Chromosomal evolution in mosquitoes – vectors of diseases

Anastasia N. Naumenko

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

Sally Paulson
Maria Sharakhova
Daniela Cimini
Ksenia Onufrieva
Zhijian Tu
Paul Marek

May 10th, 2017
Blacksburg, Virginia

Keywords: mosquitoes, chromosomes, genome mapping, evolution

Chromosomal evolution in mosquitoes – vectors of diseases

Chromosomal evolution in mosquitoes – vectors of diseases

Anastasia N. Naumenko

ACADEMIC ABSTRACT

The World Health Organization estimates that vector-borne diseases account for 17% of the global burden of all infectious diseases and has identified the mosquito as the most dangerous of all disease-transmitting insects, being responsible for several million deaths and hundreds of millions of cases each year. The study of mosquito genomics provides a deeper understanding of the molecular mechanisms involved in every aspect of vector biology, such as sex determination, host-parasite interaction, ecology, feeding behavior, immunity and evolutionary trends and can be used for the development of new strategies for vector control.

We developed the first map of the mitotic chromosomes of the major vector for West Nile fever and lymphatic filariasis, *Culex quinquefasciatus*. The map was then successfully utilized for mapping of approximately 90% of available genetic markers to their precise positions on the chromosomes. Idiograms were integrated with 140 genetic supercontigs representing 26.5% of the genome. A linear regression analysis demonstrated good overall correlation between the positioning of markers on physical and genetic linkage maps. This will improve gene annotation and help in distinguishing potential haplotype scaffolds and regions of segmental duplications. It will also facilitate identification of epidemiologically important genes that can be used as targets for the vector control and provide a better framework for comparative genomics that will help understanding of the evolution of epidemiologically important traits.

In another study, we confirmed the presence of the newly described species, *Anopheles daciae*, in regions of Russia using molecular data. Although sympatric with its sibling species, *Anopheles messeae*, five nucleotide substitutions in the internal transcribed spacer 2 of ribosomal DNA can be used to distinguish the morphologically similar species. Chromosome rearrangements have a significant impact on mosquito adaptation and speciation. Using sequencing data in combination with karyotyping, we demonstrated that significant differences in inversion frequencies distinguish *An. messeae* from *An. daciae*, suggesting that these inversions are actively involved in adaptation and speciation. It is essential to have reliable toolbox for correct identification of these species and to know their range for future possible malaria outbreaks prevention.

Chromosomal evolution in mosquitoes – vectors of diseases

Anastasia N. Naumenko

GENERAL AUDIENCE ABSTRACT

The more you study, the more you know
The more you know, the more you forget
The more you forget, the less you know
So why study?

According to the World Health Organization, mosquitoes are one of the deadliest animals in the world. They spread disease to humans resulting in hundreds of millions of illnesses and several million deaths every year. Study of the mosquito genome can help us understand vector biology and speciation and can be used to develop new strategies for vector control.

Culex quinquefasciatus, the southern house mosquito, is one of the major vectors for the West Nile virus in the U.S. and for lymphatic filariasis, a disabling and disfiguring disease, worldwide. The traditional methods of control are of limited effectiveness because of high insecticide resistance in many populations of the mosquito. To enhance our resources for the control strategies, we developed physical maps of the chromosomes for this mosquito and effectively integrated it with available genetic linkage map. This work will help to identify epidemiologically important genes that can be used as targets for the vector control.

Malaria vectors, mosquitoes from the genus *Anopheles*, are known for their ecological plasticity, which can be partially explained by chromosome rearrangements called inversion. A global malaria eradication program significantly reduced the number of deaths related to malaria, especially in Europe and the U.S. However, malaria outbreaks can occur anywhere competent vectors occur. We studied *Anopheles messeae*, one of the major European malaria vectors and its closely related species, *Anopheles danciae*. We report for the first time the presence of *An. daciae* in Russia and demonstrate that its distribution overlaps with that of *An. messeae*. Using genetic sequence data in combination with chromosome structure, we demonstrated that significant differences in inversion frequencies reliably distinguish *An. messeae* from *An. daciae*. These inversions may be involved in adaptation and speciation of these two species. It is essential to have reliable toolbox for correct identification of these species and to know their range for future possible malaria outbreaks prevention.

Dedication

I dedicate this work to my parents, Nikolay and Zinaida, and to Nikita.

Acknowledgements

I would like to thank all people who inspired and taught me during my PhD. First of all, I would like to thank my committee chair, Dr. Paulson, and my advisor, Dr. Maria Sharakhova, for their knowledgeable advices, support at every step of the PhD, reading and editing the manuscript. I am very grateful to Dr. Maria Sharakhova and Dr. Igor Sharakhov for the opportunity to work in their lab on many great projects. That was a true pleasure to be a part of your group.
I am very grateful to my committee committee members: Dr. Daniela Cimini, Dr. Paul Marek, Dr. Ksenia Onufrieva, and Dr. Jake Tu. You supported me on each and every step of my studies. Your experiences and scientific knowledge bring true enthusiasm for science.

I am very thankful to my former lab mate, Dr. Vladimir Timoshevskiy, now University of Kentucky affiliate, for his sense of humor, valuable advice and teaching me how to perform FISH. I always remember his patience and kindness.
I would like to thank my labmates, now-doctors:
Dr. Philip George, Dr. Maryam Kamali, Dr. Atashi Sharma, Dr. Ashley Peery, Dr. Nick Kinney; not-yet-doctors : Jiyoung Lee, Reem Masri , Jiangtao Liang; and my undergraduate student Jeremy Jenrette – for their company and moments we shared together.

I am thankful to Dr. Gleb Artemov and Dr. Alina Kokhanenko from Tomsk State University, Russia – it is a pleasure to work with you.
Many thanks to Dr. Mikhail Gordeev and Dr. Anton Moskaev from Moscow Pedagogical State University for providing me with karyotyping data for samples of *An. messeae* and for help with manuscript editing. Thanks to Dr. Andrey Yurchenko for his help with genome data analysis.
I would also like to mention Dr. Dave Severson, who provided BAC clone library for *Cx. quinquefasciatus* physical mapping, and Becky deBruyn, who screened BAC clones for us.

Table of contents

Chapter 4. Chromosome inversions and speciation in the European malaria vector *Anopheles messeae s.l.*

Chapter 1. Introduction

1.1.Chromosomal evolution in mosquitoes

1.1.1. Mosquitoes: valued insects for studying evolution of genomes

"Nothing in biology makes sense except in the light of evolution"
Theodosius Dobzhansky, 1973

Over one hundred and eighty five years ago, her Majesty's ship *Beagle*, led by Captain Fitz Roy, R.N., sailed from Devonport. Among others, a young naturalist, Charles Robert Darwin, was aboard. Beagle sailed around South America, stopping by the Galapagos Islands (now part of Ecuador), and returned via Tahiti and Australia. During this five-year surveillance trip, Darwin collected a number of zoological, botanical and geological specimens and, far more important, his own observations of nature. While on the Galapagos, Darwin noted differences in the morphology of finches. He later discovered 13 varieties of birds were present on the islands. Comparing birds' behavior and other traits, Darwin realized that 13 varieties were descendants of one common ancestor, which spread throughout the islands and adapted to different habitats and foods, acquiring new traits. He was then the first to introduce the term "transmutation by means of natural selection", which later transformed into "evolution" [1].

Shortly after the point when the term "evolution" was introduced, naturalists observed the genome of organisms – chromosomes - using relatively simple cytological methods. Even at this level of observation, it was noted that organisms have extensive variation in number and morphology of the chromosomes. Dramatic differences in the karyotypes were found in closely related species, and chromosomal set was shown to vary between males and females of the same species [2]. It was soon demonstrated that chromosomes inherit (Boveri-Sutton theory). Modifications of chromosomes, and the consequences and causes of chromosomal evolution were a long-standing interest to biologists. Theodosius Dobzhansky in his classic work *Genetic basis for evolution* focused on the role of chromosomal rearrangements, particularly inversions,

in evolutionary process [3]. However, the main direction in evolutionary biology soon shifted to genes and their function while chromosomal evolution fell out of focus for decades.

A new era in molecular biology aided by methods such as new high-throughput genome sequencing, enabled scientists to investigate molecular features of chromosomal structure and functions: how chromatin, centromeres and telomeres are organized; which events are evolved in chromosomal rearrangements; and how chromosomes recombine during meiosis. Our understanding of molecular mechanisms underlying chromosomal functionality led back to the classic questions of evolutionary genetics but now in perspective of molecular cytogenetics: how do different karyotypes evolve? What are the consequences of chromosomal rearrangements and whether chromosomal rearrangements play the role in speciation?

Mosquitoes are of particular interest in terms of chromosomal evolution. Despite large divergence times between clades, all the species have a karyotype except one made of three chromosomes [4-6]. However, the size of chromosomes, as well as size of the genomes, is dramatically different between species. Clade Anophelinae has smaller chromosomes with heteromorphic sex chromosomes X and Y, whereas the Culicinae possess generally larger chromosomes and lack differences between male and female karyotypes (homomorphic sex chromosomes). The most frequent chromosomal rearrangements in mosquitoes, unlike in Drosophila, are inversions [7]. In *Drosophila*, fissions and fusions are predominant [8]. In addition, the role that mosquitoes have in transmission of vector-borne diseases makes them an important group for genome sequencing and studying genomic evolution: introduction of genome-editing tools may aid in the suppression of mosquito populations and/or reduce the rates of disease transmission [9, 10].

1.1.2. Classic approaches in studying mosquito evolution

Genetic linkage maps

Genetics has been a foundation in mosquito research for a long time [11, 12].
Similar to *Drosophila* genetics, most of the pioneer studies included observations of frequencies and heredity of mutant phenotypes [13]. Various mutants have been described, primarily for *Aedes* and *Culex* species. The first sex-linked mutant, *white eye*, was described for *Cx. tritaeniorhynchus* [12]. A number of mutant phenotypes have been found in the *Cx. pipiens* mosquito complex, both autosomal and sex-linked. For example, the autosomal recessives, *curved larval antennae* and *enlarged tergum* in pupae were studied by A. Barr [14], A. Barr and S. Narang [15] and the sex-linked recessives *divided eye* and *female lethal* were described by A. Barr [16]. *Female lethal* is transmitted by males by outcrossing and eliminates almost all female progeny before hatching, producing male-only egg rafts. In *Cx. tarsalis*, *white eye* and *yellow larva* mutants have been described as autosomal and sex-linked recessives, respectively. An unusual finding was that the *white eye* mutation was not either sex-linked or allelic with a similar mutation previously found in *Cx. tritaeniorhynchus* [14].

In *Aedes*, hundreds of mutants have been investigated [17]. A phenotype of particular interest, *bronze*, has been identified by S. Bhalla and G. Craig [18]. This sex-linked recessive allele controls the bronze shade of dark parts in egg shells, larvae, pupae and adults. The males with this phenotype are fully fertile; however, females are sterile and the stock can only be maintained if heterozygous females are used in crosses with homozygous males.

Mutations described above as well as many others were successfully used as genetic markers in crosses and led to the development of first genetic linkage maps for mosquitoes [13]. A number of studies focused on the alleles for insecticide resistance [19-21]; one of the first linkage studies on *Anopheles gambiae* was done for DDT and dieldrin resistance alleles [22].

Cytogenetics

The great achievement in mosquito genetics was the discovery that the salivary glands of the Anophelinae clade contain polytene chromosomes – large chromosomes build by thousands of DNA copies [23]. The polytene chromosomes of *An. gambiae* species complex have been shown to demonstrate reproducible banding patterns and were therefore suitable for cytotaxonomy – proper distinction and identification of phylogenetic relationships in the group based on the chromosomes. The studies by M. Coluzzi and A. Sabatini [24] demonstrated that the morphologically indistinguishable sibling species of the *An. gambiae* complex could be identified based on the banding patterns of their polytene chromosomes. Most of the first cytogenetic maps were drawn for *Anopheles* species. *Culex* and *Aedes* had lengthy polytene chromosomes that posed a significant challenge in terms of slide preparation. However, in 1968 L. Dennhofer published the first cytogenetic map for *Cx. pipiens* [25]. The first detailed studies of polytene chromosomes in *Anopheles* species revealed different levels of inversion polymorphism: some species had many inversions whereas others lacked polymorphism. Few inversions were found on X chromosomes. For instance, *An. albimanus* lacked heterozygosity [26]; moderate numbers of inversions were found in *An. atropos* [26] and *An. messeae* [27], and elevated frequencies of inversions were shown for *An. darlingi* [28], *An. freeborni* [29], *An. gambiae* [30]and *An. stephensi* [31]. The early studies of inversion polymorphisms were used to trace relationships between species; however, the information that could be derived from karyotypes was fragmented and incomplete. The missing evidence of phylogenetic relationships was found much later, when the information on genes and genomes became widely accessible.

1.1.3. Genomics and post-genomics era in mosquito research

Traditional methods enabled initial broad understanding of mosquito biology, starting from morphology-based phylogenies to cytogenetics, insecticide resistance studies and karyotype-based phylogenies [13]. Soon, new information was needed to control vector-borne diseases. Sequencing technologies such as DNA sequencing, transcriptome profiling, and RNA expression arrays, offered deeper access to understand vector biology [32]. Mosquitoes as vectors for many diseases became candidates for the whole-genome sequencing projects. The primary role of genome sequencing was to identify approaches for the genetic control of species involved in the transmission of malaria, dengue, filariasis, and other arthropod-borne diseases.

In 2002, a breakthrough was made in mosquito genomics: the genome of one of the most efficient malaria vectors and major *Plasmodium falciparum* vector in Africa, *An. gambiae*, was published [33]. The shotgun sequencing covered approximately 91% of the genome. Expansions in gene families involved in cell adhesion and immunity were found. This assembly was validated by physical mapping of 2000 BAC clones to the precise positions on polytene chromosomes. This publication established a platform for further comparative genomics, studies on gene function, genomic landscapes and evolution [34].

Following the release of *An. gambiae* genome, the genomes of *Ae. aegypti*, vector for dengue and yellow fevers, and *Cx. quinquefasciatus*, vector for filariasis and West Nile fever, have been published [35, 36]. Based on Sanger sequencing technology, the efforts were time- and labor-intensive. However, the publication of annotated genomes further facilitated mosquito comparative genomics [37].

In 2007, the largest web-accessible data repository for vector genomes including mosquitoes, VectorBase, was released [38]. The initial database included information on two organisms: *An. gambiae*, vector for *P. falciparum*, and *Aedes aegypti*, a mosquito vector for flaviviruses causing yellow fever and dengue fever. VectorBase handles genomics data at all stages, starting from initial gene annotation to re-annotation and NCBI submission [37, 38]. The platform uses simple yet effective Genome Model Organism Database construction set to store information on

genomes and annotations [39]. Multiple tools have been added to VectorBase in the decade following its release, such as BioMart [40] population biology tools for storing genomic variations, expression profiles and associated metadata [37].

Molecular cytogenetics

The availability of genome data enabled molecular studies of the karyotype. Molecular cytogenetics combines molecular genetics with the cytogenetics. Fluorescent *in situ* hybridization (FISH) technique is a powerful research tool to address a range of karyotype evolution studies. FISH includes labelling of a DNA probe with fluorochrome conjugated nucleotides, denaturation of both DNA and the probe, reannealing of the probe with chromosomal preparation on a slide, removal of excessive probe, and image acquisition [41]. FISH is a useful approach to study complex chromosomal rearrangements, such as inversions, inversion breakpoints, translocations and deletions, at much higher resolution than standard G-banding analysis. Molecular cytogenetics tools are widely used in production of high-resolution physical maps of mosquito genomes.

NGS (Next Generation Sequencing) has contributed greatly to the accessibility and affordability of insect genomes [32]. The method enables deeper depth and larger coverage of the genomes along with the increased speed of the sequencing itself. In less than 10 years, the time gap between start of the genome sequencing and publication of the annotation genome has been dramatically reduced. In 2017, the VectorBase genome storage included annotated datasets on 60 arthropod species along with data on population genomics [42].

In 2015, the genomes of additional 16 *Anopheles* species, all major malaria vectors, were published [40]. This accomplishment enabled fine-scale studies of evolutionary relationships among these species. FISH and bioinformatics tools were used to identify chromosomal rearrangements with a principal finding that the synteny of the whole-arm is highly conserved despite several whole-arm translocations. Similar to *Drosophila*, the rearrangement rates are higher on the X chromosome in all species, but, unlike *Drosophila*, the rearrangement frequency on X chromosomes is, on average, 2.7 times higher than on autosomes [7]. This fine-scale

phylogenetic study revealed that two major African malaria vectors, *An. gambiae* and *An. arabiensis*, diverged before other species [7, 32]

In cases when the annotated genome is absent, RNA sequencing technologies (Illumina RNA-seq) provide sufficient information for the genome-wide analysis. For example, transcriptome assembly enabled comparative genomics for *An. funestus* even though the genome was not yet accessible [43]. Several annotated transcriptomes are available through VectorBase (*An. freeborni*, *An. quadrimaculatus*) [42]. Time- and cost-effective transcriptomes can be used along with genomes in comparative studies on gene shuffling and gene families.

Multiple previously unknown aspects of mosquito biology were discovered in parallel with the development of new molecular tools. In particular, population genomics became easily accessible. For instance, two forms of *Cx. pipiens*, wild *Cx. pipiens pipiens* and domesticated *Cx. pipiens molestus*, were compared and genetic mechanisms involved in adaptations were highlighted by calculating the divergence rates [44]. Another salient example is identification of components involved in the ecological plasticity in M and S forms of *An. gambiae* [45].
The post-genomics era in mosquito research is endorsed by multiple sources of information. The developing opportunities for fast NGS or WGA sequencing approaches along with RNA-seq create metadata flow that is organized into web-accessible data blocks.

1.1.4. Chromosomal rearrangements and their role in mosquito speciation

Organisms evolve differently because of multiple factors such as adaptation to various environments and food sources, geographic isolation and spontaneous mutations followed by natural selection. Dobzhanskiy stated that “species is a stage in a process, not a static unit”[3]. As constantly changing units of evolution, karyotypes of species are of particular interest because chromosomal rearrangements may contribute to speciation in different ways.

Major chromosomal rearrangements include inversions, translocations, and Robertsonian fusions and fissions. The suppressed-recombination model, proposed by Coluzzi in 1982 [8], is based on the evidence of speciation within the *An. gambiae* species complex. The main African malaria

vector, *An. gambiae*, dominates in rain forests whereas its close relative, *An. arabiensis*, leans towards dryer habitats. The seven species from the *An. gambiae* complex are morphologically similar but are mutually exclusive in range with only small areas of overlap. The species have differences in feeding (zoophilic versus antropophilic), breeding sites and can be distinguished by egg morphology and chromosomal patterns, primarily inversion polymorphisms on second and third chromosomes. Within the *An. gambiae* complex, *An. arabiensis* was believed to be ancestral because of the geographic evidence of Middle-East origin and fixed 2La inversion [46]. Based on chromosomal rearrangements and multigene phylogenies, this evolutionary history has been recently revised [7, 47]. Based on the unique origin of 2La inversion and taking chromosomal rearrangements into consideration, *An. gambiae, An. arabiensis* or *An. merus* were possibly the closest to the ancestral species. The ancestral species could have two polymorphic inversions, 2Rp and 2R+^{o}. As speciation progressed, one of the lineages gave rise to *An. merus* with 2Rp inversion and the other lineage, 2R+^{o}, to *An. gambiae*. *An. gambiae* acquired 2La/+ inversion after entering forest habitats in Central Africa. Multiple inversions on 2R chromosome led to *An. gambiae* expanding its range into arid West Africa. Later, through multiple X chromosome rearrangements and the fixation of 2La inversion, *An. arabiensis* evolved from the common ancestor. The fixation of 2L+a inversion gave rise to *An. quadriannulatus*. *An. bwambae* and *An. melas* originated as 3La and 2Rm inversions were acquired.

Out of the listed species, *An. gambiae* has the widest range and is highly antropophilic. Along with *An. arabiensis, An. gambiae* is held accountable for the majority of malaria cases in Africa [45, 48].

The inversion polymorphism exists within *An. gambiae* species. Three chromosomally distinct populations, Savanna, Mopti and Bamako, are different in ecotype preferences and breeding sites; yet enough assortative mating suggests that these species are incipient. Based on the evidence for inversion polymorphism and fixed SNPs, the populations partially overlap with M and S molecular forms [45, 49-51]. The S form is referred to as separated species, *An. coluzzi*, by many scientists [52, 53].

The patterns of chromosomal evolution in the *An. gambiae* complex are explained well by the suppression-recombination model [30]. According to the model, the chromosomally homomorphic population colonizes a favorable range and expands. As the numbers increase, the

population reaches the periphery of the range and faces extreme conditions such as low or high temperatures, dryness or humidity, and seasonal climate fluctuations. The mutations that increase adaptations will accumulate in the subpopulations at the periphery of the range, creating new ecotypes that are adapted to new environmental conditions. The mating between mosquitoes from subpopulations with mosquitoes from central population will eventually homogenize or even eradicate the newly acquired alleles, except for those protected by chromosomal inversions, where the recombination is suppressed. The mutations will further be accumulated within the inversions, making grounds for subsequent reproductive isolation. The process of incipient speciation produces new ecotypes that are better adapted to various environments and wide ranges of habitats. The *Anopheles* genus has at least 170 cryptic species belonging to 30 species complexes [54]. Most of the cryptic species evolved as the result of the incipient speciation and crossingover-suppressing paracentric inversions, particularly on chromosome X.

1.2.Advances and limitations in vector control

1.2.1. Overview of mosquito-borne diseases

Rapidly evolving genes and genomes, mouthparts, elegantly designed for blood feeding, and worldwide distribution make mosquito species a perfect vectors for many viruses and parasites. Mosquitoes are responsible for a number of explosive outbreaks of diseases in different regions. Changing climate contributes to distribution of pathogen vectors and changes epidemiology of outbreaks [55].

Malaria

Malaria is one of the best-known mosquito-borne diseases - an acute illness with devastating impacts on the human population [56]. Malaria in humans is caused by five species of *Plasmodium*: *P. falciparum*, the most pathogenetic species and the predominant agent of malaria in Africa, *P. vivax*, the most prevalent human parasite in Asia, *P. ovale*, morphologically and biologically similar to *P. vivax*, *P. malariae*, found worldwide, and *P. knowlesi*, which is responsible for recent zoonotic outbreaks of malaria in Southeast Asia [57]. Of the approximately 430 species of *Anopheles*, only 30-40 species transmit malaria in nature [58]. Anophelines responsible for malaria transmission can be found not only in regions with recent

malaria outbreaks; they are constantly found in areas where malaria has been eliminated. Such areas are under a threat of malaria re-introduction [59].

Despite global efforts by groups such as the WHO and the Bill and Melinda Gates Foundation to eradicate malaria, it continues to cause high incidence of mortality in resource-poor countries. Between 2010 and 2015, malaria incidence fell by 21% globally [60]. The mortality risks have been reduced by 29% in adult population and by 35% in children under 5. Sub-Saharan Africa remains the most affected region of the world: over 90% of new malaria cases are registered here. According to World Health Organization (WHO), in 2015, 91 countries and areas had ongoing malaria transmission. In 2015, there were 212 million new cases of malaria and 429000 malaria deaths worldwide [60]. Two major forms of mosquito control for malaria prevention are insecticide-treated nets and residual indoor spraying.

West Nile fever

West Nile virus (WNV) is an arbovirus from the family Flaviviridae [61]. *Culex* mosquitoes, primarily *Cx. pipiens* and *Cx. quinquefasciatus*, serve as vectors for this virus. The West Nile fever presents in both tropical and temperate regions of the world along the wide range of Culicinae mosquitoes [56]; currently, West Nile virus is the most common in the United States. The virus has its natural reservoir in bird populations, with an enzootic cycle of wild birds and mosquitoes although it can spill over into other vertebrates including humans. WNV is asymptomatic in 80% of the human infections; however, febrile illness develops in the rest and approximately 1% develop severe illness with possible neurological complications [61]. There is no vaccine available for prevention of West Nile fever, which makes it particularly dangerous threat to public health [62].

Lymphatic filariasis

Lymphatic filariasis, in some cases known as elephantiasis, is a neglected tropical disease caused by nematodes from family Filarioidea [56, 63]. Three nematodes are spread by mosquitoes: *Wuchereria bankrofti* is responsible for 90% of the cases, *Brugia malayi* almost 10% and *Brugia timori* is the least common cause of filarasis. The nematodes are spread primarily by *Culex* mosquitoes. Mosquitoes deposit infective L3 larvae that then migrate to lymphatic vessels, where

adult worms develop. Adult filarial worms stay in the lymphatic system, inducing lymphatic dilatation. Progressive lymphatic damage results from the inflammatory reaction to the parasite but also from the inflammatory response to *Wuschereria* bacterial endosymbiont, *Wolbachia* [63] and from accompanying bacterial and fungal infections. Over 947 million people in 54 countries are under the risk of being infected; worldwide, 25 million people with the infection develop hydrocele (the accumulation of fluid around the testes leading to an increase in the volume of scrotum) [64] and 15 million have lymphedema (the immune reaction to the presence of an adult worm, which causes endothelial and connective tissues proliferation, damaged valves and subsequently lymphatic dysfunction) [63]. Both conditions are disfiguring and often incapacitating. At least 36 million people develop chronic disease manifestations [60]. Vector control is a supplemental strategy used by WHO to prevent disease transmission; however, no fully effective measurements exist.

1.2.2. Traditional approaches to mosquito control

Physical control of mosquito populations

Traditional strategies for vector-borne diseases prevention have often focused on the reduction of interactions between human and mosquito through (1) using insecticides to suppress local mosquito populations and (2) using bed nets and sprays to minimalize the risk of human-mosquito exposure [65]. Classic measurements for mosquito control included early and repeated indoor application of insecticides, construction of houses away from mosquito breeding sites, elimination of breeding sites by drainage or repeated usage of larvicides, use of bed nets, window and door nets, and early awareness of new cases of disease transmission in local areas [65].

Repeated application of insecticides, or indoor residual spraying (IRS) works well in combination with other methods such as impregnated bed nets (IBNs) [66]. IBNs serve as a physical barrier between mosquito and human, and reduce human/vector contact. Starting 1980s, cheap and effective pyrethroids were utilized for the bed nets [67]. IBNs along with IRS were used as one of the major steps in malaria prevention. However, the involving resistance traits in

Anopheles in Asia and Africa could compromise the sustainability of vector control strategies [68].

Source reduction

Most mosquitoes breed in standing water; therefore, drainage and removal of water-accumulating reservoirs might be considered as a necessary step in vector population control. Open water marsh management (OWMM), used in both East and West Coasts of the USA, is an alternative to insecticide use. The network of ditches connects marsh to a pond or a canal, enabling access for predators [69].

Rotational impoundment management (RIM) controls the level of water within impounded marsh, preserving the ecosystem itself but preventing mosquitoes from breeding [70].

Biological agents for vector population control have been in use for a long time. Spores of soil bacteria, *Bacillus thuringiensis israelensis*, were found to interfere with larvae digestive system, causing larval elimination [71]. One of the effective biological agents used for mosquito control is the mosquitofish *Gambusia affinis*; however, concerns arose around placing the fish in habitats which are not native [72].

Insecticides and insecticide resistance

Four major classes of insecticides are used for mosquito population control: organochlorines, organophosphates, carbamates and pyrethroids [73]. Twelve insecticides from these classes are recommended for IRS but only pyrethroids can be used for long-lasting insecticidal nets [74]. However, the broad usage of insecticides and heavy chemical application led to the fast adaptation of mosquitoes to insecticides – insecticide resistance. There is evidence of insecticide resistance among major malaria vectors in Africa and in *Cx. pipiens* populations worldwide [75], as well as in *Aedes* [76]. If a mosquito develops resistance to one insecticide, in most cases all the insecticides from the given class become ineffective [74]. The major mechanisms underlying insecticide resistance involve either mutations in insecticide target site or faster detoxification rate of the insecticide [77]. The molecular genetics of insecticide resistance is based on gene amplification (increasing the number of gene copies in the genome [36, 77], increased gene expression for monooxygenase and glutathione –S –transferase-based resistance [78], and gene splicing – production of several enzymes with different activities from only one mRNA.

Despite the evident need for vector control, the current strategies are limited due to high diversity of mosquito species and their ability to quickly adapt to changing environment. All the methods listed in this section require costly and prolonged application and mosquito population control. The development of insecticide resistance in mosquito populations requires the design of new strategies for efficient vector control and prevention of vector-borne diseases.

1.2.3. Post-genomics era and genetic advances to vector control

Recent advances in genomics of disease vectors established grounds for identification of genes involved in insecticide resistance and new targets for insecticides – and the information was not limited to insecticide resistance. The genome projects provided us with a deeper understanding of the molecular mechanisms involved in every aspect of vector biology, such as sex determination, host-parasite interaction, ecology, feeding behavior, immunity and evolution trends. This knowledge can be also useful for the development of new strategies for the vector control.

The idea of mosquito genetic control is not new [79]; however, recently the technology has progressed and greater availability of genomics and proteomics tools has enabled genetic manipulations [32]. Techniques of genetic modification can be used for the reduction of disease transmission at any step of vector-pathogen interaction: vector population control, especially with sterile insect technique (SIT), and genetic engineering of mosquitoes resistant to pathogens in different ways. Several approaches are known as promising in vector control. One of the first approaches is SIT. The method uses the release of sterile males into the wild population in order to decrease the reproductive potential. However, the technique requires constant intervention and monitoring of the population and would be costly [73, 80, 81]. One of the novel approaches is the release of engineered males carrying dominant lethal allele OX513A. Upon release, OX513A males mate with wild females, causing population collapse. This technique has been successfully used by recently founded UK startup Oxitec in several locations; initially at Cayman Islands and later in Brazil and Malaysia [82]. However, this technique also requires the permanent release of modified insects. The other direction is the search for alleles which, when introduced into the population, would transfer the population into refractory state to the pathogen [10]. All of the techniques based on transgene (transgene – a gene or several genes that have been introduced into another organism naturally or by genetic engineering) release have several potential weaknesses. First, the transgene may be stochastically eliminated from the population [82]. Second, the male-only release requires the efficient sex-separation technique. Recent advances in genetic technology, such as development of programmed genome editing tool CRISPR/Cas9

system [83], enabled more effective and heritable introduction of the transgene. CRISPR/Cas9 has been successfully used in *Ae. aegypti* [84-86]. This technique has promising results in driving maleness in the population or creating disease refractory phenotype. Only few individuals are needed to be introduced into the population to create permanent effect [87, 88]. This may pose a potential challenge because of the somewhat unpredictable consequences of allele introduction or elimination in the local population, when even few individuals can establish the gene-drive system in the wild population [89]. However, with efficient testing systems and legislative regulations CRISPR/Cas9 remains the most powerful currently available genome editing tool[90].The use of modern technologies can improve existing vector control techniques in different ways and create novel strategies for mosquito control [91].

1.3.*Culex quinquefasciatus, the southern house mosquito: the main vector of West Nile fever and lymphatic filariasis*

1.3.1. *Culex quinquefasciatus* as diseases vector

Among mosquitoes, species of the genus *Culex* are the most taxonomically diverse and geographically widespread [92]. Mosquitoes within the *Cx. pipiens* complex are major vectors for lymphatic filariasis caused by nematode *Wuchereria bankrofti* in tropical and subtropical regions of Asia, Africa, Central and South America and Pacific Islands. *Cx. quinquefasciatus* is also a major vector for arboviruses, such as West Nile, St. Louis encephalitis, Sindbis and Rift Valley fever viruses. Members of *Cx. pipiens* complex have great variation in their host range, feeding behavior and the diapause habit. The ability to feed on wide range of hosts, such as birds and mammals, makes Culicinae mosquitoes perfect vectors for the transmission of diseases with natural reservoirs, such as West Nile fever.

1.3.2. Genome assembly and physical mapping of *Culex quinquefasciatus*

Genetic linkage mapping has been the most effective method for the genome mapping of *Cx. quinquefasciatus* so far. Among closely related *Cx. pipiens* species several morphological mutants have been described by L. Dennhofer [93]. Crosses involving different mutants permitted assignment of genes related to these mutations to three linkage groups. Use of deoxyribonucleic acid (DNA) markers as a new approach allowed the construction of a genetic map which originally consisted of 21 complementary DNA (cDNA) markers and covered 7.1, 80.4, and 78.3 cM on chromosomes 1, 2, and 3, respectively [94]. The sex determination locus was genetically mapped to the smallest linkage group 1 in *Cx. pipiens.* In addition, multiple quantitative trait loci, related to differences in reproductive diapause between species in the *Cx. pipiens* complex, were also genetically mapped [95]. The most recent genetic linkage map developed for *Cx. quinquefasciatus* includes 63 genetic loci [96]. This map covered 29.5, 88.8, and 65.6 cM on linkage groups 1, 2, and 3 and allowed integration of 10.4% of the genome with the genetic linkage map. However, this map has never been integrated with cytogenetic maps developed for this mosquito before our current study.

As compared to other mosquitoes, *Cx. quinquefasciatus* has the most fragmented genome. A total of 579 Mb is currently assembled into 3 171 supercontigs with the N50 size being ~476 Kb [36]. The N50 supercontig sizes are 12.3 Mb in the *An. gambiae* (PEST) genome and 1.5 Mb in the *Ae. aegypti* genome. Lack of a high-quality chromosome-based genome assembly for *Cx. quinquefasciatus* remains a significant impediment to further progress in *Cx. quinquefasciatus* biology and comparative genomics of mosquitoes. Fragmented unmapped genome assemblies create problems for the genome analysis. Lack of chromosome assignment and orientation of the sequencing contigs complicates study of chromosome organization and evolution [97]. Therefore, the genome sequencing project is incomplete until the majority of supercontigs is assembled and anchored onto *Cx. quinquefasciatus* chromosomes. Physical mapping in *Cx. quinquefasciatus* is challenging because of the poor quality of the polytene chromosomes. Several attempts to create a cytogenetic photomap using *Cx. quinquefasciatus* polytene chromosomes have been made. The Malpighian tubule chromosome map for *Cx. pipiens* [98] and *Cx. quinquefasciatus* [99] and, more recently, the salivary gland chromosome map for *Cx.*

quinquefasciatus [100], were developed. However, correspondence of arms and regions among these maps and the original drawn map published by L. Dennhofer [25] is uncertain. Almost no similarities between landmarks of different chromosome maps were found [100]. These problems occurred because of low levels of polyteny, high frequency of ectopic contacts or associations of nonhomologous chromosome regions, and poor spreading of *Cx. quinquefasciatus* polytene chromosomes in preparation. As a result, only two genes for esterase- and odorant-binding proteins were mapped to the polytene chromosomes of *Cx. quinquefasciatus* [100, 101].

Thus, the genome requires improvement. Indicate here the gap of knowledge that we are trying to fill.

1.4. Inversions and speciation in Palearctic members of Maculipennis group

1.4.1. *Anopheles maculipennis* group of species

Cryptic species complexes are groups of closely related species which are difficult or impossible to distinguish by morphological traits [102]. The complexes are particularly important for arthropod-borne diseases studies because the complex may include highly specialized malaria vectors along with species with lesser susceptibility to *Plasmodium*. Mosquitoes from *Anopheles maculipennis* group of species are known from the early 20th century (Falleroni 1926); sibling species from the complex are effective malaria vectors in Europe and Asia. The Maculipennis complex comprised twelve Palearctic members: *An. atroparvus, An. artemievi, An.beklemishevi, An. daciae, An. labranchiae, An. lewisi, An. maculipennis, An. martinus, An. melanoon, An. messeae s.s.* (further referred as *An. messeae*), *An. persiensis,* and *An. sacharovi* [103, 104]. *An. atroparvus*, *An. sacharovi* and *An. labranchiae* are the most effective malaria vectors in Europe [105]; *An. messeae* was the primary cause of malaria in Russia [106].

1.4.2. Anopheles messeae: vector status and newly described cryptic species, *An. daciae*

An. messeae s.l. has a widespread distribution extending from Ireland across Europe and Asia and into China and Russia [105]. The large range of this species combined with known genetic variability associated with different geographic areas suggests that area-specific biological or behavioral adaptations are likely to have occurred. The newly described species *An. daciae* is different from the closest member of Maculipennis group, *An. messeae,* by egg morphology: the eggs of *An. daciae* are generally darker in color, smaller and have tubercles that are organized in patches of slightly different shape [105]. A PCR assay developed for identification of sibling species from Maculipennis complex [107] including *An. atroparvus, An. beklemishevi, An. labranchiae, An. maculipennis, An. melanoon, An. messeae* and *An. sacharovi*, was not reliable for the identification of *An. daciae* due to the same size of PCR product for the two species. The sequencing of rDNA ITS2 of mosquitoes from Maculipennis group collected near Danube river in Romania by Linton, Nicolescu and Harbach (2004) [105] revealed five single nucleotide polymorphisms in *An. daciae* which differentiated it from the sympatric species, *An. messeae*. Using molecular approach, *An. daciae* has been later discovered in Germany [108, 109], England and Wales [110]. In all locations, *An. daciae* has been found in sympatry with its cryptic species, *An. messeae.*

The species status of *An. daciae* remains unclear. O. Bezzhonova and I. Goryacheva [111] reported in 2008 the presence of the rDNA sequences specific for *An. messeae* and *An. daciae* simultaneously in one individual, as well as high heterogeneity of ITS2 (9 variants described). However, the methodology included using of regular instead of proofreading DNA polymerase, thus increasing number of spontaneous mutations, and subsequent cloning of PCR product into vector prior to sequencing which also could lead to higher rate of mutations in general. The other explanation is that the specimens could have been obtained from an isolated population where the rate of SNPs is higher.

1.4.3. Cytogenetic variability within *An. messeae* populations

Y. Novikov in 1979 [112] postulated the presence of two cytogenetically distinct races within *An. messeae s.l.* populations. He referred to these forms as A and B. Two ways of how chromosomal variations combine in *An. messeae s.l.* populations were found: original variant X0 combined with 2R0, 3R0 and 3L0, inverted X1 tended to combine with 2R1, 3R1 and 3L1. XL0, 2R0, 3R0 and 3L0 were distributed in the west and inverted variants were found in northern and continental parts of the range. Later, he confirmed with restriction fragment length polymorphism assay and taxonprints that two forms, A and B, exist [113]. Restriction of rDNA ITS2 fragment with FokI enzyme was used to distinct A and B forms. However, there was not enough evidence to support the separate species status for each of those forms. At the same time, other authors did not consider chromosomal forms of *An. messeae* to be separate species [106, 114].

In 2014, R. Danabalan [110] supported the species status of *An. daciae* by using slightly different RFLP assay with BStUI enzyme. PCR product resulting from amplification of *An. daciae* DNA was shown to have 4 products after restriction with BStUI whereas digestion of *An. messeae* rDNA ITS2 gave three products. It was not clear, however, whether A and B forms correspond to cryptic *An. messeae* and *An. daciae*.

Chapter 2. A mitotic chromosome-based cytogenetic tool for physical mapping of the *Culex quinquefasciatus* genome

This chapter is published in PLoS ONE journal (2015). As an author, I retain the right to include the article in my dissertation.

Anastasia N. Naumenko[1], Nicholas A. Kinney[2], Alina A. Kokhanenko[3], Vladimir A. Timoshevskiy[1], Becky S. deBruyn[4], Diane D. Lovin[4], Vladimir N, Stegniy[3], David W. Severson[4], Igor V. Sharakhov[1], Maria V. Sharakhova[1]

[1]Department of Entomology and Fralin Life Science Institute, Virginia Tech, Blacksburg, VA, USA; [2]Department of Genomics, Bioinformatics, and Computational Biology, Virginia Tech, Blacksburg, VA, USA; [3]Institute of Biology and Biophysics Tomsk State University, Russia; [4]Department of Biological Sciences and Eck Institute for Global Health, University of Notre Dame, Notre Dame IN, USA

Abstract

***Background*:** The genome assembly of southern house mosquito *Cx.quinquefasciatus* is highly fragmented and does not provide the chromosome coordinates of the genomic supercontigs. Although cytogenetic maps for the polytene chromosomes of this mosquito were developed, their utilization for genome mapping remains difficult because of the low number of high-quality spreads in chromosome preparations. Therefore, a simple and robust mitotic-chromosome-based cytogenetic tool for the physical genome mapping of *Cx. quinquefasciatus* still needs to be developed.

***Methodology/Principal finding*:** In this study, we used mitotic chromosomes from imaginal discs of 4th instar larvae for idiogram development and further physical mapping. The genetic linkage map nomenclature was adopted for the chromosome numbering based on the direct positioning of 58 previously mapped genetic markers. The smallest, largest, and intermediate chromosomes were numbered as 1, 2, and 3, respectively. Chromosomes on idiograms were subdivided into 72 bands of four different intensities. Using fluorescent *in situ* hybridization, 37 genomic supercontigs from the *Cx. quinquefasciatus* genome were mapped to the chromosomes. Our study also determined the presence of polymorphism in length on the q arm of sex-determining chromosome 1 in *Cx. quinquefasciatus* related to the size of ribosomal locus.

***Conclusion*:** Our study developed a new nomenclature and idiograms for mitotic chromosomes of *Cx. quinquefasciatus.* We demonstrated the utility of these idiograms for physical mapping the *Cx. quinquefasciatus* genome to the chromosomes. This effort together with the linkage mapping resulted in the chromosomal assignment of 13% of the total genome assembly. About 50% of all currently available genetic markers were mapped to the chromosome locations. Further application of the cytogenetic tool constructed in this study will help to improve the quality of the southern house mosquito genome.

Author summary

The southern house mosquito *Culex quinquefasciatus* is one of the most widespread and medically important vectors of the lymphatic filariasis agent *Wuchereria bankrofti* and encephalitis viruses, including the West Nile virus. Construction of high-resolution genetic and physical maps is important for the development of novel vector control strategies. However, the current genome assembly for *Cx.quinquefasciatus* is highly fragmented and requires chromosome mapping. However, physical mapping for this species is challenging due to the poor quality of polytene chromosomes. A genetic linkage mapping assigned 10.4% of the genome to the chromosomes. Here we developed a new cytogenetic tool for *Cx. quinquefasciatus* genome mapping based on mitotic chromosomes from imaginal discs. Application of this tool allowed assignment of 37 genomic supercontigs and 58 genetic markers to the chromosome bands of *Cx. quinquefasciatus*. Further utilization of this tool will facilitate the development of a detailed chromosome map of the southern house mosquito genome.

Introduction

Mosquito-borne infectious diseases pose unacceptable risks to public health and welfare [56]. Among mosquitoes, species of the genus *Culex* are the most taxonomically diverse and geographically widespread [115, 116]. Mosquitoes within the *Cx. pipiens* complex are major vectors for lymphatic filariasis caused by nematode *Wuchereria bankrofti* in tropical and subtropical regions of Asia, Africa, Central and South America and Pacific Islands. *Cx. quinquefasciatus* is also a primary vector for arboviral infections, such as West Nile virus, St. Louis encephalitis, Sindbis and Rift Valley fever viruses. Members of the *Cx. pipiens* complex have great variation in their host range, feeding behavior and female diapause. Sequencing of the genomes for three major mosquito taxa, *Anopheles gambiae* [33], *Aedes aegypti* [35], and *Cx. quinquefasciatus* [36], provides important insights into genetic diversity of mosquitoes and evolution of the mosquito-pathogen interactions [117]. However, as compared to other mosquitoes, *Cx. quinquefasciatus* has the most fragmented genome. A total of 579 Mb is currently assembled into 3,171 supercontigs with the N50 size being ~476 Kb. The N50

supercontig sizes are 12.3 Mb in the *An. gambiae* (PEST) genome and 1.5 Mb in the *Ae. aegypti* genome. A lack of a high-quality chromosome-based genome assembly for *Cx. quinquefasciatus* remains a significant impediment to further progress in *Cx. quinquefasciatus* biology and comparative genomics of mosquitoes. Fragmented unmapped genome assemblies create substantial problems for genome analysis. For example, unidentified gaps cause incorrect or incomplete annotation of genomic sequences; unmapped sequences lead to confusion between paralogous genes and genes from different haplotypes, and the lack of chromosome assignment and orientation of the sequencing contigs does not allow for studying chromosome organization and evolution [118]. Therefore, utility of the genome assembly for investigations on basic biology requires that the majority of supercontigs are anchored and oriented onto *Cx. quinquefasciatus* chromosomes.

Among other approaches, genetic linkage mapping has been the most effective method for the genome mapping of *Cx. quinquefasciatus* so far. Among closely related *Cx. pipiens* species several morphological mutants have been described by Leonore Dennhofer [93]. Crosses involving different mutants permitted assignment of genes related to these mutations to three linkage groups. Use of deoxyribonucleic acid (DNA) markers as a new approach allowed the construction of a genetic map which originally consisted of 21 complementary DNA (cDNA) markers and covered 7.1, 80.4, and 78.3 cM on chromosomes 1, 2, and 3, respectively [94]. The sex determination locus was genetically mapped to the smallest linkage group 1 in *Cx. pipiens.* In addition, multiple quantitative trait loci, related to differences in reproductive diapause between species in the *Cx. pipiens* complex, were also genetically mapped [95]. The most recent genetic linkage map developed for *Cx. quinquefasciatus* includes 63 genetic loci [96]. This map covered 29.5, 88.8, and 65.6 cM on linkage groups 1, 2, and 3 and allowed integration of 10.4% of the genome with the genetic linkage map. Currently, this is the most representative map of the *Cx. quinquefasciatus* genome. However, this map has never been integrated with cytogenetic maps developed for this mosquito.

Physical mapping in *Cx. quinquefasciatus* is challenging because of the poor quality of the polytene chromosomes. Several attempts to create a cytogenetic photomap using *Cx. quinquefasciatus* polytene chromosomes have been made. The Malpighian tubule chromosome map for *Cx. pipiens* [98] and *Cx. quinquefasciatus* [99] and, more recently, the salivary gland chromosome map for *Cx. quinquefasciatus* [100], were developed. However, correspondence of

arms and regions among these maps and the original drawn map published by L. Dennhofer [25] is uncertain. Almost no similarities between landmarks of different chromosome maps were found [100]. These problems occurred because of low levels of polyteny, high frequency of ectopic contacts or associations of nonhomologous chromosome regions, and poor spreading of *Cx. quinquefasciatus* polytene chromosomes in preparation. As a result, only two genes for esterase- and odorant-binding proteins were mapped to the polytene chromosomes of *Cx. quinquefasciatus* [100, 101].

In contrast to polytene chromosomes, mitotic chromosomes do not form ectopic contacts and can be easily utilized for mapping DNA probe to the chromosome bands. A simple and robust technique for obtaining high-quality mitotic chromosomes from imaginal discs of 4th instar larvae was recently developed for the yellow fever mosquito *Ae. aegypti* [119]. This work resulted in 13%, and more recently, in 45% genome placement to the chromosomes for this mosquito [120, 121]. Mitotic chromosomes of *Cx. pipiens*, the closest relative of *Cx. quinquefasciatus*, have been described as three pairs of metacentric chromosomes and numbered in order of increasing size as chromosomes I, II, and III [122]. In some cases, chromosome I was identified as a submetacentric chromosome, meaning that the relative length of the shorter arm p was less than 35% of the total chromosome length. It was also determined that *Cx. pipiens* chromosomes are smaller than those in *Ae. aegypti.* Chromosome measurements also demonstrated that compared with *Ae. aegypti* chromosomes *Cx. pipiens* chromosome I was disproportionally smaller than chromosomes II and III. Unlike in anophelines that have heteromorphic X and Y sex chromosomes [123], sex-determining chromosomes in *Cx. quinquefasciatus* are homomorphic. Only two genes, 18S and 28S ribosomal DNA (rDNA), have been physically mapped to the smallest mitotic chromosome of *Cx. pipiens* [124]. Chromosome maps suitable for the physical mapping have not been developed for the mitotic chromosomes of *Cx. quinquefasciatus*.

In this study, we developed idiograms for mitotic chromosomes of *Cx. quinquefasciatus.* This mapping directly linked previously established genetic markers to physical positions on chromosomes by placing 27 Bacterial Artificial Chromosome (BAC) probes associated with 58 previously mapped genetic markers to the chromosome bands [125]. As a result, chromosomes were renumbered according to the existing genetic linkage groups [94, 96] as follows: 1 – smallest, 2 – largest, and 3 – intermediate. In addition we also mapped an 18S rDNA probe and

10 large genomic supercontigs to the chromosomes. Thus, our study has demonstrated that a mitotic chromosome band-based technique can be utilized to further develop a high-resolution physical map for the *Cx. quinquefasciatus* genome.

Methods

Mosquito strain and slide preparation

The laboratory strain Johannesburg (JHB), used in this study, originated from the field population of *Cx. quinquefasciatus* near Johannesburg, South Africa. The same strain was previously used for the genome sequencing project [36]. Adult mosquitoes were kept at 26^{o}C and fed on artificial membrane blood feeders 4-5 days after emerging. Approximately 4 days after feeding, the eggs were collected and hatched at 26^{o}C. After 4 days, 2nd instar larvae were transferred to 16^{o}C to obtain a high number of mitotic divisions in imaginal discs [119]. At 7-8 days, 4th instar larvae were used for slide preparation. The morphology of the imaginal discs and details of their dissection from the larvae were described for three species of mosquitoes including *Cx. quinquefasciatus* before [126]. 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 [126]. The percentage of chromosome preparations suitable for further analyses, which contained more the 50 chromosome spreads, was ~85%.

DNA probe preparation and fluorescent in situ hybridization

BAC clone DNA from the Notre Dame Johannesburg (NDJ) BAC library [125] was extracted using the Qiagen Large Construct kit (Qiagen Science, Germantown, MD, USA). BAC clone correspondence to the certain genomic supercontigs or genetic markers was determined by BAC library screening [126] or by BAC-end sequence comparison using **Basic Local Alignment Search Tool** (BLAST) against the genome assembly of *Cx. quinquefasciatus* available at Vectorbase [24]. Three PCR fragments with sizes ~1Kb from genomic supercontig 3.32 were amplified using primers: AAAACCCATCTCCCTCGTAG forward, GCTTCTCCAAAACCTTCCTC reverse; TCAAACGACCACAACTTTGA forward, TGGCCTTGTTCTTCTTCTTG reverse; and ATGAAGTTACGGTCGTCAGC forward,

AGTGCATGATGACTCCCATT reverse. Probes were labeled by nick translation with Cy3- and Cy5-deoxyuridine 5-triphosphate (dUTPs) (GE Healthcare UK Ltd., Buckingham-shire, UK) as described before [126]. An 18S rDNA probe was amplified using forward primer CCTATATGGTGGCGCTTGAT and reverse primer AACTAAGAACGGCCATGCAC. It was labeled by Cy3- and Cy5-dUTPs in a PCR reaction using PCR IMMOMIX (Bioline USA, Taunton, MA) with standard parameters. Unspecific hybridization of BAC DNA probes to the chromosomes was preventing by pre-hybridization of the probe with unlabeled repetitive DNA fractions of genomic DNA [126]. Genomic DNA was extracted from adult mosquitoes using Qiagen Blood & Cell Culture DNA Maxi Kit (Qiagen Science, USA). Approximately 500 mg of adult mosquitoes were taken for extraction. Final outcome of repetitive DNA fractions accounts for ~20% of genomic DNA. Approximately 200-300 ng of DNA probe were pre-hybridized with 4 mg of repetitive DNA. Fluorescent *in situ* hybridization **(**FISH) of DNA probes was performed using a standard protocol [126] . Slides were analyzed using Zeiss LSM 510 Laser Scanning Microscope (Carl Zeiss Microimaging, Inc., Thornwood, NY, USA) at 600X magnification. For each probe from 5 to 10 chromosome spreads were tested.

Image processing and measurements

For idiogram development, the best images of the chromosomes from imaginal discs stained with Oxasole Yellow (YOYO-1) iodide (Invitrogen Corporation, Carlsbad, CA, USA) were selected. The original images were converted into gray scale images and contrasted as described previously [127]. These chromosome images were straightened and aligned for comparison using ImageJ program [119, 128]. In total, 150 chromosomes at early metaphase were analyzed, and 25-30 images of each chromosome with reproducible banding patterns were used for idiogram development. To calculate exact proportions of chromosomes, we utilized standard curve measurements in Zen2009LightEdition software [129]. We utilized early metaphase and mid-metaphase chromosome to precisely assign signals to the particular chromosome band.

Results

Culex quinquefasciatus chromosome nomenclature

Our study utilized mitotic chromosome from imaginal discs of 4th instar larvae which develop in legs and wings at adult stage. These imaginal discs are located right under the cuticle and can be easily dissected from the larvae. The morphology of the imaginal discs in *Cx. quinquefasciatus* is similar to that in *Ae. aegypti* [119] and other mosquitoes [126]. According to an original chromosome nomenclature, three pairs of metacentric chromosomes of *Cx. pipiens*, the closest relative of *Cx. quinquefasciatus*, were numbered as I, II, and III in order of increasing size (Rai, 1963). In this study, we established correspondence between mitotic chromosomes and genetic linkage groups by direct placement of 27 genomic supercontigs associated with 58 genetic markers to the chromosomes (Table 2.1). These markers were previously mapped to smallest, largest, and intermediate linkage groups 1, 2 and 3 of *Cx. pipiens*, respectively [94]. 7 BAC clones for this mapping were identified by screening the NDJ BAC library [125] using PCR-amplified genetic markers [130]. Another 21 BAC clones were identified as belonging to the same supercontigs with known genetic markers by BAC-end sequencing with following BLAST-based alignment to the genomic sequences of *Cx. quinquefasciatus* [24]. Supercontig 3.32 containing genetic marker LF335 was mapped as 3 PCR-amplified products with sizes ~1 Kb. BAC clones corresponding to linkage group 1 carrying markers CX60, LF284 and 8 microsatellite markers were mapped to the smallest chromosome. Markers of 6 (cDNA) and 16 microsatellites from linkage group 2 were mapped to the largest chromosome. 7 cDNA and 17 microsatellite markers from linkage group 3 were mapped to the intermediate-in-size chromosome. The order of most markers on chromosomes exactly followed their positions in the genetic linkage map. We found only two discrepancies in the order of markers CX44 and LF203 on chromosome 2 and also a BAC clone with genetic marker LF108 was mapped on chromosome 3 instead of chromosome 2. Thus, we propose renumbering the mitotic chromosomes for *Cx. quinquefasciatus* in correspondence to the genetic linkage groups as follow: 1 – smallest, 2 – largest, and 3 – intermediate chromosomes.

We also determined the average chromosome lengths at mid-metaphase as 4.04 μm for chromosome 1, 6.37 μm for chromosome 2, and 5.59 μm for chromosome 3 (Table 2). Relative lengths of chromosomes were 25.3%, 39.8%, and 34.9% for each chromosome, respectively.

Centromeric indexes (the relative length of the p-arm) were 47.4% and 46.9% for chromosomes 2 and 3. Thus our data confirmed that these two chromosomes are metacentric [122]. Measurements for the centromere position in chromosome 1 varied depending on size of the ribosomal locus determined by FISH of 18S rDNA probe on the chromosome between 43.1% or 48.1%, respectively (Figure 1). Thus, both variants of chromosome 1 must be also considered as metacentric according to the modern chromosome nomenclature [131].

Idiograms of mitotic chromosomes for Culex quinquefasciatus

In addition to chromosome nomenclature, our study developed idiograms or drawn schematic representations of the banding pattern for mitotic chromosomes of *Cx. quinquefasciatus* at early-metaphase. From a whole range of the different stages of mitosis (prophase, prometaphase, metaphase, and anaphase) metaphase chromosomes have the most clear and reproducible banding patterns (Figure 2.2). Similarly to *Ae. aegypti* in *Cx. quinquefasciatus*, homologous chromosomes are paired at prophase and prometaphase (Figure 2.2A, B). At these two stages, the visible chromosome number is equal to three. Chromosomes start segregating from each other at prometaphase and become completely segregated at metaphase (Figure 2.2C). Visible chromosome number at metaphase is equal to six. For the idiogram development, we used images of the chromosomes at early metaphase stained with YOYO-1 iodide (Figure 2.2D). This dye stains euchromatin and provides more detailed banding pattern then DAPI which stains mostly AT-rich regions [119, 126]. The original chromosome pictures were converted into gray scale images. Chromosome images were then straightened and aligned for comparison. After that, unique and reproducible patterns for each chromosome were identified. Following human chromosome nomenclature [132], we determined four color intensities of the chromosome bands: intense (black), medium intensity (dark gray), low intensity (light gray), and negative (white). Chromosomes were subdivided into 20, 28, and 24 bands for chromosome 1, 2, and 3, respectively (Figure 2.2E). A total number of bands for all chromosomes of *Cx.quinquefasciatus* was equal to 72 bands. Each chromosome of *Cx. quinquefasciatus* has unique features or landmarks convenient for the arm recognition of each chromosome: large negative band containing ribosomal locus in the q arm region of chromosome 1, negative band separating intense and medium-intense bands in the p arm of chromosome 2, and large negative band in the middle of arm q on chromosome 3 (Figure 2.2).

Physical mapping on mitotic chromosomes of Culex quinquefasciatus

In addition to BAC clones associated with genetic markers, 9 BAC clones from the largest genomic supercontigs and 18S ribosomal DNA (Table 2.2) were also mapped to the bands on idiograms by FISH (Figure 2.3). Correspondence of the BAC clones to certain genomic supercontigs was determined by BAC-end sequencing and BLAST-based alignment against the *Cx. quinquefasciatus* genome assembly available at Vectorbase [24]. An 18S ribosomal DNA probe was hybridized above the 2 dark bands on the q arm of chromosome 1 on idiogram. In total, of the majority of the DNA probes (17) hybridized to the largest chromosome 2, 9 BAC clones were found in intermediate chromosome 3, and 11 DNA probes were hybridized to the smallest sex-determining chromosome 1. To simplify physical mapping, we optimized a landmark-guided approach developed for *Ae. aegypti* [121, 133] for *Cx. quinquefasciatus* chromosomes (Figure 2.3). We hybridized two BAC clones of interest in the presence of 3 landmark probes: 18S rDNA for 1q arm, telomere BAC clone with genetic marker LF334 on 2q arm, and a BAC clone with genetic marker CX112 close to telomere for 3q arm. Two BAC clones on 3q arm carrying genetic markers CX17 and CX112 were ordered within the band on 3q arm using a two-step mapping approach [120]. In addition to FISH on metaphase mitotic chromosomes (Figure 2.3D), the FISH results on prophase and polytene chromosomes were also analyzed. This additional step permitted the ordering of these genetic markers within chromosome band (Figure 2.3E, F).

Discussion

Our study established a new chromosome nomenclature for *Cx. quinquefasciatus* based on direct physical mapping of the genomic supercontigs associated with previously mapped genetic markers [94, 96]. Originally numbered as I, II, and III in order of increasing size [122], chromosomes were renumbered as 1 – smallest, 2 – largest, and 3 – intermediate chromosomes in correspondence to the genetic linkage map of *Cx. quinquefasciatus.* Similarly, the smallest, largest, and intermediate chromosomes of *Ae. aegypti* were also renumbered as 1, 2, and 3, accordingly to the genetic linkage groups [134]. According to our measurements a total chromosome length at mid-metaphase in *Cx. quinquefasciatus* is 1.5 times longer than in *An. gambiae* [126] and 1.5 times shorter than in *Ae. aegypti* (Table 2.2). It reflects the difference in

genome sizes of 264 Mb, 579 Mb and 1376 Mb in *An. gambiae*, *Cx. quinquefasciatus*, and *Ae. aegypti*, respectively.. Chromosome idiograms for *Cx. quinquefasciatus* are comparable to that previously developed for the yellow fever mosquito *Ae. aegypti* [119]. A total number of bands is also lower and equal to 72 in *Cx. quinquefasciatus* out of 94 in *Ae. aegypti* Physical mapping approach based on idigrams allowed assignment of 45% the *Ae. aegypti* genome to chromosome bands [121]. Additional physical mapping based on the idiograms developed by this study for *Cx. quinquefasciatus* have to be conducted.

Using mitotic chromosomes for physical genome mapping raises a concern about the low resolution of this mapping approach compared to traditionally used polytene chromosomes [135-138]. Unlike the subfamily Anophelinae which have well-developed polytene chromosomes, mosquitoes from the Culicinae subfamily lack high-quality polytene chromosome spreads [99, 139]. Normal polytene chromosomes in *Culex* reflect low levels of polytenization, produce multiple ectopic contacts, and chromosomal preparations are challenging, meaning that it is hard to follow the banding pattern on each particular chromosome It has been estimated that the percentage of preparations with recognizable polytene chromosomes is ~30% in salivary glands of the JHB strain of *Cx. quinquefasciatus* [100]. One of the reasons of selecting this strain for the genome sequencing was good quality of its polytene chromosomes. Among all chromosome spreads within a preparation, usually only one nucleus possessed a readable level of polytenization. However, at least 3 or more readable chromosome spreads should be present in the same preparation to prove the reproducibility of FISH results. This low yield of highly polytenized chromosomes in the preparations makes their utilization for routine chromosome-band-based mapping difficult. Nevertheless, our study determined that in addition to mitotic chromosomes low-polytenized chromosomes from salivary glands can be used for the ordering of closely located supercontigs of *Cx. quinquefasciatus* without assigning them to bands in the polytene chromosomes (Figure 2.4). This so called “two-step” mapping approach was successfully used for the ordering of 100 genomic supercontigs on *Ae. aegypti* [120]. This strategy significantly increased the final resolution of the physical map. The distance between two signals that can be distinguished from each other was estimated at 300 Kb for the polytene chromosomes of *Ae. aegypti*. The resolution of polytene chromosomes in *Cx. quinquefasciatus* may be higher due to their better polytenization and might be comparable to the 100 Kb resolution of polytene chromosomes in *An. gambiae* [136].

Previous investigations of chromosome arm homology between *Cx. quinquefasciatus*, *An. gambiae,* and *Ae. aegypti* indicated whole-arm conservation between *Cx. quinquefasciatus* and *An. gambiae,* and a whole-arm translocation between chromosomes 2 and 3 of *Cx. quinquefasciatus* and *Ae. aegypti* [36]. This conclusion was based only on 9%, 31%, and 88% genome placement to the chromosomes for *Cx. quinquefasciatus*, *Ae. aegypti*, and *An. gambiae* respectively [94, 136, 140]. The dramatic gene order reshuffling between homologous chromosomes of *Ae. aegypti* and *An. gambiae* was recently demonstrated based on 45% and 88% genome placement to the chromosomes for these two mosquitoes, respectively [121]. Additional physical mapping may provide some new insights into chromosome evolution in *Cx. quinquefasciatus*. For example, FISH result of 18S rDNA suggests an inverted position of the ribosomal locus in chromosome 1 of *Cx. quinquefasciatus* compared with *Ae. aegypti* [120]. This locus was mapped close to the centromere above the dark bands in *Cx. quinquefasciatus* (Fig.2.1) but in the middle of the 1q arm below the dark band in *Ae. aegypti*. Our chromosome measurement data of *Cx. quinquefasciatus* chromosomes support the previous observation that the proportions between sex-determining chromosome 1 and autosomes 2 and 3 differ between *Cx. pipiens* and *Ae. aegypti* (Table 2.2). It is clear that the relative length of chromosome 1 is shorter in *Cx. quinquefasciatus* than in *Ae. aegypti*. Our measurements also indicate the presence of two variants of sex-determining chromosome 1 that differ from each other by the size of ribosomal locus and the centromere position on this chromosome. These results support the idea of partial degradation of the sex-determining chromosome 1 in *Cx. quinquefasciatus* compared with chromosome 1 in *Ae. aegypti*. Degradation of sex chromosomes was described in different lineages of Drosophila [141]. However, a more advanced chromosome-based genome map for *Cx. quinquefasciatus* is required for clarifying the intimate details of chromosome evolution in mosquitoes.

Conclusion

Our study developed a new nomenclature for mitotic chromosomes of *Cx. quinquefasciatus*. Based on the genetic linkage map, the smallest, largest, and intermediate chromosomes were numbered as 1, 2, and 3, respectively. We also constructed chromosome idiograms of *Cx. quinquefasciatus* and optimized a landmark-guided two-step physical mapping approach. We

demonstrated the efficiency of this approach for genome mapping purposes on *Cx. quinquefasciatus* by placing 37 genomic supercontigs to the chromosomes. This effort together with previously conducted linkage mapping [96] resulted in the chromosomal assignment of 13% of the total genome assembly. Further application of the approach described here will improve the current highly fragmented genome assembly of *Cx. quinquefasciatus* and will also stimulate research in vector biology and comparative genomics in mosquitoes.

Acknowledgements

Authors thank Sergei Demin for the image processing, Joanne Cunningham for the technical help, and Melissa Wade for editing the manuscript.

Figures

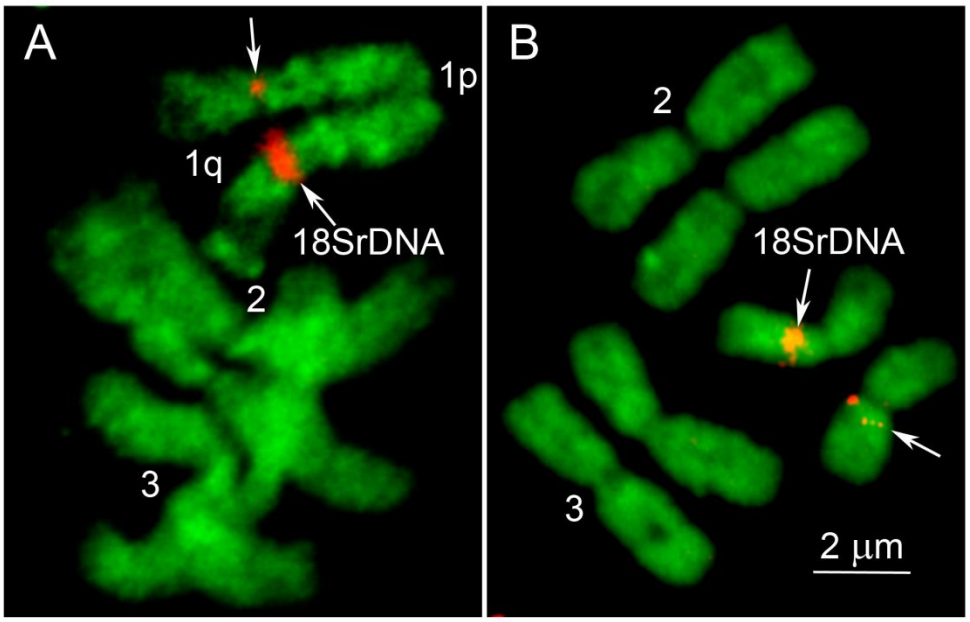


Figure 2.1. Two variants of sex-determining chromosome 1 at early- (A) and mid-metaphase (B). Chromosome 1 is defined as submetacentric with large ribosomal locus or metacentric with small ribosomal locus. Position of 18S rDNA probe on chromosome 1 is indicated by arrow.

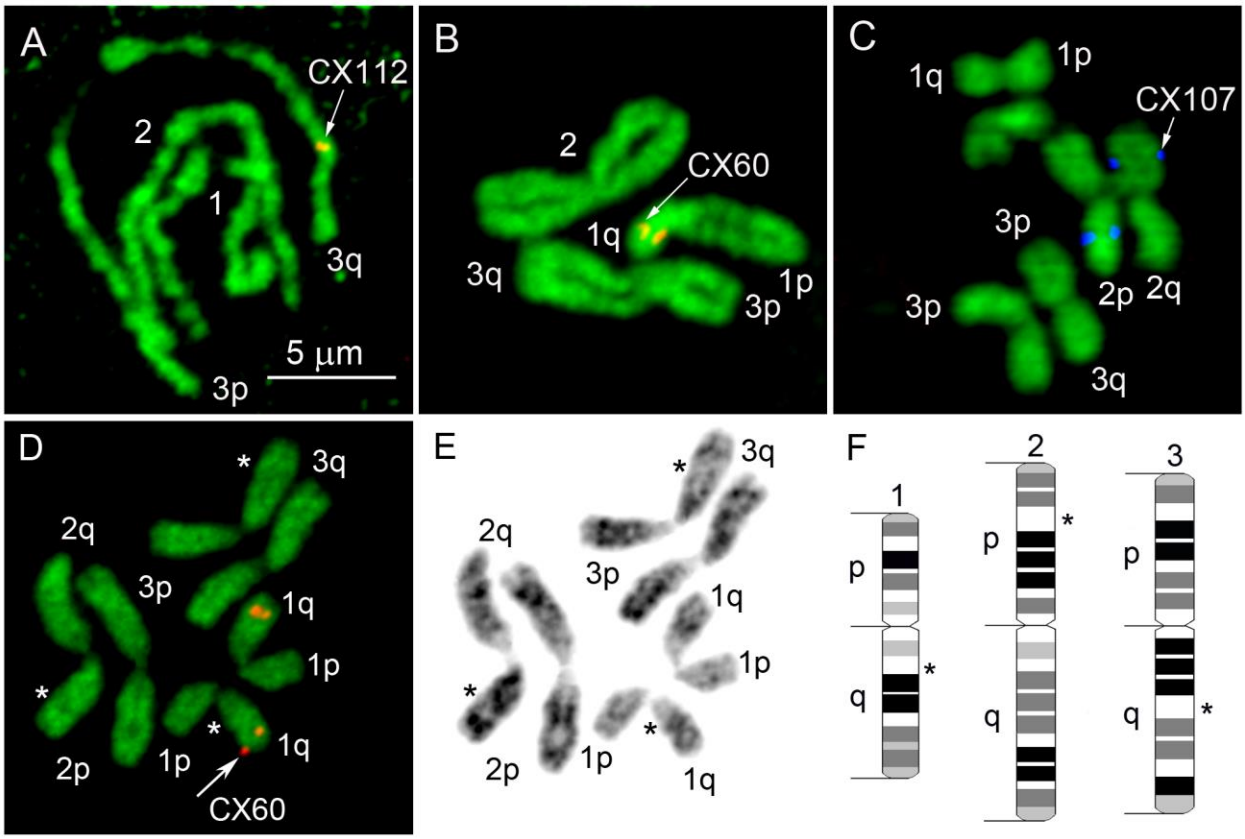


Figure 2.2. Stages of mitosis (A-C) and chromosome idiogram development (D-F) in *Cx. quinquefasciatus.* Early metaphase chromosomes (D) were chosen from prophase (A), prometaphase (B), and late-metaphase (C) chromosomes for the ideogram development. Chromosome images stained with YOYO-1 iodide were converted into gray images (E). Chromosomes on idiograms were subdivided into 72 bands with 4 different intensities (F). Arrows show chromosome positions of the genetic markers CX60 (B, D), CX 112 (A) and CX107 (C). Chromosome landmarks are indicated by asterisks.

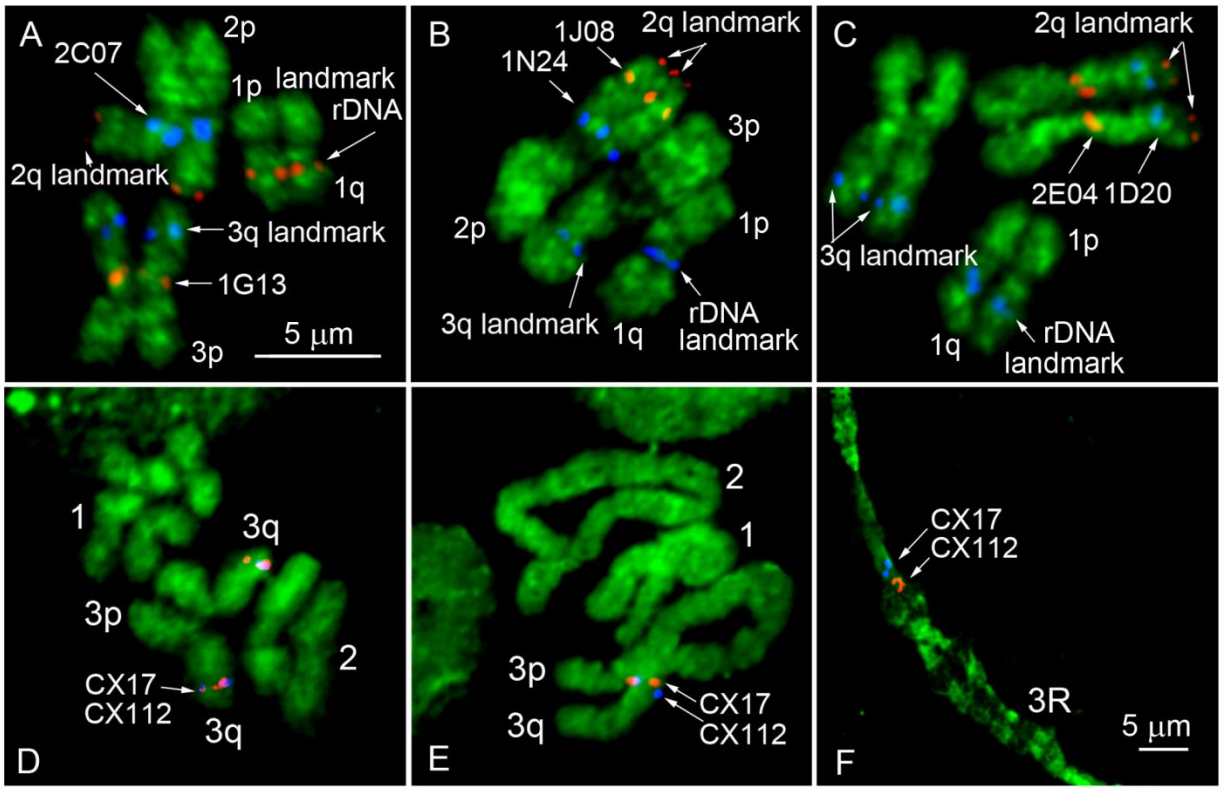


Figure 2.3. A landmark-guided two-step physical mapping approach on *Cx. quinquefasciatus* chromosomes. FISH of two BAC clones of interest was performed in the presence of 2 additional BAC clones, and 18S rDNA were used as landmarks for the chromosome arm identification (A-C). Positions of molecular landmarks and 2 BAC clones of interest on chromosomes 2 and 3 (A, C) and on chromosome 2 only (B) are indicated by arrows. Mitotic chromosomes at metaphase were used for the rapid assignment of the genomic supercontigs to the chromosome bands (D). Longer prophase (E) or polytene chromosomes (F) were further utilized for ordering the genomic supercontigs within the band.

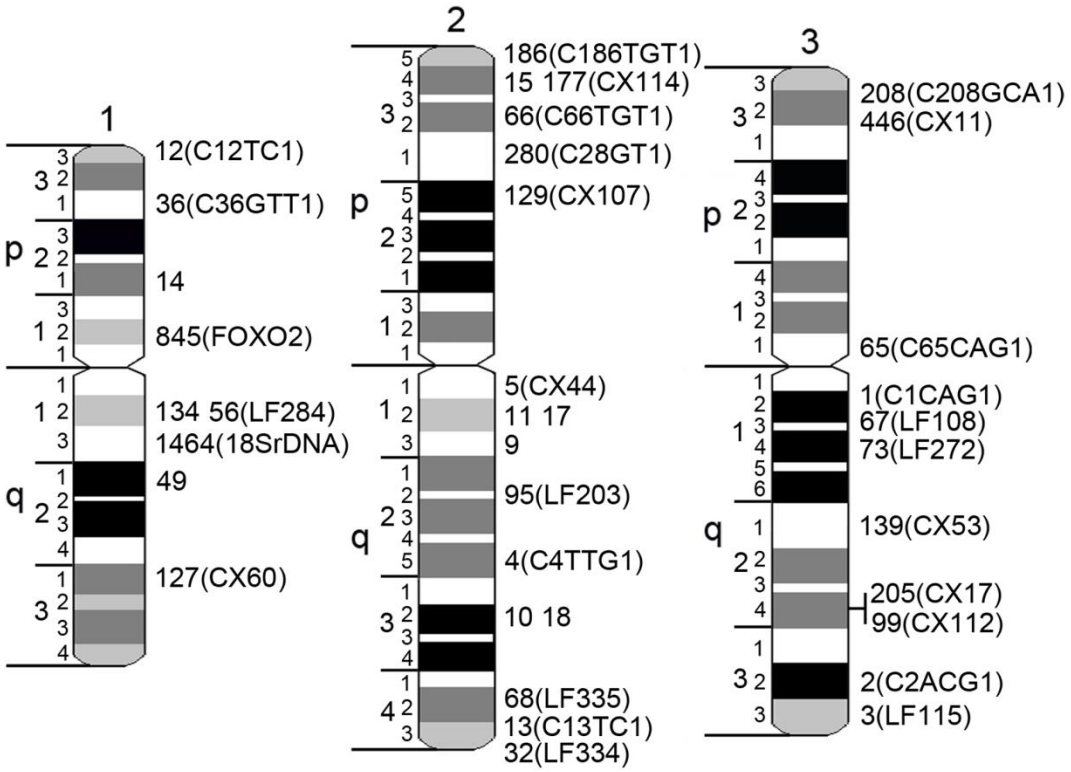


Figure 2.4. Chromosome idiograms with positions of supercontigs and genetic markers. Chromosomes 1, 2 and 3 are indicated by numbers**.** Short and long chromosome arms are indicated by letters p and q, respectively. Genomic supercontigs are indicated by the last 1 to 4 digits of their accession numbers. Genetic markers are shown in brackets.

Tables

Table 2.1. List of DNA probes mapped to the chromosome of *Cx. quinquefasciatus.*

SC	SC size	BAC well/plate	AC# (T7)	AC# (M13)	Genetic marker (AC#, if applicable)	Location
3.845	193967	NDJ.020P4	GF110931	N/A	N/A	1p12
3.14	1835525	NDJ.001J24	KG777556	N/A	N/A	1p21
3.36	1496554	NDJ.003F11	KG961588	KG961589	C36GTT1	1p31
3.12	1895535	NDJ.003N20	KG961610	KG961611	C12CCT1, C12GTT1, C12GTC1, Cxpq51*	1p33
3.56	1226699	NDJ.001O10	KG777559	KG777560	LF284 (BM005502), C56GCA1, C56TGT1	1q12
3.134	821072	NDJ.001D17	KG777561	KG777562	C134AC1b	1q12
3.1464	65637	N/A	N/A	N/A	18S rDNA	1q13
3.49	1394590	NDJ.002A07	KG961590	KG961591	N/A	1q21
3.127	851360	NDJ.048G24	N/A	N/A	CX60 (FD664718)	1q31
3.129	873558	NDJ.064F11	N/A	N/A	CX107 (FD664723), C129CT1	2p25
3.280	510693	NDJ.001J13	KG961592	KG961593	C28GT1	2p31
3.66	1106043	NDJ.002N10	KG961586	KG961587	C66CGT1, C66GT1, C66TGT1*	2p32

3.15	1741670	NDJ.002C09	KG777563	KG777564	N/A	2p34
3.177	728683	NDJ.001N13	KG961594	KG961595	CX114 (FD664728), C177CGT1	2p34
3.186	747982	NDJ.001K16	KG961596	KG961597	C186TGT1	2p35
3.5	2487969	NDJ.002C07	KG777565	KG777566	CX44 (FD664710), C5CGT1, C5GTG1	2q11
3.11	2034973	NDJ.001N24	KG777567	KG777568	N/A	2q12
3.9	2056888	NDJ.001E14	KG777571	KG777572	N/A	2q13
3.17	1689851	NDJ.001E04	KG777569	KG777570	N/A	2q17
3.95	956384	NDJ.003F24	KG961598	KG961599	LF203 (BM005503), C95CAG1, C95GCA1	2q22
3.4	2511003	NDJ.002L02	KG961600	KG961601	C4TTG1	2q25
3.18	1726395	NDJ.001D20	KG777573	KG777574	N/A	2q32
3.10	2129711	NDJ.001J08	KG777575	KG777576	N/A	2q32
3.68	1113402	N/A	N/A	N/A	LF335 (BM005505), C68TCG1	2q42
3.13	1876709	NDJ.003H21	KG961602	KG961603	C13TC1	2q43
3.32	1521851	NDJ.033G10	N/A	N/A	LF334 (BM005506), C32AC1, C32AG1, C32TC1b, C32TGC1	2q43
3.65	1116611	NDJ.005K08	N/A	KG961581	C65AC1, C65CAG1	3p11

3.446	375653	NDJ.009F06	N/A	N/A	CX11 (FD664697), C446TC1	3p32
3.208	649753	NDJ.003E21	N/A	N/A	C208GCA1, CxqTri4	3p32
3.1	3873040	NDJ.001G13	KG777581	KG777582	C1CAG1	3q12
3.67	1097170	NDJ.002F04	N/A	N/A	LF108 (T58322), C67CT1	3q12
3.73	1095011	NDJ.004O20	KG961605	KG961605	LF272 (BM005484), C73CA1, C73TCG1	3q14
3.139	823831	NDJ.013B09	N/A	N/A	CX53 (FD664714), C139CT1, C139GA1	3q21
3.205	667856	NDJ.005A07	KG961612	KG961613	CX17 (FD664699), C205GAC1, C205GTC1	3q21
3.99	949261	NDJ.024B18	N/A	N/A	CX112 (FD664727), C99GTC1	3q24
3.2	2744360	NDJ.002F15	KG961606	KG961607	C2ACG1	3q32
3.3	2758190	NDJ.001M21	KG777583	KG777584	LF115 (R67978), C3GAC1, C3TGC1	3q33

N/A – not applicable

Table 2.2. The measurements of *Cx. quinquefasciatus* mitotic chromosomes from imaginal discs in comparison to *Ae. aegypti*.

Mosquito species	***Cx. quinquefasciatus***	***Ae. aegypti***
Chromosome 1 Average length, µm	4.04	7.1
Relative length, %	25.30%	28.6%
Centromeric index, %	43.1% or 48.1%	46.9%
Chromosome 2 Average length, µm	6.37	9.5
Relative length, %	39.80%	37.9%
Centromeric index, %	47.4%	48.6%
Chromosome 3 Average length, µm	5.59	8.4
Relative length, %	34.9%	33.5%
Centromeric index, %	46.9%	47.4%

List of abbreviations

BAC – bacterial artificial chromosome

FISH - fluorescent in situ hybridization

Mb – mega base pairs

Kb – kilo base pairs

PCR – Polymerase Chain Reaction

N/A – not applicable

NDJ – Notre Dame - Johannesburg

DNA – deoxyribonucleic acid

cDNA – complementary DNA

dUTP – deoxyuridine 5-triphosphate

YOYO – Oxasole Yellow

rDNA – ribosomal DNA

BLAST – **Basic Local Alignment Search Tool**

Chapter 3. An integrated chromosome, linkage and genome map of the West Nile vector, *Culex quinquefasciatus*

Anastasia N. Naumenko[1], David W. Severson[2], Diane D. Lovin[2], Igor V. Sharakhov[1], Maria V. Sharakhova[1]

[1]Department of Entomology and Fralin Life Science Institute, Virginia Tech, Blacksburg, VA, USA; [2]Department of Biological Sciences and Eck Institute for Global Health, University of Notre Dame, Notre Dame IN, USA

Abstract

The idiograms developed for the West Nile vector, *Culex quinquefasciatus*, in the previous study were successfully integrated with 140 genetic supercontigs representing 26.5% of the *Cx. quinquefasciatus* genome. The positions of BAC clones on physical map were compared to the order of available genetic markers on genetic linkage map using linear regression model. A linear regression analysis demonstrated good overall correlation between the positioning of markers on physical and genetic linkage maps. Even with good correlation coefficient, discrepancies were found in markers positions, especially near the centromeres which have lower recombination rates. An integrated physical, genome and chromosome map can be utilized as population genomics tool. The correspondence of *Cx. quinquefasciatus* chromosomes to *Cx. tarsalis*, the primary vector of Western encephalitis and WNV, linkage groups has been established in this study.

Introduction

The southern house mosquito, *Culex quinquefasciatus*, belongs to the widely-spread *Cx. pipiens* species complex which contains a number of related species and their hybrids on multiple continents [92]. In the U.S., the species is spread from Virginia to Florida on the eastern shore and from Iowa to Southern California on the western shore. *Cx. quinquefasciatus* extensively

hybridizes with northern house mosquito *Cx. pipiens* (L.) in the zones of geographic introgression [44, 92]. *Cx. quinquefasciatus* is recognized as a principal vector of lymphatic filariasis and a number of arboviral infections, including St. Louis encephalitis and West Nile fever. Global climate change raises a concern about increase of vector-borne diseases, such as West Nile fever, which is now the most common arboviral disease in the USA [142, 143]. Control of this disease is challenging because its transmission involves multiple hosts—humans, birds, and mammals—as well as multiple species of mosquitoes, such as *Cx. quinquefasciatus*, *Cx. pipiens,* and *Cx. tarsalis,* that are domestic species in both southern and northern and urban and agricultural regions in the U.S. [55]. Lymphatic filariasis, a disabling disease, affects over 120 million people in 73 countries. In sub-Saharan region of Africa, the prevalence of lymphatic filariasis is dramatic, with over 40 million people affected [144]. As a vector of lymphatic filariasis, *Cx. quinquefasciatus* poses an important public health threat, particularly throughout the tropics and sub-tropics of Asia, Africa, the Western Pacific, and parts of the Caribbean and South America. Historically, the control over mosquito populations relied on using organophosphates or pyrethroid insecticides [67]. The usage of insecticides in the past decades led to the spread of resistance in mosquito populations, and *Cx. quinquefasciatus* is considered one of the most resistant species, with multiple resistance mechanisms described [145-147]. Extensive evolvement of resistance traits and multiple feeding sources led to the lack of effective control over widely distributed *Cx. quinquefasciatus* populations.

Sequencing of the *Cx. quinquefacsiatus* genome was done as a joint effort of Broad Institute and J. Craig Venter Institute [36]. Johannesburg (JHB) strain of *Cx. quinquefasciatus* served as the source of genomic DNA. This particular strain has been selected mainly for the expected high quality of polytene chromosome spreads, which would have had promising outcomes in physical mapping of the genome. The genome size of *Cx. quinquefasciatus* equals 579 Mb and takes an intermediate position between *Aedes aegypti* (1311 Mb) and *Anopheles gambiae* (273 Mb) [36]. Interestingly, the repertoire of protein-coding genes in *Cx. quinquefasciatus* (18883 genes in 3172 scaffolds) is 22% larger than in *Ae. aegypti* and 52% larger than in *An. gambiae* due to multiple expansions in gene families, including olfactory, odorant and taste receptors, immune-related genes and xenobiotic detoxification-associated genes such as cytosolic glutathione transferases and cytochrome P450 repertoire. Compared to

Ae. aegypti and *An. gambiae*, *Cx. quinquefasciatus* genome remains fragmented, with N50 size of approximately 476 kb [36].

Simultaneously with the sequencing of the genome, attempts to correctly place genes onto the chromosomes have been done using both genetic and physical mapping. Early efforts to establish genetic linkage groups in *Cx. quinquefasciatus* were done on mutant phenotypes [14-16]. The most recent genetic linkage maps developed for *Cx. quinquefasciatus* developed by P. Hickner and co-authors included 63 genetic loci and covered 29.5, 88.8 and 65.6 cM on linkage groups 1, 2 and 3, respectively [96].

First drawn cytogenetic maps of polytene chromosomes for mosquitoes from *Cx. pipiens* complex were developed by J. Kitzmiller in 1952 [148] and L. Dennhofer in 1968 and 1972 [25, 149]. These drawn maps were not consistent with each other and therefore could not be used for reliable physical mapping. The first high-quality standard cytogenetic photomap was published by M. Unger and co-authors in 2015 [150]. However, the routine utilization of polytene chromosomes for physical mapping of *Culex* mosquitoes remains challenging. *Cx. pipiens* and *Cx. quinquefasciatus* develop very few fully polytenized nuclei per slide. The telomeres of their long chromosomes tend to fuse with each other and, moreover, ectopic contacts between non-sister chromatids are often formed throughout the length of the chromosome. The above-mentioned reasons make *Cx. pipiens* complex mosquitoes one of the most difficult species for cytogenetic studies and therefore for physical mapping: for instance, the mapping effort of *Cx. quinquefasciatus* JHB polytene chromosomes has only resulted in placement of 16 supercontigs on the chromosomes [150].

On the contrary, mitotic chromosomes of *Cx. pipiens* complex mosquitoes do not pose significant challenges for slide preparation. The standard method developed by M. Sharakhova and V. Timoshevskiy for *Ae. aegypti* in 2011 [119] produces over 85% of high quality mid-metaphase and early metaphase spreads that can be utilized for physical mapping. Despite of lower resolution compared to polytene chromosomes of the same species, mitotic chromosomes have reproducible banding patterns and are sufficient for faster mapping of the genomic supercontigs. The idiograms for mitotic chromosomes of *Cx. quinquefasciatus* were developed in 2014 [151]. Mitotic-chromosome based physical mapping since has proven to be the most effective way of mapping *Cx. quinquefasciatus* genome. Based on fluorescent *in situ* hybridization of genetic markers, chromosomes were renumbered in correspondence with

genetic linkage groups as 1 – the smallest, 2 – the largest and 3 – intermediate-sized chromosome. This physical mapping effort resulted in the integration of genetic linkage maps with cytogenetic map and in the placement of 37 genomic supercontigs onto the chromosomal map. Altogether with previous genetic linkage map, this one increased the mapped portion of the genome to 13%. The utility of the genome assembly for investigations on basic biology requires that the majority of supercontigs are anchored onto *Cx. quinquefasciatus* chromosomes. Therefore, an additional mapping effort is required to cover evenly the chromosomal complement. In this case, further improvements of genome assembly make it possible to gain the detailed knowledge of supercontigs position and orientation on the chromosomes.

In this study, we utilized previously constructed idiograms for physical assignment of 101 BAC clones carrying major genetic markers to chromosomal bands. The total of 153 Mb (26.5%) of the genome was placed to the chromosomal map. The majority of chromosomal bands (67 out of 72) were integrated with corresponding genomic sequences. We believe that our effort on placement of genes onto chromosomes will facilitate further genome assembly improvement as well as boost studies on population genetics and QTLs inferring insecticide resistance and diapause in *Cx. pipiens* complex of mosquitoes. A high-quality genome assembly anchored to the chromosomes will improve gene annotation and will help to distinguish potential haplotype scaffolds from the regions of segmental duplications. It will also facilitate identification of epidemiologically important genes that can be used as targets for the vector control. Genome sequencing of *Cx. quinquefasciatus* along with physical mapping of genes to the chromosomes is beneficial for identification of genes evolved into insecticide resistance and pathogen transmission. An improved genome assembly will provide a better framework for comparative genomics that will help understanding of the evolution of epidemiologically important traits and other studies in vector biology.

Methods

Mosquito strain

In this study, we used Johannesburg (JHB) strain of *Cx. quinquefasciatus* which has been previously used for the genome sequencing project. The same strain has been utilized for the development of most recent genetic map and for the integration of genetic, genome and chromosome maps. Based on many inbred populations, the colony is expected to show relatively little genetic variation and is easily maintained in the lab.

BAC clone sequencing

BAC clones were sequenced at Clemson University Genomics Institute platform. All sequences (2721) were submitted to NCBI GSS database under accession numbers KS518275-KS520995.

Fluorescent in situ hybridization

Chromosomal preparations of mitotic chromosomes were obtained from 4th instar larvae as described in published protocols [120, 126]. Samples of BAC clone DNA were prepared by Clemson University Genomics Institute in 96-well plates. FISH was performed as previously described. A landmark-based two-color physical mapping approach was utilized for localizing BAC positions on mitotic chromosomes. BAC DNA for hybridization was labeled with either Cy3- or Cy5-dUTP (Enzo Life Sciences) by nick translation. Chromosomes were stained with YOYO-1 iodide (Invitrogen Corporation, Grand Island, NY, USA). Slides were analyzed using a Zeiss LSM 510 Laser Scanning Microscope (Carl Zeiss Microimaging, Inc., Thornwood, NY, USA) at 600× magnification.

Statistical analysis

To determine the relationship between the physical locations of markers and their linkage positions, genes of known physical position were assigned an integer score. These scores were 1–20 (1p3.4–1q4.4) on chromosome 1, 1–28 (2p4.4–2q4.4) on chromosome 2, and 1–24 (3p4.4–3q4.4) on chromosome 3. This integer score was then regressed upon the cM position of the gene as determined in a number of previous independent linkage mapping studies and F_1 intercross

families. cM position was identified based on the most extensive available genetic linkage map by P. Hickner [96]. Linear regression analysis was performed using Statistica 6.0.

Results and discussion

Physical mapping of Cx. quinquefasciatus genome

The genome of *Cx. quinquefasciatus*, available through VectorBase, allowed us to map 140 supercontigs to their physical locations on the chromosomes (Fig. 1).

Physical mapping also helped to identify misassembled supercontigs, those where BAC clones assigned to the same supercontig were found in different locations on the map. Our data suggests that supercontigs 3.5, 3.8, 3.57, 3.58, 3.67 and 3.73 are misassembled. The misassembled supercontigs are indicated with letters on Fig. 3. The mapped BAC clones cover evenly the whole complement: 68 out of 72 chromosomal bands have genomic supercontigs assigned to them. In total, our mapping effort placed 153.417 Mb of the supercontigs, which is equal to 26.5% of the genome, to the chromosomes. This is the most extensive physical map available for *Cx. quinquefasciatus* genome to date.

The gene densities were calculated for each chromosome based on the mapped supercontigs. The highest density of genes was found on the largest chromosome 2 (2599 genes). Chromosomes 1 and 3 had almost equal numbers of mapped genes, 1614 and 1578, respectively. 5791 protein-coding genes were mapped to the chromosomes, which is equal to 30.5% of the total number of protein-coding genes in *Cx. quinquefasciatus* genome.

Correlation of physical position of the gene with the position on linkage map

BAC clones containing genetic markers from the latest published genetic map were hybridized to chromosomes. The second (largest) and the third (intermediate) chromosome had higher densities of markers (27 and 28, respectively). The smallest chromosome had 17 markers. The whole complement was covered with the markers (Fig.3.1). However, the density of markers was generally higher near the telomeres and lowered towards the centromeres, with very few markers (C5CGT1 on chromosome 2, C45CGT1 and C45AG1 on chromosome 3) located in close proximity to the centromeres. The highest density of markers was found on the telomeres of p

arm of chromosome 2 and q arm of chromosome 3. Two areas, the middle of p arm of chromosome 2 and p arm of chromosome 3, had low densities of markers.
The positions of BAC clones on physical map were compared to the order of available genetic markers on genetic linkage map using linear regression model. A linear regression analysis demonstrated good overall correlation between the positioning of markers on physical and genetic linkage maps: correlation coefficient R^2 equaled 0.66, 0.90 and 0.93 for chromosomes 1, 2 and 3, respectively. 1 cM on a genetic linkage map corresponded to 3.12 Mb of the genome or to approximately 0.4 of the band on physical map (Fig.3.2). Even with significant positive correlation (our data were close to the fitted regression line at $p<0.05$), discrepancies were found in markers positions, especially near the centromeres which have lower recombination rates. Interestingly, the majority of markers located near the telomeres followed the order of the genetic map in most cases, except for q arm of chromosome 1. Sex determination markers were found on the same arm of chromosome 1.

Correspondence of physical map and Hi-C map

Overall, the physical map developed for *Cx. quinquefasciatus* chromosomes was in good correlation with O. Dudchenko and co-authors Hi-C map [152]. Only two markers were in the disagreement with the Hi-C map: LF108 (supercontig 3.67) and C134AC1b (supercontig 3.134). In our study, two BAC clones containing marker LF 108 were mapped to the second and third chromosomes, indicating potential misassembly. In Hi-C map, LF 108 was mapped to the second chromosome.

C134AC1b was mapped to the second chromosome in Hi-C map. Interestingly, in our study it was mapped to one of the homologs of chromosome 1 in male mosquitoes and to chromosome 2 in female mosquitoes. This discrepancy might be explained by sexual dimorphism of *Cx. quinquefasciatus* karyotype.

Establishment of corresponding linkage groups in map available for Culex tarsalis

We were able to establish identity for the linkage groups of another WNV vector, *Cx. tarsalis*. The linkage map for *Cx. tarsalis* was developed using microsatellite markers [153]. Four linkage groups, corresponding to three chromosomes, were established. Several flanking sequences of

mapped microsatellites for *Cx. tarsalis* genome have high similarity with regions of *Cx. quinquefasciatus* genome, making it possible for us anchoring linkage groups to physical chromosomes (Fig. 3.3). The longest linkage group 1 was found to have markers from the supercontigs we placed on the longest chromosome 2: TB210, TB218, and TB220 (3.22, 3.10, and 3.16, respectively). Interestingly, the sex locus in *Cx. tarsalis* is located on the longest chromosome, between microsatellite markers TB210 and TB218.
The intermediate group, linkage group 2, contained markers TB203, TD211, TD113. Supercontigs of *Cx. quinquefasciatus* with sequences similar to these markers (3.185, 3.50, 3.1) were mapped on the intermediate sized chromosome 3.

The smallest third and fourth linkage groups had limited availability of markers: TB1 corresponding to superconting 3.32 for linkage group 3 and TD120 corresponding to supercontig 3.49 for linkage group 4. On our map, these markers are located on the telomere of the long q arm of chromosome 2 and near the sex locus on q arm of chromosome 1, respectively.
In this study, we successfully established correspondence between linkage groups and chromosomes of *Cx. tarsalis* and *Cx. qiuinquefasciatus*. Longest linkage group of *Cx. tarsalis* was anchored to the longest chromosome 2. Intermediate-sized linkage group in *Cx. tarsalis* was anchored to intermediate-sized chromosome 3 in *Cx. quinquefasciatus*. Small p arm of chromosome 1 did not have available markers and therefore was left uncovered. Interestingly, the linkage group 3 was anchored to the telomere of chromosome 2, and the sex-determining locus flanked by markers TB210 and TB218 (supercontigs 3.22, 3.10) was located on q arm of chromosome 1. The smallest linkage group, 4, was anchored to the sex-determining q arm of chromosome 1. These findings should assist in validating markers positioning on the linkage map and help to explore the synteny between two species.

Conclusions

The current genome assembly for *Cx. quinquefasciatus* is highly fragmented and as such is a major impediment to further progress in *Culex* biology and comparative genomics. In this study, we integrated idiograms for the *Cx. quinquefasciatus* genome with additional genetic markers and supercontigs, covering 26.5% of the genome. 30.5% of protein-coding genes were assigned

to their chromosomal positions. This is currently the most detailed physical map for *Cx. quinquefasciatus* genome. Further efforts on genome assembly improvement should significantly increase the mapped portion of the genome as the whole chromosome complement is covered with the mapped BAC clones.

.

Figures

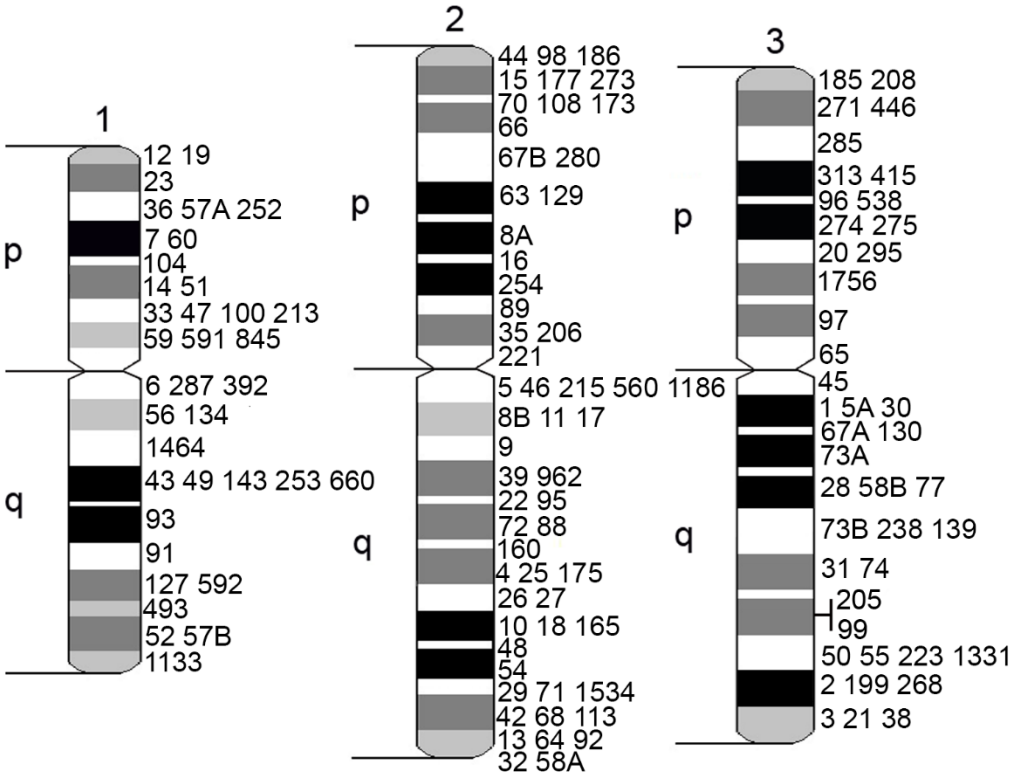


Figure 3.1. A physical map of *Cx. quinquefasciatus* genome.

Chromosomes 1, 2 and 3 are indicated by numbers**.** Short and long chromosome arms are indicated by letters p and q, respectively. Genomic supercontigs are indicated by their numbers in VectorBase *Cx. quinquefasciatus* genome assembly.

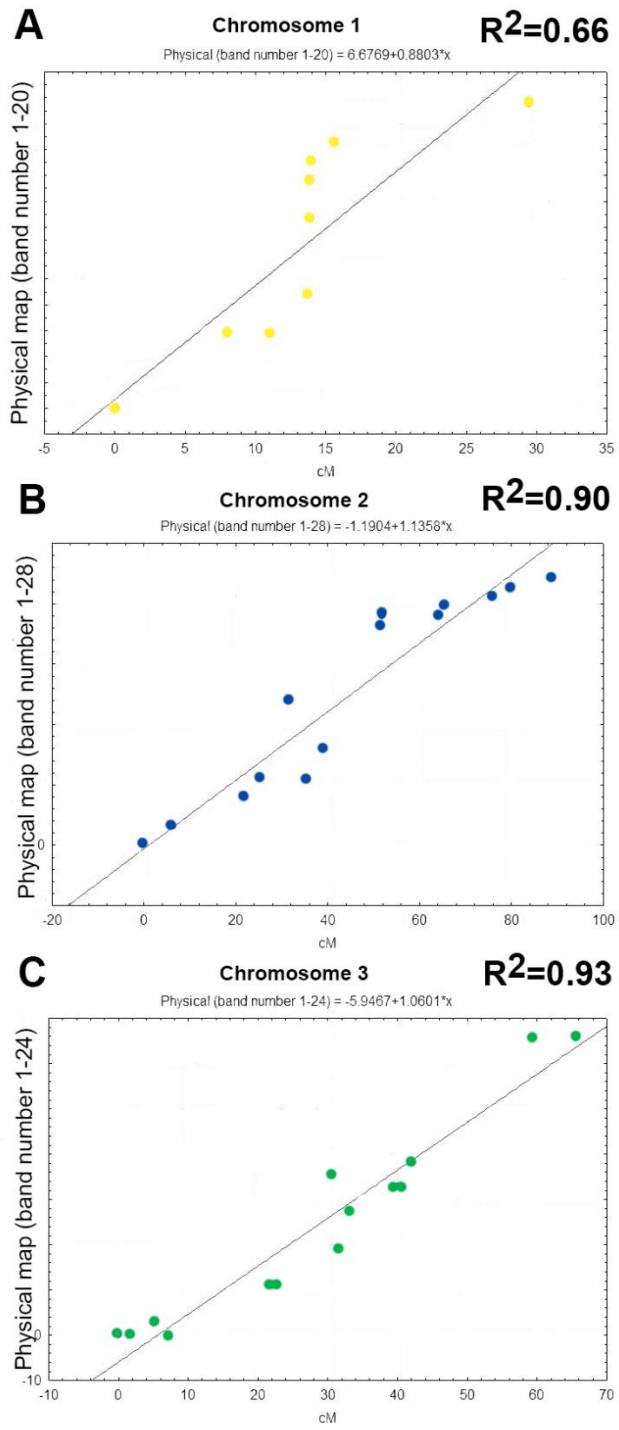


Figure 3.2. Linear regression analyses of the physical position of a gene as a function of its cM position. The relationship between the physical locations of markers and their linkage positions was assessed by assigning genes of known physical position an integer score. Linear regression model in Statistica was used for regression analysis.

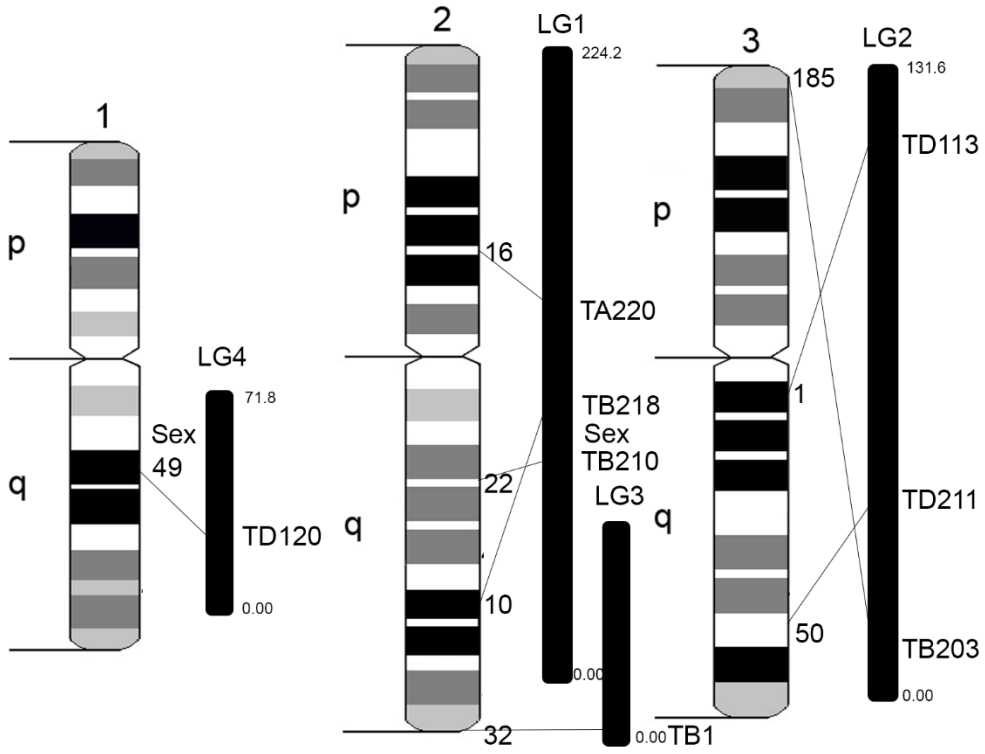


Figure 3.3. Anchoring *Cx. tarsalis* linkage groups to *Cx. quinquefasciatus* physical map. Idiograms representing *Cx. quinquefasciatus* chromosomes are located next to the schematic representations of *Cx. tarsalis* linkage groups (black). Genomic supercontigs containing genetic markers are indicated by their VectorBase numbers (CJpip1 assembly). Corresponding genetic markers of *Cx. tarsalis* are indicated by marker names.

Chapter 4. Inversions and speciation in populations of European malaria vector, *An. messeae s.l.*, in Russia

Anastasia N. Naumenko, Mikhail I. Gordeev, Andrey I. Yurchenko, Anton V. Moskaev, Gleb N. Artemov, Vladimir N. Stegniy, Igor V. Sharakhov, and Maria V. Sharakhova

Abstract

Anopheles messeae s.l. is one of Eurasian mosquito species from the Maculipennis group. It is distributed from Ireland across Europe to Asia. Recently, a cryptic species to *An. messeae s.s.*, formally named *Anopheles daciae*, was described based on the molecular data from the area near Danube river, Romania. *An. daciae* only differs from *An. messeae* by egg morphology and five nucleotide substitutions in the rDNA internal transcribed spacer 2 (ITS2) sequence. *An. daciae* was never reported in Russia. Itt is possible that *An. daciae* is potentially sympatric along the whole range of *An. messeae s.s*. In this article, we report the presence of both species, *An. daciae* and *An. messeae s.s.*, in Russia. Our findings are based on both PCR-restriction fragment length polymorphism assay and ITS2 sequencing. Species were found to be present in various proportions depending on the location, with the polymorphic ITS2 in *An. daciae* and invariable ITS2 sequence in *An. messeae*. Our findings support the hypothesis of species sympatry and further question the bionomics of *An. messeae* and *An. daciae*. The correct identification of *An. daciae* is of high importance to further determine the species status, divergence time and geography of species from the Maculipennis group.

There is known cytogenetic variability in populations of *An. messeae s.l.* Two chromosomal forms, A and B, were described in 1970s [154, 155]. It is not clear, however, if the two described chromosomal forms of *An. messeae s.l.* correspond to genotypes of *An. messeae s.s.* and *An. daciae*. Moreover, if such correspondence exists, it is not clear how widespread is the correspondence between the chromosomally defined forms and genotypes. Here, we attempt to establish the link between karyotype and genotype using cytogenetic and molecular data and to evaluate the level of reproductive isolation between chromosomal forms. If the forms are

effectively reproductively isolated from one another in nature, genetic discontinuity will exist in areas of sympatry. On the other hand, if the forms actively cross, the populations will not show significant correlation between chromosomal form and the genotype.

Introduction

Mosquitoes from Maculipennis group of species (Meigen, 1818) are represented by diverse yet morphologically similar taxa. The distribution and taxonomy of the group are continuously being revised [110]. Sibling species from the group are effective malaria vectors in Europe and Asia. The Maculipennis group comprised twelve Palearctic members: *An. atroparvus*, *An. artemievi*, *An.beklemishevi*, *An. daciae*, *An. labranchiae*, *An. lewisi*, *An. maculipennis*, *An. martinus*, *An. melanoon*, *An. messeae s.s*, *An. persiensis*, and *An. sacharovi* [103, 104]. *An. atroparvus*, *An. sacharovi* and *An. labranchiae* are the most effective malaria vectors in Europe [105]; *An. messeae s.l.* was the primary cause of malaria in Russia [106]. Members of the complex are known for the so-called paradox of *anophelism without malaria* [59], or the presence of malaria vectors in geographical areas with very few malaria cases [156]. However, the presence of such species by itself poses a threat because of the great potential to spread malaria given the changing climate and global warming [157, 158].

The discovery of new species, such as *An. artemievi* [159], *An. persiensis* [104] and *An. daciae* [105] further questioned the taxonomy and bionomics of the species from Maculipennis group.

An. messeae s.l. has a widespread distribution extending from Ireland across Europe and Asia and into China and Russia [107]. The large range of this species combined with known genetic variability associated with different geographic areas [160, 161] suggests that area-specific biological or behavioral adaptations are likely to have occurred. For a long time, *An. messeae s.l.* was considered as a highly polymorphic species representing genetic variability associated with different geographical locations. *An. messeae s.l.* is susceptible to *Plasmodium vivax* but not to the tropical *P. falciparum* malaria infection. Furthermore, its vectorial capacity has come into question with the finding of recently described species, formally named *An. daciae*, in Romania and later in Southwestern England.

There were other efforts to explain genetic polymorphisms within *An. messeae s.l.* populations. Y. Novikov and V. Kabanova postulated the presence of two cytogenetically

distinct races within *An. messeae s.l.* populations [112]. Two ways of how chromosomal variations combine in *An. messeae* populations were found: original variant X0 combined with 2R0,3R0 and 3L0, inverted X1 and X2 tended to combine with 2R1, 3R1 and 3L1.
XL0, 2R0, 3R0 and 3L0 were shown to be distributed in west, inverted variants – in northern and continental parts of the range [106, 112, 155]. These races differ in fecundity, viability of imago and larvae, food behavior, rate of development, relationship with predators and parasites, sensitivity to the toxins of *Bacillus thuringiensis subsp. israelensis*; [155, 161-166]. Y. Novikov referred to these races as forms A and B [112, 154]. Later, he confirmed with restriction fragment length polymorphism assay and taxonprints [113] that two forms, A and B, exist. Restriction of rDNA ITS2 fragment with FokI enzyme was used to distinct A and B forms. However, he never considered these forms to be independent species.

The newly described species *An. daciae* is different from the closest member of Maculipennis group, *An. messeae,* by egg morphology: the eggs of *An. daciae* are generally darker in color, smaller and have tubercles that are organized in patches of slightly different shape [105]. A PCR assay developed for identification of sibling species from Maculipennis group [107] including *An. atroparvus, An. beklemishevi, An. labranchiae, An. maculipennis, An. melanoon, An. messeae* and *An. sacharovi*, was not reliable for the identification of *An. daciae* due to the same size of PCR product for the two species. The sequencing of rDNA ITS2 of mosquitoes from Maculipennis group collected near Danube river in Romania by Linton, Nicolescu and Harbach in 2004 [105] revealed five single nucleotide polymorphisms in *An. daciae* which differentiated it from the sympatric species, An. messeae. Using molecular approach, *An. daciae* has been later discovered in Germany [108, 109], England and Wales [110]. In all locations, *An. daciae* has been found in sympatry with its cryptic species, *An. messeae.*

R. Danabalan [110] supported the species status of *An. daciae* by using slightly different RFLP assay with BStUI enzyme. PCR product resulting from amplification of *An. daciae* DNA was shown to have 4 products after restriction with BStUI whereas digestion of *An. messeae* rDNA ITS2 gave three products. It is not clear, however, whether A and B forms correspond to cryptic *An. messeae* and *An. daciae*.

Material and methods

Field collections

Mosquitoes (450 specimens) were collected between 2005 and 2016 in 6 locations in Russia: Solnechnogorsk, Egoryevsk, Novokosino and Noginsk near Moscow (European part of Russia) and Chainsk and Kandinka near Tomsk (Siberia) (Fig. 1). Larval stages were collected from the breeding sites by dipping and were then fixed in Carnoy solution (1:3 acetic acid: ethanol) or 70% ethanol. Adult specimens were collected by trapping from resting places.

Karyotyping

100 larvae from each of three populations: Egoryevsk, Novokosino and Noginsk (total 300 larvae specimens), identified as *An. messeae s.l.* by morphology, were fixed in Carnoy solution and dissected. Thorax was separated and preparations of polytene chromosomes from salivary glands were done in acetoorseic acid. Polytene chromosomes were visualized using light microscope and karyotype for each specimen was described for the whole chromosomal complement. Each inversion (inversions X1, X4, 2R1, 3R1 and 3L1) was taken into account. Each sample was numbered and the rest of the specimen was then used for DNA extraction and subsequent Sanger sequencing.

DNA extraction

Mosquito DNA was extracted from specimens using standard protocol for Qiagen DNeasy Mini Kit (Qiagen, USA) with slight modifications. Samples were homogenized in 45 μl of extraction buffer ATL with 5 μl added Proteinase K and incubated at 56ºC overnight. For homogenization, sterile 1.5 mL tips were used to prevent the risk of contamination. For each sample, fresh tip has been used. DNA elution has been performed in 100 μl of water.

PCR amplification

rDNA ITS2 has been amplified using forward universal primer designed by Proft [107] 5'-ATCACTCGGCTCTCGTGGATCG-3'(Tm=64.5 ºC) and reverse primer used by Novikov [113]: 5'-ATGCTTAAATTTAGGGGGTA-3' (Tm=54.2 ºC). Different combinations of primers have been tried: this one was chosen based on the longer amplicon length (613 bp) and higher

rate of PCR success (100%). Either hot start Immomix™ or proofreading Accuzyme™ polymerases (Bioline) were used with similar success rates. PCR mixture contained 1-2 µl of DNA template (depending on concentration), 1 µl of both forward and reverse primers at 10mM concentration and 10 µl of Immomix™ or Accuzyme™ polymerase. Water was added to the mixture up to 20 µl total volume. Amplification was performed using Eppendorf (USA) thermal cycler with programmed parameters: initial denaturation at 95 ºC for 10 min, followed by 35 cycles of 95 ºC for 15 s, 55 ºC for 30 s and 72 ºC for 30s, and a final extension step at 72 ºC for 5 min. The reaction was then placed on hold at 4 ºC.

PCR-Restriction Fragment Length Polymorphism assay

The amplified product (ITS2 of *An. daciae* and *An. messeae* has been digested with BstU I (New England Biolabs) (CG↓CG) for 30 min at 60 ºC (PCR and digestion parameters for both reactions are described in the previous protocol). Expected fragment size for *An. messeae* after BstUI digestion equals 332, 109 and 42 bp, for *An. daciae* - 332, 59, 52 and 42 bp.

In the next step, we used the same PCR product to distinct between A and B forms of *An. messeae.* The parameters of FokI enzyme (New England Biolabs) restriction were slightly different from the standard protocol. The product was digested for 60 min at 37 ºC. The reaction was inactivated for 20 min at 65 ºC. Reaction mix contained 5 µl of ITS2 PCR product without purification, 12.5 µl of ddH_2O, 2 µl of 10X CutSmart Buffer (New England Biolabs) and 0.5 µl (2.5 U) of FokI restriction enzyme.

The expected size of digestion products for A form of *An. messeae* equals 159 and 276 bp. ITS2 product of B form does not have restriction sites for FokI.

DNA sequencing

For DNA sequencing, amplicons were checked on a gel and then purified with Wizard™ PCR Clean Up Kit (Promega). Concentrations of purified PCR products were measured. PCR products were mixed with either forward or reverse primers at 3.2 pm concentration and sequenced at Virginia Bioinformatics Institute platform. The majority of samples were sequenced in both forward and reverse directions to eliminate the possibility of contamination

and to confirm the presence of same SNPs on both DNA strands. 48 samples were sequenced in forward direction only.

Sequence analysis

Sequence were trimmed, aligned and analyzed using DNASTAR Lazergene software modules EditSeq, Seqman Pro and NCBI BLAST. Assemblies' parameters were set to minimum percent match 85.

Statistical analysis

Statistical analysis was performed in JMP Pro (Fit X by Y, ANOVA) and Statistica 6.0 (correlation matrices). Correlation matrix was created for selected variables (inverted and non-inverted chromosome variants). All results level of detail was requested, and a scatterplot matrix for the selected variables was computed. Detailed results were selected to compute a detailed table of results with descriptive statistics and the parameters for the regression equation. The default p-value for highlighting was 0.05; significant t-tests with p less than or equal to this value were shown in the highlight color in the results spreadsheets. For Fit X by Y and ANOVA analyses we indicated variables as 0 – both homologs are not inverted (homozygote), 0.5 – inverted and not inverted variant together (heterozygote) and 1 – both homologs inverted (homozygote). The differences were significant at $p < 0.0001$. The t-test in oneway ANOVA was performed assuming unequal variances.

Fixation index was calculated using formula $F_{ST}=(H_T-H_S)/H_T$, where H_S = the average expected heterozygosity among organisms within subpopulations. The heterozygosity of the subpopulations was measured based on inversion frequencies in those populations. H_P was calculated by simply averaging all the subpopulations together.

H_T= The average heterozygosity among organisms within the total area considered. This is calculated by taking the average of all the frequencies of each inversion separately, and then using 2pq to calculate an expected total heterozygosity. This is the heterozygosity expected if all the individuals in all the subpopulations were panmictic.

Results

The ratio An. daciae: An. messeae in populations is variable

Based on the location, we found different ratios of *An. messeae* and *An. daciae* in each population. In general, larger proportion of *An. messeae* was noted in Moscow populations (Figure 4.1). Populations from Tomsk had significantly larger proportion of *An. daciae.* Among closely located Moscow populations, the ratio of *An. messeae: An. daciae* varies from 27.9%:72.1% in Egoryevsk population to 73.7%:26.3% in Novokosino population.

A-form corresponds to An. daciae, B-form – to An. messeae

To compare Danabalan and Novikov assays and to establish correspondence between A and B forms of *An. messeae*, *An. messeae* and *An. daciae*, we simultaneously performed BstUI and FokI RFLP assays for 10 female and 16 male mosquitoes collected in Moscow and 10 female and 10 male mosquitoes collected in Kandinka, Tomsk.

The amplified product (613 bp ITS2 amplicon) has been digested with BstU I. The same product has been digested with Fok I.

Expected fragment size for *An. messeae* after BstUI digestion equals 332, 109 and 42 bp, for *An. daciae* - 332, 59, 52 and 42 bp.

Based on BstUI PCR-RFLP assay results, all DNA samples from Kandinka belong to *An. daciae* and have uniform sizes of enzyme restriction products, whereas the length of the smaller fragments was variable for samples from Moscow. In 12 (7 male, 5 female) samples the smallest product of digestion equals approximately 100 bp when comparing to the ladder (corresponding to *An. daciae* expected fragment sizes), in other cases the size of the smaller fragment corresponded to *An. messeae s.s.* (Figure 4.2A).

All the DNA samples from Kandinka after FokI PCR-RFLP were digested as *An. messeae* form A. Therefore, *An. daciae* corresponds to cytogenetic form A of *An. messeae s.l.*

Fok I digestion of ITS2 amplified from the same samples identified as *An. daciae* after BstUI digestion were digested in the way of A-form. ITS2 of DNA samples identified as *An. messeae* in the first assay were not digested by Fok I, thus bringing us to a conclusion these mosquitoes represent B-form (Figure 4.2B).

Based on the results of simultaneous BstUI and FokI RFLP assays, A form of *An. messeae* corresponds to *An. daciae* and B form – to *An. messeae*.

An. messeae ITS2 sequence is conserved but An. daciae ITS2 sequence is polymorphic

Both *An. messeae* and *An. daciae* were identified among sequenced specimens, and in every location except for Kandinka, where only *An. daciae* have been found, the species were found in sympatry. We did not observe sequence heterogeneity which would make it impossible to differentiate between two species. However, we found that *An. messeae* ITS2 is conserved whereas *An. daciae* ITS2 is polymorphic. *An. messeae* ITS2 contains five SNPs: T in position 221, T in position 225, C in position 227, G in position 422 and G in position 442 (Figure 4.3). *An. daciae* ITS2 sequence varied. We found 4 genotypes: AATAC in positions 221, 225, 227, 422 and 442 which was originally found by Nicolescu, Linton and Harbach; ATTAC; TTTAC; and TTCAC. The first 2 genotypes were widespread in all locations (98% of all *An. daciae* samples). Genotype TTTAC had 2% frequency, and only one sample had genotype TTCAC. Based on the sequencing data peaks, we were able to identify only one hybrid between *An. daciae* and *An. messeae* (double peaks were presented in both forward and reverse sequences).

An. messeae s.s. and An. daciae are genetically isolated taxa

Each genotyped sample was assigned to a corresponding karyotype (Supplementary table 1). The chromosomal variant frequencies were then calculated for each of the species in all populations (Table 4.1).

We confirmed that two chromosomal forms exist (Table 4.2) and that these chromosomal forms correspond to *An. daciae* (A form) and *An. messeae s.s.* genotype (B form), respectively. Correlation matrices demonstrated that *An. daciae* is characterized by the combination of positively correlating X0, 2R0, 3R0 and 3L0 variants. *An. messeae s.I* has the prevalence of X1 inversion on chromosome X which positively correlates with the presence of inversions 2R1 on chromosome 2, 3R1 and 3L1 on chromosome 3. No samples with the genotype of *An. messeae s.s.* had original variant X0, even in heterozygote.

Inversions X4 and 3L1 were the least common. Variant X4 was found in 11 samples of *An. messeae s.s.* and was not found in *An. daciae* in all populations. Variant 3L1 was found in 15 *An. messeae s.s.* samples and was not observed in *An. daciae*.

Based on sequencing data analysis and the presence of double peaks on chromatograms, we have observed only one hybrid in all three populations. The only hybrid found had karyotype X04 (heterozygote of original variant, only found in *An. daciae*, and inversion X4, only observed in *An. messeae s.s.*).

For all populations, strong correlation at $p<0.0001$ was found between species and the prevalence of either original or inverted variant (Supplementary Tables 4.2-4.10). Despite the presence of heterozygotes, the original variants were dominant in *An. daciae*. The inverted variants X1, X4, 2R1, 3R1 and 3L1 were widely found in *An. messeae s.s.*

Inversion frequencies are different in An. messeae s.s. and An. daciae

We calculated the mean with 0.95 confidence interval for each of the inversions separately. For each of inversions 2R1, 3R1, X1 significant differences at $p<0.0001$ were observed for each of three locations (Figure 4.4 A-D). Inversions 3L and X4 were specific for *An. messeae s.s.*, and X0 was only found in *An. daciae*.

Inversion frequencies largely did not depend on geographical location. (Supplementary Table 11-12). However, *An. messeae s.I* population from Egoryevsk was significantly different from Noginsk population by the frequency of 2R1 inversion ($p=0.0182$).

Fixation index Fst suggests that An. messeae s.I and An. daciae are isolated

We measured the genetic distance between *An. messeae s.I* and *An. daciae* subpopulations using formula $F_{ST}=(H_T-H_S)/H_T$, where F_{ST} – fixation index, H_S - the average of each subpopulation heterozygotes and H_T - heterozygocity based on the total population karyotype frequencies (Table 4.3). Based on the fixation index, the subpopulations were found to be almost completely isolated.

Discussion

The species status of *An. daciae* remains unclear. O. Bezzhonova and I. Goryacheva report presence of the rDNA sequences specific for *An. messeae s.s.* and *An. daciae* simultaneously in one individual, as well as high heterogeneity of ITS2 (9 variants described) [111]. However, the methodology included using of regular instead of proofreading DNA polymerase and subsequent cloning of PCR product into vector prior to sequencing which could lead to higher rate of errors in general. The other explanation is that the specimens could have been obtained from an isolated population where the rate of SNPs is higher. R. Danabalan reported feeding of *An. daciae* on humans, mammals and birds [110]. The vector status as well as taxonomy, bionomics and geographical distribution of *An. daciae* remains largely unknown, and correct species identification is especially important for vector-borne disease studies. for the first time, we report presence of *An. daciae* in Russia using both sequencing and restriction fragment length polymorphism assays. Similar to previously described European populations, *An. daciae* has been found in close sympatry with the cryptic species, *An. messeae*. Other species from Maculipennis group were occasionally found among samples: *An. maculipennis* in Egoryevsk, Noginsk and Novokosino, *An. beklemishevi* in Egoryevsk and Chainsk. Interestingly, the populations containing significantly smaller proportion of *An. messeae s.s.* were found to be in sympatry with *An. beklemishevi*. The ratio of species varied even in closely located Moscow populations, suggesting possible habitat splitting between two species. The pattern might be explained by differences in habitats: Novokosino and Noginsk specimens were collected from lake and the pond accordingly, whereas Egoryevsk and Chainsk specimens were collected from low-level water sources (Figure S1). Interestingly, the same differences were found between M and S forms of *An. gambiae*, with M forms associated with semi-permanent, flooded areas and S-forms prevalent in temporary, rain-dependent sites [30, 167].

The comparison of Y. Novikov and R. Danabalan assays led to the establishment of direct correspondence between A form of *An. messeae s.l.* and *An. daciae*. B form of *An. messeae s.l.* was found to be analogous to *An. messeae s.s.* In all cases, RFLP results were confirmed by sequencing. We found Y. Novikov assay to be more reliable and faster as smaller products in R. Danabalan assay are often invisible or hard to visualize on a gel and require longer runtime.

Out of the significant number of specimens, we were able to identify only one hybrid based on sequencing data. The low hybrid proportion despite of two species living in close proximity suggests that some level of isolation exists between *An. daciae* and *An. messeae*.

For the first time, genetic polymorphism was detected in ITS2 rDNA of *An. daciae*. Upon review of all the genotypes found, we were able to confirm only two SNPs, in positions 422 and 442 that unambiguously differentiate *An. messeae s.s.* from *An. daciae*. In the only hybrid found, double data peaks suggested that SNPs characterizing both *An. messeae* and *An. daciae* are present in both strands. Our finding suggests that the polymorphisms in positions 422 and 442 in primary ITS2 sequence are sufficient to distinct *An. daciae* from *An. messeae*.

Here, we established the correspondence between chromosomal form A and *An. daciae* using the combination of karyotype and genotype data. In all populations, the ratios of *An. messeae s.s.* and *An. daciae* were different. Interestingly, in Egoryevsk population, where the highest proportion of *An. daciae* was observed, *An. maculipennis* and *An. beklemishevi* were found as co-habitants. In the populations where *An. messeae s.s.* was predominant (Noginsk and Novokosino), only two samples of *An. maculipennis* were found. The habitats are also different. Despite the same climate zone and close proximity between three populations, Egoryevsk habitat has the lowest water level among three. Populations from Noginsk and Novokosino were collected from the pond and the lake with deeper water levels, respectively (Figure S1).

In An. messeae s.s. and An. daciae, different chromosomal variants dominate

Chromosomal inversions are directly involved in ecological and behavioral plasticity of the mosquito species. An introgression of inversions that carry novel, adaptive alleles might trigger large range expansion of mosquitoes [168]. The process of incipient speciation produces new ecotypes that are better adapted to various environments and wide ranges of habitats. The *Anopheles* genus has at least 170 cryptic species belonging to 30 species complexes [54]. Cryptic species complexes are groups of closely related species which are difficult or impossible to distinguish by morphological traits [102, 169]. Most of the cryptic species in such complexes evolved as the result of the incipient speciation and crossingover-suppressing paracentric inversions, particularly on chromosome X. The complexes are particularly important for arthropod-borne diseases studies because the complex may include highly specialized malaria vectors along with species with lesser susceptibility to malaria parasite, *Plasmodium*.

Genetic analysis is used for distinguishing species that are not recognized based on their morphological features. The criteria for speciation, in this case, would be reproductive isolation, or the inability of species to cross in the areas of sympatry [3]. However, such species may have different levels of isolation. For instance, in *An. gambiae* species complex the speciation is most likely the result of partial reproductive isolation.

The suppressed-recombination model, proposed by Coluzzi in 1982 [8], describes the speciation within *An. gambiae* complex of species and can be applied as a speciation model to mosquito species complexes in general. According to the model, the chromosomally homomorphic population colonizes a favorable range and expands. As the numbers increase, the population reaches the periphery of the range and faces extreme conditions such as low or high temperatures, dryness or humidity, and seasonal climate fluctuations. The mutations that increase adaptations will accumulate in the subpopulations at the periphery of the range, creating new ecotypes that are adapted to new environmental conditions. The mating between mosquitoes from subpopulations with mosquitoes from central population will eventually homogenize or even eradicate the newly acquired alleles, except for those protected by chromosomal inversions, where the recombination is suppressed. The mutations will further accumulate within inversions, making grounds for subsequent reproductive isolation. The inversion polymorphism exists within *An. gambiae* s. s. species [48, 170]. Three chromosomally distinct populations, Savanna, Mopti and Bamako, are different in ecotype preferences and breeding sites; yet enough assortative mating suggests that these species are incipient. However, based on the evidence for inversion polymorphism and fixed SNPs [170], the S form is now referred to as separated species, *An. coluzzi*, by many scientists. Chromosomal polymorphism has been shown for other species groups as well.

Multiple polymorphic inversions are typical for *An. messeae s.s*. populations [106, 112, 114]. The standard chromosome arrangement, X0 2R0 3R0 3L0, is predominant in *An. daciae* (Table 4.1, Table S2-7). *An. messeae s.s.* has exclusively (with only one sample found to be heterozygous) inverted variant X1 (Figure 4.4 D). *An. messeae s.s.* significantly differs from *An. daciae* by the frequencies of inversions 2R1 and 3R1 (Figure 4.4 A-C, Table S2-7). X4 inversion, and 3L1 (Figure 4.4 B,D) inversions were only found in *An. messeae s.s.*

Inversion frequencies were found to depend on species but not on the breeding site, except for 2R inversion in *An. messeae s.I.*(Table S8-16). The limited sampling, however, poses serious limitation on the studies of inversion frequencies along the geographical range.

Reproductive isolation separates An. messeae s.s. and An. daciae

Out of 285 specimens, we were able to identify only one hybrid (ME21) based on chromatogram data peaks. The fixation index data supports hypothesis of assortative mating within *An. messeae s.s.* populations (Table 4.3). Hartl and Clark [171] defined fixation index classes as following: <0.05 = little genetic differentiation; 0.05-0.15 = moderate genetic differentiation, gene flow is limited; 0.15-0.25 = significant genetic differentiation, gene flow is significantly limited; >0.25 = great genetic differentiation, subpopulations are isolated. Frankham [172] described F_{ST}>0.15 as significant differentiation. Based on these classifications, all three populations demonstrate significant to great genetic isolation which supports the hypothesis of incipient speciation within *An. messeae s.s*. populations. The transcriptome phylogeny suggests that *An. messeae s.I* and *An. daciae* separated 2-3 million years ago, which is more than M and S forms of *An. gambiae* [7]. Despite great genetic distance between *An. messeae s.s.* and *An. daciae*, the species hybridize in the laboratory conditions, suggesting the possibility of prezygotic isolation.

Our findings call for more detailed cytogenetic studies on *An. messeae s.s*. chromosomes. The inversions, as one of the main mechanisms of adaptation to new environments, need to be studied along the geographic range. It is essential to identify the specific adaptive traits which are involved in the genetic mechanism of local adaptation: for instance, 2La inversion in *An. gambiae s.l.*. varies along the aridity gradient, from complete absence in humid rainforest to fixation in savannas [30, 173]. The present study is limited to three populations located in close proximity to one another; however, strong evidence for assortative mating was found in these populations. This indicates the need for broader geographical sampling of *An. messeae s.l.*, increased search for mechanisms of prezygotic isolation and field-based hybrid survival tests to evaluate the postzygotic isolation.

Conclusions

Environmental and behavioral diversity within malaria mosquito's populations pose a significant threat to the effectiveness of the malaria control. This study identified the presence of newly described species from Maculipennis group, *An. daciae*, in 6 distant locations in Russia and characterized the genetic divergence between two cryptic species using multiple polymorphic inversions as markers. we *An. daciae* was found to correspond to cytogenetic form A of *An. messeae s.l.*, *An. messeae s.s.* – to B form. Molecular approach demonstrated that *An. messeae s.s.* ITS2 sequence is conserved, whereas *An. daciae* ITS2 sequence is polymorphic. Two fixed conserved SNPs in positions 422 and 442 and low number of hybrids suggest species isolation. That was supported by inversion frequencies calculated for both species.

Analysis of the inversion frequencies has demonstrated near complete reproductive isolation between *An. messeae s.s.* and *An. daciae*. The most significant divergence was observed for inversions on chromosome 2 (2R1) and X (X1).These data clearly suggest the existence of two reproductively and genetically isolated species *An. messeae s.s.* and *An. daciae* in Russia. Chromosomal inversions are beneficial for both adaptation and ecological plasticity enabling geographical expansions of potential malaria vectors. Understanding the adaptive mechanisms of vector species will greatly enhance the effectiveness of vector control strategies.

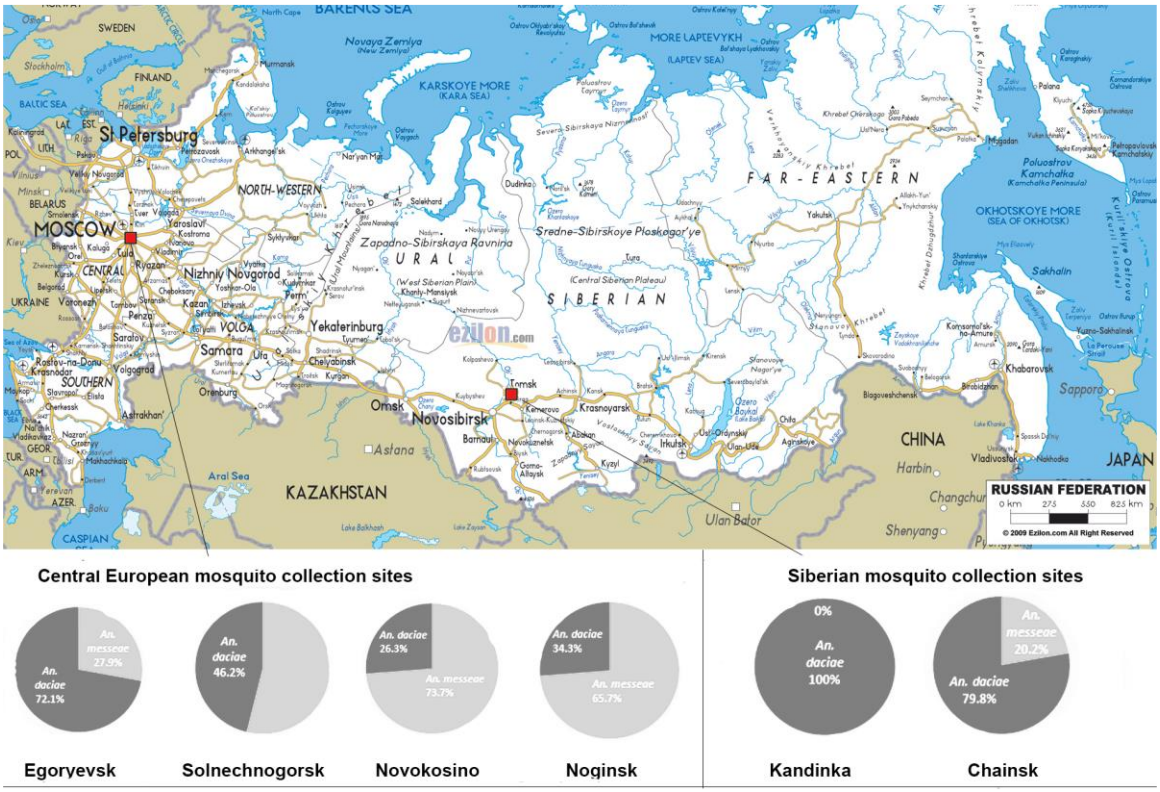


Figure 4.1. The sites of mosquito collections in Central European and Siberian parts of Russia. The ratios of *An. messeae s.s.*:*An. daciae* is shown in pie charts for each population. *An. daciae* is indicated with boulder gray, *An. messeae s.s.* with lighter gray color.

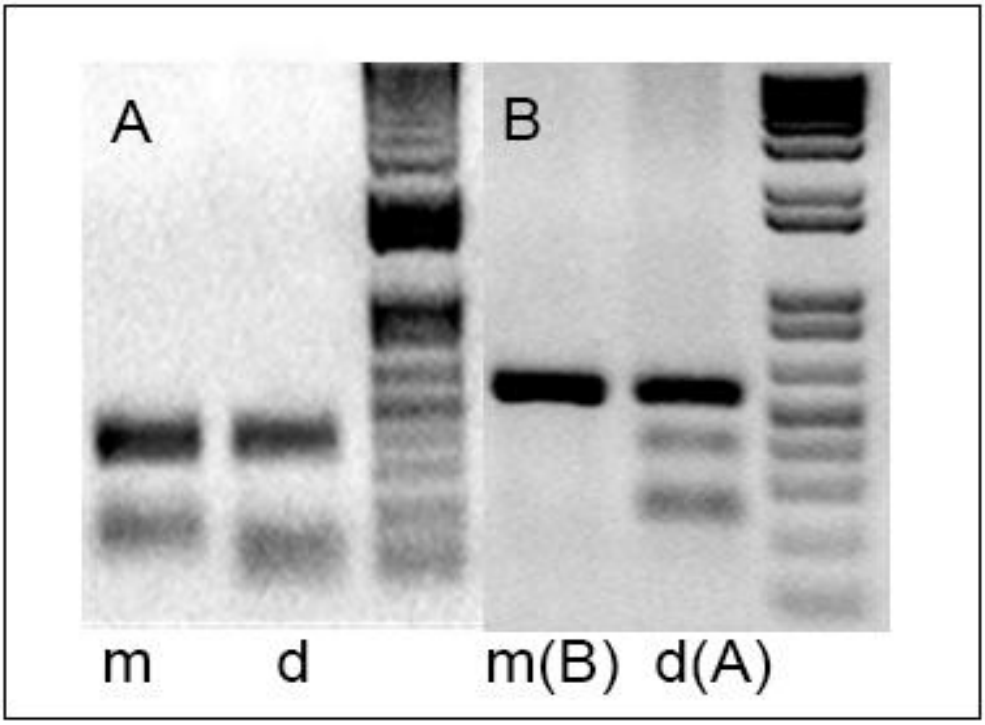


Figure 4.2, A-B. Digestion of the same two samples of *An. messeae s.l.* simultaneously with enzymes BstUI (A) and FokI (B). (A)Digestion of samples with BstUI. Fragment sizes correspond to both *An. messeae s.s.* and *An. daciae.* B) digestion of the same samples with FokI. Fragment sizes of the same products correspond to B (B) and A (A) form of *An. messeae s.l.* Letter m indicates *An. messeae s.s.*, d indicates *An. daciae*.

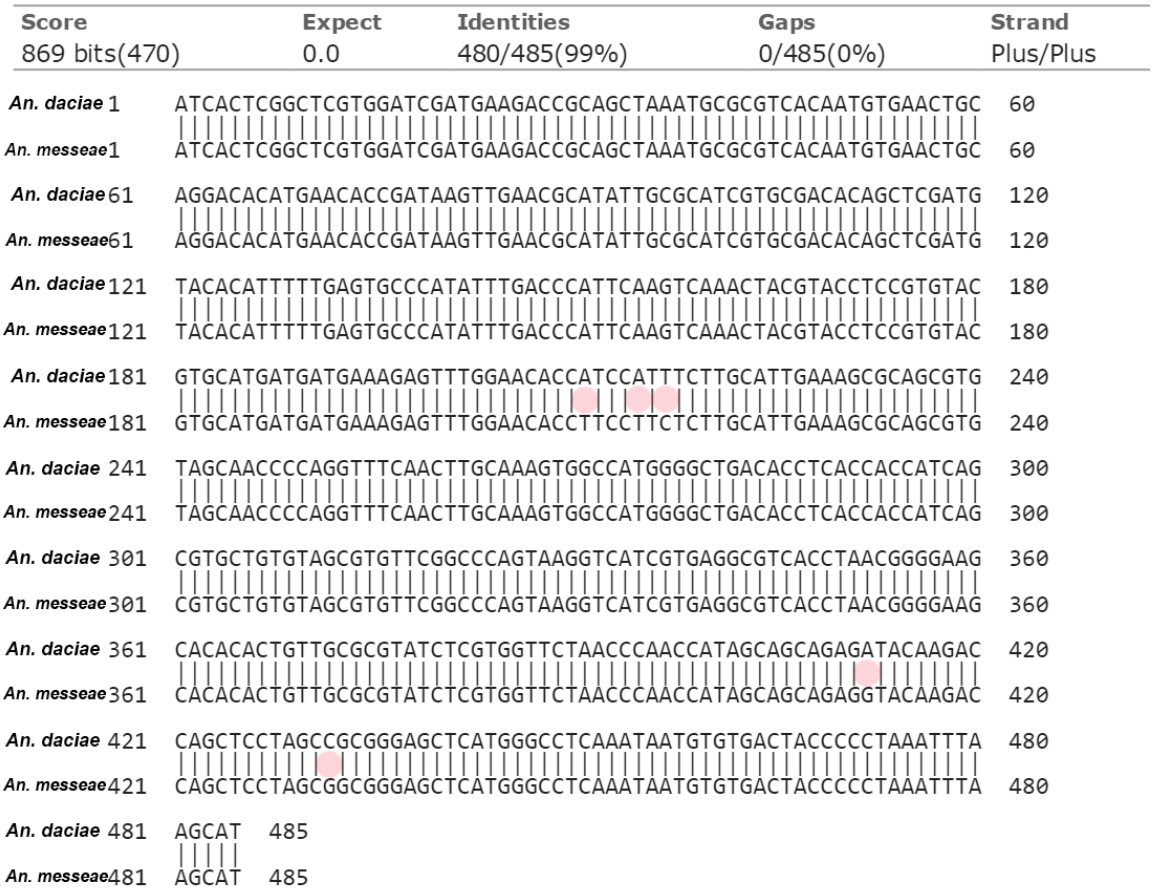


Figure 4.3. BLAST alignment of *An. daciae* ITS2 and *An. messeae s.s.* ITS2 sequences. Five SNPs are highlighted in pink.

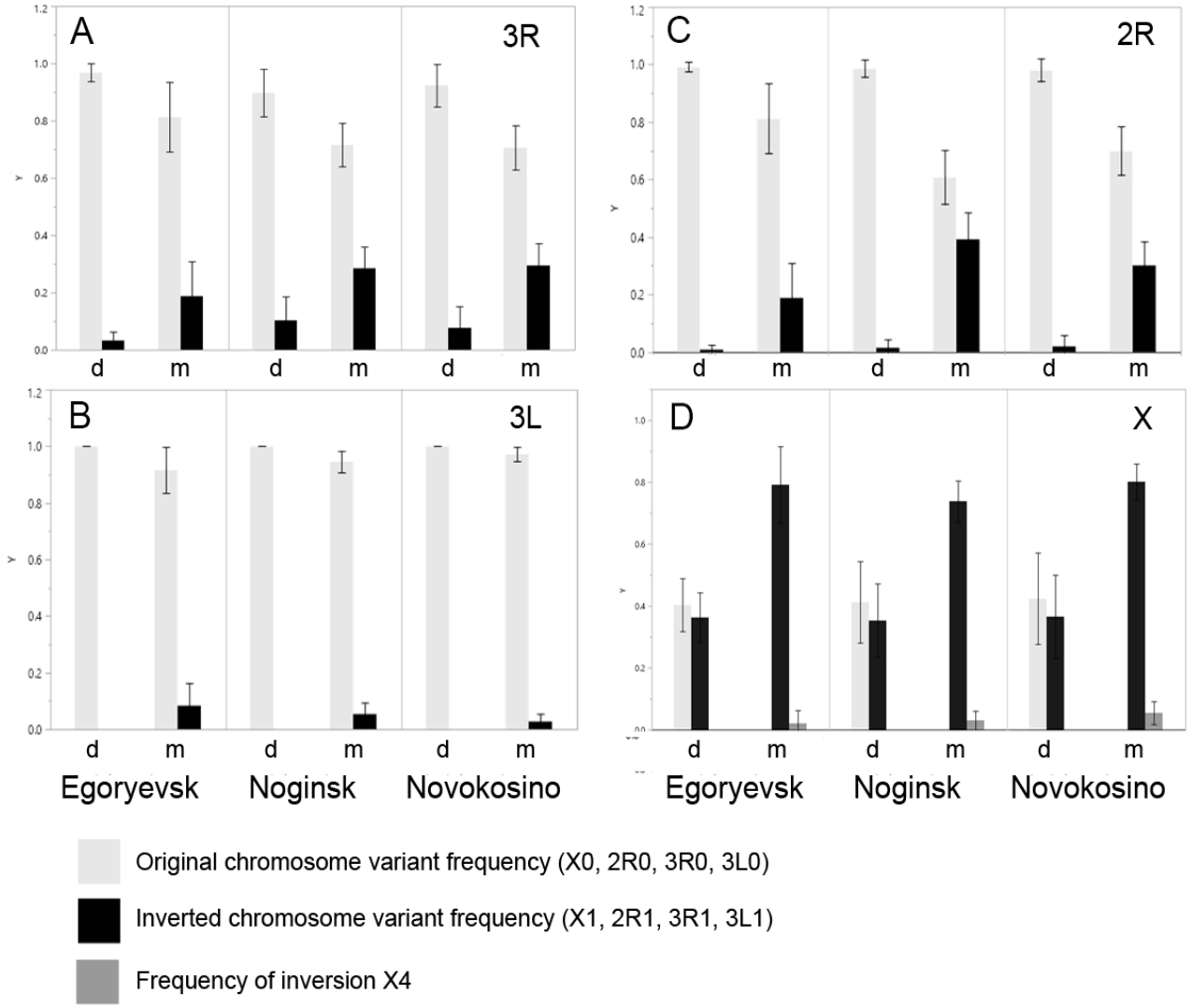


Figure 4.4 A-D. Calculated mean of inversion frequencies for An. messeae s.s. and An. daciae in three locations. (A) 3R1 inversion frequency. (B) 3L1 inversion frequency. (C) 2R1 inversion frequency. (D) X1 and X4 inversions frequency. Original chromosome variant is indicated in light gray, inverted variant is indicated in black color. Boulder gray color indicates X4 inversion. Letter d indicates *An. daciae*, letter m indicates *An. messeae s.s.* Error bars constructed using 0.95 confidence interval of the mean.

Table 4.1. Chromosomal variant frequencies for *An. messeae* and *An. daciae* in three populations.

Location	Species	X0	X1	X4	2R0	2R1	3R0	3R1	3L0	3L1
Egoryevsk	*An. messeae s.s.*	0	0.9744	0.0256	0.8125	0.1875	0.8125	0.1875	0.9167	0.0833
	An. daciae	0.5263	0.4737	0	0.992	0.008	0.9677	0.0323	1	0
Novokosino	*An. messeae s.s.*	0	0.9365	0.0635	0.7027	0.2973	0.7027	0.2973	0.973	0.027
	An. daciae	0.5366	0.4634	0	0.9808	0.0192	0.9231	0.0769	1	0
Noginsk	*An. messeae s.s.*	0	0.96	0.03	0.6077	0.3923	0.7154	0.2846	0.9453	0.0547
	An. daciae	0.5472	0.4528	0	0.9853	0.0147	0.897	0.103	1	0

Table 4.2. Correlation matrices of all observed chromosome variants. Marked correlations are significant at P<0.05000.

Karyotype	X0	X1	X4	2R0	2R1	3R0	3R1	3L0	3L1
X0	1.00000	**-0.68236**	**-0.13571**	**0.39333**	**-0.39333**	**0.27626**	**-0.27626**	**0.15139**	**-0.14632**
X1	**-0.68236**	1.00000	**-0.11998**	**-0.34096**	**0.34096**	**-0.20735**	**0.20735**	**-0.17288**	**0.15903**
X4	**-0.13571**	**-0.11998**	1.00000	0.00644	-0.00644	-0.06727	0.06727	-0.01930	0.02361
2R0	**0.39333**	**-0.34096**	0.00644	1.00000	-1.00000	**0.17283**	**-0.17283**	0.10799	-0.08922
2R1	**-0.39333**	**0.34096**	-0.00644	-1.00000	1.00000	**-0.17283**	**0.17283**	-0.10799	0.08922
3R0	**0.27626**	**-0.20735**	-0.06727	**0.17283**	**-0.17283**	1.00000	-1.00000	0,05133	-0.03072
3R1	**-0.27626**	**0.20735**	0.06727	**-0.17283**	**0.17283**	-1.00000	1.00000	-0.05133	0.03072
3L0	**0.15139**	**-0.17288**	-0.01930	0.10799	-0.10799	0,05133	-0.05133	1.00000	**-0.96274**
3L1	**-0.14632**	**0.15903**	0.02361	-0.08922	0.08922	-0.03072	0.03072	**-0.96274**	1.00000

Table 4.3. Genetic distance measurement by fixation indices (Fst) for *An. messeae s.s.* and *An. daciae* subpopulations.

	Egoryevsk	Noginsk	Novokosino
X	0.319	0.296	0.304
2R	0.26	0.242	0.245
3R	0.15	0.15	0.15

Supplementary figure S1. Mosquito collection sites in Central European part of Russia.
A. Egoryevsk breeding site.
B. Novokosino breeding site.
C. Noginsk breeding site.

Supplementary table S1. Karyotyping and genotyping of specimens based on ITS2 sequence. Accession letters indicate region (M – Moscow) and location (E – Egoryevsk, N – Novokosino, No – Noginsk. Accession numbers indicate specimen.

Accession number	Species	SNP221	SNP225	SNP227	SNP422	SNP442	XL	2R	3R	3L
ME1	*An. daciae*	T/A	T	T	A	C	OO	OO	OO	OO
ME2	*An. daciae*	A	T/A	T	A	C	O	OO	OO	OO
ME3	*An. messeae*	T	T	C	G	G	11	11	OO	OO
ME4	*An. daciae*	T/A	T/A	T	A	C	11	OO	OO	OO
ME5	*An. beklemishevi*						O1	St	St	St
ME6	*An. daciae*	T/A	T	T	A	C	11	OO	OO	OO
ME7	*An. daciae*	T	T	T	A	C	O	OO	OO	OO
ME8	*An. beklemishevi*						O	St	St	St
ME9	poor data						O1	OO	OO	OO
ME10	*An. daciae*	T/A	T	T	A	C	O1	OO	OO	OO
ME11	*An. daciae*	A	T/A	T	A	C	O1	OO	OO	OO
ME12	*An. daciae*	A	A	T	A	C	O1	OO	OO	OO
ME13	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
ME14	*An. messeae*	T	T	C	G	G	1	O1	OO	O1
ME15	*An. daciae*	A	T/A	T	A	C	O1	OO	OO	OO
ME16	*An. messeae*	T	T	C	G	G	11	OO	11	OO
ME17	*An. daciae*	A	T/A	T	A	C	O1	OO	OO	OO
ME18	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
ME19	*An. daciae*	T	T	T	A	C	OO	OO	OO	OO
ME20	*An. messeae*	T	T	C	G	G	11	OO	OO	O1
ME21	hybrid *An. messeae-An. daciae*	T	T	C	R	S	O4	OO	OO	OO
ME22	*An. daciae*	A	T	T	A	C	O1	OO	OO	OO
ME23	*An. daciae*	A	T/A	T	A	C	1	OO	OO	OO
ME24	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
ME25	*An. messeae*	T	T	C	G	G	4	OO	OO	OO
ME26	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
ME27	*An. daciae*	A	T/A	T	A	C	O	OO	OO	OO
ME28	*An. daciae*	A	T	T	A	C	O	OO	OO	OO
ME29	*An. messeae*	T	T	C	G	G	1	OO	O1	O1
ME30	*An. daciae*	A	T	C/T	A	C	1	OO	OO	OO
ME31	poor data						11	OO	OO	OO
ME32	*An. daciae*	T	T	T	A	C	1	OO	OO	OO

ME33	*An. daciae*	T/A	T	T	A	C	O1	OO	OO	OO
ME34	*An. daciae*	A	T	T	A	C	OO	OO	OO	OO
ME35	*An. maculipennis*						OO	St	St	St
ME36	*An. daciae*	T/A	T	T	A	C	1	OO	OO	OO
ME37	*An. daciae*	T	T	T	A	C	1	OO	OO	OO
ME38	*An. daciae*	A	T	T	A	C	O	OO	OO	OO
ME39	*An. daciae*	T/A	T	T	A	C	1	OO	OO	OO
ME40	*An. daciae*	A	T	T	A	C	1	OO	OO	OO
ME41	*An. daciae*	T/A	T	T	A	C	O	O4*	OO	OO
ME42	*An. daciae*	T	T	T	A	C	O1	OO	OO	OO
ME43	*An. daciae*	A	A	T	A	C	O1	OO	OO	OO
ME44	*An. daciae*	A	T/A	T	A	C	11	OO	OO	OO
ME45	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
ME46	*An. messeae*	T	T	C	G	G	1	OO	OO	O1
ME47	*An. daciae*	T	T	T	A	C	O	OO	OO	OO
ME48	not sequenced						1	OO	OO	OO
ME49	*An. daciae*	A	T	T	A	C	OO	OO	OO	OO
ME50	*An. daciae*	T/A	T	T	A	C	11	OO	OO	OO
ME51	*An. daciae*	A	T/A	T	A	C	O1	OO	OO	OO
ME52	*An. daciae*	A	T/A	T	A	C	1	OO	OO	OO
ME53	*An. daciae*	T/A	T	T	A	C	O	OO	O1	OO
ME54	*An. maculipennis*						O	St	St	St
ME55	*An. daciae*	A	T/A	T	A	C	O	OO	O1	OO
ME56	*An. daciae*	A	T/A	T	A	C	1	OO	OO	OO
ME57	*An. daciae*	T/A	T	T	A	C	O1	OO	OO	OO
ME58	*An. daciae*	A	T/A	T	A	C	11	O1	OO	OO
ME59	*An. maculipennis*						O	St	St	St
ME60	not sequenced						O1	OO	OO	OO
ME61	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
ME62	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
ME63	*An. daciae*	T/A	T/A	T	A	C	OO	OO	OO	OO
ME64	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
ME65	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
ME66	*An. daciae*	A	T/A	T	A	C	1	OO	OO	OO
ME67	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
ME68	*An. daciae*	A	T/A	T	A	C	OO	OO	OO	OO
ME69	*An. messeae*	T	T	C	G	G	11	O1	OO	OO

ME70	not sequenced						OO	OO	O1	OO
ME71	*An. daciae*	A	T	T	A	C	O1	OO	OO	OO
ME72	*An. daciae*	T/A	T/A	T	A	C	OO	OO	OO	OO
ME73	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
ME74	*An. daciae*	A	A	T	A	C	OO	OO	OO	OO
ME75	*An. daciae*	A	A	T	A	C	O1	OO	OO	OO
ME76	*An. messeae*	T	T	C	G	G	1	OO	OO	O1
ME77	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
ME78	*An. daciae*	T/A	T	T	A	C	O	OO	O1	OO
ME79	*An. daciae*	A	T/A	T	A	C	O1	OO	OO	OO
ME80	*An. daciae*	A	T	T	A	C	1	OO	O1	OO
ME81	*An. daciae*	T/A	T/A	T	A	C	O1	OO	OO	OO
ME82	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
ME83	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
ME84	*An. daciae*	A	T/A	T	A	C	O	OO	O1	OO
ME85	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
ME86	*An. daciae*	A	T	T	A	C	O1	OO	OO	OO
ME87	*An. beklemishevi*						O1	St	St	St
ME88	*An. daciae*	A	T/A	T	A	C	1	OO	OO	OO
ME89	*An. daciae*	A	T/A	T	A	C	O	OO	OO	OO
ME90	*An. daciae*	A	A	T	A	C	OO	OO	OO	OO
ME91	*An. daciae*	T/A	T/A	T	A	C	11	OO	OO	OO
ME92	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
ME93	*An. daciae*	A	A	T	A	C	O	OO	OO	OO
ME94	not sequenced						11	OO	OO	OO
ME95	*An. beklemishevi*						OO	St	St	St
ME96	*An. daciae*	A	A	T	A	C	O1	OO	OO	OO
ME97	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
ME98	*An. daciae*	A	A	T	A	C	O	OO	OO	OO
ME99	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
ME100	*An. daciae*	T/A	T/A	T	A	C	O1	OO	OO	OO
MN1	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MN2	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MN3	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN4	*An. daciae*	A	T	T	A	C	O1	OO	O1	OO
MN5	*An. daciae*	A	A	T	A	C	11	OO	OO	OO
MN6	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MN7	*An. messeae*	T	T	C	G	G	1	O1	OO	OO

MN8	*An. messeae*	T	T	C	G	G	14	OO	O1	O1
MN9	*An. messeae*	T	T	C	G	G	14	OO	OO	OO
MN10	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN11	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MN12	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MN13	*An. messeae*	T	T	C	G	G	11	OO	11	OO
MN14	*An. messeae*	T	T	C	G	G	11	OO	O1	O1
MN15	*An. messeae*	T	T	C	G	G	11	11	OO	OO
MN16	*An. messeae*	T	T	C	G	G	11	11	O1	OO
MN17	*An. daciae*	A	T	T	A	C	O	OO	OO	OO
MN18	*An. maculipennis*						O1	St	St	St
MN19	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN20	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MN21	*An. messeae*	T	T	C	G	G	11	11	O1	OO
MN22	*An. daciae*	A	T	T	A	C	O1	OO	O1	OO
MN23	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN24	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MN25	*An. daciae*	A	A	T	A	C	OO	OO	OO	OO
MN26	*An. messeae*	T	T	C	G	G	11	11	O1	OO
MN27	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MN28	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN29	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MN30	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MN31	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MN32	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MN33	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MN34	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN35	*An. daciae*	A	A	T	A	C	O1	OO	OO	OO
MN36	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MN37	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN38	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN39	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN40	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN41	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN42	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MN43	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MN44	*An. daciae*	A	T	T	A	C	OO	OO	OO	OO
MN45	*An. daciae*	A	T	T	A	C	1	OO	OO	OO
MN46	*An. messeae*	T	T	C	G	G	11	O1	OO	O1
MN47	*An. daciae*	A	A	T	A	C	1	OO	O1	OO

MN48	*An. messeae*	T	T	C	G	G	14	OO	11	OO
MN49	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN50	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN51	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
MN52	*An. daciae*	A	T	T	A	C	O1	OO	OO	OO
MN53	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN54	*An. messeae*	T	T	C	G	G	1	OO	11	OO
MN55	*An. messeae*	T	T	C	G	G	1	O1	OO	O1
MN56	*An. messeae*	T	T	C	G	G	1	O1	OO	OO
MN57	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN58	*An. daciae*	A	T/A	T	A	C	O	OO	OO	OO
MN59	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MN60	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN61	*An. messeae*	T	T	C	G	G	1	O1	OO	OO
MN62	*An. daciae*	A	A	T	A	C	O	OO	OO	OO
MN63	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN64	*An. messeae*	T	T	C	G	G	14	11	O1	OO
MN65	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MN66	*An. daciae*	A	A	T	A	C	O1	OO	OO	OO
MN67	*An. messeae*	T	T	C	G	G	11	OO	11	OO
MN68	*An. daciae*	A	T	T	A	C	11	OO	OO	OO
MN69	*An. daciae*	A	A	T	A	C	O	OO	OO	OO
MN70	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MN71	*An. daciae*	A	A	T	A	C	O1	OO	OO	OO
MN72	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN73	*An. messeae*	T	T	C	G	G	1	11	OO	OO
MN74	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN75	*An. daciae*	A	A	T	A	C	O	OO	OO	OO
MN76	*An. daciae*	A	T	T	A	C	O1	OO	OO	OO
MN77	*An. messeae*	T	T	C	G	G	14	OO	O1	OO
MN78	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MN79	*An. daciae*	T	T/A	T	A	C	OO	OO	OO	OO
MN80	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
MN81	*An. daciae*	T/A	T	T	A	C	OO	OO	OO	OO
MN82	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MN83	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MN84	*An. daciae*	T/A	A	T	A	C	O1	OO	OO	OO
MN85	*An. messeae*	T	T	C	G	G	1	O1	OO	OO
MN86	*An. daciae*	A	T/A	T	A	C	11	OO	O1	OO
MN87	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MN88	*An. messeae*	T	T	C	G	G	11	OO	OO	OO

MN89	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MN90	*An. messeae*	T	T	C	G	G	11	11	O1	OO
MN91	*An. messeae*	T	T	C	G	G	1	11	11	OO
MN92	*An. messeae*	T	T	C	G	G	14	OO	O1	OO
MN93	*An. daciae*	A	A	T	A	C	OO	OO	OO	OO
MN94	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MN95	*An. messeae*	T	T	C	G	G	14	11	OO	OO
MN96	*An. messeae*	T	T	C	G	G	1	11	O1	OO
MN97	*An. messeae*	T	T	C	G	G	14	OO	O1	OO
MN98	*An. messeae*	T	T	C	G	G	11	OO	11	OO
MN99	*An. daciae*	A	T/A	T	A	C	1	OO	OO	OO
MN100	*An. messeae*	T	T	C	G	G	11	OO	11	OO
MNo1	*An. messeae*	T	T	C	G	G	1	11	OO	OO
MNo2	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MNo3	*An. messeae*	T	T	C	G	G	11	O1	OO	O1
MNo4	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MNo5	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MNo6	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MNo7	*An. daciae*	A	A	T	A	C	O	OO	O1	OO
MNo8	*An. daciae*	T	T	T	A	C	OO	OO	OO	OO
MNo9	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MNo10	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MNo11	*An. messeae*	T	T	C	G	G	1	11	OO	OO
MNo12	*An. messeae*	T	T	C	G	G	4	11	OO	OO
MNo13	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MNo14	*An. messeae*	T	T	C	G	G	11	O1	11	OO
MNo15	*An. messeae*	T	T	C	G	G	11	11	OO	O1
MNo16	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MNo17	*An. messeae*	T	T	C	G	G	1	11	OO	OO
MNo18	*An. messeae*	T	T	C	G	G	14	OO	OO	OO
MNo19	*An. daciae*	T	T	T	A	C	11	OO	OO	OO
MNo20	*An. messeae*	T	T	C	G	G	11	11	O1	OO
MNo21	*An. daciae*	A	T	T	A	C	11	OO	OO	OO
MNo22	*An. messeae*	T	T	C	G	G	11	11	OO	OO
MNo23	*An. daciae*	A	T	T	A	C	OO	OO	OO	OO
MNo24	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MNo25	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MNo26	*An. daciae*	A	A	T	A	C	OO	OO	OO	OO
MNo27	*An. messeae*	T	T	C	G	G	11	O1	O1	O1
MNo28	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MNo29	*An. daciae*	A	A	T	A	C	OO	OO	OO	OO

MNo30	*An. messeae*	T	T	C	G	G	1	11	OO	OO
MNo31	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MNo32	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MNo33	*An. daciae*	A	T	T	A	C	11	OO	OO	OO
MNo34	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
MNo35	*An. messeae*	T	T	C	G	G	1	OO	11	OO
MNo36	*An. daciae*	T	T	T	A	C	O1	OO	OO	OO
MNo37	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MNo38	*An. messeae*	T	T	C	G	G	1	11	OO	OO
MNo39	*An. daciae*	T	T	T	A	C	O1	OO	O1	OO
MNo40	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MNo41	*An. daciae*	T	T	T	A	C	O1	OO	OO	OO
MNo42	*An. messeae*	T	T	C	G	G	11	11	OO	OO
MNo43	*An. daciae*	A	T	T	A	C	14	OO	OO	OO
MNo44	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MNo45	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MNo46	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MNo47	*An. daciae*	A	T	T	A	C	OO	OO	OO	OO
MNo48	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MNo49	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MNo50	*An. daciae*	A	T	T	A	C	O1	OO	OO	OO
MNo51	*An. daciae*	T	T	T	A	C	O	OO	OO	OO
MNo52	*An. daciae*	A	T	T	A	C	O	OO	O1	OO
MNo53	*An. daciae*	T	T	T	A	C	1	OO	OO	OO
MNo54	*An. daciae*	A	A	T	A	C	O	OO	O1	OO
MNo55	*An. messeae*	T	T	C	G	G	11	OO	OO	O1
MNo56	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MNo57	*An. messeae*	T	T	C	G	G	1	11	O1	OO
MNo58	*An. daciae*	A	T	T	A	C	OO	OO	OO	OO
MNo59	*An. daciae*	T	T	T	A	C	O	OO	OO	OO
MNo60	*An. messeae*	T	T	C	G	G	1	O1	OO	OO
MNo61	*An. messeae*	T	T	C	G	G	1	O1	O1	OO
MNo62	*An. messeae*	T	T	C	G	G	11	OO	OO	OO
MNo63	*An. messeae*	T	T	C	G	G	O1	OO	OO	OO
MNo64	*An. messeae*	T	T	C	G	G	14	OO	OO	OO
MNo65	*An. messeae*	T	T	C	G	G	1	O1	OO	OO
MNo66	*An. messeae*	T	T	C	G	G	1	11	O1	OO
MNo67	*An. daciae*	A	T	T	A	C	O1	OO	OO	OO
MNo68	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
MNo69	*An. messeae*	T	T	C	G	G	1	O1	11	OO
MNo70	*An. messeae*	T	T	C	G	G	11	OO	OO	OO

MNo71	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MNo72	*An. messeae*	T	T	C	G	G	11	O1	O1	OO
MNo73	*An. messeae*	T	T	C	G	G	11	OO	OO	O1
MNo74	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MNo75	*An. daciae*	A	A	T	A	C	O	OO	O1	OO
MNo76	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MNo77	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MNo78	*An. daciae*	A	A	T	A	C	O1	OO	OO	OO
MNo79	*An. daciae*	A	T	T	A	C	OO	OO	OO	OO
MNo80	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MNo81	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MNo82	*An. daciae*	A	T	T	A	C	1	OO	OO	OO
MNo83	*An. messeae*	T	T	C	G	G	11	OO	O1	OO
MNo84	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
MNo85	*An. daciae*	A	A	T	A	C	1	OO	OO	OO
MNo86	*An. daciae*	A	T	T	A	C	1	O1	OO	OO
MNo87	*An. messeae*	T	T	C	G	G	1	OO	O1	OO
MNo88	*An. daciae*	T	T	T	A	C	O1	OO	OO	OO
MNo89	*An. daciae*	A	T	T	A	C	1	OO	11	OO
MNo90	*An. messeae*	T	T	C	G	G	11	OO	11	OO
MNo91	*An. maculipennis*						female, St	St	St	St
MNo92	*An. messeae*	T	T	C	G	G	11	11	O1	O1
MNo93	*An. messeae*	T	T	C	G	G	1	OO	OO	OO
MNo94	*An. messeae*	T	T	C	G	G	11	OO	O1	O1
MNo95	*An. messeae*	T	T	C	G	G	11	O1	OO	OO
MNo96	*An. messeae*	T	T	C	G	G	1	O1	OO	OO
MNo97	*An. daciae*	T	T	T	A	C	O	OO	OO	OO
MNo98	*An. messeae*	T	T	C	G	G	1	O1	OO	OO
MNo99	*An. daciae*	T	T	T	A	C	11	OO	OO	OO
MNo100	*An. messeae*	T	T	C	G	G	11	O1	O1	OO

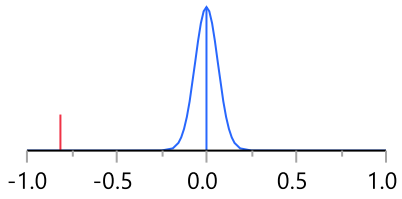
Supplementary table S2. Oneway analysis of X0 variant frequency in *An. messeae s.s.* and *An. daciae*.

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
An. daciae	122	0.819672	0.704371	0.06377	0.6934	0.94592
An. messeae	162	0.006173	0.078567	0.00617	-0.0060	0.01836

t Test

An. messeae-An. daciae
Assuming unequal variances

Difference	-0.81350	t Ratio	-12.6973
Std Err Dif	0.06407	DF	123.27
Upper CL Dif	-0.68668	Prob > |t|	<.0001*
Lower CL Dif	-0.94032	Prob > t	1.0000
Confidence	0.95	Prob < t	<.0001*



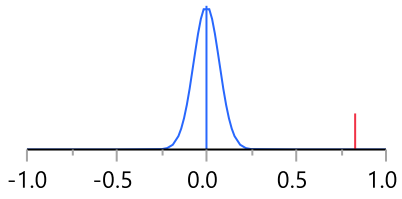
Supplementary table S3. Oneway analysis of X1 variant frequency in *An. messeae s.s.* and *An. daciae*.

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
An. daciae	122	0.72131	0.646188	0.05850	0.6055	0.8371
An. messeae	162	1.54938	0.523396	0.04112	1.4682	1.6306

t Test

An. messeae-An. daciae
Assuming unequal variances

Difference	0.828071	t Ratio	11.57986
Std Err Dif	0.071510	DF	228.2309
Upper CL Dif	0.968975	Prob > |t|	<.0001*
Lower CL Dif	0.687168	Prob > t	<.0001*
Confidence	0.95	Prob < t	1.0000



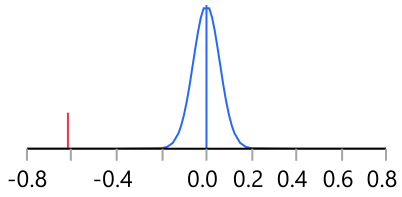
Supplementary table S4. Oneway analysis of 2R0 variant frequency in *An. messeae s.s.* and *An. daciae*.

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
An. daciae	122	1.97541	0.155511	0.01408	1.9475	2.0033
An. messeae	162	1.35802	0.727604	0.05717	1.2451	1.4709

t Test

An. messeae-An. daciae
Assuming unequal variances

Difference	-0.61739	t Ratio	-10.4865
Std Err Dif	0.05887	DF	180.2419
Upper CL Dif	-0.50121	Prob > |t|	<.0001*
Lower CL Dif	-0.73356	Prob > t	1.0000
Confidence	0.95	Prob < t	<.0001*



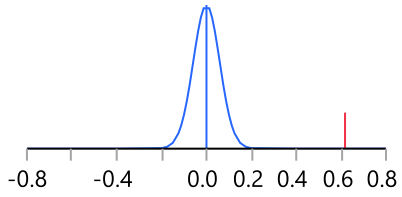
Supplementary table S5. Oneway analysis of 2R1 variant frequency in *An. messeae s.s.* and *An. daciae*.

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
An. daciae	122	0.024590	0.155511	0.01408	-0.0033	0.05246
An. messeae	162	0.641975	0.727604	0.05717	0.5291	0.75487

t Test

An. messeae-An. daciae
Assuming unequal variances

Difference	0.617385	t Ratio	10.48651
Std Err Dif	0.058874	DF	180.2419
Upper CL Dif	0.733557	Prob > |t|	<.0001*
Lower CL Dif	0.501214	Prob > t	<.0001*
Confidence	0.95	Prob < t	1.0000



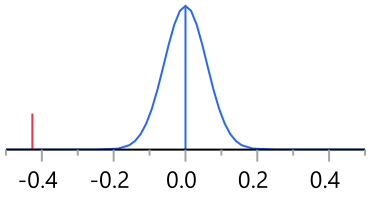
Supplementary table S6. Oneway analysis of 3R0 variant frequency in *An. messeae s.s.* and *An. daciae*.

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
An. daciae	122	1.87705	0.353912	0.03204	1.8136	1.9405
An. messeae	162	1.45062	0.631005	0.04958	1.3527	1.5485

t Test

An. messeae-An. daciae
Assuming unequal variances

Difference	-0.42643	t Ratio	-7.22403
Std Err Dif	0.05903	DF	262.624
Upper CL Dif	-0.31020	Prob > |t|	<.0001*
Lower CL Dif	-0.54266	Prob > t	1.0000
Confidence	0.95	Prob < t	<.0001*



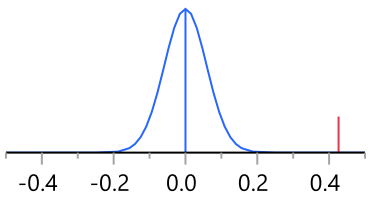
Supplementary table S7. Oneway analysis of 3R1 variant frequency in *An. messeae s.s.* and *An. daciae*.

Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
An. daciae	122	0.122951	0.353912	0.03204	0.05952	0.18639
An. messeae	162	0.549383	0.631005	0.04958	0.45148	0.64729

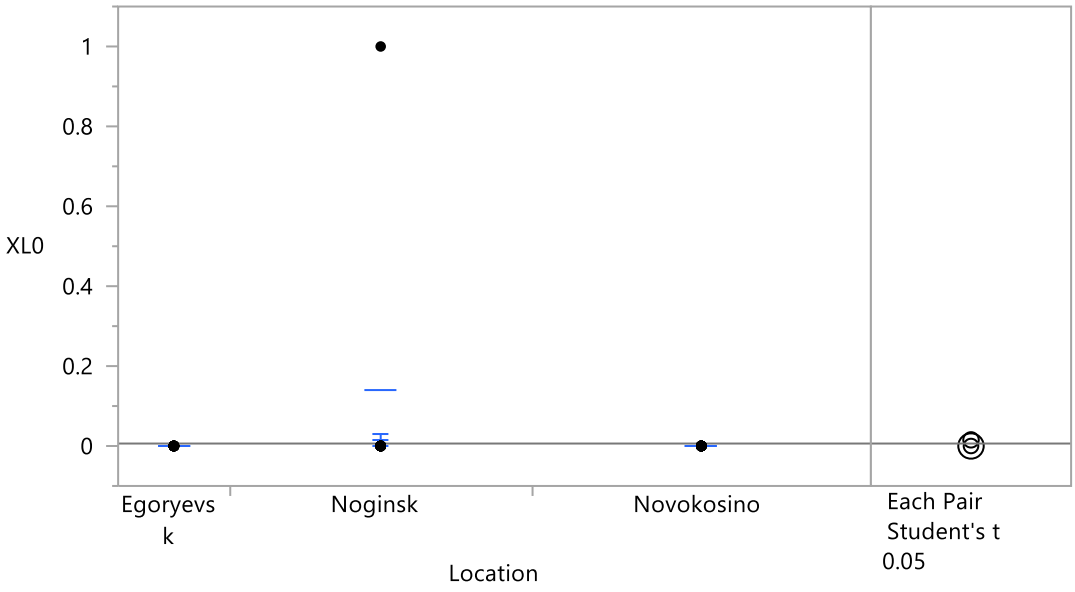
t Test

An. messeae-An. daciae
Assuming unequal variances

Difference	0.426432	t Ratio	7.224033
Std Err Dif	0.059030	DF	262.624
Upper CL Dif	0.542663	Prob > |t|	<.0001*
Lower CL Dif	0.310200	Prob > t	<.0001*
Confidence	0.95	Prob < t	1.0000

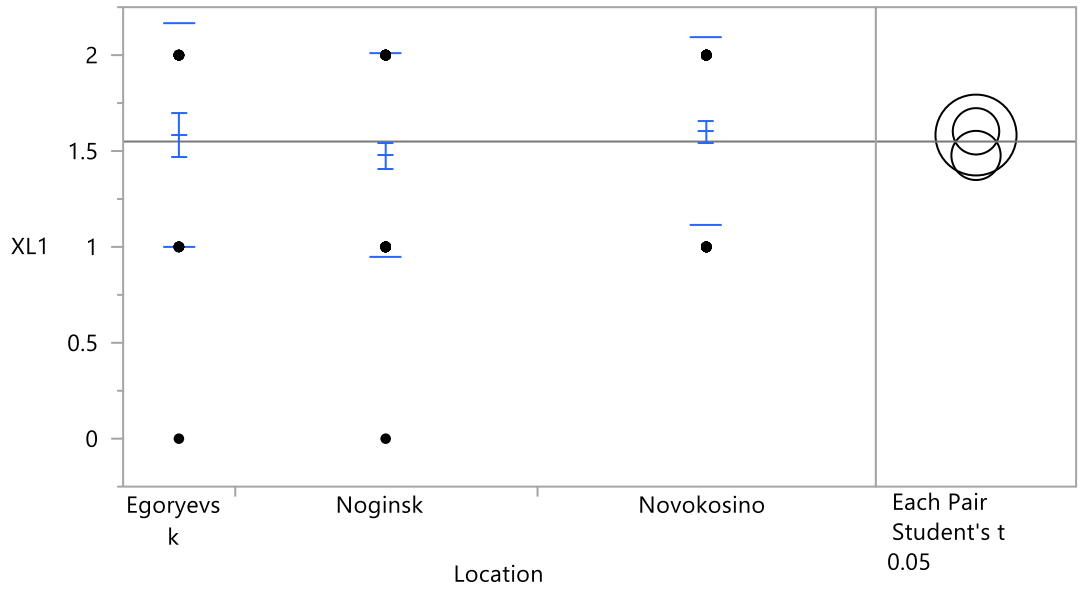


Supplementary table S8. X0 variant frequencies in *An. messeae s.l.* depending on the location.



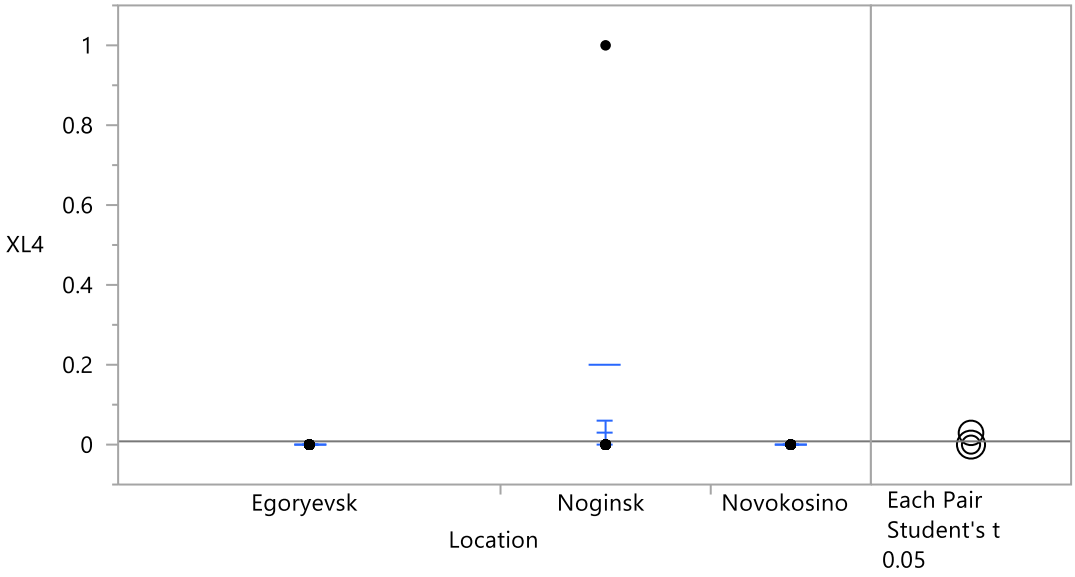
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Noginsk	Egoryevsk	0.0153846	0.0187961	-0.021738	0.0525068	0.4143
Noginsk	Novokosino	0.0153846	0.0134201	-0.011120	0.0418893	0.2534
Novokosino	Egoryevsk	0.0000000	0.0185163	-0.036570	0.0365696	1.0000

Supplementary table S9. X0 variant frequencies in *An. messeae s.l.* depending on the location.



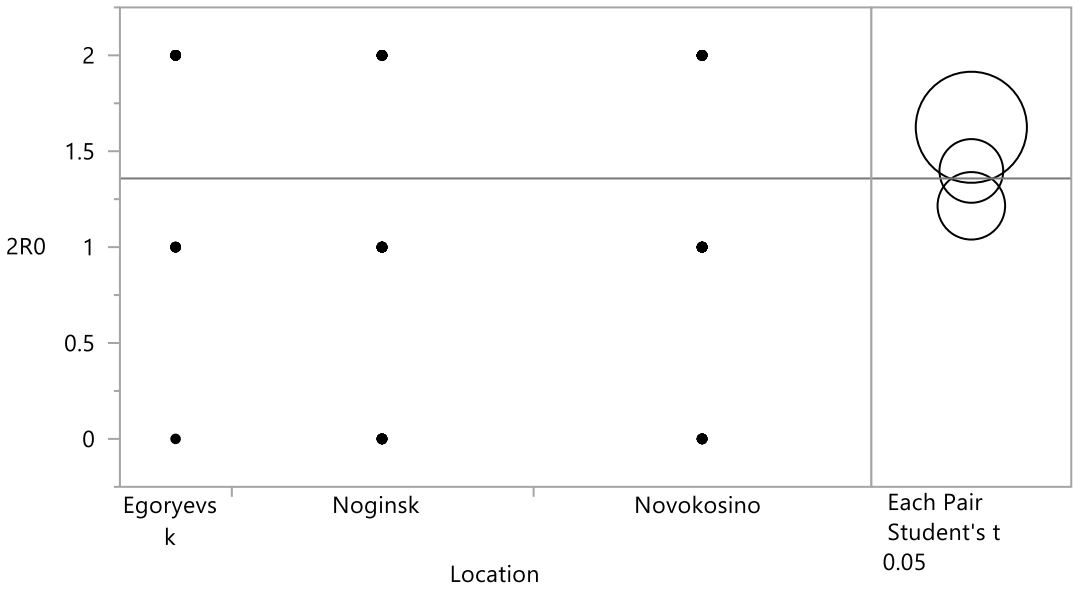
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Novokosino	Noginsk	0.1258166	0.0892294	-0.050411	0.3020444	0.1605
Egoryevsk	Noginsk	0.1064103	0.1249739	-0.140413	0.3532332	0.3958
Novokosino	Egoryevsk	0.0194064	0.1231134	-0.223742	0.2625550	0.8749

Supplementary table S10. X1 variant frequencies in *An. messeae s.l.* depending on the location.



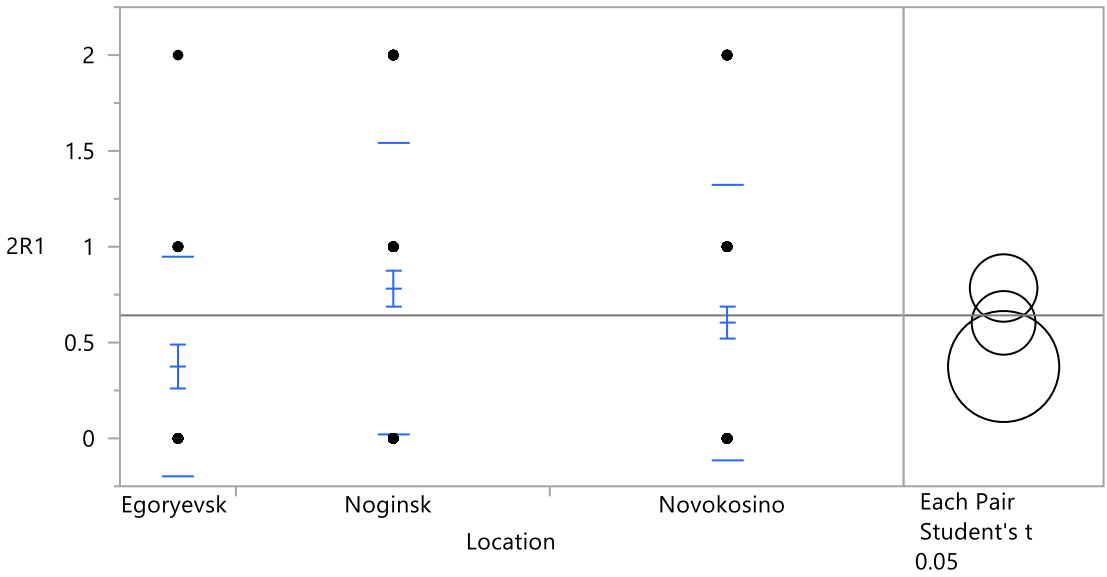
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Noginsk	Egoryevsk	0.0294118	0.0192728	-0.008750	0.0675738	0.1296
Noginsk	Novokosino	0.0294118	0.0235285	-0.017177	0.0760005	0.2137
Novokosino	Egoryevsk	0.0000000	0.0211010	-0.041782	0.0417821	1.0000

Supplementary table S11. 2R0 variant frequencies in *An. messeae s.l.* depending on the location.



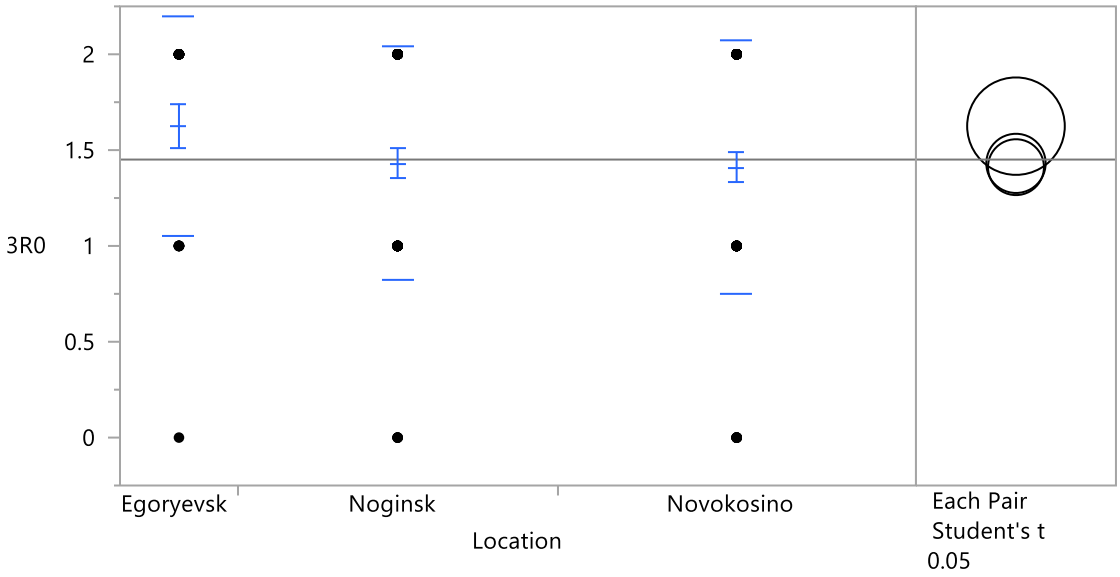
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Egoryevsk	Noginsk	0.4096154	0.1716235	0.070660	0.7485712	0.0182*
Egoryevsk	Novokosino	0.2277397	0.1690686	-0.106170	0.5616495	0.1799
Novokosino	Noginsk	0.1818757	0.1225365	-0.060133	0.4238848	0.1397

Supplementary table S12. 2R1 variant frequencies in *An. messeae s.l.* depending on the location.



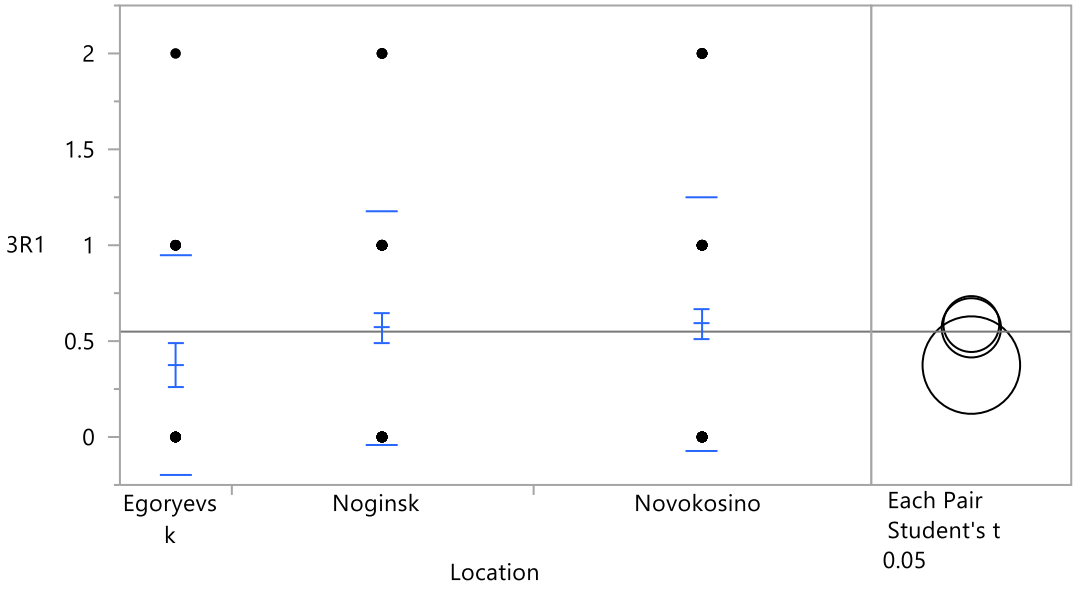
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Noginsk	Egoryevsk	0.4096154	0.1716235	0.070660	0.7485712	0.0182*
Novokosino	Egoryevsk	0.2277397	0.1690686	-0.106170	0.5616495	0.1799
Noginsk	Novokosino	0.1818757	0.1225365	-0.060133	0.4238848	0.1397

Supplementary table S13. 3R0 variant frequencies in *An. messeae s.l.* depending on the location.



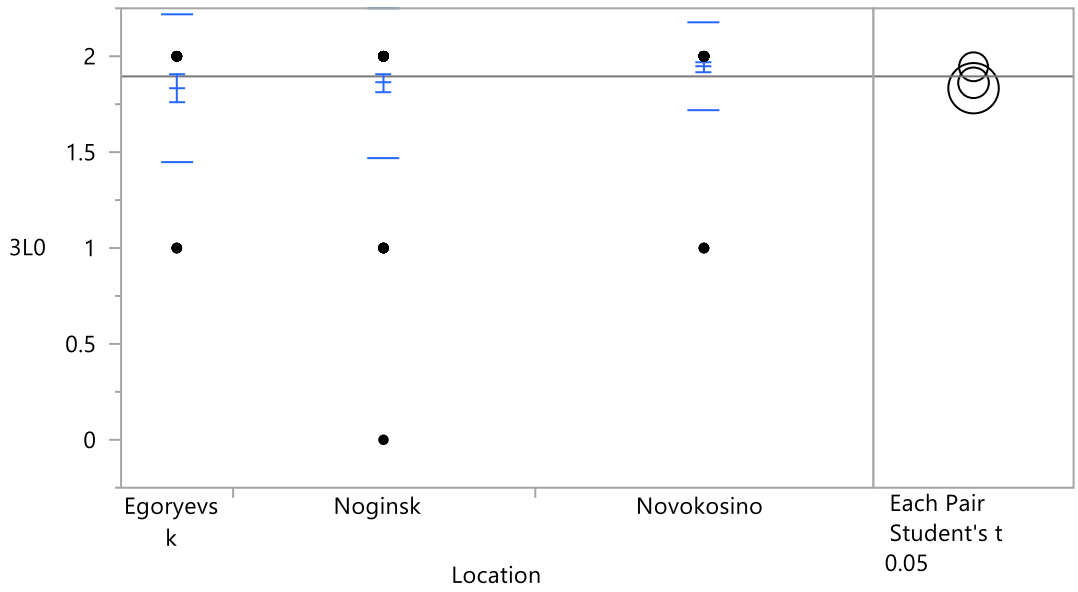
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Egoryevsk	Novokosino	0.2140411	0.1483877	-0.079024	0.5071063	0.1511
Egoryevsk	Noginsk	0.1942308	0.1506301	-0.103263	0.4917247	0.1991
Noginsk	Novokosino	0.0198103	0.1075476	-0.192596	0.2322164	0.8541

Supplementary table S14. 3R1 variant frequencies in *An. messeae s.l.* depending on the location.



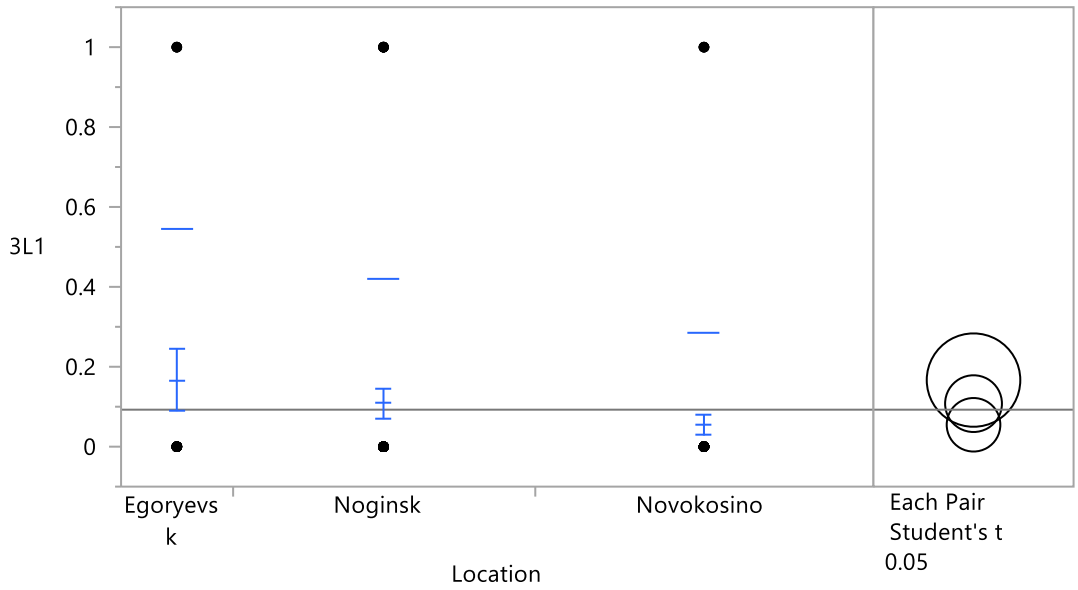
Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Novokosino	Egoryevsk	0.2140411	0.1483877	-0.079024	0.5071063	0.1511
Noginsk	Egoryevsk	0.1942308	0.1506301	-0.103263	0.4917247	0.1991
Novokosino	Noginsk	0.0198103	0.1075476	-0.192596	0.2322164	0.8541

Supplementary table S15. 3L0 variant frequencies in *An. messeae s.l.* depending on the location.



Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Novokosino	Egoryevsk	0.1118721	0.0766394	-0.039490	0.2632347	0.1463
Novokosino	Noginsk	0.0836670	0.0555462	-0.026037	0.1933706	0.1340
Noginsk	Egoryevsk	0.0282051	0.0777976	-0.125445	0.1818550	0.7174

Supplementary table S16. 3L1 variant frequencies in *An. messeae s.l.* depending on the location.



Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
Egoryevsk	Novokosino	0.1118721	0.0682069	-0.022836	0.2465806	0.1029
Egoryevsk	Noginsk	0.0589744	0.0692377	-0.077770	0.1957185	0.3956
Noginsk	Novokosino	0.0528978	0.0494346	-0.044735	0.1505310	0.2862

5. Chapter summary and conclusions

5.1. Chapter 2 review

First attempts to develop physical map for *Cx. quinquefasciatus* polytene chromosomes were done in 1970s. These efforts were pioneering for mosquito chromosome studies, but challenging at the same time. The quality of polytene chromosome spreads for this mosquito, as well as for *Aedes aegypti*, remained tricky. Large genome size, low degree of polyteny and the ability of chromosomes to form ectopic contacts presented an impediment for slide preparation. Another way to develop a chromosome map is to use mitotic chromosomes from imaginal discs of the 4th instar larvae. Using this approach, we developed map for the mitotic chromosomes of the major vector for West Nile fever and lymphatic filariasis, *Cx. quinquefasciatus*.
The nomenclature of physical and genetic linkage maps for *Cx. quinquefasciatus* was conflicting. The first nomenclature suggested by K. Rai named the chromosomes based on the size as 1 – the smallest, 2 – intermediate and 3 – the largest chromosome. The latest genetic linkage maps included three groups: smallest linkage group 1, largest linkage group 2 containing the largest quantity of the genetic markers and intermediate linkage group 3. We established the correspondence between chromosomes and genetic linkage groups by mapping markers originally used for the genetic map and renumbered chromosomes as 1 – the smallest, 2 – the largest and 3 – intermediate. We mapped approximately 50% of available genetic markers to their precise positions on the chromosomes. This mapping effort placed 13% of the genome to chromosomes. In general, the mitotic chromosomes mapping is the most reliable way of detailed physical mapping for *Cx. quinquefasciatus* due to the low numbers of successful preparations for polytene chromosomes with reproducible banding patterns.

5.2. Chapter 3 review

The role that mosquitoes have in transmission of vector-borne diseases makes them an important group for studying genes and genomes: introduction of genome-editing tools may aid in the suppression of mosquito populations and/or reduce the rates of disease transmission. The currently available genome assembly of *Cx. quinquefasciatus* consists of the large number of

small supercontigs, which makes the genome assembly fragmented and gives the possibility of gaps and misassemblies in the genome. In this study, we integrated the idiograms developed in the previous study with 140 genetic supercontigs representing 26.5% of the *Cx. quinquefasciatus* genome. The positions of BAC clones on physical map were compared to the order of available genetic markers on genetic linkage map using linear regression model. A linear regression analysis demonstrated good overall correlation between the positioning of markers on physical and genetic linkage maps. Even with good correlation coefficient, discrepancies were found in markers positions, especially near the centromeres which have lower recombination rates. An integrated physical, genome and chromosome map can be utilized as population genomics tool. Our map will greatly aid in further assembly improvement as this mapping effort evenly covered the whole chromosomal complement.

5.3. Chapter 4 review

Cryptic species complexes are groups of closely related species which are difficult or impossible to distinguish by morphological traits. The complexes are particularly important for arthropod-borne diseases studies because the complex may include highly specialized malaria vectors along with species with lesser susceptibility to *Plasmodium*. Mosquitoes from *Anopheles maculipennis* group of species are known from the early 20th century (Falleroni 1926); sibling species from the complex are effective malaria vectors in Europe and Asia.

For the first time, we confirmed the presence of the newly described species from Maculipennis group, *An. daciae*, in distant regions of Russia. The correct diagnostics of *An. daciae* poses significant challenge because of only one weak morphological character distinguish it from the sibling species *An. messeae*. Moreover, *An. daciae* and *An. messeae* are sympatric along the range. The molecular data, in particular five fixed SNPs in ITS2 ribosomal DNA, reliably distinct *An. messeae* from *An. daciae*. Genetic polymorphism found in *An. messeae* populations was attributed to *An. daciae* presence in the populations. Using sequencing and restriction fragment length polymorphism assays, we established the correspondence between cytogenetically distinct form A and *An. daciae*. Form B was found to correspond to *An. messeae*. Interestingly, the inversion frequencies and measured genetic distances suggested that *An. messeae s.s.* and *An. daciae* have nearly completed the process of speciation. Further

transcriptome analysis and genome sequencing will add to the analysis of the genomic differentiation level between the two species.

References

1. Cohen, J.I., *Exploring the nature of science through courage and purpose: a case study of Charles Darwin's way of knowing.* Springerplus, 2016. **5**(1): p. 1532.
2. Kitzmiller, J.B., *Mosquito genetics and cytogenetics.* Rev Bras Malariol Doencas Trop, 1953. **5**(4): p. 285-359.
3. Dobzhansky, T., *The genetic basis of evolution.* Sci Am, 1950. **182**(1): p. 32-41, illust.
4. Rai, K.S., *A comparative study of mosquito karyotypes.* Ann ent Soc Am, 1963. **56**: p. 160-170.
5. Rai, K.S., *Cytogenetics of Aedes aegypti.* Bull World Health Organ, 1967. **36**(4): p. 563-5.
6. Rai, K.S. and W.C.t. Black, *Mosquito genomes: structure, organization, and evolution.* Adv Genet, 1999. **41**: p. 1-33.
7. Neafsey, D.E., et al., *Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes.* Science, 2015. **347**(6217): p. 1258522.
8. Ayala, F.J. and M. Coluzzi, *Chromosome speciation: humans, Drosophila, and mosquitoes.* Proc Natl Acad Sci U S A, 2005. **102 Suppl 1**: p. 6535-42.
9. Bosio, C.F., et al., *Quantitative trait loci that control vector competence for dengue-2 virus in the mosquito Aedes aegypti.* Genetics, 2000. **156**(2): p. 687-98.
10. Burt, A., *Heritable strategies for controlling insect vectors of disease.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1645): p. 20130432.
11. Pal, R., *Chromosomes of mosquitoes.* Trans R Soc Trop Med Hyg, 1947. **40**(4): p. 361.

12. Baker, R.H. and M.G. Rabbani, *Complete linkage in females of Culex tritaeniorhynchus mosquitoes.* J Hered, 1970. **61**(2): p. 59-61.

13. Kitzmiller, J.B., *Genetics, cytogenetics, and evolution of mosquitoes.* Adv Genet, 1976. **18**: p. 315-433.

14. Barr, A.R. and C.M. Myers, *Inheritance of a lethal affecting larval antennae in Culex pipiens L. (Diptera: Culicidae).* J Parasitol, 1966. **52**(6): p. 1163-6.

15. Barr, A.R. and S.L. Narang, *The inheritance of enlarged tergum in Culex pipiens.* J Med Entomol, 1972. **9**(6): p. 560-3.

16. Barr, A.R., *Divided-eye, a sex-linked mutation in Culex pipiens L.* J Med Entomol, 1969. **6**(4): p. 393-7.

17. Craig, G.B., Jr. and W.A. Hickey, *Current status of the formal genetics of Aedes aegypti.* Bull World Health Organ, 1967. **36**(4): p. 559-62.

18. Bhalla, S.C. and G.B. Craig, Jr., *Bronze, a female-sterile mutant of Aedes aegypti.* J Med Entomol, 1967. **4**(4): p. 467-76.

19. Coker, W.Z., *The inheritance of DDT resistance in Aedes aegypti* Ann Trop Med Parasitol, 1958. **52**(4): p. 443-55.

20. Dubash, C.J., R.K. Sakai, and R.H. Baker, *Esterases in the malaria vector mosquito, Anopheles culicifacies. Genetic and linkage analyses of an alpha and beta polymorphism.* J Hered, 1982. **73**(3): p. 209-13.

21. Khan, N.H. and A.W. Brown, *Genetical studies on dieldrin-resistance in Aedes aegypti and its cross-resistance to DDT.* Bull World Health Organ, 1961. **24**: p. 519-26.

22. Haridi, A.M., *Linkage studies on DDT and dieldrin resistance in species A and species B of the Anopheles gambiae complex.* Bull World Health Organ, 1974. **50**(5): p. 441-8.

23. Coluzzi, M., et al., *A polytene chromosome analysis of the Anopheles gambiae species complex.* Science, 2002. **298**(5597): p. 1415-8.

24. Coluzzi, M. and A. Coluzzi, *[On various species of Aedes (Diptera, Culcidae) from the Apennines and Gargano]*. Riv Parassitol, 1967. **28**(1): p. 47-61.

25. Dennhofer, L., *[Salivary gland chromosomes of the mosquito Culex pipiens. I. The normal chromosome complement]*. Chromosoma, 1968. **25**(3): p. 365-76.

26. Kreutzer, R.D., et al., *The salivary gland chromosomes of Anopheles punctimacula.* Rev Bras Malariol Doencas Trop, 1969. **21**(3): p. 559-70.

27. Kabanova, V.M., N.N. Kartashova, and V.N. Stegnii, *[Karyological study of natural populations of malarial mosquitoes in the Middle Ob river. I. Characteristics of the karyotype of Anopheles maculipennis messeae]*. Tsitologiia, 1972. **14**(5): p. 630-6.

28. Kreutzer, R.D. and J.B. Kitzmiller, *Hybridization between two species of mosquitoes. Anopheles punctipennis Say and Anopheles perplexens Ludlow.* J Hered, 1972. **63**(4): p. 191-6.

29. Smithson, T.W., *Population cytogenetics of Anopheles freeborni Atiken. A preliminary report.* Proc Pap Annu Conf Calif Mosq Control Assoc, 1970. **38**: p. 91-2.

30. 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.

31. Coluzzi, M., M. Di Deco, and G. Cancrini, *Chromosomal inversions in Anopheles stephensi.* Parassitologia, 1973. **15**(1): p. 129-36.

32. Rinker, D.C., R.J. Pitts, and L.J. Zwiebel, *Disease vectors in the era of next generation sequencing.* Genome Biol, 2016. **17**(1): p. 95.

33. Holt, R.A., et al., *The genome sequence of the malaria mosquito Anopheles gambiae.* Science, 2002. **298**(5591): p. 129-49.

34. Zdobnov, E.M., et al., *Comparative genome and proteome analysis of Anopheles gambiae and Drosophila melanogaster.* Science, 2002. **298**(5591): p. 149-59.

35. Nene, V., et al., *Genome sequence of Aedes aegypti, a major arbovirus vector.* Science, 2007. **316**(5832): p. 1718-23.

36. Arensburger, P., et al., *Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics.* Science, 2010. **330**(6000): p. 86-8.

37. Megy, K., et al., *VectorBase: improvements to a bioinformatics resource for invertebrate vector genomics.* Nucleic Acids Res, 2012. **40**(Database issue): p. D729-34.

38. Lawson, D., et al., *VectorBase: a home for invertebrate vectors of human pathogens.* Nucleic Acids Res, 2007. **35**(Database issue): p. D503-5.

39. Haider, S., et al., *BioMart Central Portal--unified access to biological data.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W23-7.

40. Smedley, D., et al., *The BioMart community portal: an innovative alternative to large, centralized data repositories.* Nucleic Acids Res, 2015. **43**(W1): p. W589-98.

41. Timoshevskiy, V.A., et al., *Fluorescent in situ Hybridization on Mitotic Chromosomes of Mosquitoes.* J Vis Exp, 2012(67).

42. VectorBase, C. *VectorBase Overview*. 2017.

43. Crawford, J.E., et al., *De novo transcriptome sequencing in Anopheles funestus using Illumina RNA-seq technology.* PLoS One, 2010. **5**(12): p. e14202.

44. Price, D.C. and D.M. Fonseca, *Genetic divergence between populations of feral and domestic forms of a mosquito disease vector assessed by transcriptomics.* PeerJ, 2015. **3**: p. e807.

45. Black, W.C.t. and G.C. Lanzaro, *Distribution of genetic variation among chromosomal forms of Anopheles gambiae s.s: introgressive hybridization, adaptive inversions, or recent reproductive isolation?* Insect Mol Biol, 2001. **10**(1): p. 3-7.

46. della Torre, A., et al., *Speciation within Anopheles gambiae--the glass is half full.* Science, 2002. **298**(5591): p. 115-7.

47. Kamali, M., et al., *A new chromosomal phylogeny supports the repeated origin of vectorial capacity in malaria mosquitoes of the Anopheles gambiae complex.* PLoS Pathog, 2012. **8**(10): p. e1002960.

48. Besansky, N.J., *Complexities in the analysis of cryptic taxa within the genus Anopheles.* Parassitologia, 1999. **41**(1-3): p. 97-100.

49. della Torre, A., Z. Tu, and V. Petrarca, *On the distribution and genetic differentiation of Anopheles gambiae s.s. molecular forms.* Insect Biochem Mol Biol, 2005. **35**(7): p. 755-69.

50. Favia, G. and C. Louis, *Molecular identification of chromosomal forms of Anopheles gambiae sensu stricto.* Parassitologia, 1999. **41**(1-3): p. 115-8.

51. Wondji, C., F. Simard, and D. Fontenille, *Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of Anopheles gambiae in an area of sympatry.* Insect Mol Biol, 2002. **11**(1): p. 11-9.

52. Toe, K.H., et al., *The recent escalation in strength of pyrethroid resistance in Anopheles coluzzi in West Africa is linked to increased expression of multiple gene families.* BMC Genomics, 2015. **16**: p. 146.

53. Robert, V., et al., *Mosquitoes and malaria transmission in irrigated rice-fields in the Benoue valley of northern Cameroon.* Acta Trop, 1992. **52**(2-3): p. 201-4.

54. Harbach, R.E., *The classification of genus Anopheles (Diptera: Culicidae): a working hypothesis of phylogenetic relationships.* Bull Entomol Res, 2004. **94**(6): p. 537-53.

55. Andersen, L.K. and M.D. Davis, *Climate change and the epidemiology of selected tick-borne and mosquito-borne diseases: update from the International Society of Dermatology Climate Change Task Force.* Int J Dermatol, 2017. **56**(3): p. 252-259.

56. Tolle, M.A., *Mosquito-borne diseases.* Curr Probl Pediatr Adolesc Health Care, 2009. **39**(4): p. 97-140.

57. WHO, World Health Organization, 2012(Fact sheet 117).

58. Ramana, B., *Malaria: who is at fault?* World J Surg, 2007. **31**(11): p. 2072-4.

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

60. WHO. *WHO Malaria fact sheet*. 2016.

61. Huntington, M.K., J. Allison, and D. Nair, *Emerging Vector-Borne Diseases.* Am Fam Physician, 2016. **94**(7): p. 551-557.

62. CDC. *CDC West Nile fever data sheet*. 2016.

63. Nutman, T.B., *Insights into the pathogenesis of disease in human lymphatic filariasis.* Lymphat Res Biol, 2013. **11**(3): p. 144-8.

64. Noroes, J. and G. Dreyer, *A mechanism for chronic filarial hydrocele with implications for its surgical repair.* PLoS Negl Trop Dis, 2010. **4**(6): p. e695.

65. Tauil, P.L., *Malaria control--multisectorial approach.* Mem Inst Oswaldo Cruz, 1992. **87 Suppl 3**: p. 349-50.

66. Richards, F.O., Jr., et al., *Effects of permethrin-impregnated bed nets on malaria vectors of northern Guatemala.* Bull Pan Am Health Organ, 1994. **28**(2): p. 112-21.

67. Sexton, J.D., *Impregnated bed nets for malaria control: biological success and social responsibility.* Am J Trop Med Hyg, 1994. **50**(6 Suppl): p. 72-81.

68. Gnankine, O., et al., *Insecticide resistance in Bemisia tabaci Gennadius (Homoptera: Aleyrodidae) and Anopheles gambiae Giles (Diptera: Culicidae) could compromise the sustainability of malaria vector control strategies in West Africa.* Acta Trop, 2013. **128**(1): p. 7-17.

69. Wolfe, R.J., *Effects of open marsh water management on selected tidal marsh resources: a review.* J Am Mosq Control Assoc, 1996. **12**(4): p. 701-12.

70. Tabachnick, W.J., *Research Contributing to Improvements in Controlling Florida's Mosquitoes and Mosquito-borne Diseases.* Insects, 2016. **7**(4).

71. Boyce, R., et al., *Bacillus thuringiensis israelensis (Bti) for the control of dengue vectors: systematic literature review.* Trop Med Int Health, 2013. **18**(5): p. 564-77.

72. Rupp, H.R., *Adverse assessments of Gambusia affinis: an alternate view for mosquito control practitioners.* J Am Mosq Control Assoc, 1996. **12**(2 Pt 1): p. 155-9; discussion 160-6.

73. Borkovec, A.B., *Control and management of insect populations by chemosterilants.* Environ Health Perspect, 1976. **14**: p. 103-7.

74. *WHO procedures for certification of malaria elimination.* Wkly Epidemiol Rec, 2014. **89**(29): p. 321-5.

75. Raymond, M., et al., *An overview of the evolution of overproduced esterases in the mosquito Culex pipiens.* Philos Trans R Soc Lond B Biol Sci, 1998. **353**(1376): p. 1707-11.

76. Nasci, R.S., *Control of Aedes albopictus from the perspective of North America.* Parassitologia, 1995. **37**(2-3): p. 123-7.

77. Hemingway, J., et al., *The role of gene splicing, gene amplification and regulation in mosquito insecticide resistance.* Philos Trans R Soc Lond B Biol Sci, 1998. **353**(1376): p. 1695-9.

78. Prapanthadara, L.A., et al., *Characterization of the major form of glutathione transferase in the mosquito Anopheles dirus A.* Biochem Soc Trans, 1995. **23**(1): p. 81S.

79. Kitzmiller, J.B., *Genetic control of mosquitoes.* Am J Trop Med Hyg, 1972. **21**(5): p. 764-71.

80. Hallinan, E. and K.S. Rai, *Radiation sterilization of Aedes aegypti in nitrogen and implications for sterile male technique.* Nature, 1973. **244**(5415): p. 368-9.

81. DeMilo, A.B., A.B. Borkovec, and R.L. Fye, *Insect chemosterilants. Analogues of 2,5-dichloro-N-(2,4-dinitrophenyl)benzenesulfonamide.* J Agric Food Chem, 1976. **25**(1): p. 81-3.

82. Harris, A.F., et al., *Successful suppression of a field mosquito population by sustained release of engineered male mosquitoes.* Nat Biotechnol, 2012. **30**(9): p. 828-30.

83. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.

84. Basu, S., et al., *Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in Aedes aegypti.* Proc Natl Acad Sci U S A, 2015. **112**(13): p. 4038-43.

85. Hall, A.B., et al., *SEX DETERMINATION. A male-determining factor in the mosquito Aedes aegypti.* Science, 2015. **348**(6240): p. 1268-70.

86. Kistler, K.E., L.B. Vosshall, and B.J. Matthews, *Genome engineering with CRISPR-Cas9 in the mosquito Aedes aegypti.* Cell Rep, 2015. **11**(1): p. 51-60.

87. Gantz, V.M., et al., *Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi.* Proc Natl Acad Sci U S A, 2015. **112**(49): p. E6736-43.

88. Gantz, V.M. and E. Bier, *Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations.* Science, 2015. **348**(6233): p. 442-4.

89. Marshall, J.M., *The effect of gene drive on containment of transgenic mosquitoes.* J Theor Biol, 2009. **258**(2): p. 250-65.

90. Adelman, Z.N. and Z. Tu, *Control of Mosquito-Borne Infectious Diseases: Sex and Gene Drive.* Trends Parasitol, 2016. **32**(3): p. 219-29.

91. Marshall, J.M. and C.E. Taylor, *Malaria control with transgenic mosquitoes.* PLoS Med, 2009. **6**(2): p. e20.

92. Fonseca, D.M., et al., *Population genetics of the mosquito Culex pipiens pallens reveals sex-linked asymmetric introgression by Culex quinquefasciatus.* Infect Genet Evol, 2009. **9**(6): p. 1197-203.

93. Dennhofer, L., *Inherited preferential segregation in translocation heterozygotes of the mosquito,Culex pipiens L.* Theor Appl Genet, 1975. **45**(6): p. 250-3.

94. Mori, A., D.W. Severson, and B.M. Christensen, *Comparative linkage maps for the mosquitoes (Culex pipiens and Aedes aegypti) based on common RFLP loci.* J Hered, 1999. **90**(1): p. 160-4.

95. Mori, A., J. Romero-Severson, and D.W. Severson, *Genetic basis for reproductive diapause is correlated with life history traits within the Culex pipiens complex.* Insect Mol Biol, 2007. **16**(5): p. 515-24.

96. 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.

97. Dennhofer, L., *Genlokalisation auf den larvalen Speicheldriisen-chromosomen der Stechmuske Culex pipiens L.* Theor. and Appl. Genetics, 1975. **45**: p. 279-289.

98. Zambetaki, A., N. Pasteur, and P. Mavragani-Tsipidou, *Cytogenetic analysis of Malpighian tubule polytene chromosomes of Culex pipiens (Diptera: Culicidae).* Genome Biol, 1998. **41**: p. 751-755.

99. 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.

100. 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.

101. Heyse, D., et al., *Unconventional organization of amplified esterase B gene in insecticide-resistant mosquitoes of the Culex pipiens complex.* J Am Mosq Control Assoc, 1996. **12**(2 Pt 1): p. 199-205.

102. Collins, F.H. and S.M. Paskewitz, *A review of the use of ribosomal DNA (rDNA) to differentiate among cryptic Anopheles species.* Insect Mol Biol, 1996. **5**(1): p. 1-9.

103. White, G.B., *Anopheles gambiae complex and disease transmission in Africa.* Trans R Soc Trop Med Hyg, 1974. **68**(4): p. 278-301.

104. Sedaghat, M.M., et al., *The Anopheles maculipennis complex (Diptera: Culicidae) in Iran: molecular characterization and recognition of a new species.* Bull Entomol Res, 2003. **93**(6): p. 527-35.

105. Nicolescu, G., et al., *Mosquitoes of the Anopheles maculipennis group (Diptera: Culicidae) in Romania, with the discovery and formal recognition of a new species based on molecular and morphological evidence.* Bull Entomol Res, 2004. **94**(6): p. 525-35.

106. Stegnii, V.N., *[Inversion polymorphism of the malarial mosquito Anopheles messeae. V. The interaction of different chromosomal inversions in the spatial area]*. Genetika, 1983. **19**(3): p. 474-82.

107. Proft, J., W.A. Maier, and H. Kampen, *Identification of six sibling species of the Anopheles maculipennis complex (Diptera: Culicidae) by a polymerase chain reaction assay.* Parasitol Res, 1999. **85**(10): p. 837-43.

108. Weitzel, T., C. Gauch, and N. Becker, *Identification of Anopheles daciae in Germany through ITS2 sequencing.* Parasitol Res, 2012. **111**(6): p. 2431-8.

109. Kronefeld, M., D. Werner, and H. Kampen, *PCR identification and distribution of Anopheles daciae (Diptera, Culicidae) in Germany.* Parasitol Res, 2014. **113**(6): p. 2079-86.

110. Danabalan, R., et al., *Occurrence and host preferences of Anopheles maculipennis group mosquitoes in England and Wales.* Med Vet Entomol, 2014. **28**(2): p. 169-78.

111. Bezzhonova, O.V. and Goryacheva, II, *Intragenomic heterogeneity of rDNA internal transcribed spacer 2 in Anopheles messeae (Diptera: Culicidae).* J Med Entomol, 2008. **45**(3): p. 337-41.

112. Novikov Iu, M. and V.M. Kabanova, *[Adaptive association of inversions in a natural population of the malaria mosquito Anopheles messeae Fall]*. Genetika, 1979. **15**(6): p. 1033-45.

113. Novikov Iu, M. and A.I. Shevchenko, *[Inversion polymorphism and the divergence of two cryptic forms of Anopheles messeae (Diptera, Culicidae) at the level of genomic DNA repeats]*. Genetika, 2001. **37**(7): p. 915-25.

114. Stegnii, V.N. and V.M. Kabanova, *[Cytoecological study of natural populations of malaria mosquitoes on the USSR territory. 1. Isolation of a new species of Anopheles in Maculipennis complex by the cytodiagnostic method]*. Med Parazitol (Mosk), 1976. **45**(2): p. 192-8.

115. Reddy, B.P., P. Labbe, and V. Corbel, *Culex genome is not just another genome for comparative genomics.* Parasit Vectors, 2012. **5**: p. 63.

116. Vinogradova, E.B., *Culex pipiens pipiens mosquitoes: taxonomy, distribution, ecology, physiology, genetics, applied importance and control. 1st edition.* Pensoft Publishers, 2000.

117. Bartholomay, L.C., et al., *Pathogenomics of Culex quinquefasciatus and meta-analysis of infection responses to diverse pathogens.* Science, 2010. **330**(6000): p. 88-90.

118. Lewin, H.A., et al., *Every genome sequence needs a good map.* Genome Res, 2009. **19**(11): p. 1925-8.

119. 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.

120. 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.

121. Timoshevskiy, V.A., et al., *Genomic composition and evolution of Aedes aegypti chromosomes revealed by the analysis of physically mapped supercontigs.* BMC Biol, 2014. **12**(1): p. 27.

122. Rai, K., *A comparative study of mosquito karyotypes.* Ann Entomol Soc Amer 1963. **56**: p. 160-170.

123. Coluzzi, M. and A. Sabatini, *Cytogenetic observations on species A and B of the Anophles gambaie complex.* Parassitologia 1967. **9**: p. 73-88.

124. Kumar, A. and K.S. Rai, *Chromosomal localization and copy number of 18S+28S ribosomal RNA genes in evolutionary diverse mosquitoes (Diptera, Culicidae).* Hereditas, 1990. **113**: p. 277-289.

125. Hickner, P.V., et al., *Enhancing genome investigations in the mosquito Culex quinquefasciatus via BAC library construction and characterization.* BMC Res Notes, 2011. **4**: p. 358.

126. Timoshevskiy, V.A., et al., *Fluorescent in situ hybridization on mitotic chromosomes of mosquitoes.* J Vis Exp, 2012(67): p. e4215.

127. Demin, S., et al., *High-resolution mapping of interstitial teleomeric repeats in Syrian hamster metaphase chromosomes.* Cytogenetic and Genome Research, 2010. **DOI: 10.1159/000321676**: p. 1-5.

128. website, N., *Available at: http://rsb.info.nih.gov/ij/. Accessed 2011 March 10.*

129. website, Z., *Available at: http://www.zeiss.de* Accessed 2010 December 5.

130. Jimenez, L.V., et al., *Characterization of an Aedes aegypti bacterial artificial chromosome (BAC) library and chromosomal assignment of BAC clones for physical mapping quantitative trait loci that influence Plasmodium susceptibility.* Insect Mol Biol, 2004. **13**(1): p. 37-44.

131. Levan, A., K. Fredga, and A.A. Sandberg, *Nomenclature for centromeric positions on chromosomes.* Hereditas, 1964. **52**: p. 201-220.

132. Shaffer, L.G., M.L. Slovak, and L.J. Campbell, *An international system for human cytogenetic nomenclature*. 2009, Basel (Switzerland): Karger.

133. Brown, S.E. and D.L. Knudson, *FISH landmarks for Aedes aegypti chromosomes.* Insect Mol Biol, 1997. **6**(2): p. 197-202.

134. McDonald, P.T. and K.S. Rai, *Correlation of linkage groups with chromosomes in the mosquito, Aedes aegypti.* Genetics, 1970. **66**(3): p. 475-85.

135. 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.

136. Sharakhova, M.V., et al., *Update of the Anopheles gambiae PEST genome assembly.* Genome Biol, 2007. **8**(1): p. R5.

137. 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.

138. Sharakhov, I.V., et al., *Inversions and gene order shuffling in Anopheles gambiae and A. funestus.* Science, 2002. **298**(5591): p. 182-5.

139. Campos, J., C.F. Andrade, and S.M. Recco-Pimentel, *A technique for preparing polytene chromosomes from Aedes aegypti (Diptera, Culicinae).* Mem Inst Oswaldo Cruz, 2003. **98**(3): p. 387-90.

140. 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.

141. Bachtrog, D., *Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration.* Nat Rev Genet, 2013. **14**(2): p. 113-24.

142. Brown, H.E., et al., *Projection of Climate Change Influences on U.S. West Nile Virus Vectors.* Earth Interact, 2015. **19**.

143. Reisen, W.K., *Ecology of West Nile virus in North America.* Viruses, 2013. **5**(9): p. 2079-105.

144. Hotez, P., *Neglected diseases amid wealth in the United States and Europe.* Health Aff (Millwood), 2009. **28**(6): p. 1720-5.

145. Li, T., et al., *A G-protein-coupled receptor regulation pathway in cytochrome P450-mediated permethrin-resistance in mosquitoes, Culex quinquefasciatus.* Sci Rep, 2015. **5**: p. 17772.

146. Wang, W., et al., *Identification of proteins associated with pyrethroid resistance by iTRAQ-based quantitative proteomic analysis in Culex pipiens pallens.* Parasit Vectors, 2015. **8**: p. 95.

147. Waterhouse, R.M., et al., *Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes.* Science, 2007. **316**(5832): p. 1738-43.

148. Kitzmiller, J.B., *Inbred strains of Culex mosquitoes.* Science, 1952. **116**(3003): p. 66-7.

149. Dennhofer, L., *[Correlation of linkage-groups with chromosomes in the mosquito, Culex pipiens L]*. Chromosoma, 1972. **37**(1): p. 43-52.

150. 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.

151. Naumenko, A.N., et al., *Mitotic-chromosome-based physical mapping of the Culex quinquefasciatus genome.* PLoS One, 2015. **10**(3): p. e0115737.

152. 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.

153. Venkatesan, M., et al., *An initial linkage map of the West Nile Virus vector Culex tarsalis.* Insect Mol Biol, 2009. **18**(4): p. 453-63.

154. Pleshkova, G.N., et al., *[Inversion polymorphism in the malarial mosquito, Anopheles messeae. III. The temporal dynamics of the inversion frequencies in the population of the center of a geographic range]*. Genetika, 1978. **14**(12): p. 2169-76.

155. Gordeev, M.I. and V.N. Stegnii, *[Inversion polymorphism in the malaria mosquito Anopheles messeae. VII. Fertility and the population genetics structure of the species]*. Genetika, 1987. **23**(12): p. 2169-74.

156. Zoller, T., et al., *Malaria transmission in non-endemic areas: case report, review of the literature and implications for public health management.* Malar J, 2009. **8**: p. 71.

157. Novikov, Y.M. and O.V. Vaulin, *Expansion of Anopheles maculipennis s.s. (Diptera: Culicidae) to northeastern Europe and northwestern Asia: causes and consequences.* Parasit Vectors, 2014. **7**: p. 389.

158. Githeko, A.K., et al., *Climate change and vector-borne diseases: a regional analysis.* Bull World Health Organ, 2000. **78**(9): p. 1136-47.

159. Gordeev, M.I., et al., *[Description of the new species Anopheles artemievi sp.n. (Diptera, Culicidae)]*. Med Parazitol (Mosk), 2005(2): p. 4-5.

160. Gordeev, M.I. and A.K. Sibataev, *[Cytogenetic and phenotypic variation in central and peripheral populations of the malaria mosquito, Anopheles messeae Fall. (Diptera, Culicidae)]*. Genetika, 1996. **32**(9): p. 1199-205.

161. Gordeev, M.I., O.V. Bezzhonova, and A.V. Moskaev, *[Chromosomal polymorphism in the populations of malaria mosquito Anopheles messeae (Diptera, Culicidae) at the south of Russian Plain]*. Genetika, 2012. **48**(9): p. 1124-8.

162. Gordeev, M.I. and N. Troshkov, *[Inversion polymorphism of the malaria mosquito Anopheles messeae. IX. Cannibalism in larvae as a selection factor]*. Genetika, 1990. **26**(9): p. 1597-603.

163. Burlak, V.A. and M.I. Gordeev, *[The effect of infection by the entomopathogenic bacterium Bacillus thuringiensis on the spread of microsporidia in an inversion-polymorphic population of the malarial mosquito Anopheles messeae (Diptera: Culicidae)]*. Parazitologiia, 1998. **32**(3): p. 264-7.

164. Gordeev, M.I. and V.P. Perevozkin, *[Strategies for selection and stability to asphyxia in larvae of the malaria mosquito Anopheles messeae with various karyotypes]*. Genetika, 1995. **31**(2): p. 180-4.

165. Gordeev, M.I. and V.A. Burlak, *[Inversion polymorphism in malaria mosquito Anopheles messeae. Part X. Resistance of larvae with different genotypes to toxins of crystal-*

forming bacteria Bacillus thuringiensis subsp. israelensis (serovar H14)]. Genetika, 1991. **27**(2): p. 238-46.

166. Gordeev, M.I. and V.A. Burlak, *[Inversion polymorphism in the malaria mosquito Anopheles messeae. XI. The group effect of larval infection with Bacillus thuringiensis subsp. israeliensis bacteria]*. Genetika, 1992. **28**(7): p. 82-8.

167. Sharakhov, I.V. and M.V. Sharakhova, *Chromosome evolution in malaria mosquitoes.* Genetika, 2010. **46**(9): p. 1250-3.

168. Kirkpatrick, M. and B. Barrett, *Chromosome inversions, adaptive cassettes and the evolution of species' ranges.* Mol Ecol, 2015. **24**(9): p. 2046-55.

169. Sharakhov, I.V., *Chromosome phylogenies of malaria mosquitoes.* Tsitologiia, 2013. **55**(4): p. 238-40.

170. Gentile, G., et al., *Attempts to molecularly distinguish cryptic taxa in Anopheles gambiae s.s.* Insect Mol Biol, 2001. **10**(1): p. 25-32.

171. Balloux, F. and J. Goudet, *Statistical properties of population differentiation estimators under stepwise mutation in a finite island model.* Mol Ecol, 2002. **11**(4): p. 771-83.

172. Balloux, F. and N. Lugon-Moulin, *The estimation of population differentiation with microsatellite markers.* Mol Ecol, 2002. **11**(2): p. 155-65.

173. Matoke-Muhia, D., et al., *Decline in frequency of the 2La chromosomal inversion in Anopheles gambiae (s.s.) in Western Kenya: correlation with increase in ownership of insecticide-treated bed nets.* Parasit Vectors, 2016. **9**(1): p. 334.