# Chromosome and Genome Evolution in Culicinae Mosquitoes

Reem Abed Masri

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Entomology

Maria V. Sharakhova
Igor V. Sharakhov
Sally L. Paulson
Daniela Cimini
Zhijian (Jake) Tu

April 30, 2021
Blacksburg VA, USA

Keywords: Mosquito, *Aedes albopictus*, *Culex pipiens* complex, physical map, genome assembly, Fluorescent *in situ* hybridization

Chromosome and Genome Evolution in Culicinae Mosquitoes

Reem Abed Masri

## ABSTRACT

The Culicinae is the most extensive subfamily among the Culicidae family of mosquitoes. Two genera, *Culex* and *Aedes*, from this subfamily have world-wide distribution and are responsible for transmitting of several deadly diseases including Zika, West Nile fevers, chikungunya, dengue, and Rift Valley fevers. Developing high-quality genome assembly for mosquitoes, studying their population structure, and evolution can help to facilitate the development of new strategies for vector control. Studies on *Aedes albopitcus* as well as on species from the *Culex pipiens* complex, which are widely spread in the United States, provide excellent models on these topics.

*Ae. albopictus* is one of the most dangerous invasive mosquito species in the world that transmits more than 20 arboviruses. This species has highly repetitive genome that is the largest among mosquito genomes sequenced so far. Thus, sequencing and assembling of such genome is extremally challenging. As a result, the lack of high-quality *Ae. albopictus* genome assembly has delayed the progress in understanding its biology. To produce a high-quality genome assembly, it was important to anchor genomic scaffolds to the cytogenetic map creating a physical map of the genome assembly. We first developed a new gene-based approach for the physical mapping of repeat-rich mosquito genomes. The approach utilized PCR amplification of the DNA probes based on complementary DNA (cDNA) that does not include repetitive DNA sequences. This method was then used for the development of a physical map for *Ae. albopictus* based on the *in situ* hybridization of fifty cDNA fragments or gene exons from twenty-four scaffolds to the mitotic chromosomes from imaginal discs. This study resulted in the construction of a first physical map of the *Ae. albopictus* genome as well as mapping viral integration and polyphenol oxidase genes. Moreover, comparing our present *Ae. albopictus* physical map to the current *Ae. aegypti* assembly indicated the presence of multiple chromosomal inversions between them.

To better understand population structure and chromosome evolution in Culicinae mosquitoes, especially in the *Culex pipiens* complex, we studied genomic and chromosomal differentiation between two subspecies *Cx. pipiens pipiens* and *Cx. pipiens molestus*. For the species responsible for the spread of human diseases, understanding the population dynamics and processes of taxa diversification is important for an effective mosquito control . Two vectors of West Nile virus, *Cx. p. pipiens* and *Cx. p. molestus*, exhibit epidemiologically important behavioral and physiological differences, but the whole-genome divergence between them was unexplored. The first goal of this study was to better understand the level of genomic differentiation and population structures of *Cx. p. pipiens* and *Cx. p. molestus* from different continents. We sequenced and compared

whole genomes of 40 individual mosquitoes from two locations in Eurasia and two in North America. Principal Component, ADMIXTURE, and neighbor joining analyses of the nuclear genomes identified two major intercontinental, monophyletic clusters of *Cx. p. pipiens* and *Cx. p. molestus*. The level of genomic differentiation between the subspecies was uniform along chromosomes. The ADMIXTURE analysis determined signatures of admixture in *Cx. p. pipens* populations, but not in *Cx. p. molestus* populations. Thus, our study identified that *Cx. p. molestus* and *Cx. p. pipiens* represent different evolutionary units with monophyletic origin that have undergone incipient ecological speciation. The second goal was to study differences at the chromosome level between these two organisms. We first measured whole chromosome and chromosome arm length differences between *Cx. p. molestus* and *Cx. p. pipiens* as a basic cytogenetic approach. In addition, we used the novel Hi-C approach to detect chromosomal rearrangements between them since Hi-C was successful in detecting a known inversion in *Cx. quinquefasciatus. Cx. p. molestus* and *Cx. p. pipiens* embryos were used to perform the Hi-C technique. Analysis of the Hi-C data showed the presence of two different inversions in *Cx. p. pipiens* and *Cx. p. molestus* heatmap, which could explain their different physiology and adaptation in nature. Developing modern genomic and cytogenetic tools is important to enhance the quality of genome assemblies, improve gene annotation, and provide a better framework for comparative and population genomics of mosquitoes; also it is the foundation for the development of novel genome-based approaches for vector control.

Chromosome and Genome Evolution in Culicinae Mosquitoes

Reem Abed Masri

## GENERAL AUDIENCE ABSTRACT

Mosquitoes are medically important insects because they vector a range of diseases that infect humans. The subfamily Culicinae is responsible for transmitting such diseases as Zika, dengue, and West Nile fevers, which have triggered fatal infections and epidemics in multiple parts of the world. Since 2010-2016, studies have reported exceeding levels of insecticide resistance that slows the disease elimination process. Novel transgenic techniques have a tremendous potential for more efficiently minimizing mosquito-borne diseases and transmission. Availability of high-quality genome assemblies for mosquitoes may help to better understand their population structure and to develop effective and safe vector-control approaches that we urgently need.

For the development of high-quality genome assemblies, we need to construct a physical genome map, that shows the physical locations of genes or other DNA sequences of interest along the chromosomes. For this reason, we developed a new gene-based approach for the physical mapping of the mosquito genomes. This method was then used for the development of a physical map for *Ae. albopictus.* This study resulted in the generation of the first physical map of the *Ae. albopictus* genome.

To understand population structure in Culicinae mosquitoes, we used mosquitoes from the *Culex pipiens* complex. Species in this complex transmit different arthropod-borne viruses or arboviruses. Notable is the West Nile Virus, which has triggered fatal infections and epidemics in Eastern and Central Europe, North America and is also known in Asia, Australia, Africa, and the Caribbean. We specifically focused on two subspecies in this complex, *Cx. pipiens pipiens* and *Cx. pipiens molestus* that are morphologically identical, but are different physiologically and behaviorally. Although they are spread globally in temperate regions, their population structure and taxonomic status remains unclear. The first goal of this study was to better understand the level of genomic differentiation of *Cx. p. pipiens* and *Cx. p. molestus* from different continents. We sequenced and compared the whole genomes of 40 individual mosquitoes from two locations in Eurasia and two in North America. Our study identified that *Cx. p. molestus* and *Cx. p. pipiens* represent different evolutionary units that are currently undergoing ecological speciation. The second goal was to study differences at the chromosome level between them. Using the Hi-C approach we detected presence of two different inversions in *Cx. p. pipiens* and *Cx. p. molestus*, which could potentially explain their different physiology and adaptation.

**Acknowledgements**

I would like to acknowledge everyone who played a role in my academic accomplishments.

I would like to thank my University, the Entomology Department, and our Department Head Dr. Tim Kring.

I would like to thank my advisor Dr. Maria Sharakhova for her continuous assistance throughout my Ph.D. years. Also, great thanks to my committee members, Dr. Igor Sharakhov, Dr, Jake Tu, Dr. Sally Paulson and Dr. Daniela Cimini for their support and guidance along the way.

I would like to give a huge thanks to Sharakhova and Sharakhov Lab members. Thanks to Varvara Lukyanchikova for helping me learn the Hi-C method, Il'ya Brusentsov and Andrey Yurchenko for assistance in bioinformatics, Megan Fritz, Natalya Khrabrova and Anuarbek Sibataev for their help with mosquito collections.

Finally, I must express my very profound gratitude to my parents and to my partner, Bilal Hamzeh for providing me with support and continuous encouragement throughout my years of study, research, and writing. This journey would not have been possible without them. Thank you.

**Attributions**

**Chapters 1 and 6** were written by the graduate candidate to provide the overall background and the summary for four manuscripts in this dissertation. A brief description of the authors' contribution is included as below:

**Chapter 2**

This chapter has been published in a peer-reviewed journal *Insects:*

Conceptualization, M.V. Sharakhova. and D.A. Karagodin; data curation, R.A. Masri and A. Sharma; funding acquisition, M.V. Sharakhova; investigation, D.A. Karagodin, R.A. Masri and A. Sharma; supervision, M.V. Sharakhova; writing—original draft, D.A. Karagodin, R.A. Masri and M.V. Sharakhova

**Chapter 3**

This chapter has been published in a peer-reviewed journal *Genome Biology*:

Conceptualization, resources, and whole-genome sequencing: M. Bonizzoni and J. Powell. Assembly and Hi-C databased scaffolding: S. Koren, J. Ghurye, A. Rhie, and A.M. Phillippy. Whole-genome assembly quality control and analyses: U. Palatini, E. Pischedda, M. Bonizzoni, F. Krsticevic, and P.A. Papathanos. Physical mapping: R. A Masri, A. Sharma, D.A. Karagodin, J. Jenrette, and M. Sharakova. Cytofluorimetric based genome size analyses: J.S. Johnston. Small RNA seq libraries and sequencing: M. Marconcini and A. Failloux. nrEVE annotation: U. Palatini and E. Pischedda. piRNA clusters and miRNA annotation: R. Halbach, P. Miesen, and R.P. Van Rij. Genome annotation: F. Thibaud-Nissen and P. Masterson. Immunity genes: F. Krsticevic and P.A. Papathanos. M locus: J.K. Biedler and Z. Tu. Population genomics: L.V. Cosme, J.E. Crawford, and A. Caccone. Developmental transcriptome: I. Antoshechkin, S.

Gamez, and O.S. Akbari. Figures: U. Palatini, R.A. Masri, M.V. Sharakova, R. Halbach, P. Miesen, R.P. Van Rij, J. Biedler, Z. Tu, F. Krstivevic, P.A. Papathanos, L.V. Cosme, J.E. Crawford, A. Caccone, S. Gamez, and O.S. Akbari. Manuscript writing team: M. Bonizzoni, U. Palatini, J. Powell, P.A. Papathanos, M.V. Sharakova, R. Halbach, P. Miesen, R.P. Van Rij, J. Tu, A. Caccone, S. Gamez, and O.S. Akbari.

**Chapter 4**

This chapter has been published in a peer-reviewed journal *Scientific Reports*:

M.V. Sharakhova and A.A. Yurchenko designed the experiments. R.A. Masri, M.L Fritz, A.K. Sibataev and N.V. Khrabrova performed mosquito collections and species identification. A.A. Yurchenko conducted statistical and bioinformatics analysis. M.L. Fritz provided mosquito colonies. M.V. Sharakhova, A.A. Yurchenko and R.A. Masri wrote the manuscript.

**Chapter 5**

R.A. Masri performed physical mapping, Hi-C experiments, chromosome measurements, writing of the first manuscript draft. V. Lukyanchikova to the Hi-C experiment. V. Lukyanchikova and I. Brusentsov contributed to Hi-C analysis. M.V. Sharakhova supervised the study design, the analyses, and the results' interpretation, and finalized the manuscript.

## Table of Contents

**List of Figures and Table**

## Introduction

Culicinae is a subfamily that comprises species ranging from obligate anthropophagic major vectors to opportunists, minor vectors, and even zoophilic non vectors of arboviruses [1]. Major vectors of this subspecies belong to *Aedes* and *Culex* genera. *Aedes* mosquitoes are dominant vectors of most arboviruses that infect humans and animals worldwide. Recent Zika virus outbreaks in South and Central America, as well as yellow fever virus outbreaks in Africa and Brazil, serve as examples of how easily zoonoses that seem to be inactive or stable can flare up or spread [2]. To provide an improved vector surveillance and control, an increasing use of novel omics technologies has been seen in the last decades. However, more sequence data and cytogenetic information is needed to build and develop effective vector-control strategies. The first mosquito genome sequenced in 2002 was that of the main African malaria vector, *Anopheles gambiae*, providing the first opportunity to study insect biology [3]. Comparative studies were limited to the fruit fly, *Drosophila melanogaster*, since it was the only other insect species with a sequenced genome at the time. Genomic efforts then focused on the Culicinae subfamily, which includes *Aedes aegypti*, the main vector for yellow and dengue fevers, and *Culex quinquefasciatus,* the vector of West Nile fever. However, more studies are needed on other major mosquito vectors. These studies should include, but should not be limited to, the development of high-quality genome assemblies and the study of chromosome rearrangements, sex genes, and 3D genome structure.

The aims of this dissertation are:

1) to develop an effective approach for physical mapping of repeat rich mosquito genomes; 2) to construct the first physical map of the *Ae. albopictus* genome;

3) to determine the genomic divergence between two members of *Culex pipiens* complex *Cx. pipiens pipiens* and *Cx. pipiens molestus* from different continents;

4) to investigate if chromosomal inversions differentiate *Cx. p. pipiens* and *Cx. p. molestus*.

## Chapter 1: Literature Review

### 1.1 Vector-borne Diseases Overview:

Vector-borne diseases represent a threat for the general health of humans and cattle worldwide. Vector-borne diseases are illnesses caused by parasites, viruses, and bacteria that are transmitted by vectors [4]. Arthropod vectors include mosquitoes, ticks, sandflies, and blackflies. Mosquito-borne diseases are those spread by the bite of an infected mosquito [5]. Diseases that are spread to people by mosquitoes include Zika, West Nile fevers, chikungunya, dengue fever, and malaria [6]. According to the World Health Organization (WHO) nearly 700,000 deaths occur every year because of mosquito-borne diseases. Also, significant outbreaks of dengue fever, chikungunya, yellow fever, and Zika fever have occurred in multiple countries, infecting people, claiming lives and overwhelming health systems since 2014 [4].

Mosquitoes of medical importance belong to the family Culicidae and are extensively distributed around the world [5]. This large family includes three subfamilies: Anophelinae, Culicinae, and Toxorhynchitinae [7]. Important mosquito vectors reside within two subfamilies, Anophelinae and Culicinae. The three dominant mosquito species groups are *Anopheles gambiae* complex, *Culex pipiens* complex, and *Aedes* genus, subgenus *Stegomyia* [8]. Ongoing efforts and interventions are being tested and implemented to decrease the toll from these diseases [9]. There is still, however a major gap in our understanding of the mechanisms that allow mosquitoes to thrive as vectors. To control mosquito-borne diseases, we need to broaden the boundaries of existing knowledge to include information that is currently lacking.

### 1.2 Mosquito arboviral transmission:

Arboviral diseases, caused by viruses, spread to people by the bite of an infected arthropod, primarily mosquitoes and ticks [10]. The most reemerging arboviruses this century (Zika, dengue, yellow fever, West Nile, and chikungunya viruses) are all transmitted in urban or periurban areas by *Aedes* mosquitoes of the subgenus *Stegomyia* and *Culex* complex species [10]. Zika virus (ZIKV) became the most recent mosquito transmitted virus that is threatening the global human population. On February 1, 2016, the outbreak was declared by WHO a Public Health Emergency of International Concern as evidence grew that ZIKV can cause birth defects as well as neurological problems [11].

In a sylvatic cycle, where there is a steady transmission between animal reservoirs, these viral pathogens are widely circulated. Some of these animal reservoirs include the nonhuman primates for dengue virus (DENV) and chikungunya virus (CHIKV) [12]. Sylvatic cycles usually occur in forested areas uninhabited by humans. However, because of deforestation, human exposure to these infected vectors occurs [13]. This factor then raises the incidence of spillover into immunologically naive human populations of viruses. Though humans are typically dead-end hosts for these arboviruses, there may be human-to-human dissemination via an urban cycle involving the transmission of mosquitoes [10, 12]. Similarly, infected mosquitoes could also spread these viruses to domestic animals such as horses, cattle, and pigs [14].

### 1.3 Mosquito Lifecycle:

Mosquitoes are holometabolous and go through 4 different stages in their life. The four stages are egg, larva, pupa, and adult [15]. Mosquito eggs are generally laid on the surface of water or moist soil/leaves. Culicinae mosquito species lay their eggs either in clusters called rafts (e.g., *Culex* genus) or separately (e.g., *Aedes* genus). Each raft holds approximately 250 mosquito

eggs, which can hatch in 1 to 5 days, depending on the temperature. Larvae, known as wiggles, hatch from mosquito eggs and they are aquatic throughout their larval stage. Mosquito larvae feed on algae, bacteria, and organic debris in the water [14]. Culicinae mosquitoes actively swim to deeper water to feed off the bottom and wiggle their way back up to the water surface to get oxygen using their siphon (tubular organ of the respiratory system).

Larvae go through 4 instars. After the $4^{th}$ instar, they emerge as a pupa. The pupa stage is different in appearance compared to the larva. The pupae are known as tumblers and look like a "fat comma" with ears. Instead of a siphon, pupae have breathing tubes known as trumpets. Mosquitoes in the pupal stage do not eat and this stage lasts 1 to 2 days.

Adult mosquitoes emerge from pupae. The adult mosquito has 3 major body regions: head, abdomen, and thorax. Only the adult female mosquitoes seek a host for a blood meal. The females often find host animals by smelling the carbon dioxide they exhale along with smelling body odors and detecting body heat. The females require the blood as a source of proteins for their developing eggs. Most, but not all species of mosquitoes require a blood meal for their egg development.

Behavioral studies have shown that mosquitoes detect and respond to several host cues that allow them to identify their hosts in the landscape. These clues include carbon dioxide ($CO_2$) emissions, visual appearance and color, local increases in temperature and humidity, and odorants released from skin and breath [16, 17]. Mosquitoes must integrate multiple sensory modalities as they transition from long-range to short-range detection of cues during host-seeking [18, 19]. $CO_2$ is used as a long range cue; it activates host-seeking behaviors, allowing mosquitoes to orient and locate a host from a great distance [20]. At short range, female mosquitoes use skin odorants, movement, and color to detect the host. Detection of host body

heat and humidity guide landing, and skin odors are thought to ultimately determine host acceptance [21].

### 1.4 Intervention Methods:

Vector control has been the main method for controlling vector-borne diseases both historically and today. Moreover, for some diseases, such as Zika, chikungunya, and West Nile fevers, vector control is currently the only method available to protect from spreading disease [22]. Vector control intends to limit the transmission of pathogens by reducing or eliminating human contact with the vector. For decades, chemical interventions were successful in controlling or eradicating mosquitoes worldwide [23] and to this day, they still makeup a big portion of mosquito control strategies. However, their limitations encouraged scientists to study and develop alternative control strategies to battle vector-borne diseases.

Conventional methods have been successful in many areas of the world, such as reducing malaria incidence in 1945 [23,24] especially after the discovery of dichlorodiphenyltrichloroethane (DDT). However, an increased public rejection of the application of DDT started because of its ecological impact and resulted in a DDT ban in the USA in 1972 [24]. Currently, insecticide use still plays a significant role in mosquito control programs involving the use of pyrethroid-treated bed nets, chemical or biological larvicides targeting immatures, and indoor/outdoor residual spraying [25]. The emergence of mosquitoes displaying resistance to insecticides prompted scientists to look for alternative control methods.

Alternative control methods, or biological methods, include Sterile Insect Technique (SIT), Release of Insects carrying a Dominant Lethal gene (RIDL), and release of *Wolbachia*-infected male mosquitoes [26].

SIT is a genetics-based method that has been successfully used to suppress agricultural pest insects since the 1950s. It includes the release of radiation-sterilized male insects into wild populations, where they seek out and mate with the females, giving rise to offspring that are not viable [27]. The substantial decrease in mating competition in nature due to radiation and major labor work is a drawback of this strategy that has so far hindered its implementation in mosquito control [26]. The RIDL is a strategy related to SIT but with a dominant lethal transgene inserted into the mosquito that can be programmed to become active at any developmental stage and does not require radiation exposure [28]. A small-scale field trial of RIDL in the Brazilian state of Bahia was conducted in 2010 and reported to have a considerable level of success in suppressing wild *A. aegypti* populations [29]. Also, Oxitec- a biotechnology company that develops genetically modified insects to aid in insect control- received approval from the Florida Keys Mosquito Control District to release "Oxitec's Friendly *Aedes aegypti* mosquitoes in 2021" using the RIDL approach.

The third approach to biological control is the release of *Wolbachia*-infected male mosquitoes. *Wolbachia* is an intracellular bacteria genus that mainly infects arthropods [30] and studies demonstrated that it renders *Ae. aegypti* mosquitoes resistant to infection with dengue [31] and chikungunya [32]. *Wolbachia* infection produces a phenotype termed cytoplasmic incompatibility, whereby uninfected females do not have viable offspring when they mate with infected males [33]. Due to this sterilizing effect on mosquitoes, *Wolbachia* is considered a promising vector control strategy [34].

When released in field studies, these mosquitoes are expected to outcompete wildtype males to mate with females. Most of these approaches rely on a particular male factor being identified [25]. Since male mosquitoes do not spread diseases they are thought of as an excellent source of

transgenic or environmentally friendly insect control techniques. The effectiveness of most of these strategies, however, depends on having a complete genetic code of these species so that scientists can easily and effectively manipulate and target certain gene/genes.

**1.5 Sequenced Culicinae Genomes:**

Mosquitoes are considered the deadliest animals to humans [35] and mosquito control expenses hold a large impact on the economy [36]. Even though there have been considerable advances in the field of mosquito genomics, a complete mapped genome for the most important species in Culicinae mosquitoes is still far from complete [37]. A complete annotated chromosome-level genome assembly will enable scientists to study regulatory genome and the epigenetic regulation of gene expression, specifically in the context of an infection [38, 39].

Only a few mosquito genomes are assembled at the chromosome-scale level. Most of the mosquitoes with chromosome-level assemblies belong to malaria vectors from the subfamily Anophelinae: *An. gambiae* [3, 40, 41], *An. albimanus* [42-44], *An. Atroparvus* [44-46], *An. stephensi* [43], *An. funestus* [43,48], *An. merus* [44]*, An. arabiensis* [47], *and An. coluzzii* [47]. Only one chromosome-scale assembly has been developed for Culicinae mosquito, the major vector of arboviruses, *Ae. aegypti* [48]. Two other important species from this subfamily, *Cx. quinquesfasciatus* and *Ae. albopictus* have incomplete assemblies. *Cx. quinquesfasciatus* [51] has a chromosome-level assembly that assigns 13% of the genome to chromosomes. *Ae. albopictus* has a linkage map developed with only 73 markers assigned to the genetic map [52].

A chromosome-level assembly provides a near-complete catalogue of the genomic content almost all of which is ordered and oriented along the chromosomal arms [37]. This is important to facilitate genome-wide evolutionary analyses to build informed hypotheses on putative gene function [37]. Also, greater species sampling and improved reference genomes serve to enhance

the resolution of gene evolutionary histories as well as the confidence in interpreting the genomes [49]. For example, examining conservation across the genus helped putatively select a region under strong functional constraint of the *An. gambiae* double sex gene for targeted disruption that produced sterile females [50]. Chromosome-level assembly can also aid in finding and studying quantitative trait loci (QTL), which are sequences of DNA associated with quantitative phenotypes, such as those related to olfactory or sensory traits [51]. Moreover, chromosome-level assembly has also enabled extensive study of detoxification gene families, mainly with respect to their roles in resistance to insecticides [52, 53]. These studies will facilitate the implementation of insecticide resistance management strategies for arboviral control programs. The number of complete genome assemblies in Anophelinae today outcompetes those in Culicinae; therefore, there is an urge to improve and complete genome assemblies within the Culicinae subfamily to be able to better understand and compare the physiology of those mosquito species compared to those in Anophelinae.

Currently, extensive efforts that combine data from complementary technologies offer new possibilities to substantially improve existing draft assemblies and their annotations, as shown in the new *Ae. aegypti* genome [48]. Physical mapping using fluorescence *in situ* hybridization (FISH) to chromosomally localize scaffolds on high-resolution cytogenetic maps [48, 54, 55] has played a crucial role in assembly improvements. For example, high resolution cytogenetic photomaps with FISH-mapping together with computational homology analysis have recently produced 98.2% and 89.6% chromosome-anchored *An. albimanus* and *An. atroparus* reference genomes, respectively [42, 45]. Moreover, Hi-C genomic proximity data promise fast and accurate ordering and orienting of fragmented draft assemblies into chromosome-length scaffolds, as achieved recently for the *Ae. aegypti* and *Cx. quinquefasciatus* genomes [56].

Furthermore, long-read sequencing data are proving useful for assembling repeat-rich highly heterozygous genomes as seen in the 2.25 Gbp *Ae. albopictus* C6/36 cell line assembly [57].

**1.6 Culicinae Cytogenetics:**

The genomes of Culicinae mosquitoes are enriched in repetitive elements and transposable elements compared to Anophelinae mosquitoes [58]. The abundance of repetitive elements in the genome results in low polytenization levels and poor spreading of polytene chromosomes [59].

Most of the classical cytogenetic on *Ae. aegypti* was performed on mitotic or meiotic chromosomes from larval brain or male testis. The culicinae karyotype includes 3 pairs of chromosomes that were initially named by Rai in 1963 as 1, 2 and 3 in order of increasing size [63]. Later, in *Ae. aegypti* chromosomes were renumbered according to the linkage map as follows: 1 as the shortest, chromosome 2 as the longest, and chromosome 3 as the mid-length [77]. Unlike in the anophelines, there are no sex chromosomes in culicine mosquitoes [60]. The sex determination alleles linked to chromosome 1 were described as Mm in males and mm in females[61].

Physical mapping in culicines is challenging because of the poor quality of the polytene chromosomes. Several attempts to create a cytogenetic map for instance using *Cx. quinquefasciatus* polytene chromosomes have been made using the Malpighian tubules [62] and salivary glands [63]. However, correspondence of arms and regions among these maps and the original drawn map published by L. Dennhofer [64] were uncertain. These problems occur because low levels of polyteny led to high frequency of ectopic contacts and poor spreading of polytene chromosomes during sample preparation. Mitotic chromosomes, in contrast to polytene chromosomes, do not form ectopic contacts and can be easily used for mapping DNA probes along the chromosomes. Using the chromosomes from imaginal discs of 4$^{th}$ instar larvae for

physical mapping, in Sharakhova et al. (2011) paved the way to successful mapping of probes in Culicinae mosquitoes and was an important step for further utilization of the genome [46, 48, 55, 58, 65-69].

### 1.7 Genome Mapping:

Genome maps provide the framework for sequencing projects since they indicate the position of genes and other recognizable features in the genome and hence enable the accuracy of an assembled DNA sequence to be checked. They are also used in identifying and mapping quantitative trait loci (QTL) that have been successfully used to examine several complex phenotypes in animals and vector species [70]. Genome mapping is also important for studying genome organization and evolution, and localization and isolation of genes of interest [71]. The localization of genes such as those responsible for sex-determination, vector competence, and blood-feeding can help in the field of genetic control for mosquito suppression and disease elimination. Additionally, the availability of the genome sequence will facilitate comparative studies across mosquito species. This will result in deeper knowledge and in species-to-species transfer of methodologies and experimental procedures.

#### 1.7.1 Importance of the Physical Map

As any other map, a genome map describes the relative locations of characteristics which are of concern or may serve as reference points. The characteristics found on a genome map are collectively referred to as markers and can include both genes and non-coding sequences.

Physical maps for vectors have traditionally been based on low-resolution cytogenetic approaches [72]. Physical maps display the physical distance between genes, can be constructed using cytogenetic maps [73], and use actual physical distances measured in bp (base pairs).

Particularly important are visually distinct regions, called light and dark bands, which give each of the chromosomes a unique appearance. Cytogenetic maps can be applicable to polytene chromosomes, also known as the giant chromosomes. Polytene chromosomes have great banding pattern and aid in visualizing structural rearrangements, such as inversions [74]. Cytogenetic maps or idiograms can also be constructed using mitotic chromosomes for species that do not have great polytene chromosomes, but banding patterns are not as clear as in polytenes [75].

The first cytogenetic map was developed by Bridges [76], and the first high-quality-drawn cytogenetic maps of chromosomes from salivary glands were developed by Coluzzi and coauthors for mosquitoes of the Anopheles gambiae complex [77, 78]. A drawback of drawn chromosomal maps is that they can be subjective. Advances later in photo techniques and microscopy made it possible for research to use photo based cytogenetic maps [41]. By using DNA sequenced probes rather than pattern of inheritance (in linkage map), physical maps tend to be more accurate [73]. To be able to directly visualize the localization of the markers, fluorescence *in situ* hybridization (FISH) is used for physical mapping. Fluorescently labelled probes are hybridized to metaphase or polytene chromosomes on a glass slide, and their position is then directly observed using a fluorescence microscope [79].

A variety of probes can be used such as PCR amplified fragments, plasmids, cosmids, and Bacterial Artificial Chromosomes [80]. The resulting images clearly reveal the location of the hybridized probe. The characteristic banding pattern of the chromosomes gives the absolute location of the probe in the genome. By using different combinations of fluorophores, two or more probes may be hybridized and detected simultaneously. The resulting multicolor FISH allows to accurately determine order and orientation of the mapped probes.

### 1.7.2 The Hi-C Approach:

The Hi-C method is a step forward in the field of mapping, in that it allows researchers to study the three-dimensional folding of the chromatin. Hi-C is a form of high-throughput sequencing based on chromatin conformation for the study of 3D interactions of the genome folding by calculating the frequency of interaction between loci [81]. The contact frequency between a pair of loci directly correlates with the one-dimensional distance between them. Therefore, Hi-C data can be used for genome scaffolding and generating chromosome scale scaffolds and high-quality genome assemblies [56, 82]. Using the Hi-C approach to improve genome assembly in combination with other tools has been successful in several mosquito species: *Cx. quinquefascitus* [56], *Ae. aegypti* [48, 56], A*n. albimanus* [43], *An. funestus* [83], *An. coluzzi* [47], and *An. arabiensis* [47].

The Hi-C approach can also help detect structural rearrangements in the genome [82, 84]. The technique requires extremely low coverage compared with other Next Generation Sequencing (NGS) techniques being used for this purpose and hence may be cost-effective. It also has the additional advantage of providing copy number information from the same data as shown in multiple studies [84-86]. Hi-C overcomes the problematic detection of rearrangements that is usually poorly mappable or unseen due to repetitive regions [84].

Ultimately, integrating Hi-C with other molecular techniques will provide high quality genome assemblies as well as prove the presence/absence of structural rearrangements that is important in understanding mosquito's physiology and speciation.

### 1.7.3 The Long-Read Sequencing Technologies:

Long-read sequencing is overcoming its limitations in accuracy and cost and becoming more advantageous compared to short-read sequencing [87, 88]. Short-read sequencing is accurate,

cost-effective, and widely used [89]. However, producing reads up to 600 bases makes it very hard to reconstruct original DNA sequence, which causes a lot of fragmentation in genome assemblies [87]. Long reads can therefore enhance *de novo* assembly, detection of variants, and fixing fragmented genome [56, 88].

The two long-read technologies commonly used for genome sequencing right now are: Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) sequencing and Oxford Nanopore Technologies' (ONT) nanopore sequencing, commercially released in 2011 and 2014, respectively [57]. The first long-read sequencing technology to achieve a widespread deployment is SMRT sequencing technology. One complication of SMRT sequencing is the high error rate of this process compared to short read sequencing that ranges between 11–14% [90]. However, this error rate can be lessened by repeated measurements of the sequence [88, 90]. SMRT has been used in many mosquito genome assemblies [46, 48, 55, 91].

The next successful single molecule technology to hit the market was ONT. The read length of ONT data is very similar to that of PacBio, with a maximum length up to a few hundred thousand base pairs [90, 92]. However, ONT reads have error rates higher than PacBio reads. One of ONT devices marketed is the MinION sequencer that is attracting significant interest in the genomics community due to its small size and low equipment cost, particularly for pathogen surveillance and clinical diagnostic applications or any field that can benefit from this sequencing platform's real-time nature [92]. Genome assemblies in the arthropod field using long-read sequencing are being published more and more frequently [93, 94]. Specifically, in the world of mosquito genomics, long-read sequencing technologies have increased the integrity of genome assemblies, decreased the level of genome fragmentation, and allowed chromosome-

level genome assembly. Some of these genome assemblies include *Ae. aegypti* [48], *Cx. quinquefasciatus* [51], *An. albimanus* [42], *An. coluzzi* [47], and *An. arabiensis* [47].

**Chapter 2: A Gene-Based Method for Cytogenetic Mapping of Repeat-Rich Mosquito Genomes**

**Authors and Affiliations:**

Reem A. Masri[1], Dmitriy A. Karagodin[2], Atashi Sharma[3] and Maria V. Sharakhova[1,2]*

[1] Department of Entomology and the Fralin Life Sciences Institute, Virginia Polytechnic Institute and State

University, Blacksburg, VA 24061, USA; reemm7@vt.edu

[2] Laboratory of Evolutionary Genomics of Insects, The Federal Research Center Institute of Cytology and

Genetics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia; karagodin@bionet.nsc.ru

[3] Department of Biochemistry and the Fralin Life Sciences Institute, Virginia Polytechnic Institute and State

University, Blacksburg, VA 24061, USA; atashi04@vt.edu

***Corresponding Author**: msharakh@vt.edu

**Citation**: Masri, R.A.; Karagodin, D.A.; Sharma, A.; Sharakhova, M.V. A Gene-Based Method for Cytogenetic Mapping of Repeat-Rich Mosquito Genomes. Insects 2021, 12, 138. https://doi.org/10.3390/insects 12020138

## 2.1 Abstract

Long-read sequencing technologies have opened up new avenues of research on the mosquito genome biology, enabling scientists to better understand the remarkable abilities of vectors for transmitting pathogens. Although new genome mapping technologies such as Hi-C scaffolding and optical mapping may significantly improve the quality of genomes, only cytogenetic mapping, with the help of fluorescence in situ hybridization (FISH), connects genomic scaffolds to a particular chromosome and chromosome band. This mapping approach is important for creating and validating chromosome-scale genome assemblies for mosquitoes with repeat-rich genomes, which can potentially be misassembled. In this study, we describe a new gene-based physical mapping approach that was optimized using the newly assembled Aedes albopictus genome, which is enriched with transposable elements. To avoid amplification of the repetitive DNA, 15 protein-coding gene transcripts were used for the probe design. Instead of using genomic DNA, complementary DNA was utilized as a template for development of the PCR-amplified probes for FISH. All probes were successfully amplified and mapped to specific chromosome bands. The genome-unique probes allowed to perform unambiguous mapping of genomic scaffolds to chromosome regions. The method described in detail here can be used for physical genome mapping in other insects.

**Keywords**: genome; physical mapping; DNA hybridization

## 2.2 Introduction

Mosquito genomics is rapidly developing as a discipline that can answer a large spectrum of biological questions. However, its major focus is assisting in the development of new strategies for mosquito control [95]. Mosquito genomics is employed to address new threats in the spread of infectious diseases such as malaria, dengue, Zika fever, chikungunya, and others [96, 97]. With the help of novel and affordable sequencing technologies such as long-read technologies from Pacific Biosciences (PacBio) [98] and Oxford Nanopore Technology (ONT) [99, 100], the number of mosquito genomes available through public databases is rapidly increasing. Nevertheless, a high number of mosquito genomes are still far from complete and are represented by multiple relatively short scaffolds with low N50 statistics; only a few mosquito genomes are assembled at the chromosome-scale level. Most of the mosquitoes with chromosome-level assemblies belong to malaria vectors from the subfamily Anophelinae: *Anopheles gambiae* [3, 101, 102], *An. albimanus* [42-44], *An. atroparvus* [44, 45, 103], *An. stephensi* [44, 103], *An. funestus* [83, 103], *An. merus* [44], *An. arabiensis* [104], and *An. coluzzii* [104]. Only one chromosome-scale assembly has been developed for Culicinae mosquitoes, the major vector of arboviruses, *Aedes aegypti* [105]. Recently, a significantly improved version of the Asian tiger mosquito *Ae. albopictus* genome became available [106]. This effort placed 57% of the genome to the chromosomes. However, chromosome-level genome assembly for this species is still missing.

There are two major components that affect the quality of the genome: heterozygosity and repetitive DNA sequences. The first problem, heterozygosity, can be avoided by using inbred homozygous individuals [105] or by sequencing a single mosquito genome [91]. The second

problem, repetitive DNA sequences, leads to algorithmically compressed sequences [107]. Repetitive DNA sequences are a significant impediment for genome quality in mosquitoes from the Culicinae subfamily [53, 108]. Unlike Anophelinae mosquito genomes, Culicinae genomes are extremely enriched by different classes of transposable elements which have a tendency to localize in gene introns and spread throughout the genome [109, 110]. This problem can be resolved through additional direct mapping of the genomic scaffolds to the chromosomes. Although new physical mapping technologies such as optical mapping [111] or Hi-C scaffolding [56, 112] could significantly increase contiguity of the genomic scaffolds, only cytogenetic mapping based on fluorescence in situ hybridization (FISH) provides physical evidence of the scaffold location in particular regions of the chromosomes [113]. This method allows one to relate DNA probes of a known sequence to a particular band on the chromosomes. Thus, further development of the cytogenetic physical method is crucial.

The goal of the current study is to simplify and further optimize the cytogenetic mapping technique for highly repetitive mosquito genomes. We propose a new approach based on PCR amplification of DNA probes using complementary DNA (cDNA) that does not include repetitive DNA sequences from gene introns. This method was used for the development of a physical map for Ae. albopictus and is based on the hybridization of fifty cDNA fragments or gene exons from twenty-four scaffolds to mitotic chromosomes from imaginal discs, generating the first physical map of the *Ae. albopictus* genome [106]. The genome of this mosquito is enriched with transposable elements, making its size between 1.190 and 1.275 Gb long across populations, which is around five times larger than the genome of the major malaria vector An. gambiae [3]. Here, we describe this method in all required details so that it can be applied towards physical genome mapping in other mosquitoes or in other insect species of interest.

## 2.3 Materials and Methods

In this study, we used the Foshan strain, which was utilized for sequencing of the first and the last versions of the Ae. albopictus genome [106, 114]. Mosquito eggs were hatched at 28 ◦C, and after several days, 2nd or 3rd instar larvae were transferred to 20 ◦C to obtain a high number of mitotic divisions in the imaginal discs of the 4th instar larvae.

At the time when the experiments were planned, the *Ae. albopictus* Aalbo_primary.1 genome assembly was still unavailable [91]. Therefore, we used a transcriptome from the C6/36 cell line of *Ae. albopictus* [57] to design target probes for FISH (Figure 2.1A). A total of 15 transcripts which represent protein-coding genes with unique locations in the newly assembled genome, with sizes bigger than 5000 nucleotides (nt), were selected (Table 2.1). Primers were developed using the primer3plus program [115] with the following criteria: optimal primer size of 23 nt, melting temperature (Tm of 58 ◦C with maximum Tm difference of 5 ◦C, and GC content of 45%. The melting temperatures of the oligonucleotides were determined using the New England Biolab's Tm calculator (NEB, New England Biolabs, Ipswich, MA, USA). To visually see the FISH signal on mitotic chromosomes using microscopy, the final sizes of the PCR products were at least 3800 nt.

For cDNA synthesis, RNA was first extracted from ten 24 h old gravid female mosquito ovaries (Figure 2.1B). Female mosquitoes were placed on ice for immobilization; then, the ovaries were dissected in distilled water using a binocular microscope and immediately placed in TRIzol reagent (Thermo Fisher Scientific Inc., Waltham, MA, USA). Extraction was then performed using the Direct-zol RNA Miniprep kit (Zymoresearch, Irvine, CA, USA) following the manufacturer's protocol. cDNA was synthesized using ~200 ng RNA and Superscript III

Reverse Transcriptase following the manufacturer's protocol (Thermo Fisher Scientific Inc., Waltham, MA, USA). cDNA synthesis was performed using either oligo(dT), random hexamer, or gene-specific primers. However, after multiple attempts and troubleshooting steps, cDNA synthesized using oligo(dT) was determined to be the only successful method to give the desired band on agarose gels. RNA residuals were digested following the manufacturer's protocol by adding 1 µL of RNase H to each tube and incubating for 20 min at 37 ◦C. For later use, cDNA was diluted in Fisher DNAse/RNAse free water (Thermo Fisher Scientific Inc., Waltham, MA, USA) to 100 ng/µL and stored at −20 ◦C.

To amplify DNA fragments, PCR was performed with the specific primers mentioned above (Table1) using NEB Q5 high-fidelity DNA polymerase following the manufacturer's protocol (New England Biolabs, Ipswich, MA, USA). Briefly, 10 µL of 5 Q5 reaction buffer, 1 µL 10mM dNTPs, 2.5 µL of 10 µM forward primer, 2.5 µL of 10 µM reverse primer, 800 ng of cDNA as a DNA template, 0.5 µL 2 u/µL Q5 high-fidelity DNA polymerase, 10 µLQ5 high GC enhancer, and nuclease-free water for a 50 µL reaction were used. PCR cycle conditions following the protocol were: initial denaturation, 98 ◦C for 10 min; denaturation at 98 ◦C for 1 min; annealing at 50–72 ◦C for 30 s (optimal annealing temperature was 5 ◦C degrees below our lowest primer Tm); and 72 ◦C for 30 s per 1000 nt (2 min 30 s for an average 5000 nt probe). The optimal cycle number was 32. However, this number can be manipulated from 25 to 35 to obtain an optimal gel band. The final extension was at 72 ◦C for 2 min then held at 4 ◦C. Briefly, 5 µL of PCR product was run on 1% agarose to check for the desired DNA fragment (Figure 2.1B), and then, DNA was purified from the PCR mixture using Promega Wizard SV gel and a PCR clean-up system protocol (Promega Corporation, Madison, WI, USA).

For FISH, slides were prepared from imaginal discs of 4th instar larvae of *Ae. albopictus*. Before dissection, larvae were placed on ice for several minutes to immobilize them. Then, larvae were transferred to a slide along with a drop of a cold, hypotonic solution of 0.5% sodium citrate (Fisher scientific, Waltham, MA, USA) and dissected under an Olympus SZ microscope (Olympus America, Inc., Melville, NY, USA). The dissection procedure has been thoroughly explained in the previously published protocol [32]. Slides were analyzed under an Olympus CX41 microscope (Olympus America, Inc., Melville, NY, USA) at 400 magnification (40 × objective and 10 × eyepieces) (Figure 2.1C).

Purified DNA was labeled with Cy3- or Cy5-dUTP (Enzo Life Sciences Inc., Farmingdale, NY, USA) by nick translation using published protocols (Figure 2.1B), but probe incubation was performed for 1 h 45 min instead of 2 h 30 min [32]. The duration time was decreased to obtain the optimal fragment sizes between 300 and 500 nt. The increased time resulted in over-digestion of the probe, which reduced the effectiveness of FISH. The hybridization buffer which was used to dissolve the probes contained 50% formamide, 10% dextran sulfate sodium (DSS), and 0.1% Tween 20 in 2 Saline-Sodium Citrate, SSC (Thermo Fisher Scientific, Waltham, MA, USA). In situ hybridization (Figure1C) was performed using a modified protocol [32]. Briefly, 1 µL of 10 mg/µL RNase A in 100 µL of 2 SSC solution was placed on each slide for 30 min. Slides were then placed in the following solutions: 2 SSC for 5 min at 37 ◦C, and pretreated in pepsin solution (50 mL sterile water, 50 µL 1M HCl, 0.1 mg/mL of pepsin) for 5 min at 37 ◦C. After that, slides were washed in 1 Phosphor Buffered Saline, PBS (Thermo Fisher Scientific, Waltham, MA, USA) and formaldehyde fixation solution (1.5 mL 37% formaldehyde to 50 mL 1 PBS) for 10 min each at room temperature. They were then rinsed in 1 PBS for 5 min and dehydrated using an ethanol series (70%, 80%, and 100%) for 5 min each and dried. Then, 250

ng of probe in hybridization buffer was added to the slide, covered with a 22 22-mm glass coverslip, and sealed with Elmer's rubber cement. The slides were then placed on a Thermobrite (Abbott Molecular Inc., Chicago, IL, USA) programmer at 73 ◦C for 5 min, followed by 37 ◦C overnight. The next day, slides were washed using the following solutions: in 1 SSC, prewarmed at 60 ◦C, for 5 min and, at 37 ◦C, 4 SSC/NP40 (40 mL of water, 10 mL of 20 SSC, and 50 µL of NP40) for 10 min. After that, slides were rinsed in 1 PBS and stained in 50 µL of Oxazol Yellow, YOYO-1 stain (Thermo Fisher Scientific, Waltham, MA, USA) at a 1:1000 dilution in 1 PBS under parafilm in a dark box for 15 min. Slides were then rinsed in 1 PBS for 3 min and a drop of prolong gold antifade (Invitrogen Corporation, Carlsbad, CA, USA) was added to each slide. Slides were covered with glass coverslips and analyzed using a Zeiss LSM 880 Confocal Microscope (Carl Zeiss Microimaging, Inc., Thornwood, NY, USA) at 1000× magnification (100× objective and 10× eyepieces) (Figure 2.1C).

**2.4 Results**

In this study, we describe in detail a new physical approach for mapping repeat-rich mosquito genomes to their mitotic chromosomes. This method was used, for the first time, to physically map the *Ae. albopictus* genome [106]. Here, we describe this technique with all the methodological details required for further utilization of this method. Figure1 summarizes the entire procedure. For the probe design, we utilized transcripts from the *Ae. albopictus* cell line genome (Figure1A) [57]. Based on our genome analyses, these transcripts have unique locations in the genome and, thus, can be used for further physical genome mapping (Table 2.1). To simplify primer design, the original sizes of the transcripts varied from 5012 to 10,567 bases. The primers were designed to obtain a final product of the DNA probe between 3800 and 5000 nt. The size of the DNA probe was chosen so as to obtain a clear signal on small mitotic

chromosomes. Mosquito ovaries were utilized to obtain RNA for further development of a large amount of cDNA (Figure 2.1B). cDNA was produced from mRNA that mostly represents gene exons and does not contain gene introns, which are enriched with different classes of transposable elements in repeat-rich Culicinae genomes [109, 116]. Repetitive DNA from gene introns produces multiple unspecific signals on the chromosomes that make it impossible to detect the correct position of a gene probe. After the probes were designed, we used standard procedures [113] to obtain chromosome preparations from 4th instar imaginal discs of *Ae. albopictus* (Figure 2.1C) and proceeded with probe labeling and FISH (Figures 2.1C and 2.2). Similar to other mosquitoes, the Ae. albopictus karyotype includes three pairs of chromosomes. In correspondence with Ae. aegypti chromosomes [75], the shortest chromosome was described as chromosome 1, the longest chromosome was identified as chromosome 2, and the mid-length chromo- some was designated as chromosome 3 [106]. Unlike the anophelines, the sex-determining chromosomes in all culicine mosquitoes, including *Ae. albopictus*, are homomorphic [117]. For the physical mapping of the DNA probes (Figures 2.1C and 2.3), we utilized previously developed chromosome idiograms [106]. Chromosomes of Ae. albopictus can easily be distinguished using chromosome and arm length differences [106]. Furthermore, *Ae. albopictus* chromosomes have a clear banding pattern that aids in physical mapping. In this study, 15 PCR-amplified probes and rDNA were mapped to the bands on an idiogram determined by FISH (Table1; Figures 2.2and 2.3). An 18 S ribosomal DNA probe that hybridized to the secondary constriction on 1q22 was used as an additional marker for chromosome 1 (Figures 2.1C and 2.3). The positions of the mapped transcripts were consistent with their positions in the Ae. aegypti genome, supporting the accuracy of the physical mapping.

## 2.5 Discussion

Cytogenetic genome mapping could utilize different sources of DNA probes and chromosomes, depending on the size of the genomes and the quality of the chromosomes. This study focuses on the further development of cytogenetic physical mapping for ap- plication in highly repetitive genome research in mosquitoes such as *Ae. albopictus*. A detailed comparison of the new gene-based method and the conventional Bacterial Artificial Chromosome (BAC)-based approach for physical genome mapping is summarized in Table 2.2. Although BAC clones have been successfully used for physical genome mapping for *An. gambiae* [3], *Ae. aegypti* [105, 109], and *Cx. quinquefasciatus* [65], utilizing BAC clones for highly repetitive genomes is challenging. First, BAC library development and BAC-end sequencing are expensive and unproductive in repeat-rich genomes because a large portion of the BAC-ends are represented by repeats located in multiple places in the genome. Second, blocking unspecific BAC clone hybridization to chromosomes by unlabeled repeats, so-called Cot-1 DNA, is also an inefficient and extremely time-consuming process as it requires extraction of large amounts of genomic DNA. Utilizing cDNA as a template for the probe development, which we proposed in this study, allows for specific hybridization of the probes to the chromosomes without additional DNA sequencing and repeat blocking during FISH. Although both methods are labor-intensive, the gene-based method allows us to avoid long procedures such as extraction of multiple BAC clones. Using a PCR-based approach simplifies the probe design because it requires only a primer design and DNA amplification. The duration of the entire FISH procedure is reduced since a separate probe denaturation step is not required for the PCR-amplified probe.

Another challenge associated with physical mapping in mosquitoes with highly repetitive genomes is the poor quality of their polytene chromosomes. Thus, mitotic chromosomes became

a major resource for cytogenetic mapping in Culicinae mosquitoes [65, 117]. However, using mitotic chromosomes requires large DNA probes for FISH. For example,

~500 nt probes, which are clearly visible in FISH on large polytene chromosomes in malaria mosquito chromosomes [42, 45], are not visible on small mitotic chromosomes. Thus, in our study, only probes with sizes larger than 3800 nt were utilized for FISH. Although an alternative method based on amplification of the gene exons for the FISH probes was also used for physical mapping of the Ae. albopictus genome [106], this method is not efficient for Culicinae mosquitoes because most of the exons in these mosquito genomes are smaller than 38 nt and would be invisible on mitotic chromosomes after FISH.

In addition to the cytogenetic approaches that we describe here, anchoring the genome assembly to the chromosomes can be performed by using different methods such as classical genetic linkage mapping [118-120] or by using modern technologies such as optical mapping [111] and Hi-C scaffolding [56, 112]. Each of these mapping methods has its own advantages and limitations. For example, genetic linkage maps, which are based on recombination frequencies, offer the opportunity to connect genomic assemblies with epidemiologically important traits in mosquitoes, including their ability to transmit pathogens [58, 105]. However, due to low recombination rates around the centromeres, genetic linkage maps have very low resolution around the centromeres [118, 120]. New approaches such as op- tical mapping based on high-resolution restriction maps from single, stained molecules of DNA [111] or the Hi-C method, which uses the frequency of chromosome contacts to estimate the distance between DNA fragments [112], are highly efficient for linear ordering of genomic scaffolds at the molecular level. However, both methods are useless in connecting the genome assembly with particular chromosomes or chromosome structures such as telomeres, centromeres, and chromosome

bands. Overall, we think that the best strategy for the development of high-quality genomes is to use a combination of different genome mapping techniques in one study. For example, one of the best chromosome-scale assemblies developed for mosquitoes, the *Ae. aegypti* genome, is based on a combination of modern technologies, Hi-C and optical mapping, with classical chromosome mapping technologies and linkage mapping [105].

### 2.6 Conclusions

High-resolution cytogenetic maps are fundamental for anchoring genome sequences onto chromosomes. Using complementary DNA as a probe template for FISH, we developed a new approach for the construction of physical maps and described this method in detail using the newly assembled *Ae. albopictus* genome as an example. This method is especially beneficial for the physical mapping of the highly repetitive Culicinae mosquito genome but can also be used for development of physical genome maps for other mosquitoes or insects to better aid in their genome assembly.

**Table 2. 1 Transcript locations in the Aedes albopictus genome assembly and chromosome map.** Probes used for fluorescence in situ hybridization (FISH) are shown by transcript and gene IDs. Probe locations are indicated by scaffold number, chromosome band in the genome, or in the chromosomes.

| Transcript ID | Scaffold ID/ Number | Chromosome Band | Forward (F) and Reverse (R) Primers |
|---|---|---|---|
| 18 S rDNA | | 1q22 | F:ATGCAAAATGCAGGAACCTC<br>R:GGTAATAGCAGCTGGGCTTG |
| XM_019696654 | NW_021838798.1/ 6 | 1p25 | F:TCGTTCGTGTAGATAAAGTCCAG<br>R:ATGGTTATGAGGTTCCAACAACT |
| XM_019681895 | NW_021838576.1/ 4 | 2p12 | F:GAAGTAACGGGCTCAGTTCTGGTTTC<br>R:CTTCGAGTAGTTGGACCAGTTCGAGA |
| XM_019677471 | NW_021838576.1/ 4 | 2p25 | F:AGCTCAACCAAAGGAAGGATTTA<br>R:TGATTGTTCACCTTGTTTTCCAC |
| XM_019673127 | NW_021838576.1/ 4 | 2p25 | F:AAGTGTGCGGAAAATTGTTTACA<br>R:TAATGGATTGAAGCTGCTTTTCG |
| XM_019675272 | NW_021838153.1/ 2 | 2p32 | F:TCCCTCTTTTATGAACAGCTGTT<br>R:ACAAAACATTCATGCAGTTGTCA |
| XM_019703202 | NW_021838687.1/ 5 | 2q34 | F:AAACGACAAGAAATGTTCTGCAA<br>R:TTGTGCCGCATTATCATTCATTT |
| XM_019682530 | NW_021838153.1/ 2 | 2q44 | F:AACAGAGCGGTATCTACAAAGAG<br>R:GTAGAACACGAAGGCATTAGGTA |
| XM_019675517 | NW_021838832.1/ 63 | 3p13 | F:ACTTCGGTTATGGGTAAGGTTTT<br>R:CAAAGACATGGGATTTTCTCGTC |
| XM_019697550 | NW_021838154.1/ 20 | 3p14 | F:GGAAGTTTTGTGTCGAAGGAAAA<br>R:GTCGTCCAGATTGTACAGATCTT |
| XM_019682463 | NW_021837045.1/ 1 | 3p32 | F:GAGACCAACGCAGAGTACGTCTTCAC<br>R:CGCATAGGCTCTGATGAACTTAGTCG |
| XR_002127435 | NW_017857621.1/ 18 | 3q12 | F:CCATTAAAAACGCCATCTAGCAA<br>R:TATGAGTGTAGTGTGCTAGCAAG |
| XM_019682067 | NW_021838465.1/ 3 | 3q23 | F:AGGTACCGTACAAAAGAAGTGAA<br>R:ACGGAACTAAGAAACAAAGTCCT |
| XM_019699815 | NW_021837489.1/ 14 | 3q23 | F:TTCGGTGGAAAAATCGTTTTGAA<br>R:CGAGTTTCTCTATTGCCACTTTG |
| XM_019701500 | NW_021838465.1/ 3 | 3q43 | F:CGATGGACTGTTCTTCCAATCTA<br>R:TTTGATTTGTGTTTGTCCCAGAC |
| XM_019701398 | NW_021838465.1/ 3 | 3q44 | F:TCCCGTTACTTCTACGAAATGAG<br>R:CCATCTTCTGGTTTGCATAACAG |

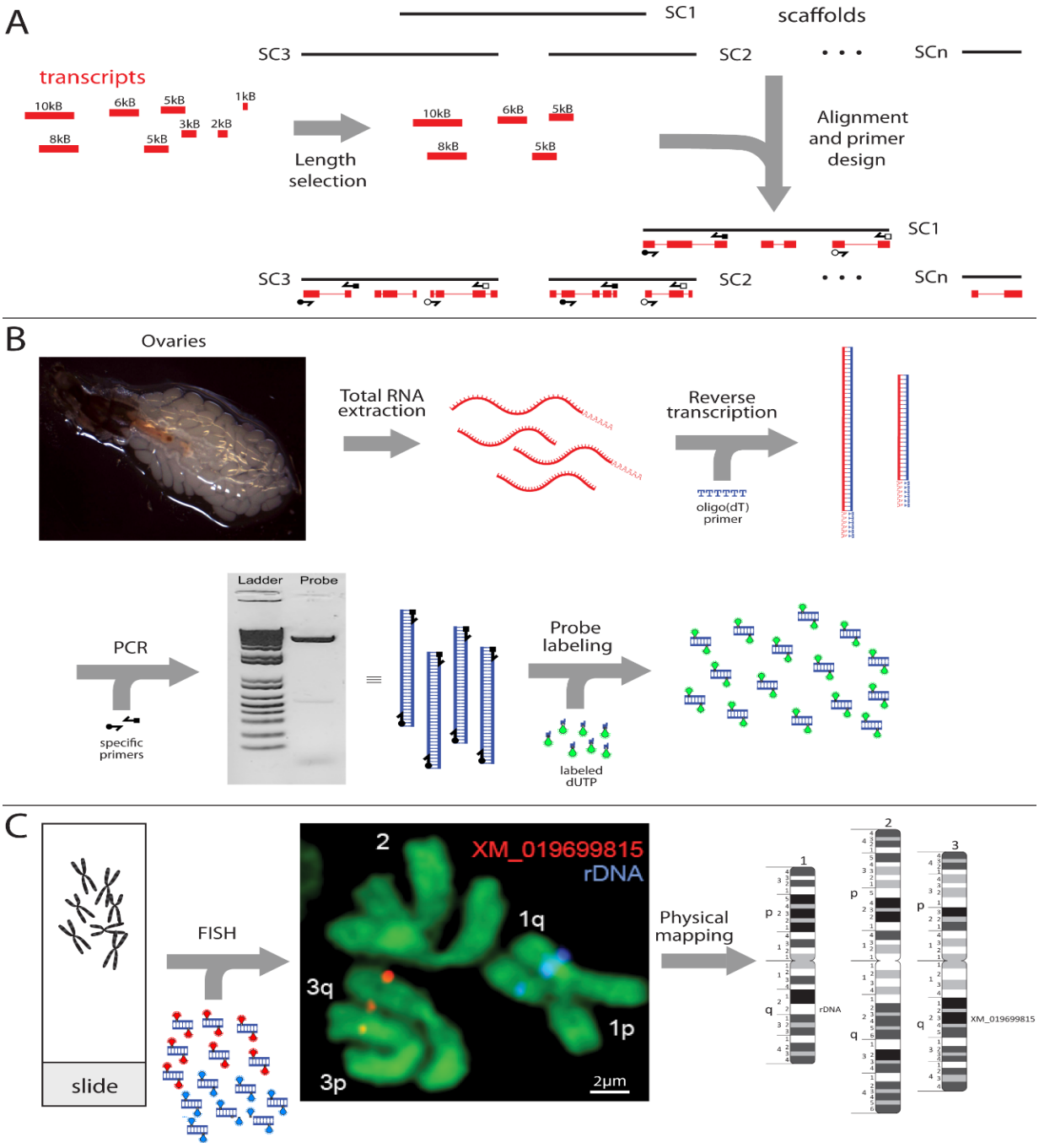


**Figure 2. 2 Genome mapping approach based on complementary DNA (cDNA).** Procedure includes: (A) primer design using transcripts from the Ae. albopictus genome; (B) extraction of RNA from female ovaries, obtaining complementary DNA based on RNA, amplifying the probe by PCR and probe labeling; (C) obtaining the chromosome preparations from mosquito imaginal discs, fluorescent in situ hybridization (FISH), and physical mapping of the probes onto chromosome idiograms. Chromosomes and chromosome arms are indicated by the numbers 1, 2, and 3 and the letters p (petite) and q (long), respectively. A transcript number XM_019699815 is indicated by the same color as signal on FISH image. rDNA stands for ribosomal DNA.

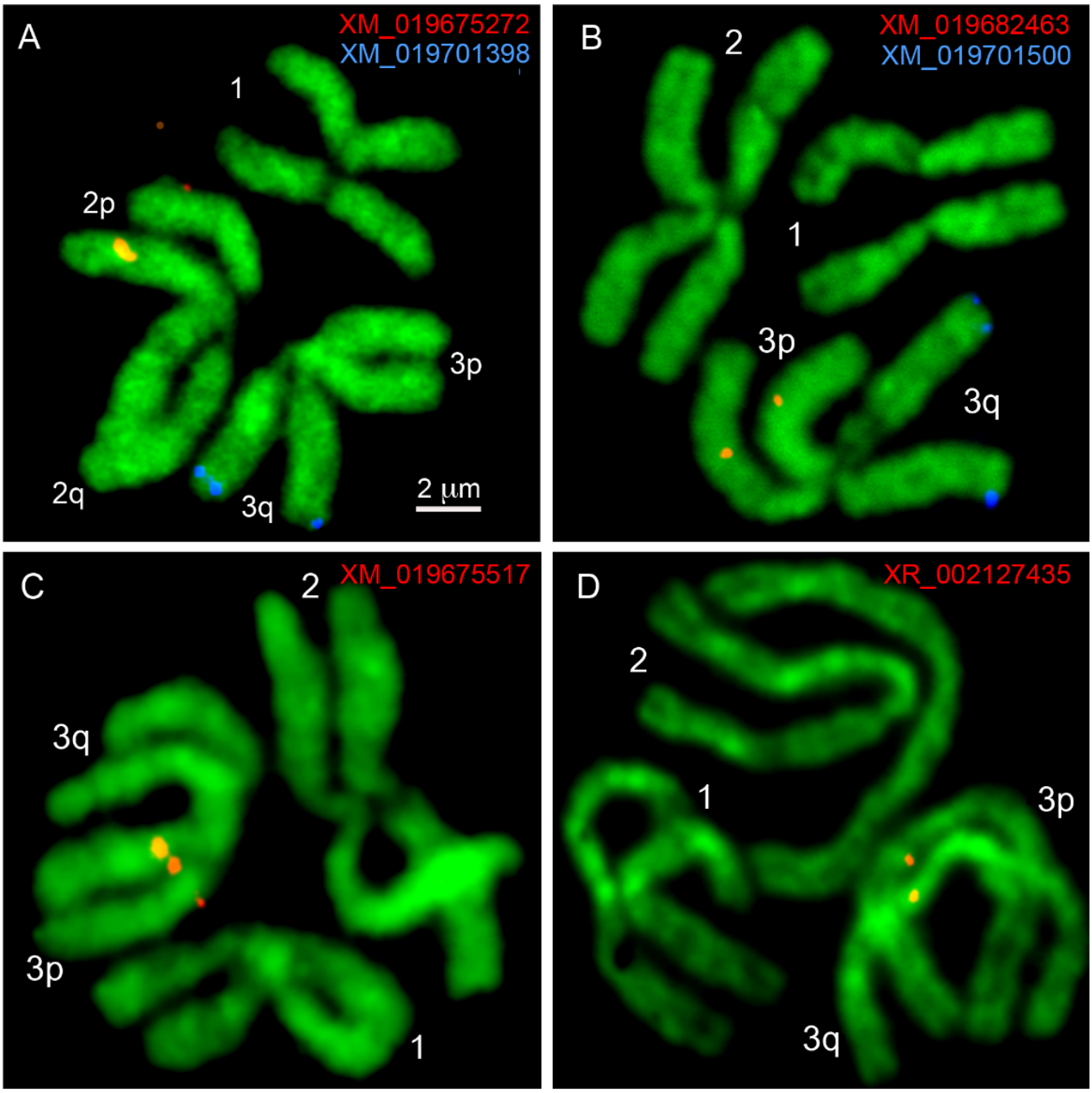


**Figure 2. 3 Examples of FISH in mitotic chromosomes of *Ae. albopictus***. The locations of transcripts XM_019675272 and XM_019701398 (A); XM_019682463 and XM_019701500 (B); XM_019675517 (C); and XR_002127435 (D) are shown. Chromosomes and chromosome arms are indicated by the numbers 1, 2, and 3 and the letters p and q, respectively. Transcripts numbers on the figure are shown with the same color as the probe signals on the chromosomes.

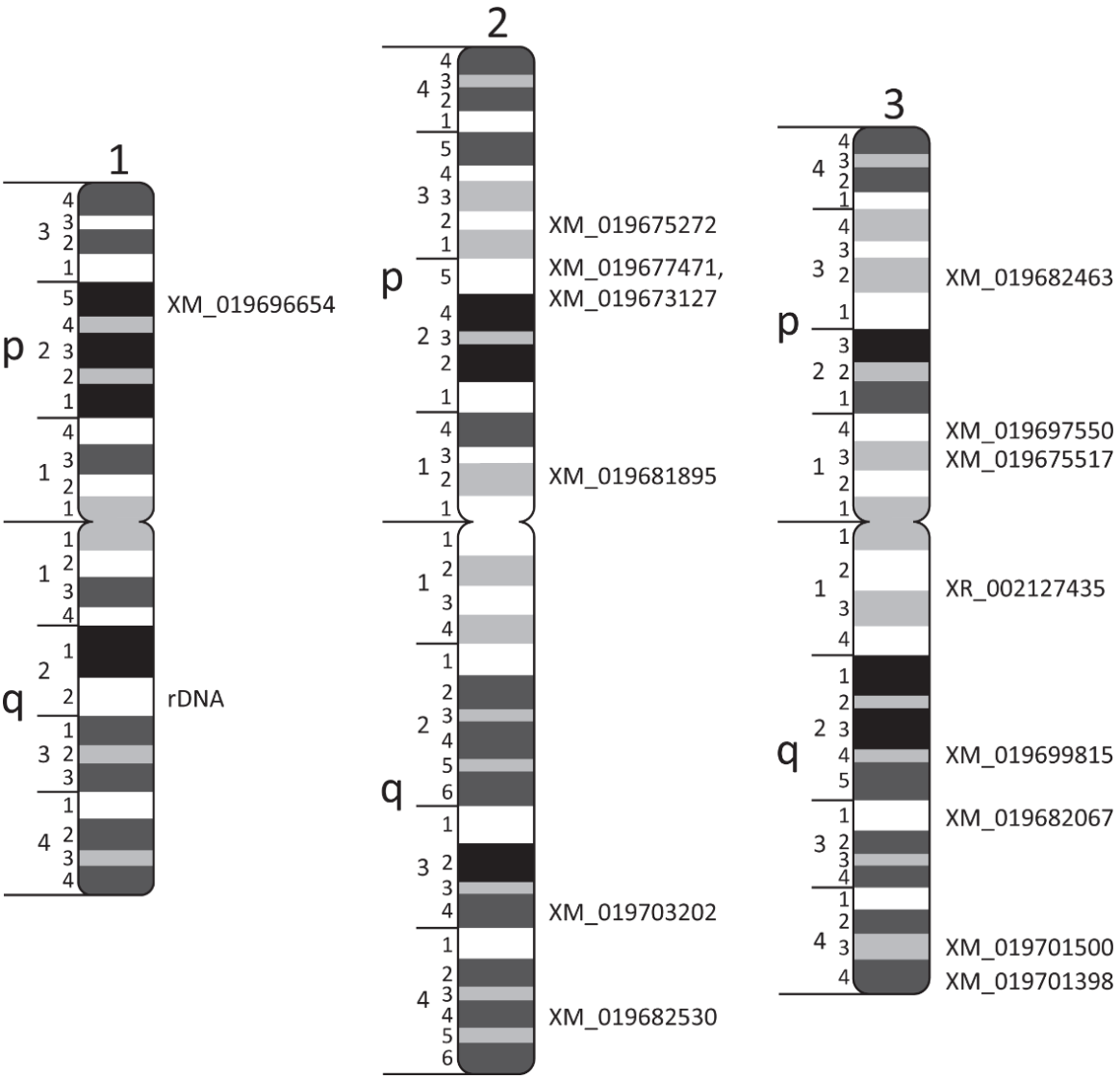


**Figure 2. 4 The locations of 15 DNA probes in the chromosome idiograms of *Ae. albopictus*.** Chromosomes and chromo-some arms are indicated by the numbers 1, 2, and 3 and the letters p and q, respectively. Chromosome divisions and subdivisions are shown on the left side of the idiograms. Transcripts are indicated on the right side of the idiograms; rDNA stands for ribosomal locus.

**Table 2. 2** Comparison of the Bacterial Artificial Chromosome (BAC)-based approach and the gene-based method for physical genome mapping.

| Comparison Parameters | BAC-Based Genome Mapping | Gene-Based Genome Mapping |
|---|---|---|
| BAC library development and BAC-end sequencing | Required | Not required |
| Primer design | Not required | Required |
| Blocking of unspecific hybridization with $C_0t_1$ DNA | Required | Not required |
| Multiple FISH signals are produced in the chromosomes | Often | Never |
| Additional DNA denaturation step in 70% formamide at 72 °C | Required | Not required |
| Expensive | Yes | No |

## Chapter 3: A Physical Map of the *Aedes albopictus* Genome (a modified chapter based on the below publication)

## Improved reference genome of the arboviral vector *Aedes albopictus*

**Authors and Affiliations:**

Umberto Palatini [#1], Reem A Masri [#2], Luciano V Cosme [3], Sergey Koren [4], Françoise Thibaud-Nissen [5], James K Biedler [2], Flavia Krsticevic [6], J Spencer Johnston [7], Rebecca Halbach [8], Jacob E Crawford [9], Igor Antoshechkin [10], Anna-Bella Failloux [11], Elisa Pischedda [1], Michele Marconcini [1], Jay Ghurye [4], Arang Rhie [4], Atashi Sharma [2], Dmitry A Karagodin [12], Jeremy Jenrette [2], Stephanie Gamez [13], Pascal Miesen [8], Patrick Masterson [5], Adalgisa Caccone [3], Maria V Sharakhova [2 12 14], Zhijian Tu [2], Philippos A Papathanos [6], Ronald P Van Rij [8], Omar S Akbari [13], Jeffrey Powell [3], Adam M Phillippy [4], Mariangela Bonizzoni [15]

# the Authors contributed equeally to this work

[1]Department of Biology and Biotechnology, University of Pavia, Pavia, 27100, Italy.

[2]Department of Entomology and the Fralin Life Science Institute, Virginia Polytechnic and State University, Blacksburg, VA, 24061, USA.

[3]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, 06511-8934, USA.

[4]Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, 20892-2152, MD, USA.

[5]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, 20894, MD, USA.

[6]Department of Entomology, Robert H Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, 7610001, Rehovot, Israel.

[7]Department of Entomology, Texas A&M University, College Station, TX, 77843, USA.

[8]Department of Medical Microbiology, Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands.

[9]Verily Life Sciences, South San Francisco, 94080, CA, USA.

[10]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, 91125, USA.

[11]Department of Virology, Arbovirus and Insect Vectors Units, Institut Pasteur, Paris, 75015, France.

[12]Laboratory of Evolutionary Genomics of Insects, The Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 630090, Russia.

[13]Division of Biological Sciences, University of California, San Diego, La Jolla, CA, 92093-0349, USA.

[14]Laboratory of Ecology, Genetics and Environment Protection, Tomsk State University, Tomsk, 634041, Russia.

[15]Department of Biology and Biotechnology, University of Pavia, Pavia, 27100, Italy.

**Corresponding Author**: m.bonizzzoni@unipv.it

**Keywords**

*Ae. albopictus*, genome, miRNAs, piRNA clusters, viral integrations, immunity, sex locus, population differentiation, developmental transcriptome

**Citation:** Palatini U, Masri RA, Cosme LV, Koren S, Thibaud-Nissen F, Biedler JK, Krsticevic F, Johnston JS, Halbach R, Crawford JE, Antoshechkin I, Failloux AB, Pischedda E, Marconcini M, Ghurye J, Rhie A, Sharma A, Karagodin DA, Jenrette J, Gamez S, Miesen P, Masterson P, Caccone A, Sharakhova MV, Tu Z, Papathanos PA, Van Rij RP, Akbari OS, Powell J, Phillippy AM, Bonizzoni M. Improved reference genome of the arboviral vector Aedes albopictus.

Genome Biol. 2020 Aug 26;21(1):215. doi: 10.1186/s13059-020-02141-w. PMID: 32847630; PMCID: PMC7448346.

### 3.1 Abstract

The Asian tiger mosquito *Aedes albopictus* is globally expanding and has become the main vector for human arboviruses in Europe. With limited antiviral drugs and vaccine available, vector control is the primary approach to prevent mosquito-borne diseases. A reliable and accurate DNA sequence of the *Ae. albopictus* genome is essential to develop new approaches that involve genetic manipulation of mosquitoes. Here, the first physical genome map for this species with 75% of the assembled genome anchored to the chromosomes was built. The AalbF2 genome assembly represents the most up-to-date collective knowledge of the *Ae. albopictus* genome. These resources represent a foundation to improve understanding of the adaptation potential and the epidemiological relevance of this species and foster the development of innovative control measures.

## 3.2 Background

Climate change, urbanization, and increased international mobility are predicted to further increase the spreading of the highly invasive mosquito *Aedes albopictus* and severely exacerbate the risk and burden of *Aedes*-transmitted human pathogens, in primis dengue, Zika, and chikungunya viruses, but also the veterinary-relevant parasite *Dirofilaria immitis* [10, 121]. As a consequence, nearly a billion people could face their first exposure to arboviral transmission within the next century, especially in subtropical and temperate regions of the world, including Europe [121].

The initial genome assembly of *Ae. albopictus* (AaloF1) from the Chinese Foshan strain represented a fundamental achievement for the genetic characterization of this mosquito [114]. From this analysis, based solely on the assembly of short DNA sequence reads, the genome of *Ae. albopictus* appears to be the largest mosquito genome sequenced to date (1.9 Gb). However, due to very high levels of repetitive DNA and reliance on short-read sequencing, AaloF1 remains highly fragmented with more than 150,000 scaffolds, limiting its utility.

## 3.3 Results

To foster continuity, the previously sequenced Foshan strain [114] was chosen to use for further genome study. PacBio sequencing and Hi-C scaffolding were performed for the development of a new genome assembly based on this strain. A total of 82 Gb of PacBio and 135 Gb of Hi-C sequence were generated. The final AalbF2 genome size of 2.54 Gb was assembled in 2,197 scaffolds with N50 of 55.7 Mb and the genome was completely reannotated.

A physical genome map of the *Ae. albopictus* genome was developed using *in situ* hybridization on mitotic chromosomes covering 57% of the genome assembly by targeting twenty of the

largest genomic scaffolds and three minor scaffolds (Fig. 3.1b). The mitotic chromosome complement of *Ae. albopictus* consists of three pairs of chromosomes [122]. The smallest, largest and intermediate chromosomes are numbered as 1, 2 and 3, respectively [91]. A total of 4, 9, and 10 scaffolds were assigned to chromosomes 1, 2, and 3, respectively. Positions of the transcript from scaffolds 15, 48, and 55 hybridized to places already covered with other large scaffolds. The positions of all tested transcripts were consistent with their positions in the *Ae. aegypti* genome, which is assembled into chromosome-size scaffolds, providing an independent confirmation of the accuracy of the *in-situ* hybridization results [105]. Based on probe mapping to the *Ae. aegypti* genome and homology between the *Ae. aegypti* and the *Ae. albopictus* chromosomes, we bioinformatically assigned the 58 longest scaffolds covering 75% of the genome to *Ae. albopictus* chromosomes.

Cytogenetic comparison (Table 3.1) between *Ae. albopictus* and *Ae. aegypti* demonstrated that the total chromosome length is 4.9 μm or 16.4% longer in *Ae. albopictus* ($P < 0.0001$), which suggests a slightly larger genome size in this species, as also suggested by cytofluorimetry (Fig. 3.1A). Chromosome proportions, such as relative chromosome and arm lengths, between the two species were also different. In *Ae. albopictus*, chromosome 1 was shorter but chromosome 2 was longer relative to *Ae. aegypti*. Besides positioning and orienting the largest scaffolds, we physically mapped the 18S rDNA and other genomic features (e.g., viral integrations and representative immunity genes) described below (Fig. 3.1c, d). The 18S rDNA mapped in the region of the secondary constriction in region 1q22. The intensity of the signal significantly varied among chromosomes from individual mosquitoes suggesting variations in numbers of ribosomal genes.

### 3.3.1 Mapping endogenous viral elements

The genome of *Ae. albopictus* harbors hundreds of integrated sequences from nonretroviral RNA viruses, called nonretroviral endogenous viral elements (nrEVE) or nonretroviral integrated RNA virus sequences (NIRVS). The largest viral integration (Canu-Flavi19) in AalbF2 reached 6593 bp. This viral integration was mapped by *in situ* hybridization to chromosome 2q close to the telomere, confirming it is integrated within the genome (Fig. 3.1b). Signals were also found in the centromeres of all three chromosomes, probably because these regions contain nrEVEs with sequence similarity to Canu-Flavi19 (Fig. 3.1b).

### 3.3.2 Mapping PPO Genes

The capacity of mosquitoes to acquire, disseminate, and transmit viruses (i.e., vector competence) is a complex phenotype which is controlled by genetic elements of both the vector and the pathogen, as well as environmental variables [123]. Understanding the complex relationship between vectors and pathogens requires understanding innate immunity in mosquitoes. A detailed analysis of the immune repertoire of *Ae. albopictus* revealed extensive expansions in 16 of the 27 functional groups relative to *Ae. aegypti*. The most extreme expansion event involves the prophenoloxidase (PPO) gene family, which in *Ae. aegypti* includes six tandemly arrayed genes, namely PPO4, PPO8, PPO7, PPO5, PPO1, and PPO2. The entire cluster of six genes has been locally duplicated twice in *Ae. albopictus*, resulting in 18 genes (Fig. 3.1d). The triplication of the clusters was confirmed using *in situ* hybridization (Fig. 3.1b). PPOs are enzymes that catalyze the production of melanin in response to infection [124]. Expansion of PPO genes is not common in insects [125], but in mosquitoes, the number of genes is higher than other insects. The high conservation of the PPO organization and order in the array in both *Ae. aegypti* and *Ae. albopictus* strongly suggests that these duplications are ancient events. Future

studies focusing on dissecting the functional importance of specific family expansions in *Ae. albopictus* may determine their significance for its biology including vector competence and ecological adaptation.

### 3.4 Discussion

This study developed the new genome assembly for *Ae. albopictus* AalbF2 which represent a significant improvement compared to AaloF1. For example, the N50 scaffold size increased from 201 kb in AaloF1 to 55.7 Mb in AalbF2. The genome reannotation led to a significant 8.3% increase in the percentage of complete, single-copy BUSCOs with respect to AaloF1. The first physical genome map of *Ae. albopictus* was developed, which consists of fifty DNA markers that cover the largest genomic scaffolds, rDNA, PPO gene clusters, and the largest viral integration in the genome. Overall, FISH data were consistent with the assembled genome, confirming its large-scale structural accuracy. Combining *in situ* and bioinformatic approaches, 58 scaffolds were anchored to the *Ae. albopictus* chromosomes, whose length sum makes 75% of the genome. This map represents a significant improvement of the genome map as compared to previously developed linkage map [126]. A linkage map was developed in 2011 and included 73 markers (1 marker/2.9 cM) with 9, 16, and 10 markers to the 3 linkage groups, respectively [126, 127]. Unlike linkage mapping that indicates the relative order of genes on a chromosome, a physical DNA map labels the absolute position of genes.

The map that we constructed here is the second genome map developed for Culinae mosquitoes. *Aedes aegypti* physical map covered 93.5% of the genome which is one of the most complete genome maps across mosquito species [105, 128]. Thus, the major disadvantage of the current physical map of the *Ae. albopictus* genome is that the chromosome-scale scaffolds are still

missing. The AalbF2 assembly still consists of the multiple genome scaffolds that anchored to different parts of the chromosomes.

Analyses of mitotic chromosomes showed that the *Ae. albopictus* chromosomes are slightly longer than *Ae. aegypti* ones, which is consistent with cytofluorimetry results. The major chromosomal proportions of the chromosomes are also slightly different (table 3.1). This observation suggests presence of potential chromosomal rearrangements between these two mosquito-vectors. Our preliminary analyses of physically mapped DNA probes suggested presence of 2 chromosomal inversion on 3p chromosomal arm and 1 inversion in 2p arm. However, we were unable to fully analyze and compare two mosquito genomes because chromosome-length scaffolds are still missing in AalbF2 assembly.

Thus, further improvement of *Ae. albopictus* genome is still necessary. Chromosomal rearrangements in the genome are often associated with adaptations of the mosquitoes to the natural environment [129, 130]. However, presence of such structures in *Ae. albopictus* genome has yet to be determined. The Asian tiger mosquito *Ae. albopictus* is the world's most successful invasive mosquito species that in the last four decades has invaded all continents except Antarctica. It is also a competent vector of various arboviral illnesses including dengue Zika fevers and chikungunya. The future studies of the chromosomal rearrangements will help us to better understand the genetic basis of the incredible phenotypic plasticity of *Ae. albopictus* and its ability in transmitting of multiple pathogens.

## 3.5 Methods

### 3.5.1 In situ hybridization and physical map construction

We developed a new mapping approach based on the amplification of DNA probes using cDNA instead of bacterial artificial chromosome (BAC) clones. DNA probes de- rived from the largest genomic scaffolds, 18S rDNA, PPO genes, and Canu-Flavi19 were mapped to the chromosomes using FISH. To identify genes that could be used for the physical mapping of the *Ae. albopictus* genome, transcripts of *Ae. albopictus* C6/36 cell lines were aligned against AalbF2 [57]. DNA fragments were amplified by PCR using a Q5 high-fidelity DNA polymerase (New England Biolab, Ipswich, MA, USA). cDNA or genomic DNA fragments were used as templates to amplify transcript fragments or large exons; cDNA or genomic DNA fragments were used as templates to amplify transcript fragments or large exons, respectively. RNA was obtained from mosquito ovaries following the Zymo Research Direct-Zol DNA/RNA mini prep protocol (Zymo Research Corporation, Irvine, CA, USA). cDNA was synthesized using ~ 200 ng RNA and primed with oligo (dT) following the Thermo Fisher Scientific Superscript III first-stand synthesis system protocol (Thermo Fisher, Ashville, NC, USA). Laboratory protocols for performing preparations and principal steps for FISH have been described earlier [58, 131]. Transcript fragments or large exons with minimal length of 3.8 kb were used as probes for FISH. PCR-amplified DNA was labeled with two fluorescence dyes Cy3- or Cy5-dUTP (Enzo Life Sciences, Farmingdale, NY, USA) by nick-translation. A pair of DNA probes was hybridized simultaneously to the chromosomes [113, 131]. Slides of mitotic chromosomes were prepared from imaginal discs of 4th instar larvae from the Foshan strain following the published protocols [58, 131, 132]. Chromosomes were stained with a YOYO-1 dye (Thermo Fisher, Ashville, NC,

USA), and slides were mounted with a Prolog Gold reagent (Thermo Fisher, Ashville, NC, USA). FISH results were analyzed using a Zeiss LSM 880 Laser Scanning Microscope (Carl Zeiss Micros- copy, LLC, White Plains, NY, USA) at × 600 magnification. Chromosome idiograms were developed using previously described protocols [58, 75]. Chromosome proportions, such as relative chromosome length and centromeric index (relative length of the p arm), were calculated based on measurements of 60 chromosomes. The statistical analysis was performed using the JPM Pro 15 software program at 95% confidence intervals [133]. One-way ANOVA was used to calculate P values for comparison chromosome proportions between *Ae. albopictus* and *Ae. aegypti.* Chromosomes were subdivided into 96 bands with 4 different intensities.

Note: Other methods are described in depth in Palatini et al., 2020

**Table 3. 1**. **Comparison between *Ae. aegypti* and *Ae. albopictus* mitotic chromosomes.**

| **Chromosome length/proportions** | ***Ae. albopictus*** | | ***Ae. aegypti*** |
|---|---|---|---|
| **Chromosome 1** | | | |
| Average length, µm | 8 | | 7.1 |
| Relative length, %; P-value* | 26.8 | P<0.0001 | 28.4 |
| Centromeric index, %; P-value | 46.7 | P=0.0044 | 46.9 |
| **Chromosome 2** | | | |
| Average length, µm | 11.7 | | 9.5 |
| Relative length, %; P-value | 39.1 | P<0.0001 | 38 |
| Centromeric index, %; P-value | 46.9 | P=0.0051 | 48.6 |
| **Chromosome 3** | | | |
| Average length, µm | 10.2 | | 8.4 |
| Relative length, %; P-value | 34.1 | P<0.0001 | 33.6 |
| Centromeric index, %; P-value | 47.2 | P=0.0079 | 47.4 |

*P-value indicates significant difference in relative length and centromeric index between Ae. albopictus and Ae. aegypti chromosomes.

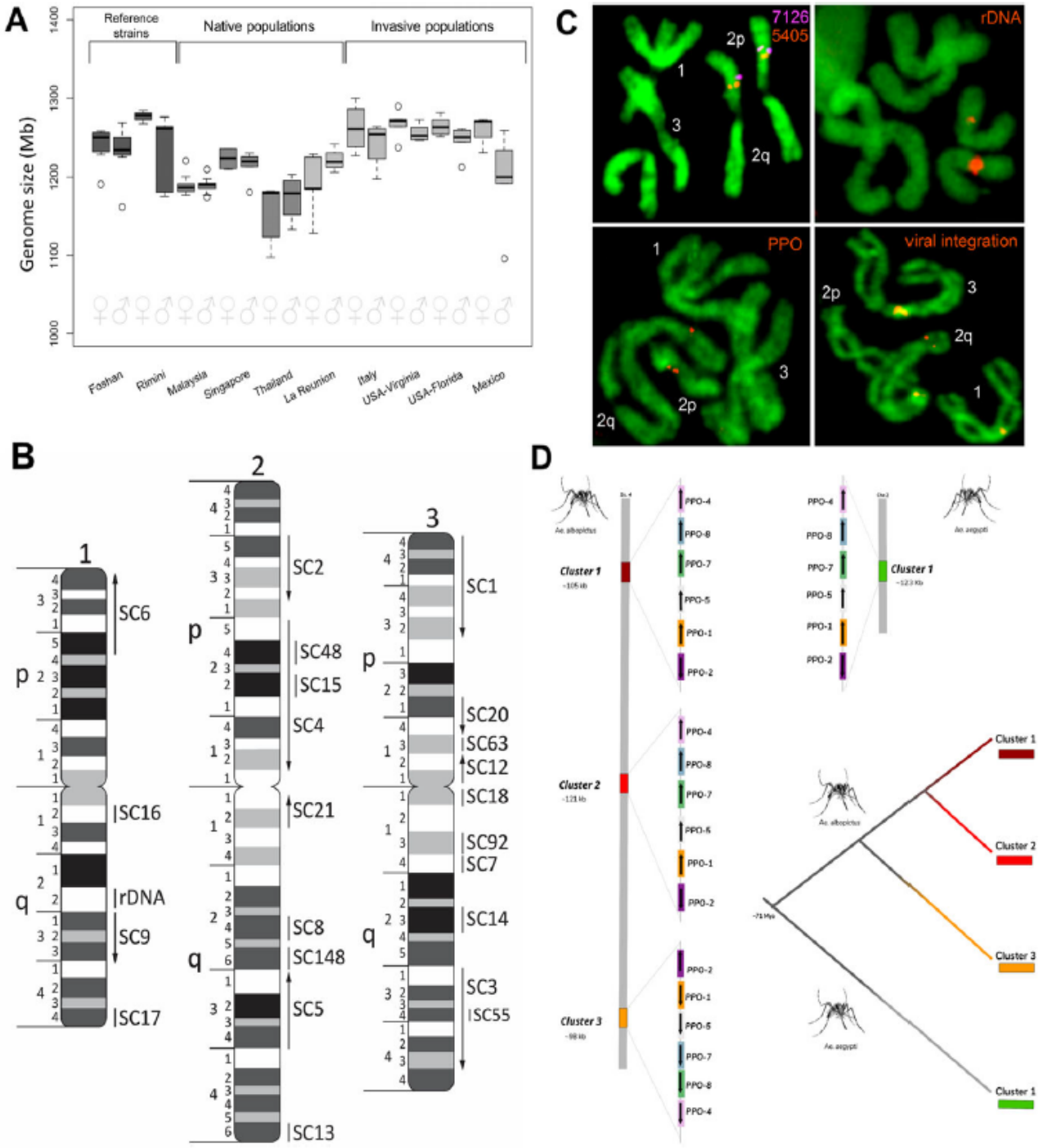


**Figure 3. 1. Size of the Aedes albopictus genome and physical map.** A cytofluorimetric-based estimates of the genome size of *Ae. albopictus* strains, including Foshan and Rimini from which genome assemblies were derived based on short-read Illumina sequencing [3, 4] and Ae. albopictus wild-collected samples from the native home range (Malaysia, Singapore, Thailand), an old-colonized region (La Reunion), and newly invaded areas (the USA, Mexico, Italy). The *Ae. albopictus* genome size is estimated to be in the range of 1095–1299 Mb, comparable or slightly larger than that of Ae. aegypti (1066–1309 Mb) [5]. b Physical genome map of *Ae. albopictus* based on 50 DNA probes hybridized in situ to mitotic chromosomes. Chromosomes

and chromosome arms are indicated by numbers 1, 2, and 3 and letters p and q, respectively. Chromosome divisions and subdivisions are shown on the left sides of the idiograms. Scaffolds are indicated by arrows or lines. Arrows indicate orientations of the scaffolds. SC stands for scaffold; rDNA stands for ribosomal locus. C Examples of fluorescence in situ hybridization. Chromosomal locations of transcripts XM_019675405 and XM_020077126 from scaffolds 4 and 48, respectively; rDNA, polyphenol oxidase (PPO) gene clusters, and the largest viral integration in the genome (Canu-Flavi19) are demonstrated. Transcripts are indicated on a figure by the last four digits of their accession numbers. d Schematic illustration of chromosomal locations of PPO cluster triplication in the new assembly of *Ae. albopictus* (GCF_006496715.1). Comparative genomics analysis of the synteny in the AaegL5 genome reveals an array of 6 genes localized in a region of 123.44 kb at chromosome 2 (2:199,230,485-199,353,929) and was locally duplicated twice in *Ae. albopictus*, resulting in 18 PPO genes. PPO gene cluster array of chromosome 2 of *Ae. aegypti* includes AAEL015116 (PPO1), AAEL015113 (PPO2), AAEL013492 (PPO5), AAEL013493 (PPO7), AAEL013501 (PPO4), and AAEL013496 (PP08)

## Chapter 4: Genomic differentiation and intercontinental population structure of mosquito vectors *Culex pipiens pipiens* and *Culex pipiens molestus*

**Authors and Affiliations:** Andrey A. Yurchenko[#1,2,5], Reem A. Masri[#1], Natalia V. Khrabrova[3], Anuarbek K. Sibataev[3], Megan L. fritz[4] & Maria V. Sharakhova[1,2,3]

[1]Department of Entomology and the Fralin Life Sciences Institute, Virginia Polytechnic and State University, Blacksburg, USA

[2]Laboratory of Evolutionary Genomics of Insects, the Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

[3]Laboratory of Ecology, Genetics, and Environment Protection, Tomsk State University, Tomsk, Russia

[4]Department of Entomology, University of Maryland, College Park, USA

[5]Kurchatov Genomics Center, the Federal Research Center Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Science, Novosibirsk, Russia

# The first two authors contributed equally to the work

**Corresponding Author**: Maria V. Sharakhova, Email: msharakh@vt.edu

**Citation:** Yurchenko AA, Masri RA, Khrabrova NV, Sibataev AK, Fritz ML, Sharakhova MV. Genomic differentiation and intercontinental population structure of mosquito vectors Culex pipiens pipiens and Culex pipiens molestus. Sci Rep. 2020 May 5;10(1):7504. doi: 10.1038/s41598-020-63305-z. PMID: 32371903; PMCID: PMC7200692.

## 4.1 Abstract

Understanding the population structure and mechanisms of taxa diversification is important for organisms responsible for the transmission of human diseases. Two vectors of West Nile virus, *Culex pipiens pipiens* and *Cx. p. molestus*, exhibit epidemiologically important behavioral and physiological differences, but the whole-genome divergence between them was unexplored. The goal of this study is to better understand the level of genomic differentiation and population structures of *Cx. p. pipiens* and *Cx. p. molestus* from different continents. We sequenced and compared the whole genomes of 40 individual mosquitoes from two locations in Eurasia and two in North America. Principal Component, ADMIXTURE, and neighbor joining analyses of the nuclear genomes identified two major intercontinental, monophyletic clusters of *Cx. p. pipiens* and *Cx. p. molestus.* The level of genomic differentiation between the subspecies was uniform along chromosomes. The ADMIXTURE analysis determined signatures of admixture in *Cx. p. pipens* populations but not in *Cx. p. molestus* populations. Comparison of mitochondrial genomes among the specimens showed a paraphyletic origin of the major haplogroups between the subspecies but a monophyletic structure between the continents. Thus, our study identified that *Cx. p. molestus* and *Cx. p. pipiens* represent different evolutionary units with monophyletic origin that have undergone incipient ecological speciation.

## 4.2 Introduction

The advent of genomics has provided new insights into population divergence between incipient taxa, changing our vision about the mechanisms of adaptation and speciation [134].Experimental data demonstrate highly heterogenous patterns of population differentiation in various groups of organisms [135]. In addition to the classical model of allopatric speciation, when incipient taxa are isolated geographically [136], it becomes obvious that ecological speciation or the development of reproductive isolation between populations as a result of adaptation to different environments is feasible and common in nature [137-139]. Understanding the population structure and the mechanisms of taxa diversification in the changing environment is extremely important if the studied organisms are related to the transmission of human diseases [140]. Members of the *Culex pipiens* complex are globally distributed throughout Europe, Asia, America, Africa, and Australia and represent competent vectors of the lymphatic filariasis parasite and encephalitis viruses, including the widely spread West Nile virus [141-144]. However, despite the fact that *Cx. pipiens* was the first mosquito species described by C. Linnaeus in his "Systema Naturae" [145], mosquitoes from the *Cx. pipiens* complex still represent "one of the major outstanding problems in mosquito taxonomy" because the members of the complex can mate and produce viable progeny in nature [146, 147]. Thus, the mechanisms of genetic differentiation in the members of this complex remain poorly understood.

The "Systematic Catalog of Culicidae" maintained by the Walter Reed Biosystematics Unit at the Smithsonian Institution [148], recognizes the following species as members of the *Cx. pipiens* complex: *Cx. pipiens*, *Cx. quinquefasciatus*, *Cx. australicus*, and *Cx. globocoxitus* [149]. Among these species, *Cx. pipiens* and *Cx. quinquefasciatus* spread globally in temperate and

tropical/subtropical regions, respectively. Distribution of *Cx. australicus* and *Cx. globocoxitus* is restricted to Australia. In addition to these species, the *Cx. pipiens* complex includes a subspecies, *Cx. p. pallens*, found in Japan and the Far East of Eurasia, and two additional members, *Cx. p. pipiens* and *Cx. p. molestus*, that, according to this catalog, represent two physiological forms. Indeed, *Cx. p. pipiens* and *Cx. p. molestus* exhibit important physiological and ecological differences [141, 149]. *Cx. p. pipiens* mates in open spaces, feeds on birds, and requires a blood meal for oviposition. During the winter, females of *Cx. p. pipiens* undergo diapause. In contrast, *Cx. p. molestus* mates in confined spaces, feeds on mammals, can lay eggs without a blood meal, and females of this subspecies cannot enter diapause. *Cx. p. molestus* is also known as an underground mosquito because it invades basements, sewers, and underground railways. This fact explains why *Cx. p. molestus* can survive in a cold climate without the ability to enter diapause [141, 150]. Since *Cx. p. pipiens* has strong bias toward feeding on birds, while *Cx. p. molestus* readily bites humans and other mammals, hybrids between the two forms can act as a bridge vector between the bird reservoirs and susceptible mammalian hosts [151-153]. Thus, understanding the abilities of *Cx. p. pipiens* and *Cx. p. molestus* to produce viable progeny in natural populations is of medical importance. Historically, P. Forskal described *Cx. molestus* as a separate species from Egyptian specimens in 1775, but, later, this species was synonymized with *Cx. pipiens* [146]. Following the concept of E. Mayr [148], P. Mattingly proposed considering the *Cx. pipiens* complex as a polytypic species; thus, the status of *Cx. molestus* was reduced to the subspecies *Cx. p. molestus* [139, 154]. Later, E. Vinogradova considered *Cx. p. pipiens* and *Cx. p. molestus* as ecophysiological forms because of the differences in their behavior [141]. R. Harbach argued that *Cx. p. molestus* is a “phenotypic, ecological, and physiological variant” of *Cx. p. pipiens* and should not be considered a subspecies because of the absence of clear

morphological differences and the presence of hybrids [146]. However, a molecular analysis based on microsatellite markers determined that *Cx. p. pipiens* and *Cx. p. molestus* from different continents cluster together in a subspecies specific manner and more likely represent two separate taxonomic units or species with a unique evolutionary history [152]. Similar results regarding the monophyletic origin of *Cx. p. pipiens* and *Cx. p. molestus* were obtained by the analysis of single nucleotide polymorphisms (SNPs) in the mitochondrial Cytochrome Oxidase I (COI) gene of European populations [155]. These conclusions were strongly supported by another work using amplified fragment length polymorphism markers that identified discrete genomic differences between the subspecies and, thus, considered *Cx. p. pipiens* and *Cx. p. molestus* to be distinct evolutionary entities that are likely in the process of incipient speciation [156].

Sequencing the *Cx. quinquefasciatus* genome offered exciting opportunities for comparative genomic studies of the *Cx. pipiens* complex [53]. Here, for the first time, we used whole-genome analysis of individual mosquitoes to investigate the level of genomic differentiation and population structures of *Cx. p. pipiens* and *Cx. p. molestus* from different continents. We compared sequences of 40 samples from field collected mosquitoes from Eurasia (Minsk, Republic of Belarus and Mailuu-suu, Kyrgyz Republic) and North America (Washington, D.C., USA) and two laboratory colonies of *Cx. p. pipiens* and *Cx. p. molestus* derived from Chicago, IL, USA. In our study, we follow the classical nomenclature that was proposed by Mattingly in 1951 and consider *Cx. p. pipiens* and *Cx. p. molestus* as subspecies [157, 158].

## 4.3 Results

### 4.3.1 Mosquito collections

In this study, we sequenced whole genomes of 40 individual mosquitoes from 4 different locations—two from Eurasia and two from North America (Table 4.1, Fig. 4.1). All mosquitoes were collected from the urban environment. *Cx. p. molestus* mosquitoes were collected at the larval stage from an underground water source in two locations (the Kyrgyz Republic and the Republic of Belarus). *Cx. p. pipiens* were also collected at the larval stage in an open water reservoir in the Kyrgyz Republic, but at the adult stage in the basement of a multi-floor building in the Republic of Belarus. All field collected specimens from Eurasia were identified as *Cx. pipiens* by larval morphology and then tested to subspecies level using the COI assay, which is based on SNPs in a specific position [159]. Mosquitoes from the Washington, D.C., USA samples were collected as egg rafts from two different urban environments—an underground parking lot and an open water reservoir. Both egg collections were used to establish mosquito colonies that were kept in the laboratory for several generations. At the larval stage, mosquitoes from both collections were identified morphologically as *Cx. pipiens.* However, molecular assays based on COI SNP [159] and polymorphism in the flanking region of a microsatellite locus [160] failed to identify them at the subspecies level. The latter method was considered unreliable for the diagnosis of *Cx. p. molestus* and *Cx. p. pipiens* in California [161]. In addition, we performed a physiological assay and identified that mosquitoes from the parking lot colony were able to lay eggs without a blood meal. Mosquitoes from the open water collection needed a blood meal for egg development. Thus, we called our samples from Washington, D.C, USA *Cx. pipiens* autogenetic (WPA) and *Cx. pipiens* anautogenic (WPN). In addition, we used two colonies of *Cx. p. molestus* and *Cx. p. pipiens* that were derived from the Chicago area in the USA [156].

### 4.3.2 Nuclear genome analysis

After the resequencing the 40 individual genomes, we performed neighbor joining analysis. The *Cx. quinquefasciatus* genome assembly [53, 56] was used as a reference genome and an outgroup. The analysis identified two major monophyletic clusters that segregated in a subspecies-specific manner with high bootstrap support (Fig. 4.2). *Cx. p. molestus* and *Cx. p. pipiens* clusters included field collected mosquitoes from the Republic of Belarus and the Kyrgyz Republic as well as colonies from Chicago, IL, USA. In addition, we determined the presence of a third cluster that had a polyphyletic nature and included mosquitoes from *Cx. pipiens* autogenic and *Cx. pipiens* anautogenic colonies from Washington, D.C., USA and *Cx. quinquefasciatus,* based on the genome assembly of this species [53, 56]. In fact, the Washington, DC area is located within the *Cx. pipiens* - *Cx. quinqufasciatus* hybrid zone (Farajiollahi, et al. 2011) and, thus, these mosquitoes may represent hybrids between the latter species not included in the analysis.

PCA analyses demonstrated groupings along the first principal component (13.57% of variance explained) with a closer relationship among *Cx. p. molestus* and *Cx. p. pipiens* subspecies rather than between geographical groups (Fig. 4.3). The second and third principal components (11.65% and 7.53%, respectively) grouped mainly geographical populations. In this analysis, both autogenic and anautogenic mosquitoes from Washington, D.C. grouped together with *Cx. p. pipiens* from other locations. The ADMIXTURE analysis demonstrated that *Cx. p. molestus* and *Cx. p. pipiens* from the Republic of Belarus, the Kyrgyz Republic, and Chicago, IL, USA formed subspecies-specific clusters at K=2, demonstrating the major pattern of population differentiation (Fig. 4.4). The specimens from Washington, DC clustered together with *Cx. p. pipiens* at K=2-3. We observed some putative signatures of admixture at K=2-3 in all *Cx. p. pipens* samples from

the Republic of Belarus, the Kyrgyz Republic, in 2 out of 5 individuals in the *Cx. p. pipens* Chicago colony, and in all autogenic and 4 out of 5 anautogenic mosquitoes from Washington, D.C. Overall, the introgression signature between the subspecies was lower in northern populations of *Cx. p. pipiens* in the Republic of Belarus and Chicago, IL than in southern locations in the Kyrgyz Republic and Washington, D.C. In the Washington, D.C. samples, the admixture levels were higher in autogenic strains overall than in anautogenic strains. Interestingly, the signature of admixture in all *Cx. p. molestus* samples was very low or even absent, suggesting restricted gene flow from *Cx. p. pipiens* to *Cx. p. molestus*. At a higher level of clustering (K=6-9), the local populations of the subspecies formed distant well-defined clusters without high overlap, especially for *Cx. p. molestus*. Thus, we concluded that both autogenic and anautogenic samples from Washington, D.C. represent *Cx. p. pipiens* or *Cx. p. pipiens* - *Cx. quinqufasciatus* hybrids with some admixture of *Cx. p. molestus,* which is more pronounced in autogenic samples.

To get more insight into the nature of genomic differentiation between subspecies/populations, we calculated genome-wide pairwise *Fst* values and found that they varied between 0.08 and 0.25 and were highly significant for all comparisons (Fig. 4.5). The most diverged group was *Cx. p. molestus* from the Chicago strain, which can be explained by the very low genetic diversity of this strain caused by its longtime cultivation and possible bottleneck, which accelerated the genetic drift (Fig. 4.7). *Fst* analysis along the chromosomes demonstrated relatively uniform patterns of genetic differentiation across the genome without clear signs of "islands of divergence" (Fig. 4.6), which can indicate infrequent hybridization in nature and/or prezygotic instead of postzygotic reproductive isolation as major mechanisms of the subspecies differentiation. The *Fst* levels dropped around centromeres probably because a small number of

markers are present in these regions due to a highly repetitive sequence composition. Overall, *Fst* values were higher between the subspecies than between autogenic and anautogenic *Cx. p. pipiens* from Washington, D.C. The level of genomic diversity did not differ significantly between the subspecies demonstrating an overall trend of slightly lower genetic diversity in *Cx. p. molestus;* however, this was not significant in pairwise comparisons except for the cultivated colonies from Chicago, IL (Fig. 4.7), which demonstrated a strong depletion of genetic diversity in *Cx. p. molestus*.

### 4.3.3 Mitochondrial genome analysis

In addition, to nuclear genome comparisons, we performed neighboring joining tree analyses using almost complete mitochondrial genomes for all 40 specimens recovered from the whole genome sequencing data. The patterns of phylogenetic structure were very different from the nuclear counterpart (Fig. 4.8) and showed the paraphyletic origin of the major haplogroups among the subspecies but the monophyletic structure between the continents. The mitochondrial genome of *Cx. quinquefasciatus* grouped inside the American haplogroup along with the samples from Washington, D.C. and did not demonstrate any significant divergence. Samples of *Cx. p. molestus* from the Republic of Belarus and the Kyrgyz Republic formed a diverged monophyletic haplogroup. Interestingly, two samples of *Cx. p. pipiens* from the Republic of Belarus represented a very diverged haplogroup, which may be reminiscent of the past species diversity or has been introgressed from other diverged populations/species. We specifically checked the quality of alignment and calling for this haplogroup but did not identify any abnormalities. Overall, we concluded that mitochondrial genome comparison does not reflect the true evolutionary history of the subspecies/species, probably due to multiple introgression events.

## 4.4 Discussion

### 4.4.1 Independent monophyletic origin of *Culex pipiens pipiens* and *Culex pipiens molestus*

Two alternative hypotheses have been proposed to explain the differences in ecological and physiological strategies of *Cx. p. pipiens* and *Cx. p. molestus* [152]. One hypothesis explains such differentiation as a rapid shift in physiological and behavior traits as an adaptation to an underground environment that is associated with human activities [157]. In this scenario, the local populations of *Cx. p. molestus* must be closely related to the local populations of *Cx. p. Pipiens* [162]. This scenario considers *Cx. p. molestus* as an eco-physiological variant of *Cx. p. pipiens*. An alternative hypothesis suggests that the difference between *Cx. p. pipiens* and *Cx. p. molestus* is a result of the distinction in their evolutionary history. Ecological and physiological strategies of *Cx. p. molestus* could have first arisen under the warm climate [152]. Later, such a strategy appeared to be relevant because of the human activities that created an underground environment and, thus, spread *Cx. p. molestus* all over the world. This scenario considers *Cx. p. molestus* and *Cx. p. pipiens* as separate evolutionary entities.

In our study, we tested these hypotheses using whole genome analysis of 40 *Cx. p. molestus* and *Cx. p. pipiens* samples collected from four locations in Eurasia and North America. Our findings rejected the hypothesis of reemergence of *Cx. p. molestus* from the local populations of *Cx. p. pipiens* and strongly supported the idea of an independent monophyletic origin of both *Cx. p. pipiens* and *Cx. p. molestus* from different continents. NJ and ML analyses separately cluster *Cx. p. pipiens* and *Cx. p. molestus* from the Republic of Belarus, the Kyrgyz Republic, and Chicago, IL, USA (Fig. 4.2). The presence of an additional hybrid cluster formed by autogenic and anautogenic field collected specimens from Washington, D.C., USA more likely reflects the existence of a *Cx. pipiens – Cx. quinquefasciatus* hybrid zone in this area16. However, based on

genetic distances, this cluster is much closer related to the Eurasian *Cx. p. pipiens* rather than to the *Cx. p. molestus* from Eurasia and the USA (Chicago, IL). Additionally, PCA (Fig. 4.3) and ADMIXTURE (Fig. 4.4) analyses identified two distinct genetic clusters that correspond to *Cx. p. pipiens* and *Cx. p. molestus*. In both cases, all the specimens from Washington, D.C. clustered together with *Cx. p. pipiens* from other locations. Genome-wide pairwise *Fst* values were highlyGenome-wide pairwise *Fst* values were highly significant for all the comparisons (Fig. 4.5). We did not identify any specific region of high divergence in the genome (Fig. 4.6). Overall, *Fst* values were much higher between *Cx. p. molestus* and *Cx. p. pipiens* than between autogenic and anautogenic samples of *Cx. p. pipiens* from Washington, D.C. The most diverged group, based on all conducted analyses, was the strain of *Cx. p. molestus* from Chicago, IL, which can be explained by the very low genetic diversity of this strain caused by its longtime colonization and possible bottleneck, which accelerated genetic drift (Fig. 4.7).

Similar observations about the independent monophyletic origin of *Cx. p. molestus* and *Cx. p. pipiens* were obtained using microsatellite analysis in ~600 samples of *Cx. pipiens* mosquitoes derived from different world-wide locations [152]. The unrooted distance tree procedure clustered aboveground populations of *Cx. p. pipiens* and underground populations of *Cx. p. molestus* from northern and southern Europe separately. Similar to our study, the aboveground populations of *Cx. p. pipiens* from the USA formed an additional cluster suggesting more intense hybridization between the members of the *Cx. pipiens* complex in North America. The admixture analysis indicated the presence of three major clusters that corresponded to *Cx. p. molestus, Cx. p. pipiens*, and *Cx. quinquefasciatus.* The latter cluster was found in this study only in the USA samples. Another work based on amplified fragment length polymorphism (AFLP) analyses compared samples from southern and northern Europe and strains from Chicago, IL, USA [24].

The North American and European populations used in this study showed a similar ADMIXTURE pattern in the AFLP genome scan. The analysis of COI genes also indicated a monophyletic origin of *Cx. p. pipiens* and *Cx. p. molestus* in Europe, Asia, and Africa [159].

Thus, the independent monophyletic origin and high level of genetic divergence between *Cx. p. molestus* and *Cx. p. pipiens* suggest that these two members of the *Cx. pipiens* complex represent distinct phylogenetic entities with independent evolutionary histories prior to human-mediated translocation.

### 4.4.2 Concepts of speciation and evolution of the *Culex pipiens* complex

Theoretical models of speciation in animals could be subdivided into two major groups: allopatric or geographical speciation and sympatric or ecological speciation. The first concept of speciation, which was intensively promoted by E. Mayr [136], suggests that incipient taxa first become isolated geographically. This situation reduces the gene flow between the populations and may lead to accumulation of mutations that cause genetic incompatibility among the hybrids. An alternative concept of speciation emphasizes ecological barriers between the emerging taxa as major drivers of evolution [163]. This scenario considers the development of reproductive isolation between the populations as a result of adaptation to different environments without geographical isolation, which usually takes place in the face of gene flow. In this situation, the hybrids between incipient taxa become less fit to the environment that promotes natural selection of any traits that reduce mating between them. We think that overall diversification of the *Cx. p. pipiens* and *Cx. p. molestus* subspecies represents a striking example of speciation through isolation-by-ecology mechanisms. In our study, the *Fst* analysis determined significant levels of genomic divergence between *Cx. p. pipiens* and *Cx. p. molestus* (Fig. 4.5 and 4.6) across the entire genome without clear islands of speciation. Surprisingly, the levels of differentiation were

lower around the centromeres, probably due to the low number of reliable SNVs in these highly repetitive regions. The differentiation was extremely high between the Chicago, IL, USA strains of *Cx. p. pipiens* and *Cx. p. molestus* but was lower between the subspecies in the Eurasian mosquito collections. Overall, these observations suggest significant restriction of gene flow between the subspecies.

Several mechanisms of reproductive isolation between *Cx. p. pipiens* and *Cx. p. molestus* have been described [164]. Two of them are prezygotic acting, before fertilization, and reduce the opportunities for mosquito mating. The first mechanism is related to habitat specialization of the larvae: *Cx. p. molestus* occupies basements or other underground environments, but *Cx. p. pipiens* prefers open aboveground water bodies for breeding sites. This reduces chances for the two subspecies to meet and mate at the adult stages. The second isolating mechanism relies on the differences in mating behavior between *Cx. p. molestus* and *Cx. p. pipiens*. *Cx. p. molestus* males usually form homogeneous swarms near the ground and require limited space for mating [164]. In contrast, *Cx. p. pipiens* males swarm near the foliage about 2-3 m above the ground. Experimental studies of mating behavior in small cages indicate that in crosses of females and males of *Cx. p. molestus* copulation success was 90% but in *Cx. p. pipiens* it was only 3.3% [165]. In inter-subspecies crosses between *Cx. p. molestus* and *Cx. p. pipiens,* the copulation success was also low and varied between 6.6% to 10% depending on the sexes of the subspecies. This study demonstrated that females of both subspecies actively avoid copulation with males from an alternative subspecies. Moreover, *Cx. p. pipiens* females were unsuccessful in receiving sperm from *Cx. p. molestus* and, as a result, produced no eggs.

Two other reproductive isolating mechanisms described for *Cx. p. pipiens* and *Cx. p. molestus* are postzygotic, they act after the mating and result in decreased fitness of the hybrids. One of

the mechanisms is related to the inheritance of diapause in hybrids of *Cx. p. molestus* and *Cx. p. pipiens* as a recessive trait [141]. F1 hybrids and a significant portion of F2 hybrids are unable to develop diapause and cannot survive under winter conditions. This mechanism, perhaps, could explain the higher introgression levels in *Cx. p. pipiens* in southern locations [152, 155, 156]. Finally, the members of the *Cx. pipiens* complex are exposed to cytoplasmic incompatability of hybrids infected with different strains the *rickettsial* parasite *Wolbachia pipientis*. Despite cytoplasmic introgression of this parasite through hybridization between the members of the *Cx. pipiens* complex [166], cytoplasmic incompatibility could significantly limit survival rates of the hybrids. For example, *W. pipientis* infection significantly reduced hybridization between *Cx. pipiens* and *Cx. quinquefasciatus* in South Africa [167]. Another study demonstrated that in Eurasian populations *Cx. p. molestus* was only infected by one strain of *W. pipientis* but *Cx. p. pipiens* by two different strains [168]. Moreover, that the specimens of *Cx. p. molestus* and *Cx. p. pipiens,* which we used in our study, were infected with the same strains of *W. pipientis* in the south of the Kyrgyz Republic but by different strains in the north of the Republic of Belarus. Thus, it may explain the differences in introgression from *Cx. p. molestus* to *Cx. p. pipiens* that was more pronounced in the Republic of Belarus than in the Kyrgyz Republic. In fact, an interesting example of infectious speciation was described in the South American *Drosophila paulistorum* complex. In this complex, six semi-species with overlapping geographic distribution became reproductively isolated as a result of premating and postmating isolation triggered by the *Wolbachiae* infection [169].

Recent genomic studies conducted on different organisms including *Drosophila simulans* [170], *Rhagoletis* fruit flies [171, 172], and *Heliconius* butterflies [173, 174] provide additional evidence that ecological speciation occurs in nature. The genomic patterns of speciation can be

very different [135] ranging from small genomic islands of speciation [174] to significant levels of divergence across the entire genome. Widespread genomic divergence was identified between the incipient species *Anopheles gambiae* and *An. coluzzii* in the *An. gambiae* complex [175]. These species were originally identified by differences in the structure of their ribosomal DNA as S and M forms [176], but later their taxonomic status was elevated to the species level [177]. *An. gambiae* and *An. coluzzii* are believed to develop reproductive barriers in sympatry as a result of the differences in their ecological preferences [178]. The *An. gambiae* larval stage is associated with small rain pools. In contrast, *An. coluzzi* exploits persistent water reservoirs associated with rice cultivation. Although, premating barriers between the species are incomplete [179], they developed differences in their swarming behavior [180, 181] and divergent song types [182].

Thus, a high level of genome-wide divergence, a striking difference in adaptation to distinct ecological environments and evidence of prezygotic and postzygotic barriers for mating suggest that *Cx. p. pipiens* and *Cx. p. molestus* represent distinct ecological units that undergo incipient ecological speciation.

### 4.4.3 Hybridization in the Culex pipiens complex

The most intriguing observation of the members of the *Cx. pipiens* complex is that, despite differences in ecology, physiology, behavior, and geographic distribution, they still can hybridize and produce viable progeny in nature, indicating that reproductive isolation between them is not complete. Our study also demonstrated significant hybridization events between the subspecies *Cx. p. molestus* and *Cx. p. pipiens.* Whole genome comparison indicated that most of the *Cx. p. pipiens* samples represent individuals with some level of introgression from *Cx. p. molestus* (Fig. 4.4). We observed high discrepancy between the nuclear and mitochondrial phylogenies (Fig.

4.2, 4.8), which indicates that, historically, the transmission of mitochondrial genomes can happen between the subspecies. At the same time, we did not observe haplogroups shared between the local populations of the subspecies. In the continents, all the mitochondrial phylogenies were strongly monophyletic, which points to male-mediated dispersal and hybridization. The admixture with *Cx. p. molestus* was higher in southern populations of *Cx. p. pipiens* in the Kyrgyz Republic and in Washington, D.C., USA, but lower in northern populations in the Republic of Belarus and Chicago, IL, USA. This could be related to an incapability of the hybrids to develop diapause in cold climates. For comparison, microsatellite analysis revealed a modest level of hybridization between *Cx. p. pipiens* and *Cx. p. molestus* of ~8% in the northern cities of the USA (Chicago, IL and New York, NY) [162, 183]. In southern Europe, where *Cx. p. pipiens* and *Cx. p. molestus* could both be found aboveground, the hybridization levels between them were similar and was estimated at 8-10 % [184]. Much higher levels of hybridization were found in southern populations in eastern USA, where ~40% of all samples were identified as hybrids between *Cx. p. molestus* and *Cx. p. pipiens* [153]. Thus, overall hybridization rates between the members of *Cx. pipiens* were higher in North America than in the Old World. For comparison, hybridization between cryptic species in the *An. gambiae* complex (*An. gambiae* and *An. coluzzi*) significantly varied between 1% in Mali [179] to >20% in Guinea Bissau [185], which is comparable to overall hybridization levels between *Cx. p. molestus* and *Cx. p. pipiens*.

Intriguingly, in our study, we did not determine a *Cx. p. pipiens* admixture signature in any *Cx. p. molestus* samples from the Republic of Belarus, the Kyrgyz Republic, and Chicago, IL, USA (Fig. 4). These findings demonstrate very limited or no gene flow from *Cx. p. pipiens* to *Cx. p. molestus*. Similar findings of asymmetric introgression from *Cx. p. molestus* to *Cx. p. pipiens*

was shown by previous studies [153, 186]. The mechanism of asymmetric introgression is currently unknown. One hypothesis suggests that males of *Cx. p. molestus*, which can mate in confined spaces, can hybridize with both *Cx. p. molestus* and *Cx. p. pipiens* females. In contrast, *Cx. p. pipiens* males, which require space for swarming, have a higher disposition to mate with *Cx. p. pipiens* females [187]. However, more population genetic and experimental studies are needed to explain this phenomenon.

Finally, our study demonstrated that *Cx. p. pipiens* can develop autogeny as a result of adaptive introgression of the genetic material from *Cx. p. molestus*. In the field collected samples from aboveground and underground environments in Washington, D.C., USA, we selected mosquitoes for autogeny. The underground mosquitoes were autogenic but the aboveground mosquitoes were anautogenic. However, mosquitoes from both autogenic and anautogenic colonies formed a single cluster when neighbor joining analysis was applied (Fig. 4.2). Moreover, PCA (Fig. 4.3) and ADMIXTURE (Fig. 4.4) approaches cluster these samples together with *Cx. p. pipiens* from other locations. Populations with mixed characteristics were found in Europe [186] and in the USA [188]. In Portugal, an unusual pattern of blood feeding behavior on birds was also found in *Cx. p. molestus* [153].

Thus, the presence of ongoing hybridization between members of the *Cx. pipiens* complex suggests that the speciation process between them is not complete and postzygotic barriers of reproductive isolation are not fully formed. Overall, we believe that members of the *Cx. pipiens* complex represent a remarkable model for studying different aspects of geographical and ecological speciation in the face of ongoing gene flow between them and local adaptations to diverse environments.

## 4.5 Materials and Methods

### 4.5.1 Mosquito collections.

In this study, we used field collected mosquitoes from Minsk, Republic of Belarus, Mailuu-suu, Kyrgyz Republic, and Washington, D.C., USA (Table 4.1, Fig. 4.1). The map on Fig. 4.1. was created using program ESRI ArcGis Pro v.2.4 (https://www.esri.com/). The species distributions are shown according to previously published data [16]. Mosquitoes from the Republic of Belarus were collected from an urban environment in Minsk, the capital city of the Republic. *Cx. p. pipiens* mosquitoes were collected at the adult stage in the basement of a multi-floor building. *Cx. p. molestus* were collected at the larval stage by dipping on the ground floor of a clinical hospital. In the Kyrgyz Republic, all samples were collected in an urban environment at the larval stage by dipping. *Cx. p. pipiens* were found in an open water pool, while *Cx. p. molestus* were found in an underground water barrel. Mosquitoes were identified as *Cx. p. molestus* and *Cx. p. pipiens* by the COI assay [155]. In Washington, D.C., USA mosquitoes were collected in an urban environment as egg rafts from an open water puddle and an underground parking lot. Mosquitoes from both locations were successfully colonized and fed on fish pellet food at the larval stage. Morphologically, mosquito larvae from both collection sites were identified as *Cx. pipiens*. Adult mosquitoes were kept at 26 °C and offered 10% sucrose solution. Mosquitoes from both sites were tested for autogeny. Individuals from the parking lot demonstrated the ability to lay eggs without blood feeding, while individuals from the open water reservoir did not. Anautogenic mosquitoes were fed on artificial membrane blood feeders 4–5 days after emerging. Five mosquitoes from the second generation of both colonies were selected for sequencing. Finally, we used laboratory colonies of *Cx. p. molestus* and *Cx. p. pipiens* from Chicago, IL, USA. These colonies were established from mosquitoes collected in the Chicago,

IL area. *Cx. p. molestus* was collected by sampling a drainage sump using aspirators and larval dipping in January 2009 (Mutebi and Savage, 2009) and *Cx. p. pipiens* was collected as egg rafts in above ground breeding sites in Evanston and Northfield, IL in August of 2016.

### 4.5.2 DNA extraction and sequencing.

DNA was extracted from individual mosquitoes for all Eurasian samples using the DNA Invisorb Spin Tissue Mini Kit (Invitek, Berlin, Germany) and for all American samples using the Qiagen Blood and Tissue Kit (Qiagen, Germantown, MD, USA). DNA concentration was determined by nanodrop; ~50 ng of DNA from each individual mosquito was sent for sequencing to the Fasteris company (Fasteris Inc., Switzerland). Sequencing was done using Illumina HiSeq 2 x 150 single index configuration with a 12-17X coverage for individual mosquito samples.

### 4.5.3 Genomic analysis: variant calling.

Individual genomes for each population were aligned against the *Cx. quinquefasciatus* reference genome [56] using BWA-MEM software [55]. The resulting BAM files were sorted and PCR duplicates were removed using Samtools [189]. To call Single Nucleotide Variations (SNVs), we used the *bcftools mpileup* and *call* multisample (*-m*) functions of the bcftools package [190] and had alignments with the quality of mapping equal to at least 40 and a base quality not less than 20. All of the BAM files were called simultaneously. The raw VCF file was filtered with vcftools [191] to remove all variants with a quality less than 500 *(--minQ 500*), more than two alleles per position, and were monomorphic with missing genotypes in more than four individuals.

To call mitochondrial DNA (mtDNA) variants, the reads were aligned to the nearly complete mitochondrial genome of *Cx. quinquefasciatus* with BWA-MEM and only unique, properly paired alignments with the highest quality (60) were left for the sorting and deduplication steps.

The mitochondrial variants were called using the bcftools consensus algorithm (*bcftools call -c)* [190].

### 4.5.4 Population genetic analysis

To reduce the influence of the linkage between neighboring loci on population genetic analysis, we removed all loci located closer than 10 Kbp from each other using vcftools (–thin 10000). The resulting pruned dataset with all individuals from both subspecies simultaneously was used to compute PCA using SNPrelate software v1.20.1 [192] (https://bioconductor.org/packages/release/bioc/html/SNPRelate.html) and unsupervised clustering using ADMIXTUREv1.3 [193] (http://software.genetics.ucla.edu/admixture/download. html) for K = 2–9. To find the most optimal number of ADMIXTURE clusters we employed Cross-Validation Error estimation and ΔLogLikelihood which uses a relative gain of LogLikelihood for each successive step (estimated as LogLikelihood K = x − LogLikelihood K = x − 1). Pairwise Fst values [194] between locations/subspecies were calculated with the Stacks package v2.4 [195] (http://catchenlab.life.illinois.edu/stacks/). To profile the Fst values between subspecies along the chromosomes, we used the sliding window Fst approach implemented in vcftools with a window size equal to 50 Kbp and steps equal to 5 Kbp. To construct neighbor-joining trees for the autosomal and mtDNA datasets, we used RapidNJ software v2.3.2 [196] (https://birc.au.dk/software/rapidnj/) with K2P distance and 1000 bootstrap replications. Additionally, we used ML approach realized in RaxML software v8.2.12 [197] (https://github.com/stamatak/standard-RAxML/ releases) with GTRGAMMA model, rooting to Cx. quinquefasciatus and 500 bootstrap replications to corroborate the NJ analysis for nuclear WGS dataset. A ML approach for mitochondrial dataset was used in MEGA X software

v10.1.766 (https://www.megasoftware.net) with K2P distances and 1000 bootstrap replicates. To find an optimal model of substitutions we used Bayesian information criteria in jModelTest 2 software [198]. Finally, all the nodes with bootstrap values lower 90% and 70% for nuclear and mitochondrial phylogenies respectively were collapsed. A part of figures was produced using R software v3.5.3 (https://cran.r-project.org/mirrors.html) [199].

**Table 4. 1. Collection of the *Cx. pipiens* mosquitoes identified and used for sequencing.**

| Sample ID | Country | City, state | GPS coordinates | Habitat | Species/subspecies, stage, physiology | Identification method | Sample size |
|---|---|---|---|---|---|---|---|
| BP | Republic of Belarus | Minsk | 54.00, 27.00 | Basement, multi floor building | *Cx. p. pipiens,* adults | COI, [159] | 5 |
| BM | Republic of Belarus | Minsk | 54.00, 27.00 | Basement, multi floor building | *Cx. p. molestus,* larvae | COI, [159] | 5 |
| KP | Kyrgyz Republic | Mailuu-suu II | 41.00, 72.00 | Aboveground open water pool | *Cx. p. pipiens,* larvae | COI, [159] | 5 |
| KM | Kyrgyz Republic | Mailuu-suu | 41.00, 72.00 | Underground barrel | *Cx. p. molestus,* larvae | COI, [159] | 5 |
| CP | USA | Chicago, IL | N/A | colony | *Cx. p. pipiens,* larvae | [156] | 5 |
| CM | USA | Chicago, IL | N/A | colony | *Cx. p. molestus,* larvae | [156] | 5 |
| WPN | USA | Washington, DC | 38.25, -76.76 | Aboveground puddle | *Cx. pipiens,* larvae, anautogenic | morphology | 5 |
| WPA | USA | Washington, DC | 38.25, -76.76 | Underground parking lot | *Cx. pipiens,* larvae, autogenic | morphology | 5 |

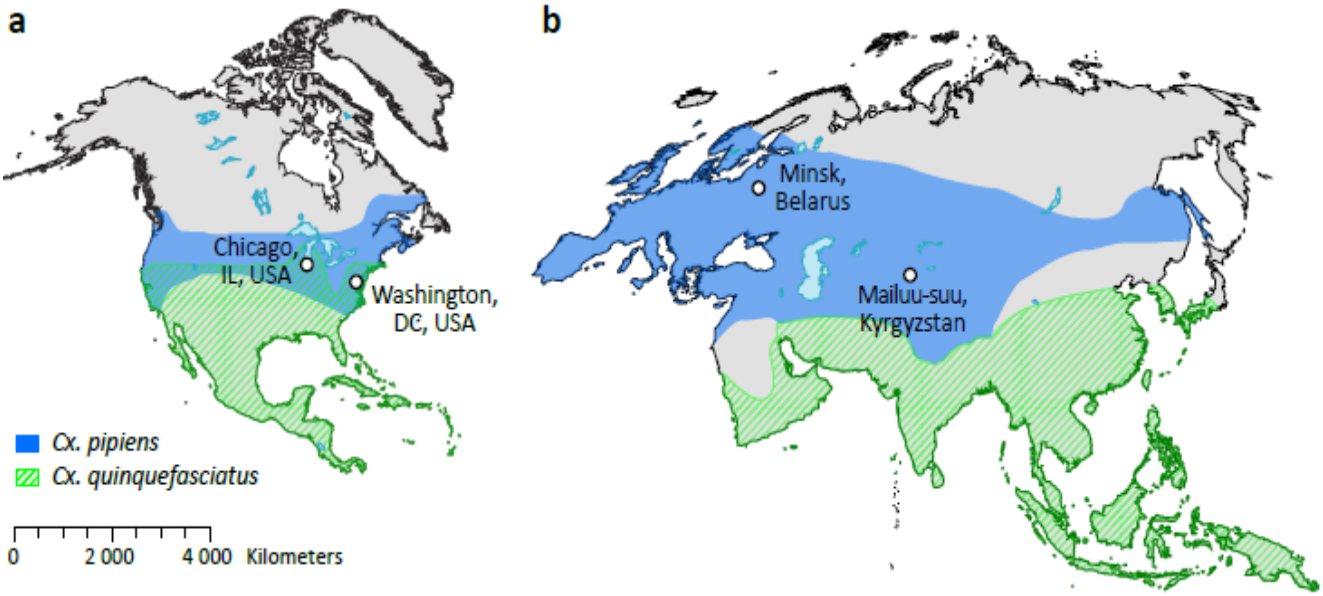


**Figure 4. 1. Mosquito collection sites in North America (a) and Eurasia (b).** Worldwide distribution of *Cx. pipiens* and *Cx. quinquefasciatus* is indicated by blue color and green lines. The overlap between the sheds represents a species hybrid zone indicating that collections in the USA (Chicago, IL and Washington, D.C.) were made within the hybrid zone. The map was created using program ESRI ArcGis Pro v.2.4 (https://www.esri.com/). The species distributions are shown according to previously published data [16].

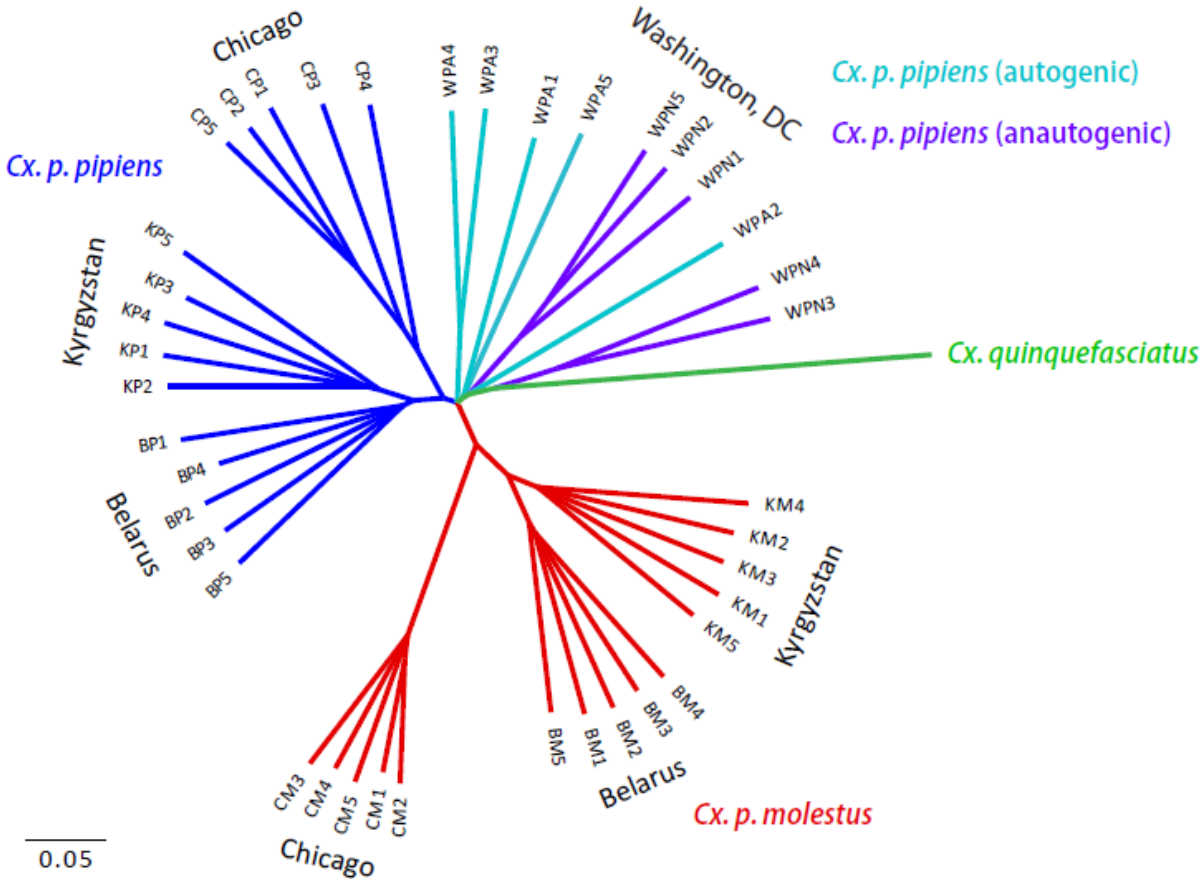


**Figure 4. 2. Neighbor joining tree based on K2P distances and autosomal genome-wide SNVs.** Samples of *Cx. p. pipiens* and *Cx. p. molestus* from the Republic of Belarus (Belarus), the Kyrgyz Republic (Kyrgyzstan), and the USA (Chicago, IL) demonstrate their monophyletic origin except for the autogenic (WPA) and anautogenic (WPN) samples from Washington, D.C., USA. *Cx. quinquefasciatus* is used for the outgroup species. All the nodes with bootstrap support less than 90% were collapsed.

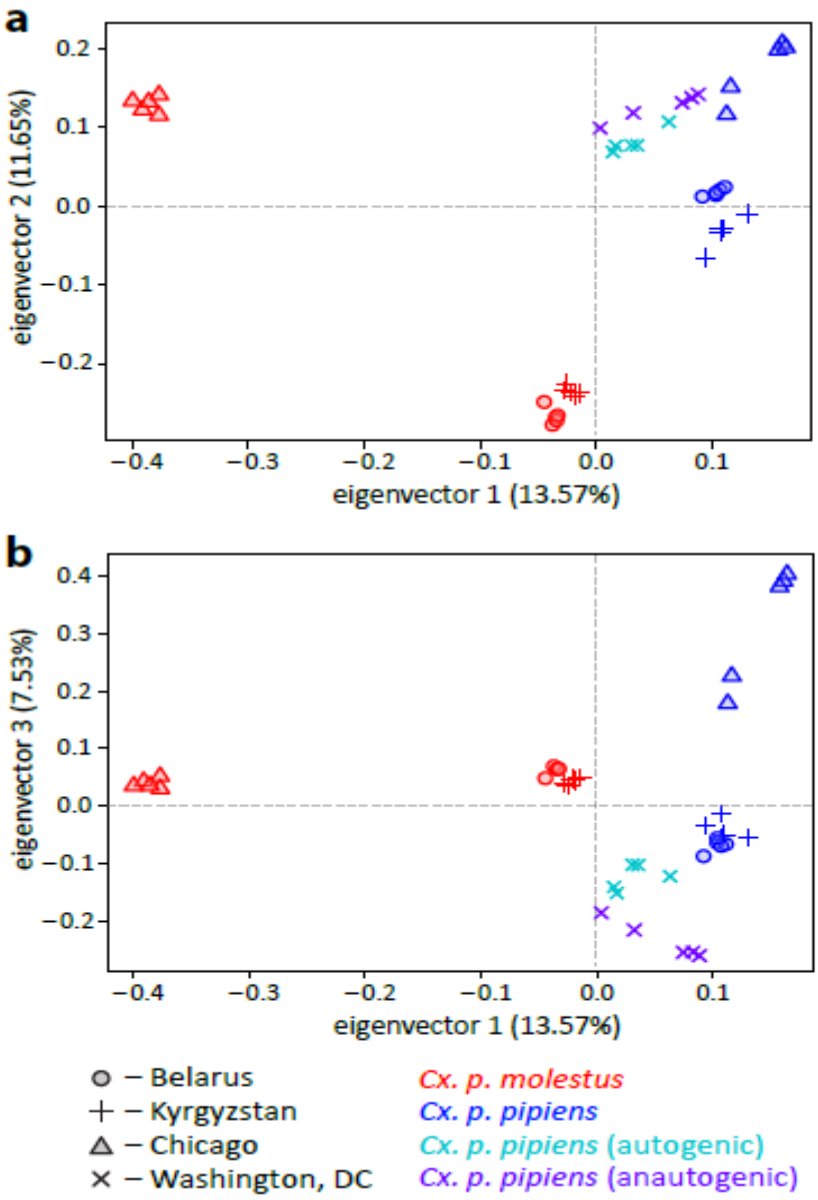


**Figure 4. 3. Principal component analysis of the individual samples for PC1-2 (a) and PC1-3 (b).** PC1 separates the subspecies *Cx. p. pipiens* and *Cx. p. molestus* and PC2 with PC3 into separate geographic regions. Samples are from the Washington, D.C., USA group with other samples of *Cx. p. pipiens.*

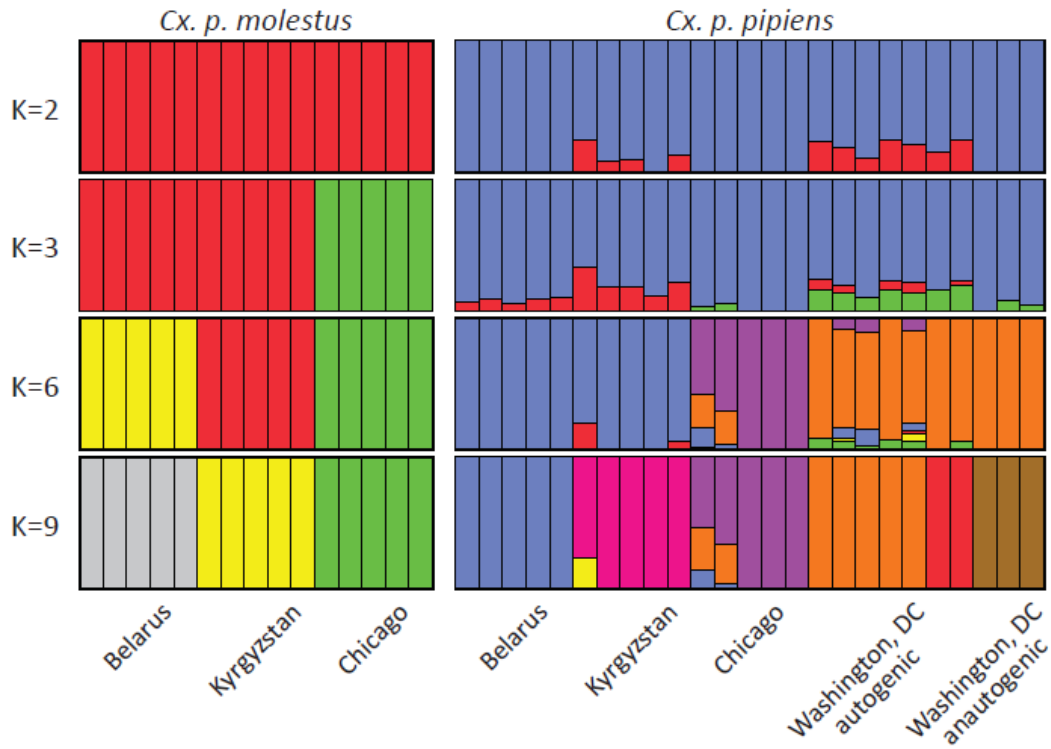


**Figure 4. 4. ADMIXTURE plot of the samples grouped by subspecies and regions.** The major pattern of clustering is formed by subspecies level at K = 2–3 and local populations at K = 6–9. Samples are from the Washington, D.C., USA group with other samples of *Cx. p. pipiens.*

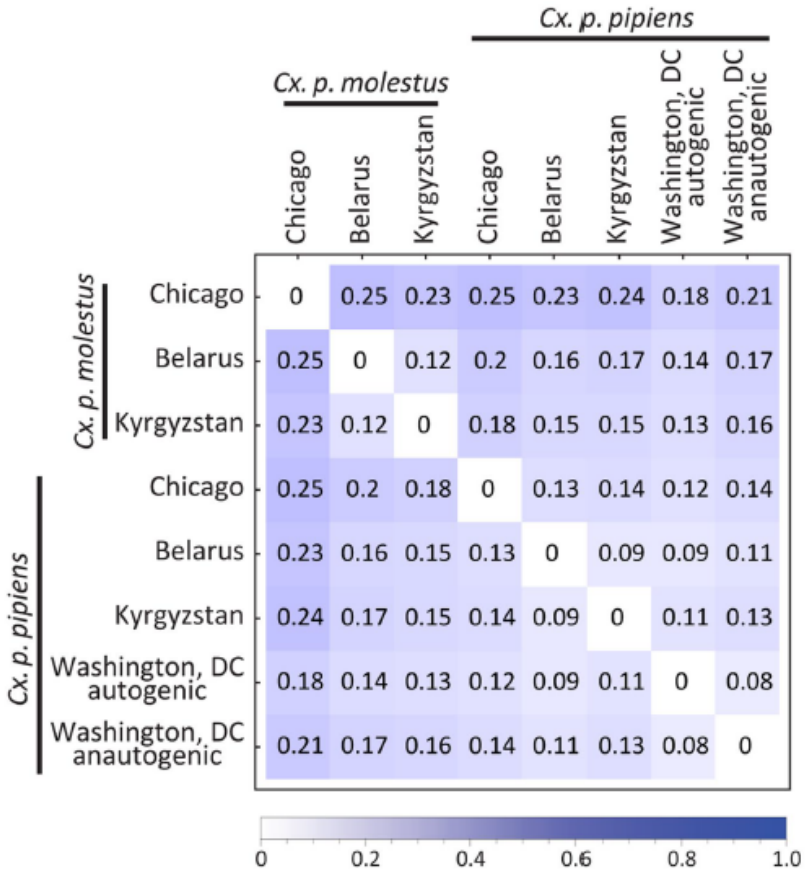


**Figure 4. 5. *Fst* matrix between the studied species/locations.** *Cx. p. molestus* is the most divergent group due to its high level of genetic drift and low diversity. All values are highly significant.

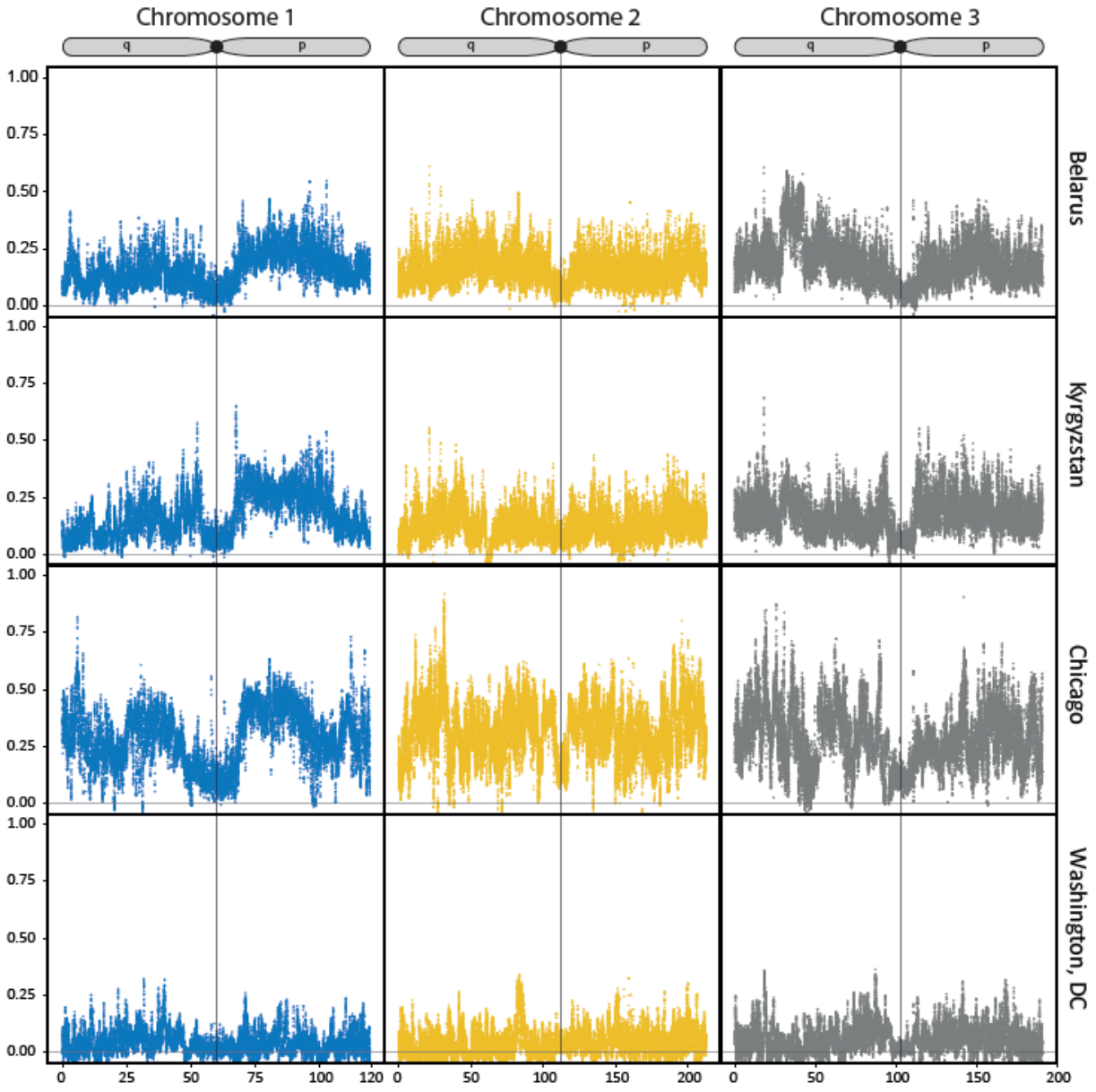


**Figure 4. 6. *Fst* values between the subspecies in different locations plotted along chromosomes.** The patterns of genetic differentiation across the genome are more or less uniform. Overall, *Fst* values are higher between the subspecies than between autogenic and anautogenic *Cx. p. pipiens* from Washington, D.C., USA. P and q stand for short and long chromosome arms, respectively.

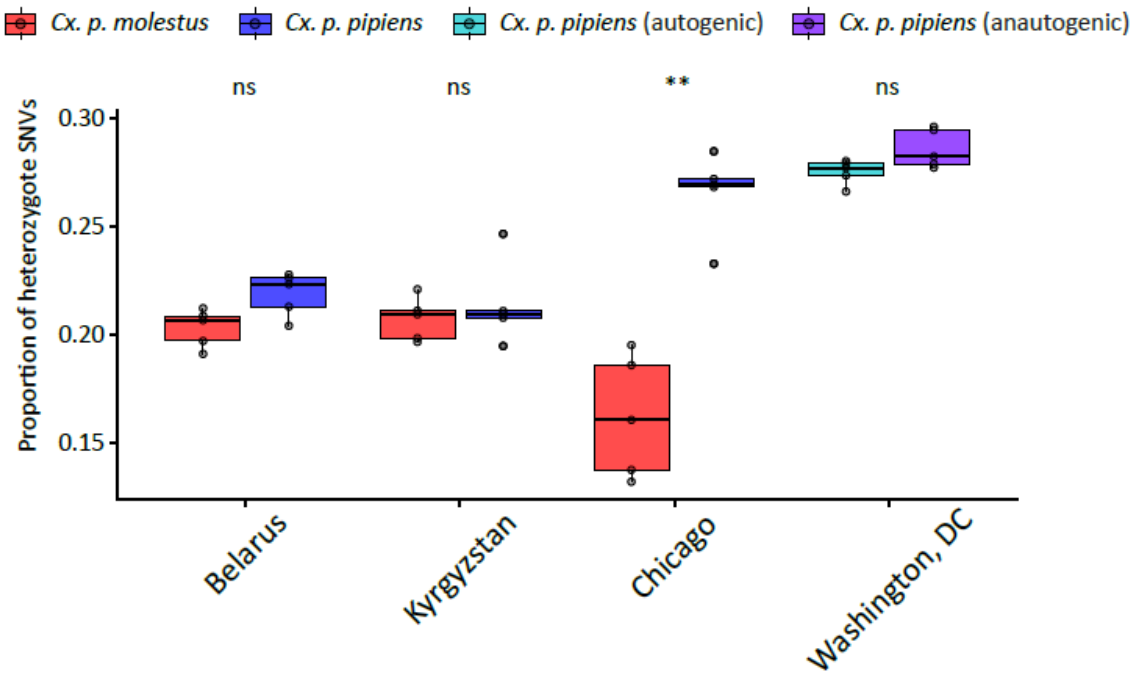


**Figure 4. 7. Level of individual genetic diversity estimated as a proportion of heterozygous SNVs per genome.** Samples of *Cx. p. molestus* from Chicago, IL, USA demonstrate the lowest level of diversity and are significantly different from *Cx. p. pipiens* from the same location. **Indicate significant differences (P-value < 0.01).

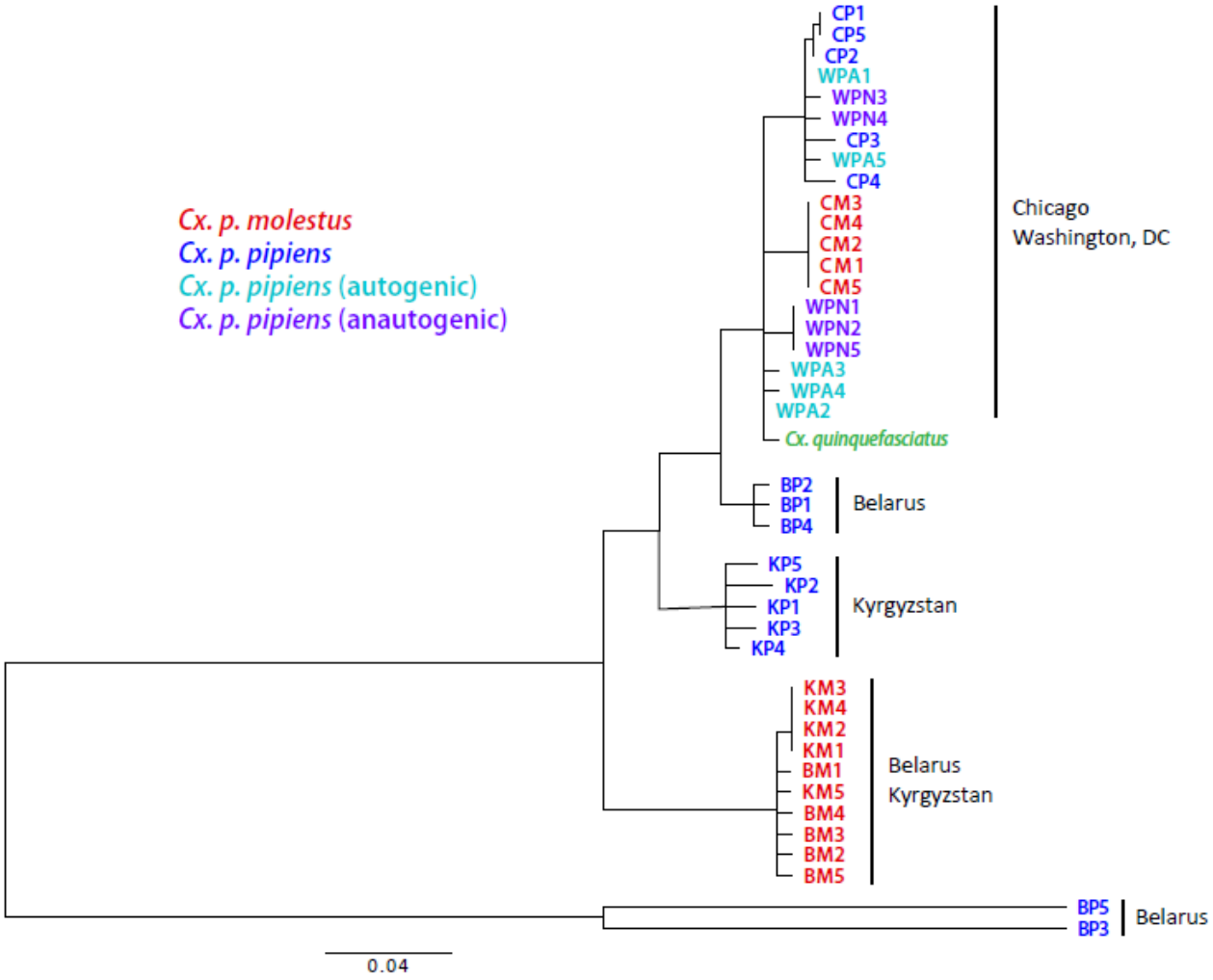


**Figure 4. 8. Neighbor joining tree based on the almost complete mtDNA genomes derived from WGS data (K2P distance)**. Patterns showed paraphyletic origins of the major haplogroups among the subspecies but monophyletic structures between the continents. All the nodes with bootstraps lower than 70% were collapsed.

**Chapter 5: Detection of Chromosomal Inversions in West Nile Vectors from *Culex pipiens* Complex Using Hi-C Approach.**

**Authors and Affiliations:** Reem A. Masri[1], Varvara Lukyanchikova[1], Ilja Brusentsov[1], Maria V. Sharakhova[1]

[1]Department of Entomology and the Fralin Life Sciences Institute, Virginia Polytechnic and State University, Blacksburg, USA

**Keywords:** *Culex pipiens* complex, Hi-C, chromosomes, inversion, rearrangement.

## 5.1 Introduction

The globally distributed *Culex pipiens* species complex consists of several related species that are morphologically similar and are involved in different viral transmissions, primarily West Nile virus (WNV). WNV, originally isolated from Uganda in 1937, was first introduced in the United States in 1999 through patients in New York City [200-202]. The geographic range of the virus presently extends into Canada, all 48 contiguous states, and south into Mexico, and Central and South America [203]. Studies conducted in the early 2000s suggested that mosquitoes of the *Cx. pipiens* complex served as "bridge vectors" biting both birds and humans and thus increasing the risk of transmission of WNV [204, 205]. Therefore, it is important to understand the taxonomic relationships between members of the *Culex pipiens* species complex to better aid in mosquito control methods against this complex. The complex includes *Culex p. pipiens* L., *Cx. p. molestus* L., *Cx. quinquefascitus* Say, *Cx. p. pallens*, *Cx. australicus*, *and Cx. globocoxitus* [146, 206]. Among these species, *Cx. pipiens* and *Cx. quinquefasciatus* spread globally, in temperate and tropical/subtropical regions, respectively [207]. Distribution of *Cx. australicus* and *Cx. globocoxitus* is restricted to Australia [208]. *Cx. p. pallens* is distributed in far East Asia and considered a subspecies of *Cx. pipiens* [209]. Also, two additional members *Cx. p. pipiens* and *Cx. p. molestus* are known as two physiological forms of *Cx. pipiens* [152]. Due to the sometimes-wide vector ability discrepancies among the taxa that are morphologically indistinguishable, complexes of sibling species poses specific challenges [158].

Despite being morphologically identical, *Cx. p. pipiens* and *Cx. p. molestus* exhibit important physiological and ecological differences [210]. *Cx. p. pipiens* mates in open spaces, feeds on birds, requires a blood meal for oviposition, and undergoes diapause. On the contrary, *Cx. p.*

*molestus* mates in confined spaces, feeds on mammals, can lay eggs without a blood meal, and females cannot enter diapause [211]. *Cx. p. molestus* is known as the underground mosquito because it invades basements, sewers, and underground railways [212]. Since *Cx. p. pipiens* prefers feeding on birds, while *Cx. p. molestus* readily bites humans and other mammals, hybrids between the two forms can act as an important bridge vector between the bird reservoirs and susceptible mammalian hosts [213]. Moreover, development of effective vector control strategies requires detailed knowledge regarding how many mosquito species to target and what the relationships between these species are.

Our previous data [214], published along with others that used microsatellite markers [152] and single nucleotide polymorphisms in the mitochondrial Cytochrome Oxidase I (COI) gene [155], determined that *Cx. p. pipiens* and *Cx. p. molestus* represent two separate taxonomic units or species with a unique evolutionary history. However, the existence of ongoing hybridization between members of the complex suggests that the process of speciation between them is not complete and that reproductive isolation barriers are not completely established [214].

Population characteristics commonly feature gradual changes in time or space associated with environmental variables such as temperature, rainfall, or altitude [215]. Mechanisms underlying and maintaining such adaptations are still poorly studied in Culicidae. Chromosomal polymorphism such as inversions, Robertsonian fusions and fissions, and translocations has been perceived as an important driving force in local adaptation, speciation processes, and evolution of sex chromosomes [216-218]. The key evolutionary significance of inversions may be that they reduce recombination in heterozygotes, shielding genomic regions from introgression [219]. Also, recombination suppression allows locally advantageous alleles to segregate together, causing higher fitness than recombinant haplotypes [220]. Moreover, reduced recombination

between heterozygotes promotes ecological divergence over time and encourages reproductive isolation between incipient species [216, 221]. Chromosomal inversions are widespread among members of the *Anopheles gambiae sensu lato* (*s.l.*) complex species [178]. Inversion polymorphisms have been speculated to be responsible for much of the adaptive ecological potential in this species complex [178]. A great example is the chromosomal polymorphisms with the largest geographical distribution involving inversions on the left and right arm of chromosome 2 (the 2L*a* and 2R*b* arrangements, respectively); these have been greatly studied in *A. gambiae sensu stricto* (*s.s.*) and found to correlate with factors such as aridity [222], insecticide resistance [202], and pathogen transmission [3]. A detailed investigation of the chromosomal differences that may explain physiological variation and adaptation in the *Cx. pipiens* complex is lacking.

Detecting chromosomal rearrangements through new high-throughput technologies can be extremely efficient and aid in detecting precise breakpoint coordinates. New approaches involving next generation sequencing (NGS) have been used to detect chromosomal inversions; however none has been effective when chromosome breakpoints are not already known [223]. Here, we use the power of Hi-C technique [224], a derivative of the chromosome conformation capture (3C) technique [225], to detect both known and novel chromosomal rearrangements from *Cx. p. pipiens* and *Cx. p. molestus*. Moreover, we report metaphase chromosomal measurements for both *Cx. p. pipiens* and *Cx. p. molestus.*

## 5.2 Materials and Methods

### 5.2.1 Mosquito Strain and Slide Preparation

In this study, we used the Buckeye strain for *Cx. p. pipiens* and Chicago strain for *Cx. p. molestus*. Mosquito eggs were hatched at 28 C and, after several days, 2$^{nd}$ or 3$^{rd}$ instar larvae

were transferred to 20 C to obtain a high number of mitotic divisions in the imaginal discs of the 4th instar larvae. For chromosomal preparations, slides were prepared from imaginal discs of 4th instar larvae of both *Cx. p. pipiens* and *Cx. p. molestus*. Chromosome preparations were made using a routine technique based on hypotonic treatment and subsequent application of Carnoy's solution (3 parts of ethanol: 1 part of acetic acid) and 50% propionic acid [131].

### 5.2.2 Chromosomal Measurements and Idiogram Development

For idiogram development, the best images of early metaphase chromosomes from imaginal discs stained with Oxasole Yellow (YOYO-1) iodide (Invitrogen Corporation, Carlsbad, CA, USA) were selected. The original gray-scale images were contrasted as previously described [226]. These chromosome images were straightened and aligned for comparison using the ImageJ program available on the NIH website at: http://rsb.info.nih.gov/ij/ [227]. In total, 100 chromosomes at early metaphase were analyzed. To calculate exact proportions of chromosomes, we utilized standard curve measurements in Zen2009LightEdition software.

### 5.2.3 18S rDNA Probe Preparation and Fluorescence *In Situ* Hybridization (FISH)

Both *Cx. p. pipiens* and *Cx. p. molestus* genomic DNA were isolated using Qiagen Blood and Tissue Kit (Qiagen, Germantown, MD, USA). DNA concentration was determined by nanodrop; ~50 ng of DNA was used in polymerase chain reaction (PCR) for labeling. For PCR, a 25 µl reaction was prepared as follows: 12 µl of IMMOMIX (Bioline USA, Taunton, MA), 1 µl of 10 µM forward primer CCTATATGGTGGCGCTTGAT, 1 µl of 10µM reverse primer AACTAAGAACGGCCATGCAC, 50ng of genomic DNA, 1 µl of Cy3- or Cy5-dUTP (Enzo Life Sciences Inc., Farmingdale, NY, USA), and remaining volume of Fisher DNAse/RNAse free water (Thermo Fisher Scientific Inc., Waltham, MA, USA). PCR cycle conditions were as

follows: initial denaturation, 95°C for 10 min; denaturation, 94°C for 15 sec; annealing, 55°C for 30 sec then 72°C for 30 sec; final extension, 72°C for 3 min and then held at 4°C . For 25 µl probe precipitation, 5 µl of 10 mg/ml salmon sperm DNA (Invitrogen Corporation, Carlsbad, CA, USA), 0.1X of 3M sodium acetate, and 2.5X of 100% ethanol were added and the mixture was placed at -20°C for the next day. FISH of the DNA probe was performed using a standard protocol [131]. Slides were analyzed using a Zeiss LSM 510 Laser Scanning Microscope (Carl Zeiss Microimaging, Inc., Thornwood, NY, USA) at 600X magnification. For each slide, 5–10 chromosome spreads were imaged and analyzed.

### 5.2.4 Hi-C Protocol for Mosquito Embryos

To collect sufficient amount of mosquito eggs, 50-100 female mosquitoes were used and fed fresh sheep blood. The egg dish was placed accordingly to strain biological schedule for overnight egg laying (~ 72 hours for Culex) and the embryos used for experiments were not older than 15 hours, to ensure sensitivity to bleach. Due to enormous loss of cells during the experiment, at least 8 egg rafts were used to generate qualitative Hi-C library. Eggs were carefully collected from egg dishes and incubated in 50% bleach-water solution for ~ 5 minutes to unseal extraembryonic membranes (amnion and serosa) (Figure 5.1). We used 6-well plastic plates and 100 µm nylon filters (MilliporeSigma, Burlington, MA, USA). Incubation time was accurately monitored and was terminated once mild discoloration was noticed, to ensure that embryos were not over digested. The procedure was adapted from the Hi-C 2.0 published protocol [228], with few changes. To ensure maximal digestion, chromatin was incubated with MboI, 25,000U/ml (New England Biolabs, Ipswich, MA, USA) overnight in a thermocycler with interval agitation to fragment DNA with restriction endonucleases. After digestion, the digested DNA forms a smear of 400–3000 bp on agarose gel. Digestion was terminated by heat

inactivation of the restriction enzyme at 62°C for 25 min. Next, the DNA was marked with biotin-14-ATP, 0.4 mM (Thermo Fisher Scientific Inc., Waltham, MA, USA) by strategically replacing one of the deoxyribonucleotides with a biotin-conjugated variant. Klenow 5,000U/ml (New England Biolabs, Ipswich, MA, USA) fragments of DNA polymerase 1 was used to fill the 5' overhang for 90 min at 37°C with 500-1000 rpm rotation. Chromatin was then ligated by thoroughly pipetting the reaction and incubating it at room temperature with slow rotation, 300-500rpm, overnight. When ligation was completed the next day, 10 µg/µl RNase A (New England Biolabs, Ipswich, MA, USA) and 20 µg/µl proteinase-K (Qiagen, Inc. Valencia, CA, USA) were used for crosslink reversal and purification. During the procedure, small aliquots of the solution were taken after 3 main steps: lysis, digestion, and ligation. DNA isolated from these steps was visiualized on 1% agarose gel for quality control. Isolated DNA dissolved in 10 mM Tris-HCl (Thermo Fisher Scientific Inc., Waltham, MA, USA) was then used to perform ultrasonic DNA shearing using COVARIS (M220 COVARIS/E220 COVARIS). The majority of the DNA fragments have to show a smear in the 300-500 bp range on the electropherogram. Next, sheared DNA was carefully transferred to a fresh tube to perform size selection using Agencourt® AMPure XP beads (Beckman Coulter, Inc. Brea, CA, USA) to create even tighter distribution of fragments. Ampure is a mixture of magnetic beads and polyethylene glycol (PEG-8000). DNA concentration was then estimated using dsDNA High Sensitivity Qubit assay (Invitrogen Corporation, Carlsbad, CA, USA) and 10 µl of purified solution was visualized on 2% agarose gel to verify proper size of the selected DNA fragments. Tubes were kept at -20°C for the next step. To enrich for Hi-C ligation junctions, streptavidin-coated beads with a high affinity for the incorporated biotin were used. This effectively eliminates any DNA without biotin which is DNA that was not properly digested. After the last wash, 50 µl of 10 mM Trish-HCl buffer was

used to resuspend the beads. These were then transferred to a PCR tube to proceed with NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® and NEBNext® Multiplex Oligos for Illumina® Index Primers Set 1 (New England Biolabs, Ipswich, MA, USA) to amplify the library directly on streptavidin beads. The final concentration was determined using dsDNA High Sensitivity Qubit assay for *Cx. p. pipiens* and *Cx. p. molestus* libraries and ranged between 17 and 49 ng/µl.

### 5.2.5 Genome Sequencing and Analysis

We sequenced pre-prepared Hi-C libraries using Hi-seq Illumina, 2 × 150 pair-ended reads (Novogene Co., Inc. Sacramento, CA, USA). Four Culex libraries were sequenced per one HiSeq lane, 2 libraries for *Cx. p. pipiens* and 2 for *Cx. p. molestus* to obtain 200 mIn paired-end reads per genome (2 libraries). DNA sequences were then exported and analyzed using Juicebox [229].

## 5.3 Results

In this study, we compared and examined chromosomal difference between *Cx. p. pipiens* and *Cx. p. molestus* that may be the basis to important phenotypes that separates them in nature. Only *Cx. quinquefasciatus* [65] chromosomes were measured before, therefore we determined the lengths of the chromosomal arms for both *Cx. p. pipiens* and *Cx. p. molestus* to provide a basic cytogenetic characterization of their chromosomes. We determined that in *Cx. p. pipiens*, the average chromosome lengths at mid-metaphase were 5.892 µm for chromosome 1, 8.485 µm for chromosome 2, and 7.637 µm for chromosome 3, whereas in *Cx. p. molestus*, the lengths were 5.465 µm for chromosome 1, 7.92 µm for chromosome 2, and 6.76 µm for chromosome 3 (table 5.1). Centromeric indeces (the relative length of the p-arm) for *Cx. p. pipiens* were 44.9%, 47.3% and 44.0% for chromosomes 1, 2 and 3, respectively. Centromeric indeces for *Cx. p. molestus*

were 44.8%, 47.2%, and 45.7% for chromosomes 1, 2, and 3, respectively (table 1). Comparison of chromosome measurements indicated that only the length of chromosome arm 3p was significantly different between the two forms, with the *Cx. p. pipiens*'s chromosome longer than that of *Cx. p. molestus* (Figure 5.2). This finding may suggest the presence of chromosomal rearrangements between the two physiological forms or accumulation of repetitive DNA sequences in *Cx. p. pipiens*.

To investigate whether chromosomal rearrangements were present, we performed more complexed Hi-C analysis. Analysis of the Hi-C data showed the presence of two inversions in *Cx. p. pipiens* and *Cx. p. molestus* (Figure 5.3). Genome sequence was analyzed against the reference genome of *Cx. quinquefasciatus* CpipJ5 [52]. Comparing both heatmaps indicated the presence of 1pA inversion in *Cx. p. molestus* but not in *Cx. p. pipiens*. On the other hand, a 3pA inversion was detected for *Cx. p. pipiens*, but was absent in *Cx. p. molestus* (Figure 5.3 C & D). These inversions appear as bow-tie patterns and are marked with a black arrow in figure 3. By aligning the genome sequences of both Culex forms we determined the chromosomal breakpoints of these inversions (Figure 5.3 C & D) that can be used later to design probes and perform fluorescence *in situ* hybridization for a physical validation of these inversions. The length of the 1pA inversion was 10.970 Mb and that of the 3pA inversion was 1.863 Mb (Fig. 5.4).

**5.4 Discussion**

Numerous efforts and studies were made to identify genetic differences between *Cx. p. pipiens* and *Cx. p. molestus*. Despite the various biological distinctions between *Cx. p. molestus* and *Cx. p. pipiens*, accurate identification of *Cx. p. molestus* by using morphology and/or molecular methods and phylogenetic placement of this form within the *Cx. pipiens* complex remains

controversial. For instance, no consistent differences between these forms in the small ribosomal subunit 12S, CO1, or ND4 were detected in the literature [210]. Moreover, morphological studies using variation in length of dorsal and ventral arms of the phallus in adult males (DV/D ratios) have failed to detect any difference [230, 231]. In 2006, a rapid molecular assay to identify *Cx. p. pipiens, Cx. p.* molestus, and their hybrid form was developed using sequence differences in the genomic regions flanking the CQ11 microsatellite locus [160]. These advances provide evidence on the molecular and genetic basis for the observable phenotypes that distinguishe *Cx. p. pipiens* and *Cx. p. molestus*; however, identification of major genetic or chromosomal differences is needed to help explain the difference in their physiology and adaptation.

Evidence from a number of models suggests that chromosome inversion polymorphism plays an important role in adaptation [232, 233]. Adaptation is suggested to occur as a result of reduced recombination between inverted and standard orientation of chromosomal arrangement, which preserves sets of locally favored alleles and limit their exchange [234]. In *An. gambiae,* for instance, a 2La inversion demonstrates a significant association with resistance to desiccation in Western and Central Africa [235, 236]. Moreover, in *An. atroparvus*, a 3L inversion is associated with DDT insecticide resistance [237]. Inversions have been previously identified in the *Cx. quinquefasciatus* JHB strain [238] using polytene chromosomes. However, detecting chromosomal inversions in this genus continues to be a challenging task as polytene chromosomes are poorly polytenized and their preparation is highly time consuming [75]. In our preliminary studies, we devloped Hi-C libraries for the Cx. *quinquefasciatus* JHB strain and aligned them to the *Cx. quinquefasciatus* genome. The analysis indicated the presence of the 3Rb inversion in the Hi-C heat map that is clearly seen as a long-distance interaction with a “bowtie”

shape outside of the major diagonal (Fig. 5.4B). The position of the inversion overlaps with the molecular marker that was mapped inside the inversion. The cytogenetic structure of the 3Rb inversion heterozygote in polytene chromosomes of the salivary glands *of Cx. quinquefasciatus* is shown in Fig. 4C. Thus, our preliminary studies have demonstrated that the Hi-C approach could be successfully used to identify chromosomal inversions in the *Culex pipiens* complex.

In this study, we measure total chromosome and arm length for both *Cx. p. pipiens* and *Cx. p. molestus*, both of which have chromosomes that are longer than those of *Cx. quinquefasciatus* [65]. Also, chromosome arm 3p was significantly different between the *Cx. p. pipiens* and *Cx. p. molestus*, which may indicate the presence of a duplication or an inversion, but further examination is needed.

Hi-C analysis reveled two inversions, a 1p-inversion in *Cx. p. molestus* and a 3p-inversion in *Cx. p. pipiens*. This is the first report describing the presence of inversions between these two forms with their specific breakpoints along the chromosomes. The systematic relationships of Cx. *pipiens* complex species are still under debate, and the vectorial capacity of each species is still under study [149, 155, 239]. Speciation, in general terms, can occur not only by gene/DNA mutations but also through chromosomal rearrangements, especially inversions. Thus, studying and examining possible landmarks of genome organization like inversions that may be associated with different taxa in the complex is extremely important. Also, having the ability to detect accurate chromosomal breakpoints for these inversions makes it easier to design primers and visualize these inversions using fluorescence *in situ* hybridization. Since chromosomal inversions can be linked to adaptation [222], this type of study can help explain why members of the *Cx. pipiens* complex differ in physiological characteristics.

## 5.5 Conclusion

Here we show for the first time the presence of different chromosomal rearrangements between *Cx. p. pipiens* and *Cx. p. molestus* using the Hi-C assay, without the requirement for deep sequencing. The ability to determine structural aberrations could prove a powerful tool in the identification and understanding of the complex chromosomal rearrangements often seen in different mosquito species.

**Figure 5. 1. Illustration of *Culex* species eggs before and after 50% bleach application.** A. Culex egg raft before bleaching. B. lightly "spotted" 5 minutes of 50%-bleaching, enough for formaldehyde solution to interact with embryos tissues.

**Table 5. 1.** The measurements of *Cx. quinquefasciatus*, *Cx. p. pipiens* and *Cx. p. molestus* mitotic chromosomes from imaginal discs

| Mosquito Species | *Cx. quinquefasciatus* | *Cx. p. pipiens* | *Cx. p. molestus* |
|---|---|---|---|
| ***Chromosome 1*** *average length, µm* | 4.787 | 5.892 | 5.465 |
| *Centromeric index, %* | 47.552 | 44.976 | 44.809 |
| *Relative length, %* | 26.325 | 26.705 | 27.157 |
| ***Chromosome 2*** *average length, µm* | 6.986 | 8.485 | 7.929 |
| *Centromeric index, %* | 46.992 | 47.357 | 47.257 |
| *Relative length, %* | 38.461 | 38.639 | 39.304 |
| ***Chromosome 3*** *average length, µm* | 6.370 | 7.637 | 6.769 |
| *Centromeric index, %* | 47.050 | 46.099 | 45.734 |
| *Relative length, %* | 35.212 | 34.690 | 33.571 |

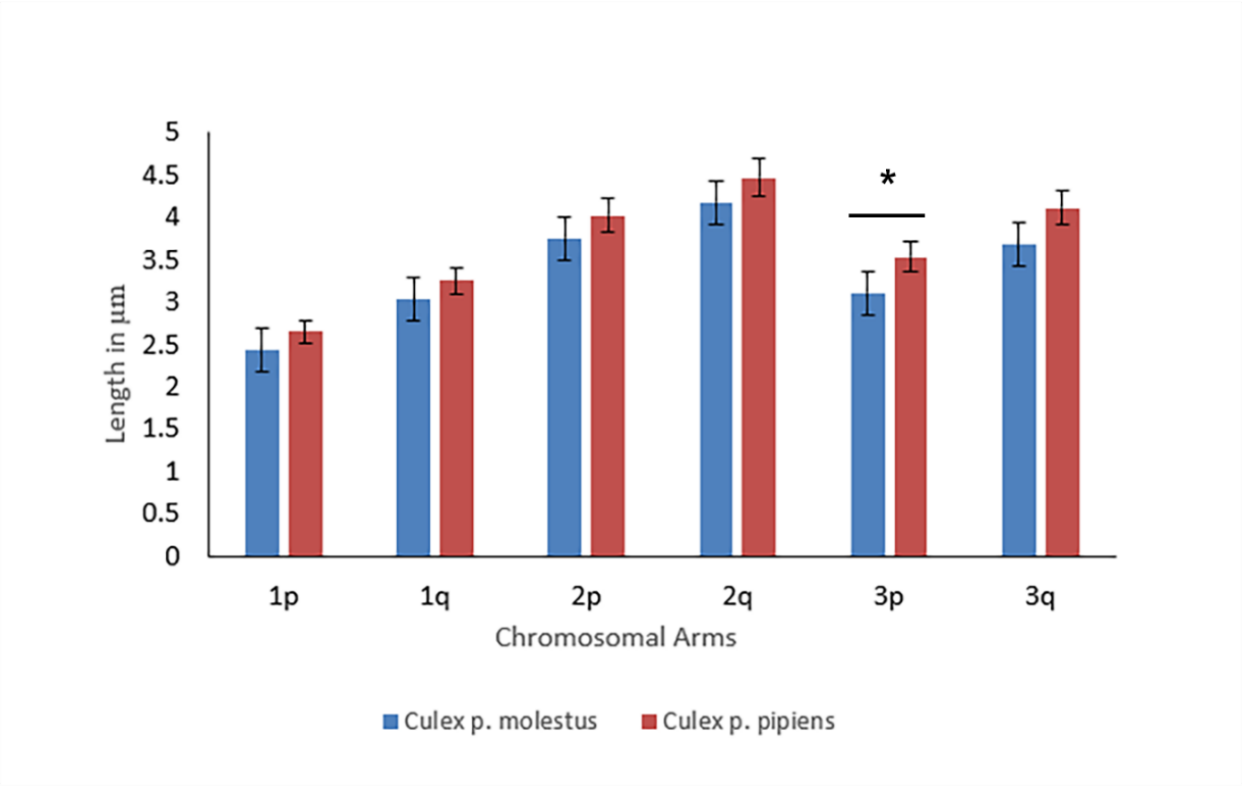


**Figure 5. 2. Chromosome arm length difference between *Culex p. molestus* (Blue) and *Culex p. pipiens* (red).** * $p<0.05$

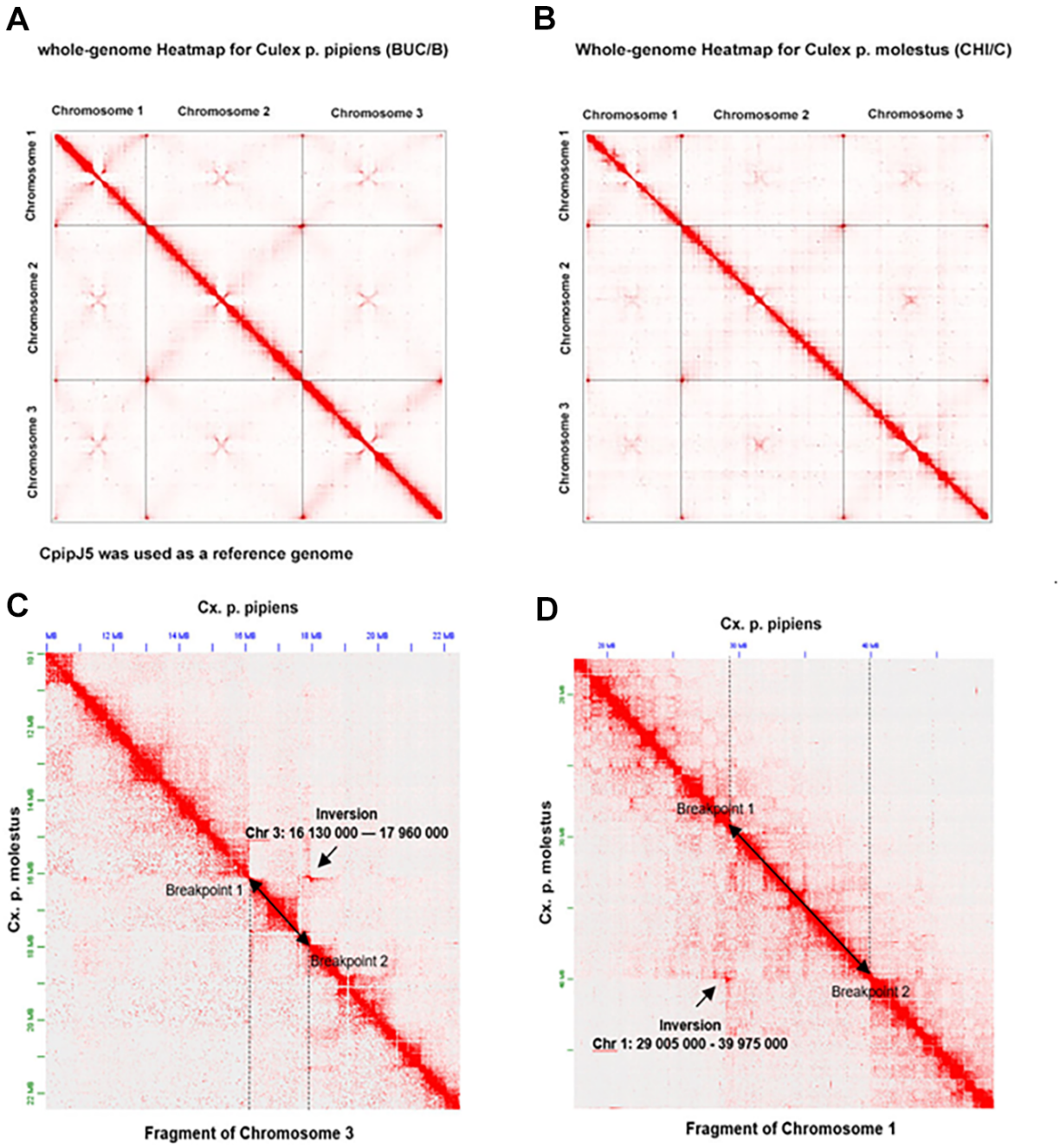


**Figure 5. 3. Whole-genome Hi-C map for Cx. p. pipiens (A) and Cx. p. molestus (B).** Inversion coordinates of breakpoints detected in Cx. p. pipiens and Cx. p. molestus. The inversion with coordinates Chr3: 16 130 000 - 17 960 000, was detected for Cx. p. pipiens (C). The inversion with coordinates Chr1: 29 005 000 - 39 975 000, was detected for Cx. p. Molestus (D). Inversion marked with black arrow demonstrating the typical bow-tie pattern.

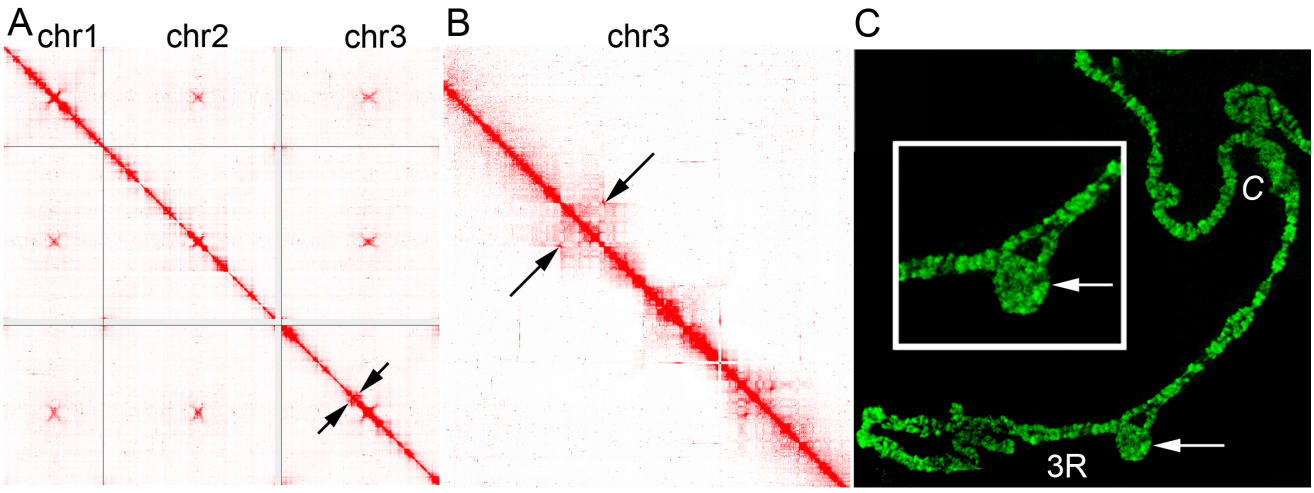


**Figure 5. 4. Chromosomal inversion 3Rb in *Cx. quinquefasciatus*.** 3Rb inversion was detected in *Cx. quinquefasciatus* genome using a Hi-C technique (A, B) and cytogenetic approach in polytene chromosomes (C). Arrows indicate the inversions. C indicates the position of the centromere.

## Chapter 6: Summary and Future Perspectives

Studies performed in Culicine mosquitoes revealed species within this subfamily tend to have large genome size and high number of repetitive and transposable elements compared to anopheline mosquitoes [240]. To develop high-quality genome assemblies, it has been shown that anchoring a physical map onto the chromosomes increases its quality and accuracy. While new genome mapping technologies such as Hi-C scaffolding and optical mapping have the potential to improve genome quality, only cytogenetic mapping using fluorescence *in situ* hybridization binds genomic scaffolds to a specific chromosome and chromosome band. Here, a new simple and robust technique using PCR amplified probes for physical mapping of repeat-rich mosquito genomes was developed. This method can be further used for physical genome mapping in other mosquitoes and insects.

Secondly, this method was used to successfully develop the first physical map of the *Ae. albopictus* genome. The physical map allowed 75% of the assembled genome to be anchored to the chromosomes. Also, the physical map validated and oriented the newly developed genome assembly. A total of 4, 9, and 10 scaffolds were assigned to chromosomes 1, 2, and 3, respectively. Moreover, we physically mapped 18 S ribosomal DNA, largest viral integration (Canu-Flavi19), and polyphenol oxidase gene clusters to chromosome regions 1q22, 2q, and 2p, respectively. To improve this assembly and anchor the whole genome onto the chromosome, additional physical mapping is needed in the future.

Other mosquito species that are widely spread in the US and worldwide belong to the *Culex pipiens* complex and are vectors capable of transmitting the commonly distributed West Nile fevers [241]. Even though *Cx. pipiens* was the first mosquito genus identified by Linnaeus,

mosquitoes from the *Cx. pipiens* complex remain one of the great outstanding issues of mosquito taxonomy because members of this complex can mate and produce viable progeny in nature [242]. Therefore, understanding the mechanisms of genetic differentiation in members of this complex is extremely important. Here, we investigated the extent of genomic differentiation and population structures of *Cx. p. pipiens* and *Cx. p. molestus* from various continents using whole-genome analysis of individual mosquitoes. Our findings supported the idea of an independent monophyletic origin of both *Cx. p. pipiens* and *Cx. p. molestus* from different continents with independent evolutionary histories. However, the presence of hybridization between members of the complex indicates that reproductive barriers between them is still not complete. In general, we conclude that members of the *Cx. pipiens* complex provide an excellent model for researching diverse aspects of geographical and ecological speciation in the face of ongoing gene flow and local adaptations to a variety of ecosystems.

Finally, we studied the chromosomal difference between *Cx. p. pipiens* and *Cx. p. molestus* using the novel 3D genome organization analysis method, Hi-C. Analysis of the Hi-C data of *Cx. p. pipiens* and *Cx. p. molestus* libraries showed the presence of two different inversions, a 1p-inversion in *Cx. p. molestus* and a 3p-inversion in *Cx. p. pipiens*. Using the Hi-C assay, we demonstrate for the first time and without the need for deep sequencing that distinct chromosomal rearrangements exist between *Cx. p. pipiens* and *Cx. p. molestus*. Future efforts are needed to identify these inversions using FISH from samples in natural populations and establish if these inversions are fixed. The ability to identify structural aberrations may be helpful in identifying and recognizing the complex chromosomal rearrangements found in various mosquito species. Mosquito vector species identification is important for implementing effective vector management strategies. The two main methods used in species recognition are

morphology-based taxonomy and molecular characterization. Data on each species' distribution, habitat, host choice, and time of biting differ to the extent where one or more species coexist can make it difficult to incorporate insecticide and control strategies. As a result, knowledge about various species is currently our most powerful tool.

Based on studies conducted here we can draw the following conclusions:

1) the development of a new approach for the construction of physical maps described here using the newly assembled *Ae. albopictus* genome as an example is beneficial for the physical mapping of the highly repetitive Culicinae mosquito genome, but can also be used for development of physical genome maps for other mosquitoes or insects to better aid in their genome assembly.

2) the first physical genome map for *Ae. albopictus* with 75% of the assembled genome anchored to the chromosomes was built. The AalbF2 genome assembly represents the most up-to-date collective knowledge of the *Ae. albopictus* genome.

3) the presence of ongoing hybridization between members of the *Cx. pipiens* complex suggests that the speciation process between them is not complete and postzygotic barriers of reproductive isolation are not fully formed. Overall, we believe that members of the *Cx. pipiens* complex represent a remarkable model for studying different aspects of geographical and ecological speciation in the face of ongoing gene flow between them and local adaptations to diverse environments.

4) the presence of different chromosomal rearrangements between *Cx. p. pipiens* and *Cx. p. molestus* using the Hi-C assay was detected for the first time. The ability to determine structural aberrations could prove a powerful aid in the identification and understanding of the complex chromosomal rearrangements often seen in different mosquito species.

**References**


1. Tandina, F., et al., *Mosquitoes (Diptera: Culicidae) and mosquito-borne diseases in Mali, West Africa.* Parasites & Vectors, 2018. **11**(1): p. 467.
2. Chang, C., et al., *The Zika outbreak of the 21st century.* J Autoimmun, 2016. **68**: p. 1-13.
3. Holt, R.A., et al., *The genome sequence of the malaria mosquito Anopheles gambiae.* Science, 2002. **298**(5591): p. 129-49.
4. *Vector-borne diseases*. World Health Organization 2020.
5. Tandina, F., et al., *Mosquitoes (Diptera: Culicidae) and mosquito-borne diseases in Mali, West Africa.* Parasit Vectors, 2018. **11**(1): p. 467.
6. Manning, J.E., et al., *Mosquito Saliva: The Hope for a Universal Arbovirus Vaccine?* J Infect Dis, 2018. **218**(1): p. 7-15.
7. Mullen, G.R. and L.A. Durden, *Medical and veterinary entomology*. 2009: Academic press.
8. Foster, W.A. and E.D. Walker, *Mosquitos (Culicidae) In: Mullen G, Durden L, editors.* Medical and Veterinary Entomology. Academic, 2002.
9. Singh, R.K., et al., *Prevention and Control Strategies to Counter Zika Virus, a Special Focus on Intervention Approaches against Vector Mosquitoes-Current Updates.* Front Microbiol, 2018. **9**: p. 87.
10. Wilder-Smith, A., et al., *Epidemic arboviral diseases: priorities for research and public health.* Lancet Infect Dis, 2017. **17**(3): p. e101-e106.
11. Sikka, V., et al., *The Emergence of Zika Virus as a Global Health Security Threat: A Review and a Consensus Statement of the INDUSEM Joint working Group (JWG)*. J Glob Infect Dis, 2016. **8**(1): p. 3-15.
12. Ong, R.-Y., F.-M. Lum, and L.F. Ng, *The fine line between protection and pathology in neurotropic flavivirus and alphavirus infections.* Future Virology, 2014. **9**(3): p. 313-330.
13. Sips, G.J., J. Wilschut, and J.M. Smit, *Neuroinvasive flavivirus infections.* Reviews in medical virology, 2012. **22**(2): p. 69-87.
14. Weaver, S.C. and A.D. Barrett, *Transmission cycles, host range, evolution and emergence of arboviral disease.* Nature Reviews Microbiology, 2004. **2**(10): p. 789-801.
15. Shapiro, L.L., et al., *Larval food quantity affects the capacity of adult mosquitoes to transmit human malaria.* Proceedings of the Royal Society B: Biological Sciences, 2016. **283**(1834): p. 20160298.
16. Friend, W. and J. Smith, *Factors affecting feeding by bloodsucking insects.* Annual review of entomology, 1977. **22**(1): p. 309-331.
17. Bowen, M., *The sensory physiology of host-seeking behavior in mosquitoes.* Annual review of entomology, 1991. **36**(1): p. 139-158.
18. Lacey, E.S. and R.T. Cardé, *Location of and landing on a source of human body odour by female Culex quinquefasciatus in still and moving air.* Physiol Entomol, 2012. **37**(2): p. 153-159.
19. van Breugel, F., et al., *Mosquitoes Use Vision to Associate Odor Plumes with Thermal Targets.* Curr Biol, 2015. **25**(16): p. 2123-9.
20. Gillies, M., *The role of carbon dioxide in host-finding by mosquitoes (Diptera: Culicidae): a review.* Bulletin of Entomological Research, 1980. **70**: p. 525-532.
21. Carde, R.T., *Multi-cue integration: how female mosquitoes locate a human host.* Current Biology, 2015. **25**(18): p. R793-R795.

22. Paixão, E.S., M.G. Teixeira, and L.C. Rodrigues, *Zika, chikungunya and dengue: the causes and threats of new and re-emerging arboviral diseases.* BMJ Glob Health, 2018. **3**(Suppl 1): p. e000530.
23. Bliman, P.-A., et al., *Implementation of control strategies for sterile insect techniques.* Mathematical Biosciences, 2019. **314**: p. 43-60.
24. CDC, *Elimination of Malaria in the United States (1947-1951).* 2018.
25. Ramirez, J.L., L.S. Garver, and G. Dimopoulos, *Challenges and approaches for mosquito targeted malaria control.* Curr Mol Med, 2009. **9**(2): p. 116-30.
26. Yakob, L., et al., *Aedes aegypti Control Through Modernized, Integrated Vector Management.* PLoS Curr, 2017. **9**.
27. Dame, D.A., et al., *Historical applications of induced sterilisation in field populations of mosquitoes.* Malar J, 2009. **8 Suppl 2**(Suppl 2): p. S2.
28. Aldridge, S., *Genetically modified mosquitoes.* Nature Biotechnology, 2008. **26**(7): p. 725-725.
29. Carvalho, D.O., et al., *Suppression of a Field Population of Aedes aegypti in Brazil by Sustained Release of Transgenic Male Mosquitoes.* PLoS Negl Trop Dis, 2015. **9**(7): p. e0003864.
30. Sinha, A., et al., *Complete Genome Sequence of the Wolbachia wAlbB Endosymbiont of Aedes albopictus.* Genome Biol Evol, 2019. **11**(3): p. 706-720.
31. Walker, T., et al., *The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations.* Nature, 2011. **476**(7361): p. 450-3.
32. van den Hurk, A.F., et al., *Impact of Wolbachia on infection with chikungunya and yellow fever viruses in the mosquito vector Aedes aegypti.* PLoS Negl Trop Dis, 2012. **6**(11): p. e1892.
33. Altinli, M., et al., *Wolbachia diversity and cytoplasmic incompatibility patterns in Culex pipiens populations in Turkey.* Parasites & Vectors, 2018. **11**(1): p. 198.
34. Sinkins, S.P., *Wolbachia and cytoplasmic incompatibility in mosquitoes.* Insect Biochemistry and Molecular Biology, 2004. **34**(7): p. 723-729.
35. Forrester, J.A., T.G. Weiser, and J.D. Forrester, *An Update on Fatalities Due to Venomous and Nonvenomous Animals in the United States (2008-2015).* Wilderness Environ Med, 2018. **29**(1): p. 36-44.
36. Breman, J.G., A. Egan, and G.T. Keusch, *The intolerable burden of malaria: a new look at the numbers*, in *The Intolerable Burden of Malaria: A New Look at the Numbers: Supplement to Volume 64 (1) of the American Journal of Tropical Medicine and Hygiene*. 2001, American Society of Tropical Medicine and Hygiene.
37. Ruzzante, L., M. Reijnders, and R.M. Waterhouse, *Of Genes and Genomes: Mosquito Evolution and Diversity.* Trends Parasitol, 2019. **35**(1): p. 32-51.
38. Shaw, W.R. and F. Catteruccia, *Vector biology meets disease control: using basic research to fight vector-borne diseases.* Nat Microbiol, 2019. **4**(1): p. 20-34.
39. Lezcano, Ó.M., et al., *Chromatin Structure and Function in Mosquitoes.* Frontiers in Genetics, 2020. **11**(1469).
40. Sharakhova, M.V., et al., *Update of the Anopheles gambiae PEST genome assembly.* Genome Biol, 2007. **8**(1): p. R5.
41. George, P., M.V. Sharakhova, and I.V. Sharakhov, *High-resolution cytogenetic map for the African malaria vector Anopheles gambiae.* Insect Mol Biol, 2010. **19**(5): p. 675-82.

42. Artemov, G.N., et al., *The Physical Genome Mapping of Anopheles albimanus Corrected Scaffold Misassemblies and Identified Interarm Rearrangements in Genus Anopheles.* G3 (Bethesda), 2017. **7**(1): p. 155-164.
43. Compton, A., et al., *The beginning of the end: a chromosomal assembly of the New World malaria mosquito ends with a novel telomere.* bioRxiv, 2020: p. 2020.04.17.047084.
44. Lukyanchikova, V., et al., *Anopheles mosquitoes revealed new principles of 3D genome organization in insects.* bioRxiv, 2020: p. 2020.05.26.114017.
45. Artemov, G.N., et al., *Partial-arm translocations in evolution of malaria mosquitoes revealed by high-coverage physical mapping of the Anopheles atroparvus genome.* BMC Genomics, 2018. **19**(1): p. 278.
46. Waterhouse, R.M., et al., *Evolutionary superscaffolding and chromosome anchoring to improve Anopheles genome assemblies.* BMC Biol, 2020. **18**(1): p. 1.
47. Zamyatin, A., et al., *Chromosome-level genome assemblies of the malaria vectors Anopheles coluzzii and Anopheles arabiensis.* bioRxiv, 2020: p. 2020.09.29.318477.
48. Matthews, B.J., et al., *Improved reference genome of Aedes aegypti informs arbovirus vector control.* Nature, 2018. **563**(7732): p. 501-507.
49. Neafsey, D.E., et al., *Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes.* Science, 2015. **347**(6217): p. 1258522.
50. Kyrou, K., et al., *A CRISPR–Cas9 gene drive targeting doublesex causes complete population suppression in caged Anopheles gambiae mosquitoes.* Nature Biotechnology, 2018. **36**(11): p. 1062-1066.
51. Members of the Complex Trait, C., *The nature and identification of quantitative trait loci: a community's view.* Nature Reviews Genetics, 2003. **4**(11): p. 911-916.
52. Strode, C., et al., *Genomic analysis of detoxification genes in the mosquito Aedes aegypti.* Insect Biochem Mol Biol, 2008. **38**(1): p. 113-23.
53. Arensburger, P., et al., *Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics.* Science, 2010. **330**(6000): p. 86-8.
54. Artemov, G.N., et al., *The Development of Cytogenetic Maps for Malaria Mosquitoes.* Insects, 2018. **9**(3).
55. Palatini, U., et al., *Improved reference genome of the arboviral vector Aedes albopictus.* Genome Biol, 2020. **21**(1): p. 215.
56. Dudchenko, O., et al., *De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds.* Science, 2017. **356**(6333): p. 92-95.
57. Miller, J.R., et al., *Analysis of the Aedes albopictus C6/36 genome provides insight into cell line utility for viral propagation.* Gigascience, 2018. **7**(3): p. 1-13.
58. Timoshevskiy, V.A., et al., *An integrated linkage, chromosome, and genome map for the yellow fever mosquito Aedes aegypti.* PLoS Negl Trop Dis, 2013. **7**(2): p. e2052.
59. de Koning, A.P., et al., *Repetitive elements may comprise over two-thirds of the human genome.* PLoS Genet, 2011. **7**(12): p. e1002384.
60. Rai, K.S., *A Comparative Study of Mosquito Karyotypes1.* Annals of the Entomological Society of America, 1963. **56**(2): p. 160-170.
61. Hall, A.B., et al., *SEX DETERMINATION. A male-determining factor in the mosquito Aedes aegypti.* Science, 2015. **348**(6240): p. 1268-70.

62. Campos, J., C.F. Andrade, and S.M. Recco-Pimentel, *Malpighian tubule polytene chromosomes of Culex quinquefasciatus (Diptera, Culicinae).* Mem Inst Oswaldo Cruz, 2003. **98**(3): p. 383-6.
63. McAbee, R.D., J.A. Christiansen, and A.J. Cornel, *A detailed larval salivary gland polytene chromosome photomap for Culex quinquefasciatus (Diptera: Culicidae) from Johannesburg, South Africa.* J Med Entomol, 2007. **44**(2): p. 229-37.
64. Dennhöfer, L., *[Salivary gland chromosomes of the mosquito Culex pipiens. I. The normal chromosome complement].* Chromosoma, 1968. **25**(3): p. 365-76.
65. Naumenko, A.N., et al., *Mitotic-chromosome-based physical mapping of the Culex quinquefasciatus genome.* PLoS One, 2015. **10**(3): p. e0115737.
66. Sharakhova, M.V., et al., *A physical map for an Asian malaria mosquito, Anopheles stephensi.* Am J Trop Med Hyg, 2010. **83**(5): p. 1023-7.
67. Jiang, X., et al., *Genome analysis of a major urban malaria vector mosquito, Anopheles stephensi.* Genome Biol, 2014. **15**(9): p. 459.
68. Sharakhov, I.V., et al., *Inversions and gene order shuffling in Anopheles gambiae and A. funestus.* Science (New York, N.Y.), 2002. **298**(5591): p. 182-185.
69. Sharakhov, I.V., G.N. Artemov, and M.V. Sharakhova, *Chromosome evolution in malaria mosquitoes inferred from physically mapped genome assemblies.* J Bioinform Comput Biol, 2016. **14**(2): p. 1630003.
70. Abiola, O., et al., *The nature and identification of quantitative trait loci: a community's view.* Nat Rev Genet, 2003. **4**(11): p. 911-6.
71. Rai, K.S. and W.C.t. Black, *Mosquito genomes: structure, organization, and evolution.* Adv Genet, 1999. **41**: p. 1-33.
72. Artemov, G.N., et al., *A standard photomap of ovarian nurse cell chromosomes and inversion polymorphism in Anopheles beklemishevi.* Parasites & Vectors, 2018. **11**(1): p. 211.
73. Brown, T.A., *Mapping genomes*, in *Genomes. 2nd edition*. 2002, Wiley-Liss.
74. Stormo, B.M. and D.T. Fox, *Polyteny: still a giant player in chromosome research.* Chromosome Res, 2017. **25**(3-4): p. 201-214.
75. Sharakhova, M.V., et al., *Imaginal discs--a new source of chromosomes for genome mapping of the yellow fever mosquito Aedes aegypti.* PLoS Negl Trop Dis, 2011. **5**(10): p. e1335.
76. Bridges, C.B., *Salivary chromosome maps: with a key to the banding of the chromosomes of Drosophila melanogaster.* Journal of Heredity, 1935. **26**(2): p. 60-64.
77. Coluzzi, M. *Experimental infections with Rubetella fungi in Anopheles gambiae and other mosquitoes*. in *Proceedings of the First International Congress of Parasitology*. 1966. Elsevier.
78. Coluzzi, M., et al., *Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex.* Transactions of the Royal Society of tropical Medicine and Hygiene, 1979. **73**(5): p. 483-497.
79. Biémont, C., L. Monti-Dedieu, and F. Lemeunier, *Detection of transposable elements in Drosophila salivary gland polytene chromosomes by in situ hybridization.* Methods Mol Biol, 2004. **260**: p. 21-8.
80. Griffiths, A.J., et al., *Modern genetic analysis: integrating genes and genomes*. Vol. 1. 2002: Macmillan.

81. van Berkum, N.L., et al., *Hi-C: a method to study the three-dimensional architecture of genomes.* J Vis Exp, 2010(39).
82. Ay, F. and W.S. Noble, *Analysis methods for studying the 3D architecture of the genome.* Genome Biol, 2015. **16**: p. 183.
83. Ghurye, J., et al., *A chromosome-scale assembly of the major African malaria vector Anopheles funestus.* Gigascience, 2019. **8**(6).
84. Harewood, L., et al., *Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours.* Genome Biology, 2017. **18**(1): p. 125.
85. Wang, S., et al., *HiNT: a computational method for detecting copy number variations and translocations from Hi-C data.* Genome Biology, 2020. **21**(1): p. 73.
86. Chakraborty, A. and F. Ay, *Identification of copy number variations and translocations in cancer cells from Hi-C data.* Bioinformatics, 2018. **34**(2): p. 338-345.
87. Pollard, M.O., et al., *Long reads: their purpose and place.* Human Molecular Genetics, 2018. **27**(R2): p. R234-R241.
88. Amarasinghe, S.L., et al., *Opportunities and challenges in long-read sequencing data analysis.* Genome Biology, 2020. **21**(1): p. 30.
89. Heather, J.M. and B. Chain, *The sequence of sequencers: The history of sequencing DNA.* Genomics, 2016. **107**(1): p. 1-8.
90. Mantere, T., S. Kersten, and A. Hoischen, *Long-Read Sequencing Emerging in Medical Genetics.* Frontiers in Genetics, 2019. **10**(426).
91. Kingan, S.B., et al., *A High-Quality De novo Genome Assembly from a Single Mosquito Using PacBio Sequencing.* Genes (Basel), 2019. **10**(1).
92. Ashton, P.M., et al., *MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island.* Nat Biotechnol, 2015. **33**(3): p. 296-300.
93. Richards, S., *Arthropod Genome Sequencing and Assembly Strategies.* Methods Mol Biol, 2019. **1858**: p. 1-14.
94. Djambazian, H., et al., *De novo genome assembly of the olive fruit fly (Bactrocera oleae) developed through a combination of linked-reads and long-read technologies.* bioRxiv, 2018: p. 505040.
95. Ruzzante, L., M.J.M.F. Reijnders, and R.M. Waterhouse, *Of Genes and Genomes: Mosquito Evolution and Diversity.* Trends in Parasitology, 2019. **35**(1): p. 32-51.
96. Lounibos, L.P. and L.D. Kramer, *Invasiveness of Aedes aegypti and Aedes albopictus and Vectorial Capacity for Chikungunya Virus.* J Infect Dis, 2016. **214**(suppl 5): p. S453-s458.
97. Vazeille, M., et al., *Zika virus threshold determines transmission by European Aedes albopictus mosquitoes.* Emerging microbes & infections, 2019. **8**(1): p. 1668-1678.
98. Berlin, K., et al., *Assembling large genomes with single-molecule sequencing and locality-sensitive hashing.* Nat Biotechnol, 2015. **33**(6): p. 623-30.
99. Goodwin, S., et al., *Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome.* Genome Res, 2015. **25**(11): p. 1750-6.
100. Miller, D.E., et al., *Highly Contiguous Genome Assemblies of 15 <em>Drosophila</em> Species Generated Using Nanopore Sequencing.* G3: Genes|Genomes|Genetics, 2018. **8**(10): p. 3131.
101. Sharakhova, M.V., et al., *Update of the Anopheles gambiaePEST genome assembly.* Genome Biology, 2007. **8**(1): p. R5.

102. George, P., M.V. Sharakhova, and I.V. Sharakhov, *High-resolution cytogenetic map for the African malaria vector Anopheles gambiae.* Insect molecular biology, 2010. **19**(5): p. 675-682.
103. Waterhouse, R.M., et al., *Evolutionary superscaffolding and chromosome anchoring to improve Anopheles genome assemblies.* BMC Biology, 2020. **18**(1): p. 1.
104. Zamyatin, A., et al., *Chromosome-level genome assemblies of the malaria vectors Anopheles coluzzii and Anopheles arabiensis.* bioRxiv, 2021: p. 2020.09.29.318477.
105. Matthews, B.J., et al., *Improved reference genome of Aedes aegypti informs arbovirus vector control.* Nature, 2018. **563**(7732): p. 501-507.
106. Palatini, U., et al., *Improved reference genome of the arboviral vector Aedes albopictus.* Genome Biology, 2020. **21**(1): p. 215.
107. Baptista, R.P., et al., *Assembly of highly repetitive genomes using short reads: the genome of discrete typing unit III Trypanosoma cruzi strain 231.* Microb Genom, 2018. **4**(4).
108. Nene, V., et al., *Genome sequence of Aedes aegypti, a major arbovirus vector.* Science, 2007. **316**(5832): p. 1718-23.
109. Timoshevskiy, V.A., et al., *Genomic composition and evolution of Aedes aegyptichromosomes revealed by the analysis of physically mapped supercontigs.* BMC Biology, 2014. **12**(1): p. 27.
110. Sharakhov, I.V. and M.V. Sharakhova, *Heterochromatin, histone modifications, and nuclear architecture in disease vectors.* Curr Opin Insect Sci, 2015. **10**: p. 110-117.
111. Deschamps, S., et al., *A chromosome-scale assembly of the sorghum genome using nanopore sequencing and optical mapping.* Nat Commun, 2018. **9**(1): p. 4844.
112. Belton, J.M., et al., *Hi-C: a comprehensive technique to capture the conformation of genomes.* Methods, 2012. **58**(3): p. 268-76.
113. Sharakhova, M.V., et al., *Physical Genome Mapping Using Fluorescence In Situ Hybridization with Mosquito Chromosomes.* Methods Mol Biol, 2019. **1858**: p. 177-194.
114. Chen, X.G., et al., *Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution.* Proc Natl Acad Sci U S A, 2015. **112**(44): p. E5907-15.
115. Untergasser, A., et al., *Primer3—new capabilities and interfaces.* Nucleic Acids Research, 2012. **40**(15): p. e115-e115.
116. Biessmann, H., et al., *Isolation of cDNA clones encoding putative odourant binding proteins from the antennae of the malaria-transmitting mosquito, Anopheles gambiae.* Insect Mol Biol, 2002. **11**(2): p. 123-32.
117. Karamjit, S.R., *Genetics of mosquitoes.* Journal of Genetics, 1999. **78**(3): p. 163-169.
118. Severson, D.W., et al., *Linkage map organization of expressed sequence tags and sequence tagged sites in the mosquito, Aedes aegypti.* Insect Mol Biol, 2002. **11**(4): p. 371-8.
119. Hickner, P.V., et al., *Composite linkage map and enhanced genome map for Culex pipiens complex mosquitoes.* J Hered, 2013. **104**(5): p. 649-55.
120. Juneja, P., et al., *Assembly of the genome of the disease vector Aedes aegypti onto a genetic linkage map allows mapping of genes affecting disease transmission.* PLoS Negl Trop Dis, 2014. **8**(1): p. e2652.

121. Kraemer, M.U.G., et al., *Past and future spread of the arbovirus vectors Aedes aegypti and Aedes albopictus.* Nat Microbiol, 2019. **4**(5): p. 854-863.
122. Xu, J., et al., *Comparative transcriptome analysis and RNA interference reveal CYP6A8 and SNPs related to pyrethroid resistance in Aedes albopictus.* PLoS Negl Trop Dis, 2018. **12**(11): p. e0006828.
123. Whitfield, Z.J., et al., *The Diversity, Structure, and Function of Heritable Adaptive Immunity Sequences in the Aedes aegypti Genome.* Curr Biol, 2017. **27**(22): p. 3511-3519.e7.
124. Dudzic, J.P., et al., *Drosophila innate immunity: regional and functional specialization of prophenoloxidases.* BMC Biol, 2015. **13**: p. 81.
125. Pischedda, E., et al., *Insights Into an Unexplored Component of the Mosquito Repeatome: Distribution and Variability of Viral Sequences Integrated Into the Genome of the Arboviral Vector Aedes albopictus.* Front Genet, 2019. **10**: p. 93.
126. Sutherland, I.W., et al., *A linkage map of the Asian tiger mosquito (Aedes albopictus) based on cDNA markers.* J Hered, 2011. **102**(1): p. 102-12.
127. Marconcini, M., et al., *Polymorphism analyses and protein modelling inform on functional specialization of Piwi clade genes in the arboviral vector Aedes albopictus.* PLoS Negl Trop Dis, 2019. **13**(12): p. e0007919.
128. Pischedda, E., et al., *ViR: a tool to account for intrasample variability in the detection of viral integrations.* bioRxiv, 2020: p. 2020.06.16.155119.
129. Ayala, D., et al., *Chromosome inversions and ecological plasticity in the main African malaria mosquitoes.* Evolution, 2017. **71**(3): p. 686-701.
130. Bonizzoni, M., et al., *The invasive mosquito species Aedes albopictus: current knowledge and future perspectives.* Trends Parasitol, 2013. **29**(9): p. 460-8.
131. Timoshevskiy, V.A., et al., *Fluorescent in situ hybridization on mitotic chromosomes of mosquitoes.* J Vis Exp, 2012(67): p. e4215.
132. Suzuki, Y., et al., *Non-retroviral endogenous viral element limits cognate virus replication in Aedes aegypti ovaries.* bioRxiv, 2020: p. 2020.03.28.013441.
133. Jabeen, R., et al. *A COMPARATIVE CHROMOSOMAL COUNT AND MORPHOLOGICAL KARYOTYPING OF THREE INDIGENOUS CULTIVARS OF KALONGI (NIGELLA SATIVA L.)*. 2012.
134. Arias, C.F., S. Van Belleghem, and W.O. McMillan, *Genomics at the evolving species boundary.* Curr Opin Insect Sci, 2016. **13**: p. 7-15.
135. Seehausen, O., et al., *Genomics and the origin of species.* Nat Rev Genet, 2014. **15**(3): p. 176-92.
136. Mayr, E., *Animal Species and Evolution.* Belknap of Harvard University Press, 1963.
137. Ortiz-Barrientos, D., J. Engelstädter, and L.H. Rieseberg, *Recombination Rate Evolution and the Origin of Species.* Trends Ecol Evol, 2016. **31**(3): p. 226-236.
138. Shafer, A.B. and J.B. Wolf, *Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology.* Ecol Lett, 2013. **16**(7): p. 940-50.
139. Butlin, R.K., J. Galindo, and J.W. Grahame, *Review. Sympatric, parapatric or allopatric: the most important way to classify speciation?* Philos Trans R Soc Lond B Biol Sci, 2008. **363**(1506): p. 2997-3007.
140. Powell, J.R., *Genetic Variation in Insect Vectors: Death of Typology?* Insects, 2018. **9**(4).
141. Vinogradova, E.B., *Culex pipiens pipiens mosquitoes: taxonomy, distribution, ecology, physiology, genetics, applied importance and control*. 2000: Pensoft Publishers.

142. Turell, M.J., *Members of the Culex pipiens complex as vectors of viruses.* J Am Mosq Control Assoc, 2012. **28**(4 Suppl): p. 123-6.
143. Turell, M.J., D.J. Dohm, and D.M. Fonseca, *Comparison of the Potential for Different Genetic Forms in the Culex pipiens Complex in North America to Transmit Rift Valley Fever Virus.* J Am Mosq Control Assoc, 2014. **30**(4): p. 253-9.
144. Calistri, P., et al., *Epidemiology of west nile in europe and in the mediterranean basin.* Open Virol J, 2010. **4**: p. 29-37.
145. Linné, C.v., *Systema naturae per regna tria naturae: secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis. Vol. 1, pt. 7*. 1789: Lugduni: Apud JB Delamolliere.
146. Harbach, R.E., *Culex pipiens: species versus species complex taxonomic history and perspective.* J Am Mosq Control Assoc, 2012. **28**(4 Suppl): p. 10-23.
147. Harbach, R.E., C. Dahl, and G.B. White, *Culex (Culex) pipiens Linnaeus (Diptera, Culicidae)-concepts, type designations, and description.* Proc Entomol Soc Wash, 1985. **87**(1): p. 24.
148. *Catalog of the mosquitoes of the world.*
149. Farajollahi, A., et al., *"Bird biting" mosquitoes and human disease: a review of the role of Culex pipiens complex mosquitoes in epidemiology.* Infect Genet Evol, 2011. **11**(7): p. 1577-85.
150. Dehghan, H., J. Sadraei, and S. Moosa-Kazemi, *The Morphological Variations of Culex pipiens Larvae (Diptera: Culicidae) in Yazd Province, Central Iran.* Iran J Arthropod Borne Dis, 2010. **4**(2): p. 42-9.
151. Ciota, A.T., P.A. Chin, and L.D. Kramer, *The effect of hybridization of Culex pipiens complex mosquitoes on transmission of West Nile virus.* Parasit Vectors, 2013. **6**(1): p. 305.
152. Fonseca, D.M., et al., *Emerging vectors in the Culex pipiens complex.* Science, 2004. **303**(5663): p. 1535-8.
153. Gomes, B., et al., *Distribution and hybridization of Culex pipiens forms in Greece during the West Nile virus outbreak of 2010.* Infect Genet Evol, 2013. **16**: p. 218-25.
154. Mattingly, P., *The systematics of the Culex pipiens complex.* Bulletin of the World Health Organization, 1967. **37**(2): p. 257.
155. Shaikevich, E.V., et al., *Genetic diversity of Culex pipiens mosquitoes in distinct populations from Europe: contribution of Cx. quinquefasciatus in Mediterranean populations.* Parasites & Vectors, 2016. **9**(1): p. 47.
156. Gomes, B., et al., *Limited genomic divergence between intraspecific forms of Culex pipiens under different ecological pressures.* BMC Evol Biol, 2015. **15**: p. 197.
157. Mattingly, P.F., *The systematics of the Culex pipiens complex.* Bull World Health Organ, 1967. **37**(2): p. 257-61.
158. Mattingly, P.F., *The Culex pipiens complex.* Transactions of the Royal Entomological Society of London, 1951. **102**(pt. 7).
159. Shaikevich, E.V., *PCR-RFLP of the COI gene reliably differentiate Culex pipiens, Cx. pipiens f. molestus, and Cx. torrentium of the Pipiens complex.* Europian Mosquito Bulletin, 2007. **23**: p. 25-30.
160. Bahnck, C.M. and D.M. Fonseca, *Rapid assay to identify the two genetic forms of Culex (Culex) pipiens L. (Diptera: Culicidae) and hybrid populations.* Am J Trop Med Hyg, 2006. **75**(2): p. 251-5.

161. Cornel, A., et al., *Culex pipiens sensu lato in California: a complex within a complex?* J Am Mosq Control Assoc, 2012. **28**(4 Suppl): p. 113-21.
162. Kothera, L., et al., *A comparison of above-ground and below-ground populations of Culex pipiens pipiens in Chicago, Illinois, and New York City, New York, using 2 microsatellite assays.* J Am Mosq Control Assoc, 2012. **28**(4 Suppl): p. 106-12.
163. Berlocher, S.H., *Can sympatric speciation via host or habitat shift be proven from phylogenetic and biogeographic evidence.* Endless forms: species and speciation, 1998. **99**: p. 113.
164. Vinogradova, E.B., *Ecophysiological and morphological variations in mosquitoes of the Culex pipiens complex (Diptera:Culicidae).* Acta Soc. Zool. Bohem., 2003. **67**: p. 41-50.
165. Kim, S., S. Trocke, and C. Sim, *Comparative studies of stenogamous behaviour in the mosquito Culex pipiens complex.* Med Vet Entomol, 2018. **32**(4): p. 427-435.
166. Dumas, E., et al., *Population structure of Wolbachia and cytoplasmic introgression in a complex of mosquito species.* BMC Evol Biol, 2013. **13**: p. 181.
167. Cornel, A.J., et al., *Differences in extent of genetic introgression between sympatric Culex pipiens and Culex quinquefasciatus (Diptera: Culicidae) in California and South Africa.* J Med Entomol, 2003. **40**(1): p. 36-51.
168. Khrabrova, N.V., et al., *The distribution of strains of endosymbiotic bacteria Wolbachia pipientis in natural populations of Culex pipiens mosquitoes (Diptera: Culiciadae).* Europian Mosquito Bulletin, 2009. **27**: p. 18-22.
169. Miller, W.J., L. Ehrman, and D. Schneider, *Infectious speciation revisited: impact of symbiont-depletion on female fitness and mating behavior of Drosophila paulistorum.* PLoS Pathog, 2010. **6**(12): p. e1001214.
170. Garrigan, D., et al., *Genome sequencing reveals complex speciation in the Drosophila simulans clade.* Genome Res, 2012. **22**(8): p. 1499-511.
171. Doellman, M.M., et al., *Genomic Differentiation during Speciation-with-Gene-Flow: Comparing Geographic and Host-Related Variation in Divergent Life History Adaptation in Rhagoletis pomonella.* Genes (Basel), 2018. **9**(5).
172. Doellman, M.M., et al., *Geographic and Ecological Dimensions of Host Plant-Associated Genetic Differentiation and Speciation in the Rhagoletis cingulata (Diptera: Tephritidae) Sibling Species Group.* Insects, 2019. **10**(9).
173. Nadeau, N.J., et al., *Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing.* Philos Trans R Soc Lond B Biol Sci, 2012. **367**(1587): p. 343-53.
174. Nadeau, N.J., et al., *Genome-wide patterns of divergence and gene flow across a butterfly radiation.* Mol Ecol, 2013. **22**(3): p. 814-26.
175. Lawniczak, M.K., et al., *Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences.* Science, 2010. **330**(6003): p. 512-4.
176. Favia, G., et al., *Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of Anopheles gambiae s.s.* Insect Mol Biol, 2001. **10**(1): p. 19-23.
177. Coetzee, M., et al., *Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex.* Zootaxa, 2013. **3619**: p. 246-74.
178. Coluzzi, M., et al., *A polytene chromosome analysis of the Anopheles gambiae species complex.* Science, 2002. **298**(5597): p. 1415-8.

179. Tripet, F., et al., *DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of Anopheles gambiae.* Mol Ecol, 2001. **10**(7): p. 1725-32.
180. Diabaté, A., et al., *Spatial swarm segregation and reproductive isolation between the molecular forms of Anopheles gambiae.* Proc Biol Sci, 2009. **276**(1676): p. 4215-22.
181. Sawadogo, P.S., et al., *Swarming behaviour in natural populations of Anopheles gambiae and An. coluzzii: review of 4 years survey in rural areas of sympatry, Burkina Faso (West Africa).* Acta Trop, 2014. **132 Suppl**: p. S42-52.
182. Pennetier, C., et al., *"Singing on the wing" as a mechanism for species recognition in the malarial mosquito Anopheles gambiae.* Curr Biol, 2010. **20**(2): p. 131-6.
183. Kothera, L., et al., *Bloodmeal, Host Selection, and Genetic Admixture Analyses of Culex pipiens Complex (Diptera: Culicidae) Mosquitoes in Chicago, IL.* J Med Entomol, 2020. **57**(1): p. 78-87.
184. Gomes, B., et al., *Asymmetric introgression between sympatric molestus and pipiens forms of Culex pipiens (Diptera: Culicidae) in the Comporta region, Portugal.* BMC Evolutionary Biology, 2009. **9**(1): p. 262.
185. Marsden, C.D., et al., *Asymmetric introgression between the M and S forms of the malaria vector, Anopheles gambiae, maintains divergence despite extensive hybridization.* Mol Ecol, 2011. **20**(23): p. 4983-94.
186. Gomes, B., et al., *Asymmetric introgression between sympatric molestus and pipiens forms of Culex pipiens (Diptera: Culicidae) in the Comporta region, Portugal.* BMC Evol Biol, 2009. **9**: p. 262.
187. Gomes, B., et al., *Asymmetric introgression between sympatric molestus and pipiens forms of Culex pipiens (Diptera: Culicidae) in the Comporta region, Portugal.* BMC Evol Biol, 2009. **9**: p. 262.
188. Nelms, B.M., et al., *Phenotypic variation among Culex pipiens complex (Diptera: Culicidae) populations from the Sacramento Valley, California: horizontal and vertical transmission of West Nile virus, diapause potential, autogeny, and host selection.* Am J Trop Med Hyg, 2013. **89**(6): p. 1168-78.
189. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
190. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.* Bioinformatics, 2011. **27**(21): p. 2987-93.
191. Danecek, P., et al., *The variant call format and VCFtools.* Bioinformatics, 2011. **27**(15): p. 2156-8.
192. Zheng, X., et al., *A high-performance computing toolset for relatedness and principal component analysis of SNP data.* Bioinformatics, 2012. **28**(24): p. 3326-8.
193. Alexander, D.H., J. Novembre, and K. Lange, *Fast model-based estimation of ancestry in unrelated individuals.* Genome Res, 2009. **19**(9): p. 1655-64.
194. Weir, B.S. and C.C. Cockerham, *Estimating F-Statistics for the Analysis of Population Structure.* Evolution, 1984. **38**(6): p. 1358-1370.
195. Catchen, J., et al., *Stacks: an analysis tool set for population genomics.* Mol Ecol, 2013. **22**(11): p. 3124-40.
196. Simonsen, M., Mailund, T. & Christian, N. S., *In Proceedings of the 8th Workshop in Algorithms in Bioinformatics (WABI), LNBI 5251*

*113–122.* Springer Verlag, 2008.

197. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.* Bioinformatics, 2014. **30**(9): p. 1312-3.
198. Darriba, D., et al., *jModelTest 2: more models, new heuristics and parallel computing.* Nature Methods, 2012. **9**(8): p. 772-772.
199. *R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria,.* 2019.
200. Rossi, S.L., T.M. Ross, and J.D. Evans, *West Nile virus.* Clin Lab Med, 2010. **30**(1): p. 47-65.
201. Lanciotti, R.S., et al., *Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States.* Science, 1999. **286**(5448): p. 2333-7.
202. Brooke, B.D., *kdr: can a single mutation produce an entire insecticide resistance phenotype?* Trans R Soc Trop Med Hyg, 2008. **102**(6): p. 524-5.
203. Sejvar, J.J., *West nile virus: an historical overview.* Ochsner J, 2003. **5**(3): p. 6-10.
204. Apperson, C.S., et al., *Host feeding patterns of established and potential mosquito vectors of West Nile virus in the eastern United States.* Vector-Borne and Zoonotic Diseases, 2004. **4**(1): p. 71-82.
205. Kilpatrick, A.M., et al., *West Nile virus risk assessment and the bridge vector paradigm.* Emerging infectious diseases, 2005. **11**(3): p. 425.
206. Smith, J.L. and D.M. Fonseca, *Rapid assays for identification of members of the Culex (Culex) pipiens complex, their hybrids, and other sibling species (Diptera: culicidae).* Am J Trop Med Hyg, 2004. **70**(4): p. 339-45.
207. Farajollahi, A., et al., *"Bird biting" mosquitoes and human disease: A review of the role of Culex pipiens complex mosquitoes in epidemiology.* Infection, Genetics and Evolution, 2011. **11**(7): p. 1577-1585.
208. Russell, R.C., *A review of the status and significance of the species within the Culex pipiens group in Australia.* J Am Mosq Control Assoc, 2012. **28**(4 Suppl): p. 24-7.
209. Liu, B., et al., *The potential distribution and dynamics of important vectors Culex pipiens pallens and Culex pipiens quinquefasciatus in China under climate change scenarios: an ecological niche modelling approach.* Pest Manag Sci, 2020. **76**(9): p. 3096-3107.
210. Kent, R.J., L.C. Harrington, and D.E. Norris, *Genetic differences between Culex pipiens f. molestus and Culex pipiens pipiens (Diptera: Culicidae) in New York.* J Med Entomol, 2007. **44**(1): p. 50-9.
211. Strickman, D. and D.M. Fonseca, *Autogeny in Culex pipiens complex mosquitoes from the San Francisco Bay Area.* Am J Trop Med Hyg, 2012. **87**(4): p. 719-26.
212. Byrne, K. and R.A. Nichols, *Culex pipiens in London Underground tunnels: differentiation between surface and subterranean populations.* Heredity, 1999. **82**(1): p. 7-15.
213. Martínez-de la Puente, J., et al., *Culex pipiens forms and urbanization: effects on blood feeding sources and transmission of avian Plasmodium.* Malaria Journal, 2016. **15**(1): p. 589.
214. Yurchenko, A.A., et al., *Genomic differentiation and intercontinental population structure of mosquito vectors Culex pipiens pipiens and Culex pipiens molestus.* Sci Rep, 2020. **10**(1): p. 7504.
215. Nuismer, S.L., J.N. Thompson, and R. Gomulkiewicz, *Coevolutionary clines across selection mosaics.* Evolution, 2000. **54**(4): p. 1102-15.

216. Hoffmann, A.A., C.M. Sgrò, and A.R. Weeks, *Chromosomal inversion polymorphisms and adaptation.* Trends Ecol Evol, 2004. **19**(9): p. 482-8.
217. King, M., *Species evolution: the role of chromosome change*. 1995: Cambridge University Press.
218. Rieseberg, L.H., *Chromosomal rearrangements and speciation.* Trends Ecol Evol, 2001. **16**(7): p. 351-358.
219. Hoffmann, A.A. and L.H. Rieseberg, *Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation?* Annu Rev Ecol Evol Syst, 2008. **39**: p. 21-42.
220. Kirkpatrick, M. and N. Barton, *Chromosome Inversions, Local Adaptation and Speciation.* Genetics, 2006. **173**(1): p. 419-434.
221. Lee, C.-R., et al., *Young inversion with multiple linked QTLs under selection in a hybrid zone.* Nature Ecology & Evolution, 2017. **1**(5): p. 0119.
222. Coluzzi, M., et al., *Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex.* Trans R Soc Trop Med Hyg, 1979. **73**(5): p. 483-97.
223. Harewood, L., et al., *Hi-C as a tool for precise detection and characterisation of chromosomal rearrangements and copy number variation in human tumours.* Genome Biol, 2017. **18**(1): p. 125.
224. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.
225. Dekker, J., et al., *Capturing chromosome conformation.* Science, 2002. **295**(5558): p. 1306-11.
226. Demin, S., et al., *High-resolution mapping of interstitial telomeric repeats in Syrian hamster metaphase chromosomes.* Cytogenetic and genome research, 2011. **132**(3): p. 151-155.
227. Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, *NIH Image to ImageJ: 25 years of image analysis.* Nat Methods, 2012. **9**(7): p. 671-5.
228. Belaghzal, H., J. Dekker, and J.H. Gibcus, *Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation.* Methods, 2017. **123**: p. 56-65.
229. Durand, N.C., et al., *Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom.* Cell Syst, 2016. **3**(1): p. 99-101.
230. Harbach, R.E., B.A. Harrison, and A.M. Gad, *Culex (Culex) molestus Forskal (Diptera: Culicidae): neotype designation, description, variation, and taxonomic status.* Proc Entomol Soc Wash, 1984. **86**(3): p. 521-542.
231. Mattingly, P. *The problem of biological races in the Culex pipiens complex*. in *Proceedings of the Linnean Society of London*. 1952. Oxford University Press.
232. Dobzhansky, T. and T.G. Dobzhansky, *Genetics of the evolutionary process*. Vol. 139. 1971: Columbia University Press.
233. Kirkpatrick, M. and N. Barton, *Chromosome inversions, local adaptation and speciation.* Genetics, 2006. **173**(1): p. 419-34.
234. Ayala, D., et al., *Association mapping desiccation resistance within chromosomal inversions in the African malaria vector Anopheles gambiae.* Mol Ecol, 2019. **28**(6): p. 1333-1342.

235. Fouet, C., et al., *Adaptation to aridity in the malaria mosquito Anopheles gambiae: chromosomal inversion polymorphism and body size influence resistance to desiccation.* PLoS One, 2012. **7**(4): p. e34841.
236. Coluzzi, M., et al., *Chromosomal differentiation and adaptation to human environments in the Anopheles gambiae complex.* Transactions of The Royal Society of Tropical Medicine and Hygiene, 1979. **73**(5): p. 483-497.
237. D'Alessandro, G., G.F. Lazzaro, and M. Mariani, *Effect of DDT selection pressure on the frequency of chromosomal structures in Anopheles atroparvus.* Bull World Health Organ, 1957. **16**(4): p. 859-64.
238. Unger, M.F., et al., *A standard cytogenetic map of Culex quinquefasciatus polytene chromosomes in application for fine-scale physical mapping.* Parasit Vectors, 2015. **8**: p. 307.
239. Vogels, C.B.F., et al., *Vector competence of northern European Culex pipiens biotypes and hybrids for West Nile virus is differentially affected by temperature.* Parasites & Vectors, 2016. **9**(1): p. 393.
240. Arensburger, P., et al., *The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs.* BMC Genomics, 2011. **12**: p. 606.
241. Andreadis, T.G., *The contribution of Culex pipiens complex mosquitoes to transmission and persistence of West Nile virus in North America.* J Am Mosq Control Assoc, 2012. **28**(4 Suppl): p. 137-51.
242. Aardema, M.L., et al., *Global evaluation of taxonomic relationships and admixture within the Culex pipiens complex of mosquitoes.* Parasites & Vectors, 2020. **13**(1): p. 8.