barr, A.R. (1957) The distribution of *Culex p. pipiens* and *C. p. quinquefasciatus* in North America. *The American journal of Tropical Medicine and Hygiene*. 6 (1), 153–165.

Bataille, A., Cunningham, A.A., Cedeño, V., Cruz, M., Eastwood, G., Fonseca, D.M., Causton, C.E., Azuero, R., Loayza, J., Martinez, J.D.C. & Goodman, S.J. (2009) Evidence for regular ongoing introductions of mosquito disease vectors into the Galápagos Islands. *Proceedings. Biological sciences / The Royal Society*. 276 (1674), 3769–3775.

Bono, J.M. & Markow, T.A. (2009) Post-zygotic isolation in cactophilic *Drosophila*: larval viability and adult life-history traits of *D. mojavensis/D. arizonae* hybrids. *Journal of Evolutionary Biology*. 22 (7), 1387–1395.

Brehm, A., Pereira, L., Bandelt, H.-J., Prata, M.J. & Amorim, A. (2002) Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Annals of Human Genetics*. 66 (1), 49–60.

Brunton, C. & Hurst, G. (1998) Mitochondrial DNA phylogeny of Brimstone butterflies (genus *Gonepteryx*) from the Canary Islands and Madeira. *Biological journal of the Linnean Society*. 63 (1), 69–79.

Báez, M. & Oromí, P. (2010) Diptera. In: M. Arechavaleta, N. Rodríguez, N. Zurita, & A. García (eds.). *Lista de especies silvestres de Canarias. Hongos, plantas y animales terrestres*. Gobierno de Canarias. pp. 318–342.

Capela, R.A. (1981) Contribution to the study of mosquitoes (Diptera: Culicidae) from the Archipelago of Madeira and the Salvages I - Madeira. *Arquivos do Museu Bocage*. Série A (1), 45–66.

Capela, R.A. (1982) Contribuição para o conhecimento dos mosquitos (Diptera, Culicidae) dos arquipélagos da Madeira e das Selvagens. *Boletim do Museu Municipal do Funchal*. 34, 105–123.

Carranza, S., Arnold, E.N., Mateo, J.A. & López-Jurado, L.F. (2001) Parallel gigantism and complex colonization patterns in the Cape Verde scincid lizards *Mabuya* and *Macroscincus* (Reptilia: Scincidae) revealed by mitochondrial DNA sequences. *Proceedings. Biological sciences / The Royal Society*. 268 (1476), 1595–1603.

Carranza, S., Arnold, E.N., Mateo, J.A. & Geniez, P. (2002) Relationships and evolution of the North African geckos, *Geckonia* and *Tarentola* (Reptilia: Gekkonidae), based on mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution*. 23 (2), 244–256.

Carranza, S. & Arnold, E.N. (2006) Systematics, biogeography, and evolution of Hemidactylus geckos (Reptilia: Gekkonidae) elucidated using mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*. 38 (2), 531–545.

Cavalli-Sforza, L.L. & Edwards, A.W. (1967) Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*. 19 (3 Pt 1), 233–257.

Chevillon, C., Raymond, M., Guillemaud, T., Lenormand, T. & Pasteur, N. (1999) Population genetics of insecticide resistance in the mosquito *Culex pipiens*. *Biological Journal of the Linnean Society*. 68 (1-2), 147–157.

Collins, F.H., Mendez, M.A., Rasmussen, M.O., Mehaffey, P.C., Besansky, N.J. & Finnerty, V. (1987) A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *The American Journal of Tropical Medicine and Hygiene*. 37 (1), 37–41.

Contreras-Diaz, H.G., Moya, O., Oromi, P. & Juan, C. (2003) Phylogeography of the endangered darkling beetle species of *Pimelia* endemic to Gran Canaria (Canary Islands). *Molecular Ecology*. 12 (8), 2131–2143.

Cornel, A.J., Mcabee, R.D., Rasgon, J., Stanich, M.A., Scott, T.W. & Coetzee, M. (2003) Differences in extent of genetic introgression between sympatric *Culex pipiens* and *Culex quinquefasciatus* (Diptera: Culicidae) in California and South Africa. *Journal of Medical Entomology*. 40 (1), 36–51.

Cox, S.C., Carranza, S. & Brown, R.P. (2010) Divergence times and colonization of the Canary Islands by *Gallotia* lizards. *Molecular Phylogenetics and Evolution*. 56 (2), 747–757.

Delatte, H., Bagny, L., Brengue, C., Bouetard, A., Paupy, C. & Fontenille, D. (2011) The invaders: phylogeography of dengue and chikungunya viruses *Aedes* vectors, on the south west islands of the Indian Ocean. *Infection, Genetics and Evolution*. 11 (7), 1769–1781.

Dimitrov, D., Arnedo, M.A. & Ribera, C. (2008) Colonization and diversification of the spider genus *Pholcus* Walckenaer, 1805 (Araneae, Pholcidae) in the Macaronesian archipelagos: evidence for long-term occupancy yet rapid recent speciation. *Molecular Phylogenetics and Evolution.* 48 (2), 596–614.

Douady, C.J., Catzeflis, F., Raman, J., Springer, M.S. & Stanhope, M.J. (2003) The Sahara as a vicariant agent, and the role of Miocene climatic events, in the diversification of the mammalian order Macroscelidea (elephant shrews). *Proceedings of the National Academy of Sciences*. 100 (14), 8325–8330.

Emerson, B.C., Oromí, P. & Hewitt, G.M. (2000a) Interpreting colonization of the *Calathus* (Coleoptera: Carabidae) on the Canary Islands and Madeira through the application of the parametric bootstrap. *Evolution*. 54 (6), 2081–2090.

Emerson, B.C., Oromí, P. & Hewitt, G.M. (2000b) Tracking colonization and diversification of insect lineages on islands: mitochondrial DNA phylogeography of *Tarphius canariensis* (Coleoptera: Colydiidae) on the Canary Islands. *Proceedings. Biological sciences / The Royal Society*. 267 (1458), 2199–2205.

Evanno, G., Goudet, J. & Regnaut, S. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*. 14 (8), 2611–2620.

Excoffier, L., Laval, G. & Schneider, S. (2005) ARLEQUIN (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics online*. 1, 47–50.

Farajollahi, A., Fonseca, D.M., Kramer, L.D. & Kilpatrick, M.A. (2011) "Bird biting" mosquitoes and human disease: A review of the role of *Culex pipiens* complex mosquitoes in epidemiology. *Infection, Genetics and Evolution*. 11 (7), 1577–1585.

Fonseca, D.M., Atkinson, C.T. & Fleischer, R.C. (1998) Microsatellite primers for *Culex pipiens quinquefasciatus*, the vector of avian malaria in Hawaii. *Molecular Ecology*. 7 (11), 1617-1619.

Fonseca, D.M., Lapointe, D.A. & Fleischer, R.C. (2000) Bottlenecks and multiple introductions: population genetics of the vector of avian malaria in Hawaii. *Molecular Ecology*. 9 (11), 1803–1814.

Fonseca, D.M., Keyghobadi, N., Malcolm, C.A., Mehmet, C., Schaffner, F., Mogi, M., Fleischer, R.C. & Wilkerson, R.C. (2004) Emerging vectors in the *Culex pipiens* complex. *Science*. 303 (5663), 1535–1538.

Fonseca, D.M., Smith, J.L., Wilkerson, R.C. & Fleischer, R.C. (2006) Pathways of expansion and multiple introductions illustrated by large genetic differentiation among worldwide populations of the southern house mosquito. *The American Journal of Tropical Medicine and Hygiene*. 74 (2), 284–289.

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.

Franco, L., Di Caro, A., Carletti, F., Vapalahti, O., Renaudat, C., Zeller, H. & Tenorio, A. (2010) Recent expansion of dengue virus serotype 3 in West Africa. *Eurosurveillance* 15 (7), pii=19490.

Garcia-del-Rey, E. (2011) *Field Guide to the Birds of Macaronesia. Azores, Madeira, Canary Islands, Cape Verde*. Bellaterra, Barcelona, Lynx Edicions.

Gillespie, R.G. & Roderick, G.K. (2002) Arthropods on islands: Colonization, speciation, and conservation. *Annual Review of Entomology*. 47, 595–632.

Goudet, J. (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. *The Journal of Heredity*. 86 (6), 485–486.

Hardouin, E.A., Chapuis, J.-L., Stevens, M.I., van Vuuren, J.B., Quillfeldt, P., Scavetta, R. J., Teschke, M. & Tautz, D. (2010) House mouse colonization patterns on the sub-Antarctic Kerguelen Archipelago suggest singular primary invasions and resilience against re-invasion. *BMC Evolutionary Biology*. 10 (1), 325.

Holm, S. (1979) A Simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 6 (2), 65–70.

Howarth, F.G. & Ramsay, G.W. (1991) The conservation of island insects and their habitats. In: N.M. Collins & J.A. Thomas (eds.). *The conservation of Insects and their habitat*. London, Academic Press. pp. 71–107.

Humeres, S.G., Almirón, W.R., Sabattini, M.S. & Gardenal, C.N. (1998) Estimation of genetic divergence and gene flow between *Culex pipiens* and *Culex quinquefasciatus* (Diptera: Culicidae) in Argentina. *Memórias do Instituto Oswaldo Cruz*. 93 (1), 57–62.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* . 23 (14), 1801–1806.

Juan, C., Emerson, B.C., Oromí, P. & Hewitt, G.M. (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology & Evolution*. 15 (3), 104–109.

Kalinowski, S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*. 5 (1), 187–189.

Keyghobadi, N., Matrone, M.A., Ebel, G.D., Kramer, L.D. & Fonseca, D.M. (2004) Microsatellite loci from the northern house mosquito (*Culex pipiens*), a principal vector of West Nile virus in North America. *Molecular Ecology Notes*. 4 (1), 20–22.

Kilpatrick, A.M., Kramer, L.D., Jones, M.J., Marra, P.P., Daszak, P. & Fonseca, D.M. (2006) West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. *PLoS Biology*. 4 (4), e82.

Kilpatrick, A.M., Kramer, L.D., Jones, M.J., Marra, P.P. & Daszak, P. (2007) Genetic influences on mosquito feeding behavior and the emergence of zoonotic pathogens. *American Journal of Tropical Medicine and Hygiene*. 77 (4), 667–671.

Kilpatrick, A.M. (2011) Globalization, land use, and the invasion of West Nile virus. *Science*. 334 (6054), 323–327.

Kodandaramaiah, U. & Wahlberg, N. (2007) Out-of-Africa origin and dispersal-mediated diversification of the butterfly genus *Junonia* (Nymphalidae: Nymphalinae). *Journal of Evolutionary Biology*. 20 (6), 2181–2191

Kothera, L., Zimmerman, E.M., Richards, C.M. & Savage, H.M. (2009) Microsatellite characterization of subspecies and their hybrids in *Culex pipiens* complex (Diptera: Culicidae) mosquitoes along a North-South transect in the Central United States. *Journal of Medical Entomology*. 46 (2), 236–248.

Langella, O. (1999) *Populations 1.2.30*. . [Online]. 1999. Available from: http://bioinformatics.org/~tryphon/populations/. [Accessed: 1 May 2012]

Lapointe, D.A., Atkinson, C.T. & Samuel, M.D. (2012) Ecology and conservation biology of avian malaria. *Annals of the New York Academy of Sciences*. 1249 (1), 211–226.

Lorosa, E.S., Faria, M.S., De Oliveira, L.C.M., Alencar, J., & Marcondes, C.B. (2010) Blood meal identification of selected mosquitoes in Rio De Janeiro, Brazil. *Journal of the American Mosquito Control Association*. 26 (1), 18–23.

Lounibos, L.P. (2002) Invasions by insect vectors of human disease. *Annual Review of Entomology*. 47 (1), 233–266

McAbee, R.D., Green, E.N., Holeman, J., Christiansen, J., Frye, N., Dealey, K., Mulligan, F.S., Brault, A.C. & Cornel, A.J. (2008) Identification of *Culex pipiens* complex mosquitoes in a hybrid zone of West Nile virus transmission in Fresno County, California. *The American Journal of Tropical Medicine and Hygiene*. 78 (2), 303–310.

McBride, C.S. & Singer, M.C. (2010) Field studies reveal strong postmating isolation between ecologically divergent butterfly populations. *PLoS Biology*. 8 (10), e1000529.

Molaei, G., Andreadis, T.G., Armstrong, P.M., Anderson, J.F. & Vossbrinck, C. R. (2006) Host feeding patterns of *Culex* Mosquitoes and West Nile virus transmission, northeastern United States. *Emerging Infectious Diseases*. 12 (3), 468–474.

Molaei, G., Andreadis, T.G., Armstrong, P.M., Bueno Jr., R., Dennett, J. A., Real, S.V., Sargent, C., Bala, A., Randle, Y., Guzman, H., da Rosa, A.T., Wuithiranyagool, T. & Tesh, R.B. (2007) Host feeding pattern of *Culex quinquefasciatus* (Diptera: Culicidae) and its role in transmission of West Nile virus in Harris County, Texas. *The American Journal of Tropical Medicine and Hygiene*. 77 (1), 73–81.

Molaei, G., Cummings, R.F., Su, T., Armstrong, P.M., Williams, G.A., Cheng, M.-L., Webb, J.P. & Andreadis, T.G. (2010) Vector-host interactions governing epidemiology of West Nile virus in southern California. *The American Journal of Tropical Medicine and Hygiene*. 83 (6), 1269–1282.

Muñoz, A.G., Salazar, C., Castaño, J., Jiggins, C.D. & Linares, M. (2010) Multiple sources of reproductive isolation in a bimodal butterfly hybrid zone. *Journal of Evolutionary Biology*23 (6), 1312–1320.

Muturi, E.J., Muriu, S., Shililu, J., Mwangangi, J.M., Jacob, B. G., Mbogo, C., Githure, J. & Novak, R.J. (2008) Blood-feeding patterns of *Culex quinquefasciatus* and other culicines and implications for disease transmission in Mwea rice scheme, Kenya. *Parasitology Research*. 102 (6), 1329–1335.

Nei, M. (1987) *Molecular Evolutionary Genetics*. New York, Colombia University Press.

van Oosterhout, C., Hutchinson, W., Wills, D. & Shipley, P. (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*. 4535–538.

Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer applications in the biosciences: CABIOS*. 12 (4), 357–358.

Peel, M.C., Finlayson, B.L. & McMahon, T.A. (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*. 4 (2), 1633–1644.

Pinto, J., Sousa, C.A., Arez, A.P., Alves, J., Modiano, D., Petrarca, V., Charlwood, J.D. & do Rosário, V.E. (1999) Assessment of malaria transmission in an area with very low mosquito density. *Research and Reviews in Parasitology*. 59 (1-2), 23–26.

Pinto, M.A., Rubink, W.L., Patton, J.C., Coulson, R.N. & Johnston, J.S. (2005) Africanization in the United States: replacement of feral European honeybees (*Apis mellifera* L.) by an African hybrid swarm. *Genetics*. 170 (4), 1653–1665.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*. 155 (2), 945–959.

Rasgon, J.L. & Scott, T.W. (2003) *Wolbachia* and cytoplasmic incompatibility in the California *Culex pipiens* mosquito species complex: parameter estimates and infection dynamics in natural populations. *Genetics*. 165 (4), 2029–2038.

Ribeiro, H., Ramos, H.C., Capela, R.A. & Pires, C.A. (1980) *Os Mosquitos de Cabo Verde (Diptera: Culicidae) – Sistemática, distribuição, bioecologia e importância Médica*. Lisboa, Junta de Investigações Científica do Ultramar.

Ribeiro, H. & Ramos, H.C. (1995) *Guia ilustrado para a identificação dos mosquitos de Angola (Diptera: Culicidae).* Lisboa, Suplemento n^{o}4 ao Boletim Sociedade Portuguesa de Entomologia.

Ribeiro, H. & Ramos, H.C. (1999) Identification keys of the mosquitoes (Diptera: Culicidae) of Continental Portugal, Açores and Madeira. *European Mosquito Bulletin*. 3, 1–11.

Ribera, I., Bilton, D.T. & Vogler, A.P. (2003) Mitochondrial DNA phylogeography and population history of *Meladema* diving beetles on the Atlantic Islands and in the Mediterranean basin (Coleoptera, Dytiscidae). *Molecular Ecology*. 12 (1), 153–167.

Savage, H.M., Aggarwal, D., Apperson, C.S., Katholi, C.R., Gordon, E., Hassan, H.K., Anderson, M., Charnetzky, D., McMillen, L., Unnasch, E.A. & Unnasch, T.R. (2007) Host choice and West Nile virus infection rates in blood-fed Mosquitoes, including members of the *Culex pipiens* complex, from Memphis and Shelby County, Tennessee, 2002–2003. *Vector-Borne and Zoonotic Diseases*. 7 (3), 365–386.

Savage, H.M., Anderson, M., Gordon, E., McMillen, L., Colton, L., Delorey, M., Sutherland, G., Aspen, S., Charnetzky, D., Burkhalter, K. & Godsey, M. (2008) Host-seeking heights, host-seeking activity patterns, and West Nile virus infection rates for members of the *Culex pipiens* complex at different habitat types within the hybrid zone, Shelby County, TN, 2002 (Diptera: Culicidae). *Journal of Medical Entomology*. 45 (2), 276–288.

Savini, G., Monaco, F., Terregino, C., Di Gennaro, A., Bano, L., Pinoni, C., De Nardi, R., Bonilauri, P., Pecorari, M., Di Gialleonardo, L., Bonfanti, L., Polci, A., Calistri, P. & Lelli, R. (2011) Usutu virus in Italy: An emergence or a silent infection? *Veterinary Microbiology*. 151 (3–4), 264–274.

Shroyer, D.A. (1981) Establishment of *Wyeomyia mitchellii* on the island of Oahu, Hawaii. *Mosquito News*. 41 (4), 805–806.

Slatkin, M. & Excoffier, L. (1996) Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity*. 76 (4), 377–383.

Smith, 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 Hygiene*. 70 (4), 339–345.

Smith, J.L., Keyghobadi, N., Matrone, M.A., Escher, R.L. & Fonseca, D.M. (2005) Cross-species comparison of microsatellite loci in the *Culex pipiens* complex and beyond. *Molecular Ecology Notes*. 5 (3), 697–700.

Solomon, T. (2004) Flavivirus encephalitis. *The New England Journal of Medicine*. 351 (4), 370–378.

Steeves, T.E., Maloney, R.F., Hale, M.L., Tylianakis, J.M. & Gemmell, N.J. (2010) Genetic analyses reveal hybridization but no hybrid swarm in one of the world's rarest birds. *Molecular Ecology*. 19 (23), 5090–5100.

Subra, R. (1981) Biology and control of *Culex pipiens quinquefasciatus* Say, 1823 (Diptera, Culicidae) with special reference to Africa. *International Journal of Tropical Insect Science*. 1 (4), 319–338.

Tardei, G., Ruta, S., Chitu, V., Rossi, C., Tsai, T.F. & Cernescu, C. (2000) Evaluation of Immunoglobulin M (IgM) and IgG enzyme immunoassays in serologic diagnosis of West Nile virus infection. *Journal of Clinical Microbiology*. 38 (6), 2232–2239.

Trari, B., Dakki, M., Himmi, O. & El Agbani, M.A. (2002) Les moustiques (Diptera Culicidae) du Maroc. Revue bibliographique (1916-2001) et inventaire des espèces. *Bulletin de la Société de Pathologie Exotique*. 95 (4), 329–334.

Urbanelli, S., Silvestrini, F., Sabatinelli, G., Raveloarifera, F., Petrarca, V. & Bullini, L. (1995) Characterization of the *Culex pipiens* complex (Diptera: Culicidae) in Madagascar. *Journal of Medical Entomology*. 32 (6), 778-786.

Vähä, J.-P. & Primmer, C.R. (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*. 15 (1), 63–72.

Vieira, H. de S.-R. (1985) *Relatório do grupo de trabalho do Instituto de Higiene e Medicina Tropical (luta contra os culicídeos) em colaboração com a Brigada de Luta contra o Paludismo da República de Cabo Verde: período de 10 de Abril a 9 de Maio de 1985*.

Vieira, V., Diaz, S. & Báez, M. (2010) Diptera (other families). In: P. Borges, A. Costa, R. Cunha, R. Gabriel, V. Gonçalves, A.F. Martins, I. Melo, M., Parente, P., Raposeiro, P., Rodrigues, R.S, Santos, L., Silva, P., Vieira & V. Vieira (eds.). *A list of the terrestrial and marine biota from the Azores*. Cascais, Princípia. pp. 219–221.

Vinogradova, E.B. (2000) *Culex pipiens pipiens mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control*. S.I. Golovatch (ed.). Sofia, Pensoft.

Zinser, M., Ramberg, F. & Willott, E. (2004) *Culex quinquefasciatus* (Diptera: Culicidae) as a potential West Nile virus vector in Tucson, Arizona: Blood meal analysis indicates feeding on both humans and birds. *Journal of Insect Science*. 4 (1), 20.

Supporting information

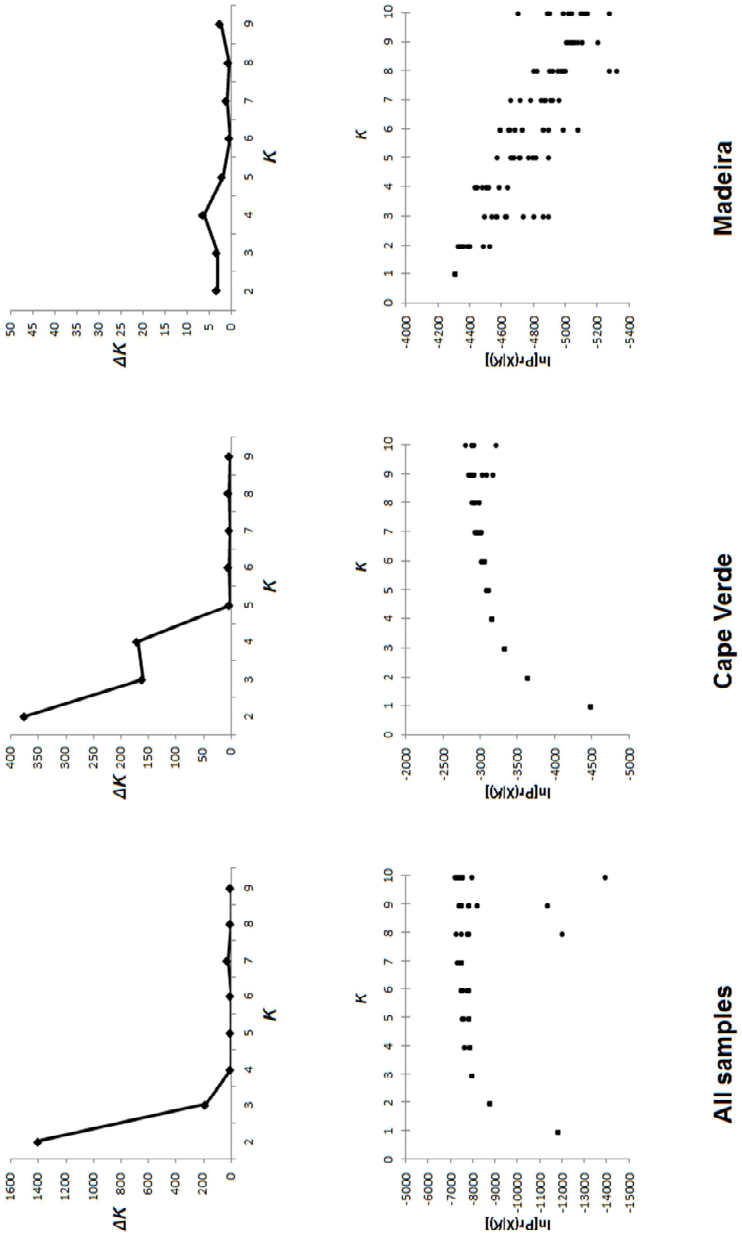

Figure S1. Graphics of *ad hoc* approaches to infer the number of clusters (*K*) in STRUCTURE analysis with all samples, Cape Verde and Madeira.

K: number of clusters; *ΔK*: see Evanno *et al.* (2005); ln[Pr(X|*K*)]: estimated log probability of the data under each *K*

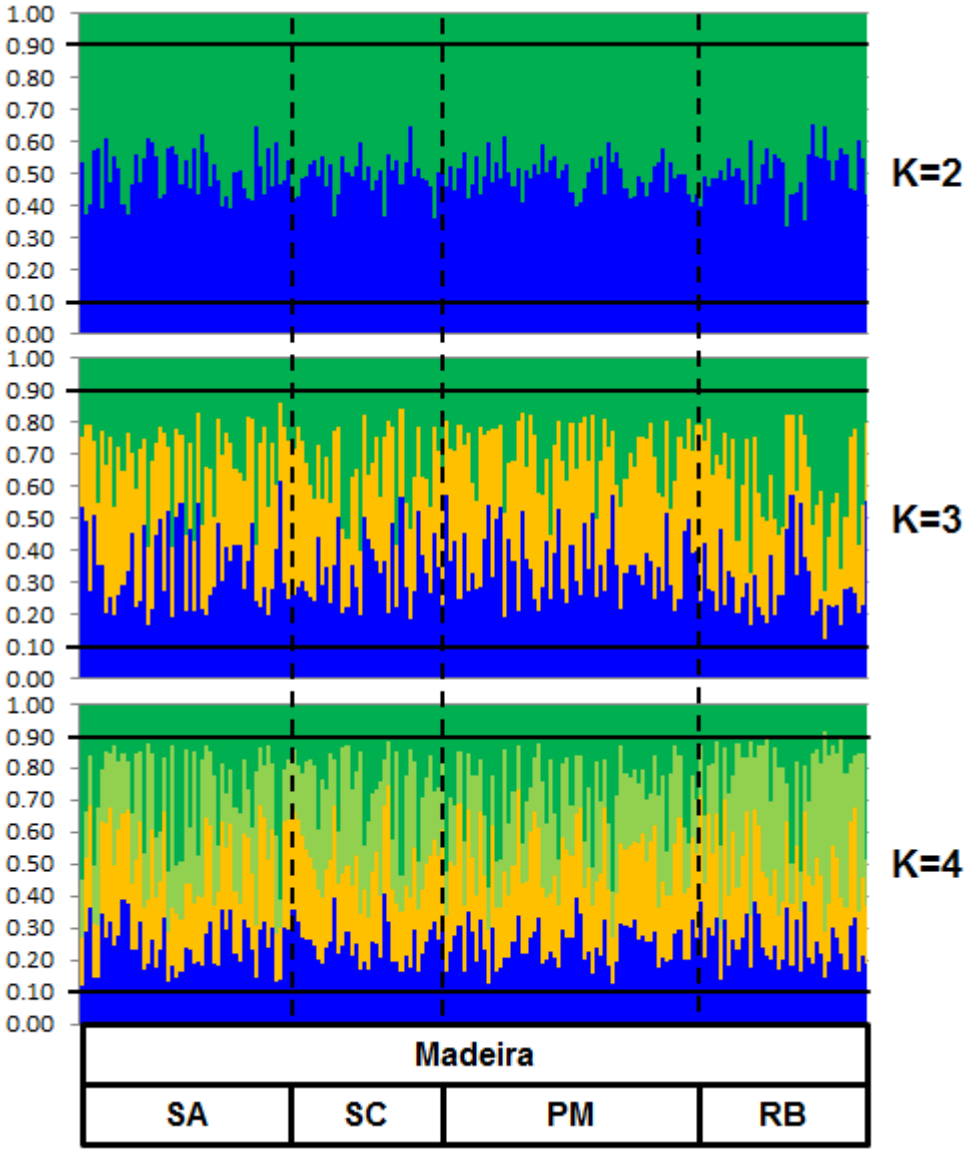

Figure S2. Bayesian cluster analysis conducted by STRUCTURE in Madeira.

K: number of clusters; SA: Santana; SC: Santa Cruz; PM: Porto Moniz; RB: Ribeira Brava. Columns correspond to the multilocus genotype of each individual, partitioned in different colours representing the probability of ancestry (q_i) to each cluster. Individuals were ordered according to their geographic information. Horizontal lines indicate the q_i threshold used to determine admixed individuals (see 'Material and Methods').

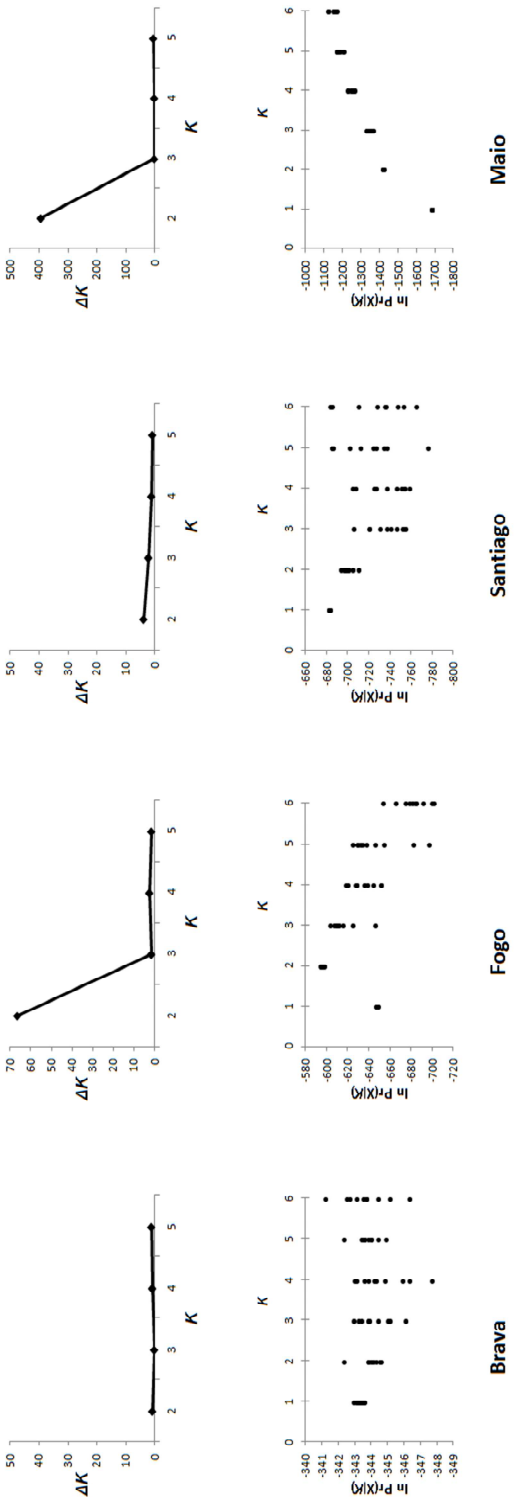

Figure S3. Graphics of *ad hoc* approaches to inference the number of clusters (*K*) in STRUCTURE analysis in each island of Cape Verde.

K: number of clusters; *ΔK*: see Evanno *et al.* 2005. Molecular Ecology 14: 2611-2620; ln[Pr(X|*K*)]: estimated log probability of the data under each *K*.

Table S1. Localities positive for *Culex pipiens* complex in Cape Verde.

Island	SS	locality	Type	PS	Coordinates	Altitude (meters)
Brava	7	Travessa	Rural	1	14º52'N 24º42'W	504
		Figueira Grande	Rural	1	14º52'N 24º42'W	565
Fogo	12	Patim	Rural	2	14º52'N 24º25'W	544
		Monte Largo	Rural	1	14º52'N 24º22'W	813
		Fonte Cabrito	Rural	1	14º51'N 24º19'W	657
Maio	18	Morro	Rural	1	15º11'N 23º13'W	21
		Vila de Maio	Urban	1	15º08'N 23º12'W	48
Santiago	131	Praia - Várzea	Urban	5	14º54'N 23º30'W	5
		Praia - Palmarejinho	Urban	2	14º54'N 23º31'W	5
		Praia – Eugénio Lima	Urban	2	14º54'N 23º30'W	5
		João Garrido	Rural	2	15º01'N 23º34'W	400*
		Achada Leite	Rural	1	15º07'N 23º45'W	20*
		Calhetona	Rural	1	15º10'N 23º35'W	44
		São Martinho Grande	Rural	1	14º55'N 23º34'W	50*

SS: total number of breeding sites sampled per island; PS: positive sites for *Cx. pipiens* sensu lato per locality; *Estimated from Google® Earth, remaining values were obtained from Global Positioning System measurements.

Table S2. Microsatellite loci used in the analysis.

Locus	Repeat	Primers	T_A (ºC)	Ref
CQ11	$(GT)_2ACTTC(GT)_9$	F: GATCCTAGCAAGCGAGAAC R: 6-fam-GAGCGGCCAAATATTGAGAC	52	[1]
CQ26	$(GTGTGTAT)_2+(GT)_{10}+(GT)_5$	F: TCCGACATGGGAAGAGCGCA R: 6-fam-ACGCGCCCTTCTTCTGCAAC	56	[1]
CQ41	$(GT)_{12}$	F: CTGCCACTGCCTGACTGAAA R: Hex-ACCACTCAGCAACATCCGGC	52	[1]
CxpGT4	$(GT)_5(GTTT)_2GC(GT)_2CT(GT)_5$	F: GTCGTCGCTAACCCTTGTT R: Ned-CGCGATAGTCGGTAATCGT	54	[2]
CxpGT9	$(GT)_{13}$	F: AATCTCCCCGTATAATTGTG R: Ned-TATAAGACCAGTGAAGCCAG	52	[2]
CxpGT12	$(TG)_{14}$	F: AACGTGAGCGTGATTGCTC R: 6-fam-CAGCTGTTGCACCAATGTC	54	[2]
CxpGT40	$(GT)_{15}$	F: CATCATCTGTCCACGATCC R: Hex-TTATGCAGTTGCTGTCATATCC	52	[2]
CxpGT46	(TG)15	F: Hex-CCGACACCGTGTTCAAAGAG R: TGACGACGACGGTACAAGAG	52	[2]
CxpGT51	$(TG)_4CG(TG)_{15}$	F: GAGTATCGCTCGTTGGAGATT R: Hex-ACCCTCTTTTCTTTCTATGTCTGT	54	[2]
CxqGT4	$(GT)_{12}$	F: ATAGAACTTGTTCGCCGTCTC R: 6-fam-TCTAAACACGCACCACGTACA	52	[3]
CxqGT6b	$(CA)_8$	F: CAACCAGCAAAACCCTCATC R: Ned-TAGCCGGGCAGATTCATTAC	54	[3]
CxqTri4	$(TGC)_7$	F: Hex-CTAGCCCGGTATTTACAAGAAC R: AACGCCAGTAGTCTCAGCAG	54	[3]

T_A: annealing temperature. Ref: References [1]: Fonseca *et al.* 1998. Molecular Ecology 7: 1613-1621. [2]: Keyghobadi *et al.* 2004. Molecular Ecology Notes 4: 20-22. [3]: Smith *et al.* 2005. Molecular Ecology Notes 5: 697–700.

Table S3. Genetic diversity at microsatellite loci of *Culex pipiens* complex from Macaronesian Islands.

Locus	Local	SC (N=34)	RB (N=39)	SA (N=51)	PM (N=66)	MAD (N=190)	B (N=31)	F (N=36)		M (N=63)		S (N=54)	CV (N=184)	Total (N=374)
	ace-2 assay	P (N=34)	P (N=39)	P (N=51)	P (N=66)		Q (N=31)	Q (N=22)	H (N=14)	Q (N=8)	H (N=42)	P (N=13)	Q (N=54)		
CQ11	$A_{R(16)}$	3.1	2.2	3.2	2.9	3.0	1.5	3.7	3.5	3.0	3.7	3.6	2.9	3.8	4.9
	H_e	0.536	**0.310***	**0.581***	**0.518***	**0.509**	0.063	0.429	0.452	0.700	0.594	0.525	0.458	**0.600**	**0.763**
	F_{IS}	0.337	0.681	**0.563**	**0.433**	**0.492**	-0.017	0.424	0.500	-0.077	-0.166	0.376	-0.133	**0.288**	**0.546**
CQ26	$A_{R(16)}$	4.2	4.5	4.5	3.0	4.0	3.0	3.9	3.6	4.0	5.6	4.8	3.4	5.9	7.4
	H_e	**0.604***	**0.723***	**0.622***	**0.523***	**0.610**	0.552	0.648	0.640	0.767	**0.773**	0.803	0.615	**0.791**	**0.850**
	F_{IS}	**0.693**	**0.518**	**0.456**	**0.708**	**0.591**	-0.232	0.018	0.226	-0.333	0.076	-0.157	-0.175	0.100	**0.435**
CQ41	$A_{R(16)}$	7.0	7.6	7.4	7.0	7.6	3.8	3.4	4.1	4.0	3.2	4.0	4.1	4.3	8.0
	H_e	**0.846***	**0.874***	**0.864***	**0.844***	**0.874**	0.635	**0.614**	0.593	0.708	0.542	0.462	0.706	**0.733**	**0.869**
	F_{IS}	**0.482**	**0.416**	**0.516**	**0.261**	**0.413**	0.036	0.339	0.161	0.125	-0.277	-0.175	0.162	**0.199**	**0.367**
CxpGT04	$A_{R(16)}$	3.5	4.3	3.4	4.1	3.9	1.3	3.2	4.0	3.0	3.8	3.8	1.1	3.7	5.4
	H_e	0.496	0.611	0.609	0.704	0.636	0.032	0.595	0.481	0.342	**0.688**	0.588	0.019	**0.474**	**0.764**
	F_{IS}	-0.008	-0.134	0.022	-0.020	-0.013	0.000	0.085	-0.195	-0.105	0.171	-0.049	0.000	**0.347**	**0.378**
CxpGT09	$A_{R(16)}$	4.5	4.7	5.0	4.6	4.8	1.0	1.9	1.8	3.0	2.7	2.0	1.8	3.3	6.0
	H_e	0.736	0.730	0.773	0.752*	0.756	NA	0.224	0.138	0.633	**0.456***	0.271	**0.123***	**0.580**	**0.827**
	F_{IS}	0.041	0.052	0.087	0.175	**0.109**	NA	0.782	1.000	0.622	**0.689**	-0.143	**0.851**	**0.858**	**0.522**
CxpGT12	$A_{R(16)}$	3.3	3.5	3.6	2.8	3.3	1.0	2.2	2.0	2.0	3.0	3.0	1.1	2.8	3.6
	H_e	0.590	0.565	0.639	0.553	0.584	NA	0.210	0.423	0.525	0.669	0.655	0.019	**0.412**	**0.669**
	F_{IS}	0.103	0.139	0.235	0.207	**0.180**	NA	0.357	-0.013	-0.207	-0.178	0.304	0.000	**0.281**	**0.418**
CxpGT40	$A_{R(16)}$	6.0	4.9	5.6	4.6	5.4	2.0	3.0	2.6	4.0	6.4	7.1	2.4	5.3	6.5
	H_e	0.785	0.738	0.785	0.668	0.745	**0.373***	0.357	0.548	0.742	**0.847**	**0.886**	0.530	**0.698**	**0.803**
	F_{IS}	0.036	0.063	-0.014	0.025	0.039	**0.744**	0.112	-0.600	-0.012	0.016	-0.043	0.147	**0.226**	**0.216**

Locus	Local	SC (N=34)	RB (N=39)	SA (N=51)	PM (N=66)	MAD (N=190)	B (N=31)	F (N=36)			M (N=63)		S (N=54)	CV (N=184)	Total (N=374)
	ace-2 assay	P (N=34)	P (N=39)	P (N=51)	P (N=66)		Q (N=31)	Q (N=22)	H (N=14)	Q (N=8)	H (N=42)	P (N=13)	Q (N=54)		
CxpGT46	$A_{R(16)}$	4.5	4.1	4.0	3.9	4.3	2.6	2.9	2.9	3.0	4.6	4.0	3.0	4.3	5.3
	H_e	0.673	0.667	**0.627***	0.586	**0.633**	0.392	0.532	0.558	0.433	**0.700**	**0.692**	0.494	**0.570**	**0.761**
	F_{IS}	0.131	0.273	0.338	-0.035	**0.167**	0.012	0.235	0.498	-0.167	-0.021	0.342	0.238	**0.186**	**0.349**
CxpGT51	$A_{R(16)}$	7.1	8.0	6.7	7.5	7.4	4.1	2.9	3.0	4.0	6.4	7.0	4.8	7.0	8.3
	H_e	0.817*	0.838	0.830	0.856	0.843	0.685	0.517	0.627	0.725	**0.830**	**0.846**	0.747	**0.844**	**0.873**
	F_{IS}	0.224	-0.010	0.007	-0.062	0.022	-0.229	-0.057	-0.509	-0.225	-0.208	0.373	0.009	**0.053**	**0.070**
CxqQGT4	$A_{R(16)}$	1.2	1.4	1.4	1.0	1.2	3.3	2.3	2.0	2.0	3.0	3.0	2.8	3.8	3.6
	H_e	0.029	0.051	0.058	NA	0.031	0.441	0.292	0.389	0.400	**0.638**	0.566	0.423	**0.534**	**0.601**
	F_{IS}	0.000	-0.013	-0.020	NA	-0.013	0.017	-0.148	0.085	-0.273	0.181	-0.379	0.109	**0.162**	**0.610**
CxqGT6B	$A_{R(16)}$	2.9	2.6	2.6	2.8	2.7	4.4	3.6	4.6	4.0	3.0	1.6	2.9	4.4	4.1
	H_e	0.510	0.505	0.384	0.496	0.472	0.757	0.511	0.730	0.617	**0.586**	0.077	0.408	**0.729**	**0.643**
	F_{IS}	-0.166	-0.095	0.031	-0.193	-0.112	-0.110	-0.162	-0.079	-0.235	-0.388	0.000	-0.229	**0.121**	**0.095**
CxqTRI4	$A_{R(16)}$	2.0	2.0	2.2	2.0	2.0	1.8	3.0	2.8	2.0	2.0	2.0	1.9	2.4	3.2
	H_e	0.506	0.506	0.441	0.490	0.488	0.178	0.623	0.561	0.325	0.496	0.369	0.124	**0.467**	**0.675**
	F_{IS}	-0.019	0.089	0.156	-0.083	0.036	-0.091	0.127	0.112	0.632	0.137	-0.263	-0.044	**0.337**	**0.421**
All loci	$A_{R(16)}$	4.1	4.2	4.1	3.9	4.1	2.5	3.0	3.1	3.2	4.0	3.8	2.7	4.3	5.5
	H_e	**0.594**	**0.593**	**0.601**	**0.636**	**0.599**	0.411	0.463	0.512	0.576	0.652	0.562	0.389	**0.619**	**0.758**
	F_{IS}	**0.183**	**0.169**	**0.219**	**0.122**	**0.177**	-0.017	0.144	0.033	-0.032	-0.011	0.051	0.043	**0.244**	**0.365**

$A_{R(16)}$*:* allelic richness for a minimum sample size of 16 genes; *He*: expected heterozygosity; F_{IS}: inbreeding coefficient; N: sample size; SC: Santa Cruz; RB: Ribeira Brava; SA: Santana; PM: Porto Moniz; MAD: Madeira island; B: Brava island; F: Fogo island; M: Maio island; S: Santiago island; CV: Cape Verde; P: *Cx. pipiens*; Q: *Cx. quinquefasciatus*; H: hybrids (*ace*-2 based identification). In bold: significant p-values for H-W tests (heterozygote deficit) after correction for multiple tests. Asterisks indicate presence of null alleles determined by Micro-Checker. Per locus and over samples H-W tests were performed in ARLEQUIN. For over loci estimates the global test available in FSTAT was used.

Chapter 3.

Asymmetric introgression between sympatric molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in the Comporta region, Portugal

Published as:

Gomes, B., Sousa, C.A., Novo, M.T., Freitas, F.B., Alves, R., Corte-Real, A.R., Salgueiro, P., Donnelly, M., Almeida, A.P. & Pinto, J. (2009) Asymmetric introgression between sympatric molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in the Comporta region, Portugal. *BMC Evolutionary Biology*. 9, 262.

Abstract

Background

Culex pipiens L. is the most widespread mosquito vector in temperate regions. This species consists of two forms, denoted molestus and pipiens, that exhibit important behavioural and physiological differences. The evolutionary relationships and taxonomic status of these forms remain unclear. In northern European latitudes molestus and pipiens populations occupy different habitats (underground *vs.* aboveground), a separation that most likely promotes genetic isolation between forms. However, the same does not hold in southern Europe where both forms occur aboveground in sympatry. In these southern habitats, the extent of hybridisation and its impact on the extent of genetic divergence between forms under sympatric conditions has not been clarified. For this purpose, we have used phenotypic and genetic data to characterise *Cx. pipiens* collected aboveground in Portugal. Our aims were to determine levels of genetic differentiation and the degree of hybridisation between forms occurring in sympatry, and to relate these with both evolutionary and epidemiological tenets of this biological group.

Results

Autogeny and stenogamy was evaluated in the F1 progeny of 145 individual *Cx. pipiens* females. Bayesian clustering analysis based on the genotypes of 13 microsatellites revealed two distinct genetic clusters that were highly correlated with the alternative traits that define pipiens and molestus. Admixture analysis yielded hybrid rate estimates of 8-10%. Higher proportions of admixture were observed in pipiens individuals suggesting that more molestus genes are being introgressed into the pipiens form than the opposite.

Conclusion

Both physiological/behavioural and genetic data provide evidence for the sympatric occurrence of molestus and pipiens forms of *Cx. pipiens* in the study area. In spite of the significant genetic differentiation between forms, hybridisation occurs at considerable levels. The observed pattern of asymmetric introgression probably relates to the different mating strategies adopted by each form. Furthermore, the differential

introgression of molestus genes into the pipiens form may induce a more opportunistic biting behaviour in the latter thus potentiating its capacity to act as a bridge-vector for the transmission of arboviral infections.

Background

The *Culex pipiens* complex includes two of the most ubiquitous mosquito species in the world, *Culex quinquefasciatus* in tropical and subtropical regions, and *Culex pipiens* L. 1958 in temperate regions. The nominal species of the complex, *Cx. pipiens s.s.* comprises two distinct forms, denoted pipiens and molestus, that are morphologically indistinguishable but exhibit important behavioural and physiological differences. The molestus form is stenogamous (mates in confined spaces, *i.e.* < 0.1 m^3; Clements, 1999), autogenous (can oviposit without a blood meal), mammophilic (prefers to feed on mammals, including humans) and homodynamic (remains active during winter). In contrast, the pipiens form is eurygamous (mates in open spaces), anautogenous (oviposition requires a blood meal), ornithophilic (feeds predominantly on birds) and heterodynamic (undergoes winter diapause) (Harbach *et al.*, 1984, 1985). In the northern regions of Europe, Russia and USA, molestus and pipiens forms occupy different habitats, underground and aboveground, respectively (Byrne & Nichols, 1999; Vinogradova, 2000; Huang *et al.*, 2008).

The taxonomic status and evolutionary relationships of these forms remain controversial. One hypothesis is that the molestus form derives from surface pipiens populations that have undergone local adaptation to underground conditions (Byrne & Nichols, 1999). Another hypothesis is that these forms may represent two distinct genetic entities (Fonseca *et al.*, 2004). Under the latter scenario, underground populations from northern Europe would have derived from southern autogenous populations that have subsequently dispersed and colonised underground habitats (Fonseca *et al.*, 2004; Kent *et al.*, 2007). If in northern regions a physical discontinuity (underground *vs.* surface) is likely to significantly reduce gene flow between molestus and pipiens, hence promoting genetic isolation, the same may not hold for southern regions, where both autogenous and anautogenous populations co-occur in surface habitats (Harbach *et al.*, 1984, 1985; Chevillon *et al.*, 1995). Moreover, individuals with

hybrid genetic signatures between molestus and pipiens have been described both in the USA and in southern Europe (Fonseca *et al.*, 2004; Kilpatrick *et* al., 2007; Huang *et al.*, 2008). These results agree with reports of hybridisation between forms that result in hybrid females with intermediate physiological and behavioural traits (Chevillon *et al.*, 1995; Spielman, 2001). Hybrids between molestus and pipiens forms are considered of great epidemiological importance. They can readily feed on both avian and mammalian hosts, including humans. This opportunistic biting behaviour will potentiate the role of *Cx. pipiens* as a bridge-vector for the transmission of arboviruses such as West Nile virus (WNV), from their amplification hosts (birds) to humans (Fonseca *et al.*, 2004; Hamer *et al.*, 2008).

Despite the conspicuous behavioural and physiological differences between molestus and pipiens, analysis of molecular markers revealed overall shallow genetic divergence and a paucity of diagnostic fixed differences between forms (Vinogradova & Shaikevich, 2007; Kent *et al.*, 2007). Exceptions are the contrasting differences in the degree of polymorphism found in the SH60 locus, a *Cx. pipiens* specific fragment originally described by Crabtree and co-workers (1997) to distinguish this species from its tropical sibling *Cx. quinquefasciatus*, and the significant differentiation detected by analysis of microsatellites (Fonseca *et al.*, 2004; Kent *et al.*, 2007). The most promising diagnostic marker so far obtained is a sequence difference in the flanking region of microsatellite CQ11, hereafter termed CQ11FL, that allows PCR-based discrimination of molestus, pipiens and putative hybrids (Bahnck & Fonseca, 2006)

In Portugal, *Cx. pipiens* is the most widespread mosquito species, reaching the highest densities in coastal estuarine areas during summer (Almeida *et al.*, 2008). Some of these areas are important sanctuaries for migratory birds and hence potential sites for the introduction of arbovirus (Rappole & Hubálek, 2003). In the summer of 2004, WNV was isolated from *Cx. pipiens* collected in the southern province of the Algarve, in a mosquito survey that followed the description of two cases of WNV fever acquired by Irish bird-watchers in the region (Connell *et al.*, 2004; Esteves *et al.*, 2005). In Portugal, autogenous/stenogamous *Cx. pipiens*, typical of the molestus form, have been described from the analysis of larvae collected in urban surface habitats (Ribeiro *et al.*, 1983). However, there is currently no information on the extent of genetic isolation between

molestus and pipiens forms when they co-occur sympatrically in southern European aboveground habitats.

In this study, we used the CQ11FL marker and microsatellite loci to analyse samples of *Culex pipiens* collected aboveground in the estuarine region of Comporta in order to: i) determine levels of differentiation between samples displaying behavioural and physiological characteristics of pipiens and molestus forms; ii) assess the degree of hybridisation between forms and relate this with the potential for arbovirus transmission in the area.

Results

Autogeny, stenogamy and molecular identification

A total of 145 F1 families were analysed in the insectary to determine autogeny and stenogamy (Table 1). Of these, 115 (79.3%) were able to lay a first egg batch without blood feeding, hence being considered autogenous. The great majority of autogenous families (109 out of the 115) laid the first egg batch within two days after the emergence of the last adult. In the remaining 30 families (20.7%), oviposition occurred only after blood feeding in 11 (36.7%) and no oviposition was seen in the other 19 (63.3%) during the 10 days of the experiment. For subsequent comparisons, these families were put together into a single group denoted as non-autogenous.

Table 1. Autogeny and insemination rates in *Culex pipiens* from Comporta, Portugal.

	INS =0%		**0%< INS <100%**		**INS =100%**		**Total**
Autogenous	1	(0.9)	30	(26.1)	84	(73.0)	115
Non-autogenous	22	(73.3)	8	(26.7)	0	(0.0)	30
Total	23	(15.9)	38	(26.2)	84	(57.9)	145

INS: proportion of inseminated females in each family.

There were significant associations of autogenous families with complete insemination and of non-autogenous families with absence of insemination (χ^2 = 100.7, d.f. = 2, P < 0.001; Table 1). In the autogenous group, the mean proportion of inseminated females was 92.9%, with 84 families (73.0%) showing 100% of inseminated females. There was a single autogenous family in which insemination was

not observed. This family oviposited without blood feeding only after the two-days period from the emergence of the last adult, after which the family was subdivided (see Methods). In this family, the level of insemination could have been too low to accurately determining the insemination rate by observing the spermathecae, but also the possibility of a parthenogenic egg batch cannot be excluded (Vinogradova, 2000). In contrast, the non-autogenous group had a mean proportion of inseminated females of 4.1% and no inseminated females were observed in 22 (73.3%) families. The remaining 8 inseminated families all laid eggs but only after blood feeding. The frequency distribution of insemination rates was bimodal, with most of the observations concentrating in the extreme values (Figure 1). More than 91% of the autogenous families had insemination rates above 80% whereas over 93% of the non-autogenous families had insemination rates below 20%.

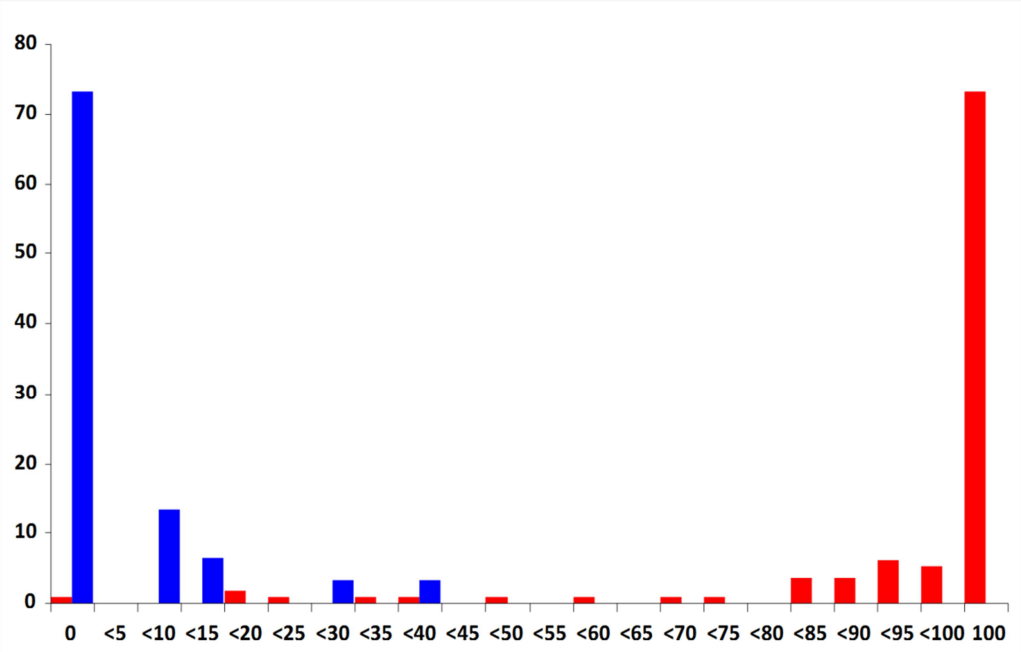

Figure 1. Frequency distribution of insemination rates in autogenous and non-autogenous families of *Culex pipiens*.

X-axis: proportion of inseminated females in each family at intervals of 5%. Y-axis: proportion of families (in percentage). Blue bars: non-autogenous; Red bars: autogenous.

A total of 145 females were molecularly analysed, representing one female per family. Of these, 134 (92.4%) were identified as *Cx. pipiens s.s.* by *ace-2* PCR (Smith & Fonseca, 2004). For the remaining 11 females no amplified product was obtained despite several attempts changing PCR conditions, possibly due to alterations in the

primers binding site. The families of these specimens were identified as belonging to *Cx. pipiens s.s.* by the observation of the genitalia of male siblings (Ribeiro & Ramos, 1999).

The genotypic frequencies of the CQ11FL marker are shown in Table 2. Overall, 78 (53.8%) females were homozygous for the 250 bp allele characteristic of the molestus form and 41 (28.3%) were homozygous for the 200 bp allele associated with the pipiens form. The remainder 26 (17.9%) females were heterozygous. There were significant associations between homozygous genotypes and alternative phenotypic traits. The "pipiens" genotype (CQ11FL200/200) predominated in non-autogenous and strictly non-stenogamous families (*i.e.* proportion of inseminated females = 0%) whereas the "molestus" genotype (CQ11FL250/250) was predominant in autogenous and strictly stenogamous families (*i.e.* proportion of inseminated females = 100%).

Table 2. Polymorphism at the flanking region of microsatellite CQ11 (CQ11FL) according to phenotypic groups of *Culex pipiens.*

	Total	**Autogeny**		**Insemination rates**		
		Autogenous	**Non-autogenous**	**INS =100%**	**0%< INS <100%**	**INS =0%**
CQ11FL250/250	78 (53.8)	77 (67.0)	1 (3.3)	60 (71.4)	17 (44.7)	1 (4.3)
CQ11FL200/250	26 (17.9)	22 (19.1)	4 (13.3)	16 (19.0)	7 (18.4)	3 (13.0)
CQ11FL200/200	41 (28.3)	16 (13.9)	25 (83.3)	8 (9.5)	14 (36.8)	19 (82.6)
Chi-square tests		$\chi 2$=58.9, d.f.=2, P<0.001		$\chi 2$ =51.7, d.f.=4, P<0.001		

INS: proportion of inseminated females in a family. Values in parenthesis refer to the relative genotypic frequencies (in percentage) within each phenotypic group. $\chi 2$: P-values of chi-square tests of homogeneity of genotypic frequencies among phenotypes.

Microsatellite analysis

Genetic diversity estimates for the 14 microsatellite loci analysed are shown in Table S1, available in the Additional File 1. Apart from the whole sample (N = 145), calculations were also made for subsamples determined by genotypes at the CQ11FL locus. Although coincidence of genotypes and phenotypes was not absolute, the significant associations between CQ11FL homozygous genotypes and alternative phenotypes justified this tentative partitioning. Diversity estimates were lower in

CQ11FL250/250 homozygotes (mean A_R = 6, mean H_e = 0.600) when compared to CQ11FL200/200 homozygotes (mean A_R = 11, mean H_e = 0.762). These differences were significant for both parameters (Wilcoxon signed-ranks tests; A_R: P = 0.001, H_e: P = 0.004). Microsatellite CQ11 was polymorphic in CQ11FL200/200 homozygous and in CQ11FL200/250 heterozygous groups. In contrast, this locus was nearly fixed for a 286 bp allele (f = 0.984) in the CQ11FL250/250 homozygous group. This allele was also the most frequent in the heterozygous group (f = 0.480) while it was absent in CQ11FL200/200 homozygotes.

Significant departures from Hardy-Weinberg proportions were detected in 10 loci (78.6%) when all specimens were analysed as a single sample (Table S1). Significant departures were seen at the same loci when analysis was repeated with pooled CQ11FL250/250 and CQ11FL200/200 homozygous specimens, *i.e.* when CQ11FL200/250 heterozygotes were excluded (data not shown). These departures were generally associated with significant positive F_{IS} values indicative of a heterozygote deficit (Table S1). However, when the sample was subdivided according to CQ11FL genotypes, significant heterozygote deficits were observed only in seven occasions (16.7% out of 42 tests). Of these, locus CxpGT9 exhibited heterozygote deficits in all three subsamples, possibly reflecting locus-specific effects such as null alleles or selective pressures. There was also one significant departure that resulted from heterozygous excess, namely for locus CQ11 in the CQ11FL200/250 heterozygous group.

Exact tests of linkage disequilibrium revealed 62 (68.1%) significant associations between pairs of loci out of 91 tests performed for the whole sample. When each form was treated in separate, significant associations were reduced to 12 in the CQ11FL250/250 homozygous group, four in CQ11FL200/200 homozygous and one in CQ11FL200/250 heterozygous. Of the total 17 significant tests detected in the subsamples nine involved locus CxpGT9, that also showed significant heterozygote deficits. This locus was therefore excluded from subsequent analyses.

Bayesian clustering analysis implemented by STRUCTURE (Pritchard *et al.*, 2000) revealed two (K = 2) genetically distinct ancestry clusters (Figure 2A). Cluster 1 grouped 96 specimens, 70 (72.9%) of which had a homozygous CQ11FL250/250

genotype and seven (7.3%) were CQ11FL200/200 homozygotes. Interestingly, all 96 specimens assigned to cluster 1 belonged to autogenous families, with nearly 80% of these having 100% insemination rates and with all families displaying at least some proportion of inseminated females, thus providing support for cluster 1 to represent the molestus form (Table 3). In contrast, cluster 2 was representative of the pipiens form, with 30 (83.3%) out of the 36 specimens assigned presenting a CQ11FL200/200 homozygous genotype and only two (5.6%) were CQ11FL250/250 homozygotes. In this cluster, 75% of females belonged to non-autogenous families and nearly 65% were from families with no insemination. None of the females assigned to cluster 2 belonged to families with 100% insemination. Very similar results were obtained when microsatellite CQ11, which exhibited the highest allelic differences between CQ11FL genotypes, was removed from the analysis (Figure 2B). With the exception of three individuals, all the remaining 142 (98%) specimens were assigned in to the same clusters as in the previous analysis, indicating that subdivision was not locus-dependent.

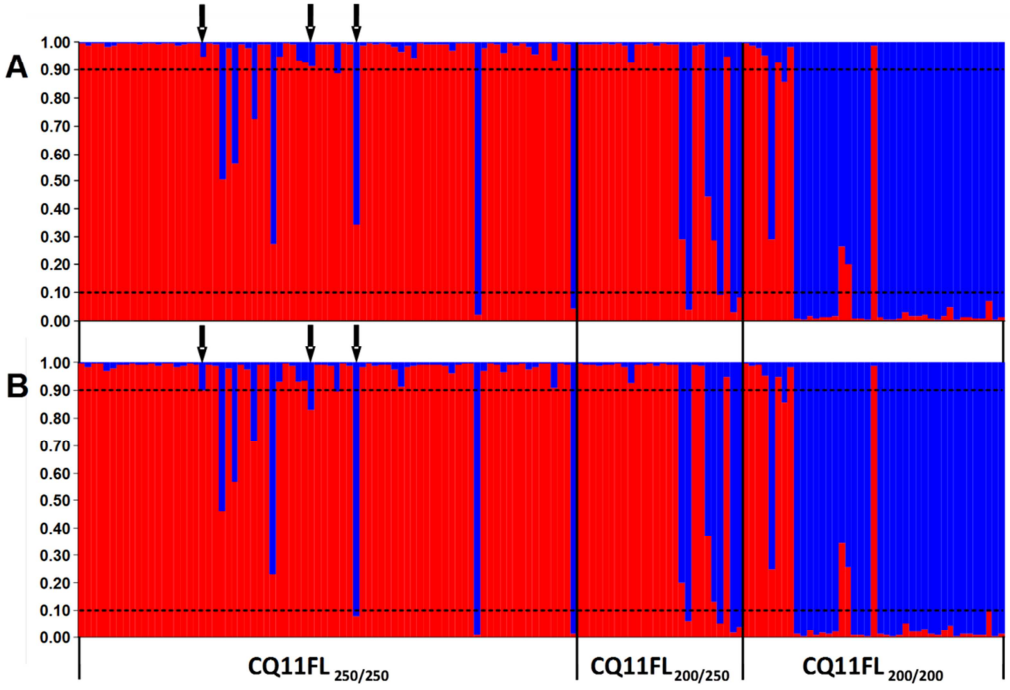

Figure 2. Bayesian cluster analysis conducted by STRUCTURE (Pritchard *et al.*, 2000).

Columns correspond to the multilocus genotype of each individual, partitioned in two colours representing the probability of ancestry (q_i) to each cluster. Red: cluster 1 (molestus); blue: cluster 2 (pipiens). Individuals were ordered according to their genotype at the CQ11FL locus. Dashed lines indicate the *qi* threshold used to determine admixed individuals (see Methods). A: analysis performed with 13 loci; B: analysis performed without locus CQ11. Arrows indicate individuals with different assignment between analyses.

There were 13 (9.0%) individuals of the total sample (N = 145) exhibiting an admixed ancestry (*i.e.* $q_i \geq 0.10$ for both clusters). Of these, only 3 (23.1%) had a heterozygous CQ11FL200/250 genotype while the majority (76.9%) were homozygous for either of the two alleles present at the CQ11FL locus. Regarding phenotypes, the proportion of admixed individuals was lower in families that displayed alternative extreme traits (*i.e.* autogenous with 100% insemination and non-autogenous with no insemination: 8 out of 106 or 7.6%) when compared to the remaining families that were either autogenous or non-autogenous with a varying proportion of insemination above 0% and below 100% (5 out of 39 or 12.8%).

Table 3. Frequencies (in percentage) of genotypes at the CQ11FL locus and phenotypes for autogeny and insemination rates in each of the ancestry clusters revealed by STRUCTURE (Pritchard *et al.*, 2000).

	N	CQ11FL genotype			Autogeny		Insemination rate		
		250/250	200/250	200/200	autogenous	non-autogenous	INS=0%	0%<INS<100%	INS=100%
Cluster 1	96	72.9	19.8	7.3	100.0	0.0	0.0	20.8	79.2
Admixed	13	46.2	23.1	30.8	76.9	23.1	0.0	38.5	61.5
Cluster 2	36	5.6	11.1	83.3	25.0	75.0	63.9	36.1	0.0

INS: proportion of inseminated females in each family.

The microsatellite allele frequency arrays together with estimates of allele richness (A_R) and private allele richness (pA_R) for the clusters representative of the molestus and pipiens forms are shown in Figure 3. Allelic diversity was higher in the pipiens cluster, with a mean A_R of 10 compared to a mean estimate of 6 for the molestus cluster. Most but not all of the alleles found in the molestus cluster were also represented in the pipiens cluster. In the molestus cluster pA_R estimates per locus varied from 0 to 3 (mean = 1) whereas in the pipiens cluster pA_R ranged from 1 to 12 (mean = 6). The pipiens and molestus clusters shared the most frequent allele at only four loci. For the remainder 9 loci, the most frequent alleles at each cluster were separated from each other on average by 8 base pairs, or four mutational steps (range: 2-12) as expected from their dinucleotide repeat constitution. The most remarkable difference was found in CQ11, with the most frequent alleles of pipiens and molestus being separated by 12 mutational steps.

Heterozygosity tests provided no evidence of recent population contraction in both molestus and pipiens clusters (Table 4). There was a single departure from

mutation drift equilibrium (MDE) in the pipiens cluster, that resulted from an apparent heterozygote deficiency ($P_{He<Heq}$ = 0.003) suggestive of population expansion and under the strict stepwise mutation model (SMM).

Table 4. Results of heterozygosity tests (Cornuet & Luikart, 1996) for molestus and pipiens clusters of *Cx. pipiens.*

		SMM	TPM (10%)	TPM (20%)	TPM (30%)
Cluster 1 (molestus)	$H_e>H_{eq}$	4	6	8	9
	$P_{He\neq Heq}$	0.027	0.685	0.736	0.497
Cluster 2 (pipiens)	$H_e>H_{eq}$	2	3	6	8
	$P_{He\neq Heq}$	0.005	0.057	0.340	0.893

$H_e>H_{eq}$: number of loci in which expected heterozygosity estimated from allele frequencies (H_e) was higher than the estimate obtained from the number of alleles and sample size under MDE (H_{eq}). $P_{He \neq Heq}$: *P*-values of Wilcoxon tests to detect if H_e differs from H_{eq} in a significant number of loci. SMM: stepwise mutation model. TPM: two-phase model. In bold: significant *P*-value after correction for multiple testing by the sequential Bonferroni procedure.

A global F_{ST} of 0.104 was obtained when subsamples were arranged according to the assignment into ancestry clusters revealed by STRUCTURE (Pritchard *et al.*, 2000; *i.e.* cluster 1, cluster 2 and admixed). The comparison between cluster 1 (molestus) and cluster 2 (pipiens) yielded a significant F_{ST} of 0.127. Differentiation was generalised, in that significant F_{ST} values were observed in 12 out of the 13 loci analysed, as shown in Table S2 of the Additional File 1. The single exception was locus CxqGT4, that was nearly monomorphic for the same allele in both forms (Figure 3). Locus CQ11 exhibited the highest F_{ST} value (0.405) compared to the remaining loci (0.002-0.272). Excluding this locus from analysis resulted in a decrease of the overall F_{ST} between molestus and pipiens to 0.103. Similar results were obtained with the R_{ST} estimator (Table S2). In comparisons between molestus and pipiens, R_{ST} was higher than F_{ST} in 6 out of 13 loci and the mean over-loci estimates were also higher, with (R_{ST} = 0.191) and without locus CQ11 (R_{ST} = 0.123).

The results of the admixture analysis performed by NEWHYBRIDS (Anderson & Thompson, 2002) on simulated genotypes generated by HYBRIDLAB (Nielsen *et al.*, 2006) are shown in Figure 4 and in Table S3 of the Additional File 1. Maximum accuracy was achieved for all *Tq* but there were variations in power. All parental individuals were correctly identified at *Tq* = 0.70 (minimum q_i = 0.724). At this

threshold, 93% of F1 hybrids were correctly assigned. Maximum power (*i.e.* 100% correct assignment) was obtained for this class at a $Tq = 0.60$.

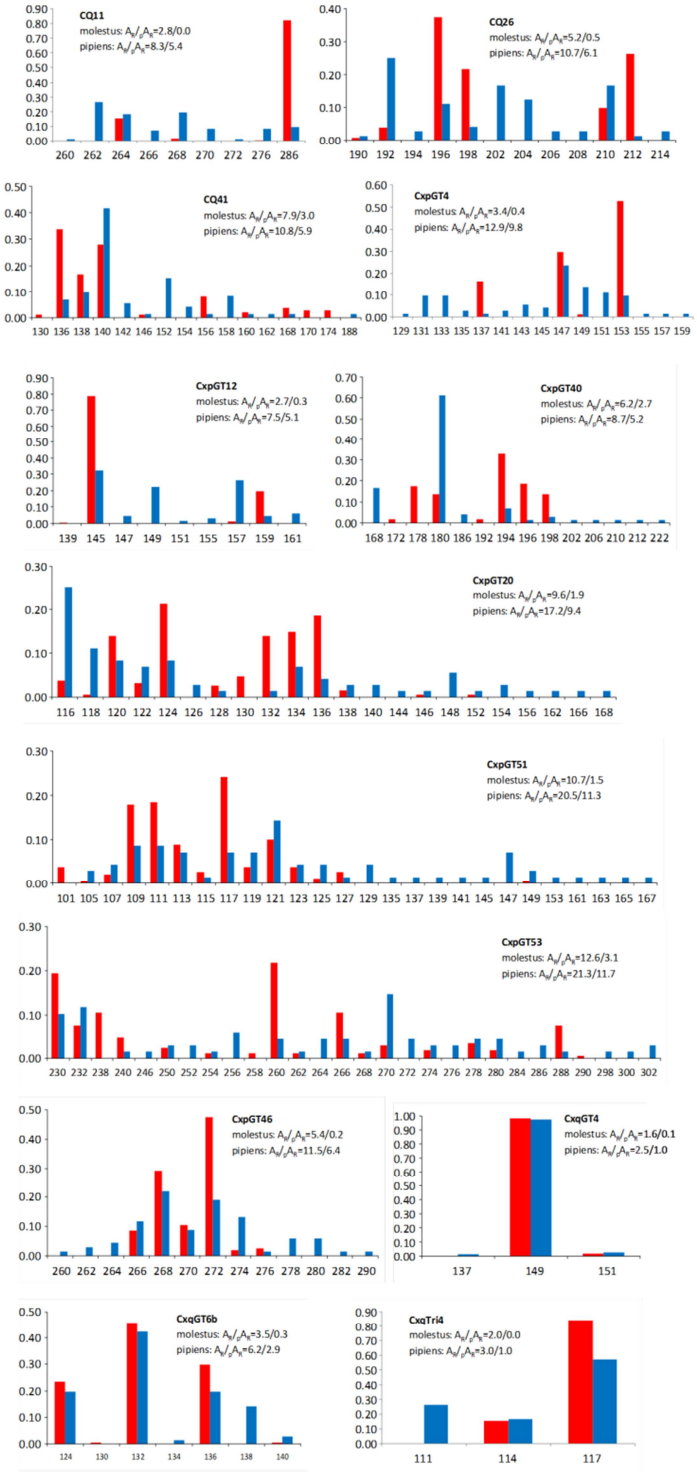

Figure 3. Microsatellite allele richness and frequency in *Culex pipiens* of Comporta, Portugal.

Allele frequencies, allele richness (A_R) and private alleles richness (pA_R) were calculated for samples of common ancestry determined by Bayesian clustering analysis using STRUCTURE. Red: molestus cluster, blue: pipiens cluster. X-axis: allele sizes in basepairs. Y-axis: allele frequency.

The analysis performed poorly in the assignment of the remaining hybrid classes, with proportions of correctly assigned individuals below 85% regardless of *Tq*. Given this poor performance, posterior probabilities of hybrid classes were summed and used as an estimate for the detection of hybrids but without definition of their admixture ancestry (Figure 4B). For this category, maximum power was achieved only for *Tq* = 0.50. Based on these results, thresholds of 0.50 and 0.70 were used for the detection of hybrids on the real dataset.

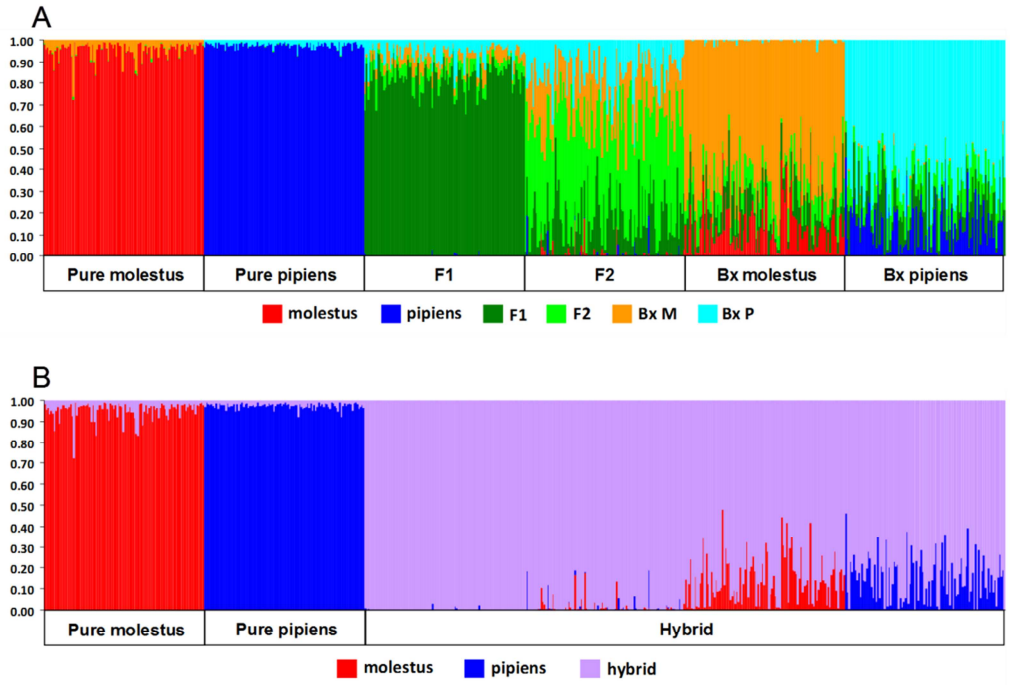

Figure 4. Bayesian assignment of simulated purebred and hybrid individuals obtained by HYBRIDLAB (Nielsen *et al.*, 2006), as implemented by NEWHYBRIDS (Anderson & Thompson, 2002).

Pure molestus, pure pipiens and hybrid (F1, F2 and backcrosses with each parental line) simulated individuals were generated from the genotypes of *Cx. pipiens* specimens that displayed by NEWHYBRIDS a q_i>0.90 of being pure molestus (N = 100) and pure pipiens (N = 11). Simulations were done using HYBRIDLAB to produce 100 simulated individuals for each class. Each vertical line represents a simulated individual. Lines are partitioned in colours according to the probabilities of assignment to each class. A: probabilities were obtained for each of the six classes. B: the "hybrid" category is the sum of probabilities of assignment to each of the four hybrid classes originally simulated.

All individuals with a molestus ancestry (N = 96) revealed by STRUCTURE (Pritchard *et al.*, 2000) were assigned to the same purebred class by NEWHYBRIDS (Anderson & Thompson, 2002) with probabilities of assignment close to 1 (minimum q_i = 0.927, Figure 5). In addition, five individuals of admixed ancestry were also included in this class. In contrast, of the 36 specimens with pipiens ancestry, only 26 (72.2%)

displayed a $q_i \geq 0.50$ of being assigned as parental pipiens (minimum $q_i = 0.510$). At $Tq = 0.70$ this number decreased to 19 (52.8%) with a minimum $q_i = 0.706$. The individual probabilities of assignment into the parental pipiens class were lower than those of purebred molestus. For individuals assigned as parental pipiens, the average proportion of assignment into a different class (*i.e.* molestus and/or hybrid) was 0.144 for $Tq = 0.70$ and 0.218 for $Tq = 0.50$.

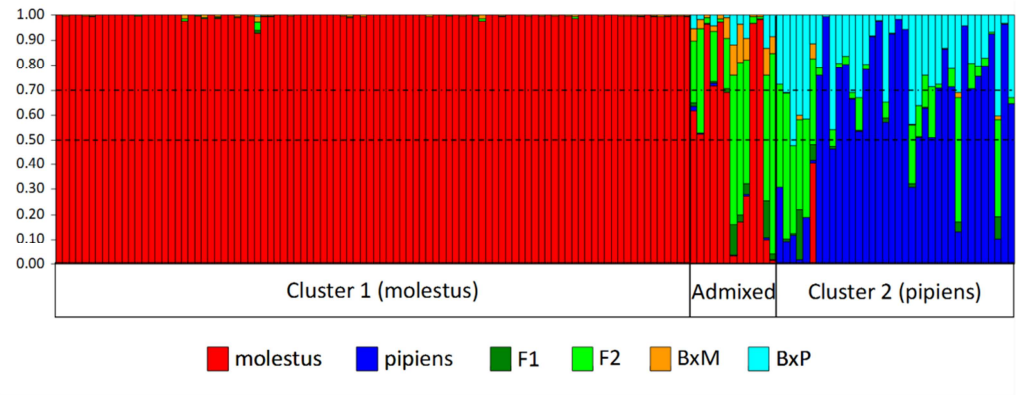

Figure 5. Bayesian assignment of individuals into pure and hybrid classes as implemented by NEWHYBRIDS (Anderson & Thompson, 2002).

Each column represents an individual analysed and it is partitioned into colours according to the probability of assignment to each of the six classes denoted in the label (pure molestus, pure pipiens, F1 hybrid, F2 hybrid, BxM: backcross with molestus, BxP: backcross with pipiens). Individuals were arranged according to their probability of ancestry obtained by STRUCTURE analysis. Dashed lines highlight the two probability thresholds (*Tq*) used to assign individuals into classes (see Methods).

Depending on the threshold, the proportion of hybrid individuals detected by NEWHYBRIDS (Anderson & Thompson, 2002) varied between 7.6% ($Tq = 0.70$) and 10.3% ($Tq = 0.50$), values comparable to the 9.0% proportion obtained by STRUCTURE (Pritchard *et al.*, 2000) analysis (Table 5).

Table 5. Proportions of pure and admixed *Culex pipiens* individuals inferred by Bayesian assignment methods implemented by STRUCTURE (Pritchard *et al.*, 2000) and NEWHYBRIDS (Anderson & Thompson, 2002).

	molestus		**admixed**		**pipiens**	
Structure	96	(66.2)	13	(9.0)	36	(24.8)
Newhybrids ($Tq = 0.50$)	104	(71.7)	15	(10.3)	26	(17.9)
Newhybrids ($Tq = 0.70$)*	101	(69.7)	11	(7.6)	19	(13.1)

* At this threshold, only 131 specimens were assigned to classes. The remainder 14 analysed individuals presented $q_i<0.70$ of belonging to any of the classes and were thus undetermined.

Discussion

Insectary experiments based on the progeny of field-caught *Cx. pipiens* females revealed strong associations between alternative traits that define molestus and pipiens forms. The highest proportions of inseminated females were seen in autogenous families. These two associated traits are expected for an autogenous/stenogamous molestus population. Conversely, non-autogenous families exhibited the lowest insemination rates suggesting that these families represent the anautogenous/eurygamic pipiens population. The non-autogenous group included families that oviposited after a blood meal and those in which no oviposition was detected throughout the experiment. Factors such as poor adaptation to insectary conditions causing gonotrophic dissociation could have resulted in the absence of oviposition in families that otherwise could in fact be autogenous. On the other hand, low insemination rates could also determine the lack of oviposition. Coincidently, no inseminated females were detected in all the 19 families that did not oviposit after blood feeding. Under the experimental conditions used, absence of insemination reflects the inability of mating in confined spaces, a trait of the pipiens form.

The observed phenotypic separation was confirmed by microsatellite analysis. Extensive heterozygote deficits and linkage between loci were detected when all individuals were treated as a single sample. These departures were greatly reduced when the sample was tentatively subdivided into subsamples defined by the CQ11FL locus, a single-locus marker available to distinguish molestus and pipiens forms (Bahnck & Fonseca, 2006). The Bayesian method of Pritchard and co-workers (2000) identifies clusters from multilocus genotypic frequencies based on the minimisation of departures from Hardy-Weinberg equilibrium and of linkage disequilibrium between loci. This analysis revealed two distinct genetic clusters that were largely coincident with the molestus and pipiens forms defined by both the phenotypic traits and the CQ11FL locus. Altogether, these results suggest that molestus and pipiens forms represent distinct gene pools of a subdivided *Cx. pipiens* population.

From the comparison with the ancestry groups revealed by STRUCTURE (Pritchard *et al.*, 2000), CQ11FL was only partially effective as a diagnostic marker. There was a good concordance between alternative homozygous genotypes and each

form but heterozygous CQ11FL genotypes performed less well in determining admixed individuals. Under conditions of continued hybridisation, recombination and independent assortment will break the linkage between alternative diagnostic genotypes and their respective genetic ancestry background. As pointed by Bahnck & Fonseca (2006), results from this marker should thus be interpreted only at the population level. Nevertheless CQ11FL still served as a good indicator of the sympatric presence of both molestus and pipiens forms in the study area.

Based on the partitioning of samples according to ancestry clusters revealed by STRUCTURE (Pritchard *et al.*, 2000), a global F_{ST} of 0.127 was obtained between molestus and pipiens forms. This estimate is slightly lower but still comparable to those reported in previous comparisons between underground molestus and aboveground pipiens populations (usually between 0.130 and 0.190) using similar microsatellite datasets (Huang *et al.*, 2008; Huang *et al.*, 2009). Although no molestus underground populations from the study area were available for comparison, it appears that gene flow between molestus and pipiens forms is not significantly increased by the sympatric co-existence of both populations in the surface. This argument plays in favour of the hypothesis of at least partial reproductive isolation between molestus and pipiens forms and that the under/aboveground physical discontinuity is not the only factor promoting genetic divergence, as previously debated (Byrne & Nichols, 1999; Fonseca *et al.*, 2004; Kent *et al.*, 2007). Under this particular situation of sympatry, positive reinforcement may play a role in counteracting the effects of gene flow (Noor, 1999), hence maintaining isolation between forms.

Microsatellite CQ11 displayed the highest differentiation between molestus and pipiens, with an F_{ST} estimate *ca.* 2-fold greater than for the other loci. This locus was close to fixation in molestus form for a 286 bp allele, but this was a low-frequency allele in the pipiens form (Figure 3). This allelic profile is not unique for the study area. High frequencies of a CQ11 allele in the same size range (283-285 bp) have been reported for underground and aboveground molestus populations from Europe and the USA (Fonseca *et al.*, 2004; Bahnck & Fonseca, 2006; Kent *et al.*, 2007). This continental-wide genetic signature is consistent with a single evolutionary origin of the molestus form, possibly arising in the southern latitudes of Europe or North Africa as a human-adapted commensal form, that later dispersed into northern latitudes as

underground suitable habitats became available (Fonseca *et al.*, 2004). Furthermore, this locus-specific differentiation may indicate that CQ11 locates in a genomic region under divergent selection. In these genomic regions, reduced recombination and selection against introgression maintain differentiation not only at loci associated with traits of ecological adaptation or reproductive isolation but also at surrounding neutral loci through genetic hitchhiking (Via & West, 2008; Nosil *et al.*, 2009). This mechanism is considered a major process of sympatric/ecological speciation and has been described in several insect species (Machado *et al.*, 2002; Turner *et al.*, 2005; Egan *et al.*, 2008). Genome-wide scans will be necessary to confirm the presence of such genomic regions in *Cx. pipiens*

Estimates of hybrid rates between molestus and pipiens forms between 7-10% were obtained by STRUCTURE (Pritchard *et al.*, 2000) and NEWHYBRIDS (Anderson & Thompson, 2002) admixture analysis. These values are similar to the estimates obtained for southern European aboveground populations (10%) using STRUCTURE, although the authors used a different *Tq* of 0.06 (Fonseca *et al.*, 2004). Adjusting ancestry assignment to this threshold still yielded a comparable hybrid rate of 15.2% for our sample. In comparisons between underground molestus and aboveground pipiens populations from the USA hybrid rates of 12% have been documented (Huang *et al.*, 2008) but up to 40% admixed individuals have been documented in USA *Cx. pipiens* populations by Fonseca and co-workers (2004). According to the authors, a more recent colonisation and posterior contact of separate Old World molestus and pipiens populations may explain the higher levels of hybridisation found in the USA. On the other hand, the low levels of hybridisation in southern European *Cx. pipiens* populations, even when both forms occur sympatrically as here demonstrated, provides additional support for reproductive/ecological barriers to gene flow other than habitat segregation.

The degree of microsatellite differentiation in our dataset was insufficient to identify hybrids beyond the F1 class, as revealed by the analysis of simulated data. This was not an unexpected result as NEWHYBRIDS (Anderson & Thompson, 2002) often requires a large number of highly diagnostic markers between populations to identify F2 and backcrossed hybrids with confidence (Gow *et al.*, 2006; Vähä & Primmer, 2006). However, this analysis revealed important differences in the proportions of admixture

within forms. Individuals with molestus ancestry were all classified as purebred molestus with probabilities of assignment above 0.92. In contrast, individuals with pipiens ancestry had a mean proportion of admixture of 0.387 (as measured by the individual posterior probabilities of belonging into a non-pipiens class) and 28-48% (depending on *Tq*) were classified as hybrids. These differences suggest a pattern of asymmetrical gene flow, in which higher proportions of molestus alleles are introgressed into the pipiens form. A similar trend has also been described in a population from Chicago IL (USA), in which the pipiens form presented higher proportions of molestus and *Cx. quinquefasciatus* ancestry (Huang *et al.*, 2009).

Another hypothesis could be raised if the molestus form would have locally evolved from the pipiens form through a recent founding event. Under this scenario, the microsatellite composition of the molestus population would be made almost exclusively of only a subset of the alleles present in the pipiens form which might result in an apparent signal of admixture in the latter. While estimates of allele and private allele richness seem to support this view, there were considerable differences between forms in the microsatellite allele arrays that are not consistent with this hypothesis. These differences are illustrated by the number of mutational steps separating the most frequent alleles at each locus. Size variance-based R_{ST} values were higher than frequency-based F_{ST} values in nearly half of the loci and also for the mean over-loci estimates. Higher R_{ST} estimates do not conciliate with a recent founding event that would otherwise imply that genetic drift rather than mutation would be the primary evolutionary force shaping genetic divergence between forms (Slatkin, 1995). Moreover, heterozygosity tests provided no evidence for the molestus form to have recently undergone any major population reduction that would be expected from a founding event. Finally, the peculiar composition of the CQ11 microsatellite in the molestus form, displaying a high frequency allele common to all other molestus populations regardless of geographic origin is also not consistent with local multiple origins of the molestus form. Altogether, these evidences render the hypothesis of the molestus population being derived from the local pipiens form unlikely. Extending the analysis to other regions of sympatry between molestus and pipiens would provide insights on whether the observed patterns of introgression are a local phenomenon or a general trend for the species in its southern distribution.

The mechanisms underlying the patterns of asymmetrical introgression between molestus and pipiens are unknown. One hypothesis can be drawn from the different mating strategies displayed by molestus and pipiens forms. Preferential introgression from molestus to pipiens could be expected if stenogamous molestus males mate readily with both molestus and pipiens females in aboveground habitats. On the other hand, pipiens males require open spaces to mate due to swarm-based mating behaviour (Downes, 1969). This more specialised behaviour may result in a higher propensity to mate with pipiens females. This hypothesis relies on two main assumptions. The first is that introgression between molestus and pipiens is mainly male-mediated and to test for this hypothesis the analysis of sex-linked markers would be required. In a recent study analysing Asian populations of two additional members of the *Cx. pipiens* complex, the allele specific of *Cx. quinquefasciatus* at the sex-linked *ace-2* locus was found to have introgressed into *Culex pipiens* pallens Coquillett 1898 through the males (Fonseca *et al.*, 2009). Patterns of male-mediated asymmetrical introgression have also been reported in several other non-insect organisms, such as tree frogs (Lamb & Avise, 1986), warbler birds (Bensch *et al.*, 2002), mouse lemurs (Gligor *et al.*, 2009) and macaque monkeys (Bonhomme *et al.*, 2009). The second assumption is that both pipiens and hybrid females have a greater propensity for seeking swarms for mating. To address this question, more studies are needed to characterise the swarming and mating behaviours in *Cx. pipiens* , in areas of sympatry between forms.

The molestus form was predominant in the study area and this trend appeared to be maintained throughout the year (data not shown). While this factor may also contribute to a higher introgression of genes from molestus to pipiens, it may also suggest fitness differences between forms. In southern regions with mild winters, the inability of the molestus form to undergo diapause during winter may be a lesser disadvantage than at northern latitudes. When occurring in sympatry with the pipiens form in surface habitats, autogeny and a more generalist mating behaviour are likely to result in a greater fitness molestus form.

Conclusion

Both physiological/behavioural and genetic data provide evidence for the sympatric occurrence of molestus and pipiens forms of *Cx. pipiens* in aboveground habitats of the study area. In spite of the sympatric occurrence, estimated hybridisation rates were not much higher than those reported in ecological settings where both forms are physically separated which suggests at least partial reproductive isolation between molestus and pipiens. More importantly, hybridisation appears not to be bidirectional and this is possibly a result of the different mating strategies exhibited by each form. The observed patterns of asymmetrical introgression may have epidemiological repercussions. In two recent studies covering three USA States, pipiens form females that have fed upon mammals (humans in particular) presented significantly higher proportions of molestus genetic ancestry (Kilpatrick *et al.*, 2007; Huang *et al.*, 2009). These findings suggest a genetic basis for host selection by *Cx. pipiens*. The introgression of molestus genes into the pipiens form may induce a more opportunistic biting behaviour thus potentiating the capacity of the latter form to act as a bridge-vector for the transmission of arbovirus such as WNV (Hamer *et al.*, 2008). Further studies focusing on the feeding habits and population dynamics of molestus and pipiens forms are required in order to clarify the impact of hybridisation in the vectorial capacity of *Cx. pipiens* and, consequently, on the potential for transmission of arboviral infections.

Material and methods

Study region and mosquito collection

Mosquito collections took place between May 2005 and August 2006 in the Comporta region (38° 22' 60 N, 8° 46' 60 W), District of Setubal, Portugal. Comporta is a low-lying area (altitude <60 m) with diverse ecotypes. Residential areas are situated along a national road that crosses the study region from north to south. The south and east is mainly occupied by pine forest (*Pinus pinaster* Aiton 1789; *Pinus pinea* L. 1753) and semi-natural agro-forest systems of cork-oak (*Quercus suber* L. 1753). In the west there are extensive areas of rice fields and a system of sand-dunes. The north and northwest is part of a protected landscape area occupied by marshland, rice fields and

saltpans. This protected area extends northwards into the national wildlife reserve of Estuário do Sado. The reserve harbours over 240 bird species. These include migratory birds such as the European starling (*Sturnus vulgaris* L. 1758), the mallard (*Anas platyrhynchos* L. 1758) and the white stork (*Ciconia ciconia* L. 1758), that have been reported as WNV hosts (Rappole & Hubálek, 2003).

The region has a warm temperate climate with a dry hot summer and a mild winter (class Csa, Köppen Classification System, Kottek *et al.*, 2006). Monthly averages of mean daily temperatures vary between 10°C and 21°C and relative humidity between 76% and 89%. Monthly averages of daily rainfall fluctuate between 0.12 and 3.4 mm.

Bimonthly mosquito collections were made by indoors resting captures and CDC light traps baited with CO_2 inside animal shelters (chicken, rabbit and pig). Collected live mosquitoes were transported to the laboratory and identified to species or complex of sibling species using morphological keys (Ribeiro & Ramos, 1999).

Determination of autogeny and stenogamy

Blood fed and gravid *Cx. pipiens* females were placed in individual cages in an Insectary (25 ± 2°C; 70 ± 10% RH) until oviposition. Individual egg rafts were reared until the adult stage to obtain F1 families. Pupae from each F1 family were transferred into cages with 20 cm side (0.008 m^3) for adult emergence. After emergence of the first adult the family was kept in the cage with access to a fructose 10% solution and an oviposition tray. Both pupae and oviposition trays where daily observed for the presence of egg-rafts. If oviposition occurred until two days after the emergence of the last adult (*i.e.* on average 14 days after the emergence of the first adult of the egg batch) the family was deemed autogenous. Families that did not lay eggs during this period were divided into two cages keeping similar sex ratios in each cage. In one of the cages mosquitoes were maintained in similar conditions as previously in order to recover eventually autogenous families that had delayed oviposition. In the other cage, females were given the opportunity to take a daily blood feed on a vertebrate host (mouse and chicken) for a period of 10 days.

After the end of the experiment, all F1 specimens were sacrificed by chilling. Females had their abdomen dissected to determine if their spermatheca was

inseminated, as an indicator of the capacity to mate in confined spaces. The head and thorax of each female were preserved in individual tubes with silica gel and kept at room temperature until DNA extraction.

Molecular analyses

DNA extraction from individual F1 females was performed by the method of Collins and co-workers (1987). Specimens were identified to species of the *Culex pipiens* complex by a multiplex PCR assay that targets species-specific polymorphisms at the intron-2 of the acetylcholinesterase-2 gene (*ace-2*), using primers specific for *Cx. pipiens s.s.*, *Culex torrentium* Martini 1925 and *Cx. quinquefasciatus* (Smith & Fonseca, 2004). The first two species have been annotated for Portugal (Ribeiro & Ramos, 1999). Although *Cx. quinquefasciatus* has not been found in Portugal, its subtropical distribution with a northern limit around 36° latitude prompted us to test this additional primer. The PCR assay described by Bahnck & Fonseca (2006) was used to detect a size polymorphism in the 5' flanking region of the CQ11 microsatellite of *Cx. pipiens*. This marker, here denoted as CQ11FL, differentiates specimens of the pipiens form, that display a PCR product of 200 bp, from the molestus form (250 bp). Hybrids exhibit both amplicons (200 bp/250 bp).

Fourteen microsatellite loci (Fonseca *et al.*, 1998; Keyghobadi *et al.*, 2004; Smith *et al.*, 2005) were analysed in this study (Table S4, Additional File 1). Each locus was amplified separately in a 20 μl PCR reaction that contained 1× GoTaq® Flexi Buffer (Promega, USA), 2.5 mM $MgCl_2$, 0.20 mg/ml Bovine Serum Albumin, 0.25 mM dNTPs, 0.20 μM of each primer and 0.5 U GoTaq® Flexi DNA polymerase (Promega, USA). For each locus, one of the primers was fluorescently labelled (NED, HEX or 6-FAM; Applied Biosystems, USA). Thermocycling conditions included an initial denaturation step of 5 min at 96°C followed by 30 cycles each with 96°C for 30 s, Annealing at 52°C-58°C (locus dependent, Table S4) for 30 s, and 72°C for 30 s. After a final extension step of 5 min at 72°C, reactions were stopped at 4°C.

Amplified products were separated by capillary electrophoresis in a genetic analyser ABI3730 (Applied Biosystems, USA) at the DNA Analysis Facility on Science Hill, Yale University (USA). Fragment sizes and genotypes were scored using the software GeneMarker 1.4. (Softgenetics, USA).

Data analysis

Pearson's Chi-square tests were used to determine associations between autogeny and stenogamy phenotypic traits and with CQ11FL genotypes.

Genetic diversity at each microsatellite locus was characterised by estimates of expected heterozygosity using Nei's unbiased estimator (Nei, 1987) and inbreeding coefficient (F_{IS}). Significance of F_{IS} values was assessed by randomisation tests. These analyses were performed using FSTAT v. 2.9.3.2. (Goudet *et al.*, 1995). In addition, estimates of allele richness (A_R) and private allele richness (pA_R) adjusted for the lowest sample size were obtained by a rarefaction statistical approach implemented by the programme HP-RARE (Kalinowski, 2005).

Departures from Hardy-Weinberg proportions were tested by exact tests available in ARLEQUIN v.3.11 (Excoffier *et al.*, 2005). The same software was used to perform exact tests of linkage disequilibrium between pairs of loci based on the expectation-maximisation approach described by Slatkin & Excoffier (1996). Cornuet & Luikart's (1996) heterozygosity tests were used to detect recent population perturbations. This method compares two estimates of expected heterozygosity, based on allele frequency (H_e) and on the number of alleles and sample size (H_{eq}), respectively. At mutation-drift equilibrium (MDE), both estimates should be similar but if a population experiences a recent bottleneck there will be a transient state in which $H_e > H_{eq}$ due to a rapid loss of rare alleles. Conversely $H_e < H_{eq}$ is an indicator of a recent population expansion. Estimates of H_{eq} under MDE were obtained assuming a strict stepwise mutation model (SMM) and two-phase models (TPM) with proportions of indels larger than one repeat of 10%, 20% and 30%. Wilcoxon tests were used to determine if there were a significant number of loci in which $H_e \neq H_{eq}$ as an indication of departure from MDE. Calculations were done using BOTTLENECK version 1.2.02 (Cornuet & Luikart, 1996).

Genetic differentiation between groups was measured by estimates of the fixation index, F_{ST}, calculated according to Weir & Cockerham (1984). Genotypic permutation tests available in FSTAT (Goudet *et al.*, 1995) were performed to infer if the estimates differed significantly from zero. The microsatellites equivalent R_{ST} (Slatkin, 1995) was estimated as implemented by ARLEQUIN (Excoffier *et al.*, 2005).

Bayesian clustering analysis as implemented by STRUCTURE 2.2 (Pritchard *et al.*, 2000) was used to infer population substructure/ancestry from the dataset without prior information of sampling groups (*i.e.* phenotypes), under the admixture model with correlated allele frequencies. Ten independent runs with 10^5 burn-in steps and 10^5 iterations were done for each value of K (K = 1 to 4 clusters). The method of Evanno and co-workers (2005) was used to determine the most likely number of clusters in the sample. Following the suggestions of Vähä & Primmer (2006), individual genetic assignment to clusters was based on a minimum posterior probability threshold (*Tq*) of 0.90. Individuals displaying $0.1 \leq q_i \leq 0.90$ were considered of admixed ancestry.

The Bayesian method implemented by NEWHYBRIDS 1.1. (Anderson & Thompson, 2002) was used to assign individuals into 6 classes: pure molestus, pure pipiens, and hybrids (F1, F2 and backcrosses with molestus or pipiens). The approach of uniform priors was used and results were based on the average of five independent runs each with 10^5 burn-in steps and 10^5 iterations.

The performance of NEWHYBRIDS to detect purebred and hybrid individuals with the present microsatellite dataset was assessed using simulated data generated by HYBRIDLAB (Nielsen *et al.*, 2006). From the initial NEWHYBRIDS analysis, pure molestus and pipiens individuals were selected based on a q_i>0.90. From this sampling, 100 simulated genotypes of each parental and hybrid class were generated. These artificial genotypes, without prior population information, were analysed in NEWHYBRIDS. Following the examples of previous works (Vähä & Primmer, 2006; Burgarella *et al.*, 2009), power (number of correctly identified individuals for a class over the actual number of individuals of that class) and accuracy (number of correctly identified individuals for a class over the total number of individuals assigned to that class) were calculated for four *Tq* values (0.50, 0.70, 0.80 and 0.90). Analysis was based on the mean of five replicates of simulated datasets.

Whenever multiple testing was performed, the nominal significance level of rejection of the null hypothesis (α = 0.05) was corrected by the sequential Bonferroni procedure (Holm, 1979).

Acknowledgments

We thank Suraya Diaz for the technical support given to the field collections and insectary experiments. We acknowledge the logistic support given by The Atlantic Company (Portugal) Turismo e Urbanização, S.A., during the mosquito collections. This study was funded by Fundação para a Ciência e a Tecnologia/FEDER, Ministério da Ciência, Tecnologia e Ensino Superior (POCI/BIA-BDE/57650/2004 and PPCDT/BIA-BDE/57650/2004).

References

Almeida, A.P.G., Galão, R.P., Sousa, C.A., Novo, M.T., Parreira, R., Pinto, J., Piedade, J. & Esteves, A. (2008) Potential mosquito vectors of arboviruses in Portugal: species, distribution, abundance and West Nile infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 102 (8), 823–832.

Anderson, E.C. & Thompson, E.A. (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 160 (3), 1217–1229.

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), 251–255.

Bensch, S., Helbig, A.J., Salomon, M. & Seibold, I. (2002) Amplified fragment length polymorphism analysis identifies hybrids between two subspecies of warblers. *Molecular Ecology*. 11 (3), 473–481.

Bonhomme, M., Cuartero, S., Blancher, A. & Crouau-Roy, B. (2009) Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). *The Journal of Heredity*. 100 (2), 158–169.

Burgarella, C., Lorenzo, Z., Jabbour-Zahab, R., Lumaret, R., Guichoux, E., Petit, R.J., Soto, A. & Gil, L. (2009) Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity*. 102 (5), 442–452.

Byrne, K. & Nichols, R.A. (1999) *Culex pipiens* in London Underground tunnels: differentiation between surface and subterranean populations. *Heredity*. 82 (1), 7–15.

Chevillon, C., Eritja, R., Pasteur, N. & Raymond, M. (1995) Commensalism, adaptation and gene flow: mosquitoes of the *Culex pipiens* complex in different habitats. *Genetical Research*. 66 (2), 147–157.

Clements, A.N. (1999) *The Biology of Mosquitoes. Sensory Reception and Behaviour Volume 2*. Wallingford, CABI Publishing.

Collins, F.H., Mendez, M.A., Rasmussen, M.O., Mehaffey, P.C., Besansky, N.J. & Finnerty, V. (1987) A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *The American Journal of Tropical Medicine and Hygiene*. 37 (1), 37–41.

Connell, J., McKeown, P., Garvey, P., Cotter, S., Conway, A., Darina O'Flanagan, P. O'Herlihy, B., Morgan, D., Nicoll, A. & Lloyd, G. (2004) Two linked cases of West Nile virus (WNV) acquired by Irish tourists in the Algarve, Portugal. *Eurosurveillance*. 8 (32), pii=2517.

Cornuet, J.M. & Luikart, G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*. 144 (4), 2001–2014.

Crabtree, M.B., Savage, B. & Miller, B.R. (1997) Development of a polymerase chain reaction assay for differentiation between *Culex pipiens pipiens* and *Cx. p. quinquefasciatus* (Diptera: Culicidae) in North America based on genomic differences identified by subtractive hybridization. *Journal of Medical Entomology*. 34 (5), 532–537.

Downes, J.A. (1969) The swarming and mating flight of Diptera. *Annual Review of Entomology*. 14 (1), 271–298.

Egan, S.P., Nosil, P., Funk, D.J. & Pfennig, D. (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution*. 62 (5), 1162–1181.

Esteves, A., Almeida, A.P.G., Galão, R.P., Parreira, R.,Piedade, J., Rodrigues, J. C., Sousa, C.A. & Novo, M.T. (2005) West Nile virus in southern Portugal, 2004. *Vector-Borne and Zoonotic Diseases*. 5 (4), 410–413.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*. 14 (8), 2611–2620.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*. 1, 47–50.

Fonseca, D.M., Atkinson, C.T. & Fleischer, R.C. (1998) Microsatellite primers for *Culex pipiens quinquefasciatus*, the vector of avian malaria in Hawaii. *Molecular Ecology*. 7 (11), 1617-1619.

Fonseca, D.M., Keyghobadi, N., Malcolm, C.A., Mehmet, C., Schaffner, F., Mogi, M., Fleischer, R.C. & Wilkerson, R.C. (2004) Emerging Vectors in the *Culex pipiens* Complex. *Science*. 303 (5663), 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.

Gligor, M., Ganzhorn, J.U., Rakotondravony, D., Ramilijaona, O.R., Razafimahatratra, E., Zischler, H., & Hapke, A. (2009) Hybridization between mouse lemurs in an ecological transition zone in southern Madagascar. *Molecular Ecology*. 18 (3), 520–533.

Goudet, J. (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. *The Journal of heredity*. 86 (6), 485–486.

Gow, J.L., Peichel, C.L. & Taylor, E.B. (2006) Contrasting hybridization rates between sympatric three-spined sticklebacks highlight the fragility of reproductive barriers between evolutionarily young species. *Molecular Ecology*. 15 (3), 739–752.

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), 125–128.

Harbach, R.E., Harrison, B.A. & Gad, A.M. (1984) *Culex (Culex) molestus* Forskål (Diptera, Culicidae) - neotype designation, description, variation, and taxonomic status. *Proceedings of the Entomological Society of Washington*. 86 (3), 521–542.

Harbach, R.E., Dahl, C. & White, G.B. (1985) *Culex (Culex) pipiens* Linnaeus (Diptera, Culicidae) - concepts, type designations, and description. *Proceedings of the Entomological Society of Washington*. 87 (1), 1–24.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 6 (2), 65–70.

Huang, S., Molaei, G. & Andreadis, T.G. (2008) Genetic insights into the population structure of *Culex pipiens* (Diptera: Culicidae) in the northeastern United States by using microsatellite analysis. *American Journal of Tropical Medicine and Hygiene*. 79 (4), 518–527.

Huang, S., Hamer, G.L., Molaei, G., Walker, E.D., Goldberg, T.L., Kitron, U. D. & Andreadis, T.G. (2009) Genetic variation associated with mammalian feeding in *Culex pipiens* from a West Nile virus epidemic region in Chicago, Illinois. *Vector-Borne and Zoonotic Diseases*. 9 (6), 637–642.

Kalinowski, S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*. 5 (1), 187–189.

Kent, R.J., Harrington, L.C. & Norris, D.E. (2007) Genetic differences between *Culex pipiens* f. molestus and *Culex pipiens pipiens* (Diptera: Culicidae) in New York. *Journal of Medical Entomology*. 44 (1), 50–59.

Keyghobadi, N., Matrone, M.A., Ebel, G.D., Kramer, L.D. & Fonseca, D.M. (2004) Microsatellite loci from the northern house mosquito (*Culex pipiens*), a principal vector of West Nile virus in North America. *Molecular Ecology Notes*. 4 (1), 20–22.

Kilpatrick, A.M., Kramer, L.D., Jones, M.J., Marra, P.P. & Daszak, P. (2007) Genetic influences on mosquito feeding behavior and the emergence of zoonotic pathogens. *American Journal of Tropical Medicine and Hygiene*. 77 (4), 667–671.

Kottek, M., Grieser, J., Beck, C., Rudolf, B. & Rubel, F. (2006) World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*. 15 (3), 259–263.

Lamb, T. & Avise, J.C. (1986) Directional introgression of mitochondrial DNA in a hybrid population of tree frogs: The influence of mating behavior. *Proceedings of the National Academy of Sciences of the United States of America*. 83 (8), 2526–2530.

Machado, C.A., Kliman, R.M., Markert, J.A. & Hey, J. (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*. 19 (4), 472–488.

Nei, M. (1987) *Molecular Evolutionary Genetics*. New York, Colombia University Press.

Nielsen, E.E.G., Bach, L.A. & Kotlicki, P. (2006) HYBRIDLAB (version 1.0): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes*. 6 (4), 971–973.

Noor, M.A. (1999) Reinforcement and other consequences of sympatry. *Heredity*. 83 (5), 503–508.

Nosil, P., Funk, D.J. & Ortiz-Barrientos, D. (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*. 18 (3), 375–402.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*. 155 (2), 945–959.

Rappole, J.H. & Hubálek, Z. (2003) Migratory birds and West Nile virus. *Journal of Applied Microbiology*. 94 (1), 47–58.

Ribeiro, H., Pires, C.A., Ramos, H.C. & Capela, R.A. (1983) Research on the mosquitoes of Portugal (Diptera, Culicicae). VIII- On the occurrence of *Culex (Culex) molestus* Forskål, 1775. *Jornal Sociedade Ciências Medicas Lisboa*. 147, 185–188.

Ribeiro, H. & Ramos, H.C. (1999) Identification keys of the mosquitoes (Diptera: Culicidae) of Continental Portugal, Açores and Madeira. *European Mosquito Bulletin*. 3, 1–11.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*. 139 (1), 457–462.

Slatkin, M. & Excoffier, L. (1996) Testing for linkage disequilibrium in genotypic data using the EM algorithm. *Heredity*. 76 (4), 377–383.

Smith, 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 Hygiene*. 70 (4), 339–345.

Smith, J.L., Keyghobadi, N., Matrone, M.A., Escher, R.L. & Fonseca, D.M. (2005) Cross-species comparison of microsatellite loci in the *Culex pipiens* complex and beyond. *Molecular Ecology Notes*. 5 (3), 697–700.

Spielman, A. (2001) Structure and seasonality of nearctic *Culex pipiens* populations. *Annals of the New York Academy of Sciences*. 951, 220–234.

Turner, T.L., Hahn, M.W. & Nuzhdin, S.V. (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*. 3 (9), e285.

Vähä, J.P. & Primmer, C.R. (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*. 15 (1), 63–72.

Via, S. & West, J. (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*. 17 (19), 4334–4345.

Vinogradova, E.B. (2000) *Culex pipiens pipiens mosquitoes: taxonomy, distribution, ecology, physiology, genetics, applied importance and control*. S.I. Golovatch (ed.). Sofia, Pensoft.

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.

Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*. 38 (6), 1358–1370.

Additional File 1

Table S1. Genetic diversity at microsatellite loci of *Culex pipiens* from Portugal.

Locus		CQ11FL$_{250/250}$ (N=78)	CQ11FL$_{200/250}$ (N=26)	CQ11FL$_{200/200}$ (N=41)	All samples (N=145)
	A_R	1.7	4.8	7.2	6.5
CQ11	H_e	0.031	**0.620**	0.801	**0.607**
	F_{IS}	-0.004	**-0.548***	0.190	**0.346**
	A_R	6.6	7.0	10.1	8.7
CQ26	H_e	**0.756**	**0.826**	0.862	**0.835**
	F_{IS}	**0.537**	**0.579**	0.001	**0.396**
	A_R	9.3	8.8	12.0	10.8
CQ41	H_e	0.789	0.789	**0.833**	**0.812**
	F_{IS}	0.144	0.088	**0.262**	**0.179**
	A_R	4.4	8.5	11.9	9.0
CxpGT04	H_e	0.637	0.727	0.879	**0.748**
	F_{IS}	0.092	-0.111	0.062	0.077
	A_R	6.1	10.0	10.3	9.3
CxpGT09	H_e	**0.743**	**0.855**	**0.847**	**0.818**
	F_{IS}	**0.521**	**0.491**	**0.366**	**0.482**
	A_R	3.9	4.9	7.4	5.9
CxpGT12	H_e	0.380	0.470	0.766	**0.545**
	F_{IS}	0.184	0.202	0.106	**0.213**
	A_R	10.8	12.5	16.7	13.2
CxpGT20	H_e	**0.867**	0.888	0.916	**0.895**
	F_{IS}	0.135	-0.082	0.121	**0.104**
	A_R	6.6	6.8	9.6	8.1
CxpGT40	H_e	0.805	0.735	0.719	**0.814**
	F_{IS}	0.038	0.238	0.185	**0.166**
	A_R	5.5	8.8	10.1	8.1
CxpGT46	H_e	0.667	0.776	0.853	**0.748**
	F_{IS}	0.110	0.034	0.194	**0.131**
	A_R	11.8	13.4	18.3	14.3
CxpGT51	H_e	0.866	0.885	0.932	0.891
	F_{IS}	0.131	0.044	0.035	0.089
	A_R	14.1	12.7	22.4	16.5
CxpGT53	H_e	**0.893**	0.888	0.954	**0.914**
	F_{IS}	**0.186**	0.009	0.106	**0.134**
	A_R	1.7	2.8	1.8	1.8
CxqQGT4	H_e	0.038	0.076	0.048	0.047
	F_{IS}	-0.013	-0.010	-0.013	-0.018
	A_R	3.3	3.9	6.9	5.0
CxqGT6B	H_e	0.635	0.686	0.754	0.681
	F_{IS}	0.079	-0.166	0.006	0.017
	A_R	2.5	3.0	3.6	3.1
CxqTRI4	H_e	0.301	0.306	0.543	0.380
	F_{IS}	0.105	0.121	0.146	0.148
	A_R	6.3	7.7	10.6	8.6
All loci	H_e	**0.601**	**0.681**	**0.765**	**0.695**
	F_{IS}	**0.192**	**0.078**	**0.135**	**0.190**

N: sample size. A_R: allelic richness; *He*: expected heterozygosity; F_{IS}: inbreeding coefficient. Values in bold indicate a significant P-value after correction for multiple tests (see Methods). *Significant *P*-value for a negative F_{IS}. Per locus and over samples Hardy-Weinberg tests were performed using ARLEQUIN. For over loci estimates the global test available in FSTAT was used.

Table S2. Estimates of F_{ST} and R_{ST} between forms of *Culex pipiens* identified by Bayesian clustering analysis performed in STRUCTURE (Pritchard *et al.*, 2000).

Locus		Cluster1 (molestus) *vs.* Admixed	Admixed *vs.* Cluster2 (pipiens)	Cluster1 (molestus) *vs.* Cluster2 (pipiens)	All samples
CQ11	F_{ST}	**0.207**	0.032	**0.405**	**0.360**
	R_{ST}	**0.266**	**0.239**	**0.612**	**0.566**
CQ26	F_{ST}	0.099	-0.017	**0.132**	**0.117**
	R_{ST}	0.079	0.018	0.004	0.017
CQ41	F_{ST}	0.038	0.014	**0.072**	**0.059**
	R_{ST}	0.059	-0.013	0.025	0.030
CxpGT4	F_{ST}	**0.043**	0.033	**0.155**	**0.120**
	R_{ST}	0.032	0.001	**0.151**	**0.110**
CxpGT12	F_{ST}	**0.072**	0.049	**0.272**	**0.216**
	R_{ST}	-0.020	0.061	**0.131**	**0.099**
CxpGT20	F_{ST}	**0.023**	0.045	**0.060**	**0.050**
	R_{ST}	0.082	0.013	-0.009	0.009
CxpGT40	F_{ST}	-0.019	**0.184**	**0.205**	**0.164**
	R_{ST}	-0.022	**0.157**	**0.267**	**0.219**
CxpGT46	F_{ST}	-0.043	0.061	**0.060**	**0.049**
	R_{ST}	-0.043	-0.058	0.010	0.001
CxpGT51	F_{ST}	0.002	0.021	**0.025**	**0.019**
	R_{ST}	**0.151**	0.063	**0.346**	**0.284**
CxpGT53	F_{ST}	0.020	0.010	**0.035**	**0.032**
	R_{ST}	0.123	-0.038	**0.076**	**0.078**
CxqGT4	F_{ST}	-0.007	-0.024	0.002	-0.004
	R_{ST}	-0.008	-0.015	0.007	0.001
CxqGT6b	F_{ST}	-0.003	0.001	**0.015**	**0.010**
	R_{ST}	0.033	0.093	0.010	0.024
CxqTri4	F_{ST}	-0.009	0.053	**0.164**	**0.119**
	R_{ST}	-0.018	**0.146**	**0.321**	**0.256**
All loci	F_{ST}	**0.031**	**0.039**	**0.127**	**0.104**
	R_{ST}	-0.003	-0.055	**0.191**	**0.135**
Without CQ11	F_{ST}	**0.021**	**0.039**	**0.103**	**0.082**
	R_{ST}	-0.016	-0.067	**0.123**	**0.079**

Individuals with a minimum posterior probability q_i<0.90 were considered admixed genotypes between the two clusters (molestus and pipiens). In bold: significant F_{ST} or R_{ST} after correction for multiple testing by the sequential Bonferroni procedure.

Table S3. Power and accuracy of NEWHYBRIDS to detect purebred and hybrid simulated individuals.

Class	*Tq* = 0.900		*Tq* = 0.800		*Tq* = 0.700		*Tq* = 0.600		*Tq* = 0.500	
	Power	**Accuracy**	**Power**	**Accuracy**	**Power**	**Accuracy**	**Power**	**Accuracy**	**Power**	**Accuracy**
Pure molestus	0.890	1.000	0.990	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Pure pipiens	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Hybrid	0.678	1.000	0.860	1.000	0.953	1.000	0.988	1.000	1.000	1.000
F1	0.110	1.000	0.720	1.000	0.930	1.000	1.000	1.000	1.000	1.000
F2	0.000	-	0.010	1.000	0.030	1.000	0.160	1.000	0.410	1.000
Bx molestus	0.010	1.000	0.110	1.000	0.410	1.000	0.620	1.000	0.840	1.000
Bx pipiens	0.000	-	0.060	1.000	0.280	1.000	0.570	1.000	0.830	1.000

Five threshold values (*Tq*) were analysed for power and accuracy to detect parental and hybrid simulated individuals. The "hybrid" category represents the sum of assignment probabilities to each of the four hybrid lineages originally tested. Power: number of correctly identified individuals for a class over the actual number of individuals of that class in the sample (*N*=100 for purebred, F1, F2, Bx molestus and Bx pipiens; *N*=400 for the hybrid class category). Accuracy: number of correctly identified individuals for a class over the total number of individuals assigned to that class.

Table S4. Microsatellite loci analysed.

Locus	Repeat	Primers	T_A (ºC)	Ref
CQ11	$(GT)_2ACTTC(GT)_9$	F:GATCCTAGCAAGCGAGAAC R:6-fam-GAGCGGCCAAATATTGAGAC	52	[1]
CQ26	$(GTGTGTAT)_2+(GT)_{10}+(GT)_5$	F: TCCGACATGGGAAGAGCGCA R: 6-fam-ACGCGCCCTTCTTCTGCAAC	56	[1]
CQ41	$(GT)_{12}$	F: CTGCCACTGCCTGACTGAAA R: Hex-ACCACTCAGCAACATCCGGC	52	[1]
CxpGT4	$(GT)_5(GTTT)_2GC(GT)_2CT(GT)_5$	F: GTCGTCGCTAACCCTTGTT R: Ned-CGCGATAGTCGGTAATCGT	54	[2]
CxpGT9	$(GT)_{13}$	F: AATCTCCCCGTATAATTGTG R: Ned-TATAAGACCAGTGAAGCCAG	52	[2]
CxpGT12	$(TG)_{14}$	F: AACGTGAGCGTGATTGCTC R: 6-fam-CAGCTGTTGCACCAATGTC	54	[2]
CxpGT20	$(TG)_{15}$	F: CAACCGCTAAATTGCCTCA R: Ned-GCAAACCCGATACCGAAT	54	[2]
CxpGT40	$(GT)_{15}$	F: CATCATCTGTCCACGATCC R: Hex-TTATGCAGTTGCTGTCATATCC	52	[2]
CxpGT46	(TG)15	F: Hex-CCGACACCGTGTTCAAAGAG R: TGACGACGACGGTACAAGAG	52	[2]
CxpGT51	$(TG)_4CG(TG)_{15}$	F: GAGTATCGCTCGTTGGAGATT R: Hex-ACCCTCTTTTCTTTCTATGTCTGT	54	[2]
CxpGT53	$(TG)_{22}$	F: 6-fam-GTCCCGTTTGGTTGGTTG R: CCATCTCCTCCTGAATCCTG	58	[2]
CxqGT4	$(GT)_{12}$	F: ATAGAACTTGTTCGCCGTCTC R: 6-fam-TCTAAACACGCACCACGTACA	52	[3]
CxqGT6b	$(CA)_8$	F: CAACCAGCAAAACCCTCATC R: Ned-TAGCCGGGCAGATTCATTAC	54	[3]
CxqTri4	$(TGC)_7$	F: Hex-CTAGCCCGGTATTTACAAGAAC R: AACGCCAGTAGTCTCAGCAG	54	[3]

T_A: annealing temperature. Ref: References; [1]: Fonseca *et al.* 1998. *Molecular Ecology* 7: 1613-1621. [2]: *Molecular Ecology Notes* 4: 20-22. [3]: *Molecular Ecology Notes* 5: 697–700.

Chapter 4.

Feeding patterns of molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in a region of high hybridisation

Submitted to *Parasites & Vectors* as:

Gomes, B., Sousa, C.A., Vicente, J.L., Pinho, L., Calderón, I., Arez, E., Almeida, A.P., Donnelly, M.J. & Pinto, J. Feeding patterns of molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in a region of high hybridisation.

Abstract

Background

Two biological forms of the mosquito *Culex pipiens s.s.*, denoted pipiens and molestus, display behavioural differences that may affect their role as vectors of arboviruses. In this study, the feeding patterns of molestus and pipiens forms were investigated in Comporta (Portugal), where high levels of inter-form admixture have been recorded. Microsatellite data were used to determine genetic backgrounds of blood-fed females collected indoors resting. Outdoor samples collected by CDC light trap and landing catches were also genotyped. The origin of the blood meal was determined by ELISA and mtDNA sequencing.

Results

The distribution of the forms differed according to collection method. The molestus form was present only in indoors collections whereas pipiens and admixed individuals were sampled both indoors and outdoors. In both forms, over 90% of blood meals were made on avian hosts.

Conclusion

The identification of blood meals taken from Passeriformes (*Passer domesticus* and *Turdus merula*) in females from both forms caught resting inside a domestic shelter highlights a potential for accidental transmission of arboviruses to humans in the region.

Background

Culex pipiens s.s. is a major vector of Japanese encephalitis serogroup arboviruses to their natural hosts birds (Medlock *et al.*, 2005) and in the accidental bridge-transmission from birds to humans and domestic mammals (Fonseca *et al.*, 2004; Hamer *et al.*, 2008). This serogroup includes West Nile virus (WNV) and Usutu virus for which human cases have been reported in the European continent (Zeller & Schuffenecker, 2004; Calzolari *et al.*, 2012).

Culex pipiens s.s. is a synanthropic mosquito with a widespread distribution in temperate regions (Vinogradova, 2000). This species occurs as two biological forms,

named molestus and pipiens, which exhibit important behavioural and physiological differences. The molestus form is stenogamous (mates in confined spaces, *i.e.* < 0.1 m^3; Clements, 1999), autogenous (can oviposit without a blood meal), homodynamic (remains active during winter) and mammophilic (prefers to feed on mammals, including humans). In contrast, the pipiens form is eurygamous (mates in open spaces), anautogenous (oviposition requires a blood meal), heterodynamic (undergoes winter diapause) and ornithophilic (feeds predominantly on birds) (Harbach *et al.*, 1984, 1985).

The degree of synanthropy also varies between forms. The molestus form is more restricted to habitats with human influence whereas the pipiens form has a greater ecological plasticity (Vinogradova, 2000). In northern temperate latitudes, molestus populations are confined to underground habitats whereas the pipiens form occupies aboveground habitats (Byrne & Nichols 1999; Vinogradova, 2000; Huang *et al.*, 2008). In southern Europe and in the Mediterranean region, populations of both forms occur sympatrically in aboveground habitats (Nudelman *et al.*, 1988; Gomes *et al.*, 2009).

Hybridisation between *Cx. pipiens s.s.* forms has been considered a major factor influencing WNV transmission (Fonseca *et al.*, 2004). Hybridisation between molestus and pipiens may result in a catholic feeding behaviour thereby increasing the risk of admixed populations to act as bridge-vectors of WNV between birds and humans (Chevillon *et al.*, 1995). An influence of different molestus and pipiens genetic backgrounds on host preference has been previously documented (Kilpatrick *et al.*, 2007). The increase of *Cx. pipiens s.s.* bites on mammals, including humans, at the end of summer in the USA, has been attributed to a peak of hybrids in aboveground habitats in this period (Spielman, 2001). However, a reduction of bird populations in the region at the end of the summer (specifically the *Turdus migratorius* L. 1766, American robin) may also potentiate a shift of the feeding behaviour in *Cx. pipiens s.s.* (Kilpatrick *et al.*, 2006).

West Nile virus surveillance studies in Europe have mainly focused on the detection of the virus (or viral antigens) in natural mosquito populations (Toma *et al.*, 2008, Almeida *et al.*, 2010; Calzolari *et al.*, 2010). Particular attention has also been given to the blood feeding preferences of these vector populations (Balenghien *et al.*, 2011; Muñoz *et al.*, 2012; Osório *et al.*, 2012). However, information about the

distribution of the *Cx. pipiens s.s.* forms and hybridisation rates is generally absent from these reports. More importantly, it remains to be determined how hybridisation between molestus and pipiens forms can affect certain behaviours that influence pathogen transmission to humans, including blood feeding preferences, and the degree of synanthropy of the mosquito populations.

A previous study carried out in 2005-2006 in Comporta, an estuarine area in south-central Portugal, described aboveground sympatric molestus and pipiens populations with incomplete genetic isolation (Gomes *et al.,* 2009). The region is home to over 240 bird species, including migratory birds that host WNV, such as the European starling (*Sturnus vulgaris* L. 1758) and the white stork (*Ciconia ciconia* L. 1758). The interaction of the migratory birds with *Cx. pipiens s.s.* mosquitoes may establish a WNV enzootic cycle with the infection of resident WNV host birds such as the house sparrow (*Passer domesticus* L. 1758) and carrion crow (*Corvus corone* L. 1758) (Rappole & Hubálek, 2003; Hubálek, 2008). Hybridisation rates of 7.6-10.3% between molestus and pipiens were recorded in this area, providing an opportunity to study the behavioural consequences of admixture between these forms (Gomes *et al.,* 2009).

In the present study, we have characterized molestus and pipiens genetic backgrounds in the Comporta region and related these to epidemiologically relevant traits, in particular their blood-meal host preferences. Results are discussed with respect to the relative contribution of the forms, and their hybrids, to the establishment of arboviral transmission cycles.

Methods

Study region and mosquito collection

The Comporta region (District of Setubal, Portugal; 38° 22' 60" N, 8° 46' 60" W) is a wet lowland (altitude <60 m) that includes a semi-natural farming ecosystem (rice production and cork-oak forest) and a protected landscape, the national wildlife reserve of Estuário do Sado. The region has a warm temperate climate with hot dry summers and mild winters (class Csa, Köppen Classification System, Peel *et al.*, 2007) with

monthly averages of mean daily temperature varying between 10°C and 21°C and daily rainfall between 0.12 and 3.4 mm.

Mosquito collections took place in two weeks in 2010 (19th - 23rd July and 7th - 13th August) in 7 localities of the region (Table S1, Additional file 1). Three sampling methods were used: i) indoor resting collections (IR) were performed inside domestic animal shelters using hand mechanical aspirators and torches. Each animal shelter was inspected for mosquitoes for a period of 10 min; ii) outdoor CDC light trap (Centers for Disease Control; Sudia & Chamberlain, 1962) collections, placed in the canopy of trees (CDC-C) and at ground level (CDC-G), were performed overnight between 19:00-09:00; iii) outdoor human landing catches (HLC) were performed between 20:00-23:00 by a team of four collectors using hand mechanical aspirators and torches.

Collected mosquitoes were killed by freezing and identified to species/complex using morphological keys (Ribeiro & Ramos, 1999). Freshly blood-fed female mosquitoes obtained by indoors resting and CDC light trap collections had their abdomens removed and preserved in 20 µl EDTA (0.125 M) at -20ºC for subsequent blood meal identification. The thorax and head of each blood-fed female was preserved individually at -20ºC until DNA extraction. Non blood-fed whole mosquitoes were preserved in the same conditions as the heads and thoraces.

Mosquito DNA extraction and molecular analysis

DNA was extracted from individual females using a phenol-chloroform method with ethanol precipitation (Donnelly *et al.*, 1999). Each specimen was identified to species by a multiplex PCR assay targeting species-specific polymorphisms in the intron-2 of the acetylcholinesterase-2 (*ace-2*) gene using primers specific for *Cx. pipiens s.s.*, *Cx. quinquefasciatus* and *Culex torrentium* (Smith & Fonseca, 2004).

Selection and analysis of microsatellite loci

The software WHICHLOCI (Banks *et al.*, 2003) was applied to the microsatellite dataset used by Gomes *et al.* (2009) to determine molestus and pipiens genetic backgrounds in Comporta, in order to select a subset of six loci to be analysed in this study. Of the 13 microsatellites genotyped in Gomes *et al.* (2009), locus CQ11 was excluded due to its linkage with the diagnostic CQ11FL marker (see below). The remaining 12 microsatellites dataset was used to create three samples of 500 simulated

individuals (molestus, pipiens and hybrids) to infer, under 10^5 iterations, which combinations of microsatellites allow to assign correctly the simulated individuals with a minimum accuracy of 90%. Bayesian clustering analysis as implemented by STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) was then used to infer population structure in the data set of Gomes *et al.* (2009) with the best six microsatellites and under the same run conditions. The results obtained for the datasets with six and 13 microsatellites were compared to establish the robustness of the analysis with the lowest battery of microsatellite loci (*i.e.* six).

Microsatellite genotyping was performed by PCR with fluorescently-labelled primers under the same conditions as in Gomes *et al.* (2009). Amplified products were separated by capillary electrophoresis in a genetic analyser ABI3730 (Applied Biosystems) at Yale DNA Analysis Facility (USA). Fragment sizes and genotypes were scored using the software GeneMarker 1.4. (Softgenetics, USA).

The multiplex PCR assay described by Bahnck & Fonseca (2006) was used to detect a size polymorphism in the 5' flanking region of the CQ11 microsatellite of *Cx. pipiens s.s.* that differentiates molestus and pipiens forms as well as their hybrids. This marker, here denoted as CQ11FL, differentiates specimens of the pipiens form (200 bp) from the molestus form (250 bp PCR product) while hybrids exhibit both pipiens and molestus amplicons (Bahnck & Fonseca, 2006). Given its relatively good performance at the population level in the region (Gomes *et al.*, 2009), this marker was used to label distinct microsatellite-based genetic clusters as belonging to the molestus or pipiens forms.

Blood meal identification

A Sandwich ELISA protocol (Simões *et al.*, 1995) was used to identify blood meals of blood-fed indoor resting mosquitoes. Blood meals were tested for the presence of chicken, cow, dog, goat/sheep, horse/donkey, human, pig, and rabbit immunoglobulin G (IgG). Four positive controls (blood from the tested species) and 14 negative controls (two blood samples from the other seven species) were used in every 96-well microplate. Absorbance values were read at 492 nm wave length in an ELISA reader (Anthos 2010 ®, Anthos Labtec Instruments). Cut-off values were calculated for each plate, as the mean plus three times the standard deviation of the negative controls.

Fragments of the mitochondrial DNA *cytochrome b* (*cyt b*) gene were sequenced to identify the blood meal source of female mosquitoes collected in the canopy of trees (CDC-C) and for a subsample of females caught indoor resting (ELISA-negative blood meals and random blood meals from all the different types of blood meal identified). DNA extraction from blood samples was performed with the DNeasy Blood &Tissue Kit (Qiagen, Valencia, CA). The vertebrate *cyt b* gene was amplified following a modified version of the protocol of Hamer *et al.* (2009) that excluded the fourth primer pair amplification. PCR products were purified with the QIAquick PCR Purification kit (Qiagen) and sequenced in a biotechnology company (StabVida, Oeiras) on an ABI3730XL automated sequencer (Applied Biosystems). Sequences were manually corrected and aligned using BioEdit 7.0.9.0 (Hall 1999). Identification of host species was performed by comparison with *cyt b* sequences deposited at NCBI GenBank.

Data analysis

Bayesian clustering analysis as implemented by STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) was used to infer population substructure/ancestry from the data set without prior information of sampling groups under the conditions of admixture (α allowed to vary between 0 and 10), and allele frequencies correlated among populations (λ was set at 1, default value). Ten independent runs with 10^4 iterations and 10^5 replications were performed for each value of K (K=1 to 10 clusters). To infer the most likely number of clusters in the sample, the ΔK statistic was used (Evanno *et al.*, 2005). Information from the outputs of each K (10 runs) was compiled by the Greedy method implemented in CLUMPP (Jakobsson & Rosenberg, 2007). Following the suggestions of Vähä & Primmer (2006), individual genetic assignment to clusters was based on a minimum posterior probability threshold (Tq) of 0.90. Individuals displaying $0.1 \leq q_i \leq 0.90$ were considered of admixed ancestry.

Genetic diversity at each microsatellite locus was characterised by estimates of expected heterozygosity, H_e (Nei, 1987) and inbreeding coefficient (F_{IS}). Significance of F_{IS} values was assessed by randomisation tests. These analyses were performed using FSTAT v. 2.9.3.2. (Goudet, 1995). Estimates of allele richness (A_R), adjusted for the lowest sample size, were obtained by a rarefaction statistical approach implemented by the programme HP-RARE (Kalinowski, 2005). Departures from Hardy–Weinberg

equilibrium were tested by exact tests available in ARLEQUIN v.3.5 (Excoffier *et al.,* 2005). The same software was used to perform exact tests of linkage equilibrium between pairs of loci based on the expectation-maximisation approach described by Slatkin & Excoffier (1996). The software Micro-Checker 2.2.3. was used to search (99% confidence interval) for null alleles at loci/samples (van Oosterhout *et al.,* 2004).

Fisher's exact tests (2x2) were performed with *"VassarStats: Website for Statistical Computation"* (Lowry, 2012) to determine associations the genetic clusters identified by STRUCTURE and the origin of blood meals.

Whenever multiple testing was performed, the nominal significance level of rejection of the null hypothesis (α= 0.05) was corrected by the sequential Bonferroni procedure (Holm, 1979).

Results

Mosquito sampling

A total of 80 IR collections were performed in 28 sites (Table 1). The majority of animal shelters found in the area were chicken coops (46.4%). Consequently, 44 (55.0%) of the IR collections were made in chicken coops whereas 19 (23.8%) were made in shelters harbouring mammalian hosts without domestics birds (*i.e.* rabbit hutches, cattle barns and pig pens). Seven (8.8%) collections were performed in shelters with both avian and mammalian hosts and 10 (12.5%) inside installations without any visible vertebrate host. The IR collections yielded a total of 235 *Cx. pipiens s.l.* females, of which 174 (74.0%) were blood fed. Of the total of females caught, 88.5% were sampled inside chicken coops, 4.3% in mammalian shelters, 3.8% in mixed avian-mammal shelters and 3.4% in installations with no domestic vertebrates (Table 1). None of the 10 females caught inside shelters exclusively with mammalian hosts was blood fed and only 6 (3.4%) engorged females were collected in mixed avian-mammalian shelters.

A total of 24 outdoor CDC light trap collections were performed (Table S1, Additional file 1). Of these, 17 were performed with traps hung at the canopy of trees (CDC-C), yielding 1,093 *Cx. pipiens s.l.* females, and 7 were placed at ground level

yielding a total of 625 females. Human landing catches were performed six times at a single site (Table S1, Additional file 1). These collections yielded a total of 155 *Cx. pipiens s.l.* females. The mean number of bites per human per hour was 2.2 for this species.

Table 1. Number of indoor resting collections and *Cx. pipiens s.l.* mosquitoes caught according to the type of shelter.

Shelters	IR sites	Collections		*Cx. pipiens s.l.*	
		N_C	N_{PC}	N_F	N_{BF}
Chicken coops	13 (46.4)	44 (55.0)	31 (70.5)	208 (88.5)	164 (94.3)
Rabbit hutches	3 (10.7)	11 (13.8)	4 (9.1)	10 (4.3)	0 (0.0)
Cattle barns	1 (3.6)	4 (5.0)	0 (0.0)	0 (0.0)	0 (0.0)
Pig pens	3 (10.7)	4 (5.0)	0 (0.0)	0 (0.0)	0 (0.0)
Mixed composition	3 (10.7)	7 (8.8)	5 (11.4)	9 (3.8)	6 (3.4)
Without vertebrates	5 (17.9)	10 (12.5)	4 (9.1)	8 (3.4)	4 (2.3)
Total	28	80	44	235	174

IR sites: number of indoor resting collection sites surveyed; N_C: number of collections performed; N_{PC}: number of collections positive for *Cx. pipiens s.l.*; N_F: number of *Cx. pipiens s.l.* females collected; N_{BF}: number of blood-fed females collected; Mixed composition: Shelter with domestic birds and domestic mammals. Values in brackets represent relative frequencies (in percentage).

Microsatellite analysis

The best combination of six microsatellites (assigned score of 92.0%) included loci CxpGT04, CQ26, CxpGT20, CxpGT12, CQ41, and CxpGT40 (see Additional file 1, Table S2). The analysis with six loci was able to split the Gomes *et al.* (2009) dataset in to two clusters with a highly similar result to that obtained with 13 loci (see Additional file 1, Table S3). Taking the 13 loci dataset as the golden standard, the analysis with six loci had an average accuracy (*i.e.* average of the number of correctly identified individuals for a class over the total number of individuals assigned to that class) of 81.6% and average power (*i.e.* average of the number of correctly identified individuals for a class over the actual number of individuals of that class) of 88.6%.

Of the IR collections, only blood-fed females caught inside shelters with vertebrate hosts were selected for molecular genotyping (*N* = 174). Of these, four specimens failed in PCR amplifications and were thus excluded. A total of 170 females from five of the seven localities (Cambado: *N* = 14; Comporta: *N* = 27; Pego: *N* = 47; Possanco: *N* = 80; Torre: *N* = 2) were analysed. In addition to IR mosquitoes, subsamples from CDC-C (*N* = 39, of which 9 were blood fed), CDC-G (*N* =42) and HLC (*N* = 40) were also included, giving a total of 291 specimens used for molecular identification and microsatellite genotyping. All specimens were molecularly identified as *Cx. pipiens s.s.* by PCR (Smith & Fonseca, 2004).

Bayesian clustering analysis implemented by STRUCTURE revealed two clusters (Figure 1A).

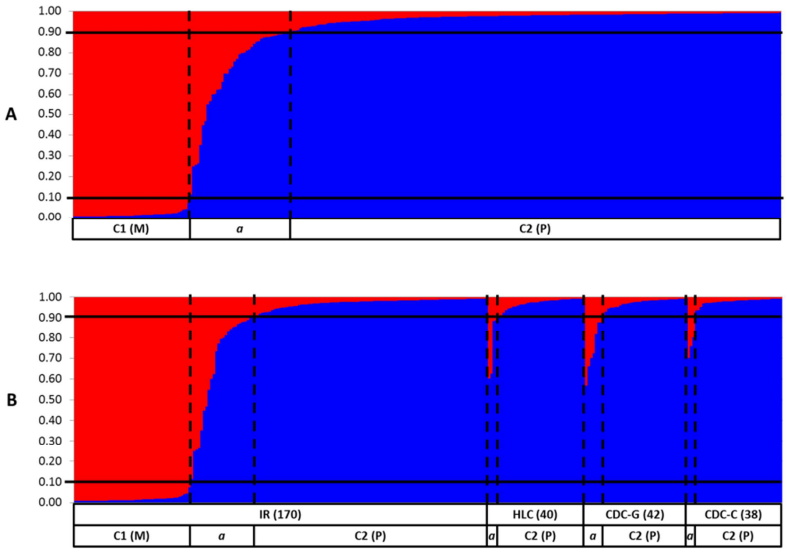

Figure 1. Bayesian cluster analysis of *Cx. pipiens s.s.* mosquitoes conducted by STRUCTURE in Comporta (2010).

A: Individuals sorted by their ancestral probability; B: Individuals sorted by collection method and ancestral probability; IR: indoor resting inside shelters; CDC-G: CDC light traps in ground level; CDC-C: CDC light traps in canopy of trees; HLC: human landing catches; *a:* admixed individuals ($0.1<Tq<0.9$). Columns correspond to the multilocus genotype of each individual, partitioned in different colours representing the probability of ancestry (q_i) to each cluster (Red: molestus; Blue: pipiens). Individuals were ordered according to their geographic information. Lines indicate the q_i threshold used to determine admixed individuals (see Methods).

Cluster 1 grouped 48 specimens of which 42 (87.5%) were classified as molestus form by the CQ11FL locus (Table 2). The majority (84.3%) of the 204 specimens in cluster 2 was classified as pipiens form by CQ11FL (Table 2). There were 39 females

exhibiting an admixed ancestry (*i.e.* $q_i \geq 0.10$ for both clusters). Of these, seven (17.9%) had a heterozygous CQ11FL200/250 genotype while the majority (*N* = 31, 79.5%) were classified as pipiens form by CQ11FL PCR (Table 2). There were twelve individuals displaying a 350 bp CQ11FL allele. Of these, 11 were grouped in the pipiens cluster, while one CQ11FL200/350 heterozygote was assigned to the admixed cluster (Table 2).

Table 2 Genotypic frequencies at the CQ11FL locus in each of the ancestry clusters revealed by STRUCTURE.

	N	**CQ11FL genotype**		
		250/250	**200/250**	**200/200**
Cluster 1 (molestus)	48	42 (87.5)	2 (4.2)	4 (8.3)
Cluster 2 (pipiens)	204[a]	6 (2.9)	25[b] (12.3)	172[c] (84.3)
Admixed	39	1 (2.6)	7 (17.9)	31[d] (79.5)
Total	291	49 (11.7)	34 (16.8)	207 (71.1)

N: number of individuals; Values in parenthesis refer to the frequencies (in percentage) within each cluster. [a] includes one specimen without CQ11FL identification; [b] includes one CQ11FL250/350 heterozygote. [c] includes one CQ11FL350/350 homozygote and nine CQ11FL200/350 heterozygotes. [d] includes one CQ11FL200/350.

Genetic diversity estimates for the 6 microsatellite loci analysed for the whole dataset (*N* = 291) and in subsamples determined by clustering analysis (STRUCTURE) and by sampling type (*i.e.* collections inside animal shelters versus outdoor collections) are shown in Table S4 (see Additional file 1). Significant departures from Hardy-Weinberg equilibrium were detected at 5 loci (83.3%) when all specimens were analysed as a single sample (see Additional file 1, Table S4). However, when the sample was subdivided according to clustering assignment and sampling site, significant heterozygote deficits were observed only on six occasions (21.4% out of 28 tests). These departures were generally associated with significant positive F_{IS} values indicative of a heterozygote deficit. Exact tests of linkage disequilibrium revealed 12 (80.0%) significant associations between pairs of loci for the whole dataset. When samples were divided by clustering assignment and type of sampling site, only one significant association was observed (1.3% out of 75 combinations). The analysis

performed by Micro-Checker did not find a consistent signal of null alleles in any loci. All microsatellite loci were maintained for subsequent analyses.

Bayesian clustering analysis showed a non-uniform distribution of the forms among collection methods (Figure 1B). All specimens with a molestus genetic background were sampled solely by IR collections, whereas individuals with pipiens or admixed ancestry were collected by both IR and outdoor collections (*i.e.* CDC-C, CDC-G and HLC). In IR collections, the proportion of molestus individuals caught inside chicken coops was 28.8% of the total catch and 16.7% in avian-mammal mixed shelters. The proportion of admixed individuals caught by IR (14.7%, 25 out of 170) was comparable to that sampled by outdoors collection methods (11.6%, 14 out of 121).

The distribution of *Cx. pipiens s.s.* forms in IR collections appeared not to be homogenous among the localities surveyed (Figure 2). Individuals of molestus ancestry were concentrated mainly in Pego (79.2%), constituting 80.9% of the total IR catch at this locality. The proportion of IR collections made at avian shelters (*i.e.* chicken coops) in Pego was 45.0% whereas it varied between 61.5% and 83.3% in the two localities where the pipiens form predominated (Comporta and Possanco; Figure 2).

Blood meal identification

Blood meal identification by ELISA revealed that most (N = 159; 93.5%) of the 170 blood feeds analysed were from avian hosts (Table 3). The proportion of blood meals taken on avian hosts by pipiens (95.9%) and molestus (91.6%) forms was not significantly different (Fisher's exact test: P = 0.108; Table 3). All admixed individuals fed on avian hosts. There were only three single blood meals taken on mammalian hosts. All consisted of human blood taken by two molestus and one pipiens females. There were also two females (one molestus and one admixed) with a mixed blood meal with cow and avian blood. The ELISA did not identify the origin of six blood meals (three pipiens, two molestus and one hybrid).

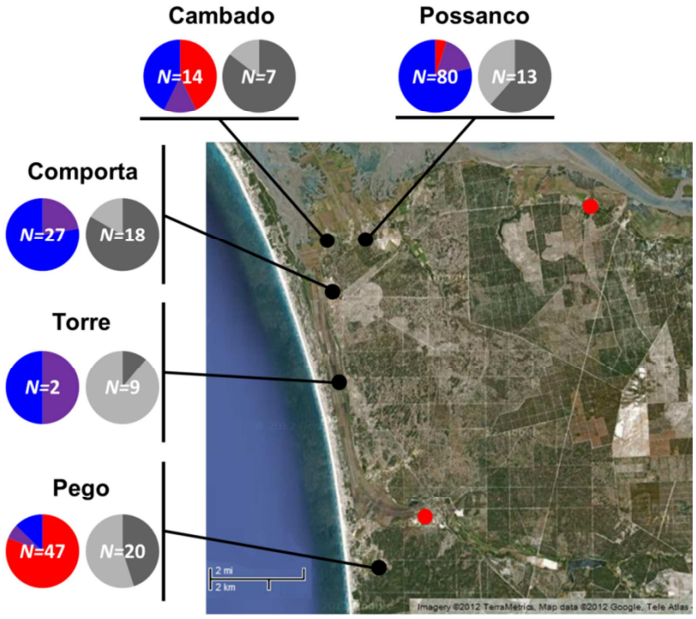

Figure 2. Frequency of the groups defined by the STRUCTURE by locality.

Black dot: positive sample site for *Cx. pipiens s.s.* (Cambado, Comporta, Pego, Possanco, Torre); Red dot: negative sample site for *Cx. pipiens s.s.* (Carvalhal, Monte Novo do Sul). Color graphics: proportion of females; Red: cluster 1 (molestus form); Blue: cluster 2 (pipiens form); Purple: Admixed (hybrids). Grey-scale graphics: proportion of mosquito collections; Dark grey: proportion of collection performed in chicken coops; Light grey: proportion of collection performed in other type of shelter.

Sequence analysis of the *cyt b* gene in blood samples was performed for the nine engorged females caught in light traps from the tree canopy (CDC-C) and 19 specimens from IR collections (the six females without ELISA identification, two females with mixed feeds, three females with only mammalian blood, and eight females with only avian blood). Two samples did not amplify *cyt b* gene of any vertebrate (one molestus female without identification and one female from canopy). The *cyt b* analysis confirmed the ELISA results for the females with single feed but identified only chicken (*Gallus gallus* L. 1758; GenBank: DQ512918.1) mtDNA in the blood of the two females with mixed feeds. Two bird species were identified in the five females caught IR without ELISA identification: house sparrow (*Passer domesticus*; GenBank: AY495393.1) in four females (two pipiens, one molestus and one hybrid), and blackbird (*Turdus merula* L. 1758; GenBank: EU154637.1) in one pipiens female. In the nine females collected by CDC-C, two other bird species were identified: long-eared owl (*Asio otus* L. 1758; GenBank: AF082067.2) blood in eight females (seven pipiens and one hybrid) and blue tit (*Cyanistes caeruleus* (L. 1758); GenBank: AF347961.1) blood in one pipiens female.

Table 3. Frequencies of blood meals identified by Sandwich ELISA in each of the ancestry clusters revealed by STRUCTURE of indoor collections.

	N	Blood feed – Indoor			
		Mammal	**Bird**	**Mixed**	**WI**
Cluster 1 (molestus)	48	2 (4.2)	43 (89.5)	1 (2.1)	2 (4.2)
Cluster 2 (pipiens)	97	1 (1.0)	93 (95.9)	0 (0.0)	3 (3.1)
Admixed	25	0 (0.0)	23 (92.0)	1 (4.0)	1 (4.0)
Total	170	3 (1.8)	159 (93.5)	2 (1.2)	6 (3.5)

N: number of individuals; Mammal: feeds in mammal (all in Human); Bird: feeds in Bird (chicken antibody); Mixed: mixed fed in mammal and bird (all in cow and chicken); WI: without positive identification. Values in parenthesis refer to the frequencies (in percentage) within each cluster.

Discussion

In this study, a notable difference was found in the distribution of molestus and pipiens forms according to collection methods. While the pipiens form was sampled by all methods, molestus individuals were caught only in IR collections. This result suggests differences between forms in biting and resting behaviours. When placed outdoors, CDC light traps are appropriate for sampling both host seeking mosquitoes and recently blood-fed mosquitoes searching for a suitable resting site (WHO, 1975). These traps have been successfully used as an alternative to outdoor resting collections in feeding pattern studies of *Cx. pipiens s.l.* conducted in the USA (Kilpatrick *et al.*, 2006, Molaei *et al.*, 2006). The absence of the molestus form from outdoor CDC light trap collections may suggest a more endophagic and endophilic behaviour of this form. A tendency of the molestus form to bite indoors was further highlighted by its absence from outdoor landing catches. These results point to a predominantly indoor and synanthropic behaviour of the molestus form, as described for populations of this form at northern latitudes where inter-form hybridisation is rare (Byrne & Nichols, 1999; Vinogradova, 2000; Spielman, 2001). Therefore, it appears that in spite of the high hybridisation levels and in addition to autogeny and stenogamy, the molestus population of Comporta maintains behavioural phenotypes typical of this form. This observation is

consistent with a pure molestus genetic background found in the region, which contrasted with a more introgressed pipiens background (Gomes *et al.*, 2009).

Also compatible with a pattern of asymmetric hybridisation, with more molestus genes introgressing the pipiens form, is an apparently more plastic resting behaviour of the pipiens form, suggested by the fact that blood-fed females of this form were collected both indoors and outdoors. However, the number of blood-fed *Cx. pipiens s.s.* females collected in outdoor CDC light traps (9 out of 1,718) was much lower than those in IR collections (174 out of 235). Furthermore, the apparent behavioural differences observed between pipiens and molestus forms should be considered with caution given the sampling design used in this study, which did not include paired collections with the same method. Additional surveys involving paired indoor/outdoor landing catches (to directly evaluate endo/exophagy) and indoor/outdoor resting collections would be required to confirm these observations.

The approach used for the selection of microsatellites to differentiate molestus and pipiens forms allowed reducing the number of loci to be genotyped from 13 to 6 whilst maintaining high accuracy and power. The efficiency of multilocus analyses tends to increase with the number of microsatellites (Vähä & Primmer, 2006). However, the use of a more limited number of loci can benefit their application in surveillance studies by minimising genotyping costs and thus allowing genotyping of larger sampling sizes. Given the importance of accurately determining the intra-specific composition of *Cx. pipiens s.s.* it is recommend that similar microsatellite-based approaches are used in epidemiological surveys to complement the information based on a single marker (CQ11FL) that has limitations in areas of continued introgression (Bahnck & Fonseca, 2006; Gomes *et al.*, 2009).

As in the survey conducted in 2005-2006 (Gomes *et al.*, 2009), sympatric molestus and pipiens populations displaying high hybridisation levels were identified aboveground in the region of Comporta. However, a higher proportion of the molestus form was found in 2005-2006 survey (66.2%; Gomes *et al.*, 2009), whereas in the present study the pipiens form prevailed (70.1%). This difference most likely reflects the outdoor sampling carried out in this study and which was not carried out in the previous survey. In addition, the survey of 2005/2006 was mainly concentrated in the

locality of Pego (*ca.* 77% of females), where 79% of molestus individuals were collected in the present survey.

Blood meal analysis revealed that the great majority of *Cx. pipiens s.s.* females fed on avian hosts. The pipiens form showed a slightly higher proportion of avian blood feeds when compared with the molestus form. However, this difference was non-significant and the proportion of avian blood meals was above 90% in both forms, suggesting an ornithophilic tendency for *Cx. pipiens s.s.* in the region. Such an ornithophilic tendency was also observed in a study analysing *Cx. pipiens s.l.* from urban and countryside areas of Portugal, in which over 70% of the females fed on birds (Osório *et al.,* 2012), and from south-west countryside areas of Spain, where over 80% of the females fed on birds (Muñoz *et al.,* 2012). The pipiens form has been described as ornithophilic whereas molestus populations were recognised as being mammophilic (Harbach *et al.,* 1984, 1985). However, the feeding patterns of *Cx. pipiens s.s.* populations depend not only on their genetic background but also on the availability of vertebrate hosts and on host defensive mechanisms (Kilpatrick *et al.,* 2007; Balenghien *et al.,* 2011). Consequently, exceptions to the general feeding pattern have been reported for both forms in the USA and in the Mediterranean region (Ribeiro *et al.,* 1983; Kilpatrick *et al.,* 2007; Huang *et al.,* 2009). Hybridisation between the two forms may also promote a catholic feeding behaviour in *Cx. pipiens s.s.* (Fonseca *et al.,* 2004). Such behaviour would thus increase the relative importance of host availability and host defensive mechanisms in the feeding pattern of the mosquito population.

While molestus and pipiens appear to be mainly ornithophilic in the Comporta region, this may reflect host availability in the region rather than an intrinsic host preference. A lower availability of mammals (including humans) is suggested by a higher proportion of chicken coops (46.4%) when compared to mammalian shelters without domestic birds (25.0%), and the well-built and protected human dwellings with door and window screens that prevent mosquito entry (Sousa, 2008). On the other hand, pipiens form mosquitoes were caught biting humans outdoors in HLC collections which play in favour of a more opportunistic feeding pattern promoted by hybridisation. Altogether, these findings suggest a closer association of both molestus and pipiens forms with avian hosts and that this ornithophilic tendency, albeit possibly genetically conditioned, is primarily modulated by host availability in the region. In this scenario,

molestus females, with preference for biting mammalian hosts, may feed more readily on the available bird hosts which may increase the odds for alternate feeding on birds and mammals. This feeding behaviour increases the risk of WNV transmission to humans and domestic mammals from birds, amplification hosts.

Blood meal host identification based on mtDNA sequencing identified bird species from Passeriformes and Strigiformes orders. Birds from these orders were identified with anti-WNV antibodies in Portugal indicating the circulation of WNV in these populations (Formosinho *et al.,* 2006). The Passeriformes are a well-known WNV reservoir (Komar *et al.,* 2003; Wheeler *et al.,* 2009). Species of this order, such as *Passer domesticus*, displayed the highest WNV prevalence in USA (Molaei *et al.,* 2007; Hamer *et al.,* 2009).

Conclusion

The presence of females from both forms collected inside domestic animal shelters with a blood meal taken from wild Passeriformes gives a clear indication of the proximity between the WNV natural cycle and the human population in the Comporta region. Species such as the house sparrow and the blackbird have tolerance for humans and the blood meal could have been taken indoors when those birds enter in human constructions searching for food or shelter. However, the combination of the genetic structure and blood meal analysis suggest that at least a proportion of pipiens form females may bite outdoors in sylvian habitats and then search for anthropogenic indoor resting sites to complete their gonotrophic cycle. In both scenarios, alternative domestic hosts and humans are available in those sites for subsequent blood feeding which may promote the accidental transmission of WNV and other arboviruses in this region.

Acknowledgments

This study was funded by Fundação para a Ciência e a Tecnologia/FEDER, Portugal (POCI/BIA-BDE/57650/2004 and PPCDT/BIA-BDE/57650/2004). BG was funded by a PhD fellowship of Fundação para a Ciência e Tecnologia/FEDER (SFRH/BD/36410/2007).

References

Almeida, A.P.G., Freitas, F.B., Novo, M.T., Sousa, C.A., Rodrigues, J.C., Alves, R. & Esteves, A. (2010) Mosquito surveys and West Nile virus screening in two different areas of southern Portugal, 2004–2007. *Vector-Borne and Zoonotic Diseases*. 10 (7), 673–680.

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), 251–255.

Balenghien, T., Fouque, F., Sabatier, P. & Bicout, D.J. (2011) Theoretical formulation for mosquito host-feeding patterns: Application to a West Nile virus focus of southern France. *Journal of Medical Entomology*. 48 (5), 1076–1090.

Banks, M.A., Eichert, W. & Olsen, J.B. (2003) Which genetic loci have greater population assignment power? *Bioinformatics*. 19 (11), 1436–1438

Byrne, K. & Nichols, R.A. (1999) *Culex pipiens* in London Underground tunnels: differentiation between surface and subterranean populations. *Heredity*. 82 (1), 7–15.

Calzolari, M., Bonilauri, P., Bellini, R., Albieri, A., Defilippo, F., Maioli, G., Galletti, G., Gelati, A., Barbieri, I., Tamba, M., Lelli, D., Carra, E., Cordioli, P., Angelini, P. & Dottori, M. (2010) Evidence of simultaneous circulation of West Nile and Usutu viruses in mosquitoes sampled in Emilia-Romagna region (Italy) in 2009. *PLoS ONE*. 5 (12), e14324.

Calzolari, M., Gaibani, P., Bellini, R., Defilippo, F., Pierro, A., Albieri, A., Maioli, G., Luppi, A., Rossini, G., Balzani, A., Tamba, M., Galletti, G., Gelati, A., Carrieri, M., Poglayen, G., Cavrini, F., Natalini, S., Dottori, M., Sambri, V., Angelini, P. & Bonilauri, P. (2012) Mosquito, bird and human surveillance of West Nile and Usutu viruses in Emilia-Romagna region (Italy) in 2010. *PLoS ONE*. 7 (5), e38058.

Chevillon, C., Eritja, R., Pasteur, N. & Raymond, M. (1995) Commensalism, adaptation and gene flow: mosquitoes of the *Culex pipiens* complex in different habitats. *Genetical Research*. 66 (2), 147–157.

Clements, A.N. (1999) *The Biology of Mosquitoes. Sensory Reception and Behaviour Volume 2*. Wallingford, CABI Publishing.

Donnelly, M.J., Cuamba, N., Charlwood, J.D., Collins, F.H. & Townson, H. (1999) Population structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa. *Heredity*. 83 (4), 408–417.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*. 14 (8), 2611–2620.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*. 1, 47–50.

Fonseca, D.M., Keyghobadi, N., Malcolm, C.A., Mehmet, C., Schaffner, F., Mogi, M., Fleischer, R.C. & Wilkerson, R.C. (2004) Emerging vectors in the *Culex pipiens* Complex. *Science*. 303 (5663), 1535–1538.

Formosinho, P., Santos-Silva, M.M., Santos, A., Melo, P., Encarnação, V., Santos, N., Nunes, T., Agrícola, R. & Portas, M. (2006) O vírus West Nile em Portugal – estudos de vigilância epidemiológica. *Revista Portuguesa de Ciências Veterinárias*. 101 (557-558), 61–68.

Gomes, B., Sousa, C.A., Novo, M.T., Freitas, F.B., Alves, R., Corte-Real, A.R., Salgueiro, P., Donnelly, M., Almeida, A.P.G. & Pinto, J. (2009) Asymmetric introgression between sympatric molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in the Comporta region, Portugal. *BMC Evolutionary Biology*. 9 (1), 262.

Goudet, J. (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. *The Journal of Heredity*. 86 (6), 485–486.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 41, 95–98.

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), 125–128.

Hamer, G.L., Kitron, U.D., Goldberg, T.L., Brawn, J.D., Loss, S. R., Ruiz, M. O., Hayes, D. B. & Walker, E.D. (2009) Host selection by *Culex pipiens* mosquitoes and West Nile virus amplification. *American Journal of Tropical Medicine and Hygiene*. 80 (2), 268–278.

Harbach, R.E., Harrison, B.A. & Gad, A.M. (1984) *Culex (Culex) molestus* Forskål (Diptera, Culicidae) - neotype designation, description, variation, and taxonomic status. *Proceedings of the Entomological Society of Washington*. 86 (3), 521–542.

Harbach, R.E., Dahl, C. & White, G.B. (1985) *Culex (Culex) pipiens* Linnaeus (Diptera, Culicidae) - concepts, type designations, and description. *Proceedings of the Entomological Society of Washington*. 87 (1), 1–24.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 6 (2), 65–70.

Huang, S., Molaei, G. & Andreadis, T.G. (2008) Genetic insights into the population structure of *Culex pipiens* (Diptera: Culicidae) in the northeastern United States by using microsatellite analysis. *American Journal of Tropical Medicine and Hygiene*. 79 (4), 518–527.

Huang, S., Hamer, G.L., Molaei, G., Walker, E.D., Goldberg, T.L., Kitron, U.D. & Andreadis, T.G. (2009) Genetic variation associated with mammalian feeding in *Culex pipiens* from a West Nile virus epidemic region in Chicago, Illinois. *Vector-Borne and Zoonotic Diseases*. 9 (6), 637–642.

Hubálek, Z. (2008) Mosquito-borne viruses in Europe. *Parasitology Research*. 103 (1), 29–43.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 23 (14), 1801–1806.

Kalinowski, S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*. 5 (1), 187–189.

Kilpatrick, A.M., Kramer, L.D., Jones, M.J., Marra, P.P., Daszak, P. & Fonseca, D.M. (2006) West Nile virus epidemics in North America are driven by shifts in mosquito feeding behavior. *PLoS Biology*. 4 (4), e82.

Kilpatrick, A.M., Kramer, L.D., Jones, M.J., Marra, P.P. & Daszak, P. (2007) Genetic influences on mosquito feeding behavior and the emergence of zoonotic pathogens. *American Journal of Tropical Medicine and Hygiene*. 77 (4), 667–671.

Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R. & Bunning, M. (2003) Experimental infection of North American birds with the New York 1999 strain of West Nile virus. *Emerging Infectious Diseases*. 9 (3), 311–322.

Lowry, R. (2012) *VassarStats: Web Site for Statistical Computation*. [Online]. 2012. Available from: http://vassarstats.net/index.html [Accessed: 5 October 2012].

Medlock, J.M., Snow, K.R. & Leach, S. (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.

Molaei, G., Andreadis, T.G., Armstrong, P.M., Anderson, J.F. & Vossbrinck, C. R. (2006) Host feeding patterns of *Culex* mosquitoes and West Nile virus transmission, Northeastern United States. *Emerging Infectious Diseases*. 12 (3), 468–474.

Molaei, G., Andreadis, T.G., Armstrong, P.M., Bueno Jr., R., Dennett, J. A., Real, S.V., Sargent, C., Bala, A., Randle, Y., Guzman, H., da Rosa, A.T., Wuithiranyagool, T. & Tesh, R.B. (2007) Host feeding pattern of *Culex quinquefasciatus* (Diptera: Culicidae) and its role in transmission of West Nile virus in Harris County, Texas. *The American Journal of Tropical Medicine and Hygiene*. 77 (1), 73–81.

Muñoz, J., Ruiz, S., Soriguer, R., Alcaide, M., Viana, D.S., Roiz, D., Vázquez, A. & Figuerola, J. (2012) Feeding patterns of potential West Nile virus vectors in south-west Spain. *PLoS ONE*. 7 (6), e39549.

Nei, M. (1987) *Molecular Evolutionary Genetics*. New York, Colombia University Press.

Nudelman, S., Galun, R., Kitron, U. & Spielman, A. (1988) Physiological characteristics of *Culex pipiens* populations in the Middle East. *Medical and Veterinary Entomology*. 2 (2), 161–169.

van Oosterhout, C., Hutchinson, W., Wills, D. & Shipley, P. (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*. 4 (3), 535–538.

Osório, H.C., Zé-Zé, L. & Alves, M.J. (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.

Peel, M.C., Finlayson, B.L. & McMahon, T.A. (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*. 4 (2), 1633–1644.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*. 155 (2), 945–959.

Rappole, J.H. & Hubálek, Z. (2003) Migratory birds and West Nile virus. *Journal of Applied Microbiology*. 94 (1), 47–58.

Ribeiro, H., Pires, C.A., Ramos, H.C. & Capela, R.A. (1983) Research on the mosquitoes of Portugal (Diptera, Culicicae). VIII- On the occurrence of *Culex (Culex) molestus* Forskål, 1775. *Jornal Sociedade Ciências Médicas Lisboa*. 147, 185–188.

Ribeiro, H. & Ramos, H.C. (1999) Identification keys of the mosquitoes (Diptera: Culicidae) of Continental Portugal, Açores and Madeira. *European Mosquito Bulletin*. 3, 1–11.

Simões, M.J., Próspero, M.I. & Henrique, R. (1995) Optimização da técnica ELISA “two sites” utilizada na identificação de refeições sanguíneas de mosquitos. *Revista Portuguesa de Doenças Infecciosas*. 18, 225–229.

Slatkin, M. & Excoffier, L. (1996) Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity*. 76 (4), 377–383.

Smith, 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 Hygiene*. 70 (4), 339–345.

Sousa, C.A. (2008) *Malaria vectorial capacity and competence of Anopheles atroparvus Van Thiel, 1927 (Diptera, Culicidae): Implications for the potential re-emergence of malaria in Portugal*. Universidade Nova de Lisboa.

Spielman, A. (2001) Structure and seasonality of nearctic *Culex pipiens* populations. *Annals of the New York Academy of Sciences*. 951, 220–234.

Sudia, W.D. & Chamberlain, R.W. (1962) Battery-operated light trap. An improved model. *Mosquito News*. 22, 126–129.

Toma, L., Cipriani, M., Goffredo, M., Romi, R. & Lelli, R. (2008) First report on entomological field activities for the surveillance of West Nile disease in Italy. *Veterinaria Italiana*. 44 (3), 499–512.

Vähä, J.-P. & Primmer, C.R. (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridisation scenarios and with different numbers of loci. *Molecular Ecology*. 15 (1), 63–72.

Vinogradova, E.B. (2000) *Culex pipiens pipiens mosquitoes: taxonomy, distribution, ecology, physiology, genetics, applied importance and control*. S.I. Golovatch (ed.). Sofia, Pensoft.

Wheeler, S.S., Barker, C.M., Fang, Y., Veronica Armijos, M., Carroll, B.D., Husted, S., Johnson, W.O. & Reisen, W.K. (2009) Differential Impact of West Nile Virus on California Birds. *The Condor*. 111 (1), 1–20.

World Health Organization (WHO) (1975) *Manual on Practical Entomology in Malaria. Part II: Methods and Techniques Geneva*. Geneva.

Zeller, H.G. & Schuffenecker, I. (2004) West Nile virus: an overview of its spread in Europe and the Mediterranean basin in contrast to its spread in the Americas. *European Journal of Clinical Microbiology & Infectious Diseases*. 23 (3), 147–156.

Additional file 1

Table S1. Localities surveyed, number of sites sampled and number of collections performed for each collection method.

Locality	latitude	longitude	Collection method							
			IR		CDC-C		CDC-G		HLC	
Cambado	38°23'39"N	8°47'16"W	3	(7)	2	(4)	1	(1)	0	(0)
Carvalhal	38º18'35"N	8º45'05"W	2	(6)	3	(3)	1	(1)	0	(0)
Comporta	38º22'42"N	8º46'59"W	7	(18)	2	(3)	0	(0)	1	(6)
Monte Novo do Sul	38°24'28"N	8°40'53"W	4	(7)	1	(1)	1	(1)	0	(0)
Pego	38°17'36"N	8°46'05"W	4	(20)	1	(2)	1	(1)	0	(0)
Possanco	38°23'48"N	8°46'23"W	4	(13)	2	(2)	1	(1)	0	(0)
Torre	38°21'09"N	8°46'50"W	4	(9)	1	(2)	2	(2)	0	(0)
Total			28	(80)	12	(17)	7	(7)	1	(6)

IR: number of sites (shelters) sampled by indoor resting collections; CDC-C: number of sites sampled by CDC light traps in canopy of trees; CDC-G: number of sites sampled by CDC light traps at ground level; HLC: number of sites sampled by human landing catches. Values in parenthesis refer to the number of collections performed.

Table S2. Loci ranking performed by WHICHLOCI with 12 microsatellites.

Rank	Locus	Score	Score (%)	A (%)
1	**CxpGT04**	0.688	10.27	92.00
2	**CQ26**	0.661	9.87	
3	**CxpGT20**	0.660	9.85	
4	**CxpGT12**	0.655	9.78	
5	**CQ41**	0.649	9.69	
6	**CxpGT40**	0.618	9.22	
7	CxpGT51	0.557	8.31	NA
8	CxqTri4	0.513	7.65	
9	CxqGT6b	0.483	7.20	
10	CxpGT53	0.463	6.90	
11	CxpGT46	0.403	6.01	
12	CxqGT4	0.351	5.24	

A: correct assignment with 6 loci; NA: Not applicable.

Table S3. Accuracy and power of the clustering analysis performed by STRUCTURE (Prichard *et al.* 2000) with 6 loci for the 13 microsatellites dataset of Gomes *et al.* (2009).

	Golden standard (13 loci)	**Assigned (6 loci)**	**Correctly assigned (6 loci)**	**Power**	**Accuracy**
Cluster 1 (molestus)	96	88	88	0.916	1.000
Cluster 2 (pipiens)	36	38	35	0.972	0.921
hybrids	13	19	10	0.769	0.526

Power: number of correctly identified individuals for a class over the actual number of individuals of that class; Accuracy: number of correctly identified individuals for a class over the total number of individuals assigned to that class. Individual assignment was based on a $Tq>0.9$, hybrids: $0.1<Tq<0.9$.

Table S4. Genetic diversity at microsatellite loci of *Culex pipiens s.s.* from Comporta.

Locus		Inside animal shelters			Outdoor		Total (*N*=291)
		P (*N*=98)	H (*N*=25)	M (*N*=48)	P (*N*=107)	H (*N*=14)	
CQ26	$A_{R(26)}$	6.4	7.1	6.0	7.4	6.9	7.7
	H_e	0.768*	**0.820***	**0.765***	0.813	0.804	**0.825**
	F_{IS}	0.151	**0.370**	**0.349**	0.078	0.207	**0.209**
CQ41	$A_{R(26)}$	9.8	9.7	6.4	10.3	8.8	10.5
	H_e	0.839*	0.840	0.795	**0.842***	0.817	**0.853**
	F_{IS}	**0.163**	0.145	-0.098	**0.273**	0.220	**0.184**
CxpGT04	$A_{R(26)}$	8.7	7.3	3.3	8.7	8.9	8.5
	H_e	0.864	0.735	0.573	0.865	0.857	0.836
	F_{IS}	-0.014	-0.091	-0.277	-0.005	0.000	-0.008
CxpGT12	$A_{R(26)}$	7.0	4.7	2.3	6.9	9.0	6.7
	H_e	**0.768***	**0.596***	0.463	**0.804***	0.840	**0.769**
	F_{IS}	**0.236**	0.422	0.237	**0.235**	0.087	**0.289**
CxpGT20	$A_{R(26)}$	15.3	13.8	8.3	14.4	12.5	14.4
	H_e	0.945	0.926	0.854	0.925*	0.915	**0.937**
	F_{IS}	0.062	0.095	0.074	**0.107**	0.226	**0.107**
CxpGT40	$A_{R(26)}$	4.4	8.2	5.1	8.9	9.0	6.0
	H_e	0.425	0.718	0.514	0.513	0.712	**0.619**
	F_{IS}	0.176	0.111	0.110	0.070	-0.004	**0.262**
All loci	$A_{R(26)}$	8.6	8.2	5.1	8.9	9.0	9.0
	H_e	**0.768**	**0.773**	0.661	**0.794**	0.824	**0.806**
	F_{IS}	**0.121**	**0.169**	0.066	**0.129**	0.127	**0.167**

P: pipiens cluster; H: *admixed* individuals; M: molestus cluster; $A_{R(26)}$: allelic richness for a minimum sample sizes of 26 genes (13 individuals); H_e: expected heterozygosity; F_{IS}: inbreeding coefficient. Values in bold indicate a significant *P*-value after correction for multiple tests (see Methods). Asterisks indicate presence of null alleles determined by Micro-Checker. Per locus and over sample Hardy-Weinberg tests were performed using ARLEQUIN. For over loci estimates the global test available in FSTAT was used.

Chapter 5.

Low levels of genomic divergence among forms of the *Culex pipiens* under different ecological pressures

To be submitted as:

Gomes, B., Wilding, C.S., Weetman, D., Sousa, C., Novo, M.T., Savage, H.M., Almeida, A.P.G., Pinto, J. & Donnelly, M.J. Low levels of genomic divergence among forms of the *Culex pipiens* under different ecological pressures.

Abstract

The West Nile virus vector *Culex pipiens s.s.* is divided into two intraspecific forms termed pipiens and molestus, characterized by differing ecological traits. Whilst in northern Europe and the USA these forms occupy distinct habitats (aboveground and underground), in southern Europe they are found sympatrically aboveground. Previous molecular studies have shown common ancestry of geographically distinct populations of each form. However, the levels and patterns of genetic differentiation across the genome remain unknown. Here, an amplified fragment length polymorphism (AFLP) based genome scan was undertaken on samples collected from both sympatric and allopatric populations from Europe and USA in order to quantify the extent and consistency of differentiation between the two forms. The forms pipiens and molestus were clearly distinct but with major sub-structuring between continents within each form, and also more marked differentiation among European molestus than pipiens populations. Three outlier analyses applied to 810 loci showed low genomic divergence between pipiens and molestus (1.4% – 3.1%), which is consistent with sympatric speciation with gene flow. Only two outlier common loci (0.25%) were detected in both Europe and the USA suggesting a low number of genomic regions involved in the typological traits (*i.e.* autogeny, stenogamy, ability for diapause) that influence the adaptation of molestus to anthropogenic habitats and the speciation process between pipiens and molestus forms.

Introduction

Divergent selection is a major driving force in speciation models involving taxa with overlapping geographic distributions, either as sympatric speciation *per se* or via reinforcement of isolation between allopatric incipient species after secondary contact (Nosil *et al.,* 2009, 2012; Hopkins & Rausher 2011). The capacity for divergent selection to promote reproductive isolation among populations depends on the strength of selection, the number of traits upon which it is acting and rates of realised gene flow (Nosil *et al.,* 2009). Multifarious selection (divergent selection acting upon multiple traits) only appears sufficient to cause speciation when gene flow is low (*i.e.* allopatric speciation; Nosil *et al.,* 2009). However, strong selection concentrated on a few traits

may overcome substantial gene flow, at least in specific genomic regions, which initiate sympatric speciation (Wu, 2001). This mechanism is considered a major process of sympatric/ecological speciation and has been described in several insect species (Machado *et al.,* 2002; Turner *et al.,* 2005; Egan *et al.,* 2008; Weetman *et al.*, 2012).

In insect groups of medical importance, the evolutionary relevance of the speciation process also has a public health dimension. *Culex pipiens s.s.* is a widespread mosquito species with an important medical and veterinary impact owing to its role in the transmission of arthropod-borne viruses (arboviruses) such as the potentially fatal West Nile virus (Solomon, 2004). *Culex pipiens s.s.* comprises two distinct forms, denoted pipiens and molestus, which are morphologically indistinguishable but exhibit behavioural and physiological differences that may impact their ability to transmit pathogens. The molestus form is differentiated from the pipiens form by four ecological/physiological characteristics: autogeny (the capacity to lay eggs without taking a blood meal), stenogamy (the capacity to mate in confined spaces), homodynamy (a continuous life cycle without diapause), and mammophily (a preference to bite mammals, including humans) (Harbach *et al.,* 1984, 1985).

In southern European/Mediterranean regions, the two *Cx. pipiens s.s.* forms are sympatric in aboveground habitats, but in northern regions of Europe, Russia and the USA, molestus and pipiens forms segregate into underground and aboveground habitats, respectively (Vinogradova, 2000; Fonseca *et al.,* 2004; Gomes *et al.,* 2009). A continuous life cycle may be a limitation for surviving in colder climates which may restrain the habitat choice of molestus, while autogeny and stenogamy are important traits for survival in underground and confined habitats with restricted access to blood meal sources. The consequences of the different ecological pressures in underground (northern latitudes) and aboveground (southern latitudes) habitats at the genomic level remain unknown.

Populations with mixed characteristics between molestus and pipiens have been found in southern European regions (Callot & Van Ty, 1943; Pasteur *et al.*, 1977; Gomes *et al.,* 2009). In these regions, inter-form gene flow has been detected, resulting in a pattern of asymmetric introgression from molestus into pipiens (Gomes *et al.,* 2009). Moreover, a catholic feeding behaviour displayed by admixed molestus and

pipiens populations may increase the chance of accidental transmission of West Nile virus (WNV) from birds to mammals, including humans (Fonseca *et al.,* 2004; Hamer *et al.,* 2008).

The evolutionary origin of the molestus populations in northern latitudes has been under debate. In one hypothesis, the molestus form derived from the pipiens form by multiple independent adaptations to underground anthropogenic habitats (Vinogradova, 2000). The second hypothesis considers molestus as an evolutionarily independent entity. Under this scenario, colonization of northern underground habitats would have been made by molestus populations from southern latitudes (Fonseca *et al.,* 2004). Molecular studies with microsatellite loci showed common ancestry among geographically distinct populations of molestus, reinforcing its status as a single evolutionary entity (Fonseca *et al.,* 2004; Kothera *et al.,* 2010). The common ancestry revealed by microsatellite analyses suggests an on-going incipient speciation process, which may imply the existence of divergent genomic regions between molestus and pipiens forms. If involved in the speciation process, these divergent genomic regions are likely to be consistent among different geographic populations of the forms. Other genomic regions may vary among different geographic regions due to other selective pressures not involved in the speciation process or due to genetic drift.

In this study, an AFLP-based genome scan was performed in geographically distinct *Cx. pipiens s.s.* samples in order to identify outlier loci between molestus and pipiens, candidates to be under divergent selection. To identify outlier loci and reduce false positives caused by population substructure (Excoffier *et al.,* 2009), we explicitly tested for substructure within the data and used three outlier detection approaches. The objectives were to determine the extent of genomic divergence between molestus and pipiens forms and to infer about the implications of the divergence in the speciation process and in the adaptation to anthropogenic habitats by the molestus form.

Material and Methods

Mosquito samples

Six field samples analysed in this study were collected in three regions of Portugal and one in the United Kingdom, while two North American samples were obtained from laboratory colonies (Table 1).

North American form-specific colonies derived from field mosquitoes collected in the area of Chicago IL. The molestus colony was established with mosquitoes from a drainage sump collected by backpack aspirator and larval dipping in January 2009 (Mutebi & Savage, 2009) while the pipiens colony was established with overwintering adults collected from a large culvert by aspiration in January 2010. The mosquitoes used in this study were taken from the colonies in February 2011.

In Portugal, indoor resting females were collected using mechanical aspirators between May 2005 and August 2006 in Comporta (Gomes *et al.*, 2009), in Alqueva during June 2007, and in Sandim during August 2010. The sample from Comporta was split into molestus and pipiens based on different genetic signatures defined by clustering analysis with microsatellites (Gomes *et al.*, 2009). A second collection in Comporta was performed by CDC-light traps placed in trees between July and August 2010 (Comporta-Tree). The individuals of this sample were provisionally identified as pipiens by a diagnostic size polymorphism at the 5' flanking region of the CQ11 microsatellite (CQ11FL, Bahnck & Fonseca, 2006). The same marker classified all individuals collected in Sandim and Alqueva as molestus. However, it should be noted that under in areas of inter-form hybridisation, the CQ11FL marker is only partially effective in discriminating molestus and pipiens forms at the individual level (Gomes *et al.*, 2009).

The sampling in UK took place in March 2010, at the veterinary facility of the University of Liverpool, in Wirral. Adults overwintering inside farm buildings (a typical behaviour of the pipiens form) were collected by Pyrethrum Spray Collection and were provisionally classified as pipiens by the CQ11FL marker.

Table 1. Localities of the samples used in the AFLP protocol

Country	Locality	Latitude	Longitude	Method	Form	Insectary	Ref
Portugal	Alqueva	38°17'54"N	7°35'17"W	IR	molestus*	Au/St	-
	Comporta	38°21'09"N	8°46'51"W	IR	molestus	Au/St	[1]
					pipiens	N-Au/N-St	
				CDC	pipiens*	-	-
	Sandim	41°01'19"N	8°30'20"W	IR	molestus*	-	-
UK	Wirral	53°17'24"N	3°02'01"W	IR-I	pipiens*	-	-
USA	Chicago	41°43'09"N	87°45'23"W	MA	pipiens	N-Au/N-St	-
		41°39'49"N	87°36'30"W	BA/LC	molestus	Au/St	[2]

IR: Indoor resting collection with mechanical aspirators; IR-I: Indoor resting collections using insecticide spraying; CDC: collections performed by CDC light traps in trees; MA: hand-held mechanical aspirators (Clarke, Roselle, IL); BA: Collections performed by backpack aspirator (Model 1412; BioQuip, Rancho Dominguez, CA); LC: larvae collections using dippers; Insectary: insectary experiments performed to determine autogeny and stenogamy (Gomes *et al.*, 2009; Mutebi & Savage 2009) Au: autogenous; N-Au: non-autogenous; St: stenogamous; N-St: non-stenogamous; Ref: References; [1]: Gomes *et. al* 2009; [2]: Mutebi & Savage 2009. *: specimens provisionally identified by the CQ11FL marker.

AFLP genotyping

DNA extraction was performed using the DNeasy blood and tissue kit (Qiagen, Inc., Manchester, UK). The DNA concentration of each sample was fluorometrically quantified by the Quant-iT™ PicoGreen® dsDNA reagent and kit (Invitrogen™, Paisley, UK) as recommended by Wilding *et al.* (2009).

For each specimen, 100 ng of genomic DNA was used as template in the AFLP protocol described by Wilding *et al.* (2001), but without a dilution step between the ligation and the pre-selective PCRs. Primers used in the amplification are provided in Table S1 (see Supplementary Materials). Selective primers were labelled to allow separation of amplified products on a CEQ^{TM} 8000 capillary sequencer (Beckman Coulter Inc., CA, USA) using a DNA size standard kit – 600bp to quantify fragments between 50 and 700 base pairs. Peaks were only scored if they exceeded thresholds of both 3% of the maximum fluorescence peak height and 500 Relative Fluorescence Units of intensity. A raw matrix of the marker peak data was defined using a bin width of 1.0 bp.

The recommendation of Whitlock *et al.* (2008) was followed in order to determine which peaks from the raw matrix could be reliably scored. A two-step

approach using two relative thresholds in the fluorescence peak height (20% to select the markers and 15% to score the chosen markers) was performed by AFLPscore (Whitlock *et al.,* 2008) in order to score the peaks of the raw matrix obtained by the CEQ^{TM} 8000 Genetic analysis system.

AFLP analysis was repeated on a sub-set of samples for all the primer combinations (Table S2, Supplementary Materials) in order to assess the error of this approach by mismatch rates and Bayesian AFLPscore error analysis (proportion of mismatch; probability of mis-scoring allele 1 as allele 0, denoted *E1;* and probability of mis-scoring allele 0 as allele 1, denoted *E2*; Whitlock *et al.,* 2008).

Population genetic structure and genetic diversity

Bayesian cluster analysis as implemented by STRUCTURE 2.3.3 (Pritchard *et al.,* 2000) was used to infer population substructure/ancestry from the AFLP data set without prior information of sampling groups, under the conditions of admixture (α allowed to vary between 0 and 10) with allele frequencies correlated among populations (λ was set at the default value of 1). Ten independent runs, with 10^4 iterations during burn-in followed by 10^5 replications, were performed for each value of *K* (*K*=1 to 10 clusters for all samples). Information from the outputs of each *K* (10 runs) was compiled by the Greedy method implemented in CLUMPP (Jakobsson & Rosenberg, 2007).To infer the most likely number of clusters in the sample, we used two *ad hoc* approaches: an estimation of ln[Pr(X|*K*)], described in the original publication (Pritchard *et al.,* 2000) and the *ΔK* statistic (Evanno *et al.,* 2005).

Divergence among the sampled populations was assessed by an analysis of molecular variance (AMOVA, Excoffier *et al.,* 1992) using GENALEX 6.41 (Peakall & Smouse, 2006).

Principal Coordinates Analysis was used to visualise patterns of genetic differentiation among samples in two-dimensional plots. Calculations were performed in GENALEX 6.41 (Peakall & Smouse, 2006) using the standardised covariance method for the distance matrix conversion.

To test for significant genetic structure (F_{ST}) between collection sites, the data were randomly resampled (10,000 iterations) in AFLP-SURV (Vekemans *et al.,* 2002). To construct a bootstrapped, neighbor-joining tree, 10,000 replicates of pairwise F_{ST}

tables (based on all loci) were calculated in AFLP-SURV. These tables were used as input for PHYLIP 3.68 (Felsenstein, 2008), in which the programs NEIGHBOR and CONSENSE were used to produce the bootstrapped neighbor-joining tree. Figtree v.1.3.1 (Rambaut, 2009) was used to visualize the tree.

The proportion of polymorphic loci at the 5% level and expected heterozygosity (Lynch & Milligan, 1994) were estimated assuming Hardy-Weinberg equilibrium by AFLP-SURV (Vekemans *et al.,* 2002). Mann-Whitney tests were performed by SPSS statistic 19 (IBM®, NY, USA) to test for differences between pipiens and molestus forms in the genetic diversity estimates.

Detecting outlier loci

BAYESCAN 2.1 (Foll & Gaggiotti, 2008; Foll *et al.*, 2010) was used to compare neutral models with models including selection and to estimate Posterior Odds (PO) in support of selection over neutrality for each locus. BAYESCAN was applied to the binary code (*i.e.* allele presence/absence) typical for dominant markers. A second approach was implemented using the amplification intensity matrix which can enhance the information obtained from the AFLP markers and yield similar power to co-dominant markers (Fischer *et al.,* 2011). We employed $\log_{10}(PO)>1.5$, equivalent to 96.9% of PO, as the threshold for the rejection of the null hypothesis of neutrality. We conducted 20 pilot runs with a length of 5,000 iterations each and a burn-in of 50,000 iterations, as preceding tests indicated that this was sufficient to achieve convergence in the MCMC. Default values were used for sample size (5,000) and thinning interval (10). For the amplification intensity matrix we used 0.10 as threshold for the recessive genotype as a fraction of maximum band intensity.

The third approach for outlier detection used the DFDIST algorithm (Beaumont & Balding, 2004), as implemented in the software MCHEZA (Antao & Beaumont, 2011). The DFDIST method compares the empirical F_{ST} values to a null distribution derived from coalescent simulations and determines the probability that observed F_{ST} values are as large as, or larger than, the observation under neutrality. Runs were conducted under 'neutral mean F_{ST}', which involves computing the initial mean F_{ST} uninfluenced by outliers, with the following settings: 50,000 simulations; false discovery rate, theta, beta-a, beta-b at the default values of 0.1, 0.1, 0.25 and 0.25,

respectively. The significance threshold for outlier detection was set at ≥0.95 percentile of simulations.

Detection of outlier loci was conducted differently according to the geographic origin of samples. For European samples, outliers were first identified over all six samples and then within molestus and pipiens samples. Outliers identified among all populations but not among either of the within-form analyses were considered as candidate loci under divergent selection between pipiens and molestus. This indirect approach was not possible to apply in the USA samples since only one sample from each form was analysed. Therefore, outliers were identified from the direct comparison between pipiens and molestus samples. The direct approach between two population samples requires a cautious interpretation since outlier detection methods are known to be less robust with a small number of populations (Foll & Gaggiotti, 2008).

Results

Dominant markers and error rates

A total of 894 dominant markers were obtained from 12 primer combinations used in the selective amplification. The markers obtained by the primer combinations EcoRI-ACG/MseI-CGA (Mix1D3) and EcoRI-ACG/MseI-ACC (Mix3D3) presented high proportions of mismatches between replicates (12.50% and 19.58%, respectively) and were removed prior to subsequent analyses. The proportion of mismatch from the remaining 810 dominant markers varied between 0.00% and 1.02% (mean: 0.33%). Error rates for these 10 primer combinations averaged 1.41% and 0.04% for the probabilities of mis-scoring a peak as absent if present, and *vice versa*. Error rates for each primer combination are detailed in Table S2 (Supplementary Materials).

Population structure analysis

STRUCTURE analysis with the 810 loci indicated an optimum of two clusters (Figure S1, Supplementary Materials). Division of the 327 females into the two clusters closely matched the molestus and pipiens forms provisionally identified by CQ11FL. However, the assignment of individuals by STRUCTURE placed eight individuals from two molestus populations (Sandim and Comporta) into the cluster representing the

pipiens form (Figure 1A). Principal component analysis (PCA) confirmed the division between the two forms and the misidentification of the eight molestus females (Figure 2). These eight individuals were excluded from the subsequent analysis. Clustering analysis was also performed within each molecular form separately; each indicated a division into two clusters, which split the Chicago samples from European samples within both forms (Figure 1B and Figure S1, Supplementary Materials).

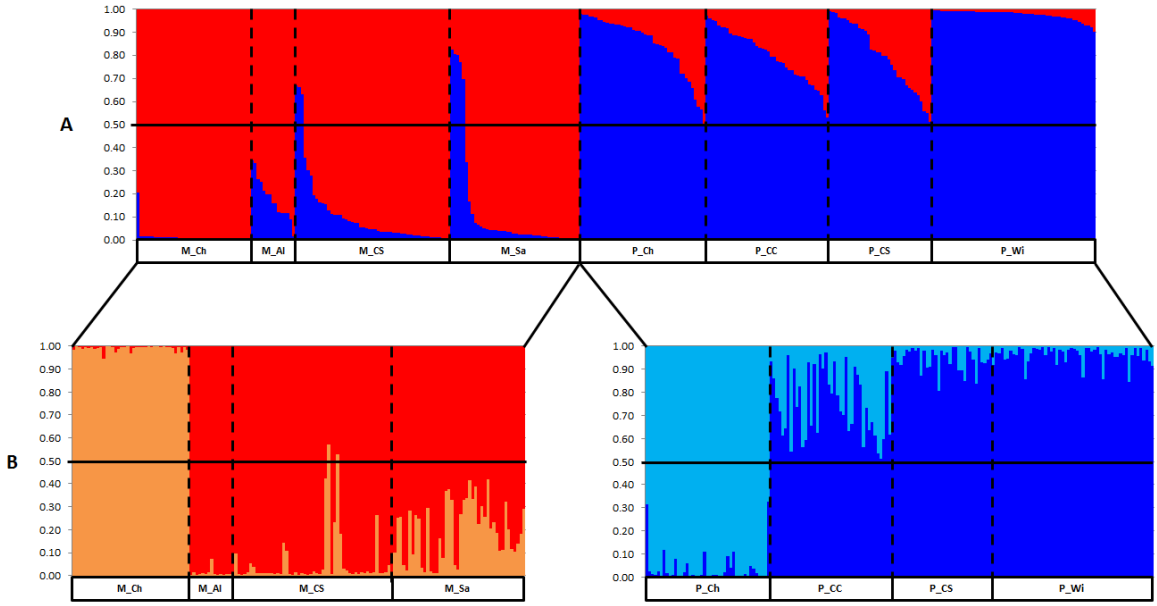

Figure 1. Bayesian cluster analysis conducted by STRUCTURE

A: analysis with the eight populations of *Cx. pipiens s.s.* B: analysis within the populations of each form. M_Ch: molestus from Chicago; M_Al: molestus from Alqueva; M_CS: molestus from Comporta, collected inside shelters; M_Sa: molestus from Sandim; P_Ch: pipiens from Chicago; P_CC: pipiens from Comporta, collected in trees by CDC light traps; P_CS: pipiens from Comporta, collected inside shelters; P_Wi: pipiens from Wirral. Columns correspond to the multilocus genotype of each individual, partitioned in different colours representing the probability of ancestry (q_i) to each cluster. Individuals were grouped according to their geographic location. Lines indicate the q_i threshold (0.50) used to assign individuals.

PCA supported the geographic (continental) division within molestus (Figure 2A) and pipiens (Figure 2B), with European samples of each form comprising a single group but the samples from Chicago (USA) separated from all the other samples. A neighbor-joining tree based on F_{ST} supported the division between the forms and also high differentiation between the European and American samples, especially in the molestus form (Figure 3).

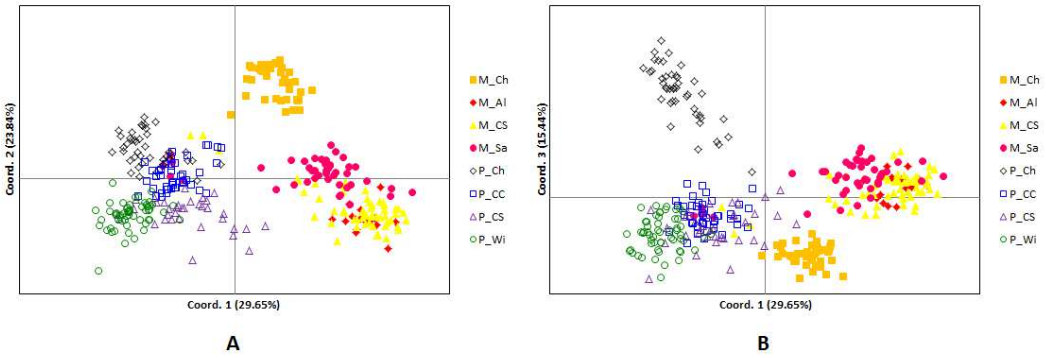

Figure 2. Principal Coordinates Analysis of the eight *Cx. pipiens s.s.* populations

A: two-dimensional plots with coordinates 1 and 2; B two-dimensional plots with coordinates 1 and 3; M_Ch: molestus from Chicago; M_Al: molestus from Alqueva; M_CS: molestus from Comporta, collected inside shelters; M_Sa: molestus from Sandim; P_Ch: pipiens from Chicago; P_CC: pipiens from Comporta, collected in trees by CDC light traps; P_CS: pipiens from Comporta, collected inside shelters; P_Wi: pipiens from Wirral. Coord: coordinate (percentage of variation explained by each coordinate).

AMOVA showed 17.6% molecular variance among populations, of which only 5.9% of the variation was distributed between the two forms. When the analysis was repeated with only European samples, the molecular variance between forms increased to 8.4%, whereas that among each form fell to 5.6%.

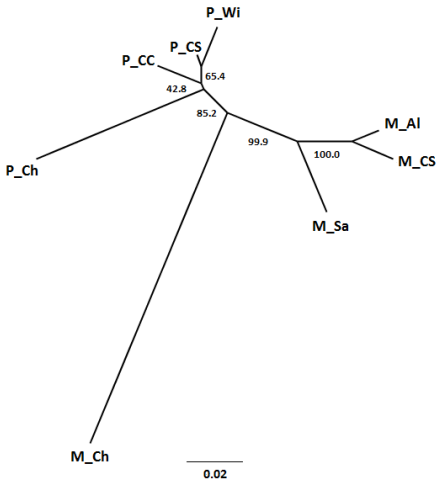

Figure 3. Unrooted Neighbor-joining tree, based on F_{ST} values obtained from 810 dominant loci. Bootstrap (%) support of each branch is given.

M_Ch: molestus from Chicago; M_Al: molestus from Alqueva; M_CS: molestus from Comporta, collected inside shelters; M_Sa: molestus from Sandim; P_Ch: pipiens from Chicago; P_CC: pipiens from Comporta, collected in trees by CDC light traps; P_CS: pipiens from Comporta, collected inside shelters; P_Wi: pipiens from Wirral.

The expected heterozygosity (H_e) and proportion of polymorphic loci at the 5% level (PLP) are shown in Table S3 (Supplementary Materials). The calculations were made for each population analysed and no significant differences were found between

pipiens and molestus forms (Mann-Whitney test; H_e: P = 0.248, PLP: P = 0.248; Table S3, Supplementary Materials).

Detecting loci under selection

Due to the geographic genetic structure, the detection of outlier loci was performed within samples of the same continent. Results of the outlier analysis, performed using three approaches among all the European populations (N = 6) and within each form in Europe (N = 3) are shown in Figure 4 and Figure S2 (Supplementary Materials).

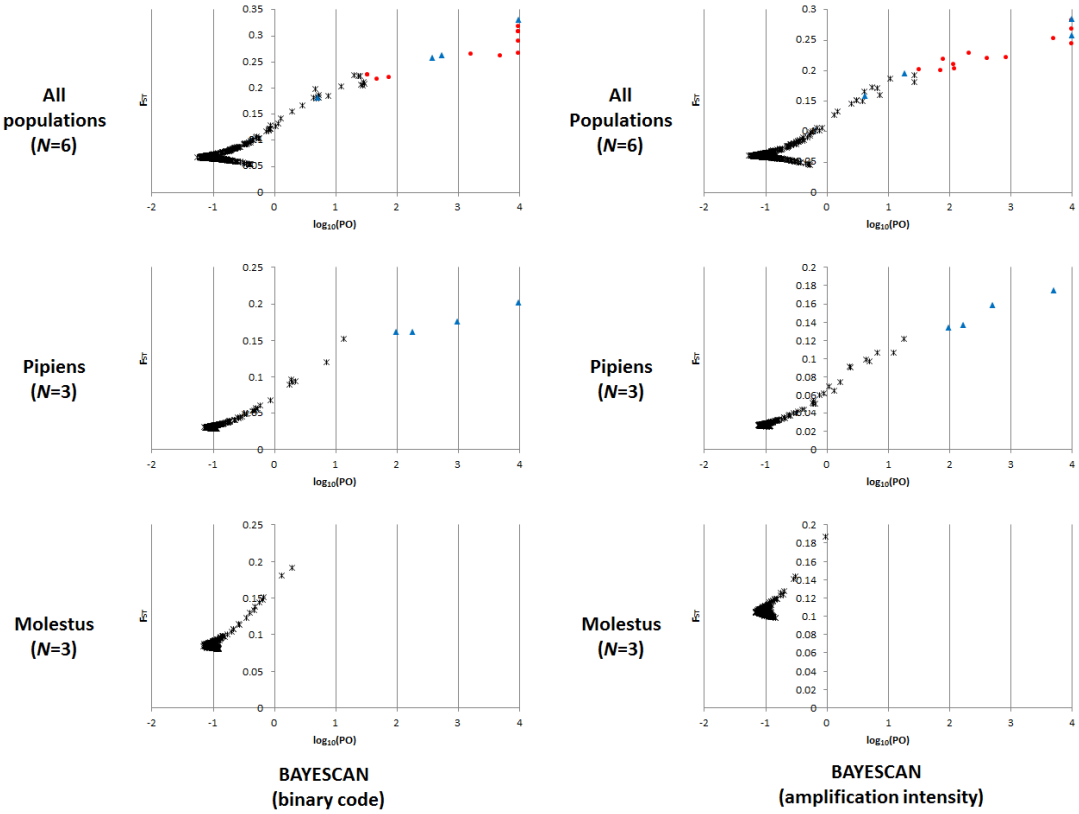

Figure 4. Outlier detection results from BAYESCAN analyses for European populations.

N: number of samples; Black loci: non-outlier loci ($\log_{10}(PO)<1.5$); Blue triangle: outlier loci within form analysis ($\log_{10}(PO)\geq1.5$); Red dot: candidate loci for divergent selection between pipiens and molestus ($\log_{10}(PO)\geq1.5$ only for all populations outlier analysis). Note that logarithm of Posterior Odds to base 10 ($\log_{10}(PO)$) is arbitrarily fixed to 4 when the posterior probability is 1 (should be infinity).

A total of 25 (3.1%) outlier loci were detected by the three methods. However, this number varied among the methods: BAYESCAN binary code analysis (N = 11; 1.4%), BAYESCAN amplification intensity analysis (N = 12; 1.5%), and MCHEZA (N= 20; 2.5%). Only six out of the 25 outlier loci were detected by all methods:

Mix1D2_022, Mix3D4_041, Mix4D2_004, Mix4D3_016, Mix4D4_011, and Mix4D4_037 (Figure 5).

MCHEZA detected 13 (1.6%) outlier loci between the molestus and pipiens samples from Chicago but the approaches implemented by BAYESCAN did not detect any outlier (Figure 5). Of the total 36 outlier loci found either in Europe or in USA, only two (0.25%), Mix3D4_041 (outlier by all methods in Europe) and Mix4D4_027 (outlier only by MCHEZA), were consistently detected in both continents (Table S4, Supplementary Materials).

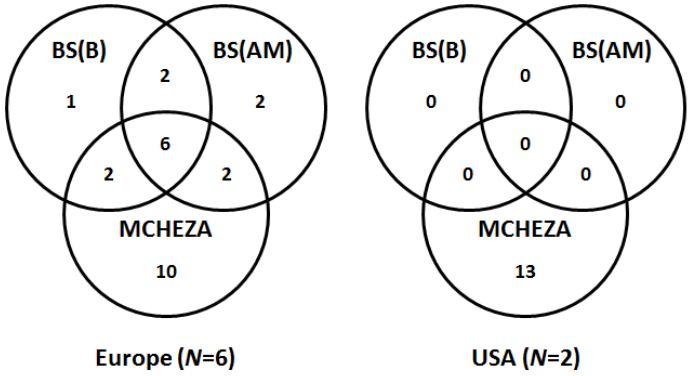

Figure 5. Number of loci detected as outliers in Europe and USA by each method and replicated as outliers in multiple methods.

BS(B): BAYESCAN with binary code; BS(AM) BAYESCAN with amplification intensity matrix; *N:* number of samples.

Discussion

The AFLP genome scan here conducted provided additional evidence for molestus and pipiens forms corresponding to evolutionarily distinct entities (Fonseca *et al.*, 2004). Independently of geographic origin, molestus samples clustered together and were genetically separated from the pipiens samples. This result was consistent in all analyses of population structure conducted. In addition to the molestus/pipiens partitioning, population sub-structuring was found between continents, within each form. However, inter-continental differentiation was higher within the molestus form, possibly due to two factors: 1) a recent underground habitat colonization of molestus in USA, possibly associated with bottlenecks. In this context, high differentiation was also observed between molestus natural populations of Chicago and New York (Kothera *et al.*, 2010); and 2) the laboratory colonization and maintenance for two years may also

have inflated differentiation from the natural populations, such as observed in other insect taxa (Kim *et al.*, 2007).

Outlier analyses have been used in order to estimate genomic divergence and its pattern across the genome among incipient species (Michel *et al.*, 2010; Weetman *et al.*, 2012). On average, estimates have varied between 5–10% of outlier loci found (range: 0.4– 24.5%; Michel *et al.*, 2010). In some studies, outlier loci were found to cluster in a low number of genomic regions (Turner *et al.*, 2005; Via & West, 2008). The genomic divergence between pipiens and molestus appears to be lower, 1.6% in USA to 3.1% in Europe, than in the majority of those studies. However, these estimates are comparable to those obtained for the M and S molecular forms of the malaria vector *Anopheles gambiae s.s.* (3.6%), in an area of high hybridisation in Guinea-Bissau (Oliveira *et al.*, 2008; Weetman *et al.*, 2012). The low proportion of outlier loci found in Guinea Bissau contrasted with estimates obtained in Ghana (12.6%) and Cameroon (16.6%), areas of low gene flow/hybridisation between M and S forms (Weetman *et al.*, 2012). Similarity between the outlier rates found in the present analysis in Europe (*Cx. pipiens*) and Guinea Bissau (*An. gambiae s.s.*) might be expected since both included the analysis of sympatric mosquito populations with elevated hybrid rates (at least for the samples of Comporta, in Portugal).

The low divergence (<3.0%) found between the USA molestus and pipiens samples was not expected since they were derived from allopatric populations (aboveground versus underground) and maintained isolated in laboratory colonies for at least one year. These results may be explained by two hypotheses: 1) molestus/pipiens habitat segregation in USA is not complete, which may allow genetic exchange in intermediate habitats at least during a short period of time; and 2) the molestus form has derived from the pipiens form through local adaptation to underground anthropogenic habitats. Previous reports of hybridisation between allopatric populations molestus and pipiens populations in the USA (Spielman, 2001; Fonseca *et al.*, 2004) and the common ancestry found among geographically distinct molestus populations (Fonseca *et al.*, 2004), including the ones in this genome scan, suggest the first hypothesis as the more plausible explanation for the lower divergence found between the two USA samples. Further analysis, with additional field-collected samples from USA and northern Europe

would be required to fully understand the genomic variation associated with the colonization of underground habitats by the molestus form.

The detection of divergent selection by outlier analysis-based genome scans is dependent upon the genetic architecture of the traits, the sampling density of the molecular markers and the effects of linkage disequilibrium between the genomic regions under selection and marker loci (Tice & Carlon, 2011). The *Cx. pipiens s.l.* genome is ~580Mb distributed over three chromosomes (Arensburger *et al.,* 2010). Assuming a uniform distribution in the genome of all loci, this suggests there is an average distance of 700Kb between the 810 loci used in this study. The lower number of outlier loci detected between the USA samples of molestus and pipiens when compared to European samples might be at least partly attributable to marker differences; with the USA samples exhibiting only approximately half (N = 450) as many positive bands as the European samples. In fact, for six of the within-Europe outliers (M1aD4_006, M1aD4_063, M2aD2_039, M4aD2_023, M4aD2_049, M4aD3_044) no positive band was detected in the USA samples, precluding its inclusion in the USA analysis. This difference reduces the coverage of the analysis between USA samples and increases the average distance among loci (1.4 Mb) thus reducing the odds for finding divergent selection. When accounting for this, the divergence rates obtained by MCHEZA between USA samples (N = 13; 2.9%) were actually higher than the rates between the European samples (N = 20; 2.5%). However, this contrasts with the BAYESCAN results where no loci were considered as outliers in USA samples. The underestimate of the outlier rates may lead to wrong conclusions, which should be avoided by adjusting the outlier rates to the number of loci used in each analysis to allow the comparison of the results. The 450 loci used in the USA analysis are in the range of loci used in previous AFLP genome scans (Buckley *et al.,* 2012; Mattersdorfer *et al.,* 2012) but the reduction of the number of loci can have consequences in the power of the analysis, which would require cautious when different data sets are compared.

Only two loci (0.25%), Mix3D4_41 and Mix4D4_027, were found with a consistent outlier signal in both Europe and USA. These loci are likely to be associated with genomic regions involved in ecological sympatric speciation and/or in the adaptation to anthropogenic habitats by the molestus form. The capacity of molestus of

occupying underground habitats associated with humans, such as subways, sewers and caves (Byrne & Nichols, 1999; Vinogradova, 2000), has been promoted by stenogamy and autogeny, which allow a continuous permanence in confined habitats with low availability of blood meal source. These traits were found to be kept even when the molestus form coexists with the pipiens form in aboveground habitats, such as in the case of Comporta, Portugal (Gomes *et al.*, 2009). Likewise, there was a tendency for molestus individuals to occupy aboveground indoor habitats in this region (Gomes *et al.*, unpublished observations). Genomic regions involved in these traits may be located at the vicinity of those outlier loci consistently detected in both European and USA samples.

This genetic study reinforces the status of the forms as distinct evolutionary entities of an on-going process of incipient speciation. The low genomic divergence between pipiens and molestus forms is comparable to other mosquito incipient species that coexist with genetic exchange (Weetman *et al.*, 2012). Most of the divergent loci varied between geographic regions, which suggest a low number of genomic regions underlying the major typological traits involved in habitat adaptation. Further studies focusing on natural molestus and pipiens populations, using wider genomic scans with high-throughput technologies (Stölting *et al.*, 2012) and identification of candidate genes and their function (Wood *et al.*, 2008; Nunes *et al.*, 2012) are required in order to fully understand the extent of the genomic regions under selection and the genes involved in this process of incipient speciation.

Acknowledgments

This study was funded by Fundação para a Ciência e a Tecnologia/, Portugal (POCI/BIA-BDE/57650/2004 and PPCDT/BIA-BDE/57650/2004). Bruno Gomes was funded by a PhD fellowship of Fundação para a Ciência e Tecnologia/MCTES (SFRH/BD/36410/2007).

References

Antao, T. & Beaumont, M. A. (2011) Mcheza: a workbench to detect selection using dominant markers. *Bioinformatics*. 27 (12), 1717–1718.

Arensburger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E., Camara, F., Campbell, C.L., Campbell, K.S., Casola, C., Castro, M.T., Chandramouliswaran, I., Chapman, S.B., Christley, S., Costas, J., Eisenstadt, E., Feschotte, C., Fraser-Liggett, C., Guigo, R., Haas, B., Hammond, M., Hansson, B.S. Hemingway, J., Hill, S.R., Howarth, C., Ignell, R., Kennedy, R.C., Kodira, C.D., Lobo, N.F., Mao, C., Mayhew, G., Michel, K., Mori, A., Liu, N., Naveira, H., Nene, V., Nguyen, N., Pearson, M.D., Pritham, E.J., Puiu, D., Qi, Y., Ranson, H., Ribeiro, J.M.C., Roberston, H.M., Severson, D.W., Shumway, M., Stanke, M., Strausberg, R.L., Sun, C., Sutton, G., Tu, Z.J., Tubio, J.M.C., Unger, M.F., Vanlandingham, D.L., Vilella, A.J., White, O., White, J.R., Wondji, C.S., Wortman, J., Zdobnov, E.M., Birren, B., Christensen, B.M., Collins, F.H., Cornel, A., Dimopoulos, G., Hannick, L.I., Higgs, S., Lanzaro, G.C., Lawson, D., Lee, N.H., Muskavitch, M.A.T., Raikhel, A.S. & Atkinson, P.W. (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science*. 330 (6000), 86–88.

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), 251–255.

Beaumont, M.A. & Balding, D.J. (2004) Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*. 13 (4), 969–980.

Buckley, J., Butlin, R.K. & Bridle, J.R. (2012) Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology*. 21 (2), 267–280.

Byrne, K. & Nichols, R.A. (1999) *Culex pipiens* in London Underground tunnels: differentiation between surface and subterranean populations. *Heredity*. 82 (1), 7–15.

Callot, J. & Van Ty, D. (1943) Sur quelques souches françaises de *Culex pipiens* L. *Bulletin de la Societe de pathologie exotique et de ses filiales*. 36 (7/8), 229–232.

Egan, S. P., Nosil, P., Funk, D. J. & Pfennig, D. (2008) Selection and genomic differentiation during ecological speciation: Isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution.* 62 (5), 1162–1181.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*. 14 (8), 2611–2620.

Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to Human mitochondrial DNA restriction data. *Genetics.* 131 (2), 479–491.

Excoffier, L., Hofer, T. & Foll, M. (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*. 103 (4), 285–298.

Felsenstein, J. (2008) *PHYLIP (phylogeny inference package).* [Online]. Available from: http://evolution.genetics.washington.edu/phylip/getme.html. [Accessed: 13 July 2012]

Fischer, M. C., Foll, M., Excoffier, L. & Heckel, G. (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*) *Molecular Ecology*. 20 (7), 1450–1462.

Foll, M. & Gaggiotti, O. (2008) A Genome-Scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics*. 180 (2), 977–993.

Foll, M., Fischer, M. C., Heckel, G. & Excoffier, L. (2010) Estimating population structure from AFLP amplification intensity. *Molecular Ecology*. 19 (21), 4638–4647.

Fonseca, D.M., Keyghobadi, N., Malcolm, C.A., Mehmet, C., Schaffner, F., Mogi, M., Fleischer, R.C. & Wilkerson, R.C. (2004) Emerging vectors in the *Culex pipiens* complex. *Science.* 303 (5663), 1535–1538.

Gomes, B., Sousa, C.A., Novo, M.T., Freitas, F.B., Alves, R., Corte-Real, A.R., Salgueiro, P., Donnelly, M., Almeida, A.P.G. & Pinto, J. (2009) Asymmetric introgression between sympatric molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in the Comporta region, Portugal. *BMC Evolutionary Biology*. 9 (1), 262.

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), 125–128.

Harbach, R.E., Harrison, B.A. & Gad, A.M. (1984) *Culex (Culex) molestus* Forskål (Diptera, Culicidae) - neotype designation, description, variation, and taxonomic status. *Proceedings of the Entomological Society of Washington*. 86 (3), 521–542.

Harbach, R.E., Dahl, C. & White, G.B. (1985) *Culex (Culex) pipiens* Linnaeus (Diptera, Culicidae) - Concepts, type designations, and description. *Proceedings of the Entomological Society of Washington*. 87 (1), 1–24.

Hopkins, R. & Rausher, M.D. (2011) Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature*. 469 (7330), 411–414.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 23 (14), 1801–1806.

Kim, K.S., French, B.W., Sumerford, D.V. & Sappington, T.W. (2007) Genetic diversity in laboratory colonies of western corn rootworm (Coleoptera: Chrysomelidae), including a nondiapause colony. *Environmental Entomology*. 36 (3), 637–645.

Kothera, L., Godsey, M., Mutebi, J.-P. & Savage, H.M. (2010) A comparison of aboveground and belowground populations of *Culex pipiens* (Diptera: Culicidae) mosquitoes in Chicago, Illinois, and New York City, New York, using microsatellites. *Journal of Medical Entomology*. 47(5), 805–813

Lynch, M. & Milligan, B.G. (1994) Analysis of population genetic-structure with RAPD markers. *Molecular Ecology.* 3 (2), 91–99.

Machado, C.A., Kliman, R.M., Markert, J.A. & Hey, J. (2002) Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*. 19 (4), 472–88.

Mattersdorfer, K., Koblmüller, S. & Sefc, K.M. (2012) AFLP genome scans suggest divergent selection on colour patterning in allopatric colour morphs of a cichlid fish. *Molecular Ecology*. 21 (14), 3531–3544.

Michel, A.P., Sim, S., Powell, T.H.Q., Taylor, M.S., Nosil, P. & Feder, J.L. (2010) Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences of the United States of America*. 107 (21), 9724–9729.

Mutebi, J.-P. & Savage, H.M. (2009) Discovery of *Culex pipiens pipiens* form molestus in Chicago. *Journal of the American Mosquito Control Association*. 25 (4), 500–503.

Nosil, P., Harmon, L. J. & Seehausen, O. (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*. 24 (3), 145–156

Nosil, P., Gompert, Z., Farkas, T.E., Comeault, A.A., Feder, J.L., Buerkle, C.A. & Parchman, T.L. (2012) Genomic consequences of multiple speciation processes in a stick insect. *Proceedings. Biological sciences / The Royal Society*. 279 (1749), 5058-5065.

Nunes, V.L., Beaumont, M.A., Butlin, R.K. & Paulo, O.S. (2012) Challenges and pitfalls in the characterization of anonymous outlier AFLP markers in non-model species: lessons from an ocellated lizard genome scan. *Heredity*. 109 (6), 340–348.

Oliveira, E., Salgueiro, P., Palsson, K., Vicente, J.L., Arez, A.P., Jaenson, T.G., Caccone, A. & Pinto, J. (2009) High levels of hybridization between molecular forms of *Anopheles gambiae* from Guinea Bissau. *Journal of Medical Entomology*. 45 (6), 1057–1063.

Pasteur, N., Rioux, J.A., Guilvard, E. & Pech-Perieres, J. (1977) Nouvelle mention pour le “Midi” méditerranéen, de populations naturelles anautogènes. *Annales de Parasitologie Humaine et Comparée*. 11 (3), 187–193.

Peakall, R. & Smouse, P. E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes.* 6 (1), 288–295.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*. 155 (5), 945–959.

Rambaut, A. (2009) *FigTree*. [Online]. Available from: http://tree.bio.ed.ac.uk/software/figtree/. [Accessed: 13 July 2012]

Solomon, T. (2004) Flavivirus Encephalitis. *The New England journal of medicine.* 351 (4), 370–378.

Spielman, A. (2001) Structure and seasonality of nearctic *Culex pipiens* populations. *Annals of the New York Academy of Sciences*. 951, 220–234.

Stölting, K.N., Nipper, R., Lindtke, D., Caseys, C., Waeber, S., Castiglione, S. & Lexer, C. (2012) Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology*. doi:10.1111/mec.12011

Tice, K.A. & Carlon, D.B. (2011) Can AFLP genome scans detect small islands of differentiation? The case of shell sculpture variation in the periwinkle *Echinolittorina hawaiiensis*. *Journal of Evolutionary Biology.* 24 (8), 1814–1825.

Turner, T.L., Hahn, M.W. & Nuzhdin, S.V. (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology.* 3 (9), e285.

Vekemans, X., Beauwens, T., Lemaire, M. & Roldán-Ruiz, I. (2002) Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology.* 11 (1), 139–51.

Via, S & West, J. (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, 17 (19), 4334–4345.

Vinogradova, E.B. (2000) *Culex pipiens pipiens mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control*. S.I. Golovatch (ed.). Sofia, Pensoft.

Weetman, D., Wilding, C.S., Steen, K., Pinto, J. & Donnelly, M.J. (2012) Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Molecular Biology and Evolution*. 29 (1), 279–291.

Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R.K. & Burke, T. (2008) An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology resources*. 8 (4), 725–735.

Wilding, C.S., Butlin, R.K. & Grahame, J. (2001) Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology.* 14 (4), 611–619.

Wilding, C.S., Weetman, D., Steen, K., & Donnelly, M.J. (2009) Accurate determination of DNA yield from individual mosquitoes for population genomic applications. *Insect Science*. 16 (4), 361–363.

Wood, H.M., Grahame, J.W., Humphray, S., Rogers, J. & Butlin, R.K. (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. *Molecular Ecology*. 17 (13), 3123–3135.

Wu, C.-I. (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*. 14 (6), 851–865.

Supplementary Materials

Table S1. Primers used in the AFLP protocol

Primer	Sequence	Extra Base	fluorescence	Stage
EcoRI	GACTGCGTACCAATTC	A		Pre-Selective PCRs
		C		
		ACG	D3 (HEX)	Selective PCRs
		AGT	D2 (NED)	
		CAG	D2 (NED)	
		CTC	D2 (NED)	
		CTC	D4 (6-fam)	
MseI	GATGAGTCCTGAGTAA	A		Pre-Selective PCRs
		C		
		ACC		Selective PCRs
		AGT		
		ATC		
		CAA		
		CCA		
		CGA		
		CTC		
		AGTC		
		CGAG		

Labels D2-D4 are Beckman-Coulter WellRED dyes, with their Applied Biosystem equivalents shown in parentheses

Table S2. Error rates of the loci obtained by each primer combination in the selective amplification

Group	Fluorescence	Primer EcoRI	Primer MseI	SR	Markers	E1 (%)	E2 (%)	M (%)
Mix1	D2 (NED)	EcoRI-CTC	MseI-CTC	16	69	3.02	0.03	0.72
	D3 (HEX)	EcoRI-ACG	MseI-CGA	16	40	4.33	0.06	12.50
	D4 (6-fam)	EcoRI-CTC	MseI-CGA	16	80	1.30	0.06	0.39
Mix2	D2 (NED)	EcoRI-AGT	MseI-CGAG	14	46	0.46	0.03	0.00
	D3 (HEX)	EcoRI-ACG	MseI-ATC	14	59	0.67	0.03	0.12
	D4 (6-fam)	EcoRI-CTC	MseI-CAA	14	108	0.34	0.02	0.07
Mix3	D2 (NED)	EcoRI-CTC	MseI-CCA	19	100	1.08	0.03	0.26
	D3 (HEX)	EcoRI-ACG	MseI-ACC	19	44	0.67	0.03	19.58
	D4 (6-fam)	EcoRI-CTC	MseI-ATC	19	85	0.95	0.04	0.25
Mix4	D2 (NED)	EcoRI-CAG	MseI-CGA	17	81	1.88	0.03	0.44
	D3 (HEX)	EcoRI-ACG	MseI-AGTC	17	46	4.16	0.08	1.02
	D4 (6-fam)	EcoRI-CTC	MseI-AGT	17	136	0.26	0.01	0.04

SR: samples repeated to infer the error analysis; E1: probability of miss-scoring allele 1 as allele 0; E2: probability of miss-scoring allele 0 as allele 1; M: mismatch. Labels D2-D4 are Beckman-Coulter WellRED dyes, with their Applied Biosystem equivalents shown in parentheses

Table S3. Population diversity of the eight populations used in the study

Population	*N*	Form	#loc	#loc_P	PLP	*He*	S.E.(*He*)
M_Ch	39	M	810	129	15.9	0.053	0.004
M_Al	15	M	810	268	33.1	0.113	0.005
M_CS	49 (50)	M	810	253	31.2	0.093	0.005
M_Sa	39	M	810	179	22.1	0.074	0.005
P_Ch	43	P	810	200	24.7	0.080	0.005
P_CC	42	P	810	249	30.7	0.091	0.004
P_CS	34 (35)	P	810	320	39.5	0.116	0.004
P_Wi	55 (56)	P	810	356	44	0.120	0.004

N: number of individuals without missing data (with missing data); #Loc: number of loci; #Loc_P: number of loci with positive bands; PLP: proportion of polymorphic loci at the 5% level; He: expected heterozygosity; S.E.(*He*): standard error of *He*; M: molestus form; P: pipiens form; M_Ch: molestus from Chicago; M_Al: molestus from Alqueva; M_CS: molestus from Comporta, collected inside shelters; M_Sa: molestus from Sandim; P_Ch: pipiens from Chicago; P_CC: pipiens from Comporta, collected in trees by CDC light traps; P_CS: pipiens from Comporta, collected inside shelters; P_Wi: pipiens from Wirral.

Table S4. Loci detected as outliers in each comparative analysis (Europe and USA).

Loci	Europe	USA
Mix3D4_041	X	X
Mix4D4_027	X	X
Mix1D2_011	X	
Mix1D2_021		X
Mix1D2_022	X	
Mix1D2_024	X	
Mix1D4_006	X	
Mix1D4_007	X	
Mix1D4_009		X
Mix1D4_024		X
Mix1D4_054		X
Mix1D4_063	X	
Mix2D2_039	X	
Mix2D3_001	X	
Mix2D4_012	X	
Mix2D4_026	X	
Mix2D4_042	X	
Mix2D4_059	X	
Mix2D4_062	X	
Mix2D4_076	X	
Mix3D2_006		X
Mix3D4_007		X
Mix3D4_017		X
Mix3D4_026	X	
Mix4D2_002	X	
Mix4D2_004	X	
Mix4D2_023	X	
Mix4D2_025		X
Mix4D2_049	X	
Mix4D3_011		X
Mix4D3_016	X	
Mix4D3_044	X	
Mix4D4_011	X	
Mix4D4_026		X
Mix4D4_037	X	
Mix4D4_063		X

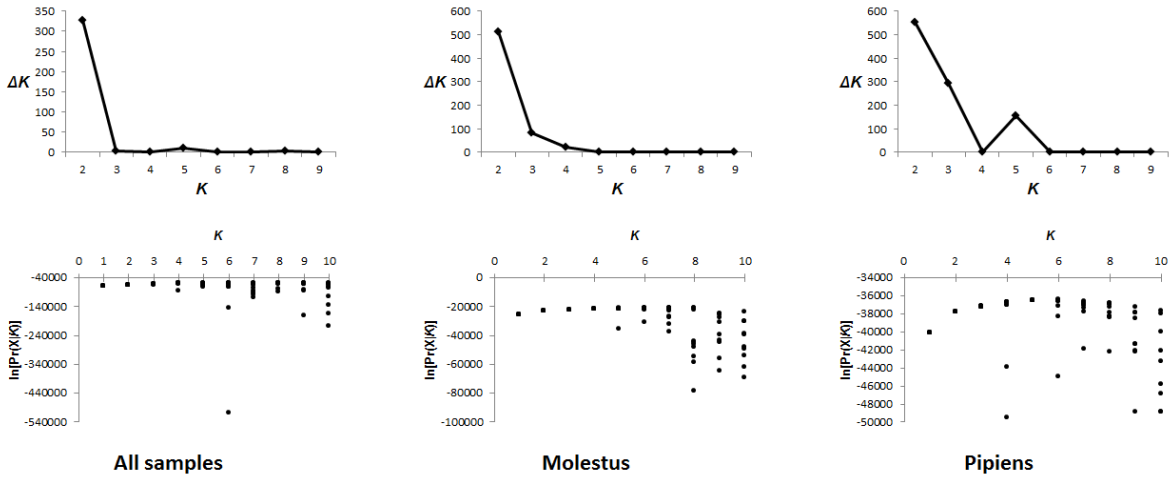

Figure S1. Graphics of *ad hoc* approaches to infer the number of clusters (*K*) in STRUCTURE analysis with all samples

K: number of clusters; Δ*K*: see Evanno *et al.*, (2005); ln[Pr(X|*K*)]: estimated log probability of the data under each *K*.

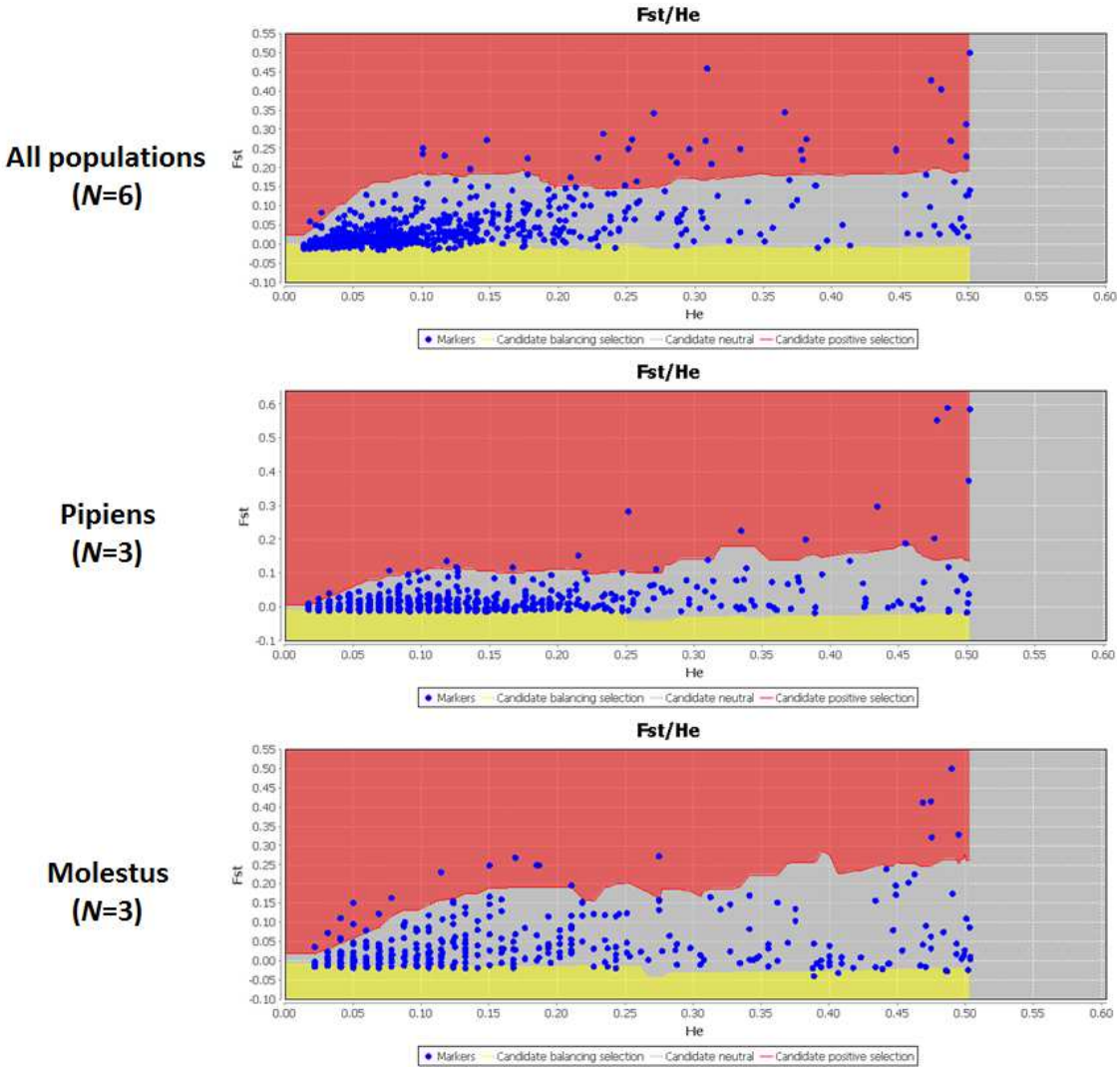

Figure S2. Outlier detection results from MCHEZA analyses

N: number of samples; Plots show F_{ST} values, conditional on heterozygosity, of the 810 AFLP loci studied. Blue dot: locus; Yellow area: candidate for balancing selection; Red area: Candidate for positive selection (Outliers); Grey area: candidate for neutrality.

Chapter 6.

Distribution and hybridisation of *Culex pipiens* forms in Greece during the West Nile virus outbreak of 2010

Submitted to *Infection Genetics and Evolution* as:

Gomes, B., Kioulos, E., Papa, A., Almeida, A.P.G., Vontas, J. & Pinto J. Distribution and hybridization of *Culex pipiens* forms in Greece during the West Nile virus outbreak of 2010.

Abstract

In 2010, an outbreak of West Nile virus (WNV) infections occurred in the region of Thessaloniki, Central Macedonia, in northern Greece. During this period, *Culex pipiens sensu stricto* mosquitoes were found infected by WNV lineage 2. *Culex pipiens s.s.* presents two distinct biological forms, denoted molestus and pipiens and hybrids between the two forms may potentiate the accidental transmission of WNV to humans. We have genetically characterized the form composition of *Cx. pipiens s.s.* samples collected during the outbreak from the region of Thessaloniki, where WNV cases occurred, and from the region Schinias-Marathonas, with no reported cases at the time. Information on bird fauna was also obtained for the two regions. Application of the CQ11FL diagnostic marker revealed a 350 bp variant of the pipiens-specific allele. Sympatric pipiens and molestus populations were detected in Thessaloniki, whereas Schinias-Marathonas presented a more genetically homogenous molestus population. A pattern of asymmetric introgression between molestus and pipiens was also observed in Thessaloniki. The presence of hybrids between molestus and pipiens forms suggests a greater receptivity of the Thessaloniki region for the establishment of WNV zoonotic cycles. However, the Schinias-Marathonas region also displayed characteristics to sustain WNV transmission cycles. These observations highlight the importance of maintaining active surveillance systems in selected regions geographically located within the range of major migratory bird flyways.

Introduction

Culex pipiens s.s. is considered a main vector of West Nile virus (WNV) in Europe (Hubálek, 2008). This mosquito species comprises two distinct forms, denoted pipiens and molestus, which are morphologically indistinguishable but exhibit important behavioural and physiological differences. The molestus form is stenogamous (mates in confined spaces, *i.e.* < 0.1 m^3; Clements, 1999), autogenous (can oviposit without a blood meal), homodynamic (remains active during winter) and mammophilic (prefers to feed on mammals, including humans). In contrast, the pipiens form is eurygamous (mates in open spaces), anautogenous (oviposition requires a blood meal),

heterodynamic (undergoes winter diapause) and ornithophilic (feeds predominantly on birds) (Harbach *et al.*, 1984, 1985; Vinogradova, 2000).

While in northern latitudes molestus and pipiens forms are physically separated by occupying underground and surface habitats, respectively, in southern European regions, populations of both forms have been found at the surface (Chevillon *et al.*, 1995; Vinogradova, 2000; Gomes *et al.*, 2009). This sympatric occurrence in aboveground habitats promotes hybridisation between forms (Fonseca *et al.*, 2004; Gomes *et al.*, 2009). Populations with intermediate behaviour between the two forms have been described in southern Europe (Callot & Van Ty, 1943; Pasteur *et al.*, 1977; Gomes *et al.*, 2009). Hybrids are considered of great epidemiological importance as they may display a more opportunistic biting behaviour. This behaviour may potentiate the role of *Cx. pipiens s.s.* as a bridge-vector for the transmission of WNV, from their avian amplification hosts to humans (Fonseca *et al.*, 2004; Hamer *et al.*, 2008).

Bird migrations have been associated with the spread of WNV. High infection rates in migratory birds have been described and this has been considered a possible cause for virus introduction in Europe and in North America (Rappole & Hubálek, 2003; Hubálek, 2004; Figuerola *et al.*, 2008). Bird migrations normally follow a north-south axis, linking breeding regions (arctic and temperate) with non-breeding regions (temperate and tropical). Eight well-established migration routes (flyways) have been identified (Si *et al.*, 2009). Of these, the Mediterranean/Black sea flyway is the largest bird migration system in the world, linking a vast area from Africa to west Siberia. The Bosphorus strait is the main entrance for African bird populations in Europe and it is the major migratory bottleneck of this flyway (Birdlife International, 2012). After the passage of the strait, migratory birds find their first European refuge in Greece and Bulgaria, where they rest and breed. In these locations, migratory birds may be bitten by local *Cx. pipiens s.s.* mosquitoes, or other WNV vectors (*e.g. Culex modestus*), which may lead to the establishment of local WNV transmission cycles.

In 2010, northern Greece experienced one of the largest WNV outbreaks described in Europe, with 262 human cases of WNV infection: 197 with neuroinvasive disease (encephalitis, meningitis, or acute flaccid paralysis) and 65 with West Nile fever; 35 (13.4%) cases were fatal (Danis *et al.*, 2011). The outbreak was restricted to

the north of the country and most human cases were observed in Central Macedonia, in wetland areas located between four major rivers, west to the city of Thessaloniki (Valiakos et al. 2011). Molecular analyses identified the WNV lineage 2 strain in birds, sentinel chickens and *Cx. pipiens s.s.* mosquitoes in this region (Chaskopoulou *et al.*, 2011; Papa *et al.*, 2011; Valiakos *et al.*, 2011). However, WNV epidemiologic studies conducted during this period have treated *Cx. pipiens s.s.* as a single entity, without determining the relative composition of molestus and pipiens forms and their relative impact in WNV transmission. Furthermore, it is still not fully understood why the outbreak was largely confined to the region of Central Macedonia in northern Greece.

In this study, Bayesian model-based clustering methods were applied to multi-locus microsatellite genotypes to infer the genetic structure of *Cx. pipiens s.s.* in Thessaloniki during the 2010 WNV outbreak and also in Schinias-Marathonas, southern Greece, a region without WNV transmission. Information on the wild avian fauna was also collected for each region, with particular attention to trans-Saharan migratory birds and species for which WNV infection has been reported in previous studies in Europe and North America. Our objectives were: i) to assess differences in the molestus/pipiens form composition of *Cx. pipiens s.s.*, as well as in hybrid frequency, between northern and southern Greece: ii) to compare the distribution of migratory bird species and/or species with previous records of WNV infection between the two regions; iii) to determine if differences in both *Cx. pipiens s.s.* form composition and avian fauna could be consistent with a higher receptivity of northern Greece (Thessaloniki) for the establishment of a WNV transmission cycle when compared to the region of Schinias-Marathonas in southern Greece.

Material and Methods

Study regions and mosquito collection

Mosquito collections in the region of Thessaloniki (northern Greece) were performed between 20th August and 15th September 2010 by CDC light traps baited with CO_2 (Sudia & Chamberlain, 1962). Traps were hung outdoors at *ca.* 1.5 m height and approximately 20 m away from human dwellings. Sampling was carried out in the villages of Chalastra, Anatoliko, Kimina, Malgara, Adendro, Brachia, Vathilakos,

Eleousa and Nea Xalkidona, located *ca.* 10-15 km west of Thessaloniki city (Figure 1). These villages lie in the region where most human cases of WNV were reported during the outbreak and also where both mosquitoes and avian hosts were found infected (Papa et al. 2011; Valiakos et al. 2011). This region has a warm temperate climate with hot dry summers and mild winters (class Csa, Köppen Classification System; Peel *et al.*, 2007). The villages are situated close to the delta of rivers Axios and Aliakmonas. Irrigation channels derived from these rivers feed *ca.* 20,000 hectares of rice fields, the main crop in the area, and provide suitable breeding sites for mosquito larvae. Additional breeding sites such as open sewages and cesspits are found in or close to the villages. A population of *ca.* 70 wild horses is present in an isolated part of the river delta. Cattle, sheep and domestic birds are common in most villages of the region and domestic horses are found in several horse-riding schools.

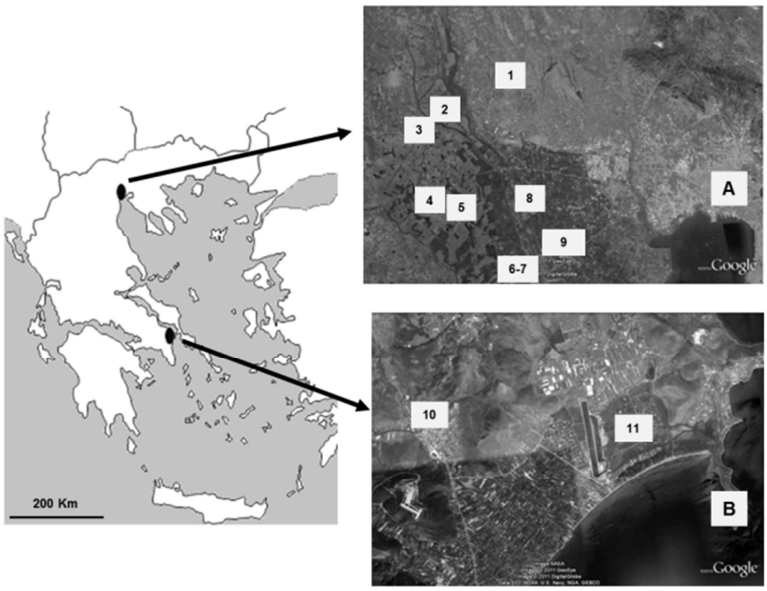

Figure 1. Map of Greece showing collection sites and samples sizes.

A: Thessaloniki region (Total *N*=37); B: Schinias-Marathonas region (Total *N*=40); 1: Vathilakos (*N*=2); 2: Eleousa (*N*=5); 3: Nea Xalkidona (*N*=5); 4: Adendro (*N*=5); 5: Brachia (*N*=4); 6: Malgara (*N*=4); 7: Kimina (*N*=5); 8: Anatoliko (*N*=3); 9: Chalastra (*N*=4); 10: Marathonas Village (*N*=40); 11: Schinias-Marathonas wetland (*N*=20).

Mosquito collections in Schinias-Marathonas region (southern Greece) were performed in the same time period as in Thessaloniki. Two collection methods were used. CDC light traps baited with CO_2 were placed outdoors at an average height of 1.5m and approximately 20 m away from human dwellings in the wetland areas of

Marathonas-Schinias. Collections of resting mosquitoes were performed with mouth aspirators inside and around houses at the village of Marathonas (Figure 1). The region has the same warm temperate climate with hot dry summers and mild winters as in Thessaloniki (class Csa; Peel *et al.,* 2007). A large number of annual vegetable crops and greenhouses are present in the area. The extensive wetland of Schinias-Marathonas, as well as neighboring streams and cesspits in villages constitute the major mosquito breeding sites of the area. No cases of WNV infection during 2010 were reported in the Schinias-Marathonas area, as well as in the whole district of Attiki, to which the area belongs. Sheep, chickens and a few horses are the main domestic animals found in this region.

Migratory birds and WNV avian hosts

Information on bird species diversity for each study region was obtained from the OrnithoTopos database, hosted at the Hellenic Ornithological Society website (OrnithoTopos, 2012). This database displays bird species observation records from birdwatchers, organized by geo-referenced locations. For each species recorded, the maximum number of individuals counted at a single observation is also given. Information from location summary reports, between January 2008 and December 2010, was collected from three locations in Thessaloniki (Axios, Loudias and Aliakmon estuaries) and two locations in Schinias-Marathonas (Schinias and Schinias marsh). Species without information on maximum bird count (*i.e.* corresponding to non-visual records) were excluded from the analysis. Recorded species from each region were classified as trans-Saharan migratory birds according to the list of Walther (2005). Information about records of WNV infection in bird populations of Europe and North America was also obtained for each species based on published reports (Figuerola *et al.,* 2007, 2008, 2009; Formosinho *et al.,* 2006; Hubálek, 2004, 2008; Jourdain *et al.,* 2008; Malkinson *et al.,* 2002; Rappole & Hubálek, 2003; Valiakos *et al.,* 2012).

Molecular analysis

DNA extraction from individual females was performed using the DNAzol method (Invitrogen). Each specimen was identified to species by multiplex PCR assay targeting species-specific polymorphisms at the second intron of the *ace*-2 gene using

primers specific for *Cx. pipiens s.s.*, *Culex quinquefasciatus* and *Culex torrentium* (Smith & Fonseca, 2004).

Fourteen microsatellites (Fonseca *et al.,* 1998; Keyghobadi *et al.,* 2004; Smith *et al.,* 2005) were analysed following the procedures described in Gomes et al. (2009). Amplified products were separated by capillary electrophoresis in a genetic analyser ABI3730 (Applied Biosystems) at Yale DNA Analysis Facility (USA). Fragment sizes were scored using the software GeneMarker 1.4. (Softgenetics, USA).

Mosquitoes were also genotyped for the molestus/pipiens diagnostic marker described by Bahnck & Fonseca (2006). This marker detects a size polymorphism in the 5' flanking region of the CQ11 microsatellite of *Cx. pipiens s.s.* that differentiates specimens of the pipiens form (200 bp PCR product) from the molestus form (250 bp PCR product). Hybrids exhibit both amplicons. Individuals displaying doubtful genotypes were clarified by sequencing of the microsatellite CQ11 (that contains the CQ11FL polymorphism), using primers CQ11R3 and CQ11F2 and conditions described by Smith et al. (2005). The PCR products were purified with the QIAquick PCR Purification kit (Qiagen) and sequenced (forward and reverse) using the same primers. Sequencing reaction products were electrophoresed on an ABI3730XL automated sequencer (Applied Biosystems). Sequences were aligned using BioEdit 7.0.9.0 (Hall, 1999) and identified by comparison with CQ11 sequences deposited in GenBank.

Data analysis

Chi-square tests on contingency tables available in VassarStats (Lowry, 2012), were performed to assess differences between regions on the proportion of trans-Saharan migratory bird species and species reported as WNV carriers. As a measure of diversity, Shannon's *H'* index (Shannon, 1948) was calculated using the records of maximum bird count for each species as an approximation of abundance. Differences in *H'* between regions were determined by Student's *t*-tests, according to Jayaraman (1999).

Bayesian clustering analysis as implemented by STRUCTURE 2.3.3 (Pritchard *et al.*, 2000) was used to infer population substructure/ancestry from the data set without prior information of sampling groups under the conditions of admixture (α allowed to vary between 0 and 10), and allele frequencies correlated among populations (λ was set

at 1, default value). Ten independent runs with 10^4 iterations and 10^5 replications were performed for each value of K (K=1 to 10 clusters). Information from the outputs of each K (10 runs) was compiled by the Greedy method implemented in CLUMPP (Jakobsson & Rosenberg, 2007). To infer the most likely number of clusters in the sample, we used a combined approach with an estimation of ln[Pr(X|K)], described in the original publication (Pritchard *et al.,* 2000) and the ΔK statistic (Evanno *et al.,* 2005). Following the suggestions of Vähä & Primmer (2006), individual genetic assignment to clusters was based on a minimum posterior probability threshold (Tq) of 0.90. Individuals displaying $0.1 \leq q_i \leq 0.90$ were considered of admixed ancestry.

Genetic diversity at each microsatellite locus was characterised by estimates of expected heterozygosity (Nei, 1987) and inbreeding coefficient (F_{IS}). Significance of F_{IS} values was assessed by randomisation tests. These analyses were performed using FSTAT v. 2.9.3.2. (Goudet, 1995). Estimates of allele richness (A_R), adjusted for the lowest sample size, were obtained by a rarefaction statistical approach implemented by the programme HP-RARE (Kalinowski, 2005). Departures from Hardy–Weinberg proportions were tested by exact tests available in ARLEQUIN v.3.5 (Excoffier *et al.,* 2005). The same software was used to perform exact tests of linkage equilibrium between pairs of loci based on the expectation-maximisation approach described by Slatkin & Excoffier (1996). The software Micro-Checker 2.2.3. was used to search (99% confidence interval) for null alleles at loci/samples (van Oosterhout *et al.,* 2004).

The Bayesian method implemented by NEWHYBRIDS 1.1. (Anderson & Thompson, 2002) was used to assign individuals into six classes: Two pure and four hybrid (F1, F2 and backcrosses with the two pure classes). The approach of uniform priors was used because it down-weighs the influence of low frequency alleles, thus preventing sampling and genotyping errors in closely related populations. Results were based on the average of five independent runs of 10^5 iterations. Following the suggestions of Anderson & Thompson (2002), individual genetic assignment to classes was based on a minimum posterior probability threshold (Tq) of 0.50.

Whenever multiple testing was performed, the nominal significance level of rejection of the null hypothesis (α=0.05) was corrected by the sequential Bonferroni procedure (Holm, 1979).

Results

Bird species

A list of the bird species retrieved from OrnithoTopos database can be found in Table S1 of the Supplementary Materials. The list was compiled based on 144 data records available for the region of Thessaloniki and 157 for the region of Schinias-Marathonas. Number of bird species and diversity indexes obtained by region are shown in Table 1. Overall, there were 172 species recorded for Thessaloniki and 167 species in Schinias-Marathonas. The proportion of trans-Saharan migratory bird species was approximately 62% in both regions (χ^2= 0.02; d.f.=1; P=0.888) and the proportion of species with records of WNV infection was also similar, around 22% in both regions (χ^2= 0.01; d.f.=1; P=0.920). However, there were significant differences between regions in the Shannon's *H'* index estimates. Overall bird diversity was higher in Thessaloniki (Student's *t*-test: P<0.001) and so was the diversity with the group of WNV carriers (Student's *t*-test: P<0.001). Conversely, Schinias-Marathonas displayed a higher diversity within the group of species with a trans-Saharan migration route (Student's *t*-test: P<0.001).

Table 1. Number of bird species and diversity indexes recorded in the study regions.

		Total	T-S	WNV
Thessaloniki	*N*	172	108 (62.8%)	40 (22.3%)
	H´	2.817	2.091	2.458
Schinias-Marathonas	*N*	167	103 (61.7%)	38 (22.8%)
	H´	1.772	3.249	0.763

Total: total number of species recorded; T-S: number of trans-Saharan species; WNV: number of species with record of WNV infection; *N*: number of species (in brackets: proportion relative to the total number of species recorded); *H'*: Shannon's index.

Molecular identification of Cx. pipiens s.s. forms

All 77 individuals analysed in this study were identified as *Cx. pipiens s.s.* by the *ace*-2 marker of Smith & Fonseca (2004).

Bayesian clustering analysis implemented by STRUCTURE revealed two clusters, as determined by both ln[Pr(X|*K*)] (Pritchard *et al.,* 2000) and the Δ*K* statistic (Evanno *et al.*, 2005) (Figure 2A; Supplementary Figure S1). Cluster-1 grouped 29 specimens from Thessaloniki, while most of the specimens from Schinias-Marathonas (*N*=39) were grouped in cluster-2 that also included four individuals from Thessaloniki. Five females (four from Thessaloniki and one from Schinias-Marathonas) of the total sample (*N*=77) exhibited admixed ancestry (*i.e.* $0.1 \leq q_i \leq 0.90$; Figure 2A).

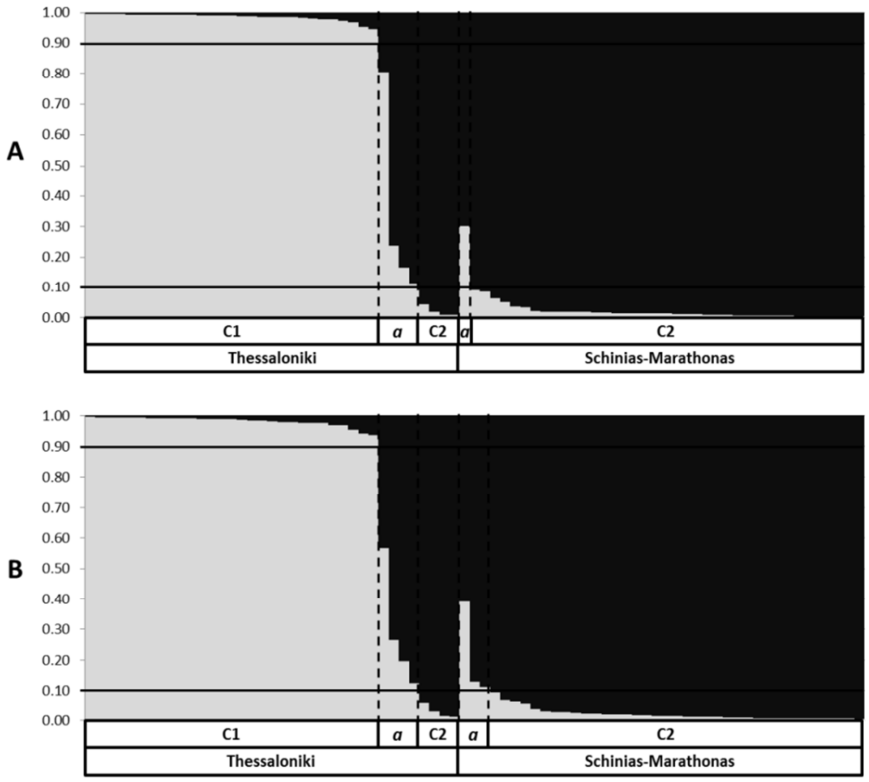

Figure 2. Bayesian clustering analysis conducted by STRUCTURE.

A: Individuals sorted by region and ancestry probability (all loci). B: Individuals sorted by region and ancestry probability (excluding locus CQ41). C1: Cluster 1; *a*: admixed individuals ($0.1<Tq<0.9$); C2: Cluster 2. Columns correspond to the multilocus genotype of each individual, partitioned in different colours representing the probability of ancestry (q_i) to each cluster. Lines indicate the q_i threshold used to determine admixed individuals (see Methods).

Estimates of genetic diversity and tests for Hardy-Weinberg equilibrium for the 14 microsatellite loci in the whole sample (*N* = 77) and in subsamples determined by clustering analysis (STRUCTURE) and geographic location are shown in Supplementary Table S2. Locus CQ41 exhibited heterozygote deficits in the cluster-1 sample of Thessaloniki and cluster-2 sample of Schinias-Marathonas (as well as in the

total sample), possibly reflecting locus-specific effects. The analysis performed by Micro-Checker suggests the possibility of null alleles at locus CQ41. The exclusion of locus CQ41 from Bayesian clustering conducted in STRUCTURE shifted the assignment of two Schinias-Marathonas specimens from cluster-2 to admixed (Figure 2B).

The results of the molestus/pipiens diagnostic CQ11FL marker are shown in Table 1. The combination of STRUCTURE and CQ11FL analyses suggest that each Bayesian cluster represents a different form of *Cx. pipiens s.s.* A higher proportion of the pipiens CQ11FL genotype (79.3%) was observed in cluster 1 while the majority of individuals in cluster 2 (78.0%) had the molestus genotype (Table 2).

Table 2. Association between ancestry clusters revealed by STRUCTURE with molestus/pipiens genotypes at the CQ11FL marker and NEWHYBRIDS pedigree classes.

		N	**CQFL11**			**NEWHYBRIDS**		
			M	**H**	**P**	**M**	**H**	**P**
STRUCTURE	Cluster-1	29	2 (6.9)	4 (13.8)	23 (79.3)	0 (0.0)	1 (3.4)	28 (96.6)
	Admixed	7	6 (85.7)	1 (14.3)	0 (0.0)	7 (100.0)	0 (0.0)	0 (0.0)
	Cluster-2	41	32 (78.0)	7 (17.1)	2 (4.9)	41 (100.0)	0 (0.0)	0 (0.0)
	Total	77	40 (51.9)	12 (15.6)	25 (32.5)	48 (62.3)	1 (1.3)	28 (36.4)

N: number of individuals; M: molestus CQ11FL genotype or NEWHYBRIDS pure class; H: hybrid CQ11FL genotype or NEWHYBRIDS class; P: pipiens CQ11FL genotype or NEWHYBRIDS pure class.

The results obtained by NEWHYBRIDS were similar to those of STRUCTURE (Table 2). Of the 29 individuals of cluster-1, 28 were assigned to a purebred pipiens class (96.6%) and all 41 cluster-2 individuals were assigned to a purebred molestus class. However, the seven admixed individuals revealed by STRUCTURE were classified as pure molestus. All individuals classified as molestus by NEWHYBRIDS showed a $qi>0.850$ for this class. There were 15 individuals (53.6%) assigned to the pure pipiens class with $qi<0.850$ for that class (Figure 3). There was a single hybrid individual (backcross pipiens) identified by NEWHYBRIDS, collected in Thessaloniki.

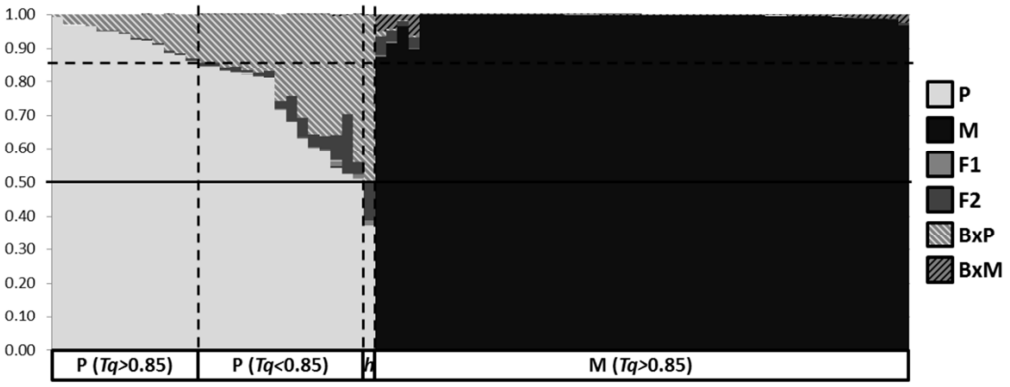

Figure 3. Bayesian cluster analysis conducted by NEWHYBRIDS in Thessaloniki and Schinias-Marathonas.

Each column represents an individual analyzed and it is partitioned into colors according to the probability of assignment to each of the six classes denoted in the label (P: pure pipiens; M: pure molestus; F1 hybrid; F2 hybrid; BxP: backcross with pipiens; BxM: backcross with molestus). *h*: assigned as hybrids. Line indicates the q_i threshold used to assign individuals into classes (see Methods).

CQ11 sequencing

One specimen of Thessaloniki (collected in the village of Anatoliko) showed a previously unknown fragment of 350 bp in the CQ11FL assay. This novel fragment, here termed as Anatoliko_350bp, was amplified by the primer combination pipCQ11R and CQ11F2 that normally amplifies the 200 bp pipiens-specific fragment. Excluding the variable microsatellite TG repeat (139-156 bp), there was an almost complete similarity between Anatoliko_350bp and the CQ11 pipiens-specific sequences (Table 3; Supplementary Figure S2). The single exception was a 149 bp insertion starting at position 219 found in Anatoliko_350bp that was absent from both molestus and pipiens sequences. This unique insertion explains the size difference of the Anatoliko_350bp that could otherwise be considered as a pipiens-specific allele.

Table 3. Nucleotide differences at microsatellite CQ11 between *Cx. pipiens s.s.* forms and the Anatoliko_350bp sequence.

Sequence	Form	34/ 40	41	45	55/ 66	81/ 85	91/ 99	131/ 134	139/ 156	159	219/ 367
DQ470141.1 DQ470148.1	pipiens	-	C	-	-	-TTTC	-	GTGAA	(TG)9 (TG)6	T	-
DQ470149.1 DQ470150.1	molestus	7bp	T	C	17bp	AGATT	9bp	-	(TG)4 (TG)4	C	-
Anatoliko_350bp		-	C	-	-	-TTTC	-	GTGAA	(TG)5	T	149bp

C: Cytosine; T: Thymine; G: Guanine; A: Adenine; -: without nucleotide; bp: base pair of nucleotide.

Discussion

The genetic analysis conducted in this study revealed important differences in *Cx. pipiens s.s.* between the two geographic regions studied. The combined results of the two Bayesian analyses and of the CQ11FL genotyping suggest a sympatric occurrence of molestus and pipiens forms in Thessaloniki, northern Greece, albeit with a higher frequency of pipiens. In this region, admixed individuals were consistently identified by the three analyses. Conversely, results point to a more homogeneous molestus form population in the region of Schinias-Marathonas, southern Greece. NEWHYBRIDS did not detect any hybrid and no pipiens individual was identified by any of the methods in this region. The detection of a higher proportion of admixed individuals by STRUCTURE (9.1%) when compared with NEWHYBRIDS (1.3%) may reflect differences in power and accuracy of the methods in detecting contemporary hybridization (Vähä & Primmer, 2006).

If the initial subdivision of *Cx. pipiens s.s.* samples into two clusters obtained by STRUCTURE was coincident with the geographic origin of samples, subsequent genotyping of the diagnostic CQ11FL marker disclosed a further partitioning corresponding to molestus and pipiens forms. This result agrees with the usefulness of the CQ11FL marker in identifying the presence of molestus and pipiens forms at the population level, even when inter-form hybridisation occurs (Bahnck & Fonseca, 2006; Gomes *et al.*, 2009; Kothera *et al.*, 2010). In this context, it is worth mentioning the finding of a previously unknown 350bp allele at this locus. Sequence analysis supports that the novel allele is a variant of the pipiens-specific 200 bp allele. Moreover, the specimen harbouring this allele was consistently assigned to a pipiens form genetic background in both Bayesian analyses. Future analyses of *Cx. pipiens* populations with the CQ11FL marker should take into consideration this novel allele as a variant of the pipiens form specific allele.

In the analysis conducted by NEWHYBRIDS, a higher degree of genetic backcrossing (mean: 16.2%) was detected in pipiens form individuals when compared to molestus form individuals (mean: 2.6%), in the region of Thessaloniki. This result suggests a pattern of asymmetric introgression, with more molestus genes being introgressed into the pipiens form. This result is consistent with previous observations

carried out in a wetland region of Portugal with an ecological landscape (river delta with rice fields) and climate similar to Thessaloniki (Gomes *et al.,* 2009). Therefore, it appears that a pattern of asymmetric introgression between molestus and pipiens forms is not a localised phenomenon and may reflect intrinsic reproductive isolation mechanisms (*e.g.* different mating strategies, Gomes *et al.,* 2009) shared among southern European *Cx. pipiens s.s.* populations. It would be interesting to investigate whether this happens in other south European settings, such as Spain, Italy and France, and whether this bears any correlation with WNV circulation.

The occurrence of hybridisation between sympatric molestus and pipiens forms found in Thessaloniki suggests the possibility of a more opportunistic feeding behaviour of *Cx. pipiens s.s.*, which may potentiate its role as bridge-vector for WNV (Fonseca *et al.,* 2004; Hamer *et al.,* 2008). This could thus be a major factor contributing for the establishment of WNV transmission in this region. It may also explain the apparent confinement of the 2010 WNV outbreak to northern Greece. To confirm this hypothesis, genotyping of samples from geographic locations intermediate to those here analysed is required in order to determine the extent of molestus/pipiens sympatry and hybridisation southwards of northern Greece. Furthermore, bioecological studies focusing on biting behaviour and host preference are necessary in order to further characterize the vector potential of molestus and pipiens forms in northern Greece. These studies should also include different mosquito sampling strategies as these may affect the correct estimation of the relative abundance of both forms. In the region of Comporta (Portugal), the molestus form was predominant in samples collected indoors resting while the pipiens form prevailed in outdoor collections performed by CDC light traps.

Thessaloniki and Schinias-Marathonas are lowland humid areas with excellent conditions for wild bird fauna. These areas can serve as stopover places (bird sanctuaries) for migratory birds following the Mediterranean/Black Sea flyway during their travel between Africa and their breeding sites in Europe. The analysis performed on the avian fauna composition did not disclose any major differences in species richness (*i.e.* number of different species recorded) between the regions studied in northern and southern Greece. The proportion of trans-Saharan migratory birds and of species with records of WNV infection was similar in both regions. However,

differences were found in the estimates of Shannon's *H'* index which suggest that levels of bird diversity may differ between regions. The higher Shannon's *H'* estimate obtained in Thessaloniki for the group of WNV carriers, indicating a more balanced number of species in terms of relative abundance, is consistent with a higher receptivity of this region in northern Greece for the introduction of WNV. On the other hand, the higher diversity of trans-Saharan migratory species evidenced by the Shannon's *H'* estimate in Schinias-Marathonas also suggests a potential of this region in southern Greece for receiving WNV infected birds from African locations. However, care should be taken when interpreting these diversity estimates, since values of absolute abundance for each species were not available (only maximum bird counts).

Conclusion

The results obtained on the genetic structure of *Cx. pipiens s.s.* suggest a greater receptivity of the Thessaloniki region for the establishment of WNV zoonotic cycles. The region has a *Cx. pipiens s.s.* population with a predominance of the pipiens form. This form, being ornitophilic, should be able to maintain the enzootic viral cycle among avian hosts. Furthermore, the sympatric presence of the molestus form and the occurrence of hybrids may promote a more opportunistic biting behaviour that would facilitate the accidental transmission of the virus to mammalian hosts, including humans. This coupled with the presence of an important fraction of trans-Saharan migratory bird species, some of which have been previously described as potential WNV amplification hosts, could have favoured the onset of the 2010 epidemic that occurred in the region. Although the situation of Schinias-Marathonas, in southern Greece, appears to be different, with a more genetically homogenous molestus form population, the region is still likely to meet conditions to sustain WNV transmission. In fact, the last data on WNV cases reported by the European Centre for Disease Prevention and Control show an increase in the number of confirmed cases in this region (ECDC, 2012). These observations highlight the importance of maintaining active surveillance systems in selected regions such as wetlands that lie within the range of major migratory bird flyways.

Acknowledgements

We thank Spyros Skareas, for providing data on bird fauna abundance, as well as Spyros Mourelatos (Eco-Development, Greece) and Panagiotis Pergantas (Bioapplications, Greece) for their assistance in collecting mosquito specimens. This work received financial support from Fundação para a Ciência e a Tecnologia/FEDER, Portugal (POCI/BIA-BDE/57650/2004 and PPCDT/BIA-BDE/57650/2004). BG was funded by a PhD fellowship of Fundação para a Ciência e Tecnologia/FEDER (SFRH/BD/36410/2007).

References

Anderson, E.C. & Thompson, E.A. (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*. 160 (3), 1217–1229.

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), 251–255.

Birdlife International (2012) *Factsheet about the Mediterranean and Black sea flyway*. . [Online]. 2012. Available from: http://www.birdlife.org/datazone/userfiles/file/sowb/flyways/8_EastAsia_Australasia_NEW.pdf [Accessed: 5 September 2012].

Callot, J. & Van Ty, D. (1943) Sur quelques souches françaises de *Culex pipiens* L. *Bulletin de la Societe de pathologie exotique et de ses filiales*. 36 (7/8), 229–232.

Chaskopoulou, A., Dovas, C., Chaintoutis, S., Bouzalas, I., Ara, G. & Papanastassopoulou, M. (2011) Evidence of enzootic circulation of West Nile virus (Nea Santa-Greece-2010, lineage 2), Greece, May to July 2011. *Eurosurveillance*. 16 (31), pii=19933.

Chevillon, C., Eritja, R., Pasteur, N. & Raymond, M. (1995) Commensalism, adaptation and gene flow: mosquitoes of the *Culex pipiens* complex in different habitats. *Genetical Research*. 66 (2), 147–157.

Clements, A.N. (1999) The Biology of Mosquitoes. Sensory Reception and Behaviour Volume 2. Wallingford, CABI Publishing.

Danis, K., Papa, A., Theocharopoulos, G., Dougas, G., Athanasiou, M., Detsis, M., Baka, A., Lytras, T., Mellou, K., Bonovas, S. & Panagiotopoulos, T. (2011) Outbreak of West Nile virus infection in Greece, 2010. *Emerging Infectious Diseases*. 17 (10), 1868–1872.

European Centre for disease Prevention and Control (ECDC) (2012) *West Nile fever maps*. [Online]. 2012. Available from: http://www.ecdc.europa.eu/en/healthtopics/west_nile_fever/West-Nile-fever-maps/Pages/index.aspx [Accessed: 5 September 2012].

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*. 14 (8), 2611–2620.

Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*. 147–50.

Figuerola, J., Soriguer, R., Rojo, G., Tejedor, C.G. & Jimenez-Clavero, M.A. (2007) Seroconversion in wild birds and local circulation of West Nile virus, Spain. *Emerging Infectious Disease*. 13 (12), 1915–1917.

Figuerola, J., Jiménez-Clavero, M.A., López, G., Rubio, C., Soriguer, R., Gómez-Tejedor, C. & Tenorio, A. (2008) Size matters: West Nile virus neutralizing antibodies in resident and migratory birds in Spain. *Veterinary Microbiology*. 132 (1-2), 39–46.

Figuerola, J., Baouab, R.E., Soriguer, R., Fassi-Fihri, O., Llorente, F. & Jímenez-Clavero, M.A. (2009) West Nile virus antibodies in wild birds, Morocco, 2008. *Emerging Infectious Diseases.* 15 (10), 1651–1653.

Fonseca, D.M., Atkinson, C.T. & Fleischer, R.C. (1998) Microsatellite primers for *Culex pipiens quinquefasciatus*, the vector of avian malaria in Hawaii. *Molecular Ecology*. 7 (11), 1617-1619.

Fonseca, D.M., Keyghobadi, N., Malcolm, C.A., Mehmet, C., Schaffner, F., Mogi, M., Fleischer, R.C. & Wilkerson, R.C. (2004) Emerging Vectors in the *Culex pipiens* Complex. *Science*. 303 (5663), 1535–1538.

Formosinho, P., Santos-Silva, M.M., Santos, A., Melo, P., Encarnação, V., Santos, N., Nunes, T., Agrícola, R. & Portas, M. (2006) O vírus West Nile em Portugal – estudos de vigilância epidemiológica. *Revista Portuguesa de Ciências Veterinárias*. 101 (557-558), 61–68.

Gomes, B., Sousa, C.A., Novo, M.T., Freitas, F.B., Alves, R., Corte-Real, A.R., Salgueiro, P., Donnelly, M., Almeida, A.P.G. & Pinto, J. (2009) Asymmetric introgression between sympatric molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in the Comporta region, Portugal. *BMC Evolutionary Biology*. 9 (1), 262.

Goudet, J. (1995) FSTAT (version 1.2): A computer program to calculate F-statistics. *The Journal of Heredity*. 86 (6), 485–486.

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), 125–128.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*. 41, 95–98.

Harbach, R.E., Harrison, B.A. & Gad, A.M. (1984) *Culex (Culex) molestus* Forskål (Diptera, Culicidae) - neotype designation, description, variation, and taxonomic status. *Proceedings of the Entomological Society of Washington*. 86 (3), 521–542.

Harbach, R.E., Dahl, C. & White, G.B. (1985) *Culex (Culex) pipiens* Linnaeus (Diptera, Culicidae) - concepts, type designations, and description. *Proceedings of the Entomological Society of Washington*. 87 (1), 1–24.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 6 (2), 65–70.

Hubálek, Z. (2004) An Annotated checklist of pathogenic microorganisms associated with migratory birds. *Journal of Wildlife Diseases*. 40 (4), 639–659.

Hubálek, Z. (2008) Mosquito-borne viruses in Europe. *Parasitology Research*. 103 (1), 29–43.

Jakobsson, M. & Rosenberg, N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 23 (14), 1801–1806.

Jayaraman, K. (1999) *A statistical manual for forestry research*. Bangkok, Food and Africulture Organization of the United Nations Regional Office for Asia and the Pacific.

Jourdain, E., Zeller, H.G., Sabatier, P., Murri, S., Kayser, Y., Greenland, T., Lafaye, M. & Gauthier-Clerc, M. (2008) Prevalence of West Nile virus neutralizing antibodies in wild birds from the Camargue area, Southern France. *Journal of Wildlife Diseases*. 44 (3), 766–771.

Kalinowski, S.T. (2005) HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*. 5 (1), 187–189.

Keyghobadi, N., Matrone, M.A., Ebel, G.D., Kramer, L.D. & Fonseca, D.M. (2004) Microsatellite loci from the northern house mosquito (*Culex pipiens*), a principal vector of West Nile virus in North America. *Molecular Ecology Notes*. 4 (1), 20–22.

Kothera, L., Godsey, M., Mutebi, J.-P. & Savage, H.M. (2010) A comparison of aboveground and belowground populations of *Culex pipiens* (Diptera: Culicidae) mosquitoes in Chicago, Illinois, and New York City, New York, using microsatellites. *Journal of Medical Entomology*. 47 (5), 805–813.

Lowry, R. (2012) *VassarStats: Web Site for Statistical Computation*. [Online]. 2012. Available from: http://vassarstats.net/index.html [Accessed: 5 October 2012].

Malkinson, M., Banet, C., Weisman, Y., Pokamunski, S., King, R., Drouet, M.-T. & Deubel, V. (2002) Introduction of West Nile virus in the Middle East by migrating white storks. *Emerging Infectious Diseases*. 8 (4), 392–397.

Nei, M. (1987) *Molecular Evolutionary Genetics*. New York, Colombia University Press.

OrnithoTopos, (2012) [Online]. Available from: http://www.worldbirds.org/v3/greece.php [Accessed: 30 November 2012].

van Oosterhout, C., Hutchinson, W., Wills, D. & Shipley, P. (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*. 4 (3), 535–538.

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.

Pasteur, N., Rioux, J.A., Guilvard, E. & Pech-Perieres, J. (1977) Nouvelle mention pour le “Midi” méditerranéen, de populations naturelles anautogènes. *Annales de Parasitologie Humaine et Comparée*. 11 (3), 187–193.

Peel, M.C., Finlayson, B.L. & McMahon, T.A. (2007) Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*. 4 (2), 1633–1644.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*. 155 (2), 945–959.

Rappole, J.H. & Hubálek, Z. (2003) Migratory birds and West Nile virus. *Journal of Applied Microbiology*. 94 (1), 47–58.

Si, Y., Skidmore, A. K., Wang, T., Boer, W. F. de, Debba, P., Toxopeus, A. G., Li, L. & Prins, H.H.T. (2009) Spatio-temporal dynamics of global H5N1 outbreaks match bird migration patterns. *Geospatial Health*. 4 (1), 65–78.

Shannon, C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*. 27 (3), 379–423.

Slatkin, M. & Excoffier, L. (1996) Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity*. 76 (4), 377–383.

Smith, 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 Hygiene*. 70 (4), 339–345.

Smith, J.L., Keyghobadi, N., Matrone, M.A., Escher, R.L. & Fonseca, D.M. (2005) Cross-species comparison of microsatellite loci in the *Culex pipiens* complex and beyond. *Molecular Ecology Notes*. 5 (3), 697–700.

Sudia, W.D. & Chamberlain, R.W. (1962) Battery-operated light trap. An improved model. *Mosquito News*. 22, 126–129.

Valiakos, G., Touloudi, A., Iacovakis, C., Athanasiou, L., Birtsas, P., Spyrou, V. & Billinis, C. (2011) Molecular detection and phylogenetic analysis of West Nile virus lineage 2 in sedentary wild birds (Eurasian magpie), Greece, 2010. *Eurosurveillance*. 16 (18), pii=19862.

Valiakos, G., Touloudi, A., Athanasiou, L. V, Giannakopoulos, A., Iacovakis, C., Birtsas, P., Spyrou, V., Dalabiras, Z., Petrovska, L. & Billinis, C. (2012) Serological and molecular investigation into the role of wild birds in the epidemiology of West Nile virus in Greece. *Virology journal*. 9 (1), 266.

Vähä, J.-P. & Primmer, C.R. (2006) Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology*. 15 (1), 63–72.

Vinogradova, E.B. (2000) *Culex pipiens pipiens mosquitoes: taxonomy, distribution, ecology, physiology, genetics, applied importance and control*. S.I. Golovatch (ed.). Sofia, Pensoft.

Walther, B.A. (2005) *List of Western Palearctic bird species migrating within Africa (last updated March 2005).* [Online]. Available from: http://www.zmuc.dk/verweb/staff/bawalther/migratoryBirdsList.htm [Accessed: 30 November 2012].

Supplementary Materials

Table S1. List of the bird species in Thessaloniki and Schinias-Marathonas for which the presence of WNV infections or trans-Saharan migration were described.

Scientific name[1]	Common Name[1]	TS	WNV	Max. Count	
				T	SM
Accipiter brevipes	Levant Sparrowhawk	X		1	-
Accipiter nisus	Eurasian Sparrowhawk	X		2	3
Acrocephalus arundinaceus	Great Reed-warbler	X		9	16
Acrocephalus melanopogon	Moustached Warbler			1	6
Acrocephalus schoenobaenus	Sedge Warbler	X		-	6
Acrocephalus scirpaceus	Reed Warbler	X		3	7
Actitis hypoleucos	Common Sandpiper	X		6	2
Aegithalos caudatus	Long-tailed Tit			18	-
Alauda arvensis	Eurasian Skylark			35	10
Alcedo atthis	Common Kingfisher			8	1
Anas acuta	Northern Pintail	X	X	214	10
Anas clypeata	Northern Shoveler	X		46	30
Anas crecca	Common Teal	X		18140	30
Anas penelope	Eurasian Wigeon	X		312	3
Anas platyrhynchos	Mallard	X	X	202	122
Anas querquedula	Garganey	X	X	85	50
Anas strepera	Gadwall	X		15	29
Anthus campestris	Tawny Pipit	X		-	4
Anthus cervinus	Red-throated Pipit	X		4	1
Anthus pratensis	Meadow Pipit			3	20
Anthus spinoletta	Water Pipit			25	12
Anthus trivialis	Tree Pipit	X		-	1
Apus apus	Common Swift	X		35	75
Apus pallidus	Pallid Swift	X		7	-
Aquila clanga	Greater Spotted Eagle	X		6	2
Aquila pomarina	Lesser Spotted Eagle	X		1	-
Ardea cinerea	Grey Heron	X		53	14
Ardea purpurea	Purple Heron	X		22	11
Ardeola ralloides	Squacco Heron	X	X	74	20
Arenaria interpres	Ruddy Turnstone	X	X	25	-
Athene noctua	Little Owl			6	2
Aythya ferina	Common Pochard	X	X	17	3
Aythya nyroca	Ferruginous Duck	X		11	28
Botaurus stellaris	Great Bittern	X	X	1	1
Bubulcus ibis	Cattle Egret		X	9	-

Scientific name[1]	Common Name[1]	TS	WNV	Max. Count	
				T	SM
Burhinus oedicnemus	Eurasian Thick-knee	X		97	-
Buteo buteo	Common Buzzard	X	X	40	5
Buteo rufinus	Long-legged Buzzard	X		2	1
Calandrella brachydactyla	Greater Short-toed Lark	X		2	-
Calidris alba	Sanderling	X	X	15	4
Calidris alpina	Dunlin	X		220	-
Calidris ferruginea	Curlew Sandpiper			200	4
Calidris melanotos	Pectoral Sandpiper			2	-
Calidris minuta	Little Stint	X		300	30
Calidris temminckii	Temminck's Stint	X		4	4
Carduelis cannabina	Eurasian Linnet			145	10
Carduelis carduelis	European Goldfinch			18	40
Carduelis chloris	European Greenfinch			8	20
Carduelis spinus	Eurasian Siskin			6	-
Casmerodius albus	Great Egret			115	1
Certhia brachydactyla	Short-toed Treecreeper			-	2
Cettia cetti	Cetti's Warbler		X	7	30
Charadrius alexandrinus	Kentish Plover	X		125	-
Charadrius dubius	Little Ringed Plover	X		20	20
Charadrius hiaticula	Common Ringed Plover	X		11	-
Chlidonias hybrida	Whiskered Tern	X		7	17
Chlidonias leucopterus	White-winged Tern	X		-	4
Chlidonias niger	Black Tern	X		600	3
Ciconia ciconia	White Stork	X	X	14	-
Circaetus gallicus	Short-toed Snake-eagle	X	X	3	3
Circus aeruginosus	Western Marsh-harrier	X		24	9
Circus cyaneus	Northern Harrier			3	2
Circus macrourus	Pallid Harrier	X		-	1
Circus pygargus	Montagu's Harrier	X		-	3
Cisticola juncidis	Zitting Cisticola			-	11
Coccothraustes coccothraustes	Hawfinch			2	-
Columba livia	Rock Pigeon		X	-	20
Columba palumbus	Common Wood-pigeon			-	1
Coracias garrulus	European Roller	X		1	-
Corvus corone	Carrion Crow		X	600	20
Corvus frugilegus	Rook		X	70	-
Corvus monedula	Eurasian Jackdaw			10	-
Coturnix coturnix	Common Quail	X		-	2

Scientific name[1]	Common Name[1]	TS	WNV	Max. Count	
				T	SM
Cuculus canorus	Common Cuckoo	X		1	2
Cygnus cygnus	Whooper Swan			-	3
Cygnus olor	Mute Swan		X	242	29
Delichon urbicum	Northern House-martin	X		120	20
Egretta garzetta	Little Egret	X		90	100
Emberiza caesia	Cretzschmar's Bunting			-	3
Emberiza cia	Rock Bunting			-	4
Emberiza cirlus	Cirl Bunting			-	3
Emberiza citrinella	Yellowhammer			6	-
Emberiza schoeniclus	Reed Bunting			35	30
Erithacus rubecula	European Robin		X	100	25
Eudromias morinellus	Eurasian Dotterel			1	-
Falco columbarius	Merlin		X	4	-
Falco peregrinus	Peregrine Falcon	X		2	1
Falco subbuteo	Eurasian Hobby	X		1	2
Falco tinnunculus	Common Kestrel	X	X	18	6
Falco vespertinus	Red-footed Falcon	X		1	1
Fringilla coelebs	Chaffinch			300	200
Fringilla montifringilla	Brambling			-	1
Fulica atra	Common Coot	X	X	600	675
Galerida cristata	Crested Lark			26	20
Gallinago gallinago	Common Snipe	X		60	6
Gallinula chloropus	Common Moorhen	X		120	25
Gavia arctica	Arctic Loon			7	1
Gavia stellata	Red-throated Loon			1	-
Glareola pratincola	Collared Pratincole	X		200	3
Grus grus	Common Crane	X		2	-
Gyps fulvus	Eurasian Griffon	X	X	-	3
Haematopus ostralegus	Eurasian Oystercatcher			32	-
Haliaeetus albicilla	White-tailed Eagle			1	-
Hieraaetus fasciatus	Bonelli's Eagle			-	1
Himantopus himantopus	Black-winged Stilt	X		215	40
Hirundo daurica	Red-rumped Swallow	X		4	5
Hirundo rupestris	Eurasian Crag-martin	X		-	10
Hirundo rustica	Barn Swallow	X		60	111
Ixobrychus minutus	Little Bittern	X	X	3	2
Jynx torquilla	Eurasian Wryneck	X		1	1
Lanius collurio	Red-backed Shrike	X		50	10

Scientific name[1]	Common Name[1]	TS	WNV	Max. Count	
				T	SM
Lanius excubitor	Northern Grey Shrike			-	1
Lanius isabellinus	Isabelline Shrike	X		-	1
Lanius minor	Lesser Grey Shrike	X		3	2
Lanius senator	Woodchat Shrike	X	X	1	2
Larus canus	Mew Gull			54	-
Larus fuscus	Lesser Black-backed Gull	X		-	1
Larus genei	Slender-billed Gull	X		124	-
Larus melanocephalus	Mediterranean Gull			1640	1
Larus michahellis	Yellow-legged Gull			2000	200
Larus minutus	Little Gull			3	5
Larus ridibundus	Common Black-headed Gull	X	X	460	50
Limicola falcinellus	Broad-billed Sandpiper	X		19	-
Limosa lapponica	Bar-tailed Godwit	X		2	-
Limosa limosa	Black-tailed Godwit	X		320	5
Lullula arborea	Wood Lark			-	4
Luscinia megarhynchos	Common Nightingale	X	X	3	2
Luscinia svecica	Bluethroat	X		-	1
Lymnocryptes minimus	Jack Snipe	X		-	1
Melanocorypha calandra	Calandra Lark			50	-
Mergus serrator	Red-breasted Merganser			64	-
Merops apiaster	European Bee-eater	X		135	4
Miliaria calandra	Corn Bunting			400	50
Monticola solitarius	Blue Rock-thrush	X		-	1
Motacilla alba	White Wagtail	X	X	7	10
Motacilla cinerea	Grey Wagtail	X		2	1
Motacilla flava	Yellow Wagtail	X		60	26
Muscicapa striata	Spotted Flycatcher	X		5	10
Netta rufina	Red-crested Pochard			-	2
Numenius arquata	Eurasian Curlew	X		409	-
Numenius phaeopus	Whimbrel	X		4	-
Nycticorax nycticorax	Black-crowned Night-heron	X	X	142	8
Oenanthe hispanica	Black-eared Wheatear	X		-	4
Oenanthe oenanthe	Northern Wheatear	X		7	8
Oriolus oriolus	Eurasian Golden-oriole	X		4	-
Otus scops	Common Scops-owl	X		-	2
Pandion haliaetus	Osprey	X		1	1
Parus ater	Coal Tit			-	1
Parus caeruleus	Blue Tit			12	4

Scientific name[1]	Common Name[1]	TS	WNV	Max. Count	
				T	SM
Parus major	Great Tit			40	9
Passer domesticus	House Sparrow		X	140	40
Passer hispaniolensis	Spanish Sparrow			200	-
Passer montanus	Eurasian Tree Sparrow			-	5
Pelecanus crispus	Dalmatian Pelican			112	-
Pelecanus onocrotalus	Great White Pelican	X		17	-
Phalacrocorax aristotelis	European Shag			1	1
Phalacrocorax carbo	Great Cormorant	X		4500	5
Phalacrocorax pygmeus	Pygmy Cormorant			300	-
Phalaropus fulicarius	Grey Phalarope	X		1	-
Phalaropus lobatus	Red-necked Phalarope	X		6	-
Philomachus pugnax	Ruff	X		300	60
Phoenicopterus roseus	Greater Flamingo	X	X	2000	1
Phoenicurus ochruros	Black Redstart		X	6	10
Phoenicurus phoenicurus	Common Redstart	X	X	-	1
Phylloscopus collybita	Common Chiffchaff	X		3	16
Phylloscopus trochilus	Willow Warbler	X	X	3	26
Pica pica	Black-billed Magpie		X	41	100
Platalea leucorodia	Eurasian Spoonbill	X		69	-
Plegadis falcinellus	Glossy Ibis	X	X	66	85
Pluvialis apricaria	Eurasian Golden-plover	X		350	-
Pluvialis squatarola	Grey Plover	X		150	-
Podiceps auritus	Horned Grebe			1	-
Podiceps cristatus	Great Crested Grebe			50	2
Podiceps nigricollis	Black-necked Grebe			65	-
Porzana parva	Little Crake	X		-	1
Porzana porzana	Spotted Crake	X		-	1
Prunella modularis	Hedge Accentor			-	1
Puffinus yelkouan	Yelkouan Shearwater			-	160
Rallus aquaticus	Water Rail			2	8
Recurvirostra avosetta	Pied Avocet	X		220	-
Remiz pendulinus	Eurasian Penduline-tit			14	11
Riparia riparia	Sand Martin	X		215	11
Saxicola rubetra	Whinchat	X		13	10
Saxicola torquatus	Common Stonechat	X		15	25
Serinus serinus	European Serin			-	15
Sitta neumayer	Western Rock-nuthatch			-	3
Somateria mollissima	Common Eider			1	-

Scientific name[1]	Common Name[1]	TS	WNV	Max. Count	
				T	SM
Sterna albifrons	Little Tern	X	X	100	-
Sterna caspia	Caspian Tern	X		6	-
Sterna hirundo	Common Tern	X		930	40
Sterna nilotica	Gull-billed Tern			80	3
Sterna sandvicensis	Sandwich Tern	X		90	5
Streptopelia decaocto	Eurasian Collared-dove			13	30
Streptopelia turtur	European Turtle-dove	X	X	9	5
Sturnus vulgaris	Common Starling		X	1000	8000
Sylvia atricapilla	Blackcap	X	X	-	4
Sylvia cantillans	Subalpine Warbler	X		-	1
Sylvia communis	Common Whitethroat	X	X	11	3
Sylvia hortensis	Orphean Warbler	X		-	1
Sylvia melanocephala	Sardinian Warbler			-	7
Tachybaptus ruficollis	Little Grebe			111	30
Tachymarptis melba	Alpine Swift	X		-	3
Tadorna tadorna	Common Shelduck			1350	6
Tringa erythropus	Spotted Redshank	X		57	3
Tringa glareola	Wood Sandpiper	X		55	50
Tringa nebularia	Common Greenshank	X		20	15
Tringa ochropus	Green Sandpiper	X	X	14	20
Tringa stagnatilis	Marsh Sandpiper	X		54	20
Tringa totanus	Common Redshank	X		200	8
Troglodytes troglodytes	Winter Wren			-	2
Turdus iliacus	Redwing			-	1
Turdus merula	Eurasian Blackbird		X	25	20
Turdus philomelos	Song Thrush			2	3
Turdus pilaris	Fieldfare			-	8
Upupa epops	Eurasian Hoopoe	X	X	6	10
Vanellus spinosus	Spur-winged Lapwing			1	-
Vanellus vanellus	Northern Lapwing		X	1200	3
Xenus cinereus	Terek Sandpiper	X		5	-

[1]OrnithoTopos (2012); T: Thessaloniki; SM: Schinias-Marathonas; TS: performed trans-Saharan migration route (Walther, 2005); WNV: detection of WNV (Figuerola *et al.*, 2007, 2008, 2009; Formosinho *et al.*, 2006; Hubálek 2004, 2008; Jourdain *et al.*, 2008; Malkinson *et al.*, 2002; Rappole & Hubálek 2003; Valiakos *et al.*, 2012); Max. Count: the maximum number of individuals counted at a single observation; T: Thessaloniki; SM: Schinias-Marathonas.

Table S2. Genetic diversity at microsatellite loci of *Culex pipiens s.s.* from Greece.

Locus	Region	Thessaloniki				Schinias-Marathonas			Total (N=77)
	STRUCTURE	C1(P) (N=29)	Ad (N=4)	C2(M) (N=4)	Total (N=37)	Ad (N=1)	C2(M) (N=39)	Total (N=40)	
CQ11	$A_{R(6)}$	3.9	1.0	1.0	3.9	NA	1.4	1.4	3.0
	H_e	0.800*	NA	NA	**0.813***	NA	0.148	0.144	**0.609**
	F_{IS}	**0.314**	NA	NA	**0.425**	NA	-0.041	-0.040	**0.513**
CQ26	$A_{R(6)}$	3.2	1.0	2.0	2.9	NA	3.0	3.0	3.5
	H_e	0.693	NA	0.536	0.641	NA	0.682	0.677	**0.767**
	F_{IS}	0.157	NA	0.571	0.222	NA	0.270	0.283	**0.358**
CQ41	$A_{R(6)}$	4.0	2.9	3.3	3.9	NA	4.4	4.4	4.3
	H_e	**0.820***	0.714	0.643	**0.806***	NA	**0.866***	**0.866***	**0.852**
	F_{IS}	0.331	1.000	-0.200	**0.366**	NA	**0.476**	**0.476**	**0.431**
CxpGT04	$A_{R(6)}$	4.4	3.5	3.7	4.4	NA	2.9	2.9	4.0
	H_e	0.872	0.750	0.821	0.879	NA	0.670	0.670	**0.810**
	F_{IS}	0.132	0.700	-0.263	0.172	NA	0.137	0.137	**0.194**
CxpGT09	$A_{R(6)}$	4.1	2.7	2.7	4.0	NA	2.8	2.8	3.8
	H_e	0.846	0.607	0.607	0.833	NA	**0.616***	**0.622***	**0.773**
	F_{IS}	-0.019	-0.286	-0.286	-0.006	NA	**0.494**	**0.468**	**0.239**
CxpGT12	$A_{R(6)}$	3.8	2.7	2.8	3.7	NA	3.0	3.0	3.4
	H_e	**0.803***	0.607	0.679	**0.795***	NA	0.665	0.656	**0.736**
	F_{IS}	**0.446**	0.200	0.294	**0.425**	NA	0.037	0.048	**0.261**
CxpGT20	$A_{R(6)}$	5.1	4.2	2.0	4.8	NA	5.0	5.0	5.1
	H_e	0.931	0.857	0.571	0.915	NA	**0.929***	**0.929***	**0.934**
	F_{IS}	0.038	0.143	1.000	0.145	NA	**0.363**	**0.363**	**0.263**
CxpGT40	$A_{R(6)}$	5.1	3.5	3.5	4.9	NA	3.3	3.3	4.2
	H_e	0.931	0.750	0.750	0.926	NA	0.707	0.708	**0.843**
	F_{IS}	0.037	0.000	0.000	0.066	NA	0.058	0.047	0.091
CxpGT46	$A_{R(6)}$	3.2	3.5	2.7	3.3	NA	2.7	2.8	3.0
	H_e	0.700	0.786	0.607	0.726	NA	0.636	0.651	0.687
	F_{IS}	0.015	0.400	0.625	0.146	NA	-0.035	-0.024	0.062
CxpGT51	$A_{R(6)}$	5.4	4.2	3.5	5.3	NA	4.8	4.8	5.3
	H_e	0.954	0.857	0.786	0.951	NA	0.908	0.910	**0.949**
	F_{IS}	0.025	0.143	-0.333	0.034	NA	-0.046	-0.045	0.015
CxpGT53	$A_{R(6)}$	5.4	4.9	2.5	5.2	NA	4.4	4.4	5.0
	H_e	0.954	0.929	0.464	0.942	NA	0.870*	0.866*	**0.920**
	F_{IS}	0.025	0.217	-0.091	0.083	NA	**0.281**	**0.297**	**0.197**
CxqQGT4	$A_{R(6)}$	1.3	1.0	1.0	1.2	NA	1.1	1.2	1.2
	H_e	0.102	NA	NA	0.080*	NA	0.026	0.073	0.076
	F_{IS}	-0.014	NA	NA	-0.014	NA	0.000	0.661	0.321
CxqGT6B	$A_{R(6)}$	3.5	3.5	3.0	3.4	NA	3.1	3.2	3.4
	H_e	0.761	0.786	0.750	0.752	NA	0.710	0.724	0.751
	F_{IS}	0.049	-0.333	-0,412	-0.043	NA	0.062	0.104	0.049
CxqTRI4	$A_{R(6)}$	2.6	1.0	1.0	2.3	NA	1.1	1.2	1.9
	H_e	0.574	NA	NA	0.488	NA	0.026	0.074	**0.300**
	F_{IS}	-0.144	NA	NA	-0.054	NA	0.000	0.664	0.134
All loci	$A_{R(6)}$	3.9	2.8	2.5	3.8	NA	3.1	3.1	3.7
	H_e	**0.767**	0.764	0.656	**0.753**	NA	**0.604**	**0.612**	**0.715**
	F_{IS}	**0.109**	0.242	0.074	**0.154**	NA	**0.191**	**0.205**	**0.215**

C1(P): pipiens cluster; Ad: Admixed; C2(M): molestus cluster; $A_{R(6)}$: allelic richness with 6 genes; *He*: expected heterozygosity; F_{IS}: inbreeding coefficient. Values in bold indicate a significant *P*-value after correction for multiple tests (see Methods). Asterisks indicate presence of null alleles determined by Micro-Checker (Populations Thessaloniki/Ad, Thessaloniki/C2, and Schinias-Marathonas/Ad showed insufficient number of individuals for this analysis). Per locus and over samples Hardy-Weinberg tests were performed using ARLEQUIN. For over loci estimates the global test available in FSTAT was used.

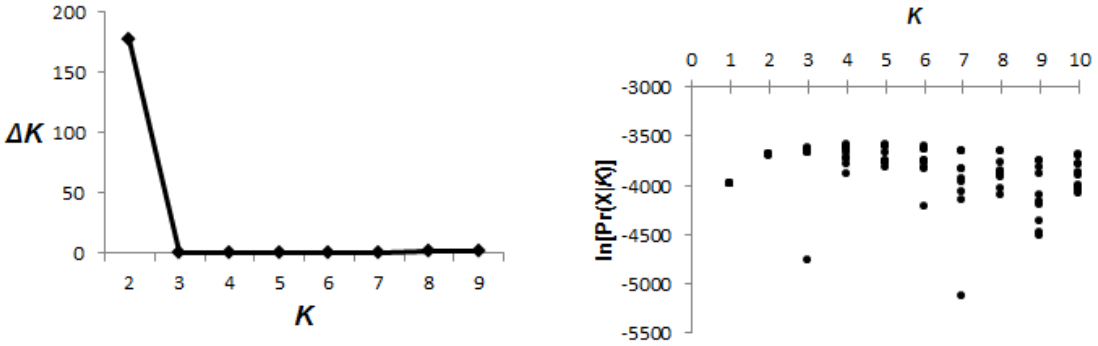

Figure S1. Graphics of *ad hoc* approaches to infer the number of clusters (*K*) in STRUCTURE analysis.

K: number of clusters; Δ*K*: see Evanno *et al.* (2005); ln[Pr(X|*K*)]: estimated log probability of the data under each *K*.

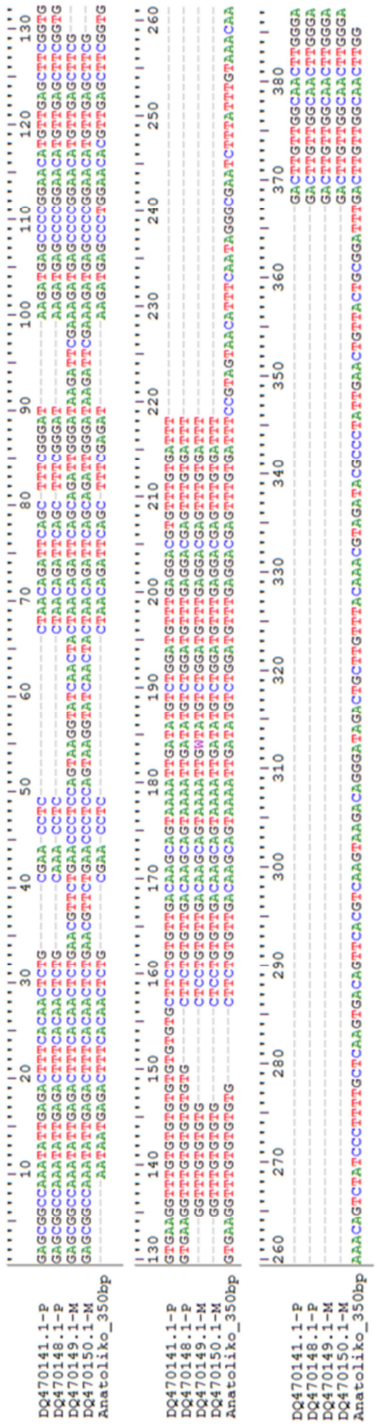

Figure S2. Sequence alignment of the CQ11 genomic region.

P: sequence from pipiens form; M: sequence from molestus form; Anatoliko 350bp: sequence from the CQ11FL350/350 homozygote (sample collected in Anatoliko); A: Adenine; G: Guanine; C: Cytosine; T: Thymine. TG repeat: between 139-156 bp; INDEL: between 219-367 bp.

Chapter 7.

General Discussion

Discussion

Hybridisation was found in all the sympatric populations of the members of the *Cx. pipiens* complex. In the sympatric *Cx. quinquefasciatus* and *Cx. pipiens s.s.* populations of Cape Verde, interspecific hybridisation reached ~40%. A high percentage of these hybrids (~50%) were shown to involve second generation hybrids (F2 and backcross *Cx. pipiens s.s.*), highlighting the potential for introgression of adaptively important genes between the two species. Cape Verde archipelago is located at the extreme limits of the geographic distribution of both species. Given its geographic location, the volcanic nature of the archipelago and the history of human peopling of the islands, the sympatric occurrence of *Cx. pipiens s.s.* and *Cx quinquefasciatus* most likely reflects a case of recent secondary contact between the two species. Secondary contact has led to high hybridisation rates in other groups of insects and also in mammals (*e.g.* Gompert *et al.,* 2010; Latch *et al.,* 2011). The existence of hybridisation between different species in secondary contact areas does not necessarily imply a scenario of admixture. Hybrids may display lower fitness which would allow selective pressures to counteract gene flow (Hopkins & Rausher, 2011). However, fitness studies were not possible to carry out, which would help clarifying the consequences of hybridisation between the two species in Cape Verde (Chapter 2).

In the case of *Cx. pipiens s.s.* in southern Europe, coexistence with hybridisation (7-16%) between molestus and pipiens forms was found in aboveground habitats. A pattern of asymmetric introgression from molestus into pipiens was also observed in two geographic areas (*ca.* ~2700 km from each other), which may be a common phenomenon in areas of southern Europe with similar ecological characteristics (Chapter 3 and 6). The maintenance of differentiated genetic backgrounds for each of the *Cx. pipiens s.s.* forms under asymmetric introgression may be explained by divergent selection acting on some genomic regions, which would promote a heterogeneous genetic divergence across the genome (Wu, 2001). This phenomenon has been considered a crucial factor in speciation models involving sympatric taxa or in the reinforcement of isolation between allopatric incipient species after secondary contact (Nosil *et al.,* 2009, 2012; Hopkins & Rausher, 2011). The outlier analysis performed using AFLPs indicate that only 1.4-3.1% of loci were under divergent selection in

European populations (Chapter 5). The relatively small number of genomic regions under strong divergence is consistent with a model of ecological sympatric speciation at an early stage (Nosil *et al.*, 2009). Such a low genomic divergence was also observed in another mosquito species undergoing sympatric speciation with gene flow (Weetman *et al.*, 2012).

In Comporta (Portugal), the *Cx. pipiens s.s.* forms displayed differences according to the type and location of the collections performed. The lack of molestus individuals in collections performed outside animal shelters suggests possible differences in the selection of aboveground habitats by the adult females, which may lead to habitat segregation (Chapter 4). This pattern may be influenced by the typological traits of this form, namely autogeny and stenogamy, which may favour a propensity for exploring more confined indoor environments. Habitat segregation has been considered a major factor underlying the divergence between incipient species of *An. gambiae s.s.* in sympatry (Diabaté *et al.*, 2008, 2009), and may also play an important role in speciation and heterogeneous genomic divergence between the pipiens and molestus form in southern Europe.

In addition to genetic determinants, the behavioural/physiological traits that define molestus and pipiens forms are also influenced by environmental factors. In the study conducted in Comporta (Portugal), a precise definition of the feeding patterns of molestus and pipiens forms was more challenging than the definition of autogeny and stenogamy (Chapter 3 and 4). Of the blood meals analysed, the great majority was made on avian hosts regardless of form. This observation may result not only by intrinsic factors but also from external factors such as the availability and behaviour of the hosts (Balenghien *et al.*, 2011). In addition, *Cx. pipiens s.s.* forms from Comporta (Portugal) also displayed differences in resting behaviour, which is considered a determinant behaviour for indoors-based vector control. Results point for a capacity of pipiens form females to explore both indoors and outdoors resting sites, suggesting a higher plasticity for this trait when compared to a more endophilic molestus form.

In conclusion, the characterization of the *Cx. pipiens* complex members is an important requirement for understanding the epidemiology of West Nile virus (WNV). Their evolutionary relationships and ecological traits are essential for evaluating the

potential of these species to act as bridge-vectors between avian amplification hosts and mammals, including humans. Risk assessment for WNV transmission should also include information on other vector species, such as *Cx. modestus* and *Cx. perexiguus*, and on the distribution and migration patterns of bird populations.

Future perspectives

Further ecological studies on Cx. pipiens s.s. populations.

The ecological and physiological differences between molestus and pipiens forms have been well established in allopatric populations of northern latitudes. The habitat segregation between forms at these latitudes (*i.e.* underground *vs.* aboveground) facilitated this distinction. The analyses performed in this thesis with samples from southern Europe, have shown that broadly the form specific differences in egg development, life cycle and mating behaviour described at northern latitudes are kept, albeit with some exceptions (*e.g.* small proportions of autogenous and non-stenogamous females). It is tempting to speculate that this is a result of the low but significant gene flow between the two forms documented in this thesis. However the established differences in host selection were not recorded for sympatric aboveground molestus and pipiens populations. This result may have been more influenced by host availability at the study site rather than reflecting intrinsic host preference. Further studies, preferably in other study areas with similar hybridisation rates but with a more balanced mammal/avian host availability, will be required to further clarify the effect of hybridisation and introgression in the feeding behaviour of molestus and pipiens forms. Given that this is a key trait for predicting the potential role of *Cx. pipiens s.s.* as a bridge-vector for transmission of WNV and other arboviral infections there is some urgency to these studies.

Another relevant aspect that was left unexplored in these studies was larval ecology. In the malaria vector *An. gambiae s.s.*, there is evidence for habitat segregation at the larval stage between M and S forms within this species (Diabaté *et al.* 2008). The S-form is more adapted to small temporary breeding puddles whereas the M-form explores more successfully larval habitats of more permanent nature such as rice fields, arguably due to a superior predator avoidance capacity (Diabaté *et al.* 2008; Gimonneau

et al., 2012). This difference has been considered as a potential mechanism of divergent selection between these two incipient species. In northern latitudes, the larval habitat of the allopatric populations of pipiens and molestus forms varies between open air and underground, respectively (Vinogradova, 2000). However, few studies have specifically addressed the larval ecology of the forms in southern latitudes. No systematic characterization of larval biotopes was made in this study. However, based on the adult surveys, it appears that the pipiens form may prevail over molestus in humid areas with rice fields. Such was the case for Thessaloniki (Greece) and Comporta (Portugal), both rice cultivation areas, when compared to other regions such as Alqueva and Sandim (Portugal), and Schinias-Marathonas (Greece). Studies focusing on the larval ecology study should be implemented in order to unveil potential differences between molestus and pipiens in southern latitudes. If a common evolutionary origin for northern and southern latitude populations of the molestus form is assumed, one would predict a tendency for this form to explore more readily larval habitats of anthropogenic nature. Larval ecology studies would also be important for the application of vector control strategies based either in ecological management of breeding sites or in biolarvicides (*e.g. Bacillus thuringiensis*).

Correlation between ecological traits and genomic regions

The results presented in this thesis showed an association between ecological traits such as autogeny and stenogamy with the genetic background of the forms. However, it is still unknown which genomic regions are involved in the expression of the typological traits of the forms. Data collected from the *Cx. quinquefasciatus* genome suggest an expansion, when compared with other mosquito species, of several gene families that may be involved in these traits, such as juvenile hormone genes and olfactory-receptor genes (Arensburger *et al.*, 2010). However, no comparative studies between pipiens and molestus forms have yet been made. The application of Next Generation Sequencing methods would allow the collection of genomic information from both forms that would refine the data obtained by AFLP about divergent regions putatively involved in the speciation process. The large amount of DNA sequence data would also be useful for identifying differences in gene families that may be involved in the expression of ecological traits, as well as genes that influence the capacity for pathogen transmission, such as mosquito immune-related genes.

The role of endosymbiotic bacteria in the evolution of the Cx. pipiens complex

Populations of the *Cx. pipiens* complex may harbour maternally-inherited *Wolbachia pipientis* endosymbiotic bacteria. This bacterium reduces reproductive success by cytoplasmic incompatibility in the progeny of mating pairs that present different infection states (*i.e.* infected male with uninfected female or infection with different *W. pipientis* strains) (Morningstar *et al.*, 2012). This mechanism has been proposed to be involved in the isolation of *Cx. pipiens s.s.* and *Cx. quinquefasciatus* in East Africa, where uninfected of *Cx. pipiens s.s.* populations were found in sympatry with infected *Cx. quinquefasciatus* (Cornel *et al.*, 2003; Walker *et al.*, 2009). In this context, it would be interesting to analyse the presence of *W. pipientis* in *Cx. pipiens s.s.* and *Cx. quinquefasciatus* from Cape Verde islands, in order to verify if there is an association between incomplete isolation of these species and the presence and genetic composition of their endosymbionts.

Wolbachia pipientis was also associated with low reproductive success among populations of *Cx. quinquefasciatus*, especially from Asia (Sinkins *et al.*, 2005). However, it is still unclear whether this infection plays a role in the divergence between the intraspecific forms of *Cx. pipiens s.s.* Surveys for endosymbiotic bacterial fauna in molestus and pipiens forms would be important to clarify the mechanisms involved in the asymmetric introgression in Comporta (Portugal) and Thessaloniki (Greece), which may be induced by unidirectional cytoplasmic incompatibility. Furthermore, given that *W. pipientis* seems to affect the infection of *Cx. quinquefasciatus* with WNV (Glaser & Meola, 2010), it would be interesting to determine whether infection with this endosymbiont also affects vector competence for WNV in *Cx pipiens s.s.*

References

Arensburger, P., Megy, K., Waterhouse, R.M., Abrudan, J., Amedeo, P., Antelo, B., Bartholomay, L., Bidwell, S., Caler, E., Camara, F., Campbell, C.L., Campbell, K.S., Casola, C., Castro, M.T., Chandramouliswaran, I., Chapman, S.B., Christley, S., Costas, J., Eisenstadt, E., Feschotte, C., Fraser-Liggett, C., Guigo, R., Haas, B., Hammond, M., Hansson, B.S. Hemingway, J., Hill, S.R., Howarth, C., Ignell, R., Kennedy, R.C., Kodira, C.D., Lobo, N.F., Mao, C., Mayhew, G., Michel, K., Mori, A., Liu, N., Naveira, H., Nene, V., Nguyen, N., Pearson, M.D., Pritham, E.J., Puiu, D., Qi, Y., Ranson, H., Ribeiro, J.M.C., Roberston, H.M., Severson, D.W., Shumway, M., Stanke, M., Strausberg, R.L., Sun, C., Sutton, G., Tu, Z.J., Tubio, J.M.C., Unger, M.F., Vanlandingham, D.L., Vilella, A.J., White, O., White, J.R., Wondji, C.S., Wortman, J., Zdobnov, E.M., Birren, B., Christensen, B.M., Collins, F.H., Cornel, A., Dimopoulos, G., Hannick, L.I., Higgs, S., Lanzaro, G.C., Lawson, D., Lee, N.H., Muskavitch, M.A.T., Raikhel, A.S. & Atkinson, P.W. (2010) Sequencing of *Culex quinquefasciatus* establishes a platform for mosquito comparative genomics. *Science.* 330 (6000), 86–88.

Balenghien, T., Fouque, F., Sabatier, P. & Bicout, D.J. (2011) Theoretical formulation for mosquito host-feeding patterns: Application to a West Nile virus focus of southern France. *Journal of Medical Entomology.* 48 (5), 1076–1090.

Cornel, A.J., Mcabee, R.D., Rasgon, J., Stanich, M.A., Scott, T.W. & Coetzee, M. (2003) Differences in extent of genetic introgression between sympatric *Culex pipiens* and *Culex quinquefasciatus* (Diptera: Culicidae) in California and South Africa. *Journal of Medical Entomology*. 40 (1), 36–51.

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.

Diabaté A, D.A., Yaro, A.S., Adamou, A., Gonzalez, R., Manoukis, N.C., Traoré, S.F., Gwadz, R.W. & Lehmann, T. (2009) Spatial swarm segregation and reproductive isolation between the molecular forms of *Anopheles gambiae*. *Proceedings. Biological sciences / The Royal Society*. 276 (1676), 4215-4222.

Gimonneau, G., Pombi, M., Choisy, M., Morand, S., Dabiré, R.K. & Simard, F. (2012) Larval habitat segregation between the molecular forms of the mosquito *Anopheles gambiae* in a rice field area of Burkina Faso, West Africa. *Medical and Veterinary Entomology*. 26 (1), 9–17.

Glaser, R.L. & Meola, M.A. (2010) The native *Wolbachia* endosymbionts of *Drosophila melanogaster* and *Culex quinquefasciatus* increase host resistance to West Nile virus infection. *PLoS ONE*. 5 (8), e11977.

Gompert, Z., Lucas, L.K., Fordyce, J.A., Forister, M.L. & Nice, C.C. (2010) Secondary contact between *Lycaeides idas* and *L. melissa* in the Rocky Mountains: extensive admixture and a patchy hybrid zone. *Molecular Ecology*. 19 (15), 3171–92.

Hopkins, R. & Rausher, M.D. (2011) Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature*. 469 (7330), 411–414.

Latch, E.K., Kierepka, E.M., Heffelfinger, J.R. & Rhodes, O.E. (2011) Hybrid swarm between divergent lineages of mule deer (*Odocoileus hemionus*). *Molecular Ecology*. 20 (24), 5265–5279.

Morningstar, R.J., Hamer, G.L., Goldberg, T.L., Huang, S., Andreadis, T.G. & Walker, E.D. (2012) Diversity of *Wolbachia pipientis* strain wPip in a genetically admixtured, above-ground *Culex pipiens* (Diptera: Culicidae) population: association with form molestus ancestry and host selection patterns. *Journal of Medical Entomology*. 49 (3), 474–481.

Nosil, P., Harmon, L.J. & Seehausen, O. (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*. 24 (3), 145–156.

Nosil, P., Gompert, Z., Farkas, T.E., Comeault, A.A., Feder, J.L., Buerkle, C.A. & Parchman, T.L. (2012) Genomic consequences of multiple speciation processes in a stick insect. *Proceedings Biological sciences / The Royal Society*. 279 (1749), 5058-5065.

Sinkins, S.P., Walker, T., Lynd, A.R., Steven, A.R., Makepeace, B.L., Godfray, H.C. & Parkhill, J. (2005) *Wolbachia* variability and host effects on crossing type in *Culex* mosquitoes. *Nature*. 436 (7048), 257–260.

Vinogradova, A.N. (2000) *Culex pipiens pipiens mosquitoes: Taxonomy, Distribution, Ecology, Physiology, Genetics, Applied Importance and Control*. Pensoft Publishers.

Walker, T., Song, S. & Sinkins, S.P. (2009) *Wolbachia* in the *Culex pipiens* group mosquitoes: Introgression and superinfection. *Journal Heredity*. 100 (2), 192–196.

Weetman, D., Wilding, C.S., Steen, K., Pinto, J. & Donnelly, M.J. (2012) Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Molecular Biology and Evolution*. 29 (1), 279–291.

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*. 14 (6), 851–865.

Annex 1

Ecology and Evolution

Open Access

Hybridization and population structure of the *Culex pipiens* complex in the islands of Macaronesia

Bruno Gomes[1,2], Joana Alves[1,2,3], Carla A. Sousa[1,4], Marta Santa-Ana[5], Inês Vieira[1,2], Teresa L. Silva[1,2], António P.G. Almeida[1,4], Martin J. Donnelly[6] & João Pinto[1,2]

[1]Unidade de Parasitologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008, Lisbon, Portugal
[2]Centro de Malária e outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008, Lisbon, Portugal
[3]Direcção-Geral da Saúde Ministério da Saúde, Palácio do Governo, CP 47, Praia, Cabo Verde
[4]Unidade de Parasitologia e Microbiologia Médicas, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008, Lisbon, Portugal
[5]Centro de Estudos da Macaronésia, Universidade da Madeira, Campus da Penteada, 9000-390, Funchal, Portugal
[6]Vector Group, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

Keywords
Culex pipiens, *Culex quinquefasciatus*, hybridization, Macaronesian islands, West Nile virus.

Correspondence
João Pinto, Unidade de Parasitologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008 Lisbon, Portugal. Tel: +351 213652626 / +351 213622458; Fax: +351 213622458; E-mail: jpinto@ihmt.unl.pt

Funding Information
This study was funded by Fundação para a Ciência e a Tecnologia, Portugal (POCI/BIA-BDE/57650/2004 and PPCDT/BIA-BDE/57650/2004; PPCDT/SAU-ESP/55110/2004). Joana Alves and Bruno Gomes were funded by a PhD fellowship of Fundação para a Ciência e Tecnologia/MCTES (SFRH/BD/153451/2005, SFRH/BD/36410/2007).

Received: 16 May 2012; Revised: 31 May 2012; Accepted: 5 June 2012

Ecology and Evolution **2012; 2(8): 1889–1902**

doi: 10.1002/ece3.307

Abstract

The *Culex pipiens* complex includes two widespread mosquito vector species, *Cx. pipiens* and *Cx. quinquefasciatus*. The distribution of these species varies in latitude, with the former being present in temperate regions and the latter in tropical and subtropical regions. However, their distribution range overlaps in certain areas and interspecific hybridization has been documented. Genetic introgression between these species may have epidemiological repercussions for West Nile virus (WNV) transmission. Bayesian clustering analysis based on multilocus genotypes of 12 microsatellites was used to determine levels of hybridization between these two species in Macaronesian islands, the only contact zone described in West Africa. The distribution of the two species reflects both the islands' biogeography and historical aspects of human colonization. Madeira Island displayed a homogenous population of *Cx. pipiens*, whereas Cape Verde showed a more intriguing scenario with extensive hybridization. In the islands of Brava and Santiago, only *Cx. quinquefasciatus* was found, while in Fogo and Maio high hybrid rates (~40%) between the two species were detected. Within the admixed populations, second-generation hybrids (~50%) were identified suggesting a lack of isolation mechanisms. The observed levels of hybridization may locally potentiate the transmission to humans of zoonotic arboviruses such as WNV.

Introduction

The biological diversity of islands with recent volcanic origin and high isolation from mainland is a result of the colonizers ability to break the isolation and survive the island's environmental conditions. The highly stochastic nature of colonizing events means that only a very limited number of taxa may be present in each archipelago (Gillespie and Roderick 2002). For example, in Hawaii, only 15% of the known insect families were observed (Howarth and Ramsay 1991), and a similar scenario occurs in the Macaronesian region (Juan et al. 2000; Gillespie and Roderick 2002). This region is formed by four archipelagos of volcanic islands located in the north-

BMC Evolutionary Biology

Research article

Asymmetric introgression between sympatric molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in the Comporta region, Portugal

Bruno Gomes[1], Carla A Sousa[2], Maria T Novo[2], Ferdinando B Freitas[1,2], Ricardo Alves[1,2], Ana R Côrte-Real[1,2], Patrícia Salgueiro[1], Martin J Donnelly[3], António PG Almeida[2] and João Pinto*[1]

Address: [1]Centro de Malária e outras Doenças Tropicais, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 96, 1349-008 Lisboa, Portugal, [2]Unidade de Entomologia Médica/Unidade de Parasitologia e Microbiologia Médicas, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 96, 1349-008 Lisboa, Portugal and [3]Vector Group, Liverpool School of Tropical Medicine, Pembroke Place L3 5QA, Liverpool, UK

Email: Bruno Gomes - bsilva@ihmt.unl.pt; Carla A Sousa - casousa@ihmt.unl.pt; Maria T Novo - tenovo@ihmt.unl.pt; Ferdinando B Freitas - ferdinando@ihmt.unl.pt; Ricardo Alves - ricardo.alves@zoo.ox.ac.uk; Ana R Côrte-Real - anarutecreal@gmail.com; Patrícia Salgueiro - psalgueiro@ihmt.unl.pt; Martin J Donnelly - M.J.Donnelly@liverpool.ac.uk; António PG Almeida - palmeida@ihmt.unl.pt; João Pinto* - jpinto@ihmt.unl.pt

* Corresponding author

Published: 6 November 2009

BMC Evolutionary Biology 2009, **9**:262 doi:10.1186/1471-2148-9-262

Received: 9 July 2009
Accepted: 6 November 2009

This article is available from: http://www.biomedcentral.com/1471-2148/9/262

Abstract

Background: *Culex pipiens* L. is the most widespread mosquito vector in temperate regions. This species consists of two forms, denoted molestus and pipiens, that exhibit important behavioural and physiological differences. The evolutionary relationships and taxonomic status of these forms remain unclear. In northern European latitudes molestus and pipiens populations occupy different habitats (underground *vs*. aboveground), a separation that most likely promotes genetic isolation between forms. However, the same does not hold in southern Europe where both forms occur aboveground in sympatry. In these southern habitats, the extent of hybridisation and its impact on the extent of genetic divergence between forms under sympatric conditions has not been clarified. For this purpose, we have used phenotypic and genetic data to characterise *Cx. pipiens* collected aboveground in Portugal. Our aims were to determine levels of genetic differentiation and the degree of hybridisation between forms occurring in sympatry, and to relate these with both evolutionary and epidemiological tenets of this biological group.

Results: Autogeny and stenogamy was evaluated in the F1 progeny of 145 individual *Cx. pipiens* females. Bayesian clustering analysis based on the genotypes of 13 microsatellites revealed two distinct genetic clusters that were highly correlated with the alternative traits that define pipiens and molestus. Admixture analysis yielded hybrid rate estimates of 8-10%. Higher proportions of admixture were observed in pipiens individuals suggesting that more molestus genes are being introgressed into the pipiens form than the opposite.

Conclusion: Both physiological/behavioural and genetic data provide evidence for the sympatric occurrence of molestus and pipiens forms of *Cx. pipiens* in the study area. In spite of the significant genetic differentiation between forms, hybridisation occurs at considerable levels. The observed pattern of asymmetric introgression probably relates to the different mating strategies adopted by each form. Furthermore, the differential introgression of molestus genes into the pipiens form may induce a more opportunistic biting behaviour in the latter thus potentiating its capacity to act as a bridge-vector for the transmission of arboviral infections.