Comparative Analysis of Heterochromatin in the *Anopheles gambiae* Complex

Atashi Sharma

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

Igor V. Sharakhov, Chair
Maria V. Sharakhova, Co-chair
Zhijian (Jake) Tu
Troy D. Anderson
Pawel Michalak

March 25, 2016
Blacksburg, Virginia

Keywords: *Anopheles gambiae*, heterochromatin, speciation, epigenetics, Y chromosome

Comparative Analysis of Heterochromatin in the *Anopheles gambiae* Complex

Atashi Sharma

Abstract

Mosquito borne diseases continue to be a big threat to human health worldwide. Despite using various vector control methods, we lose a great number of lives to this malicious disease in tropical and subtropical countries each year. Not surprisingly, mosquito is considered as the deadliest animal on the earth, because mortality rates from mosquito-vectored infections only lag behind other major diseases such as HIV and tuberculosis. Current approaches of vector control are mostly limited to using insecticidal bed nets, thus novel techniques are required to prevent a staggering loss to human health and quality of life. Advances in the genome sequencing in the past decade have helped to uncover numerous secrets of diverse genomes. The genome of malaria mosquito *Anopheles gambiae* was first sequenced in 2002 and since then has been updated to include additional scaffolds, their orientations and correction of mis-assemblies. Yet, the greatest challenge remains in assembling the heterochromatin regions, that are repeat rich part and contain relatively low-gene density. Although previously neglected by scientific studies due to its characteristic paucity of genes, heterochromatin is now recognized to be crucial for several processes such as cell viability, chromosome pairing, meiosis, longevity etc. It is therefore not surprising that heterochromatin comprises of a significant portion of the genome in many species. The efforts to analyze the genome of malaria mosquito in order to identify potential new leads for vector control warrant a better understanding of the heterochromatin.

Mosquitoes diploid chromosome number equal 6. While autosomes 2 and 3 are submetacentric and present in both sexes, females are homogametic with XX and males are heterogametic with XY sex chromosomes. To achieve a better understanding of the *Anopheles* heterochromatin, we

investigated heterochromatic region of the X chromosome. Despite one arm of the X chromosome being completely heterochromatic, few studies have investigated the molecular content of this region. Protocols were developed for performing fluorescent *in situ* hybridization (FISH) on mitotic X chromosomes in *An. gambiae*. Using cytogenetics and molecular biology techniques, we characterized the X chromosome heterochromatin in members of the *An. gambiae* complex. Specific satellite DNA and 18S ribosomal DNA probes (major components of heterochromatin) were mapped to X chromosomes enabling their differentiation and characterization in the *An. gambiae* complex. Microarray studies have highlighted the importance of X chromosome during investigation of nascent species *An. gambiae* and *An. coluzzii*. Here for the first time qualitative differences in heterochromatin in between nascent species are described. Cytogenetic idiograms are developed as to include the molecular and qualitative differences between the species of the *An. gambiae* complex. These idiograms are expected to provide a better resolution of the X chromosome heterochromatin for comparison in major malaria vectors, closing some of the gaps present due to poor sequencing of unassembled repeat rich regions in *An. gambiae* complex.

The current understanding of Y chromosome for transgenic manipulation is poor and limited to very few genes. Due to its near total heterochromatic composition, it is the hardest part of the genome to assemble. In collaboration with other researchers, the Y chromosome content was characterized among sibling species of the *An. gambiae* complex. Our data revealed the swift changes the Y chromosome has undergone in a relatively short evolutionary time period. These include a rapid rate of turnover not only in heterochromatin but also in euchromatin. In addition to previously described repeats, a novel highly repetitive element called Zanzibar was discovered and mapped to the males of various *Anopheles* sibling species. Our data can form the basis for

evolutionary studies in heterochromatin for male mosquitoes within the *An. gambiae* complex while also help identify novel targets to create successful transgenic male populations. Along with the X chromosome heterochromatin, to our knowledge this is the most extensive contribution to improve the understanding of mitotic chromosome heterochromatin in malaria mosquitoes.

This study also investigated if epigenetics play role in mosquito development, fecundity and heterochromatin formation. DNA methylation, histone modifications and small noncoding RNAs are among the epigenetic mechanisms scrutinized in mammals. However, knowledge about epigenetic mechanisms and their effects is sparse in mosquitoes. A protocol for testing the various effects of epigenetics on different stages of malaria mosquito was developed. An epigenetic drug was utilized to probe the effects on immature and adult malaria mosquitoes. Different concentrations of DZNep, a histone methyltransferase inhibitor, were administered to *An. coluzzii* larvae. Total survivorship and pupation were compared for treated and untreated groups. The drug was also administered to adult blood feeding females to determine any effects on fecundity and egg morphology, revealing a negative association with an increase in drug concentration. A dose dependent decrease in SAH hydrolase concentration in *An. coluzzii* was also noticed. These results suggest epigenetics plays a critical role in mosquito pupation and ovarian development. Our work lays the groundwork for future investigations into the field of epigenetics in mosquitoes by revealing its effect on several important developmental stages in malaria mosquitoes.

Although genomics and next-gen sequencing technology have come a long way in the last decade since the first *Anopheles* genome was sequenced, considerable gaps still exist in case characterization of heterochromatin function in an organism. Through our work, we have

endeavored to elucidate a few of the major roles that heterochromatin may play in organization, evolution and adaptation of the malaria mosquitoes.

Comparative Analysis of Heterochromatin in the *Anopheles gambiae* Complex

Atashi Sharma

General Audience Abstract

Mosquito borne diseases result in a large number of deaths across the world, particularly in developing countires. Currently, almost 700 million people are at risk to one of the diseases transmitted by mosquitoes. Among these, malaria is a chief concern, with over 214 million documented cases resulting in about 500,000 deaths in 2015 alone. A majority of these occur in Sub- Saharan Africa, with thehigest casualities in the group of children less than 5 years of age. Thus it produces a huge economic and social burden on our society. Despite using various methods like insecticide treated bednets and other traps, we lose a great number of lives to this malicious disease in tropical and subtropical countries each year. Not surprisingly, mosquito is considered as the deadliest animal on the earth, because mortality rates from mosquito-vectored infections only lag behind other major diseases such as HIV and tuberculosis. In order to arm ourselves against this enemy, we need a better understanding of what makes the malaria mosquito of genus *Anopheles*, a deadly insect vector. Interestingly, only the female mosquitoes transmit malaria as they need the protein from a blood meal to lay eggs. So determining how male and female mosquitoes differ can offer us some novel solutions for this problem. Mosquitoes are diploid with three pairs of chromosomes, one of which is the sex- determining chromosome – XX in females and XY in males. My focus is to determine how the sex chromosomes evolve in closely related species of *Anopheles* mosquitoes in both males and females, to obtain a better understanding of the malaria problem. This was achieved by placing parts of the *An. gambiae* genome that are highly repetitive called heterochromatin, to the X and Y chromosomes of sister species in the *An. gambiae* complex, which comprises of eight closely

related species. Using a combination of bioinformatics and fluorescence microscopy, we mapped repeat rich segments called satellite DNA to several Anopheline species and constructed a map for species comparison as well. Further, we developed a protocol to test the changes caused by facultative heterochromatin (repetitive DNA which may be transcribed under some circumstances) on mosquito development, survivorship and fecundity. This was done using an anti- cancer drug called DZNep. In mammals, this drug has shown effects on cell lines by causing changes to heterochromatin. We used DZNep to determine if changes in facultative heterochromatin level also impact important phenotypes of malaria mosquitoes, particularly the survivorship and fecundity as these are the traits targeted for vector control. Overall, our results show how sex chromosomes have undergone rapid evolution in the last 2 million years between closely related mosquito species in Africa, and these changes occur at the molecular level in the heterochromatin. We also show that manipulation of heterochromatin could lead to novel targets for mosquito control using already established drugs. Our resultswill provide a platform for identifying and improving existing vector control approaches.

Dedication

I dedicate this dissertation to my parents, Mrs. Mridula Sharma and Mr. Kalipada Sharma. It is not easy to see a child struggle through times when she is homesick and lonely so far away from home, and you have always encouraged me to work hard and be strong. I would also like to dedicate this work to all the other doctoral students out there, current, past and future. Graduate school and research can be extremely trying, and I hope that you all have the kind of friends' support and guidance that I did.

Acknowledgements

I would like to thank my adviser, Dr. Igor Sharakhov, for his help and guidance throughout these past five years. When I first started my rotation in his lab, I didn't know much about insect vectors and designing the right experiments, he has guided me and taught me to ask the right questions, making me a better scientist. Thanks for all the help and patience that you embodied during my time as a student in your lab, I hope to continue learning from you in future.

I also thank my co- chair, Dr. Maria Sharakhova for her help as an established cytogeneticist on one of my major projects. During the past few years I have learned a lot from her, and hope to continue our collaboration in future as well.

A huge part of being a graduate student is to learn and share along with other members of the lab, and it was my good fortune to have a team of learned members in the lab. I would like to thank Dr. Phillip George for his guidance and support in my professional as well as personal life. As a senior member of the lab, I have often picked your brains and my respect for your scientific rationale has increased with every passing day. You are also a great friend, thank you for helping with my final presentations and most importantly, getting me to start writing my dissertation well in advance. I wish you all the luck with your future endevors in the army, Captain Dr. George!

Ashley Peery, thanks for being my PhD buddy all these years. You made sure I wasn't feeling left out when I first joined the lab, and as we progressed through our research together, you were the single most trusted source for me to recount my scientific results, frustrations as well as personal dilemmas. Thanks a ton for being there, I am going to miss our fights for the lab equipment and hoping to be in touch with you always!

Nick Kinney, thanks for being a good friend and engaging in several debates over God vs science in the past few years! I enjoyed our time together in lab, it is heartwarming to see the happiness that you get from teaching young inquisitive minds.

Dr. Maryam Kamali, you are one of the kindest souls I have ever come across, and I really appreciate your calm demeanor and prayers for me in tough times. I wish you the every best of luck in life, and thank you for being there for me when I needed a big sister.

Jiyoung Lee and Jiangtao Liang, though you have not been in the lab too long, you are both great co workers and friends to be around in lab. I really enjoy our conversations about both science as well as other topics, and as international students we all have common adjustments which you both provide a perspective on.

Kerry Meguschak, you are a very hard working and talented student. In my short time of mentoring you, I see great potential in you as a scientist or a doctor, and I hope that you will be able to apply the team working skills you have acquired over the past year in our lab.

Perhaps the biggest thanks go to my roommates and friends –Meeta, Tina, Mitu, Poulomi, Vishwas, Manisha and Hazitha. PhD comes with it's own set of challenges, and being away from family for long periods of time makes it that much harder. I don't think I would have been able to successfully navigate through the last few years without all of your support, help and encouragement along with occasional doses of free yummy food! In particular, a huge shout out to Meeta aka mommy for all your words of wisdom and nautanki as my roommate, Tina aka hot ladki for your inspiration to get fitter (even though it didn't quite work on me) and Poulomi as my beloved roomie who stoppd me from eating icecream for dinner and made sure I ate healthy to complete my dissertation in time. You are all great friends, and I am lucky to have you in my graduate life and beyond.

Finally, I am thankful to my parents, particularly my mother. As a young woman who had to give up her career to raise me and my brother, she has always imparted in me the importance of being self- sufficient. With this encouragement and her prayers, I believe she was instrumental in me completing my degree successfully. Thank you, mom and dad!

Attribution

Several of my colleagues aided in the writing and research behind the chapters I present here. A brief description of the authors' contribution in included below:

Chapter 2 - Fluorescent in-situ hybridization on mitotic chromosomes of mosquitoes.

Chapter 2 was published in Journal of Visual Experiments (JoVE).

Dr. Vladimir Timoshevskiy, former post-doctoral researcher in our lab and currently a post-doctoral associate at University of Kentucky, performed experiments contributed to writing the manuscript on *Aedes aegypti* imaginal discs.

Dr. Igor V. Sharakhov, currently an associate professor in the Department of Entomology, Virginia Tech, helped with experiment designs and proofread the manuscript.

Dr. Maria V. Sharakhova, currently an assistant professor at the Department of Entomology, Virginia Tech, was the principal investigator (PI) on this project. She designed the experiments, analyzed results and wrote the manuscript.

Chapter 3 – Molecular restructuring of heterochromatin in evolution of the *Anophles gambiae* complex.

Nicholas Kinney is a fellow graduate student at our lab and wrote the MATLAB script which helped map the positions of stDNA in the project.

Dr. Igor V. Sharakhov conceived the project and designed the experiments. He was the PI on this project.

Dr. Maria V. Sharakhova helped analyze the results, construct idiograms, and in editing the manuscript.

Chapter 4 – Radical remodeling of the Y chromosome in the *Anopheles gambiae* complex

Chapter 4 was accepted in Proceedings of the National Academy of Sciences (PNAS).

Andrew Brantley Hall (G.B.C.B program, Virginia Tech), Dr. Philippos Pappathanos (Imperial College) and Dr. Changde Cheng (University of Texas, Austin) were co-first authors on this project. They wrote the scripts required for bioinformatics experiments and analyzed the data.

Omar S. Akbari, Lauren Assour, Nicholas H. Bergman, Alessia Cagnetti, Andrea Crisanti, Tania Dottorini, Elisa Fiorentini, Roberto Galizi, Jonathan Hnath, Xiaofang Jiang, Sergey Koren, Tony Nolan, Diana Radune, Maria V. Sharakhova, Aaron Steele, Vladimir A. Timoshevskiy, Nikolai Windbichler, Simo V. Zhang, Matthew W. Hahn, Adam M. Phillippy, Scott J. Emrich, Igor V. Sharakhov, Zhijian Tu, Nora J. Besansky were also co – authors on this paper. They helped with the experimental design and analysis for the data.

Chapter 5 Toxicological assays for testing effects of an epigenetic drug on development, fecundity and survivorship of malaria mosquitoes.

Chapter 5 was published in Journal of Visual Experiments.

Dr. Troy D. Anderson is an assistant professor in the Department of Entomology, Virginia Tech. He helped with the experimental design and analysis for toxicological assays. He also contributed to writing the manuscript.

Dr. Igor V. Sharakhov was the PI on this project. He designed the experiments and contributed to writing the manuscript.

TABLE OF CONTENTS

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 1 Literature Review

1.1 Overview – The malaria burden

Insect borne diseases pose a significant concern in overall human and cattle health worldwide. The association of insects with diseases is not recent. In fact, some may argue that the relationship between mammals, insects and parasites is as old as civilization itself. As human settlements and man- made environmental changes occur, we see changes in the paradigm of these relationships as well. Mosquito borne diseases such as dengue, malaria, Chikungunya and Zika collectively cost millions of lives and are capable of debilitating a country's economy if not controlled in time. The chief among them, malaria is a devastating disease resulting in almost 200 million cases with over half a million deaths estimated across the world every year, most of them children under the age of five. This disease causes a huge financial burden estimating to over billions of dollars in healthcare and rehabilitation, in Africa alone the total costs of this disease amounts to an estimated $12 billion. The WHO malaria report 2014 stated the number of people at risk to be at a colossal 3.3 billion [1]. Over the past few decades, interventions from various groups have tried and succeeded in bringing down the toll from this disease significantly by 31.5% [2]. Efforts ranging from indoor residual spraying (IRS), long lasting insecticidal bed nets (LLIN), developing novel insecticides to periodic release of sterile or transgenic male mosquitoes have greatly reduced the number of people suffering from such diseases. However, a considerable gap still exists in our understanding of the extent and reasons behind why the mosquitoes are thriving as vectors. If mosquito-borne diseases were to be controlled, we must extend the boundaries of current knowledge to include information missing at present.

Malaria is an acute febrile infection, caused by the parasite of genus Plasmodium. Five Plasmodium species are involved in the infection – *P. falciparum, P. vivax, P. malariae, P. ovale*

and *P. knowlesi*. *P. falciparum* is the primary malarial agent in Sub Saharan Africa, while *P. vivax* has a wider distribution in other regions. Though *P. knowlesi* is not a direct agent, it is emerging as a leading risk factor for acquiring zoonotic human malaria through the primary hosts long tailed macaques in South-East Asia [3]. The classical symptoms include high fever followed by chills and profuse sweating leading to extreme weakness. The susceptibility of a particular strain of plasmodium varies with mosquito genus and various factors influence host competence.

1.2 Current limitations in controlling malaria

Despite commendable advances in genomics and drug development, we lack a preventative vaccine against malaria at present. Several candidate vaccines such as RTS,S (Mosquirix) are in clinical trial phases. Additional targets for vaccine development include antigens based on the repeat region of *P. falciparum'*s circumsporozoite protein (CSP), library of recombinant *P. vivax* ectodomain proteins homologous to *P. falciparum* vaccine candidates and recently, isolation of human monoclonal antibodies for broadly reactive antibodies against epitopes of *P. falciparum* variant antigens. [4] [5, 6] A combination of CSP and viral-vectored TRAP based on *P. berghei* was also shown to increase the protection against malaria significantly [7]. However, most of the leads pose the problem of intermediate to low levels of protection, largely due to the polymorphisms in antigen diversity across endemic regions as well as diversity in Plasmodium populations in endemic regions [8]. For example, Phase 1 and Phase 2 clinical trials of RTS,S provide modest protection - 34% in infants and <50% in young kids under 2 years of age which are the biggest group at risk of malaria [9] [10] Additional major hindrances in vaccine development include target protein expression and purification, inefficacy of target vaccines to produce antibodies and the short number of bio-assays for validating a target vaccine [11]. One major limitation with vaccines based on *P. falciparum* is their failure to produce a response with

other *Plasmodium* agents, rendering the vaccination less effective in areas where multiple Plasmodium species exist. Even if a successful vaccine is developed, challenges with implementing a vaccine remain large. The diversity in parents' view and need necessitates a wider study of factors affecting vaccination, chief among them the risk, cost and neighbors views [12]. Vaccines interrupting malaria transmission (VIMT) or transmission blocking vaccines (TBV) are another area of intervention where immunity against the infectious or sexual stage of parasite results in lowering the human to mosquito transmission, leading to less morbidity within a community [13]. Several candidate TBV leads are currently under research or undergoing early clinical trials, chief among them *P. falciparum* PfGCS1, Pfs25, *P. vivax* Pvs48/45 and P230 [14] [15] [16]. A recent study developed a mathematical model of the potential effects of vaccination and bednet usage, and found a synergistic relationship between TBV and bednets that would lower the morbidity [17]. However, the genetic diversity of these candidates in the endemic region is still under evaluation [18] [19]. Additionally, most of these candidate vaccines may take years before they are launched and used by the target public. Hence, the most feasible approach currently remains vector control by indoor residual spraying (IRS) and use of long lasting insecticidal nets (LLINs).

As different mosquito species exhibit differences in resting and biting behavior, targeting a specific species is often not the solution. Health workers found this out the hard way, when implementation of the Garki malaria project failed in Nigeria [20]. The major aim of this study was to determine the effect of malaria intervention programs involving a combined usage of insecticidal spray with Propoxur along with mass drug administration of two drugs- sulfalene and pyrimethamine along with studying the epidemiology of malaria in Garki region of North Nigeria. Major vectors *An. gambiae* and *An. arabiensis* were sympatric in this region. The results

of the survey revealed that despite spending a considerable amount of money on intervention programs, the level of malaria occurrence had not changed. This was because indoor spraying only killed endophagic or endophilic (mosquitoes which take a blood meal inside houses or rest indoor) whereas proper measures were not undertaken to control the exophilic populations. One of the major concerns in fight against malaria includes increasing reports of parasitic resistance to artimesin and other anti-malarial drugs [1]. Meanwhile, several populations are reported to be increasingly resistant to drugs currently in use including carbamate, DDT, bendiocarb and other pyrethroids [21] [22] [23] [24] Moreover, the long term use of pesticides has resulted in changed vector behavior leading to a change in biting times and insecticide resistance in many areas [25]. Taken together, these factors could pose a grave situation for the future and lead to a reemergence of malaria in regions under control at present.

1.3 Alternate approaches of vector control

With the recent advances in the field of translational biology, biological control and release of genetically modified male mosquitoes to crash a local population are increasingly implemented in many parts of the world. Sterile insect technique (SIT), Release of insects with dominant lethal (RIDL) and *Wolbachia* infected genetically modified male mosquitoes are among the most popular current approaches within the realm of transgenic arsenal to fight mosquito-borne diseases. SIT involves irradiating male mosquitoes to produce sterile males with several lethal genes and subsequent mass release. Irradiated sterile male mosquitoes are released to compete with wild type males in a population. When these males mate with wild type females, they pass on lethal genes to progeny resulting in unviable eggs, affecting the female fecundity. Over time, this phenomenon potentially leads to targeted mosquito population crash [26] [27]. An extension of this is the RIDL technique, where sterile male mosquitoes carry a cell-lethal gene. Upon

mating, wild-type female mosquitoes produce offsprings in which either no females are produced or they do not progress beyond the aquatic stage. In contrast, the male progeny survives and helps spread the transgene in the local population. RIDL has been successfully implemented in field trials for controlling the dengue mosquito *Aedes aegypti* [28] [29]. Efforts are ongoing to replicate the success in malaria mosquitoes [30, 31]. Another biological control approach involves infecting male mosquitoes with endosymbiotic bacteria *Wolbachia* sp. The *Wolbachia* strain (wMelPop) was first shown to have negative effects on the egg viability and mortality in dengue mosquito *Ae. aegypti* [32]. Male *Aedes* mosquitoes infected with this strain could successfully mate and shorten the life span of wild type females. As only older females carry the bacteria, shortening of life span lowers the number of infectious females. With desirable features including cytoplasmic incompatibility and a high rate of maternal inheritance, this method could pave way for the new wave of biological vector control. Since then, several cage and semi-field trials have shown that this is indeed a promising approach to biological vector control [33] [34] [35, 36] [37]. Attempts to establish a *Wolbachia* colony in malaria mosquito strains were not successful until recently [38], when the first such reports were published in a lab with Asian malaria vector *An. stephensi* [39]. If extended to *An. gambiae* complex successfully, this could provide a novel way to crash local mosquito populations and significantly decrease the risk of contracting malaria.

When released in field studies, these mosquitoes are expected to outcompete wildtype males to mate with females. Most of these approaches depend upon identifying a male specific factor. As male mosquitoes do not transmit malaria, they are an excellent source of transgenic or environment friendly insect control techniques, as they leave no chemical or toxic residues. However, the success of most of these techniques depends on developing a fool–proof method so

that only male mosquitoes are released in the environment. Currently, knowledge about the Y chromosome in mosquitoes is limited to few sequences. A more comprehensive understanding of male specific sequences and genes in the *Anopheles* mosquitoes, particularly in *An. gambiae* complex, could potentially help enhance the effects of SIT and RIDL techniques and make them more broad-spectrum with respect to areas harboring multiple malaria mosquito species.

1.4 Diversity within malaria vector mosquitoes

As one of the most successful group of insects to inhabit various niches of the world, it is important to analyze how the specific species causing these diseases diverge, adapt and successfully populate such wide variety of climates. Mosquitoes are found in various distinct habitats e.g. *An. gambiae* and *An. arabiensis* dwell in freshwater habitats while *An. merus* and *An. quadriannulatus* prefer marshy saltwater regions. Rapid adaptation to changes in weather, changes created by human activities to habitats along with different geographical barriers have led to speciation of mosquitoes into species that populate a particular region e.g. *An. funestus* (Africa), *An. stephensi* (Asia) and *An. quadrimaculatus* (Europe) are all major vectors of malaria in different parts of the world. Global warming and frequent travelling is also suggested to increase the possibility of malaria in areas which did not have it earlier, eg. rapid rise of *An. maculipennis* complex members in Europe [40]. Further, new species have recently been discovered, for example *An. orieosis* as part of *A. farauti* complex [41], *An. punctulatus* in Papa New Guinea [42]. At the same time, established mosquito species are also undergoing further speciation (notably in Africa where currently *An. gambiae* is undergoing diversification into *An. gambiae* Giles and *An. coluzzii* [43] [44]).

Keeping track of all these species can be daunting, and yet all the various species and sub-groups of mosquitoes seem to be adapting successfully to their corresponding environment rapidly. In order to fight against the mosquito-borne diseases, it is important that novel ways to debilitate them are discovered in addition to the established methods of vector control. One way to do this is to investigate the source of their adaptations and the mechanisms leading to their successful evolution. Mutations in the genome involving chromosomal inversions, gene deletions, duplications, SNPs and epigenetic effects would help identify the factors behind the rapid success story of mosquitoes worldwide and enable us to exploit these for potential vector control measures. The combined advents of next gen sequencing, Illumina, PacBio sequencing, ChIP seq, RNA-seq have helped identify several of these mutations, but characterization, annotation and assembly of heterochromatin is yet to be performed satisfactorily.

1.5 The *Anopheles gambiae* complex

The *An. gambiae* complex is responsible for the maximum number of malaria cases in Sub-Saharan Africa. Currently it comprises of eight species which differ in genome size, host choice preferences, biting behavior, distribution and contribution to human malaria in Africa. Initially termed as a cryptic species *An. gambiae* with several sub populations, crossing experiments along with comparisons of inversion polymorphism in polytene chromosomes revealed this to be a group of morphologically indistinguishable mosquitoes known as the *An. gambiae* complex which includes both freshwater as well as marshy land inhabitants [45] [46]. Major vectors *An. gambiae* and *An. arabiensis* inhabit fresh water areas and are found in sympatry in some regions. While *An. gambiae* is endophillic and primarily bites indoors, *An. arabiensis* is exophillic and bites outdoors in regions it is anthropophillic [25]. In contrast, several other studies report *An. arabiensis* to be primarily zoophilic but resting indoors in the regions surveyed [47]. Minor

vectors *An. melas*, *An. merus* and *An. bwambae* dwell in brackish salt water while non human vector *An. quadriannulatus* is zoophilic and thus not considered a malaria vector. In addition, two members were recently renamed – *An. amharicus* (formerly *An. quadriannulatus* B) and *An. coluzzii* (formerly *An. gambiae* M form) thus increasing the number of species in the complex to eight [44]. The species' distribution differs along with host-preference and two or more species are often found in sympatry. As all of the species are morphologically indistinguishable, necessitating a quick diagnostic molecular assay to distinguish between the species vectoring malaria present in any region.

1.6 Cytogenetics of malaria mosquitoes –mitotic and polytene chromosomes

Polytene chromosomes are large well-banded chromosomes produced in many Dipteran species as a result of multiple DNA replications without undergoing cell division. As a result, they are much bigger than the mitotic chromosomes. The light and dark bands produced as a result of chromatin compaction on the polytene chromosomes provide an easily recognizable pattern of land marks that help in identifying each chromosomal arm. The original *An. gambiae* genome assembly used a map based on polytene chromosomes to identify and orient important genes and scaffolds [48]. In *Anopheles* mosquitoes, polytene chromosomes are usually obtained from salivary glands, ovaries and Malphigian tubules with five distinct arms X, 2R, 2L, 3R and 3L. In contrast, mitotic chromosomes result from DNA duplication followed by an equatorial cell division. They are much smaller than polytene chromosomes. Owing to their size, they have been largely ignored in mapping and assembly of the *An. gambiae* complex.

1.7 *An. gambiae* PEST genome

The first mosquito to be sequenced was *An. gambiae* PEST strain in 2002 [48]. This strain was selected based on the standard chromosomal arrangement and lack of inversion polymorphism on chromosome 2. The colony also had an X linked pink eye mutation, helpful for identifying any contamination during breeding. Furthermore, clones from the BAC (Bacterial Artificial Chromosome) libraries had previously been end-sequenced and mapped to polytene chromosomes, thus facilitating a cytogenetic map which could be build further with the sequence information. 91% of the 278 Mbp genome was mapped to 304 large scaffolds, with additional smaller scaffolds accounting for the remaining sequences. However, the sequencing revealed the presence of two haplotypes with equal frequency and 400,000 SNP variation, suggesting that the strain was not completely inbred and was a mixture of two molecular forms, M and S, which were identified by the fixed differences in their 18S internal transcribed region (ITS) [43]. SNP variation was determined to be lowest on the X chromosome in this assembly, possibly due to lower introgression between the forms or hemizygous nature of male mosquitoes. The *An. gambiae* PEST genome was further improved notably by the inclusion of many additional scaffolds [49] and a detailed chromosomal map [50]. However, a significant portion of the genome remains unmapped till date owing to the repeat rich nature of these sequences. Among these, of particular interest are the short arm of X chromosome and entire Y chromosome. Although some Y linked sequences were described [51] [52], they have not been mapped on the chromosome map based on polytene chromosomes. Recently, *An. gambiae* M and S forms were elevated to species levels and are now known as *An. gambiae* s.s. and *An. coluzzii* [44]. These two nascent species were independently sequenced and now additional information is available regarding each in through *An. gambiae* Pimperena and *An. coluzzii* Mali on VectorBase.

1.8 Heterochromatin

In any living organism, the genetic material is present in a combination of DNA and histone proteins known as chromatin. The basic building block of chromatin is nucleosome, which is a complex package of DNA wrapped around 4 histone proteins [53]. In eukaryotes, genome comprises of two kinds of chromatin. The gene rich, lightly packed euchromatin is easily accessible to proteins for transcription and is often actively expressed. In contrast, the gene poor, late replicating heterochromatin is tightly packed, rendering it inaccessible to most proteins that are responsible for transcription. As a result, it is seldom transcribed. Cytological dyes such as DAPI, Hoechst or C-banding are often useful to distinguish the two chromatin states. While euchromatin stains lighter and less intense, heterochromatin stains more intensely in comparison owing to the tighter packing. Due to its repeat-rich, gene poor nature heterochromatin was often termed as the 'junk DNA'. However, latter studies show that far from being useless 'junk', it constitutes an important part of the genome and plays role in genome regulation, DNA replication, spindle formation, meiotic chromosome segregation and nuclear architecture [54] [55]. Heterochromatin can be further subdivided into constitutive and facultative forms. Constitutive heterochromatin is tightly packed, repetitive and usually present in centromeres or telomeres (regions of low recombination). As a result of its highly compact nature, genes present in this region are seldom expressed. Facultative heterochromatin does not consist of many repeats, and can be transcriptionally active under certain circumstances. In *An. gambiae*, heterochromatin is enriched in genes with DNA binding and regulatory functions. Pericentric and intercalary heterochromatin constitute of tandem repeats and segmental duplications respectively [50]. Different components of heterochromatin in an organism include ribosomal DNA (rDNA), transposable elements (TE) and satellite DNA (stDNA). In addition, small

noncoding RNA, Piwi-interacting RNA (piRNA) and long non coding RNA (lncRNA) also forms a significant part of heterochromatin. While sequencing and genomics have led to a rapid increase in the knowledge of heterochromatin constituents in various organisms, the role of most of these repeats are still under investigation.

1.9 Heterochromatin and speciation

The AT-rich satellite DNA rapidly evolves between closely related populations by expanding or contracting in copy number, rendering potential source of lower recombination and gene conversion between these populations [56]. Bulk of these studies have been performed on the model insect *Drosophila melanogaster* and its sister species', with many exhibiting post- zygotic isolation where a cross between closely related species produce sterile progeny in one or the other direction. In F1 females obtained from crossing *D. melanogaster* X *D. mauritiana*, homologous chromosomes showed difference in both type and location of stDNA blocks on different chromosomes [57]. These differences may not affect the offspring viability if taking place in the super-numerary B chromosomes [58], but are responsible for a host of segregation abnormalities when present in sex chromosomes. Disruption or lowering in chromosomal pairing was shown in post-zygotic isolated species producing sterile hybrids among closely related Drosophila species. A 359 bp satellite DNA repeat region mapping to *Zygotic hybrid rescue Zhr* locus present on *D. melanogaster* was found to be the cause of mitotic defect and abnormal chromosomal segregation during early embryo stages leading to F1 hybrid female lethality between two sibling species *D. melanogaster* and *D. simulans* [59]. The maternally inherited X from *D. simulans* lacks the 359 bp repeat present in *D. melanogaster* X. Using a combination of genetic crosses and cytological experiments, the authors provided evidence for the role of heterochromatin inherited maternally in genetic background of a hybrid. During the early

embryogenesis, abnormal nuclear spacing, lagging chromatin during anaphase and atypical enrichment of protein Topoisomerase II (associated with 359 bp repeat) was seen in the F1 females produced from this cross while the F1 hybrid male embryos segregated normally. Using a simple repeat in addition to the stDNA block, the authors demonstrated that the lagging chromatin mapped to pericentric heterochromatin region of the X chromosome, indicating this region was crucial for chromatid separation. Translocation of the 359 bp block on the Y chromosome led to defective Y segregation in hybrid males. Interestingly, the reverse cross (*D. melanogaster* female X *D. simulans* males) produced viable F1 female offsprings, as *D. melanogaster* X contains the satellite DNA block implicated in this instance of hybrid incompatibility (HI), indicating reduced viability of a hybrid population. HI may also evolve as a result of adaptive interaction between heterochromatin and specific proteins. The genes Hybrid male rescue (*Hmr*) and Lethal hybrid rescue (*Lhr*) are responsible for causing hybrid lethality in F1 male hybrids between *D. melanogaster* and *D. simulans*. Both *Hmr* and *Lhr* interact with heterochromatin HP1a, but instead of the X linked 359 bp repeat they localize on Chromosome 3 and early embryos [60]. RNA-seq comparison of mutant and parental or heterozygous *Lhr* and *Hmr* lines revealed that several TE families were up-regulated in the mutant lines. A shared set for of TE's was shown to be regulated by both *Hmr* and *Lhr*, suggesting co-regulation responsible for telomere lengthening. The study also found that TEs were substantially up-regulated in the lethal hybrid male larvae compared to any of the parental male larvae, suggesting a greater role of TE and heterochromatin in hybrid lethality.

1.10 Hybrid incompatibility and heterochromatin

A common emerging theme from several of the recent studies performed in Drosophila suggests that heterochromatin plays an important role in hybrid lethality and hybrid incompatibility (HI)

[61] [59, 62] [63] [60]. This is especially evident in case of pericentric heterochromatin present in *Drosophila* X chromosome. Hybrid lethality in F1 females obtained from crosses between closely related species *D. simulans* or *D. mauritiana* and *D. melanogaster* was investigated [64]. Only F1 hybrid males produced from X mau/Y mel cross were viable, suggesting a causative dominant X-linked factor in *D. melanogaster* incompatible with the recessive factor in *D. mauritiana*. Using lines bearing small segments of the X duplicated on the *Drosophila* Y chromosome, the study found the factor *heterochromatin hybrid lethal* (*hhl*) maps to pericentromeric heterochromatin region h26-h32 in *D. melanogaster*. The strength of this factor depended on the parental *D. mauritiana* stock as well as the temperature, with increased viability of hybrids noted at lower temperature (18^{o}C). A detailed review of how satellite DNA changes may lead to speciation included examples from *Drosophila* species exhibiting postzygotic isolation. These mechanisms involve disruption of chromosome pairing due to the lack of compatible satellite DNA blocks, diverged satellite DNA binding proteins between sibling species, disruption of heterochromatin through epigenetic changes, abnormal heterochromatin packaging and selfish postmeiotic drive resulting in high frequency of transfer of selfish allele (e.g. the Segregation Disorder system SD in *Drosophila*) [57]. Heterochromatin has also been implicated in hybrid male sterility and genome evolution in *Drosophila* via abnormal segregation of circularized ring chromosomes [65]. While heterochromatin divergence may lead to irregular segregation of chromosomes in some instances, certain species of *Drosophila* show little or no impact of divergence on chromosome segregation in the hybrid background [66]. However, the possibility of this scenario being due to the involvement of chromosome 4 cannot be ruled out. Composed of euchromatin as well as heterochromatin, the supernumerary chromosome 4 in *Drosophila* is often an enigma in itself. Being a mere 3.5% of the genome, it possesses domains

characteristic of euchromatin but interspersed with heterochromatin domains as well [58]. It was suggested that the small size and peculiar composition of chromosome 4 may hinder its expression, thus not revealing any substantial effect of heterochromatin divergence in hybrid backgrounds, particularly if the other chromosomes do not show significant divergence between the species involved.

In other insects, heterochromatin has proven to be a useful tool to investigate evolutionary relationships. *Tripanosoma infestans*, the major vector of Chagas disease, has a wide region of distribution in South American countries. Two different lineages called the Andean and Non Andean lineage encompass a host of different populations of the vector insect. These species differ in genome size corresponding to differences in the heterochromatin content. The dispersal history and pyrethroid resistance in *T. infestans* was investigated by examining the heterochromatin, karyotype and 45S rDNA location in several different populations in S. America [67]. Interestingly, a high amount of variability was discovered in the location and number of rDNA loci among the populations tested in the Andean region. In contrast, the non-Andean populations were monomorphic for rDNA location and cluster. A third Intermediate form was also revealed at lower altitudes, with characteristics of both the Andean and Non Andean population and exhibiting pyrethroid resistance. Cytogenetic differences between populations emphasized upon the Andean origin of *T. infestans*, while suggesting that the intermediate group was an example of genomic variability leading to formation of a drug resistant population. Using C-banding, the variation in constitutive heterochromatin in several different species of triatomine vectors have also been determined [68]. Some of these species differ in the location or presence of C band blocks, while others exhibit differences in

chromosome number. These revelations are helpful for taxonomic purposes for understanding Chagas disease spread across various tropical countries.

While available literature explains how heterochromatin varies in species exhibiting post-zygotic isolation, role of heterochromatin in pre-zygotic isolation is not clear. Few studies have investigated whether divergence in stDNA or other heterochromatin may itself act as a major factor in speciation. Red floor beetle *Tribolium castaneum* is a widespread pest of stored food grains, and hence a source of concern for many markets. It also possesses large blocks of pericentromeric heterochromatin, making it ideal as a subject of cytogenetic analysis for heterochromatin studies. A previous study had established the role of TCAST1 as a major stDNA which makes up a significant portion of the genome [69]. Research into various natural populations of *T.castaneum* revealed the presence of subfamilies of a minor satellite DNA TCAST2 [70]. Differences in copy number, monomer size (359 bp and 180 bp) and genomic location were revealed between the natural populations examined, suggesting the role of stDNA in genome divergence at population level. However, when strains exhibiting differences in TCAST2 were crossed, no significant effects were noted on the number of offspring compared to control crosses. Combined with the observation that TACST1 was conserved in both populations, it was suggested that preservation of major family of stDNA was sufficient to maintain fertility and overcome any effects of TCAST2 divergence between these strains. At present, insufficient evidence exists to answer the extent of role played by heterochromatin in divergence. Understanding the role of heterochromatin in pre-zygotic isolation of species would help lessen the existing gap in our knowledge.

1.11 X chromosome pericentric heterochromatin in *An. gambiae*

Malaria mosquito *Anopheles gambiae* provides a unique opportunity to investigate the role of heterochromatin in pre-zygotic isolation. Although many species have well developed large polytene chromosomes, which enable identification of chromosomal inversions and regions under positive selection, heterochromatin in *An. gambiae* is not as well explored as in *Drosophila*. Polytene chromosomes are usually obtained from salivary glands, ovaries and malphigian tubules with five distinct arms X, 2R, 2L, 3R and 3L. The Y chromosome is highly heterochromatic and underpolytenized, therefore not easily visible in polytene chromosomes. In *Anopheles* mosquitoes, heterochromatin occurs in the pericentromeric, telomere and intercalary positions. Morphological identification of pericentromeric heterochromatin based on level of compaction further divides it into proximal condensed alpha (repeat rich) and distal diffused beta (involved in nuclear architecture) type [71].

Physical maps based on polytene chromosomes correlate the available genome assembly data with a specific location on a chromosome. Chromosomal locations of epidemiologically important markers strengthen our knowledge of the vector while potentially identifying parts of the genome involved in rapid adaptation to changing climates host seeking behavior among others. The first physical map published after sequencing the genome of *An. gambiae* had 2000 BAC clones hybridized *in-situ* to chromosomes making up a total of 227 megabase pairs (Mbp) [48]. However, this assembly had large gaps due to the difficulty with assembling the repeat rich heterochromatin. It was later improved by the addition of 16.6 MB heterochromatin containing 232 genes [50].

The rapidly evolving heterochromatin in malaria vectors has been suggested to be an important factor in genome variability [72]. Comparison of different genomic features also illustrates the turnover of genes, chromosomes and heterochromatin between closely related species, furthering the understanding of population genetics. At present, major vector *An. gambiae* is undergoing sympatric diversification into two molecular forms - M and S. These forms were originally detected based on differences in internal transcribed spacer (ITS) and intergenic spacer (IGS) region [43, 73] and differ in larva habitat, rate of development, response to the presence of predators, chemical tolerance, assortative mating, mate selection etc. [74-76]. Using microarray-based sequencing, it was demonstrated that the regions of high differentiation between these forms were present in centromere-proximal heterochromatic regions on 2L and X, with an additional region in euchromatic 2R. These regions were termed as genomic "speciation islands" and were suggested to contribute significantly by harboring genes responsible for ecological and behavioral diversification [77]. Later a third region on centromere of chromosome 3L was identified as a potential island of divergence [78]. Physically mapping these islands revealed them to be present in areas of pericentric heterochromatin with sizes of 4.4Mb in X, 2.4 Mb in 2L and 1.8 Mb in 3L arm, signifying that the heterochromatin in malaria mosquito genome is rapidly evolving [72]. A caveat of these findings however, is that centromere-proximal regions with the low recombination may be incidental rather than the cause of speciation [78]. Later studies sampled various regions across a cline and demonstrated that the frequency of hybrids depended upon the region sampled. This ranges from 1% to as high as 42% [79]. Genome wide differences between the two genomes were demonstrated, negating the previous belief that fixed differences concentrated in a few islands were responsible for this diversification [79-81]. At the same time, the approx. 4 MB region of X chromosome with the highest divergence cannot be

ignored. Various studies revealed this to be the region of prime differentiation between the two forms [77, 81] [79]. Role of X chromosome heterochromatin has also been highlighted in assortative mating, a prime area of focus for prezygotic isolation that is the case in differentiation of nascent species [82]. Based on the available details regarding genomic, ecological and behavioral divergence the forms are now recognized as incipient species – *An. gambiae* (former S form) and *An. coluzzii* (former M form) [83].

Technical challenges with mapping limits researchers from investigating the heterochromatic arm making up half of X chromosome in detail, therefore important information could be missing from this region. Due to its late replicating nature, heterochromatin is underpolytenized, rendering the mapping of repetitive elements incomplete. In light of this predicament, it is crucial to find new means to investigate the heterochromatic arm in X and uncover potential qualitative or quantitative differences that may lead to differentiation between the two nascent species. Mitotic chromosomes provide a solution to this problem. Previous work has shown that inter- and intra- species differences in heterochromatin exist between *Anopheles* and its sibling species [84]. Unlike polytene chromosomes, mitotic chromosomes provide an increased level of resolution for the repetitive parts of the genome compared to the euchromatin region. A more detailed approach using mitotic chromosomes would help in understanding the role of heterochromatin in speciation.

1.12 Heterochromatin in the Y chromosome within the *An. gambiae* complex

Major malaria vector *An. gambiae* possesses heterogametic sex chromosomes where by XX denotes female while XY result in a male mosquito. Y chromosome in *Anopheles* is almost completely heterochromatic and a good sequence assembly for the Y still eludes researchers.

Although reported to comprise almost 10% of the genome, and bearing important male determining genes, the highly repetitive nature of Y renders sequencing and assembly inadequate when compared to the other chromosomes [48]. It is also underpolytenized, making mapping of Y chromosome difficult using polytene preparations in other *Anopheline* species [85]. Comprised of dense repeat rich heterochromatin, Y chromosome remains refractory to both sequencing and assembly. Polytene chromosomes are also not a desired tool to map the few known sequences. It is crucial to investigate the Y due to two main reasons: Evolutionary framework within repeats in malaria mosquitoes which comprise a young recently diverged (and diverging) group of insects, and transgenic mosquito control methods which constitute releasing thousands of genetically modified male mosquitoes to target a local population crash.

Previously, Y linked stDNA families were described [51, 52]. Differing in period size and copy number, these included the stDNA AgY53A, AgY53B and AgY477. A stDNA related to AgY477 but not Y linked was also discovered and termed Ag367, mapping to X chromosome. However, as Y does not polytenize well, information regarding mapping of these elements may have been incomplete. Another study identified 54 scaffolds which were Y –linked, increasing the portion of Y mapped to 182 Kb. A novel method called Chromosome Quotient (CQ) was recently used to identify six Y chromosome genes in male mosquitoes, three each from *An. gambiae* and *An. stephensi* (Asian malaria vector) [86]. However, these sequences and genes remain unmapped. While polytene chromosomes are frequently utilized to map sequences on X or autosomes in *Anopheles*, highly heterochromatic nature of Y renders it unsuitable for identifying or mapping using this approach. Mitotic chromosomes provide a better method to map repeats as polytenization plays no role in comparative analysis during mitosis. Thus, we can validate the available information and map the sequences on an ideogram. This in turn would

help characterize the molecular content of this rapidly evolving chromosome better, particularly by comparing sibling species of *An. gambiae* complex which has recent radiation history.

With near complete heterochromatin composition, the Y chromosome in *Anopheles* mosquitoes has posed the greatest challenge in uncovering the malaria mosquito genome. Although a few studies have attempted to elucidate the molecular contents of this chromosome, it still remains the biggest piece of puzzle in comprehending the sexual dimorphism in *Anopheles*. This puzzle is further complicated by the huge interspecific and intraspecific variations in Y chromosomes within the *An. gambiae* complex [87] [84]. Comprehensive knowledge of molecular composition and mapping of the Y chromosome will facilitate the implementation of current transgenic male mosquito release programs. In addition, it also presents an opportunity to study the potential role of Y chromosome and heterochromatin in evolution of malaria mosquitoes.

1.13 Epigenetics

The term "epigenetics" originally referred to heritable changes in the cellular phenotype which manifest themselves without changes in an organism's DNA sequence [88]. Thus it is the study of changes in gene expression resulting from phenomenon other than DNA sequences. Chief among these are changes in DNA methylation, acetylation and histone modifications. DNA and histone proteins are packed into the basic unit known as "nucleosome". Each nucleosome unit comprises of 147 bp of DNA wrapped around a histone octamer – further subdivided into two units of each histone protein H2A, H2B, H3 and H4 [89]. Alteration in an epigenome of an organism involves highly coordinated enzymes which tightly regulate these modifications, activating or deactivating gene expressions at specific locations and times. Various DNA and histone modifications alter gene expression by changing the nucleosome-nucleosome

interactions, behaving as docking sites or recruiting other chromatin modifiers to specific sites [90] [91]. DNA methylation, DNA acetylation and histone post-transational modifications (PTMs) are among the chief epigenetic mechanisms that affect gene expression in multicellular organisms. In addition, increasing focus is now on researching effects of chromatin modeling, non-coding RNA and histone variant regulators as well [92] [93].

Currently, epigenetics includes changes in expression or transcription potential which may be caused by external or environmental factors. RNA signaling involving DNA methyltransferases are capable of modifying chromatin complexes as well. Several non-coding RNA such as micro RNAs, Piwi-interacting RNA or piRNA, and long non coding RNA (lncRNA) are also regulated by epigenetic modifications. Thus these mechanisms may regulate important functions including development, homeostasis, environment stimuli and transgenerational expression of genes [94].

1.14 Epigenetic effects in organisms

During the last decade, a number of studies have highlighted the connection between epigenetics and vital processes in mammals, including but not limited to brain function [95] [96], pregnancy [97], spermatogenesis [98], aging [99] [100] [101] and most notably, cancer [102]. In plants, recent work in epigenetics has revealed the cascade of effects occurring during drought in rice through DNA methylation [103], effect on gene targeting of a vector in *Arabidapsis* [104], expression of flowering locus [105] and regulation of photosynthesis in maize [106]. Failure of plants to be able to move and escape deleterious alleles or effects makes epigenetics especially important for survival in their case [107]. Studies performed on other organisms have also elucidated the wide variety of roles played by epigenetic mechanisms in a myriad of important processes. These range from identifying novel biomarkers for aquatic toxicology in *Daphnia*, a

freshwater microcrustacean [108] and environmental and temperature effects on flatfish *Cynoglossus semilaevis* which exhibits sex-reversal leading to development of ZW juvenile females into fertile phenotypic 'pseudomales' [109] to regulation of tissue and organ regeneration in adult flatworms [110].

Epigenetics has also played an important role in extending our understanding of various phenomena in invertebrates, especially insects. For instance, epigenetic stabilization involving regulation of histone H3K4 methyltransferase levels is required to initiate differentiation in follical cell progenitors of *D. melanogaster* [111]. Silk gland methylome comparison of wild and domesticated silkworm species *Bombyx mori* and *B. mandarina* revealed higher expression of DNA-methylated genes, possibly leading to enlarged silk glands in the domesticated populations [112]. Interestingly, a novel view involving the role epigenetics may play in formation of social insects such as honey bees was recently published. Darwin's theory of natural selection stipulates that in sexually reproducing organisms, the fittest individuals would usually survive and mate to pass their traits, leading to formation of a population that adapts better to its surroundings and produces further generations that survive. However; insects of the order Hymenoptera (bees, ants) and Isoptera (termites) present a case of social grouping with sterile-fertile dimorphism, whereby genetically similar females are divided into a fertile queen and the rest get categorized into worker sub-classes. Ruden *et al.* [113] suggested that epigenetics might be the perfect approach to accommodate this anomaly. Focusing on a DNA modification in 5-methylcytocine (5mc), they explained that this environmentally affected epigenetic modification may well have paved the way for strong variable evolution of traits in genetically similar females [114] [113]. This view is further supported by the fact that honey bees use royal jelly to shape the development of their offspring. The active ingredient in royal jelly is an HDACi (histone

deacetylase inhibitor), supporting the view that epigenetics plays a greater role in evolution of social insects than previously thought. In another social insect carpenter ant *Camponotus floridanus*, histone acetylation was shown to regulate the diverse behavior of genetically identical but morphologically different ants directing them to perform different foraging and scouting minors [115].

Pharmacological studies of drug effects on cell lines and animal models have established the use of epigenetics and epigenetic drugs as a useful tool for modulating the genetics and physiology of cells and organisms. Epigenome characterization and manipulation can be achieved by a combination of several techniques such as Next Generation Sequencing (NGS), Chromatin Immunoprecipitation (ChIP), ChIP-seq, qRT-PCR, knowledge of nucleosome positioning maps, protein recruitment maps, transcription binding sites, DNA and histone modification site regulators and the respective pathways involved in these mechanisms [116] [117] [118] [99] [95] [119]. Latest technologies enable us to investigate the epigenome of interest and factors affecting it using these tools to answer questions which were often enigmatic in the past.

Plasmodium epigenetics have been explored in recent years, focusing on the infectious stage during malaria and the antiviral response [120]. Suppression of microRNA (miRNA) was shown to result in increased resistance to *P. falciparum* infection [121]. Brancucci *et. al.* [122] showed that conditional knockdown of *P. falciparum* heterochromatin protein 1 (PfHP1) led to de-repression of a select group of genes, among them the virulence *var* genes and the gene required for switching to the sexual cycle of parasite. Transcriptional regulation of *P. falciparum* by distinct histone modifications, particularly global repressive mark H3K36me2 and its putative role in pathogenesis was recently described [123]. Interestingly, epigenetic silencing of *clag* genes responsible for ion transport and nutrient uptake was shown to be associated with an

increased drug resistance of *P. falciparum*, reducing the efficacy of common water-soluble anti-malarial drugs [124]. Yet in an important disease vector like the malaria mosquitoes, research in epigenetics is surprisingly limited to a few studies.

Majority of the current control measures focus on insecticide development based on mosquito physiology. The ever-increasing problem of rapid resistance development in mosquito populations against established insecticides requires additional effort to identify novel control techniques. Despite the current focus on genetic advances, phenotypic variation in vector competence are difficult to rationalize using these venues alone [125]. A better understanding of epigenetic effects on mosquito development, survivorship and physiology could help close the gap of epigenetic interactions between human host, *Plasmodium* and mosquito in addition to identifying potential novel vector control leads.

Transcriptome of a new world malaria vector *An. albimanus* was prepared from cDNA libraries obtained from salivary gland, midgut, cuticle and whole body and 16,699 putative protein coding transcripts potentially involved in malaria transmission were identified [126]. Two key histone PTMS; H3K27me3 and H3K27ac were explored from the midgut and salivary gland of adult *An. gambiae* mosquitoes, revealing a mutually exclusive pattern across the epigenome [127]. Interestingly, H3K27ac, which is a transcriptionally active marker localizing mostly to active genes and promoters, covered only 6.81% of the X chromosome, while H3K27me3 (repressive marker expressed in genes with low expression) covered 28.45% of the X chromosome, the maximum among all chromosomes. This is probably due to the high heterochromatin content on the X chromosome. Uncovering the epigenetic mechanisms of H3K27me3 on X chromosome could lead to a better understanding of X heterochromatin and female mosquitoes, the major target areas of vector control. Exploration of piRNA pathway in *An. stephensi* revealed the

transposon protector role of three conserved piRNA genes - Ago3, Auburgine and Piwi. These genes were upregulated after a blood meal, suggesting their role in ovarian development [128]. A comparison of piRNA loci between *D. melanogaster* and *An. gambiae* revealed differences in their distribution. In contrast to the clusters mostly present at telomeres and centromeres in *D. melanogaster*, in *An. gambiae* these major clusters were also found to be in regions outside the pericentric heterochromatin. In addition, the piRNA production was seen from genes important for germline and embryonic development [129]. Transgenerational effects due to environmental triggers were studied using epigenetic compounds in *Ae. aegypti*, a major vector of dengue. Two methylation modifying agents including a fungicide were applied to parental generation and subsequent generations were studied for phenotype variation. Significant lowering of susceptibility to commonly used insecticide imidacloprid was revealed [130].

Despite extensive research in various insect taxa, epigenetic studies in mosquitoes are limited. This is surprising considering mosquitoes constitute a major threat to human health and vector control is still the major means to decrease the impact of diseases vectored by mosquitoes. As the awareness regarding epigenetics in different life systems rises, we see an increase in the efforts to investigate the diverse roles played by epigenetics in mosquito immunity, insecticide resistance and characterization of epigenetic genes. Recently, 169 genes involved in epigenetic control in *Anopheles* which are orthologous to 215 gene ensemble in *D. melanogaster* were identified [94]. Changes in environment can also trigger epigenetic effects. In the Asian tiger mosquito *Ae. albopictus*, a study performed evaluated the effect of DNA methylation altering agents. Exposed individuals and subsequent generations were analyzed for changes in the DNA methylation levels. The generations never exposed to the agent also showed significant decrease in insecticide sensitivity, suggesting the importance of epigenetic mechanisms involved in a

process vital to success of vector control programs worldwide [130]. With the increasing resistance to established pesticides, we need new armor to fight against the deadliest insect. As efforts to create a successful vaccine for malaria are still underway, the potential of exploiting histone regulators or DNA modifiers as a tool to create new vector control strategy seems promising. Investigating the effect of epigenetics in mosquitoes has the potential to provide us with novel leads to achieve this target, thus helping in our fight against mosquito borne diseases like malaria, dengue, Chikungunya and Zika.

1.15 Drugs used in epigenetic treatments

Disruption of epigenetic processes may lead to changes in gene function, expression and malignant cellular transformation in some cases [131] [132]. Surprisingly, methylation-associated silencing has been found in almost 50% of the genes causing various cancers till date [133] [134]. Both hyper and hypo-methylation of CpG islands (regions with greater than 50% of C-G sequence and longer than 200bp length) are associated with silencing of tumor-suppressor genes and overexpression of oncogenes respectively in mammals. As epigenetic changes are often reversible, this makes epigenetic therapy an exciting and feasible approach for development of anti- cancer drugs. Chief among those undergoing clinical trials are DNA methyltransferase inhibitors (DNMTis) such as 5AC (Vidaza), DAC (Decitabine) and SGI-110 in low doses. Histone Deacetylase inhibitors (HDACs) are another class of compounds showing promise in cancer therapeutics. Interestingly, some families of HDACs act upon nuclear proteins in addition to histones, opening up new avenues for exploiting this trait. The HDAC drugs currently in clinical trial include broad acting inhibitors such as vorinostat or specific inhibitors such as romidepsin. However, these are toxic in high doses and may lead to double strand DNA break or cellular death [135]. Deazaneplanocin-A or DZNep is a methyl transferase inhibitor of

EZH2 (Enhance of Zeste Homolog 2), part of the Polycomb Repressive Complex 2 (PRC2) associated with many cancers [102] [136]. Additionally, it is also an S-adenocyl homocysteine (SAH) hydrolase inhibitor. In recent years, DZNep has shown promise against a variety of cancers including gastric [137], breast [102], lung [138] and colon [136] cancer with different sensitivity and effects associated with each and is currently undergoing clinical trials to determine its potential as an anti-cancer drug.

Chapter 2 - Fluorescent *in situ* hybridization on mitotic chromosomes of mosquitoes

Vladimir A. Timoshevskiy, Atashi Sharma, Igor V. Sharakhov, Maria V. Sharakhova

Vladimir A. Timoshevskiy

Department of Entomology

Virginia Tech

timvl@vt.edu

Atashi Sharma

Department of Entomology

Virginia Tech

atashi04@vt.edu

Igor V. Sharakhov

Department of Entomology

Virginia Tech

igor@vt.edu

Maria V. Sharakhova

Department of Entomology

Virginia Tech

msharakh@vt.edu

Corresponding author: Maria V. Sharakhova

Keywords: imaginal discs, mitotic chromosomes, genome mapping

2.1 Abstract

Among the three mosquito genera, namely *Anopheles, Aedes,* and *Culex,* physical genome mapping techniques were established only for *Anopheles*, whose members possess readable polytene chromosomes. For the genera of *Aedes* and *Culex*, however, cytogenetic mapping remains challenging because of the poor quality of polytene chromosomes. Here we present a universal protocol for obtaining high-quality preparations of mitotic chromosomes and an optimized FISH protocol for all three genera of mosquitoes.

2.2 Introduction

Fluorescent *in situ* hybridization (FISH) is a technique routinely used by many laboratories to determine the chromosomal position of DNA and RNA probes. One important application of this method is the development of high-quality physical maps useful for improving the genome assemblies for various organisms. The natural banding pattern of polytene and mitotic chromosomes provides guidance for the precise ordering and orientation of the genomic supercontigs. Among the three mosquito genera, namely *Anopheles, Aedes,* and *Culex,* a well-established chromosome-based mapping technique has been developed only for *Anopheles,* whose members possess readable polytene chromosomes [139]. As a result of genome mapping efforts, 88% of the *An. gambiae* genome has been placed to precise chromosome positions [140], [49]. Two other mosquito genera, *Aedes* and *Culex,* have poorly polytenized chromosomes because of significant overrepresentation of transposable elements in their genomes [141], [142], [143]. Only 31 and 9% of the genomic supercontings have been assigned without order or orientation to

chromosomes of *Ae. aegypti* [144] and *Cx. quinquefasciatus* [145], respectively. Mitotic chromosome preparation for these two species had previously been limited to brain ganglia and cell lines. However, chromosome slides prepared from the brain ganglia of mosquitoes usually contain low numbers of metaphase plates [146]. Also, although a FISH technique has been developed for mitotic chromosomes from a cell line of *Ae. aegypti* [147] , the accumulation of multiple chromosomal rearrangements in cell line chromosomes {Steiniger, 1975 #912} makes them useless for genome mapping. Here we describe a simple, robust technique for obtaining high-quality mitotic chromosome preparations from imaginal discs (IDs) of 4th instar larvae which can be used for all three genera of mosquitoes. A standard FISH protocol {Garimberti, 2010 #926} is optimized for using BAC clones of genomic DNA as a probe on mitotic chromosomes of *Ae. aegypti and Cx. quinquefasciatus,* and for utilizing an intergenic spacer (IGS) region of ribosomal DNA (rDNA) as a probe on *An. gambiae* chromosomes.. In addition to physical mapping, the developed technique can be applied to population cytogenetics and chromosome taxonomy/systematics of mosquitoes and other insect groups.

2.3 Protocol:

1) Chromosome preparation

Mosquito larvae were reared using a standard protocol described in Methods in Anopheles Research available at the website of the Malaria Research and Reference Reagent Resource Center (MR4) { #4003}. The temperatures of mosquito rearing were modified to provide the highest number of chromosomes in imaginal discs and lowest mortality of the larvae. The stages of mosquito larvae development were determined based on the sizes of their head capsules {, #4003}.

1.1) Hatch mosquito eggs at 28°C, and after 2-3 days, transfer 2nd or 3rd instar larvae to 16°C for *Ae. aegypti* and *Cx. quinquefasciatus* and to 22°C for *An. gambiae.*

1.2) Place 4th instar larvae on ice for several minutes for immobilization.

1.3) Transfer larva to a slide with a drop of cold hypotonic solution (0.5% sodium citrate or 0.075 M potassium chloride), and place it under the stereo microscope.

1.4) Select larva with oval IDs (Figure 1B) for further dissection.

1.5) Decapitate larva, and cut the cuticle from the ventral side of the larval thorax using dissecting scissors (Figure 2A). Make additional cut in second or third abdominal segment to dissect the gut from the larva. The directions of the cuts are shown by arrows.

1.6) Open the cuticle, and remove the gut and fat body from the larva. Remove the hypotonic solution from the slide using filter paper, and add a fresh drop of hypotonic solution directly to the IDs (Figure 2B). Keep larva in hypotonic solution for 10 min at RT.

1.7) Remove hypotonic solution using filter paper, and apply Carnoy's solution (ethanol/acetic acid in 3:1 ratio). After adding fixative solution, IDs immediately turn white and become easily visible under the microscope (Figure 2C).

1.8) Using dissecting needles, remove IDs from the larva (Figure 2D), and transfer them to a drop of 50% propionic acid. Remove any other tissues, such as the gut and fat body, from the slide. Cover IDs with an unsiliconized 22x22 cover slip, and keep for 10 min at RT.

1.9) Cover the slide with filter paper, and squash the tissue by tapping the eraser of a pencil on the perimeter of the cover slip.

1.10) Briefly analyze the quality of the slide using the phase-contrast microscope at 100x or 200x magnification (Figure 3). Preparations with >50 chromosome spreads can be considered suitable for FISH.

1.11) Dip and hold the slide in liquid nitrogen until it stops bubbling. Remove the cover slip from the slide using a razor blade, and transfer the slide immediately to a container of 70% ethanol chilled at -20°C. Store at 4°C for at least 1 hour for the best dehydration result (if necessary, slides can be stored at this step from several minutes to several days).

1.12) Dehydrate slides in a series of ethanol (70%, 80%, 100%) at 4°C for 5 min each, and air dry at RT.

1.13) Store dry slides at -20°C before utilizing them for FISH.

2) Extraction of repetitive DNA fractions

Performing FISH of the BAC clone DNA probe on chromosomes from Ae. aegypti and Cx. quinquefasciatus requires using unlabeled repetitive DNA fractions to block unspecific hybridization of the DNA repeats to the chromosomes. The reassociation of single-strand DNA fragmented into pieces of several hundred bp follows a C_0t curve where C_0 is the initial concentration of single-stranded DNA and t is the reannealing time. DNA fractions with C_0t values equal to 10^{-4}-10^{-1} or 10^{0}-10^{2} are considered as highly and moderately repetitive, respectively.

2.1) Extract 400-500 µg of the genomic DNA from entire adult mosquito using Qiagen Blood and Cell Culture Maxikit, and prepare 100-1,000 ng/µl DNA solution in 1.2x SSC.

2.2) Denature DNA by placing a safe-lock tube with genomic DNA into a heating block prewarmed to 120°C for 2 min. High temperature helps to range DNA into 200-500 bp fragments.

2.3) Depending on the DNA concentration, reassociate DNA by placing the tube at 60°C for 15-150 min to obtain C_0t DNA fractions up to C_0t3 (Table 1).

2.4) Place the tube with DNA on ice for 2 min.

2.5) Transfer the DNA to 42°C, add preheated 10x S1 nuclease buffer and S1 nuclease to a final concentration of 100 U per 1 mg of DNA, and incubate for 1 hour.

2.6) Precipitate DNA by adding 0.1 volume of 3 M sodium acetate and 1 volume of isopropanol at RT.

2.7) Centrifuge at 14,000 rpm for 20 min at 4°C.

2.8) Wash DNA in 70% ethanol, and centrifuge again at 14,000 rpm for 10 min at 4°C.

2.9) Air-dry and dissolve DNA pellet in TE buffer.

2.10) Measure the DNA concentration, and visualize by gel electrophoresis. Usually the final quantity of repetitive DNA fractions represents 35-50% of the original DNA amount.

3) DNA probe labeling

Two different protocols were used for the labeling BAC clone DNA probe and IGS rDNA probe.

3.1) BAC clone labeling using nick-translation

3.1.1) Extract BAC clone DNA from the BAC library using Qiagen Large Construct Kit.

3.1.2) Prepare reaction mixture for nick-translation labeling on ice with final volume of 50 µl: 1 µg isolated BAC clone DNA, 0.05 mM each of unlabeled dATP, dCTP, and dGTP and 0.015 mM of dTTP; 1 µl of Cy3-dUTP (or another fluorochrome); 0.05 mg/ml of BSA, 5 µl of 10x nick-translation buffer, 20 U of DNA-polymerase I, and 0.0012 U of DNase.

3.1.3) Incubate at 15°C for 2.5 hour.

3.1.4) Stop reaction by adding 1 µl of 0.5 M EDTA.

3.1.5) Store probe at -20°C in a dark place.

3.2) IGS rDNA labeling using PCR

3.2.1) Prepare reaction mixture on ice with final volume of 50 µl: 200 ng of genomic DNA; 0.05 mM each of unlabeled dATP, dCTP, and dGTP; 0.015 mM of dTTP; 1 µl of Cy3-dUTP (or another fluorochrome); 5 µl of 10x PCR-buffer; 50 pmol of forward; UN (GTGTGCCCCTTCCTCGATGT) and reverse; GA (CTGGTTTGGTCGGCACGTTT) primers for IGS amplification; and 10 U of Taq DNA polymerase [148].

3.2.2) Perform PCR reaction using standard PCR parameters for IGS amplification: 95°C /5 min x 1 cycle; (95°C /30 sec, 50°C /30 sec, 72°C /30 sec) x 30 cycles; 72°C /5 min x 1 cycle; and 4°C hold [148].

3.2.3) Store probe at -20°C in a dark place.

4) Fluorescent *in situ* hybridization

This FISH protocol includes two variations: the first for using BAC clone DNA as a probe on mitotic chromosomes of Ae. aegypti and Cx. quinquefasciatus and the second for using IGS rDNA on mitotic chromosomes of An. gambiae. If using BAC clone DNA probes, skip RNase treatment steps 4.3, 4.4, and simultaneous slide/probe denaturation step 4.19. If using IGS rDNA probe, prepare hybridization mixture without C_0t DNA fractions, and skip separate slide/probe denaturing steps 4.10, 4.11, 4.16, and 4.17.

4.1) Incubate slides in 2x SSC for 30 min at 37°C.

4.2) Dehydrate slides in series of 70%, 80%, and 100% ethanol for 5 min each at RT, and air dry.

If performing FISH with BAC clone DNA, proceed directly to step 4.5.

4.3) Incubate chromosome preparation in 0.1 mg/ml RNase solution under parafilm for 30 min at 37°C**.**

4.4) Wash twice in 2x SSC for 5 min each at 37°C.

4.5) Put slides in a jar with 0.01% pepsin and 0.037% HCl solution, and incubate for 5 min at 37°C.

4.6) Wash slides in 1x PBS for 5 min at RT.

4.7) Fix chromosome preparation in a jar with 1% formalin in 1x PBS prepared from 10% neutral-buffered formalin for 10 min at RT.

4.8) Wash slides in 1x PBS for 5 min at RT.

4.9) Dehydrate slides in series of 70%, 80%, and 100% ethanol for 5 min each at RT, and air dry preparations at 37°C.

If performing FISH with IGS, proceed directly to step 4.12

4.10) Denature slides in a jar with prewarmed 70% formamide for 2 min at 72°C.

4.11) Dehydrate slides in series of cold (-20°C) 70%, 80%, and 100% ethanol for 5 min each, and air dry at 37°C.

4.12) Prepare hybridization mixture: 5 µl of labeled probe DNA from step 3, 10 µl of C_0t DNA from step 2 with final concentration of 0.5 ng/µl, and 5 µl of 1 µg/µl sonicated salmon sperm DNA. *For FISH with IGS rDNA, prepare hybridization mixture without C_0t DNA fractions.*

4.13) Precipitate DNA by adding 0.1 volume of 3 M sodium acetate and 2 volumes of ethanol. Keep at -20°C for 1-3 hrs.

4.14) Centrifuge at 14,000 rpm at 4°C for 20 min, remove the ethanol, and air dry the pellet at RT.

4.15) Thoroughly dissolve the pellet in 10 µl of hybridization buffer: 50% formamide, 20% dextran sulfate, 2x SSC.

If performing FISH with IGS, proceed directly to step 4.18

4.16) Denature hybridization mixture for 7 min at 97°C, and immediately put on ice for 1 min.

4.17) Prehybridize mixture at 37°C for 30 min to prevent unspecific hybridization of repetitive DNA to the chromosomes.

4.18) Place 10 µl of the hybridization mixture on the slide, and cover with a 22x22 cover slip. Prevent bubble formation - air bubbles should be removed with gentle pressure to the coverslip.

If performing FISH with BAC clone DNA, proceed directly to step 4.20

4.19) Denature the probe and chromosome DNA simultaneously using a heating block at 75°C for 5 min.

4.20) Glue cover slip around the perimeter using rubber cement.

4.21) Perform overnight hybridization in a humid chamber at 37°C.

4.22) Remove rubber cement and coverslip from the slide.

4.23) Wash slide 2 min in prewarmed Solution 1 (0.4x SSC, 0.3% Nonidet-P40) at 73°C.

4.24) Wash slides in Solution 2 (2x SSC, 0.1% Nonidet-P40) for 5 min at RT.

4.25) Counterstain slide using 0.001 mM YOYO-1 in 1x PBS for 10 min in humid chamber at RT.

4.26) Mount in a small amount of Prolong Gold antifade reagent with a cover slip.

4.27) Analyze preparations under a fluorescent microscope using appropriate filter sets at 1,000x magnification (Figure 4).

2.4 Representative Results:

Insect IDs are located in each segment of the larva. Depending on the position, they transform into different tissues at the adult stage of the insect. The IDs, which are used for the chromosome preparation in this protocol, develop into legs at the adult stage of the mosquito. These IDs are located at the ventral side of the larval thorax and are clearly visible through the cuticle under the microscope (Figure 1). At the early 4^{th} instar larval stage, IDs have a round shape (Figure 1A). The largest numbers of mitosis, ~175 in one ID [146], are accumulated at a later "oval shaped" stage (Figure 1B), which must be considered the optimal stage for slide preparation. At this time, the intermediate ID splits into two: one transforms into a leg and another one transforms into a wing. We prefer using the large leg IDs at the "oval-shaped" stage for the chromosome slide preparation. Figure 1C represents IDs at the latest stage of 4^{th} instar larva development. At this stage, the IDs are already developed into legs and wings, and contain a significant amount of differentiated tissues and a low number of mitosis. This stage of ID development should be avoided for chromosome slide preparation. We also recommend rearing mosquito larvae at low temperatures: 16°C for *Aedes* and *Culex* and 22°C for *Anopheles.* This helps to increase the amount of mitosis in IDs [146].

Figure 2 illustrates ID dissection from the thorax of 4th instar larva. Because the cuticle of a live insect is hard to dissect, we recommend using dissecting scissors instead of the needles commonly used for larva preparation. The most crucial procedure for obtaining high-quality chromosome preparation is the hypotonic solution treatment.

For best results, we remove the gut and fat body from the larval thorax before this treatment. Swelling of the ID cells during this procedure helps to spread chromosomes on a slide (Figure 3A). The appropriate quality of the hypotonic solution treatment can be easily recognized by the round shape of cells in the preparations (Figure 3A, B). Cells with an oval shape indicate insufficient hypotonic solution treatment (Figure 3C). To be selected for FISH, chromosome preparation should contain at least 50 high-quality chromosome spreads. Normally, ~90% of the slides prepared using this protocol have sufficient quality for FISH [146].

We present two slightly different FISH protocols: an advanced protocol for FISH using genomic BAC clone DNA probe on mitotic chromosomes of *Aedes* and *Culex* and a simple FISH protocol for IGS rDNA probe on mitotic chromosomes of *Anopheles*. The genomes of *Aedes* and *Culex* are highly repetitive because of the overrepresentation of transposable elements [144], [145]. Thus, performing FISH, which utilizes genomic BAC clone DNA as a probe, requires adding unlabeled repetitive DNA fractions to the probe to block unspecific hybridization of the DNA repeats to chromosomes. For the extraction of the repetitive DNA fractions, genomic DNA is denatured at 120°C for 2 min. Boiling DNA at a high temperature also helps to obtain DNA in fragments of 200-500 bp. DNA is allowed to reassociate after this treatment. The highly repetitive

DNA fragments tend to find their mate for reassociation faster than DNA with unique sequences does. As a result, the reassociation of DNA follows a C_0 xt curve where C_0 is the initial concentration of single-stranded DNA, and t is the reannealing time. DNA fractions with C_0t values equal to 10^{-4}-10^{-1} or 10^{0}-10^{2} are considered highly and moderately repetitive, respectively. The time of the reassociation for different C_0t DNA fractions can be calculated using the formula $t = C_0t\text{X} \times 4.98/C_0$; where t – time of incubation, C_0tX – C_0t fraction (C_0t1=1, C_0t2=2, etc.) and C_0 - initial DNA concentration in µg/µl {Trifonov, 2009 #927} (Table 1). After reassociation, the single-stranded DNA is digested using S1 nuclease. We prefer using all C_0t DNA fractions up to C_0t3 together instead of the commonly used C_0t1 DNA fraction. These C_0t fractions include some of the moderately repetitive DNA sequences and together usually represent 35-50% of the original amount of the genomic DNA in *Ae. aegypti*. The correct proportion between labeled DNA probe and unlabeled C_0t DNA fraction depends on the repetitive DNA component in each particular BAC clone. On average, we use 1:20 probe to C_0t DNA fraction proportion for obtaining an acceptable signals/background ratio of the FISH result. Prehybridization of the DNA probe with C_0t DNA fractions in a tube for 30 min before the actual hybridization on the slide also helps to reduce background. Labeling, hybridization itself, and washing in this protocol are performed using standard conditions {Garimberti , 2010 #926}.

The FISH results of two differently labeled BAC clone DNA probes on mitotic chromosomes of *Ae. aegypti* and *Cx. quinquefasciatus* are shown in Figures 4A and B, respectively. The BAC clone DNA probes produce strong signals in a single position on the chromosomes. Chromosomes shown in Figure 1 are counterstained with YOYO-1 iodide. This dye produces the best banding patterns on *Ae. aegypti* chromosomes [146]. Alternatively, other fluorescent dyes, such as DAPI or propidium iodide, can be utilized for the chromosome

counterstaining. For suppressing photobleaching of the slides, we use Prolong Gold antifade mounting medium. This reagent has good signal preservation abilities and also can be easily removed from the slide by rinsing in 1x PBS if it is necessary to use the same slide for several hybridizations.

A simple version of the FISH protocol is designed for hybridization of IGS rDNA probe on mitotic chromosomes of *Anopheles*. Ribosomal genes in *Anopheles* are represented as a polymorphic cluster of genes located on sex chromosomes {Collins, 1987 #6}. A DNA probe in this protocol is labeled using standard PCR reaction by adding fluorescently labeled Cy3 or Cy5 dNTPs. Because blocking unspecific hybridization of repetitive DNA in euchromatin is not needed, all steps related to using C_0t DNA fractions are omitted. Instead, chromosome preparations are pretreated with RNase for preventing hybridization of the IGS rDNA probe to the nucleolus. Chromosomes and the DNA probe are denatured simultaneously by heating the slide together with a probe in a hybridization system at 75°C for 5 min. Hybridization and washing in this protocol are also performed using standard conditions for FISH {Garimberti, 2010 #926}. The result of FISH is demonstrated in Figure 4C: the polymorphism of the IGS rDNA hybridization between two X chromosomes is clearly visible.

Tables and figures:

Table 1. DNA concentration and reannealing times for preparation of C_0t2 and C_0t3 fractions

	DNA concentration μg/μl	Reannealing time, min
C_0t 2	0.1	100

	0.3	33
	0.5	20
	0.7	14
	0.9	11
	1	10
C_0t 3	0.1	150
	0.3	50
	0.5	30
	0.7	21
	0.9	17
	1	15

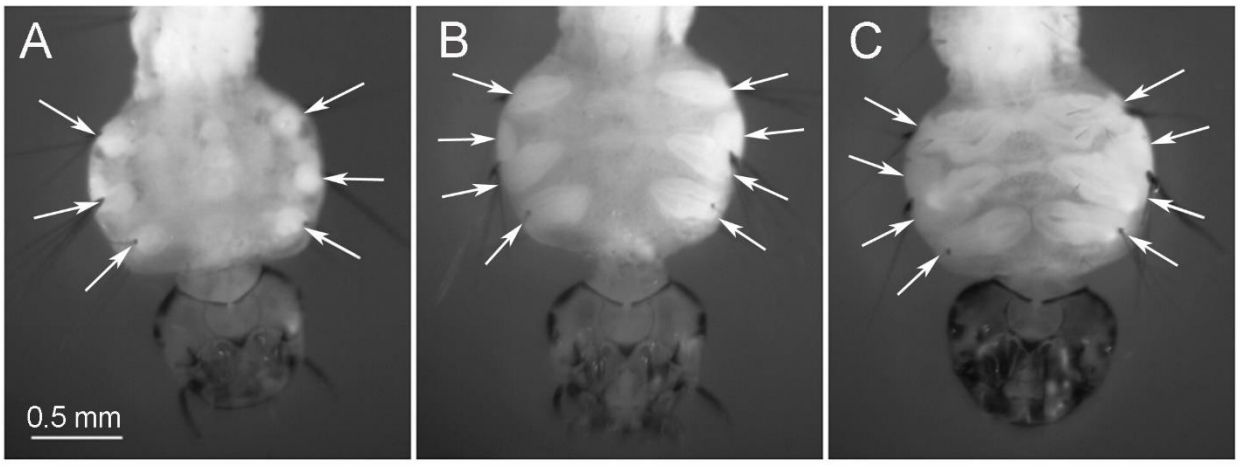

Figure 1. Stages of the ID development in 4th instar larva: A) an early "round shape" stage; B) an intermediate "oval shape" stage – optimal for the chromosome preparation; C) a late stage – inappropriate for chromosome preparations. The positions of IDs are indicated by arrows on the ventral side of the larval thorax.

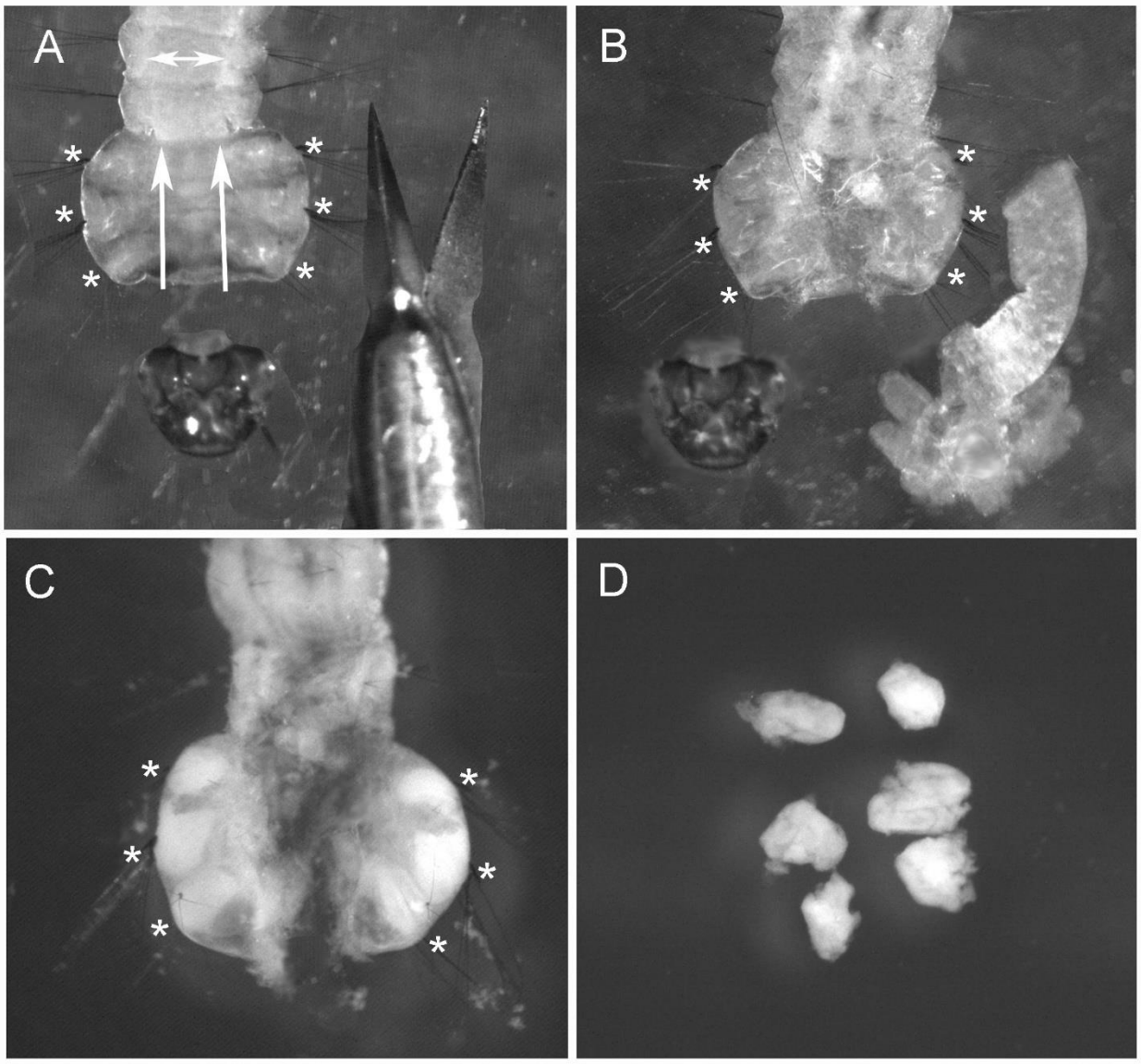

Figure 2. Steps of ID dissection in *Aedes aegypti*: A) decapitated larva (the direction of cuts are indicated by arrows); B) larvae with dissected gut under hypotonic solution treatment (IDs swell and become almost invisible); C) larva after Carnoy's solution application (IDs become white and clearly visible); D) dissected IDs in Carnoy's solution. Positions of IDs in larva are indicated by asterisks.

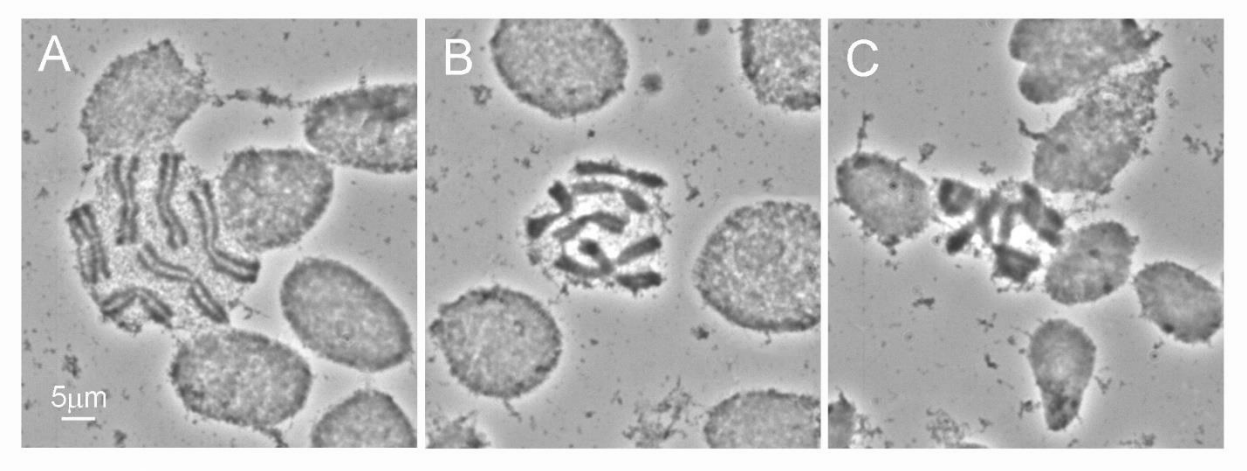

Figure 3. Different qualities of the chromosome spreads: A) a perfect chromosome spread - round shape of the cells demonstrates sufficient treatment of the IDs in hypotonic solution; B) a perfect hypotonic treatment - chromosomes are slightly undersquashed; C) a poor chromosome spread – the result of insufficient hypotonic treatment is indicated by oval shape of the cells.

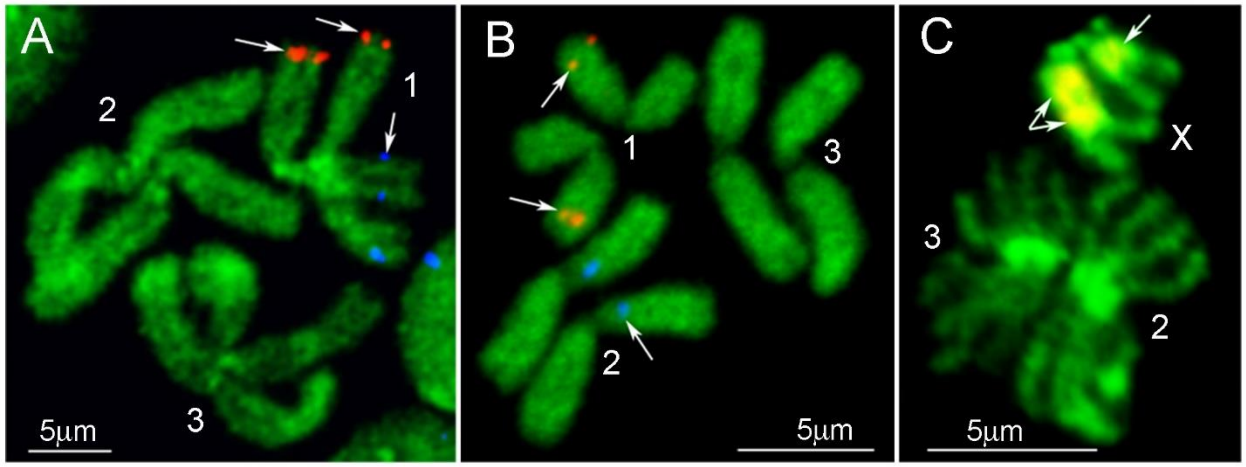

Figure 4. Examples of FISH with BAC clones and IGS rDNA (A, B) and IGS rDNA (C) in the chromosomes of *Ae. aegypti* (A), *Cx. quinquefasciatus* (B), and *An. gambiae* (C). 1, 2 and 3 – are numbers of chromosomes; X – female sex chromosome in *An. gambiae.*

2.5 Discussion:

Nonfluorescent *in situ* hybridization on mitotic chromosomes of mosquitoes was performed for the first time in 1990 by A. Kumar and K. Rai [149]. In that study, 18S and 28S ribosomal DNA genes, cloned together in one plasmid, were placed to the chromosomes of 20 species of

mosquitoes. The DNA probe was radioactively labeled and hybridized to the chromosomes from brain ganglia. Among three mosquito genera, a FISH technique has been developed only for mitotic chromosomes from the cell line of *Ae. aegypti* [147], [150], [151] and has never been performed on mitotic chromosomes from live mosquitoes. Recently, we developed a simple, robust technique for obtaining high-quality chromosome preparations from IDs of 4^{th} instar larvae [146]. This method allows a high number of chromosomes to be obtained in one slide and can be universally used for all species of mosquitoes. The necessity of using only larval, not pupal or adult stages of mosquitoes, for slide preparation is probably the only limitation of the method. The standard FISH method {Garimberti, 2010 #926} was optimized for using genomic BAC clone and IGS rDNA as probes for the mitotic chromosomes of *Aedes, Culex,* and *Anopheles*.

In addition to these specific applications, the FISH protocols described here can also be used for other purposes. The advanced FISH protocol, which utilizes C_0t DNA fractions for blocking unspecific hybridization, can also be applied for the hybridization of BAC clones or any other large DNA fragments in heterochromatic regions of *Anopheles*. Heterochromatic regions are enriched with transposable elements and other repeats, and probes from these regions normally produce strong background on the chromosomes [49]. Using unlabeled C_0t DNA fractions will help to reduce unspecific hybridization of the probe to the chromosomes. The simple version of the FISH protocol can be used for any rDNA or repetitive DNA probes on mitotic chromosomes of mosquitoes and other insects. In addition, it also can be applied for the hybridization of BAC clone DNA in species with low repetitive DNA content in euchromatic regions such as *Anopheles* or *Drosophila*. The protocol proposed here will help to obtain highly-finished chromosome-based genome assemblies for mosquitoes and can be broadly used for various cytogenetic applications in other groups of insects.

2.6 Acknowledgments: We thank Sergei Demin and Tatyana Karamysheva for their help with chromosome preparation and FISH on *Anopheles*. We also thank David Severson for providing us *Aedes* and *Culex* genomic DNA BAC clones and Melissa Wade for editing the text. This work was supported by two grants from the National Institutes of Health: 1R21 AI88035-01 to Maria V. Sharakhova and 1R21 AI094289-01 to Igor V. Sharakhov.

Disclosures: We have nothing to disclose.

Table 2. List of specific reagents and equipment to perform FISH:

Name of the Reagent/Equipment	Company	Catalogue number	Comments
MZ6 Leica stereomicroscope	Leica	VA-OM-E194-354	A different stereomicroscope can be used
Olympus CX41 phase microscope	Olympus	CX41	A different phase microscope can be used
Olympus BX61 fluorescent microscope	Olympus	BX61	A different fluorescent microscope can be used
ThermoBrite™ Slide Denaturation/Hybridization System	Abbott Molecular	30-144110	Serves as a heating block and a humid chamber
Dissecting needles	Fine ScienceTools	10130-10	
Needle holders	Fine Science Tools	26018-17	
Dissecting scissors	Fine Scienc Tools	15000-03	
75x25 double frosted micro slides	Corning	2949-75x25	
22x22 mm microscope coverslips	Fisher Scientific	12-544-10	
Parafilm	Fisher Scientific	13-374-10	
Rubber Cement	Fisher Scientific	50-949-105	

Acetic acid	Fisher Scientific	A491-212	
Alcohol 200 Proof	Decon Laboratories	2701	
Propionic acid	Sigma-Aldrich	402907	
Hydrochloric acid	Fisher Scientific	A144-500	
Sodium citrate dihydrate	Fisher Scientific	S279-500	
Sodium acetate trihydrate	Fisher Scientific	BP334-500	
Potassium chloride	Fisher Scientific	BP366-500	
EDTA	Fisher Scientific	S311-500	
Tris base	Fisher Scientific	BP152-1	
10x PBS	Invitrogen	P5493	
10% NBF (neutral buffered formalin)	Sigma-Aldrich	HT501128	
99% formamide	Fisher Scientific	BP227500	
Dextran sulfate sodium salt	Sigma-Aldrich	D8906	
20x SSC buffer	Invitrogen	AM9765	
1 mM YOYO-1 iodide (491/509) solution	Invitrogen	Y3601	
Antifade Prolong Gold reagent	Invitrogen	P36930	
dATP, dCTP, dGTP, dTTP	Fermentas	R0141, R0151, R0161, R0171	
Cy3-dUTP, Cy5-dUTP	GE Healthcare	PA53022, PA5502	
BSA	Sigma-Aldrich	A3294	
DNA Polymerase I	Fermentas	EP0041	
DNase I	Fermentas	EN0521	
S1 Nuclease	Fermentas	EN0321	

Taq DNA Polymerase	Invitrogen	18038-042	
RNase	Sigma-Aldrich	9001-99-4	
Pepsin	USB	9001-75-6	
Salmon sperm DNA	Sigma-Aldrich	D7656	
Nonidet-P40 (NP40)	US Biological	NC9375914	
Qiagen Blood and Cell Culture Maxikit	Qiagen	13362	
Qiagen Large Construct Kit	Qiagen	12462	

Chapter 3 - Molecular restructuring of heterochromatin in evolution of the *Anopheles gambiae* complex

Atashi Sharma, Nick Kinney, Maria Sharakhova and Igor Sharakhov

3.1 Abstract

Heterochromatin has been identified as a potential factor driving evolution between closely related species. Previous cytogenetic analyses demonstrated variations in the amount and positions of X chromosome heterochromatin blocks among and within species of malaria mosquitoes from the *Anopheles gambiae* complex. However, the molecular basis of these differences has not been identified. In this study we mapped DAPI stained heterochromatin blocks, ribosomal DNA (rDNA), two satellite DNA (stDNA) repeats, and euchromatin-heterochromatin boundary in mitotic chromosomes of *An. gambiae*, *An. coluzzii*, *An. arabiensis*, *An. quadriannulatus*, and *An. merus*. We also developed idiograms for metaphase chromosomes of species from the *An. gambiae* complex. Here we demonstrate that pericentric X chromosome heterochromatin not only exhibits different patterns of DAPI-positive bands but also variations in the amount of ribosomal DNA and positions of satellite DNA probes across sibling species. A custom MATLAB script identified the differences in relative positions of intensity peaks from stDNA probes and the proximal DAPI band on X chromosomes among six strains of *An. gambiae* and *An. coluzzii*, the most closely related species within *An. gambiae* complex. The alternating positions of these peaks indicate a rearrangement or putative inversions of stDNA repeats within the heterochromatin band. The 'inversion' differentiates strains but segregates within both sibling species indicating that heterochromatin undergo restructuring during early speciation events in malaria mosquitoes. The idiograms developed here will aid in the study of heterochromatin evolution in malaria vectors.

3.2 Introduction

The major malaria vector in Africa, *Anopheles gambiae* has been a subject of extensive research over the past few decades. Initially recognized as a cryptic species, *An. gambiae* was later subdivided into a complex of eight morphologically indistinguishable species by crossing experiments, fixed differences in polytene chromosome arrangement [45], differences in Intergenic Sequence (IGS) and Internal Transcribed Spacer (ITS) regions in the ribosomal DNA (rDNA) [43] [73] and most recently, by whole genome divergence [80] [44]. The current list of members of the *An. gambiae* complex include *An. arabiensis, An. amharicus, An. bwambae, An. coluzzii, An. gambiae, An. merus, An. melas, and An. quadriannulatus* at present. The more recently diverged species *An. gambiae* and *An. coluzzii*, until recently, have been considered as the S and M molecular forms, respectively. Initial identification of these molecular forms was based on specific single nucleotide polymorphism (SNP) and differences of ITS2 sequences [152] [153]. Later studies demonstrated that two forms also differ with respect to larval ecology and habitat segregation [154], gene expression [155], swarming [75], behavior in the presence of predators [156] [157]; and mate recognition [158] [82]. Thus recent literature showing evidence for ecological, behavioral and genome wide differentiation argued the case for the M form to be elevated to species status – named *An. coluzzii* while the S form retained the name *An. gambiae s.s* [44]. However, differences in the percentage of hybrids between incipient species found in areas of sympatry posed questions on their species status [159] [79]. Incomplete reproductive barrier between *An. gambiae* and *An. coluzzii* as they readily mate and lay viable eggs in laboratory conditions, and their recent divergence (~500 MYA) [160] make them a convenient system to study incipient speciation.

Cytogenetic analysis of the chromosomal banding patterns is an established tool for distinguishing species. In addition to fixed chromosomal inversions visible in polytene chromosomes, cytogenetic analysis of mitotic chromosomes demonstrated interspecific differences in sex chromosome heterochromatin between *An. gambiae* and *An. arabiensis* [87]. Additionally, staining with Hoechst enabled identification of these species based on the cytological differences in their respective mitotic karyotype banding patterns [84]. The autosomes in both species reflected a similar fluorescence pattern, with Hoechst staining only the centromere. In contrast, the presence and brightness of heterochromatic blocks differed between the two species. Intraspecific polymorphism in sex chromosome heterochromatin was also observed for both natural and lab populations of *An. gambiae* and *An. arabiensis*. Further, it was suggested that polymorphism in sex chromosome heterochromatin may affect fertility and sexual behavior, but the lack of understanding of heterochromatin function at the time prevented any evolutionary significance of this evident polymorphism. Polymorphism in the amount of heterochromatin of mitotic chromosomes within *An. gambiae* was shown before this species was taxonomically separated into *An. gambiae* and *An. coluzzii.* Therefore, it is unclear if these observed variations in heterochromatin differences can be related to the differences between *An. gambiae* and *An. coluzzii*.

Sequencing of the *An. gambiae* genome provided important information regarding its organization [48]. It has been demonstrated that repetitive DNA component represent a substantial portion in of the *An. gambiae* genome (33%) that is higher than in *Drosophila melanogaster* genome where repeats make up approximately 24% of the genome [161, 162]. The majority of repetitive DNA component in the *An. gambiae* genome is tightly packed in heterochromatin around the centromeres [163]. Difficulty with sequencing of heterochromatin

led to underrepresentation of heterochromatic sequences in the *An. gambiae* sequence assembly [48]. Moreover the repetitive nature of these sequences poses an impediment in mapping them correctly to chromosomes [49]. Subsequent attempts to characterize heterochromatin led to an addition of 16 Mb of heterochromatin to the *An. gambiae* PEST genome. Further improvement was made by predicting 232 novel heterochromatin genes and mapping some genes on the heterochromatin-euchromatin boundary on the polytene chromosome map [164]. However, bioinformatics analysis of so called 'unknown chromosome' or 40 Mb of unmapped sequences in *An. gambiae* PEST genome demonstrated that it has heterochromatic characteristics. Thus significant portion of the *An. gambiae* heterochromatin still remains unassembled and unmapped to the chromosomes. Comparison of *An. gambiae* and *An. coluzzii* genomes demonstrated that sequences corresponding to the mapped portion of the heterochromatin in PEST assembly are the most rapidly evolving. Multiple studies identified regions of high genomic divergence, termed "speciation islands" corresponding to heterochromatin in PEST [77] [78]. Later studies emphasized that the highest genome divergence between these nascent species occurred on the 4 Mb of mapped heterochromatin on the X chromosome [80] [79]. Although *An. gambiae* and *An. coluzzii* genome assemblies are now available, corresponding information of their heterochromatin on a chromosome map is still missing. Likewise, recent sequencing of four species from *An. gambiae* complex also excluded the majority of heterochromatic regions from the genome assembly [165]. Within the *Anophelines*, X chromosome is the fastest evolving [166] [165]. Yet, current bioinformatics approaches alone are insufficient to map the molecular basis of heterochromatin variability within the *An. gambiae* complex on account of high density repeats.

In this study, we examined and evaluated the differences in heterochromatin patterns within X chromosomes of mitotic chromosomes in major vectors *An. gambiae*, *An. coluzzii*, *An. arabiensis*, minor vector *An. merus* and zoophilic non vector *An. quadriannulatus*. Based on chromosome preparations from imaginal discs of early 4th instar larvae from over 50 individuals, we constructed idiograms for each species. Quantitative differences in the number and position of heterochromatin bands between sibling species were described. The position of heterochromatin bands, density and overall length of X chromosome were compared for several lab strains of nascent species *An. gambiae* and *An. coluzzii*. Finally, we mapped the position of satellite DNA repeats and rDNA locus. Here we report qualitative and quantitative differences between molecular organization of heterochromatin and rDNA loci among members of the *An. gambiae* complex and highlight the reorganization of heterochromatin among strains of incipient species *An. gambiae* and *An. coluzzii*.

3.3 Methods

Mosquito strains

Laboratory colonies examined for this study were provided by the BEI Malaria Research and Reference Reagent Resource Center (MR4, Atlanta). These included multiple colonies of *An. gambiae* (Pimperena, Kisumu, Zanu), *An. coluzzii* (Mali, Mopti, SUA 2La), *An. arabiensis* Dongola, *An. quadriannulatus* Sangwe, and *An. merus* MAF. Mosquitoes were reared at 27°C, with 12:12 light:dark cycle and 70% relative humidity. Any possible colony contamination during the experiment was ruled out by verifying all the strains for their respective species using the IMP primers described in Wilkins *et al*. [152]. All the strains showed the expected band size,

An. gambiae strains around 330bp and *An. coluzzii* strains around 460 bp. The species and strains used in this study are presented in Table 1.

Chromosome preparation

Preparations from early 4th instar larvae of lab populations were made from leg and wing imaginal discs as previously described [167]. Larvae were immobilized on ice for 10 minutes. Larvae were dissected in a drop of cold freshly prepared hypotonic solution (0.075M KCl). After decapitating the head, thorax was cut using dissecting scissors (Fine Science tools, USA) followed by removal of gut and fat body. A fresh drop of hypotonic solution was added to the preparation for 10 minutes, followed by fixation in a drop of modified Carnoy's solution (ethanol:glacial acetic acid, 3:1) for 1 minute. Next, a drop of freshly prepared 50% propionic acid was added and the imaginal discs were covered with a 22x22 mm coverslip. After 5 minutes, the preparation was squashed using the flat rubber end of a pencil and dipped in liquid nitrogen until the bubbling stopped. Coverslip was removed using a sharp blade and slides were transferred to cold 50% ethanol stored at -20^{o}C. After 2 hours, slides were serially dehydrated in 70%, 80% and 100% ethanol. Preparations with the highest number of metaphase plates were chosen for *in situ* hybridization. Both female and male larvae were used. At least 5 metaphase plates were measured from each individual larva and 5-7 larvae were used per strain. About 50 X chromosomes were measured and mapped for each strain. Altogether a total of 300 X chromosomes representing over 150 individual mosquitoes were compared for *An. gambiae* and *An. coluzzii*. For sibling species *An. arabiensis, An. quadriannulatus* and *An. merus*, 50 chromosomes from 20 individuals for each species were compared.

Fluorescence *in situ* hybridization (FISH)

Suitable slides with >10 metaphase plates were selected for FISH as described [167]. Primers from previously characterized satellite DNA repeats from *An. gambiae* AgY53A, AgY53B, AgY477and AgY53C were obtained as previously described [51]. Immomix (Bioline, USA) was used to label satellites by incorporating Cy3 and Cy5 fluorescently labeled nucleotides (Enzo Life Sciences, USA) directly in the PCR reaction. Each 25 μl PCR mix consisted of 35-40 ng genomic DNA, 0.3 U Taq polymerase, 1× PCR buffer, 200 μM each of dATP, dCTP, and dGTP, and 65 μM dTTP, and 0.5 μl Cy3-dUTP or 0.5 μl Cy5-dUTP (*Enzo* Life Sciences, *Inc.*, Farmingdale, NY, USA). Thermocycling was performed using ImmoMix ™ (Bioline USA Inc., Taunton, MA, USA) beginning with a 95°C incubation for 10 minutes followed by 35 cycles of 95°C for 30 sec, 52°C for 30 sec, 72°C for 45 sec; 72°C for 5 min, and a final hold at 4°C. Images were counter stained with DAPI-antifade (Life Technologies, Carlsbad, CA, USA) and kept in the dark for at least 2 hours before visualizing.

C_0t DNA preparation

Due to the high repeat content of the X chromosome, the location of centromere is unknown, although previous cytogenetic studies have suggested the X to be acrocentric. Based on DNA reassociation kinetics, C_0t is a method for isolating the highly repetitive parts of the genome. When subjected to high temperatures, DNA separates into single strands which reassociate based on a C_0t curve where C_0 is the initial concentration of single stranded DNA and t is the reannealing time [167]. The more repetitive fragments tend to reassociate quicker than the rest of the genome. As centromere consists of dense repeats, we labelled C_0t 1 DNA (moderate – high repeat fraction) and mapped to chromosome preparations to identify the exact location of *An. gambiae* centromere. Genomic DNA was isolated from 500g non bloodfed adult *An. coluzzii* Mopti mosquitoes using Qiagen Blood and Cell culture DNA Maxi kit (Qiagen Science, USA).

C_0t DNA was extracted as described earlier [168]. Isolated repetitive fractions were precipitated with isopropanol and estimated to be corresponding to C_0t1 repeat fraction. It was then labelled with Nick Translation (Thermo Fisher Scientific, USA) and FISH was performed as previously described.

Image acquisition and chromosome measurements

Slide preparations were viewed with Olympus BX61 fluorescent light microscope using BioView software (BioView Inc., Billerica, MA, USA) at 1000x magnification. Individual channels of each image were exported. A custom MATLAB script (MATLAB 2010) was written and used to measure the position of stDNA AgY53B, AgY477 and proximal DAPI band in each X chromosome corresponding to respective channel. The results were compared within and between species. The program was written in MATLAB (MATLAB 2010). The program was automated to measure the position of each stDNA and DAPI peak along the length of X chromosome. The output provided the maximum likelihood of a combination of probe positions. For density and length measurement, images were inverted in Adobe Photoshop CS6 and measured using the ruler tool at 1000X magnification. Statistical analysis was performed using JMP 10 software.

3.4 Results –

Heterochromatin organization in sibling species of the *An. gambiae* complex

For the continence of heterochromatin comparison, we first developed idiograms for X chromosomes and autosomes of *An. gambiae*, *An. coluzzii*, *An. arabiensis*, *An. merus*, and *An. quadriannulatus* (Figure 1). We utilized mid-metaphase chromosomes obtained from imaginal discs of the 4th instar larvae. Chromosomes at this stage provide reproducible heterochromatin

patterns verified across multiple individuals and strains. We used DAPI, a counter-stain which preferentially stains AT-rich heterochromatin, to augment the natural banding patterns resulting from heterochromatin variation between and within species. Autosomes in all the species tested were of similar lengths with heterochromatin only in the centromere. In contrast, the X chromosome length, heterochromatin location and amount varied considerably across sibling species. Differences in the length of X chromosomes were clearly visible between all reproductively isolated species. The polytene complement of X chromosome in these species are comparable in length; hence the difference in mitotic X chromosomes can be attributed to the variation in heterochromatin content. The number and position of heterochromatin bands differ between species as well. While *An. gambiae* and *An. coluzzii* possess two clear heterochromatin bands, *An. arabiensis* has a single band on the X chromosome. In contrast, *An. quadriannulatus* and *A. merus* contain multiple heterochromatin dense bands (Fig 1b). *An. gambiae* and *An. coluzzii* possess a large heterochromatin block on the end proximal to the centromere, we termed this as the proximal band. This is followed by a dull fluorescent area to which 18S rDNA locus was later mapped. A second small distal band is present immediately next to the rDNA locus, marking the beginning of the current *An. gambiae* PEST genome assembly. A single heterochromatin block is visible in *An. arabiensis*; although the location of heterochromatin is more distal compared to *An. gambiae*. Our results for *An. gambiae* and *An. arabiensis* corroborated with those for the natural populations of these species in a previous study [84]. The length of the X chromosomes measured were similar across freshwater vectors *An. gambiae*, *An. coluzzii* and *An. arabiensis*. In contrast, *An. quadriannulatus*, the zoophilic non vector of the *An. gambiae* complex, has X chromosomes longer than the freshwater species and possesses two distinct heterochromatin bands larger than others. However, the longest X chromosome was

found to be present in *An. merus*, the marshy saltwater resident of the *An. gambiae* complex. *An. merus* X contains three distinct heterochromatin blocks making it comparable in length to the autosomes, a feature not present in any other species of this complex.

X chromosomes differ in the integrated density and distance between heterochromatin bands

Mitotic X chromosomes of incipient species *An. gambiae* and *An. coluzzii* possess two distinct heterochromatin bands. Based on the distance from the centromere we named these bands as proximal and distal. The density and distance between the bands were measured for three different laboratory strains for each species – Pimperena, Kisumu, Zanu (*An. gambiae*) and Sua2La, Mali, Mopti (*An. coluzzii).* Statistical analysis revealed that while the overall length of the mitotic X chromosome was not significantly different between the two species, distance between proximal and distal bands within *An. gambiae* (p=0.0589) differs from *An. coluzzii* (p=0.0001*) (Fig 2a). Density of the proximal (D1) and distal (D2) bands were also measured across strains for both the species, with D1 differing significantly between the species (p= 0.0054*) suggesting a higher level of compaction of repeats in *An. coluzzii*. Intraspecific comparison of strains revealed less differentiation within *An. gambiae* for density (p= 0.0645) while *An. coluzzii* strains were significantly different from each other (p = 0.0001*) (Fig 2b). Based on the DAPI staining, length and density comparison between the laboratory populations, our results suggest that *An. coluzzii* is potentially undergoing further differentiation into sub-forms.

Interspecific and intraspecific variation in rDNA amount occurs within the *An. gambiae* complex

Ribosomal DNA (rDNA) sequences are highly conserved between close relatives and have been utilized for tracing evolutionary history in natural populations of several taxa [67] [169]. Within *Anophelines*, 18S rDNA locus constitutes an important part of the heterochromatin along with stDNA repeats. It is also conserved between *An. gambiae* and outgroup species *An. stephensi* [166]. We mapped this locus in all the sibling species tested. Within all strains of *An. gambiae* and *An. coluzzii*, the rDNA locus mapped between the proximal and distal heterochromatin bands (Fig 3). Interestingly, variation in size of rDNA locus was noted between multiple laboratory strains of *An. gambiae* and *An. coluzzii* as well. Additionally, F1 hybrid females resulting from *An. gambiae* X *An. coluzzii* crosses were analyzed for heterochromatin content and stDNA position mapping. Qualitative differences were clearly observed in X chromosome heterochromatin pattern and rDNA amount within the hybrids (Fig 7). Variation in 18S rDNA copy number could be a probable source of X chromosome length differences within the species. The fact that it is present in cis-position to X heterochromatin also suggests a possible role of heterochromatin organization in the variability between species. Whether or not this leads to dosage compensation between sexes or affects the speciation process in other manner is yet to be determined.

C_0t DNA analysis does not help in mapping centromere on X chromosome

The heterochromatic nature of one half of the X chromosome have prevented cytogeneticists from identifying the exact location of the centromere in *An. gambiae*. C_0t analysis is the process of renaturation of single stranded DNA to its complimentary sequence. The analysis is based on the observation that the more repetitive a sequence, lesser will be the time required for the

sequences to anneal following denaturation. To this end, we tried utilizing the C_0t DNA method to identify the densest parts of the X chromosome, presumably the location of the centromere. DNA fractions with C_0t values equal to 10^{-4}-10^{-1} or 10°-10^2 are considered as highly and moderately repetitive, respectively. In order to achieve that, we extracted C_0t1 fraction consisting of moderate and highly repetitive elements, fluorescently labelled and mapped it to chromosome preparations. C_0t1 DNA represents the highest repetitive analysis. As the centromere comprises of highly repetitive fractions of DNA, we expected C_0t1 to map to narrow regions correlating to centromeres in respective chromosomes. While this proved to be the case for autosomes, in X chromosomes C_0t1 DNA probe localized to a wide region including the proximal band and overlapping with some region of rDNA locus (Fig 4). Thus C_0t analysis is not useful to determine the exact location of centromere on X chromosome in *An. gambiae*.

Variability in stDNA locations on mitotic X chromosomes in *An. gambiae* and *An. coluzzii*

Previously published as Y enriched *An. gambiae* stDNA repeat primers were obtained [52]. However, upon performing FISH on mitotic chromosomes, the probes hybridized to both Y as well as the proximal band of pericentric heterochromatin in X chromosomes. Interestingly, this is also the region of pericentric heterochromatin on the X chromosome implicated in speciation (citation), warranting further exploration of the stDNA content here. Incongruence with the previous study may be accounted by the difference in FISH stringency and the lack of use of mitotic chromosomes which are suitable to map stDNA repeats compared to polytene chromosomes. The observed pattern of stDNA repeats in hybridization was different for AgY53B and Ag477 in *An. gambiae*. Further, FISH results suggested that stDNA AgY53A, AGY53B and AgY477 mapped to both the incipient species (Fig 5a,b). Further, FISH results suggested that AgY53B and AgY477 mapped to different locations relative to the proximal

DAPI block between *An. gambiae* and *An. coluzzii*. Due to the sequence similarity between X enriched AgY367 [51] and AgY477, it is not technically possible to differentiate between the two satellites by FISH, hence we excluded AgY367 from this study. Ag53C, the reference marker corroborated with the previous published result and mapped to the autosomal centromere in mitotic chromosomes (Fig 5c).

Mapping AgY53B and AgY477 in *An. gambiae* and *An. coluzzii*

To better resolve the stDNA position on the X chromosome, we compared these repeats between multiple laboratory strains of *An. gambiae* (Pimperena, Kisumu and Zanu) and *An. coluzzii* (Mopti, Sua2La and Mali). Using a custom MATLAB script, the position of stDNA repeats with respect to the proximal DAPI band was mapped and the results were compared across strains. Both AgY53B and AgY477 map extremely close to each other, rendering it technically unfeasible to distinguish between the signals when mapping their distance from the centromere. For simplicity, we combined the signals as a single unit and mapped them across all the strains. The MATLAB program consisted of two user guided steps followed by a third step of automated analysis. In the first step, sex chromosomes were identified by the user in the imaged mitotic slide preparations from *An. gambiae* and *An. coluzzii* strains. In the second step, the boundary of each identified sex chromosome was traced by a trained user. In an automated step, the pattern of heterochromatin and stDNA fluorescence was averaged longitudinally along the user defined boundary. Program output displayed fluorescent intensity graphically from chromosome centromere to telomere. Peak fluorescent intensity was used to automatically infer the order of the heterochromatin and stDNA probes. In some images the distance between fluorescence peaks was less than the microscope resolution. This led to a few cases of ambiguity in the order of the heterochromatin and stDNA markers. Despite this limitation, the order of peaks could often be

recovered from the aggregate of multiple images averaged for multiple individuals of each strain. Our mapping results revealed that strains differed in the pattern of stDNA and DAPI peak positions on the X chromosome. *An. gambiae* Pimperena (Fig 6 a) and *An. coluzzii* Mali depicted the same pattern while *An. coluzzii* Mopti (Fig 6 b), *An. coluzzii* SUA and *An. gambiae* Zanu clustered together and the DAPI peak was followed by the stDNA locus. Interestingly, *An. gambiae* Kisumu showed an almost 1:1 ratio of the stDNA: DAPI patterns for both *An. gambiae* and *An. coluzzii*. Previous studies have documented how Kisumu can behave like either *An. gambiae* or *An. coluzzii*, although on the basis of molecular fixed differences it is classified as *An. gambiae*. *An. gambiae* Zanu produced a stDNA orientation similar to *An. coluzzii*. Subsequently, we tested preparations from both Zanu males and females, and reaffirmed the stDNA orientation to be similar to *An. coluzzii*. MATLAB was also used to generate a cluster analysis taking into account the stDNA peak width and distance from the DAPI proximal band. For comparison, we included data from *An. arabiensis* here. Our results show a clustering of An. gambiae Pimperena, *An. coluzzii* Mali and *An. arabiensis* in one group while *An. coluzzii* Mopti, *An. coluzzii* SUA and *An. gambiae* Zanu grouped together. Consistent with our stDNA peak mapping data, Kisumu segregated in the middle(6c). When strains were grouped according to the stDNA and DAPI bands, two clear clusters emerged with Kisumu in the middle (6d). Thus our data shows the dynamic nature of heterochromatin on a molecular level in important malaria vectors undergoing speciation and suggests potential inversions within heterochromatin between the two nascent species.

Heterochromatin variation within the *An. gambiae* complex

The length of X chromosome and autosomes was measured in addition to obtaining mapping information for stDNA repeats for *An. gambiae, An. coluzzii, An. arabiensis, An.*

quadriannulatus and *An. merus*. Along with the position of pericentromeric heterochromatin bands and stDNA probes and rDNA locus, these were combined into a mitotic idiogram map for sibling species of the *An. gambiae* complex. The euchromatin/heterochromatin boundary was mapped based on the current assembly. FISH with stDNA AgY53B revealed a species specific pattern of this satellite between the sibling species, with multiple locations corresponding to heterochromatin bands in *An. merus* but mapping to the pericentric heterochromatin in other sibling species. In contrast, AgY477 mapped to the fresh water vectors but not to *An. quadriannulatus* or *An. merus*. These features were placed on respective idiograms for comparison of molecular features within heterochromatin (Fig 8). The idiograms in this study enhance our understanding of reorganization of repeats during evolution in chromosome X within the *An. gambiae* complex, and particularly between *An. gambiae* and *An. coluzzii*.

3.5 Discussion

Mosquitoes with their small number of diploid chromosomes, 2n = 6 [170], and their role as medically important disease vectors present a convenient model to investigate the molecular organization of their genomes and chromosomal evolution among the insect species [149]. In this study, we compared the mitotic karyotype of several sibling species in the *An. gambiae* complex and found qualitative as well as quantitative differences in the X chromosome pericentric heterochromatin. Variation in the position and number of heterochromatin blocks led to substantial differences in the X mitotic karyotype between the sibling species. Previously, inter and intra-specific differences between X chromosome heterochromatin in natural and lab populations of *An. gambiae* and *An. arabiensis* were described. Polymorphism in sex chromosome heterochromatin was also found in natural populations by Hoechst staining [84].

However, these studies were restricted to a few populations with no details regarding the molecular basis of this heterochromatin variation. Further, the characterization of heterochromatin organization and content between *An. gambiae* and recent species *An. coluzzii* is also lacking. In order to create a platform for heterochromatin comparison between species, we first developed preliminary idiograms representing the X chromosome for five members of the *An. gambiae* complex.

Recently formed species *An. coluzzii* is an offshoot of *An. gambiae s. l.* The region of highest divergence between the two incipient species is the 4 Mb pericentric heterochromatin, which is not represented in the current assembly as heterochromatin is unassembled. In this study we focused on the unassembled part of X chromosome and compared the mitotic X chromosome of *An. gambiae* and *An. coluzzii* for the heterochromatin density, distance between the heterochromatin proximal and distal bands and position of stDNA repeats in addition to the 18S rDNA locus. Based on cytological mapping of specific probes, intraspecific polymorphism in position of stDNA repeats AgY53B, AgY477 and DAPI within *An. gambiae* and *An. coluzzii* was revealed. We also mapped *An. gambiae* Kisumu, which shows heterozygosity in the stDNA and DAPI pattern within the population. Our results suggest a shifted amplification or putative inversion in the X chromosome short arm between some of the strains of *An. gambiae* and *An. coluzzii*, implicating satellite DNA repeats in the process of prezygotic speciation. Using an average of about 150 X chromosomes from both male and female larvae imaginal discs for each species, we developed idiograms mapping the X chromosome heterochromatin reorganization between *An. gambiae* and *An. coluzzii*.

A common emerging theme from several of the recent studies performed in *Drosophila* suggests that heterochromatin plays an important role in hybrid lethality and hybrid

incompatibility (HI) [61] [59] [62] [64] [60]. This is especially evident in case of pericentric heterochromatin present in *Drosophila* X chromosome. Rapid evolutionary forces lead to a great variation in the copy number and types of heterochromatin repeats between closely related species in nature. In *Drosophila*, several studies have highlighted the different stDNA content and localization of heterochromatin repeats between reproductively isolated sibling species [57, 60]. As repetitive DNA elements replicate mechanisms such as unequal cross-over, rolling circle replication, segmental duplication [171, 172] they are also a potential source of genome evolution. Hybrid lethality in F1 females obtained from crosses between closely related species *D. simulans* or *D. mauritiana* and *D. melanogaster* was investigated [64]. Only F1 hybrid males produced from Xmau /Y mel cross were viable, suggesting a causative dominant X-linked factor in *D. melanogaster* incompatible with the recessive factor in D. mau. Components of heterochromatin, particularly satellite DNA (stDNA) have been implicated in disruption of chromosome pairing, abnormal heterochromatin packaging and selfish postmeiotic drive [57]. Among other insects, evolutionary relationship between various natural populations was inferred using the location and number of rDNA loci (important componenets of heterochromatin) in the Chagas disease vector *Trypanosoma infestans* [67]. Reproductive barrier between *An. gambiae* and *An. coluzzii* is incomplete as they readily mate and lay viable eggs in lab conditions [75]. Incomplete reproductive isolation along with ecological and molecular divergence makes them an excellent model to study heterochromatin variation during prezygotic evolution in an important disease vector. A new *Anopheles* species in Africa would further compound the prevalent malaria problem with different biting behavior, host – recognition mechanisms and variation in ability of malaria transmittance. Hence knowledge about the genome of these incipient species is extremely crucial in combating malaria.

Since the first genome of *An. gambiae* was released, efforts were undertaken to improve the heterochromatin assembly [49] [164]. However, the repeat content of heterochromatin rendered it unsuitable for mapping. With the current advances in bioinformatics, we have information previously unavailable about the dense repeat regions [52], although successfully mapping them has been a challenge till now. Here we utilized metaphase mitotic chromosomes for heterochromatin comparison between species and mapped molecular markers. We showed that mitotic chromosomes may be used to assemble and compare heterochromatin between species in the *An. gambiae* complex.

A custom MATLAB script helped us identify the correct position as well as sequence of satellite repeats with much higher precision than can be achieved with human eye. User guidance was introduced into our MATLAB (MATLAB 2010) analysis pipeline; however, this approach does not compromise objectivity. Rather, user guidance was deliberately introduced to eliminate the pitfalls of full automation. In particular, user guidance was critical for accurate identification of sex chromosomes in the imaged mitotic slide preparations. In addition, the user defined boundary of each mitotic chromosome has no bearing on the width of heterochromatin fluorescence, rDNA fluorescence, or the distance between the peak fluorescence intensity of the signals respectively. Longitudinal averaging of the fluorescence intensity during programs automated analysis was robust to the boundary of each chromosome traced by the user. The advantages of our analysis are twofold. First, fluorescent signals in each image possess a degree of overlap which hides from observation the order of fluorescent peaks from centromere to telomere; however, the order of peaks is easily detected using our automated analysis (figure 6a,6b). Second, observation alone is insufficient to determine the distance between fluorescent peaks which are generally separated by fewer than 10 pixels. Determining the distance between

nearby peaks is made tractable in our pipeline; however, even this computational approach remains limited by microscope resolution. The images of the mitotic sex chromosome were captured using 1000x magnification: neighboring pixels at this resolution represent a distance of 0.05µ. Thus, peaks spaced closer than 0.05µ were prone to falsely inverted order (see also results). Nonetheless, peak orders were generally resolvable in the aggregate when the results from multiple images were combined.

In this study we used the laboratory colonies maintained under standard rearing conditions. However, we expect to see a higher level of polymorphism in satellite DNA content between natural populations subjected to varying environmental pressures. Further investigation of natural populations for the stDNA and DAPI pattern on the X chromosomes to determine if they correlate with specific environment adaptations would aid in our understanding of heterochromatin. At the same time, other satellite DNA or additional repeats that may contribute to speciation or ecological differences between species may have escaped our notice in the current study.

3.6 Conclusion

As polytene chromosomes are large and well developed in *An. gambiae*, they are often preferred over mitotic chromosomes for mapping unique genes or intercalary heterochromatin [164, 173]. However, the repeat rich nature of one half of X chromosome makes it virtually negligible in polytene preparations from salivary glands or ovarian nurse cells. This study utilized mitotic mid metaphase chromosomes obtained from the imaginal discs of early 4^{th} instar larvae to investigate the heterochromatin rich arm of X chromosome. We found qualitative and quantitative differences in the X pericentric heterochromatin between sibling species of the NA.

gambiae complex. Further, we described qualitative differences in the molecular organization of heterochromatin between sibling species *An. gambiae* and *An. coluzzii* which are currently undergoing sympatric diversification. Finally, we mapped rDNA locus for several sibling species. The information obtained was presented in idiograms aiding in better comparison of the under-represented heterochromatin in the current *An. gambiae* assembly. Idiograms based on mitotic chromosomes have been previously developed for other prominent mosquito vectors *Aedes aegypti* [168]and *Culex quinquefasciatus* [174]. The idiograms in this study will enhance the understanding of repeat rich heterochromatic region between two prominent malaria vectors. Our findings will facilitate an increased resolution into the region currently underrepresented in the assembly and enhance the efforts in discovering the candidate heterochromatin genes or repeats potentially driving the speciation between nascent species in Sub-Saharan Africa. To the best of our knowledge, this is the first study characterizing the qualitative differences in the position of repeats present in pericentric heterochromatin region of X chromosome between two major malaria mosquito species. This study was aimed at characterizing the parts of genome which are not easily accessible through only polytene chromosomes or bioinformatics. Hence we combined cytogenetics with bioinformatic tools like MATLAB to investigate the reorganization of repetitive elements in species that have not achieved postzygotic isolation. Heterochromatin with its characteristic low gene density has long been the black box of malaria mosquito evolution chapter, and this study will contribute to elucidating some of the box contents.

Table 1. Strains, isolation place and genome data used in the study of heterochromatin of the *An. gambiae* complex.

Strain	Species	MR4 ID	Isolation place	Isolation date	Depositors to MR4	Genotype/ phenotype	Genome data
DONG2Ra 2Rb3R	*Anopheles arabiensis*	MRA-1235	Dongola, Sudan	2009	E Dotson >> Nora Besansky	2Ra/a, 2Rb/b and 3R	Neafsey *et al* 2015
Mali-NIH	*Anopheles coluzzii*	MRA-860	Niono, Mali	2005	Tovi Lehmann >> Nora Besansky	M rDNA form, 2Rbc/bc, 2La/a	Lawniczak MK *et al* 2010
MOPTI	*Anopheles coluzzii*	MRA-763	N'Gabacoro Droit, near Bamako, Mali	2003	Greg Lanzaro	2Ru/+	Aboagye-Antwi *et al* 2015
SUA2La	*Anopheles coluzzii*	MRA-765	Suakoko, Liberia	1987	Allesandra della Torre	Xag, 2R+, 2La, 3R+, 3L+. M rDNA form	Sharakhov et al 2006
KISUMU1	*Anopheles gambiae*	MRA-762	Kisumu, Kenya	1975	G. Davidson >> Vincent Corbel	Permethrin susceptible	Gste2, Vgsc in Mitcheli et al 2013, Aboagye-Antwi *et al* 2015
Pimperena	*Anopheles gambiae*	MRA-861	Pimperena region, Mali	2005	Nora Besansky	S rDNA form, 2Rb/b, 2La/+	Lawniczak MK *et al* 2010
ZANU	*Anopheles gambiae*	MRA-594	Zanzibar, Tanzania	1982	CF Curtis >> H Ranson, FH Collins	DDT Resistance conferred by elevated GST activity	Gste2, Vgsc in Mitcheli *et al* 2013
MAF	*Anopheles merus*	MRA-1156	Kruger National Park, South Africa	1991	Maureen Coetzee	2Rop	Neafsey *et al* 2015
SANGWE	*Anopheles quadriannulatus*	MRA-1155	Sangwe, Zimbabwe	1998	Richard Hunt >> Willem Takken	wild-type, 2L+, Xf/+	Neafsey *et al* 2015

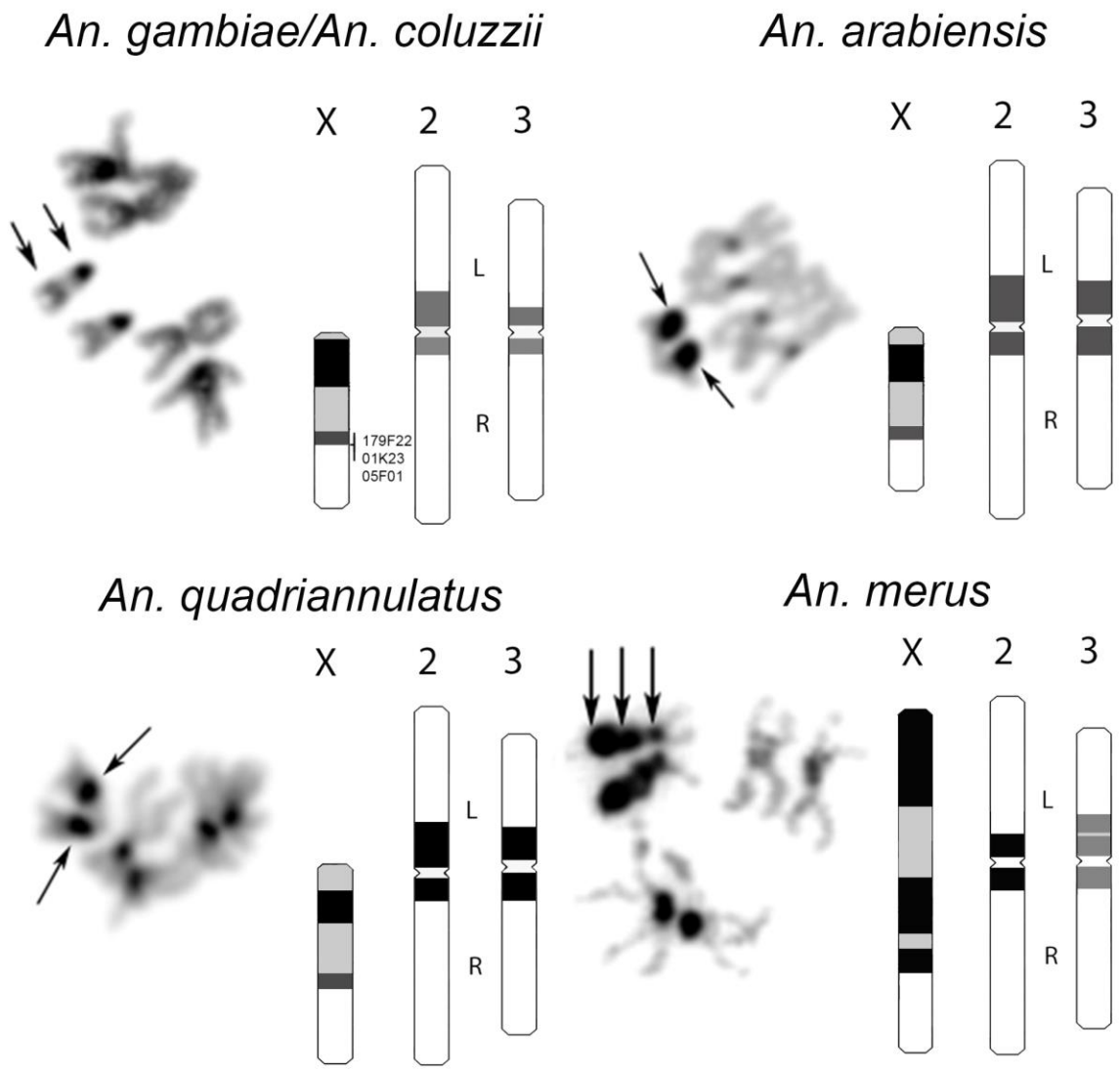

Fig 1. Grayscale images of mitotic karyotype from sibling species. A. *An. gambiae*/*An. coluzzii*; B. *An. arabiensis;* C. *An. quadriannulatus;* D. *An. merus*. Arrows indicate heterochromatin blocks. 179F22 indicates the start of the current assembly in *An. gambiae* PEST. A-C are mid metaphase stage; D is early metaphase. Black- condensed heterochromatin, Dark gray- less condensed heterochromatin, light gray – areas lighter than euchromatin, white - euchromatin.

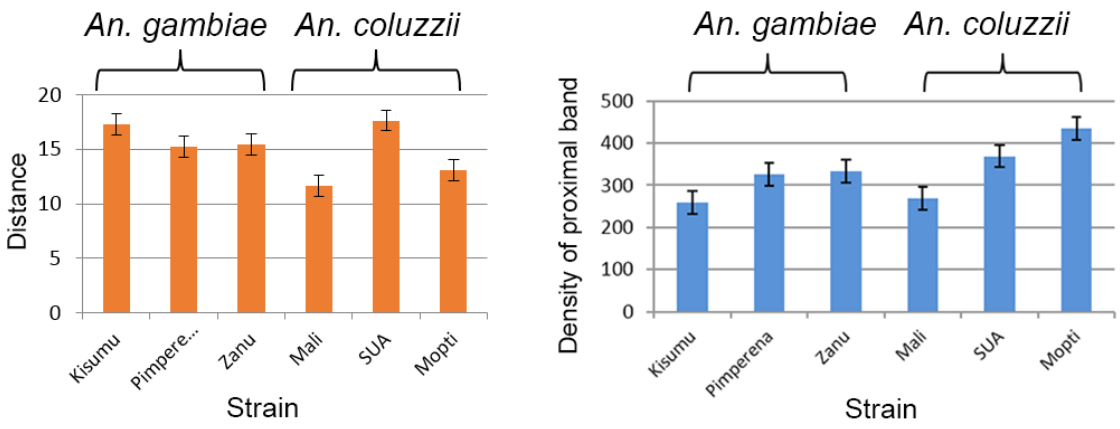

Fig 2. Intraspecific comparison of the distance between proximal and distal heterochromatin bands (2a) and integrated densities of the proximal DAPI band (2b) among laboratory strains of *An. gambiae* and *An. coluzzii*. When compared within species, *An. gambiae* showed no significant difference in distance between the heterochromatin blocks (p= 0.0589) whereas *An. coluzzii* strains differed significantly (p= 0.0001*). Density of the heterochromatin blocks within the species revealed similar results with *An. gambiae* (p= 0.0645) and *An. coluzzii* (p= 0.0001*). Asterisk denotes significant difference. Distance and density are in pixels.

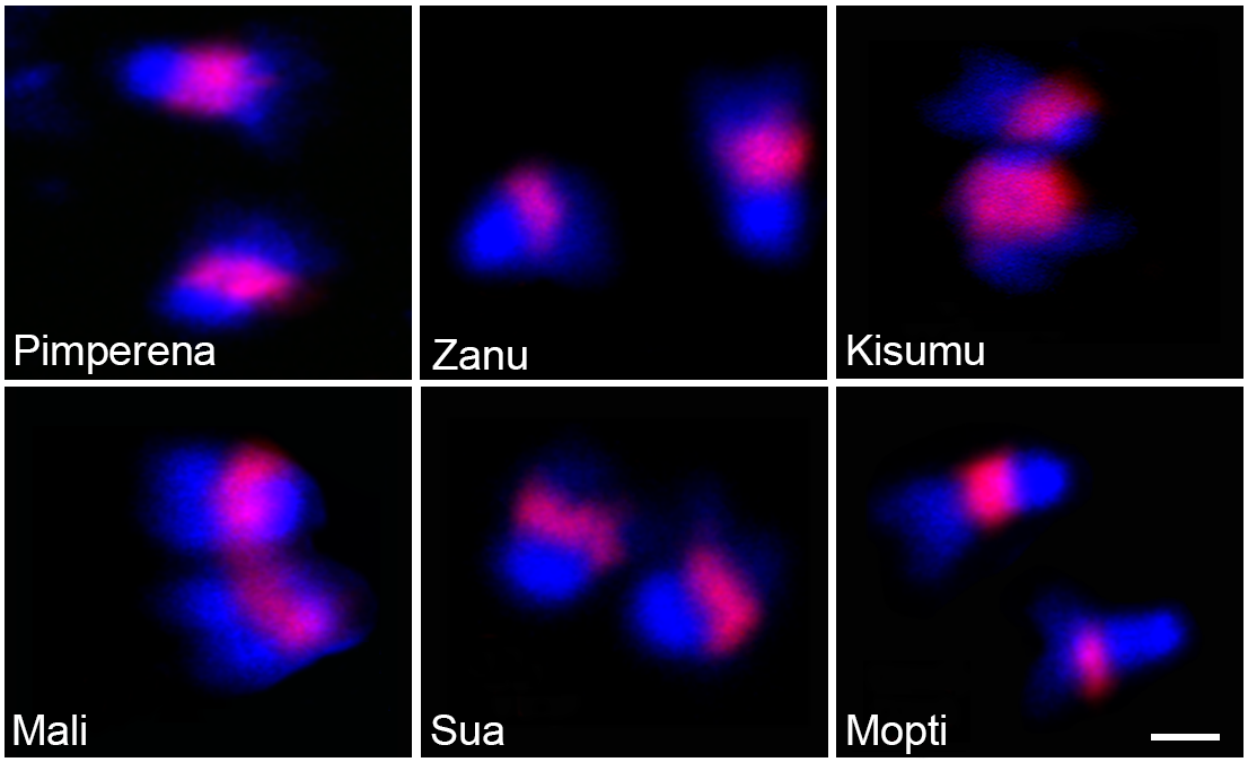

Fig 3. Multicolor FISH showing rDNA locus maps between the proximal and distal band in *An. gambiae and An. coluzzii* X chromosomes. Top – *An. gambiae* a) Pimperena b*)* Zanu c) Kisumu; Bottom - *An. coluzzii* - d) SUA e) Mali f) Mopti. Polymorphism in rDNA locus size can be seen in Kisumu and Mopti. Space bar – 1 uM. Red- 18S rDNA, Blue – DAPI.

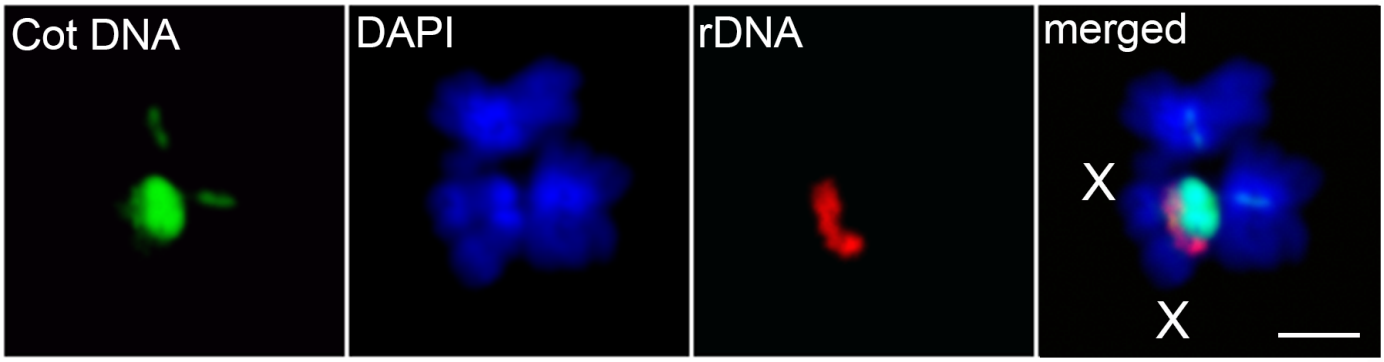

Fig 4. Multicolor FISH with C0t1 DNA and rDNA on *An. gambiae.* Image is counter stained with DAPI. Scale bar is 2 µM.

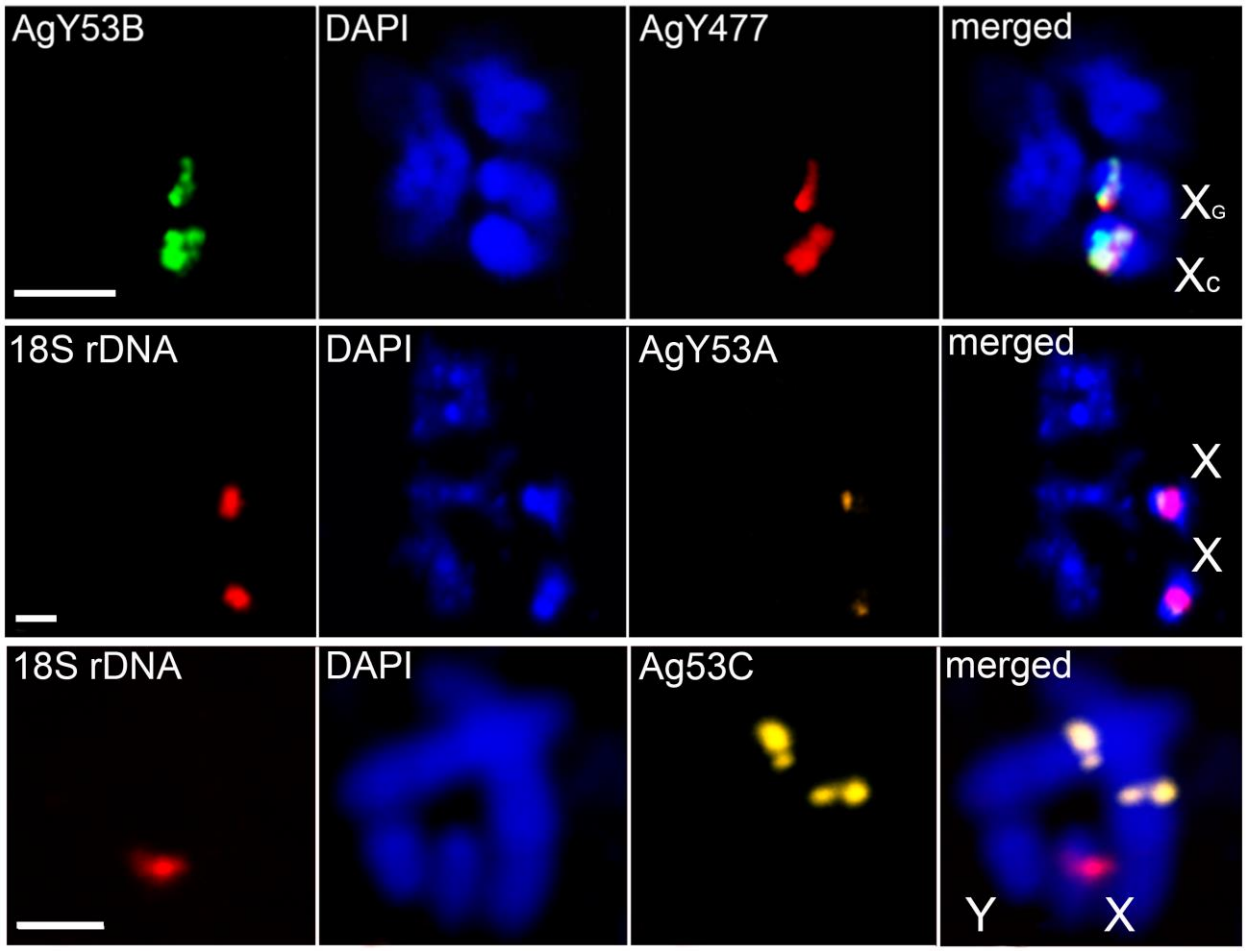

Fig 5. Multicolor FISH showing individual channel signals for AgY53B, AgY477, AgY53A, Ag53C and 18S rDNA on the X chromosome of *An. gambiae* Pimperena and *An. coluzzii* Mopti. Top panel – F1 female hybrid from *An. gambiae* Pimperena X *An. coluzzii* Mopti cross. X_C and X_G represent *An. coluzzii and An. gambiae* respectively. Middle panel - *An. gambiae* Pimperena. Bottom panel - *An. coluzzii* Mopti. Scale bar - 2 µM.

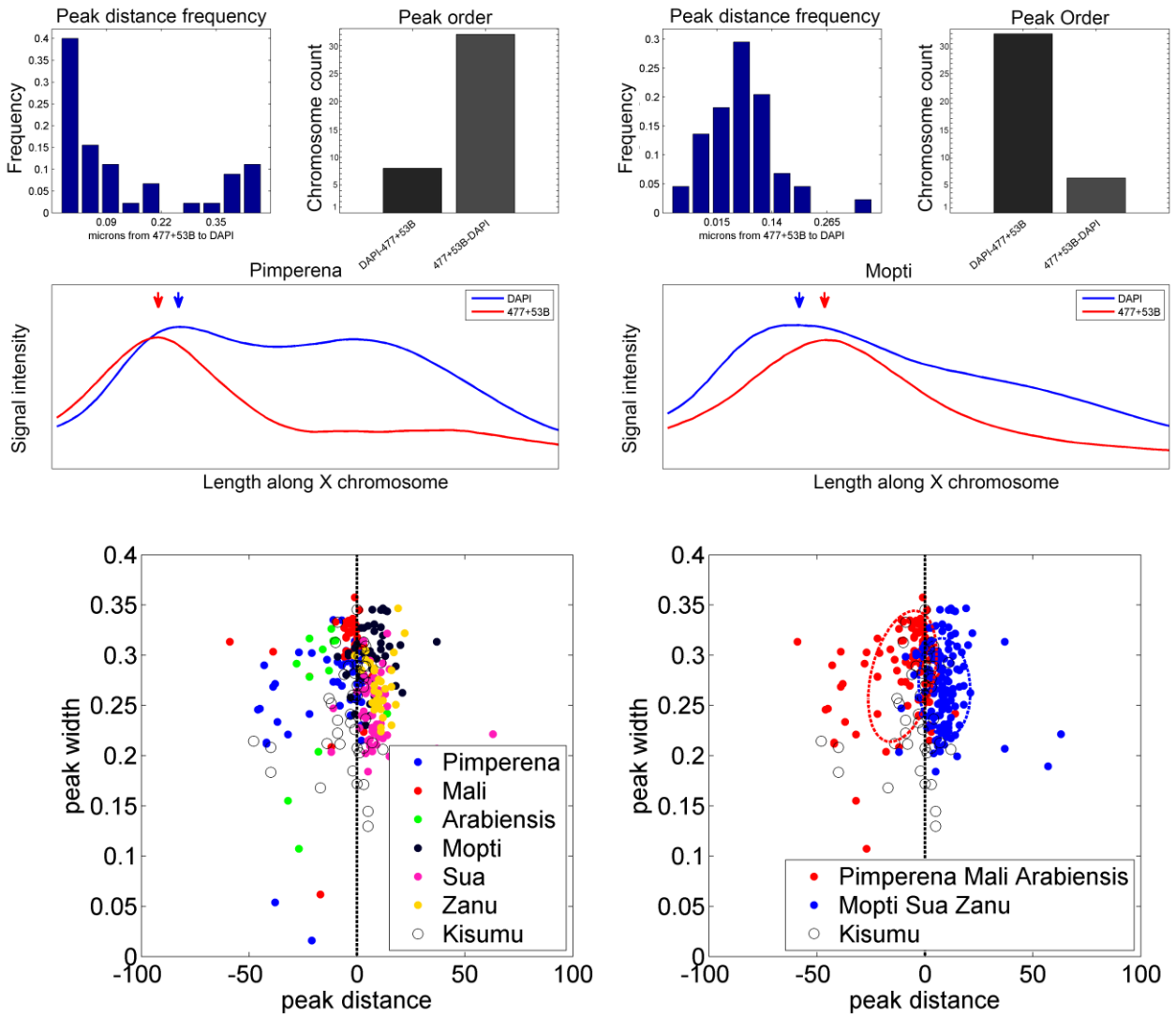

Fig 6. Using MATLAB to map the location of stDNA AgY53B+AgY477 in *An. gambiae* Pimperena and *An. coluzzii*. Mopti. MATLAB mapped the location of stDNA AgY53B+AgY477 to be present before the DAPI peak in the proximal heterochromatin band in *An. gambiae* Pimperena (6a) while this orientation was reversed in some strains eg. *An. coluzzii* Mopti (6b. Clustering analysis grouped Pimperena, Mali and Arabiensis together while Mopti, SUA and Zanu were in another group with some observations in the intersection (6d).

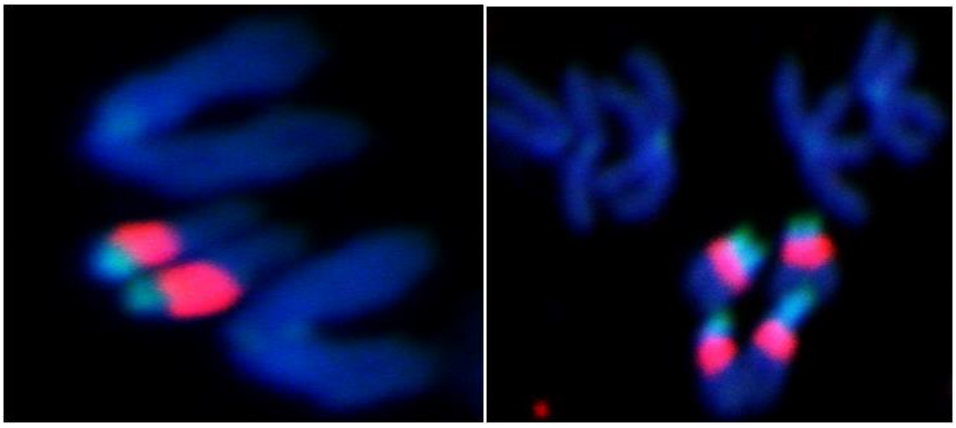

Fig 7. Multicolor FISH showing rDNA variation between homeologus X chromosomes from *An. gambiae* X *An. coluzzii* F1 hybrid females. 7a. Prometaphase, 7b part of mid metaphase showing chromosomes from two nuclei. Red- rDNA, green- YOYO-1, blue –DAPI.

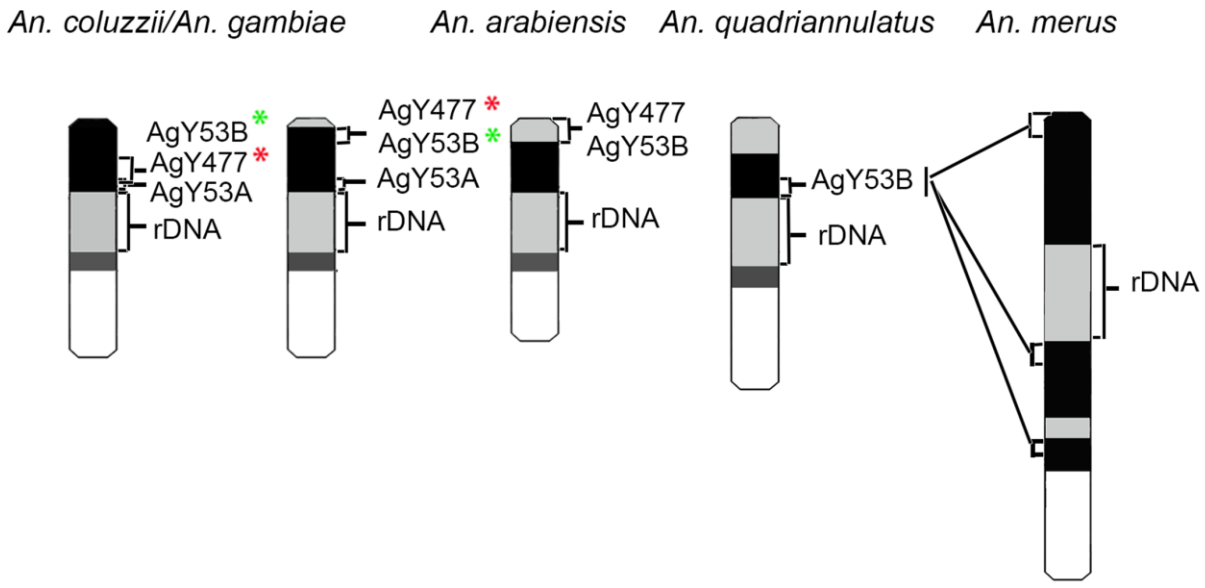

Fig 8. Comparative idiograms for X chromosome pericentric heterochromatin comparison among sibling species of the *An. gambiae* complex based on mitotic chromosomes. *An. gambiae* Pimperena *and An. coluzzii* Mopti are used here. The positon of stDNA AgY477 and AgY53B is indicated by colored * to highlight the differences in their positon between species. Black- condensed heterochromatin, Dark gray- less condensed heterochromatin, light gray – areas less fluorescent than euchromatin, White - euchromatin.

Chapter 4 – Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes

Andrew B. Hall[a,1], Philippos-Aris Papathanos[b,c,1], Atashi Sharma[d,1], Changde Cheng[e,f,1,2], Omar S. Akbari[g], Lauren Assour[h], Nicholas H. Bergman[i], Alessia Cagnetti[b], Andrea Crisanti[b,c], Tania Dottorini[c], Elisa Fiorentini[c], Roberto Galizi[c], Jonathan Hnath[i], Xiaofang Jiang[a], Sergey Koren[j], Tony Nolan[c], Diana Radune[i], Maria V. Sharakhova[d,k], Aaron Steele[h], Vladimir A. Timoshevskiy[d], Nikolai Windbichler[c], Simo V. Zhang[l], Matthew W. Hahn[l,m], Adam M. Phillippy[j], Scott J. Emrich[e,h], Igor V. Sharakhov[a,d,k,3], Zhijian Tu[a,n,3], Nora J. Besansky[e,f,3]

[a]The Interdisciplinary PhD Program in Genetics, Bioinformatics, and Computational Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA; [b]Section of Genomics and Genetics, Department of Experimental Medicine, University of Perugia, 06132 Perugia, Italy; [c]Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom; [d]Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA; [e]Eck Institute for Global Health, University of Notre Dame, Notre Dame, Indiana 46556, USA; [f]Department of Biological Sciences, University of Notre Dame, Notre Dame, Indiana 46556, USA; [g]Department of Entomology, University of California, Riverside Center for Disease Vector Research, Institute for Integrative Genome Biology, University of California, Riverside, CA, 92521, USA; [h]Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, Indiana 46556, USA; [i]National Biodefense Analysis and Countermeasures Center, Frederick, MD 21702, USA; [j]Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda,

Maryland 20892, USA; [k]Laboratory of Evolutionary Cytogenetics, Tomsk State University, Tomsk 634050, Russia; [l]School of Informatics and Computing, Indiana University, Bloomington, Indiana 47405, USA; [m]Department of Biology, Indiana University, Bloomington, Indiana 47405, USA; [n]Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061, USA.

[1]These authors contributed equally to this work.

[2]Present address: Department of Integrative Biology, University of Texas, Austin, Texas 78712, USA

[3]Corresponding author. E-mail: Igor@vt.edu (I.V.S.); Jaketu@vt.edu (Z.T.); nbesansk@nd.edu (N.J.B)

Short title: The Y chromosome of *Anopheles*

Key words: Anopheles gambiae, PacBio, RNA-Seq, tandem repetitive DNA, Y-chromosome

4.1 Abstract

Y chromosomes control essential male functions in many species, including sex determination and fertility. However, due to obstacles posed by repeat-rich heterochromatin, knowledge of Y chromosome sequences is limited to a handful of model organisms, constraining our understanding of Y biology across the tree of life. Here, we leverage long single-molecule sequencing to determine the content and structure of the non-recombining Y (NRY) chromosome of the primary African malaria mosquito, *Anopheles gambiae*. We find that the *An. gambiae* Y consists almost entirely of a few massively amplified, tandemly arrayed repeats, some of which can recombine with similar repeats on the X chromosome. Sex-specific genome re-sequencing in a recent species radiation, the *An. gambiae* complex, revealed rapid sequence turnover within *An. gambiae* and among species. Exploiting 52 sex-specific *An. gambiae* RNA-Seq datasets representing all developmental stages, we identified a small repertoire of Y-linked genes that lack X gametologs and are not Y-linked in any other species except *An. gambiae*—with the notable exception of *YG2*, a candidate male-determining gene. *YG2* is the only gene conserved and exclusive to the Y in all species examined, yet sequence similarity to *YG2* is not detectable in the genome of a more distant mosquito relative, suggesting rapid evolution of Y chromosome genes in this highly dynamic genus of malaria vectors. The extensive characterization of the *An. gambiae* Y provides a long-awaited foundation for studying male mosquito biology, and will inform novel mosquito control strategies based on the manipulation of Y chromosomes.

4.2 Significance Statement

Interest in male mosquitoes has been motivated by the potential to develop novel vector control strategies, exploiting the fact that males neither feed on blood nor transmit diseases such as malaria. However, genetic studies of male *Anopheles* mosquitoes have been impeded by the lack of molecular characterization of the Y chromosome. Here we show that the *An. gambiae* Y chromosome contains a very small repertoire of genes, with massively amplified tandem arrays of a small number of satellites and transposable elements constituting the vast majority of the sequence. These genes and repeats evolve rapidly, bringing about remodeling of the Y even among closely related species. Our study provides a long-awaited foundation for studying mosquito Y chromosome biology and evolution.

4.3 Introduction

Sex chromosomes carry a master switch gene responsible for sex determination [175] 1983. They are thought to derive from an ordinary pair of autosomes, and have multiple independent origins across the tree of life [176] [177]. In animals with morphologically distinct (heterogametic) sex chromosomes, the Y has ceased crossing over with the X across some or all of its length and the non-recombining region is transmitted clonally by males [178] [179]. The absence of recombination initiates progressive genetic decay—gene loss and accumulation of repetitive sequences—but there is increasing recognition that even relatively old and otherwise highly degenerate Y chromosomes retain functional importance not only for sexual reproduction but for their contributions to global gene regulation affecting health and survival bellott [180] cortez [181], lemos [182], lemos [183], sackton [184]. Notwithstanding these critical roles, the Y chromosome remains one of the most recalcitrant and poorly characterized portions of any genome more than a decade into the post-genomic era, with current knowledge resting largely on only two animal groups: mammals and *Drosophila* bachtrog [177], hughes [185](2, 11).

Mosquitoes in the genus *Anopheles* are the exclusive vectors of human malaria, a disease that claimed nearly 600,000 lives globally in 2013—the majority in sub-Saharan Africa WHO [1]2014 (12). Although fifteen years of intensified vector control efforts (mainly insecticide-impregnated bed nets) have successfully averted an estimated 663 million clinical cases of malaria bhatt [186] (13), further progress toward elimination in the most malarious regions will depend upon the development of novel methods of vector control complementary to existing approaches (14). One method currently under development entails genetic modification of the mosquito to bias the population sex ratio toward males (which do not bite), with the goal of local population reduction or elimination galizi [187] [188] [189](15-17). Modeling has shown that

the most efficient means toward this end is the engineering of a driving Y chromosome deredec [190] (18). A molecular-level understanding of the *Anopheles* Y chromosome is important to inform and optimize such a strategy.

Historical cytogenetic studies established that the *Anopheles* Y chromosome is entirely heterochromatic, but also suggested that, contrary to *Drosophila* and in common with mammals, it bears partial homology to the X chromosome and plays a male-determining role sakai [191], redfern [192], Mitchell [193], marchi [194] white [195], fraccaro [196] (19-24). However, efforts to characterize Y chromosome sequences in *Anopheles* have been thwarted by a lack of directed resources and effective tools. The unsurpassed medical importance of the African malaria mosquito *Anopheles gambiae* motivated its early selection for whole genome sequencing holt [140] (25), making it second only to the model organism *D. melanogaster* as a fully sequenced insect genome. Yet because of formidable obstacles to assembling repeat-dense Y chromosome sequences carvalho [197] (26), efforts to assemble or even to assign Y chromosome sequences in the framework of the *An. gambiae* genome project were largely unsuccessful despite separate sequencing of males and females krzywinski [198] (27), leaving its content and organization obscure. Here, we leverage the increased length and reduced bias of single-molecule sequencing ross [199] (28), along with Illumina-based sex-specific transcriptional profiling and whole-genome sequencing, to identify an extensive data set of Y chromosome sequences and explore their organization and evolution in the young species radiation known as the *An. gambiae* complex white [200], which contains some of the most important vectors of human malaria. We find massive remodeling of the repeat-dense Y chromosome and remarkably few genes, none of which have counterparts on the X. Only the *YG2* gene—the earliest to be expressed in 3-hour male embryos—is conserved and exclusive to the Y across species in the

complex, and thus is a possible male-determining factor. Yet no sequence similarity to *YG2* can be detected in the genome of an Asian malaria vector from the same subgenus hall [201], underscoring the rapid evolution of Y-linked genes in the evolutionarily dynamic genus *Anopheles* neafsey [165]. Contrary to the species branching order, the *YG2* gene tree from the *An. gambiae* complex supported the grouping of two major malaria vectors with a history of substantial autosomal introgression fontaine [160], consistent with the hypothesis that even the Y chromosome may have crossed species boundaries. Although a Y chromosome assembly awaits further technological developments, the compilation and comparative analysis of Y chromosome sequences in the *An. gambiae* complex substantially advances our understanding of the composition, organization and evolution of the *Anopheles* Y chromosome and lays the groundwork for exploiting the Y chromosome to control disease transmission.

4.4 Results

Identification of *An. gambiae* Y chromosome sequences. Gross cytological estimates suggest that the *An. gambiae* Y chromosome constitutes ~10% of the 264 Mb *An. gambiae* genome [198] [49], yet a mere 0.18 Mb of unordered sequences have previously been assigned to the Y in the PEST reference genome assembly [www.vectorbase.org; [202]]. Similarly, a recent *An. stephensi* genome project identified only 57 short unordered Y sequences spanning ~50 kb from genomic reads, and 11 contigs spanning ~200 kb from BAC clones that were assigned to the Y chromosome [166]. To overcome this impediment, we developed a strategy based on long-read, PacBio single molecule real-time (SMRT) sequencing eid [203]. Template genomic DNA was extracted from male siblings of a single-pair mating that inherited the same paternal *An. gambiae* Y chromosome and was sequenced to 70X autosomal (35X heterosomal) coverage with PacBio SMRT sequencing [*Supporting Information* (*SI*) *Appendix*, Text S1.1]. Consensus-based error

correction with PBcR [204](37) resulted in 40X autosomal (20X heterosomal) coverage of PacBio corrected reads with an N50 size of 2,799 bp (*SI Appendix,* Text S1.2). A whole-genome assembly of the entire PacBio dataset was performed with Celera Assembler [205] {Myers, 2000 #150, resulting in a 294 Mb assembly with an N50 contig size of 101,465 bp (*SI Appendix,* Text S1.2). However, the moderate coverage and average raw read length of 2,479 bp proved insufficient for *de novo* reconstruction of the Y chromosome. Initial analysis of this assembly revealed highly fragmented heterosomal contigs and evidence that autosomal and heterosomal sequence had been incorrectly joined. Due to known limitations in assembling heterochromatic sequence and concerns about potential mis-assemblies, we decided to focus exclusively on individual (unassembled) PacBio corrected reads from genomic DNA for all subsequent Y chromosome analysis.

As a complementary strategy to investigate the organization of large (100-kb) contiguous pieces of the *Anopheles* Y chromosome, PacBio sequencing of individual *An. gambiae* BAC clones was performed (*SI Appendix,* Text S1.3). These BACs were deemed potentially Y-linked based on initial computational analysis of available BAC-end sequences (*SI Appendix* S1.3). Directed, high-coverage PacBio sequencing, ranging from 300X to 2,000X per BAC, yielded sufficient information to completely assemble each BAC without ambiguity. Successful BAC assembly was due to localization of the repeat structure as well as the tremendous sequencing depth attained. Because PacBio read lengths are exponentially distributed, such deep sequencing increases the probability of obtaining some very long reads (e.g. >20 kb), which were necessary for the assembly of these highly repetitive sequences. Comprehensive computational analysis supported three of four assembled BACs as originating from the Y chromosome (*SI Appendix* S1.3, figs. S1-S2).

To identify presumptive Y-linked sequences among the unassembled genomic PacBio reads, we implemented two recent computational approaches (*SI Appendix,* Text S2) that exploit short-read (Illumina) genomic sequencing from sex-specific DNA pools (*SI Appendix,* Text S1.4, Table S1). The Y chromosome genome scan (YGS) approach was designed to operate on scaffolds from a genome assembly derived from mixed sexes or males; after identification and masking of identical repeats, scaffolds are classified as Y-linked if they have few or no kmer-length matches to female Illumina sequences {Carvalho, 2013 #1509}. As applied to *An. gambiae* male PacBio corrected reads, which were treated as "scaffolds," YGS failed to unambiguously classify Y-linked sequences, apparently due to the extremely small fraction of Y chromosome sequence that is exclusive to the Y in *An. gambiae* as opposed to highly enriched there (see below; *SI Appendix*, Table S8). A second approach, the chromosome quotient (CQ) method hall [201], infers Y-linkage based on the female-to-male ratio of sequence alignments to a reference—in this case *An. gambiae* female-to-male Illumina sequences aligned to PacBio reads. At a conservative threshold value ($CQ \leq 0.2$) imposed across the length of a PacBio read, the CQ method classified 79,475 unassembled reads (246 Mb) as presumptive Y chromosome sequences, which populate a database that we denote Ydb (*SI Appendix*, Text S2.2, Tables S4-S7; Other Supporting File 1). Although the rate of false positives in Ydb should be low hall [86] (30) (*SI Appendix,* Text S2.1), the conservative CQ threshold necessarily means that Ydb is incomplete with respect to possible Y chromosome sequences that share extended sequence identity with other chromosomes, as would be expected for pseudoautosomal regions or sequences recently acquired from elsewhere in the genome. However, it is likely that Ydb represents much of the non-recombining (male-limited) Y chromosome (NRY) (*SI Appendix*, Table S4). Greater than 94% of sequence classes comprising Ydb were validated as Y-linked in

An. gambiae (Fig. 1; *SI Appendix* Text S3, Table S9) through genomic PCR and physical mapping by fluorescent *in situ* hybridization (FISH) of representative sequences to mitotic chromosomes of male *An. gambiae* larvae (*SI Appendix,* Text S4).

The *An. gambiae* Y contains massively amplified satellites and retrotransposons. We conducted a detailed computational assessment of Y chromosome repeat content based on analysis of Ydb. Our inferences should be minimally affected by redundant and overlapping reads, as they are based on proportional content (relative abundance), and PacBio coverage has limited bias [199] [205]. Initially, both PacBio Ydb reads and assembled BAC sequences were screened for interspersed repeats and low complexity DNA with RepeatMasker 4.0.3 smit (41), using the *An. gambiae* PEST RepeatMasker library augmented with previously characterized Y chromosome satellite and retrotransposon sequences [52] [206] (*SI Appendix,* Text S3). Anticipating that the *An. gambiae* Y chromosome contains previously unknown repeats, or repeats whose structures differ from those represented in the reference repeat library, we characterized both annotated and unclassified output from RepeatMasker through iterative clustering and consensus building of sequences in Ydb and the Y-linked BACs. This strategy ultimately revealed that ~98% of bases in Ydb belong to a very few repetitive sequence classes, amplified extensively (Fig. 2A; *SI Appendix,* Text S3, Table S9).

Satellite DNA accounts for ~49% of all bases in Ydb (*SI Appendix*, Table S9). Yet only six different satellite monomers were identified and two—AgY477 and AgY373—predominate, comprising 93% of all satellite DNA sequence in Ydb. Moreover, the satellite sequences are found as long tandem arrays in Ydb, largely devoid of transposable elements (TEs). These data

suggest that satellite DNA is an abundant and homogeneous constituent of the NRY, a major Y chromosome sequence feature that we refer to as the SAR, for Satellite Amplified Region (Fig. 2A). The absence of other repetitive sequence classes interspersed within the SAR suggests limited genetic exchange between satellites and other repeats, but we find evidence of recombination and higher-order repeat structures among satellites within the SAR. Different satellite monomers that share extensive sequence similarity (AgY477 and AgY373; AgY280 and AgY53D) frequently co-occur on the same PacBio read, often as interspersed or even chimeric monomers, indicative of sister chromatid or intrachromatid exchange (Fig. 2C; *SI Appendix,* Text S3.1, Figs. S5-S8).

Another 43.5% of bases in Ydb are TE-related, of which only eight distinct TE types, mainly retrotransposons, were identified (*SI Appendix,* Text 3.2, Table S9). Remarkably, one particular element alone—a 6.9 kb Ty3/Gypsy LTR retrotransposon that we designate *zanzibar*—comprises almost 27% of bases in Ydb as a whole, and accounts for more than 61% of the bases classified as TEs (Fig. 2, *SI Appendix,* Table S9). Unlike the typical interspersed arrangement of TEs in genomic euchromatin, *zanzibar* is arranged on the Y in massively amplified head-to-tail tandem arrays, in which one gag/pol region followed by one LTR is repeated in succession like beads on a string: $(gag/pol+LTR)_n$. From its abundance and organization, we infer that this *Zanzibar* Amplified Region (designated ZAR; Fig. 2A,B)—like the SAR—is another prominent organizational feature of the NRY. *Zanzibar* monomers (gag/pol+LTR) in the array may carry insertions of a variety of other TEs or TE fragments. Remarkably, every copy of *zanzibar* carrying a particular TE type (*e.g.*, *mtanga*) contains precisely the same TE sequence inserted into precisely the same *zanzibar* sites (Fig. 2, *SI Appendix,* Text S3.2), as though replica insertions in different *zanzibar* copies are not the result

of independent transposition events. Taken together, these data—the precise tandem organization of the ZAR and the peculiar clonal nature of insertions—strongly suggest that *zanzibar* retrotransposons no longer function in the manner expected of autonomous transposable elements. Instead, *zanzibar* sequences appear to have been the substrate for illegitimate recombination and megabase-spanning tandem amplifications on the Y chromosome, analogous to the process described for the genesis of centromeric repeats in maize [207]. Despite its evident origin as an autonomous transposable element, the present ZAR structure most closely resembles satellite sequence and thus may reflect a general pattern of sequence amplification and evolution on the *Anopheles* Y chromosome.

Of the remaining ~7.5% of Ydb bases, we were able to classify ~5.5% as repetitive, but the last ~2% could not be clustered (*SI Appendix*, Table S9). This small unclustered fraction contains a heterogeneous mixture of less abundant types of repetitive sequences, degenerate copies of repeats categorized above, and Y chromosome genes, many of which also appear to be multicopy (see below; *SI Appendix*, Text S7, Fig. S11). Only four Ydb reads were classified by a metagenomic analysis as originating from another organism (*SI Appendix*, Text S2.2), suggesting that nearly all of Ydb is legitimate *An. gambiae* sequence.

Extensive structural dynamism of the Y chromosome in a young species radiation.

Cytological observations conducted in the 1970s revealed striking differences in sex chromosome heterochromatin among populations and between species in this complex [84] [87]. Not only did the staining intensity and pattern vary, but also length of the Y chromosome, ranging from less than half the length of the X in one *An. gambiae* population to almost the same

length as the X in others [84]. However, a mechanistic understanding of the phenomenon was lacking. Our finding that ~98% of the bases in Ydb constitute highly repetitive sequence organized into tandem arrays suggests that the cytological observations may have their basis in rapid expansion and contraction of tandem repeats on the Y chromosome, through unequal crossover and a variety of other mechanisms [198] [206] [208] [209] [210]. We applied computational and cytogenetic methods to assess the nature and degree of Y chromosome remodeling within *An. gambiae* and among sibling species during the relatively brief (2 MY) evolutionary history of the species complex.

Intraspecific variation in the SAR and ZAR of *An. gambiae* was assessed computationally among three laboratory colonies and among 85 male and female mosquitoes sampled from a natural population in Cameroon, available through MalariaGEN's *An. gambiae* 1000 Genomes Consortium (Ag1000G) phase 1 AR3 data release (2015) (http://www.malariagen.net/data/ag1000g-phase1-AR3; *SI Appendix*, Text S1.4-1.5, Table S1-S2). The sample set from Cameroon is one of the few available that contains both sexes. From the *An. gambiae* Pimperena colony—our reference—and two additional colonies (G3 and Asembo), we generated Illumina sequences from sex-specific genomic DNA pools, and aligned them to the consensus sequences of *An. gambiae* monomer repeat units compiled from Ydb. Alignments were performed twice, using either a strict read-mapping protocol (for CQ calculations) or a less stringent mapping protocol used to produce a metric analogous to CQ, termed "relaxed CQ" (RCQ) (*SI Appendix*, Text S5, Tables S6-S7). Presence/absence and relative abundance of each major repeat was estimated from the number of mapped reads; male-bias was estimated from the female-to-male ratio of sequences reflected by CQ and RCQ (*SI Appendix*, Tables S6-S7). Using similar strategies, we also interrogated individually sequenced

wild-caught *An. gambiae* mosquitoes of both sexes (40 males; 45 females) from Cameroon (*SI Appendix,* Figs. S11-13, Tables S12-14).

Overall, the pattern of Y-linkage as inferred by CQ and RCQ was qualitatively similar among *An. gambiae* samples. However, the corresponding copy number of Y-linked sequences was much more labile (Fig. 3; *SI Appendix*, Text S5, Figs. S11-S13, Tables S6-7, S13-S15). The most dramatic copy number variation of any SAR or ZAR component was displayed by satellite sequences AgY53D and AgY280 in male samples (Fig. 3). Although numbers of read alignments are not precise reflections of copy number, they can convey a rough approximation of relative abundance if copy number varies across orders of magnitude. Indeed, counts of male alignments to AgY53D and AgY280 spanned four to five orders of magnitude from the Asembo to the Pimperena colony (*SI Appendix,* Tables S6-S7), and even within the natural population from Cameroon, normalized alignment counts among individual males spanned three orders of magnitude (*SI Appendix*, Table S13), suggesting major expansions or contractions in array length. Copy number variation of this magnitude on the Y chromosome would be expected to affect its length. We estimated the combined effect on chromosome length of copy number variation of all SAR and ZAR components among the male *An. gambiae* sampled from Cameroon. Taken together, copy number variation can account for the gain or loss of up to ~30 Mb of the Y chromosome. Broadly consistent with this *in silico* estimate, our cytogenetic length estimates also varied by as much as 22 Mb among individual males from the *An. gambiae* Pimperena colony (from ~25.9 Mb to ~47.8 Mb; *SI Appendix*, Text S4.2, Table S11), in conformity with prior studies[211] .

The sibling species complex to which *An. gambiae* belongs radiated rapidly and recently, within the last 2 MY [160]. To examine the extent of structural divergence of the Y

chromosome between species over this relatively short time frame, we generated Illumina sequences from sex-specific pools of three additional members of the complex (*An. arabiensis*, *An. quadriannulatus*, *An. merus*) (*SI Appendix*, Table S1), and assessed male bias and relative abundance as described above for *An. gambiae* (*SI Appendix*, Text S5.2, Tables S6-S7, S15), and by FISH (Fig. 4). Rapid and extensive remodeling of the Y chromosome between species is evidenced by dramatic examples of turnover in both the SAR and ZAR. Satellite AgY477, heavily male-biased and a major component of the Y in *An. gambiae*, is abundant but not strongly sex-biased in *An. merus*, and is not detected in *An. arabiensis* or *An. quadriannulatus* of either sex (Fig. 1; *SI Appendix* Text S5.2, Table S15). Similarly, *zanzibar* is neither strongly sex-biased nor abundant in *An. arabiensis* and *An. merus*, yet in *An. quadriannulatus* (not a sister species of *An. gambiae*), this retrotransposon has an *An. gambiae*-like pattern of expanded tandem arrays on the Y chromosome (Fig. 1, Fig. 4, *SI Appendix*, Text S5.2, table S15).

The *Anopheles* Y recombines with the X chromosome. Meiotic pairing, chiasma formation, and crossing-over between the sex chromosomes have been reported for three anopheline species in two different subgenera [191] [193] [196], but to our knowledge, similar observations have not been reported in *An. gambiae*. Indeed, the apparent stability of X- and Y-linked translocations in *An. gambiae* [212] suggests that legitimate crossing-over between the X and Y chromosomes does not occur. However, our data are difficult to explain without some form of X-Y genetic exchange since the evolution of the sex chromosomes from an ancestral pair of autosomes, an event that must predate anopheline diversification ~100 MYA. First, there is a very high degree of sequence similarity between Y-associated and X-associated repetitive DNA in *An. gambiae*. Our initial indication of this was the near-complete failure of the YGS

computational method to identify Y-associated sequences in *An. gambiae*, presumably arising from the fact that such sequences are only rarely exclusive to the Y chromosome, as opposed to highly enriched there. Strongly supporting empirical evidence comes from physical (FISH) mapping of individual components of the SAR (*e.g*., AgY53B; Fig. 4, *SI Appendix*, Fig. S10), or of fluorescently labeled sequences derived from the entire microdissected Y chromosome (Fig. 5A), which reveals extensive cross-hybridization of Y repeats with X chromosome heterochromatin due to sequence similarity between satellite monomers (*SI Appendix*, Text 4.1.2). Detailed sequence analysis indicates that satellite monomers normally abundant only on the Y (AgY373, AgY477) share ~93% pairwise sequence similarity with a satellite monomer from the X chromosome [AgX367; (42)] (*SI Appendix*, Fig. S6). Moreover, we found the footprints of recombination on individual *An. gambiae* PacBio reads. These contained AgX367 monomers together with AgY477 and AgY373 monomers, including AgX367-AgY373 and AgX367-AgY477 recombinants (Fig. 5C), confirming previous evidence of recombination based on PCR amplicon sequencing [213]. A second line of evidence for occasional X-Y genetic exchange emerged from individually sequenced male and female *An. gambiae* from Cameroon (Fig. 5B; *SI Appendix*, Text S6, Fig. S11). In a small subset of females, normalized counts of Illumina reads mapping to consensus satellite monomers normally abundant only on the Y chromosome (*e.g*., AgY53A) were almost two orders of magnitude larger than the median for all females (*SI Appendix*, Tables S12, S14). As this female subset lacked correspondingly high copies of other Y chromosome sequences, contamination by male genomic DNA (whether in the laboratory or via sperm stored in the spermatheca) is an unlikely alternative explanation. Although intrachromosomal genetic exchange on the X may contribute to the

strongly bimodal pattern of satellite abundance observed in Cameroon females, this phenomenon must have its basis in periodic X-Y genetic exchange.

Rapid turnover of the small Y chromosome genic repertoire. Only three Y-linked genes had been identified previously in *An. gambiae* [86], designated *gYG1* to *gYG3* (hereafter, *YG1* to *YG3*). Aiming for comprehensive gene discovery on the *An. gambiae* Y chromosome, we performed extensive transcriptional profiling of developmentally staged *An. gambiae* embryos (nine time points), sexed larvae (three time points), and adults (whole and dissected males and females) through mRNA sequencing (RNA-Seq)—52 data sets in total (*SI Appendix*, Text S1.6, Table S3)—and integrated complementary approaches to gene finding (*SI Appendix*, Text S7). Gene candidates bearing significant similarity to known TEs or bacterial sequences were discounted. Arising from the combined approaches were eight presumptive genes (*YG1-8*), including the three previously identified (Fig. 1; *SI Appendix*, Text S7).

To be considered valid Y genes, we required that they exhibit male-biased or male-specific expression from RNA-Seq as well as male-specific amplification by genomic PCR, conditions met by *YG1-5*. With the exception of *YG4*, this validated set was further confirmed by male-specific RT-PCR. Moreover, we were able to physically localize *YG5* to the Y chromosome by FISH (a homolog was also detected on chromosome 3; Fig. 4, *SI Appendix*, Text S7, Fig. S19). As we could not identify male-specific SNPs distinguishing *YG6-8* from their autosomal homologs, these genes could not be validated despite exclusive expression of *YG6* in male accessory glands (*SI Appendix*, Fig. S20), and elevated numbers of normalized read

alignments to *YG8* from individually sequenced males versus females in our population sample from Cameroon (*SI Appendix*, Fig. S11, Tables S12-S14).

Beyond the strikingly small total number of genes identified as Y-linked in *An. gambiae* following this intensive search, it is noteworthy that gene number varies even between strains. Both *YG3* and *YG4* are Y-linked exclusively in the G3 strain of *An. gambiae*, not in Asembo or Pimperena (*SI Appendix*, Tables S6-S7). None of these genes have recognizable gametologs on the X, yet all have partial or complete homologs on the autosomes (*SI Appendix*, Text S7, Fig. S14), suggesting that they have been gained on the Y chromosome since its divergence from the X (see below).

To screen for candidate Y-linked genes in three *An. gambiae* sibling species (*An. arabiensis*, *An. merus*, and *An. quadriannulatus*), we used the male and female Illumina sequences from each species in conjunction with corresponding genome assemblies, mixed-sex transcript sets and *de novo* RNA-Seq assemblies [165](31), as well as *An. gambiae* genomic resources (*SI Appendix*, Text S7). After merging the results of these approaches, we made two surprising observations. First, among all the Y-linked genes identified in *An. gambiae*, only one—*YG2*—was computationally detected and confirmed (by male-specific genomic PCR) as Y-linked in each of the other three very closely related species (Fig. 1). None of the other *An. gambiae* Y genes, with the sole exception of *YG1*, was Y-linked in any other species examined. For *YG1*, Y-linkage was validated in *An. arabiensis* and *An. quadriannulatus*, but in *An. merus* the ratio of female-to-male alignments was inconclusive and male-specific genomic PCR amplification was not possible owing to highly similar sequence elsewhere in the genome (Fig. 1, *SI Appendix*, Text S7, Tables S6-S7, Fig. S14). Thus *YG2* is the only gene both conserved on, and exclusive to, the Y chromosome in all four species examined. As it is the earliest to be

expressed, at 3 hours of male embryonic development (*YG1* is not expressed until 4 hours), *YG2* is a possible male determining gene. Surprisingly, *YG2* is not a single-copy gene. The first suggestion that this might be the case was hinted by the number of read alignments to *YG2* from individual males in the Cameroon sample (Fig. 3). However, we have more definitive evidence for multiple, nearly identical copies of *YG2* in *An. gambiae*. Four distinct haplotypes were sampled repeatedly in Ydb PacBio reads (which derived from the same paternal Y chromosome; *SI Appendix*, Text S7.1.2, Fig. S17). Variant positions among *YG2* copies were not only validated by sequencing of genomic PCR amplicons from individual male *An. gambiae* derived from natural populations, but also through RNA-Seq data, which further indicates that multiple *YG2* copies are expressed (*SI Appendix*, Table S17).

With the exception of *YG2*, the near-complete absence of conserved Y-linkage between *An. gambiae* genes and corresponding genes in the sibling species was reinforced by the converse result, our second surprising observation: all genes identified as Y-linked in any one sibling species could not be assigned to the Y chromosome in any of the other species (*SI Appendix*, Text S7.2, Tables S6-S7). In *An. quadriannulatus*, we found three novel Y-linked candidate genes (*SI Appendix*, Text S7, Table S20). In *An. arabiensis*, no genes other than *YG1* and *YG2* were detected on the Y chromosome. In *An. merus*, we found evidence supporting the duplication of a multi-gene segment from chromosome 3R onto the Y since its split from other *An. gambiae* complex lineages (*SI Appendix*, Text S7, Table S19). Among seven sequential genes on 3R in this segment (corresponding to AGAP009631-37 in *An. gambiae*), the first three (AGAP009631-33) and last two (AGAP009636-37) have detectable copies on the Y chromosome in *An. merus* (based initially on the relative number of normalized read alignments in females and males, later validated by male-specific PCR). Our data are consistent with the

two intervening genes (corresponding to AGAP009634-35 on 3R) having been lost from the Y, and the five flanking genes becoming amplified, although further experimental evidence will be required to confidently reconstruct these events.

Possible Y chromosome introgression between hybridizing malaria vectors. The *YG2* gene potentially encodes a short, 56-aa peptide whose possible role in determining maleness is under investigation. In the Asian malaria vector *An. stephensi*,a Y-linked gene (*Guy1*) implicated in male determination also encodes a 56-aa sequence whose predicted secondary structure resembles that of the putative *YG2* peptide [214] even though primary sequence similarity is not detectable [86] over the relatively short evolutionary span since these lineages separated, ~30 MYA [215]. The fact that *YG2* expression is detected in early embryos before any other *An. gambiae* Y-linked gene, taken together with its uniquely conserved Y chromosome location in all four *An. gambiae* complex species—in the face of otherwise rampant structural dynamism and genic turnover on the Y—is consistent with a primary role in male determination in this group. For this reason, we predicted that a gene tree reconstructed from *YG2* would reflect the known species branching order [160]. Although sibling species in the *An. gambiae* complex are not completely reproductively isolated, contemporary interspecific gene flow is possible only though female F1 hybrids; as their brothers are sterile, the Y chromosome cannot introgress. Contrary to this expectation, a *YG2* tree built from sequences derived from population samples of the four species considered in this study supported *An. gambiae* and *An. arabiensis* as most closely related (Fig. 6; *SI Appendix*, Text S8), an arrangement previously shown to be the result of massive historical introgression between these two species that involved most of the autosomes and the proximal ~10 Mb of the X chromosome [160]. The simplest explanation for a gene tree disagreeing with the species tree in a rapid radiation such as the *An. gambiae* complex is

incomplete lineage sorting (ILS). We performed coalescent simulations to assess the likelihood that the grouping of *An. gambiae* with *An. arabiensis* in the *YG2* tree is due to ILS alone. In 62 out of 1000 simulations under the species tree, we recovered *An. gambiae* and *An. arabiensis* as sister lineages (i.e. P=0.062; *SI Appendix*, Text S8), indicating that non-introgressing lineages could produce the observed tree a small fraction of the time. Although we cannot formally reject the null hypothesis at the 0.05 significance level, these results certainly do not rule out Y chromosome introgression. Introgression of the Y chromosome between species is conventionally viewed as unlikely payseur [216], but it is important to consider that the pair of malaria vectors in question have historically exchanged the vast majority of the rest of their genomes, including part of the X chromosome [160]. In this context, introgression of the Y chromosome is possible if not likely, as long as the introgression event(s) predated the development of male F1 hybrid sterility barriers between this species pair.

4.5 Discussion

From studies of mammals [180] [181] [185] [217] [218]and *Drosophila* [219] [220] [221], it is known that the Y chromosomes in both groups have lost most of their ancestral gene repertoires and have acquired copious amounts of repetitive and ampliconic/palindromic DNA. In the ~250 MY since *Drosophila* and *Anopheles* last shared a common Dipteran ancestor, there has been parallel evolution of heteromorphic sex chromosomes from the same ancestral linkage group [222] (61), implying that the *Anopheles* Y must have undergone a similar fate of massive ancestral gene loss and genomic degradation. In one main characteristic—its male determining role—the *Anopheles* Y resembles the mammalian Y more than it does *Drosophila*, in which XO flies are (sterile) males and the scant Y chromosome genes are crucial only for male fertility [177]. Yet in other respects, the *Anopheles* and *Drosophila* Y are much more similar. More than

one-third of the human Y chromosome and 99.9% of the mouse Y is euchromatic [217] [223], whereas *Drosophila* and *Anopheles* Y chromosomes are entirely heterochromatic. Although relatively few in number, some ancestral X-Y gene pairs have been conserved throughout mammalian evolution due to their vital role as dosage-sensitive regulators of global gene expression [180] [181]. Crucially, although cases are known in which a mammalian Y chromosome has acquired autosomal genes [*e.g*., [224]], most extant mammalian Y-linked genes have an X-linked gametolog. By contrast, what little gene content exists on the Y in *Drosophila* or *Anopheles* is not only poorly conserved between species, but there are no recognizable ancestral gametologs; all known Y-linked genes in *Drosophila* seem to have an autosomal origin [219]. The recent DNA-based duplication of a gene from chromosome 3R to the Y chromosome in *D. melanogaster* following its split from *D. simulans* ~4 MYA (64) mirrors our finding of a similar event in *An. merus* since the radiation of the *An. gambiae* complex, < 2 MYA. We conclude that the most salient factor uniting *Anopheles* and *Drosophila* Y chromosomes may be the continuous gain of genes and functions from the autosomes [225], in contrast to the conservation of remaining ancestral gametologs seen on the mammalian Y. However, the *Anopheles* NRY appears to stand apart from both *Drosophila* and mammalian Y chromosomes in the relative paucity of male-specific content.

One of our main findings was rapid turnover in quantity and type of repetitive DNA on the Y chromosome within and between species in the *An. gambiae* complex. It is known that both satellite and ampliconic DNA regions are prone to rapid divergence in length, structure and sequence, due to unequal sister chromatid exchange between out-of-register repeat units and other mechanisms [210]. On the Y chromosome such regions may be subject to accelerated rates of divergence compared to the rest of the genome. Between humans and chimps whose lineages

diverged ~5 MYA, orthologous satellite arrays in the X centromere are collinear and share 93% sequence identity, while collinearity declines and sequence conservation drops to 78% between orthologous satellites in Y centromeres [226]. Additional evidence of rapid length, structure and sequence evolution of satellites and ampliconic structures on the Y chromosome has been reported in mice species 1-2 MY diverged and mice subspecies separated by only ~900,000 years[226]; between *D. melanogaster* and *D. simulans* that split ~4 MYA [221] [227]; and among human males worldwide [228]. Despite such pervasive remodeling of the *Anopheles* Y chromosome over short evolutionary distances, a transgene randomly inserted onto the Y chromosome of an *An. gambiae* strain in 2014 is transcriptionally active and has been stably integrated ever since, establishing that the Y chromosome is amenable to the molecular manipulation required for Y-linked genetic vector control strategies [229].

The high level of satellite DNA polymorphism within species could have important phenotypic consequences for fitness related traits [182, 183]. Moreover, the dramatic degree of satellite DNA turnover on the Y between closely related species has been implicated in hybrid incompatibility in *Drosophila* [59] [230] [60]. Intriguingly, two genes known to cause hybrid incompatibility between *D. melanogaster* and *D. simulans* (*Hmr* and *Lhr*) function within species to repress transcripts from satellite DNAs and TEs [60]. These species differ drastically in satellite DNA content; *D. simulans* contains four-fold less satellite DNA overall (5% versus 20% of the genome), and is particularly depauperate of the two most abundant satellite types in *D. melanogaster* lohe [231]. In *An. gambiae*, we found that AgY477 and AgY373 are the most abundant satellites on the NRY, and they are both expressed exclusively in adult male reproductive tissues; these satellite sequences are absent or altered in the other sibling species

investigated (Fig. 1). Whether hybridization leads to misregulation of satellite DNA remains to be explored in the *An. gambiae* complex.

Laborious single-haplotype iterative mapping and sequencing has previously revealed the structure of mammalian Y chromosomes [185]. In contrast, single-molecule sequencing now provides individual reads tens of kilobases in length, promising a resource-efficient alternative for characterizing Y chromosomes. Here, we were able to determine the content and structural characteristics of the heterochromatic *An. gambiae* NRY using this approach. Single-molecule sequencing reads were able to reveal complex repeat structures from whole-genome data and completely assemble heterochromatic BACs without manual finishing. However, the complete reconstruction of heterochromatic Y chromosomes remains a challenging open problem, as a recent PacBio assembly of *D. melanogaster* failed to completely assemble the Y chromosome [232], but it did successfully resolve the complex regions Mst77Y [233] and FDY [225]. The minimum read length, accuracy, and coverage required for the successful assembly of heterochromatin from whole-genome data is currently unknown and will vary by species, but as demonstrated here, *An. gambiae* Y chromosome BACs can be successfully reconstructed using long sequences collected from deep, directed single-molecule sequencing. These results suggest that continued single-molecule read length and throughput improvements may soon enable the complete reconstruction of Y chromosomes from whole-genome data alone.

4.6 Materials and Methods

Please see *SI Appendix* for detailed information about (1) genomic and transcriptomic datasets; (2) identification of Y chromosome sequences from PacBio whole genome sequencing; (3) repetitive DNA content of the *An. gambiae* Y; (4) fluorescence *in situ* hybridization and size estimation of the Y chromosome; (5) copy number variation on the Y chromosome among individual *An. gambiae* mosquitoes and across the *An. gambiae* complex; (6) Y chromosome recombination; (7) genic repertoire of the Y chromosome; and (8) phylogeny reconstruction and coalescent simulations.

The supplementary figures and tables are listed on www. pnas.org

Part of the supplementary information (my work) is included below.

S4. Fluorescence *in situ* hybridization and estimating size of the Y chromosome (fig. S10, tables S10-S11)

S4.1. Fluorescence *in situ* hybridization

S4.2. Estimated size of the Y chromosome.

S4. Fluorescence *in situ* hybridization and estimating size of the Y chromosome

S4.1. Fluorescence *in situ* hybridization

S4.1.1. Methods

Mosquito strains: Laboratory colonies examined for this study were provided by the Malaria Research and Reference Reagent Resource Center (MR4). These included multiple colonies of *An. gambiae* (Pimperena, Asembo, Kisumu, Zanu), and *An. arabiensis* Dongola, *An.*

quadriannulatus Sangwe, and *An. merus* MAF. Mosquitoes were reared at 27°C, with 12:12 light:dark cycle and 70% relative humidity.

Chromosome preparation: Most of the physical mapping was done on metaphase and prometaphase mitotic chromosomes sourced from leg and wing imaginal discs of early 4th instar male larvae. However, because the Y and X chromosomes are morphologically similar in *An. merus*, it was more difficult to verify whether imaginal disc preparations were from males (i.e., contained both X and Y chromosomes). Accordingly, for *An. merus* our preparations were derived from testes of late stage pupae, and mapping was done on metaphase I meiotic chromosomes. Chromosomes were prepared following ref. [167](26). Larvae or pupae were immobilized on ice for 10 minutes. Dissections were performed on microscope slides in a drop of cold freshly made hypotonic solution (0.075M KCl). A fresh drop of hypotonic solution was added to the preparation for 10 minutes, followed by fixation in a drop of modified Carnoy's solution (ethanol:glacial acetic acid, 3:1) for 1 minute. Next, a drop of freshly prepared 50% propionic acid was added and the preparation was immediately covered with a 22x22 mm coverslip. After 5 minutes, the preparation was squashed and dipped in liquid nitrogen for cover slip removal, followed by sequential dehydration steps in 70%, 80% and 100% ethanol. Slides with the highest number of metaphase plates were chosen for FISH.

Probe preparation and FISH: Genomic DNA was isolated from virgin males of the *An. gambiae* Pimperena strain using DNeasy Blood and Tissue Kit (Qiagen Inc., Valencia, CA, USA). Probes to perform FISH were generated by incorporating fluorescent labels during a PCR reaction. Primers for amplifying satellites AgY53A, AgY53B and AgY477 were obtained from ref. (22). For other targets, primers were designed using Primer 3 (27) (**Table S10**). For PCR labeling of satellites or TEs, each 25 μl PCR mix consisted of 35-40 ng genomic DNA, 0.3 U

Taq polymerase, 1× PCR buffer, 200 μM each of dATP, dCTP, and dGTP, and 65 μM dTTP, and 0.5 μl Cy3-dUTP or 0.5 μl Cy5-dUTP (*Enzo* Life Sciences, *Inc.*, Farmingdale, NY, USA) or 0.5 μl Fluorescein-dNTP (Thermo Scientific, Waltham, MA, USA). Thermocycling was performed using ImmoMix ™ (Bioline USA Inc., Taunton, MA, USA) beginning with a 95°C incubation for 10 minutes followed by 35 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec; 72°C for 5 min, and a final hold at 4°C. To prepare the *YG5* probe, a fragment of *YG5* was first amplified by PCR. Primers were then removed using Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA). The PCR product was labeled by nick-translation: each 50 μl reaction contained 1 μg of DNA; 0.05 mM each of dATP, dCTP, and dGTP, and 0.015 mM of dTTP (Fermentas, Inc., Glen Burnie, MD, USA); 1 μl of Cy3-dUTP or 0.5 μl Cy5-dUTP (Enzo Life Sciences, Inc., Farmingdale, NY, USA; 0.05 mg/ml of BSA (Sigma, St. Louis, MO, USA); 5 μl of 10x nick translation buffer; 20 U of DNA polymerase I (Fermentas, Inc., Glen Burnie, MD, USA); and 0.0012 U of DNAse I (Fermentas, Inc., Glen Burnie, MD, USA). FISH was performed as described previously (26).

Image acquisition and processing: After FISH, chromosomes were counter-stained with DAPI-antifade (Life Technologies, Carlsbad, CA, USA), kept in the dark for at least 2 hours, and visualized on an Olympus BX61 fluorescent microscope using BioView software (BioView Inc., Billerica, MA, USA) at 1000x magnification. Individual channels of the same plate were merged and the brightness and contrast were adjusted (applied to an entire image) using Adobe Photoshop.

Laser capture microdissection of Y chromosomes and chromosome painting: Imaginal discs from 4th instar larvae of *An. gambiae* Pimperena were dissected in a drop of hypotonic

solution (0.5% sodium citrate) and incubated for ~10 min. Imaginal discs were transferred to methanol:glacial acetic acid (3:1) fixative solution. Fixative solution was replaced by 60% acetic acid for tissue maceration. The solution was pipetted to homogenize the tissue to cell suspension condition. A membrane slide PET (Zeiss USA, Thornwood, NY, USA) was placed on a cold metal plate (-20°C) and ~10 µl of cell suspension was dropped onto the slide, allowing liquid to spread. The slide was placed on a heating table (45°C) to condense the drop. Cooling and heating were repeated five times, and then slides were left on the heating table to complete evaporation of liquid. Slides were dehydrated in 100% ethanol. After air-drying, Giemsa staining solution (1 ml of Giemsa to 10 ml sterile H_2O) was added for 5 min, then excess stain was washed with sterile H_2O. Slides were air-dried and stained in a controlled sterile climate to avoid contamination. Microdissection was performed using the PALM MicroBeam Laser Microdissection system (Zeiss USA, Thornwood, NY, USA) and the PALMRobo software as described previously (28). About 10-15 Y-chromosomes were microdissected from each slide by catapulting individual chromosomes without prior cutting. The captured Y chromosome material was dissolved in 10-15 µl of 1x PBS buffer and processed with the GenomePlex Single Cell WGA4 Kit (Sigma-Aldrich Co. LLC., St. Louis, MO, USA) protocol to produce the first library of amplified DNA. DNA was purified using the Genomic DNA Clean & Concentrator Kit (Zymo Research Corporation, Irvine, CA, USA). Amplified DNA was labeled for FISH using the GenomePlex WGA3 Reamplification Kit (Sigma-Aldrich Co. LLC., St. Louis, MO, USA) using dNTP mix, described previously for labeling in PCR (28). FISH of labeled microdissected Y chromosomal DNA to mitotic chromosome squash preparations of *An. gambiae* Pimperena was performed according to ref. (26).

S4.1.2. Results.

Y satellites and *zanzibar*. Satellite DNA repeats AgY53A and AgY53B showed similar patterns of hybridization to the *An. gambiae* Y chromosome (Fig. 4 and **Fig. S10**). Both probes were interspersed throughout ~70% of the Y chromosome on FISH images. In *An. arabiensis*, AgY53B showed an *An. gambiae*-like pattern of hybridization to the Y chromosome, but in *An. quadriannulatus*, hybridization was limited to a small region of the Y, and in *An. merus* AgY53B hybridized to multiple heterochromatin blocks that were stained by DAPI. AgY477 also showed an interspersed pattern in *An. gambiae* and *An. arabiensis*, but coverage and signal intensity was less than that observed for AgY53B. All the satellite DNA probes hybridized to both X and Y chromosomes. This can be due to the fact that the X and the Y share the same satellite monomers. However, this result can also be due to sequence similarity between different satellite monomers, which we could not control for in our FISH experiments, as our FISH probes contained the entire satellite monomer (in contrast to our computational analyses; see SI Appendix S5). For example, there is extensive sequence identity between AgY373/AgY477 and the X-biased satellite AgX367 (see **Fig. S6**) which would allow for cross-hybridization. In contrast to the satellite DNA, *zanzibar* hybridized exclusively to the Y chromosome in *An. gambiae* and *An. quadriannulatus*, and the signal covered ~50% of the Y chromosome. In *An. arabiensis*, *zanzibar* hybridized to a much smaller area of the Y chromosome (~20%), and *zanzibar* hybridization was not detected in *An. merus*.

S4.2. Estimated size of the Y chromosome.

We measured sex chromosomes of metaphase plates from *An. gambiae* Pimperena males. Mitotic chromosome spreads were prepared from imaginal discs as described above. Images of DAPI-stained chromosomes obtained from confocal microscope Zeiss LSM 880 (Zeiss USA, Thornwood, NY, USA) were measured using ZenLite software. An important source of chromosome length variation is the degree of X chromosome condensation, which increases from prometaphase to metaphase. In order to minimize this variation, we excluded prometaphase stages from our calculations. Metaphase chromosomes were selected if they had well-separated homologous chromosomes and identifiable sister chromatids. Prometaphase chromosomes, which had joined homologs and merged chromatids, were discarded.

Our strategy was to estimate the size of the Y chromosome in relation to the size of the assembled portion of the X chromosome. First, we measured the X chromosome from the telomere to the start of the ribosomal locus, which roughly corresponds to the portion (~25 Mb) that is assembled in the *An. gambiae* AgamP4 reference genome. We refer to this measurement as ($X_{assembled}$). The rest of the X chromosome is unassembled, as it is heterochromatic like the Y chromosome. Next, we measured entire (full-length) X and Y chromosomes from each metaphase plate. The formula used to estimate the size of the Y chromosome based on this set of measurements was (25 Mb*Y length in µm)/($X_{assembled}$ length in µm). Our calculations were based on 50 mitotic metaphase plates obtained from 8 *An. gambiae* Pimperena males. We found that the size of metaphase Y is approximately 36.7 Mb on average. The estimated size of the Y chromosome varied from 25.9 Mb to 47.8 Mb (**Table S11**). The size variation could be due to intraspecific polymorphism in the amount of the Y chromosome repeats and/or varying degree of chromosome condensation within metaphase.

4.7 Acknowledgements: We thank F. Catteruccia and S.N. Mitchell for sharing unpublished data, J. Pease for assistance with simulations, M. Kern, M. Menichelli, M.K. Lawniczak, I. Antoshechkin, T. Persampieri, R. Carballar, and R. D'Amato for technical assistance and discussion, and two anonymous reviewers for helpful suggestions. Sequencing data and assemblies have been submitted to NCBI under two umbrella BioProject IDs: PRJNA254152 and SRP044019. Genomic sequencing was funded in part by a grant from the Eck Institute for Global Health, University of Notre Dame. RNA-Seq was funded in part from a European Community Seventh Framework Programme (FP7/2007–2013) under grant agreement N° 228421 (INFRAVEC). Individual laboratories were funded by the NIH [R01AI076584 (NJB, MWH), R21AI112734 (NJB, SJE), R21AI101459 (NJB), R21AI094289 and R21AI099528 (IVS), R21AI105575 (ZT), HHSN272200900039C (SJE)], FNIH through the VCTR program of the Grand Challenges in Global Health Initiative (NJB, ACr, PAP, TN), European Commission and Regione Umbria Grant I-MOVE (RG, EF, PAP), Rita-Levi Montalcini Career Development Award (PAP), Marie Curie Intra-European Fellowship for Career Development (IEF) PIEFGA-273268 (TD), European Research Council Grant 335724 (NW), NSF Graduate Research Fellowship grant DGE-1519168 (ABH), Department of Education GAANN Fellowship (AS), and Fralin Life Science Institute of Virginia Tech (IVS, ZT). This research was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health (SK, AMP). The contributions of NHB, JH, and DR were funded under Agreement No. HSHQDC-07-C-00020 awarded by the Department of Homeland Security Science and Technology Directorate (DHS/S&T) for the management and operation of the National Biodefense Analysis and Countermeasures Center (NBACC), a Federally Funded Research and Development Center. The views and conclusions contained in this document are

those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security. In no event shall the DHS, NBACC, or Battelle National Biodefense Institute (BNBI) have any responsibility or liability for any use, misuse, inability to use, or reliance upon the information contained herein. The Department of Homeland Security does not endorse any products or commercial services mentioned in this publication.

Author Contributions: Conceived project: NJB, SJE, IVS, ZT; Coordinated project: NJB; Genome and BAC sequencing: NHB, JH, DR, AMP, ABH, ZT, NJB; PacBio read correction and assembly: SK, AMP; RNA-Seq, RT-PCR and gene validation: PAP, CC, OSA, ACa, TD, EF, RG, TN, NW, ACr; Computational analysis of Y linkage: ABH, PAP, CC, LA, XJ, ASt, SZ, MWH, SJE, ZT; Cytogenetics and fluorescence in situ hybridization: ASh, MVS, VAT, IVS; Phylogeny reconstruction and simulations: CC, MWH. Wrote paper: NJB, ZT, ABH, PAP with input from other authors.

Figures and figure legends

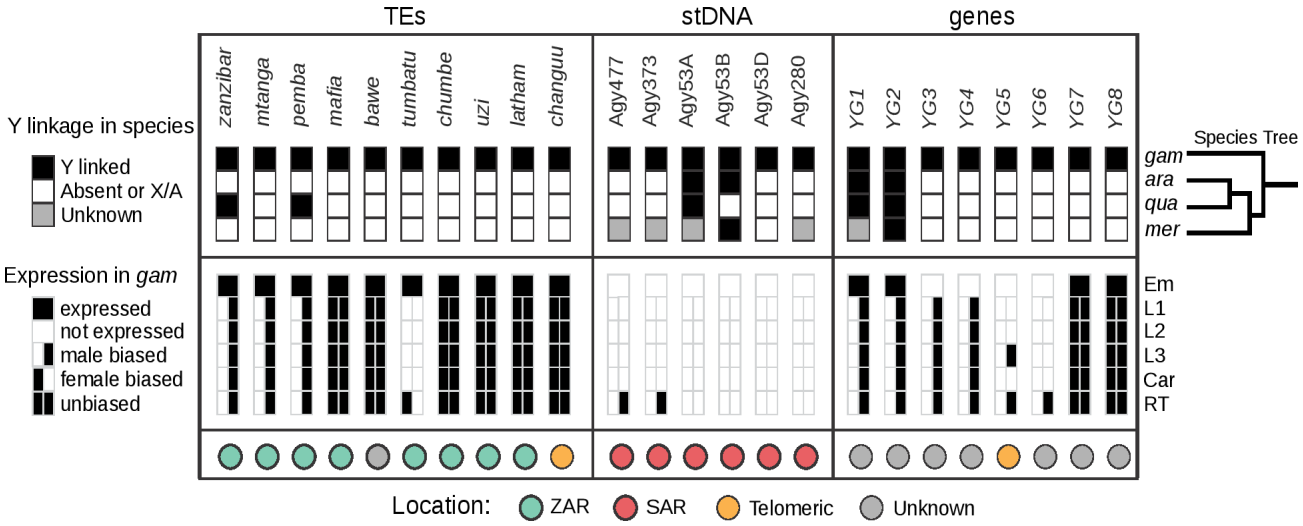

Fig. 1. Summary of major Y chromosome loci, showing rapid turnover of the Y chromosome content and expression patterns in the *Anopheles gambiae* species complex. In the **top panel**, black boxes indicate Y-linkage, white boxes indicate either total absence from the species or absence from its Y chromosome, and gray boxes indicate unknown status with regard to Y-linkage. Typically, sequences indicated by gray showed CQ or RCQ values of ~1, suggesting that they are either on both sex chromosomes, or on autosomes. Details are provided in *SI Appendix*, Tables S6-7, S15. At right, the species branching order provides an evolutionary context of the changes in Y chromosome content within the past 2 MY. Only *YG2* is conserved and exclusively on the Y chromosome in all four species of the *An. gambiae* complex. In the **middle panel**, sex-specific transcription in *An. gambiae* was assessed at different developmental stages and tissues, except for embryos (Em). The **bottom panel** shows the organization of the Y chromosome loci in *An. gambiae*, if known. TEs, transposable elements; stDNA, satellite DNA;

gam, *An. gambiae; ara*, *An. arabiensis; qua*, *An. quadriannulatus; mer*, *An. merus*. Em, embryo; L1-L3, first to third instar larvae; RT, adult reproductive tissues; Car, adult carcass.

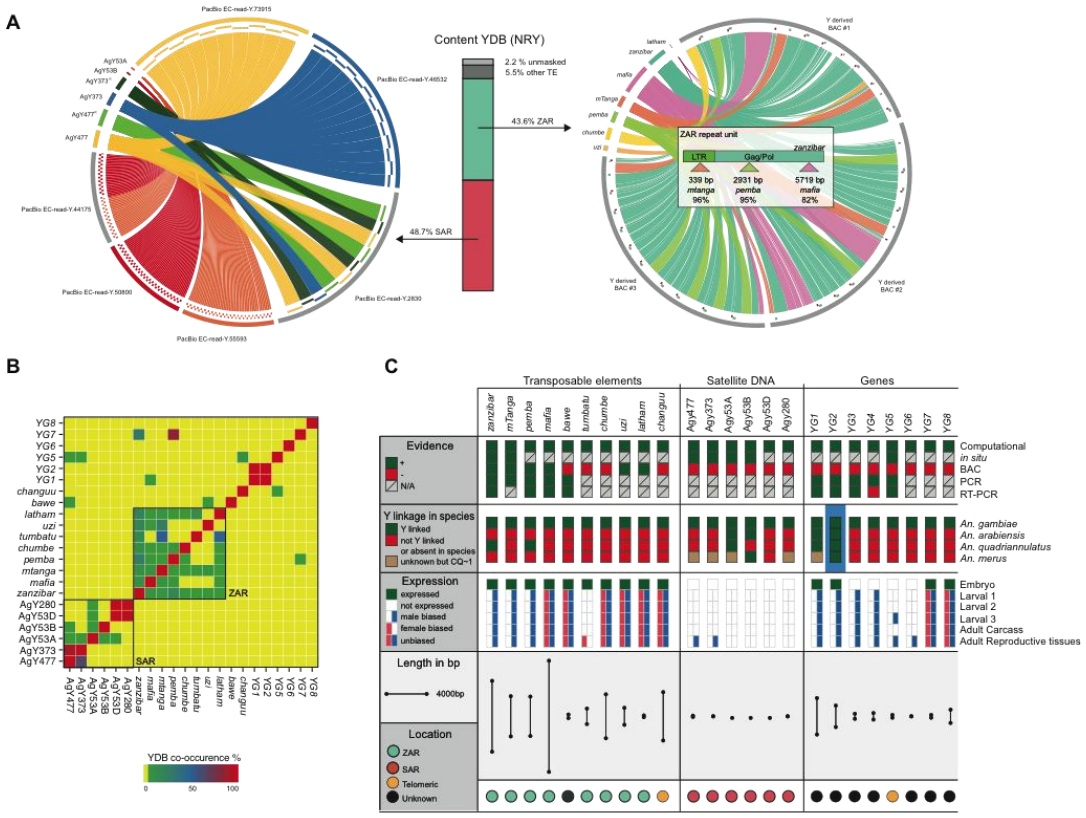

Fig. 2. The non-recombining Y (NRY) of *An. gambiae* mainly consists of massively amplified tandem arrays of a small number of satellites and transposable elements (TEs). A) Two major regions of the *An. gambiae* Y, the *zanzibar* amplified region (ZAR) and satellite amplified region (SAR), represent 92.3% of the sequences in Ydb (vertical bar plot). Ydb reflects the content of NRY in *An. gambiae*. Percentages were calculated by masking Ydb using annotated Y chromosome loci. The left circos plot, created by homology mapping of TEs on three Y

chromosome BAC clones, shows the organization of the ZAR in the three BACs. As seen in these BACs, and as independently confirmed in PacBio reads, the ZAR consists of head-to-tail tandem arrays of *zanzibar* which sometimes have other transposons inserted. The arrays of *zanzibar* units inside each BAC are shown schematically directly inside the BAC ideograms (blue semicircular lines) enclosing the circus plot. The dark green arrows of each *zanzibar* unit (shown enlarged in panel B) represent the single LTR; lines breaking *zanzibar* units indicate insertions of other TEs. A few small insertions (~200 bp) into *zanzibar* are too small to be visible in this plot. The asterisk in BAC10L19 corresponds to an atypical *zanzibar* unit that could result from recombination or misassembly. The circos plot at right, constructed by homology mapping of satellite monomers on PacBio reads from Ydb, shows the organization of the satellite amplified region (SAR). Shown are representative examples of the occurrence of homo-monomeric tandem arrays (Y73915, Y46532, Y55593), junctions between homo-monomeric tandem arrays (Y44175), and recombinant arrays (Y2830). The recombinant arrays are interspersed with recombinant and non-recombinant versions of AgY477 and AgY373 satellites (*SI Appendix*, Fig. S5). B) Schematic of a single *zanzibar* unit, consisting of a gag/pol domain and a single LTR; each unit is organized in a head-to-tail tandem array (see 2A, left circus plot). Shown by colored triangles are the canonical insertion sites of three other transposons (*mtanga*, *pemba*, *mafia*) into different *zanzibar* units. Percentages indicate the fraction of Ydb PacBio reads observed to carry TE insertions into *zanzibar* units at the precise insertion site illustrated (coordinates shown above the gag/pol domain). For example, we observed 243 of 256 (95%) PacBio reads in which *pemba* was inserted into *zanzibar* at position 2931. This phenomenon was independently confirmed in WGS Illumina reads. C) Co-occurrence matrix of Y chromosome loci in PacBio reads from Ydb. These results show that satellite sequences co-occur

(in the SAR), as do TEs (in the ZAR), but that the ZAR and SAR regions are largely independent.

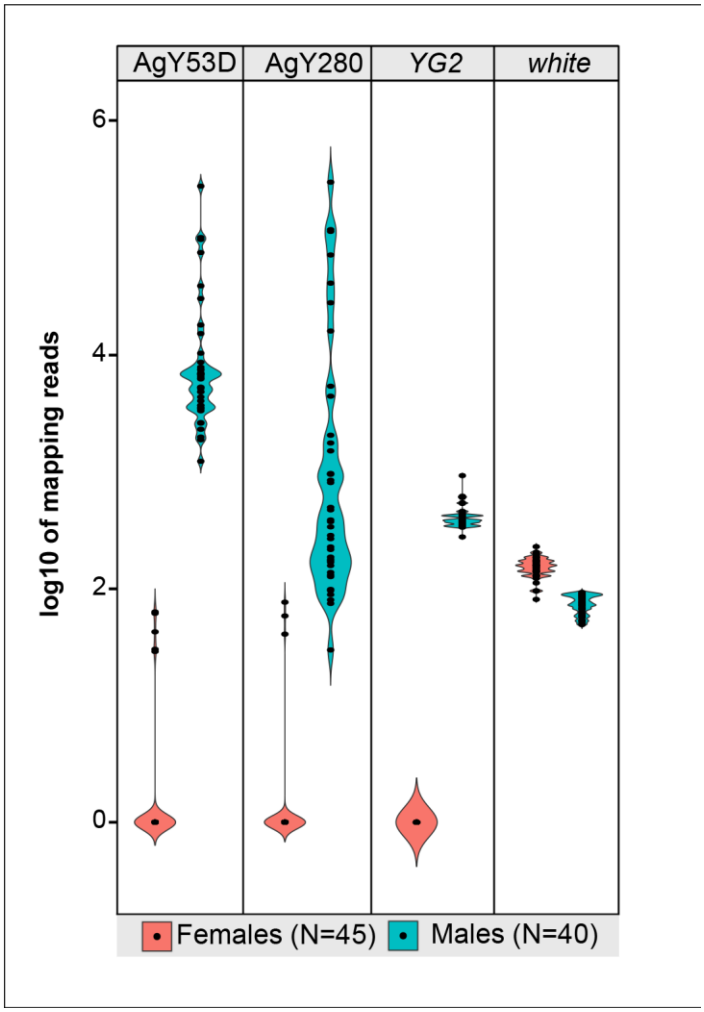

Fig. 3. Satellites AgY53D and AgY280 show extensive structural dynamism in males from a natural population of *An. gambiae*. Shown are violin plots of the $\log_{10}$ numbers of normalized read alignments from Illumina genomic sequence derived from 40 individual male (blue) and 45 individual female (pink) mosquitoes from Cameroon, mapped to satellite monomers of AgY53D and AgY280. For comparison are similar plots of reads mapping to the presumptive male-

determining gene, *YG2*, and to the single-copy X-linked gene, *white*. Numbers of read alignments to the satellite monomers varies drastically between individuals, in contrast to *YG2* and *white*, suggesting large within-population differences in satellite abundance on the Y. Mapping reads were normalized to library size and locus length.

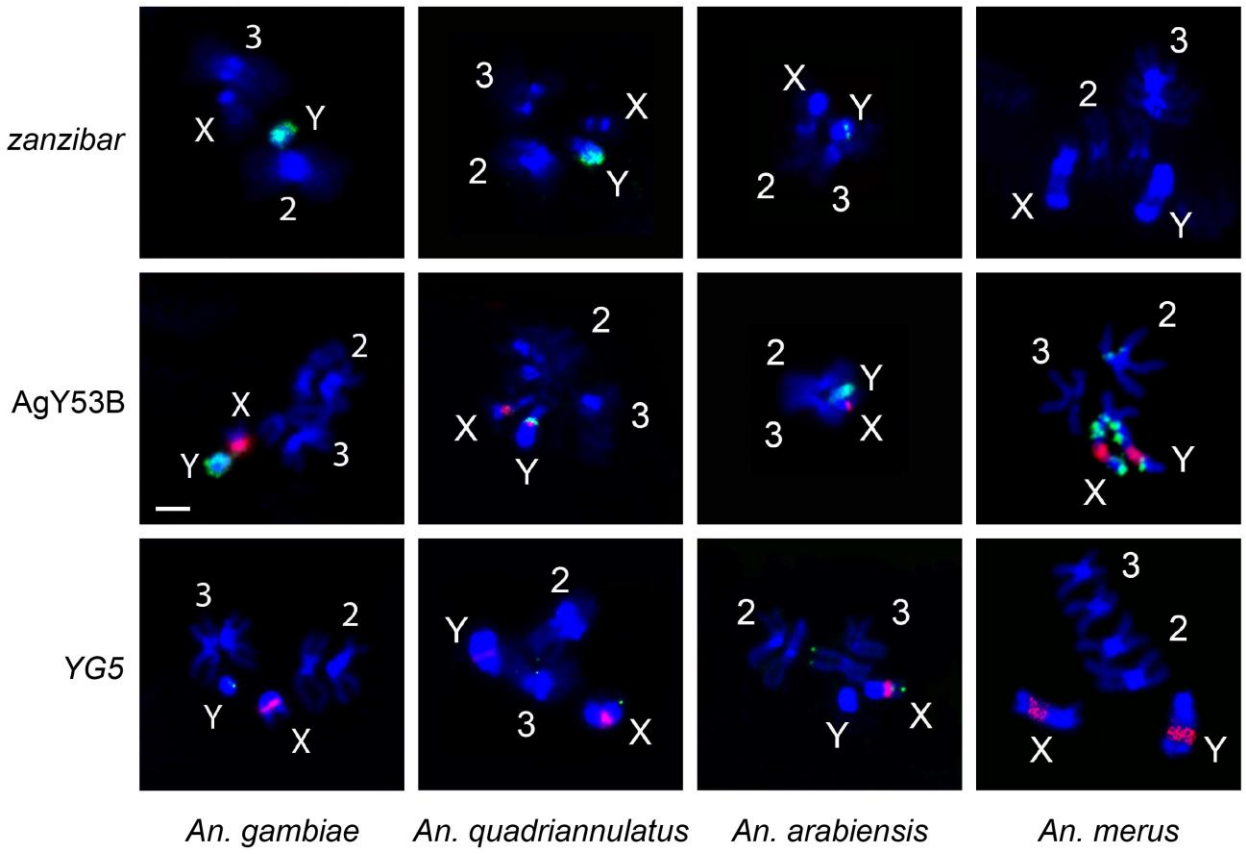

Fig. 4. Physical mapping supports structural dynamism of Y chromosome sequences in the *An. gambiae* complex. FISH of retrotransposon *zanzibar*, satellite AgY53B, and gene *YG5* (green signals) was performed on chromosomes of male *An. gambiae* Kisumu (*zanzibar*, *YG5*), *An. gambiae* Asembo (AgY53B), *An. quadriannnulatus* SANGWE, *An. arabiensis* Dongola, and *An. merus* MAF. Chromosomes were obtained from imaginal discs except for *An. merus*

chromosomes hybridized to AgY53B, which were obtained from testes. The 18S rDNA probe (red signal) was used in all experiments except with *zanzibar*. Chromosomes were counterstained with DAPI (blue). The scale bar of 2 μm applies to all images.

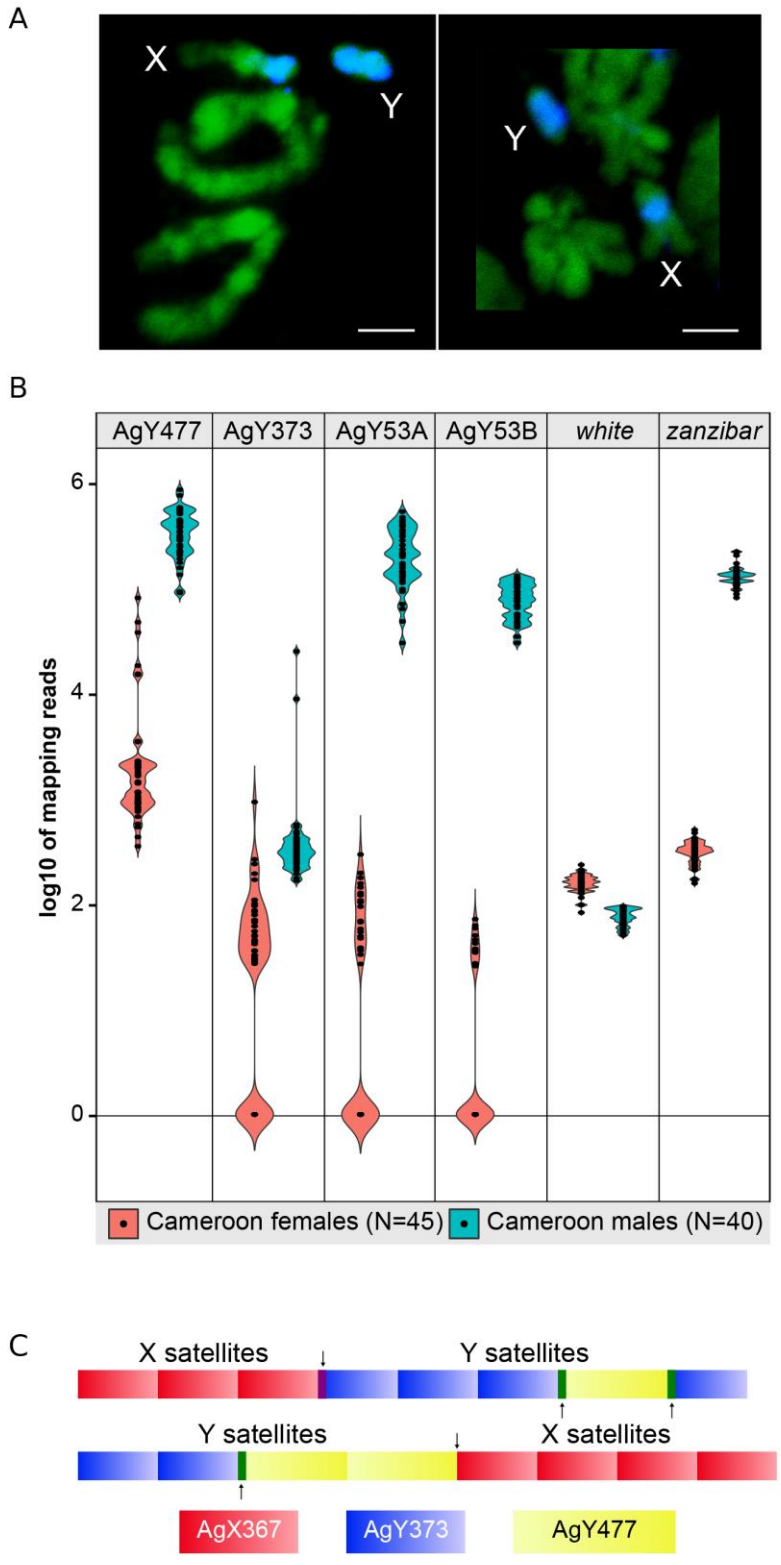

Fig. 5. The *An. gambiae* X and Y chromosomes are not genetically isolated. A) Painting of

prometaphase (left panel) and metaphase (right panel) chromosomes from male larval imaginal discs of the *An. gambiae* Pimperena strain with a probe generated from microdissected Y chromosomes (labeled blue by the WGA3 kit with dNTP-Cy3). Chromosomes are counterstained with YOYO-1 (green). Scale bar: 2 μm. B) Violin plots showing the $\log_{10}$ number of normalized read alignments from 40 individual *An. gambiae* males (blue) and 45 females (pink) from the Cameroon population, to satellite AgY477, AgY373, AgY53A and AgY53B monomers, compared to numbers of reads from these sources aligning to the *white* gene (single-copy and X-linked) and *zanzibar* (heavily Y-biased). Mapping reads were normalized to library size and locus length. C) Two examples of single PacBio reads (pacbio_7224704_1 and pacbio_5551309_1) where predominantly X-linked (AgY367; shown in red) and predominantly Y-linked (AgY373, shown in blue; AgY477, shown in yellow) satellites occur in the same PacBio read. Black arrows indicate the junction between the predominantly X-linked and predominantly Y-linked satellites. The purple and orange boxes indicate inferred recombination, between AgX367-AgY373 and AgX367-AgY477, respectively. Green boxes indicate recombination between AgY477-AgY373.

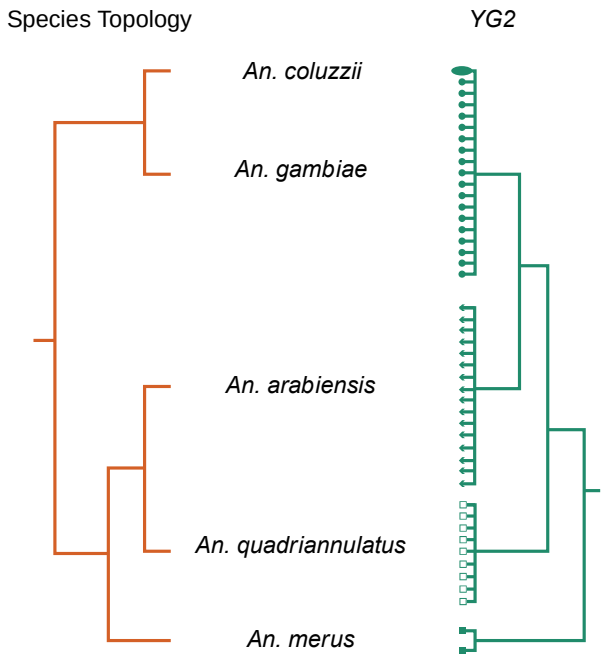

Fig. 6. Phylogeny inferred from a candidate male-determining gene on the Y chromosome, *YG2*, differs from the species branching order. Species topology (32) of five members of the *An. gambiae* complex (red branches) compared to a maximum likelihood phylogeny inferred from a Y chromosome-specific region of the *YG2* gene (green branches) sequenced from male *An. coluzzii* (ellipse), *An. gambiae* (filled circles), *An. arabiensis* (triangles), *An. quadriannulatus* (open squares) and *An. merus* (filled squares). Samples were drawn from colonies and natural populations (see text). The *YG2* tree was rooted at the midpoint; all nodes are supported by ≥ 95% bootstrap replicates. The topological disagreement involves *An. arabiensis;* in the species topology *An. arabiensis* is sister to *An. quadriannulatus*, while the *YG2* topology indicates a sister group relationship of *An. arabiensis* and *An. gambiae* +*An. coluzzii*.

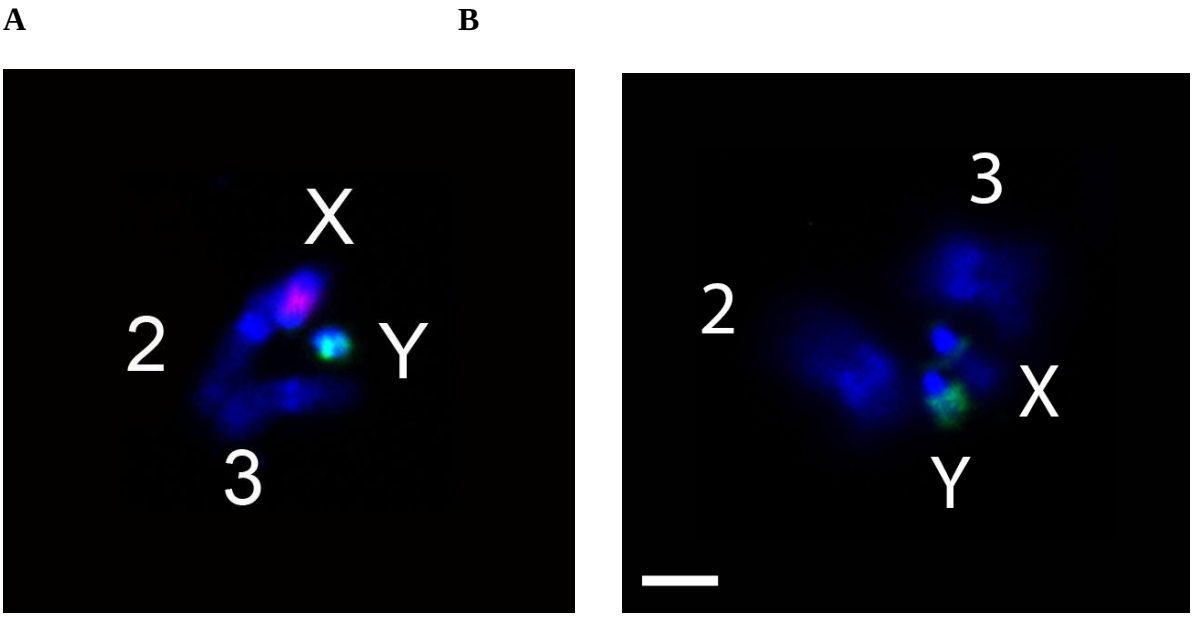

Figure 7. Physical mapping of Y chromosome-biased satellite DNA sequences in *An. gambiae.* (A) Satellite AgY53A (green signal) hybridized to most of the Y chromosome but not much to the X chromosome where the 18S rDNA probe (red signal) was seen in the *An. gambiae* Kisumu strain. (B) Satellite AgY477 (green signal) hybridized to a larger portion of the Y chromosome and to a smaller portion of the X chromosome in the *An. gambiae* Zanu strain. Chromosomes were counterstained with DAPI (blue). The scale bar is 2 μm.

Table 1. Primers designed to amplify DNA probes for FISH.

Target	Primer Name	Sequence (5'-3')
mtanga	YAR3F1	CATGGGTTTCATCCTGCTCT
	YAR3R1	TCAATACTGCCCTTCCGAAC
zanzibar	YAR2F2	TTCTTCGATGTTGTGCTGGA
	YAR2R2	ATGGAGAAACAGGGCAACAA
zanzibar	YAR4F1	ATGCATGCTTGGATTCCTTC
	YAR4R1	GGTTTCTATGATCGCCTGGA
zanzibar	YAR5F1	TTGGCATTCATCTGTCCAAA
	YAR5R1	GCACCCTTGATCTCATGTCA
YG5	*YG5*F	GACAGACCGACGGAGTAAGC
	*YG5*R	CATGCCCGAGTGCATAAGTA
AgY53A	AgY53AF2	ATGAAGAATATGGATAATGGAT
	AgY53AR3	ACGGGAGAGAGCAAGAACA
AgY53B	AgY53BF	CCTTTAAACACATGCTCAAATT
	AgY53BR	GTTTCTTCATCCTTAAAGCCTAG
AgY477	AgY477F2	TTTGAGCATGTGTTTAAAGG
	AgY477	AGGTTTTCCCGAGTACAAT

The primers for AgY53A, AgY53B, and AgY477 amplify satellite monomers as described in ref. (22). The PCR probes included fragments for a total size of ~400-500 bp for AgY53A, AgY53B and 477-900 bp for AgY477.

Table 2. Estimated size of the *An. gambiae* Y chromosome.

Male No.	**Repeated size estimates per male (Mb)**										**Mean**
1	26.62	28.05	29.96	32.73	32.93	33.22	35.17	35.54	35.64		32.21
2	29.27	30.38	36.44	36.97	38.17	44.12					35.89
3	27.42	36.38	37.59	38.56	38.99	41.18					36.69
4	33.48	33.70	35.62	35.94	36.17	47.83					37.12
5	38.87	40.84	42.00	42.45	42.51						41.33
6	43.04	43.17	43.37	45.60	46.85						44.41
7	44.28	47.52	47.61								46.47
8	25.93	27.40	30.14	30.61	31.60	33.22	33.73	33.74	34.68	38.10	31.92
Overall											36.71

Chapter 5 - Toxicological assays for testing effects of an epigenetic drug on development, fecundity and survivorship of malaria mosquitoes

AUTHORS: Atashi Sharma, Troy D. Anderson, Igor V. Sharakhov

Atashi Sharma
Department of Entomology
Virginia Tech
Blacksburg, VA 24060, USA
atashi04@vt.edu

Troy D. Anderson
Department of Entomology
Virginia Tech
Blacksburg, VA 24060, USA
anderst@vt.edu

Igor V. Sharakhov
Department of Entomology
Virginia Tech
Blacksburg, VA 24060, USA
igor@vt.edu

CORRESPONDING AUTHOR: Igor V. Sharakhov

KEYWORDS: *Anopheles gambiae*, malaria mosquito, DZNep, SAH, toxicological assay, epigenetics, vector control

5.1 Abstract:

Insecticidal resistance poses a major problem for malaria control programs. Mosquitoes adapt to a wide range of changes in the environment quickly, making malaria control an omnipresent problem in tropical countries. The emergence of insecticide resistant populations warrants the exploration of novel drug target pathways and compounds for vector mosquito control. Epigenetic drugs are well established in cancer research, however not much is known about their effects on insects. This study provides a simple protocol for examining the toxicological effects of 3-Deazaneplanocin A (DZNep), an experimental epigenetic drug for cancer therapy, on the

malaria vector, *Anopheles gambiae*. A concentration-dependent increase in mortality and decrease in size was observed in immature mosquitoes exposed to DZNep, whereas the compound reduced the fecundity of adult mosquitoes relative to control treatments. In addition, there was a drug-dependent decrease in *S*-adenosylhomocysteine (SAH) hydrolase activity in mosquitoes following exposure to DZNep relative to control treatments. These protocols provide the researcher with a simple, step-by-step procedure to assess multiple toxicological endpoints for an experimental drug and, in turn, demonstrate a unique multi-prong approach for exploring the toxicological effects of water-soluble epigenetic drugs or compounds of interest against vector mosquitoes and other insects.

5.2 Introduction:

Malaria is responsible for the highest number of insect-related deaths in the world. An estimated 219 million cases occur annually worldwide, resulting in approximately 660,000 deaths, primarily in Africa [234]. Despite concerted efforts, malaria programs face several challenges. While insecticidal treated bednets and indoor residual spraying form key components of the program, resistance to insecticides in local populations impede these efforts[235]. The rapid increase in insecticide resistant mosquito populations is largely attributed to the ability of the malaria mosquitoes to adapt quickly to changes in their environment and exploit different niches [236],[237],[238]. To overcome the existing mechanisms of insecticide resistance, the exploration of novel insecticide targets and next generation compounds is warranted. A simple, step-by-step protocol for determining the efficacy of experimental insecticides on the various life stages of malaria mosquitoes would significantly enhance these efforts.

Pharmacological studies of drug effects on cell lines and animal models have established the use of epigenetic drugs as a useful tool for modulating the genetics and physiology of cells and organisms. DNA methylation and histone modification are two epigenetic mechanisms that affect gene expression in multicellular organisms without changing the underlying DNA sequence [239]. Post translational modifications such as methylation play a crucial role in maintaining cellular integrity and gene expression, and may affect several fundamental processes [240],[241],[89]. Research in some insect species have highlighted the importance of epigenetics in processes involving oogenesis and stem cell maintenance [242], as well as dosage compensation [243] . However, such aspects in disease vectors are yet to be explored. Using a compound to modulate this system in mosquitoes may provide us with insights into the novel insecticide target pathways. 3-Deazaneplanocin A (DZNep) is a known histone methylation inhibitor, which impact on various types of cancers have been studied [102],[244],[245],[246],[247]. DZNep is a stable water-soluble epigenetic drug that indirectly inhibits histone lysine *N*-methyltransferase (EZH2), a component of the polycomb repressive complex 2 (PRC2) in mammalian cells. PRC2 plays an important role in regulating the growth of stem cells in multicellular organisms, and histone methylation is a key aspect of the PRC2 mediated gene silencing. In immunocompromised mice, cells pre-treated with DZNep have been shown to be less tumorigenic [248]. This drug is becoming used for studying other diseases, such as non-alcoholic fatty liver disease, in which EZH2 is implicated [249]. DZNep is an established *S*-adenosylhomocysteine (SAH) hydrolase inhibitor [250],[251]. The inhibition of SAH hydrolase results in an accumulation of SAH and, in turn, leads to the inhibition of methyltransferase activity by limiting available methyl donor groups. SAH is an amino acid derivative utilized by many organisms, including insects, in their metabolic pathways. A recent

study has shown that DZNep in low doses may affect diapause and delay development in insects [252].

Here, a robust protocol to investigate the effects of a water-soluble compound on various life stages of mosquitoes is developed. The three parts of this protocol include instructions for examining the effects of a water-soluble compound on immature mosquitoes, adult blood-feeding females, and enzyme activity of adult male and female mosquitoes. First, DZNep is dissolved in water to study immature mosquito development and survival. This is performed at two concentrations to compare any differences arising from 10-fold increase in drug exposure. To explore the effect of drug on adult female mosquitoes, DZNep is added to defibrillated sheep blood and fed the blood artificially to females. Subsequently, the outcome of the drug on fecundity is examined. Finally, an enzyme activity assay is performed using 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB) as an indicator to determine the effect of DZNep on SAH hydrolase inhibition in adult male and female mosquitoes. While this protocol is developed with a malaria mosquito, *Anopheles gambiae*, it can be easily adapted to studying the effects of compounds of interest on any species of mosquito or other insects. The techniques detailed in this protocol may not be efficiently applied to a drug with limited or no solubility in water or aqueous media.

5.3 Protocol:

The three parts of the protocol describe the aqueous exposure of the DZNep drug to larval mosquitoes, a blood-based exposure of DZNep to adult females to study its effect of fecundity, and SAH hydrolase inhibition by DZNep measured using a simple colorimetric technique. A schematic representation of these assays is shown in **Figure 1**.

1. Immature mosquito development and survivorship assays

This section explains the use of a water-soluble drug that affects mosquito larvae development and survival.

1.1) Hatch *An. gambiae* eggs at 28 ºC in distilled water (dH_2O) and rear them to 2nd instar larvae.

1.2) Take a six-well DNAse, RNAse free cell culture plate and pour 10 ml dH_2O to each well.

1.3) Select mosquito larvae at 2nd instar and add 15 larvae per well. Before adding 2nd instar larvae to plate, put them on a paper towel for 2-3 seconds to remove excess water.

1.4) Label wells with color-coded tapes to randomize experiment. Label 2 wells per test concentration on each plate (**Figure 2**).

1.5) Prepare a 1 mM DZNep hydrochloride (MW = 298.73) stock solution by dissolving 0.3 mg DZNep in 1ml of dH_2O. Add stock solution of DZNep.HCl to labeled wells to achieve a final concentration of 0.5 µM and 5.0 µM. Note: The stock solution can be stored at 4 ºC for short term (1 day to a month) or -20 ºC for long term (longer than one month).

Note: Addition of the stock solution for aforementioned concentrations will not change the volume of water in the wells significantly. However, in case of much higher concentrations, it is advisable to dissolve the compound in the larval medium to avoid any unnecessary dilution of the compound.

1.6) Add equal amount of flake fish food to each well. Cover the plates and incubate them at 28 ºC in an incubator.

1.7) Record the mortality for the DZNep-treated and untreated larval mosquitoes after a 24-hour exposure period. Remove and discard any dead larvae. Repeat this recording every day until

all larvae die or pupate and emerge as adults. Record the size of larvae every second day i.e. 0, 2, 4, 6, 8 day until they pupate.

1.8) Analyze the larval development and mortality data using an adequate statistical analysis software. Note: A bar graph can be used to represent the mortality data in Microsoft Excel. Multivariate Analysis of Variance (MANOVA) performed with a software such as JMP or SPSS will reveal if different doses of the drug result in significantly different survival of mosquito larvae.

2. Fecundity assay by administering drug through artificial feeder to female mosquitoes

This section details the addition of the drug to blood and feeding it to adult female mosquitoes using an artificial feeder system.

2.1) Set up adult mosquito cages with equal number of male and female mosquitoes for control and test cage. Provide 10% sugar solution soaked in cotton balls to both the cages till the females are ready for feeding (3-5 days is optimum for *Anopheles gambiae*).

2.2) 2 hours prior to the blood feeding, remove the cotton balls to ensure females are hungry.

2.3) Assemble two artificial blood feeders that consist of an inverted wide bottom glass funnel (diameter = 50 mm, length = 70 mm) and a surrounding plastic sealed with industrial glue. Attach plastic tubes (diameter = 7 mm) to two feeders' outlets (diameter= 10 mm) via connectors for inflow and outflow of water. Label the two feeders as "Control" and "5 μM DZNep."

2.4) Using a commercially available heating element, set up the program for feeding. In "Menu" go to "Settings" and select "Setpoints"; preloaded programs will appear. Select "SP1" and set the temperature for 37 ^{o}C. Note: The heating element ensures that the blood administered is maintained at a constant temperature. Additional program settings may be used for different mosquito or insect species.

2.5) Fill a bucket with water and immerse the heating element into the water.

2.6) Cut a square piece of parafilm (50 mm X 50 mm) and stretch it to make a thin membranous film. Cover the bottom of the artificial feeders with the parafilm. Cut another piece of parafilm (50 mm X 10 mm), stretch it and seal the edges of the feeders.

2.7) Assemble the system, connecting the feeders and heating element by the tubes. Once completed, switch on the system and select "SP1". Press "Enter" to start the system. Note: The water will get heated to 37 ^{o}C and maintained at that temperature.

2.8) Monitor the temperature of water using a thermometer in the water bucket. Note: The connectors and tubes help maintain a constant circulation of water through the system so that the temperature of the blood remains at 37 ^{o}C.

2.9) Fill a flat bottom tray with 10% bleach solution that will be used to decontaminate any blood spills.

2.10) Add 2 ml of defibrillated sheep's blood to a microcentrifuge tube. Label this as "Control". Repeat the process for the experimental tube labeled as "5 µM DZNep". Add 6 µl of DZNep stock solution (see step 1.5) to the experimental tube labeled "5 µM DZNep" Gently mix the blood and drug by inverting it several times. Incubate the tube 10 min at RT.

2.11) Using a pipette, add 2 ml of control and test blood to the top of correspondingly labeled feeders. Discard the pipette in the bleach solution. Cover the cage with a dark bag and breathe air on cage periodically to encourage females for feeding.

2.12) Let the mosquitoes feed for approximately 30 minutes. Once feeding is completed, switch off the heating element, take the feeders out and strip the parafilm. Soak the feeder in bleach solution for about 5 minutes and rinse well in deionized water.

2.13) Remove the feeders and put cotton balls with sugar water on the cages for 48 hours. Place an egg dish containing water and filter paper in the cages overnight to facilitate egg laying. Label each egg dish with the respective test or control concentration.

2.14) Remove the egg dish next day and examine the number and structure of eggs under a stereo microscope. Obtain images using the Q-colors5 software. Count the number of eggs. Analyze difference in the number of eggs obtained from test and control cages. Note: An unpaired t-test can be used to determine if the number of eggs are significantly different ($p \leq 0.05$).

3. Biochemical assay of mosquito enzyme inhibition by drug

This section describes an enzyme activity assay using DTNB as an indicator to determine the effect of DZNep on SAH hydrolase inhibition in adult male and female mosquitoes.

3.1) Prepare a 0.1 M Na_2HPO_4 (dibasic sodium phosphate) solution by adding 14.19 g of Na_2HPO_4 to 1 L of dH_2O. Next, prepare a 0.1 M NaH_2PO_4 (monobasic sodium phosphate) solution by adding 13.79 g of NaH_2PO_4 to 1 L of dH_2O.

3.2) Titrate the 0.1 M Na_2HPO_4 solution with the 0.1 M NaH_2PO_4 solution to pH 8.5. Note: This will be the working stock solution of 0.1 M Na_2HPO_4 (pH 8.5).

3.3) Prepare a homogenization solution of 0.1 M Na_2HPO_4 (pH 8.5) by adding 0.3% Triton X-100. Note: The homogenization solution can be stored at 4 ^oC.

3.4) Prepare a 1.6 mM SAH (MW = 384.4) solution by dissolving 6.15 mg of SAH in 10 ml of 0.1 M Na_2HPO_4 (pH 8.5).

3.5) Prepare a 1.6 mM DTNB (MW = 396.4) solution by dissolving 6.3 mg of DTNB in 10 ml of 0.1 M Na_2HPO_4 (pH 8.5). Keep the fresh DTNB solution on ice until use.

3.6) Prepare a 1 mM DZNep.HCl (MW = 298.73) solution by dissolving 0.3 mg of DZNep in 1ml dH_2O. Next, dilute the 1 mM DZNep solution into a series of concentrations from 1,000 µM to 0.002 µM in dH_2O.

3.7) To measure the SAH hydrolase inhibition by DZNep, prepare a crude enzyme extract of mosquitoes: Homogenize 10 non-blood-fed adult mosquitoes in 1 ml of ice-cold 0.1 M NaH_2PO_4 (pH 8.5), containing 0.3% Triton X-100, using a glass tissue homogenizer. Transfer the homogenate to a 1.5-ml microcentrifuge tube.

3.8) Centrifuge the homogenate for 5 min at 10,000 x *g* at 4 ^oC. Transfer the supernatant to a clean 1.5-ml microcentrifuge tube. Use the supernatant as the enzyme source for the SAH biochemical assay.

3.9) For the blank treatment (i.e., no DZNep, no enzyme), add 50 µl 1.6 mM SAH, 50 µl 1.6 mM DTNB, and 100 µl 0.1 M Na_2HPO_4 (pH 8.5) to the individual wells of a 96-well, flat bottom microplate.

3.10) For the control (i.e., no DZNep), add 50 µl 1.6 mM SAH, 50 µl 1.6 mM DTNB, 50 µl 0.1 M Na_2HPO_4 (pH 8.5) and 50 µl enzyme (SAH hydrolase obtained from crude extract) to the individual wells.

3.11) For DZNep treatments, add 50 µl 1.6 mM SAH, 50 µl 1.6 mM DTNB, 50 µl enzyme and 50 µl selected DZNep concentration to individual wells. Prepare 4 replicates per treatment and control.

3.12) Read the optical density (OD) of the SAH enzyme samples at 405 nm for 5 min at 20 sec intervals using a 96-well microplate reader.

3.13) Subtract the blank OD from control and treatment OD obtained from each well. Calculate the percent remaining SAH hydrolase activity using the following equation: % Residual Activity = (Treatment OD / Control OD) x 100.

5.4 Representative Results:
(Figures and figure legends follow).

Figure 1 is a schematic representation of the assays; it describes the various steps of the procedure listed in this article. As the protocol is based on different life stages of mosquitoes, there is no particular sequence to be followed for the experiments detailed here. The user may choose to conduct one or more assays at the same time, depending upon sample availability.

Figure 2 demonstrates the plate set up for the immature mosquito development and survivorship assays. The number of test concentrations and replicates may be altered depending upon the user's requirements and drug availability. For the immature mosquito assay, 1 mM stock solution of DZNep was added to water containing mosquitoes to achieve a final concentration of 0.5 μM and 5 μM along with a control pan with wells containing a drug-free population. The experiment demonstrated that more larvae and pupae survived in the control wells than in the DZNep-treated wells.

Figure 3 depicts the effect of DZNep on overall survivorship of malaria mosquitoes during the experiment. The results demonstrate that the epigenetic drug DZNep suppresses the growth and development and induces the mortality of immature mosquitoes. The majority of the mosquito exposed to 0.5 μM died by day 10 of the experiment, whereas a large number of the individuals exposed to 5 μM died by day 8. In contrast, the mortality of the mosquitoes in the control treatment (no drug) remained unaffected and several emerged as adults on day 8 of the

experiment. An inverse relation was observed between drug concentration and mosquito body size (**Figure 4**).

Figure 5 illustrates the effect of DZNep on mosquito fecundity. Adult female mosquitoes fed with blood containing DZNep exhibited a significant reduction in number of viable eggs. **Figure 5A** shows eggs obtained from control cage (no drug). **Figure 5B** shows the eggs obtained from test cage. A large number of eggs obtained from test cage were darker in color and lacked floats (air-filled expansions of exochorion) when compared with eggs obtained from normal blood-fed female mosquitoes (**Figure 6**). The absence of floats along with darker colored eggs indicates the potential role of epigenetics in exochorion formation.

Figure 7 demonstrates the effect of DZNep on SAH hydrolase activity in adult male and female malaria mosquitoes. A DZNep-dependent decrease in OD is observed for each drug treatment compared to the control treatment (no drug) indicating that DZNep inhibits SAH hydrolase activity.

Figures and figure legends:

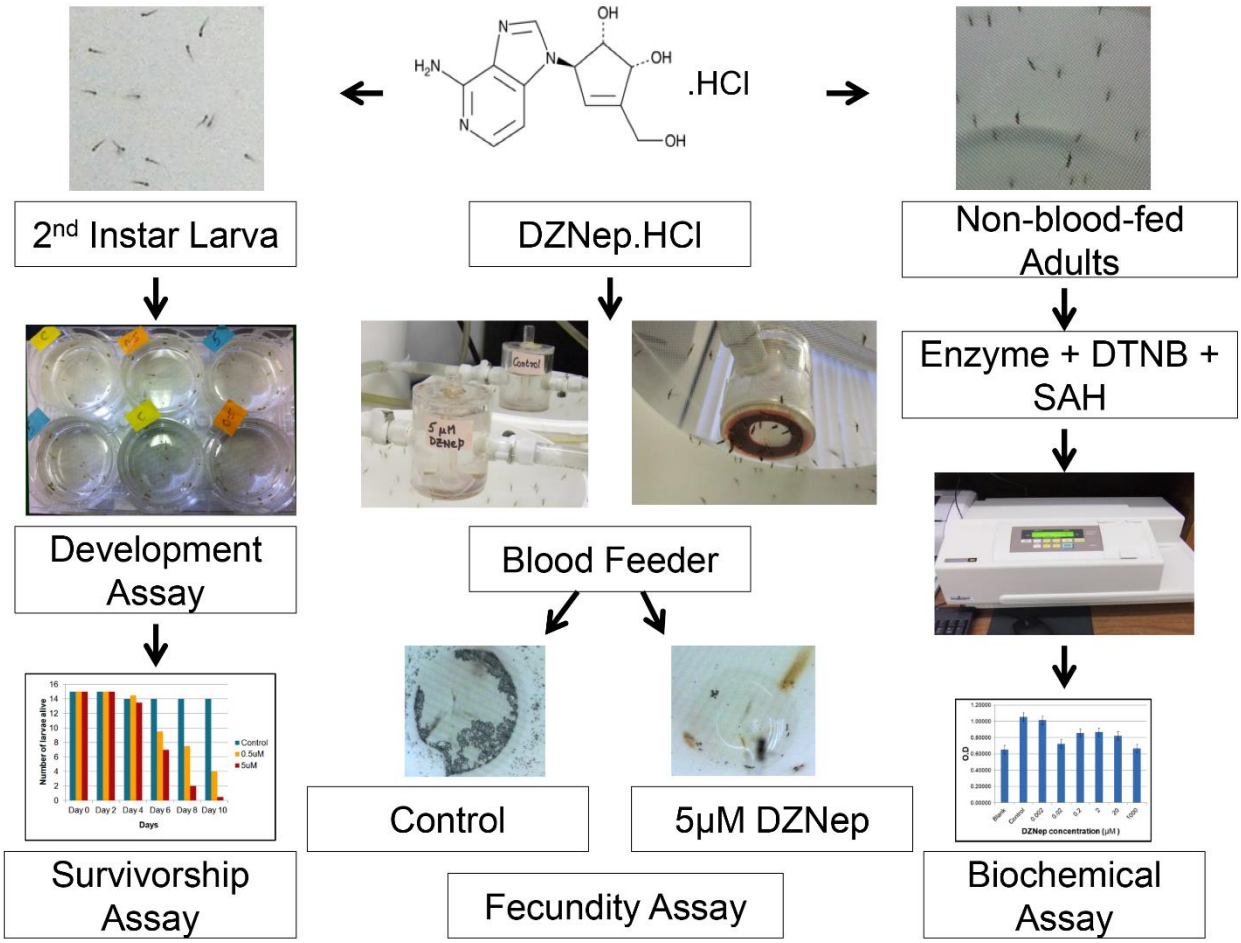

Figure 1. Schematic representation of assays using an epigenetic drug on malaria mosquitoes. The left part of the scheme shows the development and survivorship assays using immature mosquitoes. The central part demonstrates the fecundity assay via blood-feeding the drug to adult females. The right part depicts the biochemical assay of enzyme inhibition by the drug.

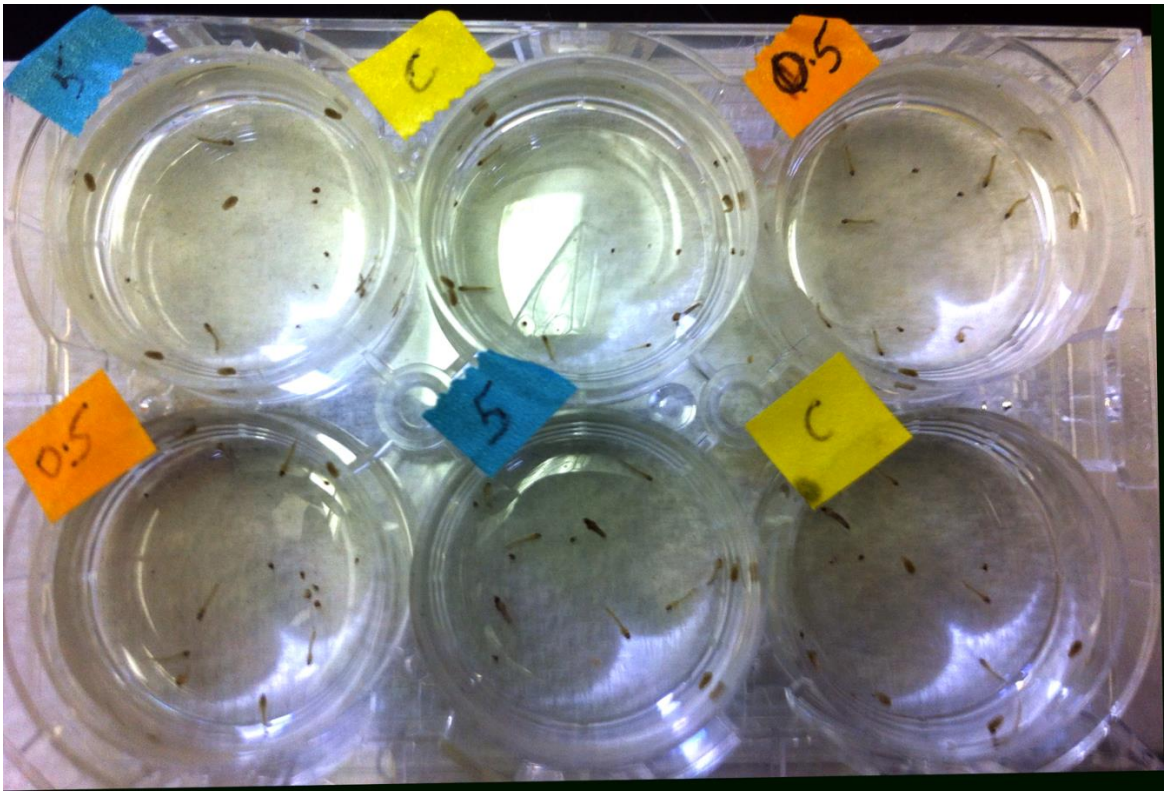

Figure 2. Plate depicting the immature mosquito development and survivorship assay. Label "C" marks control wells. Label "0.5" shows wells with 0.5 µM DZNep. Label "5" indicates wells with 5 µM DZNep.

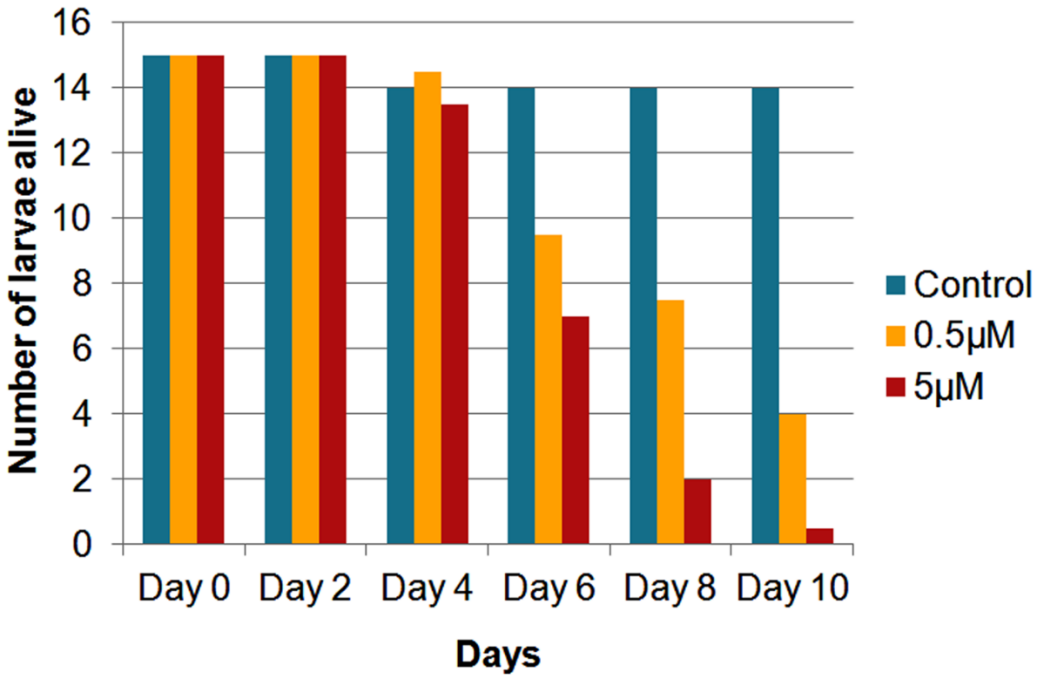

Figure 3. Effect of DZNep on survivorship of malaria mosquitoes. Concentration-dependent mortality is observed with immature mosquitoes.

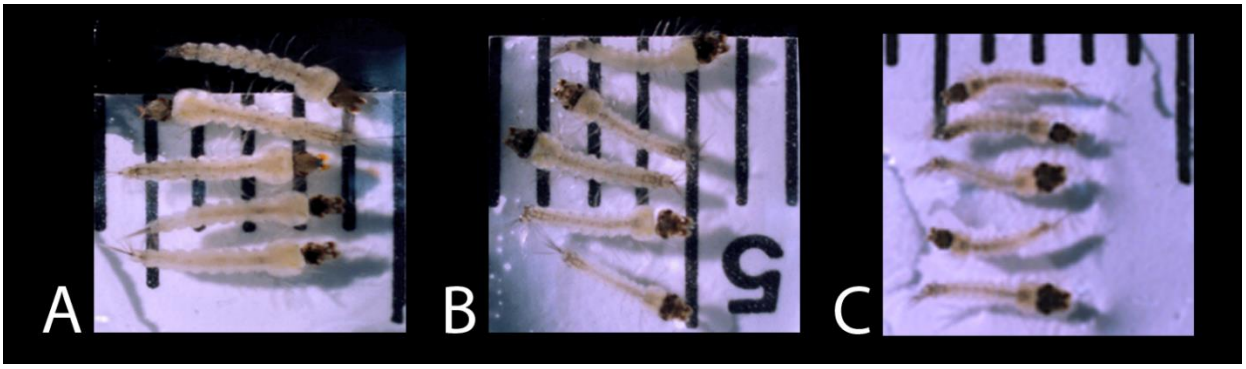

Figure 4. Effect of DZNep on development of immature mosquitoes. A) Larva from the control treatment (size = 5 mm). B) Larva treated with 0.5 µM DZNep (size = 4 mm). C) Larva treated with 5 µM DZNep (size = 3 mm). All larvae were measured on the same day of the experiment.

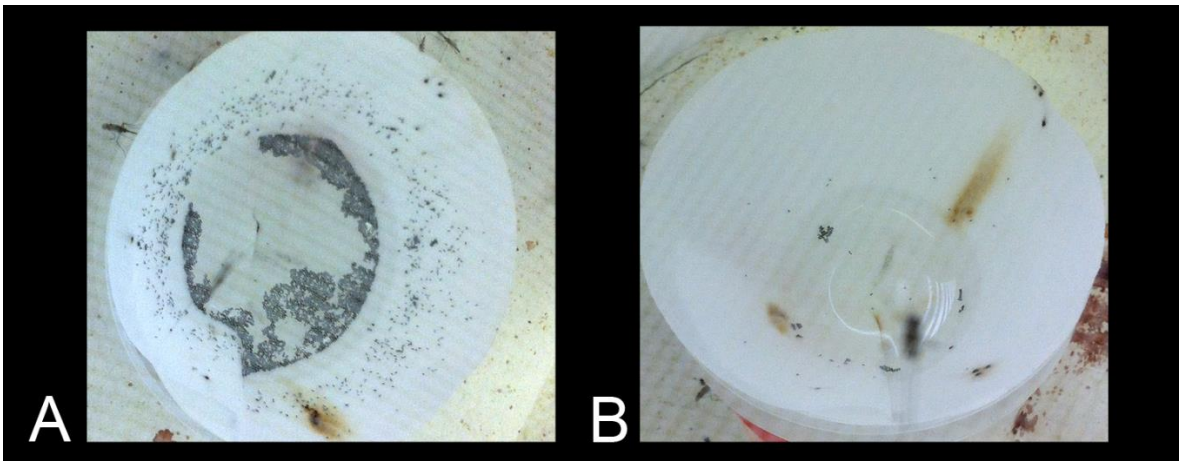

Figure 5. Effect of DZNep on mosquito fecundity. Egg dish from the control treatment (no drug) mosquitoes contains much greater number of eggs (A) compared with the egg dish of females treated with 5 µM DZNep (B).

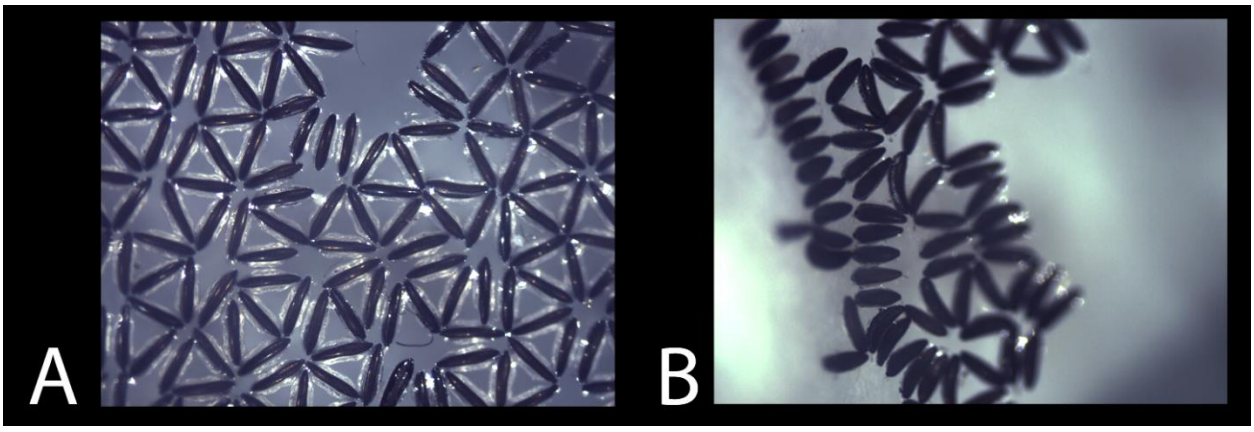

Figure 6. Effect of DZNep on egg structure in malaria mosquitoes seen under a microscope. A) Eggs from the control treatment (no drug) mosquitoes display floats and rosette-like assemblies. B) Eggs from the mosquitoes treated with 5 µM DZNep lack floats and rosette-like assemblies.

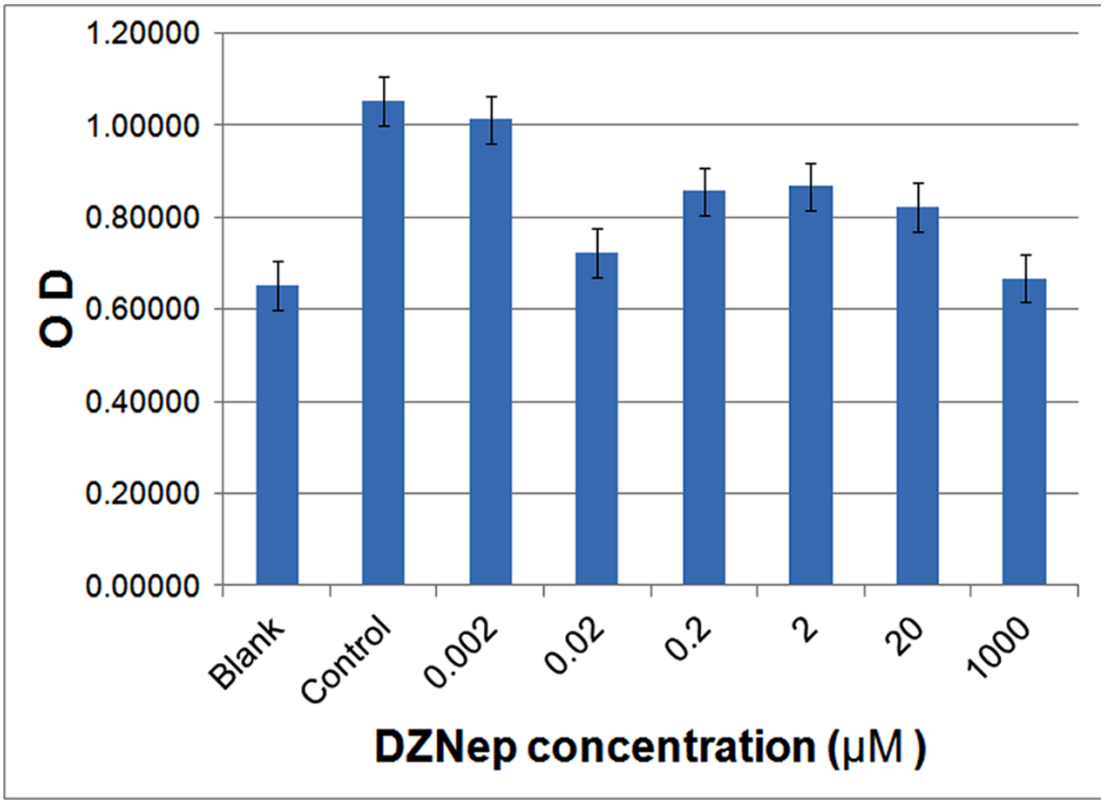

Figure 7. SAH hydrolase inhibition by DZNep in malaria mosquitoes. DZNep causes a decrease in OD compared with control treatment (no drug).

5.5 Discussion

There are several steps vital to the successful application of this protocol. For the larval assay, care should be taken to correctly label and replicate each test concentration. Randomizing test samples and adding the designated amount of drug to respective test wells is an important part of this experimental set up. Before adding 2nd instar mosquito larvae to a 96-well microplate, each larva can be put on a paper towel for 2-3 seconds to remove excess water. To prevent desiccation, larvae must be immediately transferred to the microplate wells using a blunt pair of forceps. This ensures the volume of water in each microplate well and concentration of drug remains as intended. When transferring larvae for measuring their body length, it is recommended to put the larvae on ice for few minutes in order to keep them stationary. When adapting this protocol for a different compound, care should be taken to determine the amount of

stock solution added to each well. In such cases, it might be better to dissolve the calculated amount of compound in water directly prior to adding the larvae in each microplate well.

For blood feeding assays, it is vital that mosquitoes are starved appropriately by removing sugar water at least 2 hours before of experiment. Failure to do so may result in incomplete feeding and little effect on the adult female egg production. Thorough mixing and uniform distribution of the drug in the blood is an important step as well, as DZNep takes a few minutes to dissolve. If the drug treated blood is fed immediately, DZNep does not get equally distributed in the mosquito system. It is strongly recommended that the defibrillated blood is fresh (i.e. less than a week since extraction date) for better results.

There are multiple steps critical to the success of a biochemical assay with a compound of interest. Concentrations of each stock solution used in this protocol have been optimized for DZNep. To adapt the protocol for a different compound, we suggest testing a range of different pH, SAH, DTNB concentrations and insect sample sizes to determine the most reliable results. Preparation of crude enzyme extract is the most time consuming step in this procedure. During homogenate preparation it is recommended to use not more than 10 adult mosquitoes, as lipids present in the supernatant may hinder absorbance reading. To avoid this issue, the top layer of lipids in the supernatant (using a single channel micropipette) can be removed after step 3.8. Supernatant must be transferred to a clean microfuge tube. Step 3.8 should be repeated to discard any remaining viscous lipids. The remaining clear supernatant must be pooled from all tubes and mixed gently. A multi-channel pipette can be used for increased efficiency to add SAH in the wells. DTNB is used as an indicator in the assay, and should be prepared last and kept on ice.

This protocol provides the user with simple steps to test the effects of a compound on different life stages of mosquitoes. However, there are some limitations to the use of this protocol. This technique depends primarily on the water solubility of DZNep. Immature larvae are exposed to the drug dissolved in their surrounding media. If the compound being tested is insoluble in aqueous media, it may render this protocol less efficient. For testing the fecundity, we subjected the adult females to drug dissolved in defibrillated sheep blood. This would not be possible if a lab rears mosquito colonies on mammalian blood directly (i.e., uses mice or guinea pigs for feeding mosquitoes).

The combination of different assays detailed in this article provides the user with a novel way of testing the effects of a water-soluble compound on insects. DZNep is mostly added to cell lines *in vitro* for cancer research; however, this protocol enables the user to examine any effects on various life stages of an insect *in vivo*. In addition, it also provides the researcher with step-by-step guidelines for toxicological assays. In the current literature, knowledge is limited with regard to the effects of epigenetic drugs on insects. Using DZNep as an example, we provide a procedure that may be utilized as a pre-requisite step towards exploring other epigenetic drugs or novel compounds as potential insect control agents. Although this protocol is developed for the malaria mosquito, it may be adapted to test the effect of DZNep, or any water-soluble compound of interest, in other mosquito or insect species.

5.6 Acknowledgments:
We thank Victor Marquez for providing DZNep.HCl and Scotty Bolling for manufacturing the bloodfeeders. The Mopti and SUA2La strains of *An. gambiae* were obtained from **the Malaria**

Research and Reference Reagent Resource Center (MR4). This work was supported by the Fralin Life Science Institute and the grant from National Institutes of Health 1R21AI094289 to Igor V. Sharakhov.

DISCLOSURES:
The authors have nothing to disclose.

Chapter 6 Summary

6.1 General discussion and overview

An. gambiae complex is a biological paradox comprising of species ranging from obligate anthropophagic major vectors to opportunists, minor vectors and even zoophilic non vectors of malaria. Current sequence data, biological crosses and cytogenetic information puts it as a group of eight species, however in future they may well be other members added to the complex. At the same time, information regarding each species' distribution, habitat, host- choice and time of biting differs to the extent where one or more species present in sympatry can prove to be a major obstacle in incorporating malaria programs successfully. Thus information about the different species is our most powerful weapon at present. Since the sequence of the malaria mosquito was first published [48], efforts from multiple groups have added to the sequence data and the current resolution of the reference map stands at nearly 90% [49]. However, despite many attempts the heterochromatin in *An. gambiae* is not assembled as well as the rest of the genome. The goal of this dissertation was to characterize and map the heterochromatin content among the members of *An. gambiae* complex, with special emphasis on their sex chromosomes. The wide range of attributes among the sibling species makes it a useful model to study heterochromatin. Idiograms for all the sister species within the complex were constructed and major components of heterochromatin were mapped on to these for the X chromosome, facilitating easy identification of differences in heterochromatin composition between the species. The size of Y chromosomes varies highly between and within *Anopheles* species. Mapping the repeat content of the Y chromosomes across sibling species resulted in information regarding the rapid turnover in both heterochromatin as well as euchromatin content across species. Finally, the effects of an epigenetic drug on survivorship, fecundity and

transgenerational processes were described in major vector *An. coluzzii*. Our collaborative results provide the most comprehensive mapping and characterization of sex chromosome heterochromatin in the *An. gambiae* complex and paves new ways to exploit epigenetics in the fight against malaria.

6.2 Review of Chapter 2

The polytene chromosome complement of many sibling species is comparable, yet there is a significant difference among the genome size of these species. This may be due to the difference in heterochromatin or repeat element content which differs vastly between species as a result of different recombination methods. As heterochromatin in under-polytenized, mitotic chromosomes are more suited to investigating the heterochromatin content of the different species within the *An. gambiae* complex. However, the limited number of mitotic chromosomes obtained from brain ganglia make this a labor intensive task. We developed a robust protocol for obtaining high number of mitotic chromosomes using imaginal discs from early fourth instar mosquito larvae. Imaginal discs were dissected using fine needles and squashed in appropriate media before flash freezing to obtain slides with best quality of chromosomes. Our protocol can be applied to both male and female larvae successfully to map heterochromatin repeats on mitotic chromosomes in *Anopheles* mosquitoes.

6.3 Review of Chapter 3

Cytological techniques such as Hoechst staining and C- banding had previously revealed inter- and intra- species differences between heterochromatin content of sex chromosomes between *An. gambiae* and *An. arabiensis*. However, an in depth characterization of sex chromosome heterochromatin was lacking between *An. gambiae* and *An. coluzzii*, a nascent species recently

added to the *An. gambiae* complex. A protocol for FISH with mitotic chromosomes obtained from imaginal discs was developed. Idiograms were constructed for five sibling species – *An. gambiae, An. coluzzii, An. arabiensis, An. merus and An. quadriannulatus*. 18S rDNA locus along with stDNA repeats AgY53B and AgY477 were mapped in all the species. Further, three lab strains of *An. gambiae* and *An. coluzii* were compared for stDNA position and X chromosomes were mapped using a script in MATLAB. We found that stDNA were involved in a potential inversion within heterochromatin. Currently the heterochromatin in not assembled or part of the chromosome map for *An. gambiae*. Our results will help distinguish the part of heterochromatin between the two nascent species which is involved in speciation.

6.4 Review of Chapter 4

As a part of heterogametic sex in malaria mosquitoes, the Y chromosome is the most rapidly evolving chromosome in the *An. gambiae* complex. Despite the genome sequence release of major malaria vector *An. gambiae* more than a decade ago, information regarding the Y chromosome is limited to few sequences. Based on previous cytogenetic studies, Y is estimated to be around 10% of the *An. gambiae* genome and contains the chief sex determination signal, yet information regarding Y lacks the efforts compared to other parts of the genome. Comprised of dense repeat rich heterochromatin, Y chromosome remains refractory to both sequencing and assembly. Polytene chromosomes are also not a desired tool to map the few known sequences. It is crucial to investigate the Y due to two main reasons: Evolutionary framework within repeats in malaria mosquitoes which comprise a young recently diverged (and diverging) group of insects, and transgenic mosquito control methods which constitute releasing thousands of genetically modified male mosquitoes to target a local population crash.

A large variation in Y chromosome morphology and heterochromatin content is documented based on previous studies, but the attempts to characterize molecular content of Y have been thwarted in the past due to its dense repeat nature. As a result, current *An. gambiae* assembly does not contain an assembled Y chromosome. Using mitotic chromosomes, we mapped various heterochromatin repeats on sibling species *An. gambiae, An. arabiensis, An. quadriannulatus and An. merus,* revealing species-specific patterns of different stDNA repeats. FISH with mitotic chromosomes was used to map sequences which are a) Y specific b) present on both sex chromosomes in comparative amounts and c) Y biased i.e. present on sex chromosomes but amplified on the Y. Lab strains of *An. gambiae* and sibling species *An. arabiensis, An. quadriannulatus* and *An. merus* were utilized to map the sequences of interest. Additionally, the pattern of inheritance/evolution of these elements were examined on Y chromosomes in sibling species, which reflected the rapid variation in heterochromatin between closely related species. We also mapped a novel highly amplified repeat named Zanzibar to all species except *An. merus*, in which too few copies resulted in failure of confirmation of its presence by FISH. Measurements from *An. gambiae* Pimperena Y chromosome were obtained to reveal the average size of Y to be approximatelly 36.7 Mb, suggesting that as per the current assembly Y chromosome may make up approximatelly 15% of the genome. In addition, we investigated the process of meiosis in sibling species *An. gambiae*, *An. coluzzii*, *An. arabiensis* and *An. merus*. Our data revealed that meiosis between sex chromosomes in *An. gambiae* complex depends on the species, whereby species in which X and Y are morphologically similar and possess similar amounts of heterochromatin were more likely to undergo crossing over and potential recombination than species with distinct sex chromosomes. Overall, our data reflects the high turnover rate that the *Anopheles* Y chromosome has undergone in a relatively short evolutionary

span of 2 million years. Previous cytogenetic studies were limited to the study of Y based on either simple maps showing size differences or bioinformatic methods which were not necessarily validated physically. In contrast, we combined the latest molecular biology, cytogenetic and bioinformatic approaches to identify the regions of rapid evolution in the Y chromosome within the *An. gambiae* complex. While there have been noteworthy advances in the finding of genes on the *Anopheles* Y chromosomes in recent years, our work significantly improves the Y chromosome heterochromatin content and provides a comparative map for molecular characterization of the Y within the *An. gambiae* complex. This will help further our understanding of the malaria mosquito genome and also elucidate the evolution of one of the most repeat dense chromosomes in insects. Our collective data does not merely reveal the differences between males of important vector and non-vector species, but also accounts for the source of this size variation. Our findings have important implications in evolutionary biology and potentially, in transgenic malaria control.

6.5 Review of Chapter 5

Epigenetics plays a role in many crucial processes, including cell replication, nuclear architecture etc. The three corners of malaria triangle consist of human (host) - *Plasmodium* (parasite) - mosquito (vector). While a depth of knowledge exists regarding epigenetics in humans and plasmodium, the mosquito falls short of comparable literature. To address this gap, we first identified the effects of such changes using an established histone modifier. Our goal was to determine the epigenetic effects on vital life processes of malaria mosquito. To achieve this, we used an established epigenetic drug called Deazaneplanocin-A or DZNep. It is a promising anti-cancer drug and an effective histone methyltransferase inhibitor of H3K27. By inhibiting the H3K27me3, it works antagonistically to the EZH2 component of PRC (polycomb repressive

complex) which is incriminated in several different cancers in mammals. This is especially important in case of cancer stem cells, which are special cells with tumorigenic properties and are regulated by histone modifiers such as EZH2. We first developed a protocol for using DZNep in aqueous solution to determine the effect on malaria mosquito survivorship and female fecundity. We used DZNep involved in histone H3K27 methylase transfer inhibition to determine the effects of histone modification on survivorship and fecundity of malaria mosquitoes. As low as 0.5 µM concentration had significant effects on larva mortality and pupation rate when exposed to the drug in aqueous solution. Larva body size was also affected negatively. Adult female mosquitoes administered DZNep with blood meal at 5 µM concentration showed difference in egg morphology accompanied by a significant lower hatching rate. We also characterized the SAH hydrolase inhibitor activity of DZNep in adult mosquitoes. A dose dependent decrease in optical density corresponding to a reduction in the SAH hydrolase activity was observed. A study of epigenetic changes and effects on malaria mosquitoes would further our understanding of the effects on mosquito physiology and insecticide susceptibility. Our results reveal significant effects of epigenetic changes on crucial life stages of malaria mosquito *An. coluzzii* and opens potential novel avenues for drug discovery to combat malaria.

Bibliography

1. WHO, *World Malaria Report 2014*. 2014.
2. (CDC), C.f.D.C.a.P., *Malaria Report.* 2015.
3. Grigg, M.J., et al., *Factors that are associated with the risk of acquiring Plasmodium knowlesi malaria in Sabah, Malaysia: a case-control study protocol.* BMJ Open, 2014. **4**(8): p. e006004.
4. Hostetler, J.B., et al., *A Library of Plasmodium vivax Recombinant Merozoite Proteins Reveals New Vaccine Candidates and Protein-Protein Interactions.* PLoS Negl Trop Dis, 2015. **9**(12): p. e0004264.
5. Tan, J. and P.C. Bull, *Agglutination Assays of the Plasmodium falciparum-Infected Erythrocyte.* Methods Mol Biol, 2015. **1325**: p. 115-29.
6. Tan, J., et al., *A LAIR1 insertion generates broadly reactive antibodies against malaria variant antigens.* Nature, 2015.
7. Bauza, K., et al., *Tailoring a combination pre-erythrocytic malaria vaccine.* Infect Immun, 2015.
8. Li, S., et al., *Editorial: Why Vaccines to HIV, HCV, and Malaria Have So Far Failed-Challenges to Developing Vaccines Against Immunoregulating Pathogens.* Front Microbiol, 2015. **6**: p. 1318.
9. Rts, S.C.T.P., *Efficacy and safety of the RTS,S/AS01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 African sites.* PLoS Med, 2014. **11**(7): p. e1001685.
10. Campo, J.J., et al., *Duration of vaccine efficacy against malaria: 5th year of follow-up in children vaccinated with RTS,S/AS02 in Mozambique.* Vaccine, 2014. **32**(19): p. 2209-16.
11. Sauerwein, R.W. and T. Bousema, *Transmission blocking malaria vaccines: Assays and candidates in clinical development.* Vaccine, 2015.
12. Kpanake, L., P.C. Sorum, and E. Mullet, *The potential acceptability of infant vaccination against malaria: A mapping of parental positions in Togo.* Vaccine, 2015.
13. Nunes, J.K., et al., *Development of a transmission-blocking malaria vaccine: progress, challenges, and the path forward.* Vaccine, 2014. **32**(43): p. 5531-9.
14. Kang, J.M., et al., *Limited sequence polymorphisms of four transmission-blocking vaccine candidate antigens in Plasmodium vivax Korean isolates.* Malar J, 2013. **12**: p. 144.
15. Outchkourov, N.S., et al., *Correctly folded Pfs48/45 protein of Plasmodium falciparum elicits malaria transmission-blocking immunity in mice.* Proc Natl Acad Sci U S A, 2008. **105**(11): p. 4301-5.
16. Doi, M., et al., *Worldwide sequence conservation of transmission-blocking vaccine candidate Pvs230 in Plasmodium vivax.* Vaccine, 2011. **29**(26): p. 4308-15.
17. Artzy-Randrup, Y., A.P. Dobson, and M. Pascual, *Synergistic and antagonistic interactions between bednets and vaccines in the control of malaria.* Proc Natl Acad Sci U S A, 2015. **112**(10): p. 3014-9.
18. Feng, H., et al., *Genetic diversity of transmission-blocking vaccine candidate Pvs48/45 in Plasmodium vivax populations in China.* Parasit Vectors, 2015. **8**: p. 615.
19. Mehrizi, A.A., S. Zakeri, and N.D. Djadid, *Limited genetic diversity and purifying selection in Iranian Plasmodium falciparum Generative Cell Specific 1 (PfGCS1), a potential target for transmission-blocking vaccine.* Infect Genet Evol, 2014. **22**: p. 150-6.
20. Molineaux L, G.G., *The Garki Project: research on the epidemiology and control of malaria in the Sudan Savanna of West Africa.* . 1980, World Health Organization: Geneva.
21. Glunt, K.D., et al., *Long-lasting insecticidal nets no longer effectively kill the highly resistant Anopheles funestus of southern Mozambique.* Malar J, 2015. **14**: p. 298.

22. Riveron, J.M., et al., *Rise of multiple insecticide resistance in Anopheles funestus in Malawi: a major concern for malaria vector control.* Malar J, 2015. **14**(1): p. 344.
23. Cisse, M.B., et al., *Characterizing the insecticide resistance of Anopheles gambiae in Mali.* Malar J, 2015. **14**(1): p. 327.
24. Antonio-Nkondjio, C., et al., *Rapid evolution of pyrethroid resistance prevalence in Anopheles gambiae populations from the cities of Douala and Yaounde (Cameroon).* Malar J, 2015. **14**: p. 155.
25. Wamae, P.M., et al., *Early biting of the Anopheles gambiae s.s. and its challenges to vector control using insecticide treated nets in western Kenya highlands.* Acta Trop, 2015. **150**: p. 136-42.
26. Benedict, M.Q. and A.S. Robinson, *The first releases of transgenic mosquitoes: an argument for the sterile insect technique.* Trends Parasitol, 2003. **19**(8): p. 349-55.
27. Alphey, L., et al., *Sterile-insect methods for control of mosquito-borne diseases: an analysis.* Vector Borne Zoonotic Dis, 2010. **10**(3): p. 295-311.
28. 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.
29. Harris, A.F., et al., *Field performance of engineered male mosquitoes.* Nat Biotechnol, 2011. **29**(11): p. 1034-7.
30. Gentile, J.E., S.S. Rund, and G.R. Madey, *Modelling sterile insect technique to control the population of Anopheles gambiae.* Malar J, 2015. **14**: p. 92.
31. Marinotti, O., et al., *Development of a population suppression strain of the human malaria vector mosquito, Anopheles stephensi.* Malar J, 2013. **12**: p. 142.
32. McMeniman, C.J., et al., *Stable introduction of a life-shortening Wolbachia infection into the mosquito Aedes aegypti.* Science, 2009. **323**(5910): p. 141-4.
33. Moreira, L.A., et al., *A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium.* Cell, 2009. **139**(7): p. 1268-78.
34. Hoffmann, A.A., et al., *Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission.* Nature, 2011. **476**(7361): p. 454-7.
35. Walker, T., et al., *The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations.* Nature, 2011. **476**(7361): p. 450-3.
36. Caragata, E.P., H.L. Dutra, and L.A. Moreira, *Exploiting Intimate Relationships: Controlling Mosquito-Transmitted Disease with Wolbachia.* Trends Parasitol, 2016. **32**(3): p. 207-18.
37. Dutra, H.L., et al., *From lab to field: the influence of urban landscapes on the invasive potential of Wolbachia in Brazilian Aedes aegypti mosquitoes.* PLoS Negl Trop Dis, 2015. **9**(4): p. e0003689.
38. Hughes, G.L., et al., *Native microbiome impedes vertical transmission of Wolbachia in Anopheles mosquitoes.* Proc Natl Acad Sci U S A, 2014. **111**(34): p. 12498-503.
39. Bian, G., et al., *Wolbachia invades Anopheles stephensi populations and induces refractoriness to Plasmodium infection.* Science, 2013. **340**(6133): p. 748-51.
40. Ivanescu, M.L., et al., *PCR identification of five species from the Anopheles maculipennis complex (Diptera: Culicidae) in North-Eastern Romania.* Acta Parasitol, 2015. **60**(2): p. 283-9.
41. Bangs, M.J., et al., *The mosquito Anopheles (Cellia) oreios sp. n., formerly species 6 of the Australasian Anopheles farauti complex, and a critical review of its biology and relation to disease.* Med Vet Entomol, 2015. **29**(1): p. 68-81.
42. Logue, K., et al., *Whole-genome sequencing reveals absence of recent gene flow and separate demographic histories for Anopheles punctulatus mosquitoes in Papua New Guinea.* Mol Ecol, 2015. **24**(6): p. 1263-74.
43. della Torre, A., et al., *Molecular evidence of incipient speciation within Anopheles gambiae s.s. in West Africa.* Insect Mol Biol, 2001. **10**(1): p. 9-18.

44. Coetzee, M., et al., *Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex.* Zootaxa, 2013. **3619**(3): p. 246-274.
45. Davidson, G., *The Five Mating-Types in the Anopheles Gambiae Complex.* Riv Malariol, 1964. **43**: p. 167-83.
46. Coluzzi, M.S., A., *Cytogenetic observations on species A and B of the Anopheles gambiae complex.* Parassitologia 1967. **9**(2): p. 73-88.
47. Massebo, F., et al., *Blood meal origins and insecticide susceptibility of Anopheles arabiensis from Chano in South-West Ethiopia.* Parasit Vectors, 2013. **6**: p. 44.
48. Holt, R.A., et al., *The genome sequence of the malaria mosquito Anopheles gambiae.* Science, 2002. **298**(5591): p. 129-49.
49. Sharakhova, M.V., et al., *Update of the Anopheles gambiae PEST genome assembly.* Genome Biol, 2007. **8**(1): p. R5.
50. Sharakhova, M.V., et al., *Genome mapping and characterization of the Anopheles gambiae heterochromatin.* BMC Genomics, 2010. **11**: p. 459.
51. Krzywinski, J., et al., *Isolation and characterization of Y chromosome sequences from the African malaria mosquito Anopheles gambiae.* Genetics, 2004. **166**(3): p. 1291-302.
52. Krzywinski, J., D. Sangare, and N.J. Besansky, *Satellite DNA from the Y chromosome of the malaria vector Anopheles gambiae.* Genetics, 2005. **169**(1): p. 185-96.
53. Oberdoerffer, P. and D.A. Sinclair, *The role of nuclear architecture in genomic instability and ageing.* Nat Rev Mol Cell Biol, 2007. **8**(9): p. 692-702.
54. Dernburg, A.F., et al., *Perturbation of nuclear architecture by long-distance chromosome interactions.* Cell, 1996. **85**(5): p. 745-59.
55. Rose, M.R. and W.F. Doolittle, *Molecular biological mechanisms of speciation.* Science, 1983. **220**(4593): p. 157-62.
56. Charlesworth, B., P. Jarne, and S. Assimacopoulos, *The distribution of transposable elements within and between chromosomes in a population of Drosophila melanogaster. III. Element abundances in heterochromatin.* Genet Res, 1994. **64**(3): p. 183-97.
57. Ferree, P.M. and S. Prasad, *How can satellite DNA divergence cause reproductive isolation? Let us count the chromosomal ways.* Genet Res Int, 2012. **2012**: p. 430136.
58. Sun, F.L., et al., *The fourth chromosome of Drosophila melanogaster: interspersed euchromatic and heterochromatic domains.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5340-5.
59. Ferree, P.M. and D.A. Barbash, *Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila.* PLoS Biol, 2009. **7**(10): p. e1000234.
60. Satyaki, P.R., et al., *The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats.* PLoS Genet, 2014. **10**(3): p. e1004240.
61. Brideau, N.J., et al., *Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila.* Science, 2006. **314**(5803): p. 1292-5.
62. Cattani, M.V. and D.C. Presgraves, *Genetics and lineage-specific evolution of a lethal hybrid incompatibility between Drosophila mauritiana and its sibling species.* Genetics, 2009. **181**(4): p. 1545-55.
63. Cattani, M.V., S.B. Kingan, and D.C. Presgraves, *Cis- and trans-acting genetic factors contribute to heterogeneity in the rate of crossing over between the Drosophila simulans clade species.* J Evol Biol, 2012. **25**(10): p. 2014-22.
64. Cattani, M.V. and D.C. Presgraves, *Incompatibility between X chromosome factor and pericentric heterochromatic region causes lethality in hybrids between Drosophila melanogaster and its sibling species.* Genetics, 2012. **191**(2): p. 549-59.

65. Ferree, P.M., *Mitotic misbehavior of a Drosophila melanogaster satellite in ring chromosomes: insights into intragenomic conflict among heterochromatic sequences.* Fly (Austin), 2014. **8**(2): p. 101-7.
66. Gilliland, W.D., et al., *Normal segregation of a foreign-species chromosome during Drosophila female meiosis despite extensive heterochromatin divergence.* Genetics, 2015. **199**(1): p. 73-83.
67. Panzera, F., et al., *Evolutionary and dispersal history of Triatoma infestans, main vector of Chagas disease, by chromosomal markers.* Infect Genet Evol, 2014. **27**: p. 105-13.
68. Guerra, A.L., et al., *Distribution of constitutive heterochromatin in species of triatomines with fragmentation of sex chromosomes X.* Genet Mol Res, 2014. **13**(4): p. 10279-84.
69. Feliciello, I., G. Chinali, and D. Ugarkovic, *Structure and population dynamics of the major satellite DNA in the red flour beetle Tribolium castaneum.* Genetica, 2011. **139**(8): p. 999-1008.
70. Feliciello, I., et al., *Satellite DNA as a driver of population divergence in the red flour beetle Tribolium castaneum.* Genome Biol Evol, 2015. **7**(1): p. 228-39.
71. Sharakhova, M.V. and I.V. Sharakhov, *Organization and evolution of heterochromatin in malaria mosquitoes.* Genetika, 2010. **46**(10): p. 1417-20.
72. Sharakhov, I.V. and M.V. Sharakhova, *Chromosome evolution in malaria mosquitoes.* Genetika, 2010. **46**(9): p. 1250-3.
73. della Torre, A., et al., *Speciation within Anopheles gambiae--the glass is half full.* Science, 2002. **298**(5591): p. 115-7.
74. Lehmann, T. and A. Diabate, *The molecular forms of Anopheles gambiae: a phenotypic perspective.* Infect Genet Evol, 2008. **8**(5): p. 737-46.
75. Tripet, F., et al., *DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of Anopheles gambiae.* Mol Ecol, 2001. **10**(7): p. 1725-32.
76. Pombi, M., et al., *Chromosomal plasticity and evolutionary potential in the malaria vector Anopheles gambiae sensu stricto: insights from three decades of rare paracentric inversions.* BMC Evol Biol, 2008. **8**: p. 309.
77. Turner, T.L., M.W. Hahn, and S.V. Nuzhdin, *Genomic islands of speciation in Anopheles gambiae.* PLoS Biol, 2005. **3**(9): p. e285.
78. White, B.J., et al., *Genetic association of physically unlinked islands of genomic divergence in incipient species of Anopheles gambiae.* Mol Ecol, 2010. **19**(5): p. 925-39.
79. Nwakanma, D.C., et al., *Breakdown in the process of incipient speciation in Anopheles gambiae.* Genetics, 2013. **193**(4): p. 1221-31.
80. Lawniczak, M.K., et al., *Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences.* Science, 2010. **330**(6003): p. 512-4.
81. Neafsey, D.E., et al., *SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes.* Science, 2010. **330**(6003): p. 514-7.
82. Aboagye-Antwi, F., et al., *Experimental swap of Anopheles gambiae's assortative mating preferences demonstrates key role of X-chromosome divergence island in incipient sympatric speciation.* PLoS Genet, 2015. **11**(4): p. e1005141.
83. Coetzee, M., et al., *Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles gambiae complex.* Zootaxa, 2013. **3619**: p. 246-74.
84. Bonaccorsi, S., et al., *Intraspecific polymorphism of sex chromosome heterochromatin in two species of the Anopheles gambiae complex.* Chromosoma, 1980. **76**(1): p. 57-64.
85. Sharakhova, M.V., I.V. Sharakhov, and V.N. Stegnii, *[Chromocenter organization in salivary glands of the malaria mosquito Anopheles messeae Fall].* Genetika, 1997. **33**(2): p. 196-201.
86. Hall, A.B., et al., *Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females.* BMC Genomics, 2013. **14**: p. 273.

87. Gatti, M., et al., *Fluorescence banding techniques in the identification of sibling species of the anopheles gambiae complex.* Heredity (Edinb), 1977. **38**(1): p. 105-8.
88. Waddington, C.H., *The epigenotype. 1942.* Int J Epidemiol, 2012. **41**(1): p. 10-3.
89. Dawson, M.A. and T. Kouzarides, *Cancer epigenetics: from mechanism to therapy.* Cell, 2012. **150**(1): p. 12-27.
90. Dawson, M.A., T. Kouzarides, and B.J. Huntly, *Targeting epigenetic readers in cancer.* N Engl J Med, 2012. **367**(7): p. 647-57.
91. Meier, K. and A. Brehm, *Chromatin regulation: how complex does it get?* Epigenetics, 2014. **9**(11): p. 1485-95.
92. Chen, Y.A. and A.A. Aravin, *Non-Coding RNAs in Transcriptional Regulation: The review for.* Curr Mol Biol Rep, 2015. **1**(1): p. 10-18.
93. Turinetto, V. and C. Giachino, *Histone variants as emerging regulators of embryonic stem cell identity.* Epigenetics, 2015. **10**(7): p. 563-73.
94. Jenkins, A.M. and M.A. Muskavitch, *Evolution of an epigenetic gene ensemble within the genus Anopheles.* Genome Biol Evol, 2015. **7**(3): p. 901-15.
95. Blaze, J., A. Asok, and T.L. Roth, *Long-term effects of early-life caregiving experiences on brain-derived neurotrophic factor histone acetylation in the adult rat mPFC.* Stress, 2015: p. 1-9.
96. Senut, M.C., et al., *Epigenetics of early-life lead exposure and effects on brain development.* Epigenomics, 2012. **4**(6): p. 665-74.
97. Babenko, O., I. Kovalchuk, and G.A. Metz, *Stress-induced perinatal and transgenerational epigenetic programming of brain development and mental health.* Neurosci Biobehav Rev, 2015. **48**: p. 70-91.
98. Aarabi, M., et al., *High dose folic acid supplementation alters the human sperm methylome and is influenced by the MTHFR C677T polymorphism.* Hum Mol Genet, 2015.
99. Rando, T.A. and H.Y. Chang, *Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock.* Cell, 2012. **148**(1-2): p. 46-57.
100. Tammen, S.A., et al., *Aging alters hepatic DNA hydroxymethylation, as measured by liquid chromatography/mass spectrometry.* J Cancer Prev, 2014. **19**(4): p. 301-8.
101. Grammatikakis, I., et al., *Long noncoding RNAs(lncRNAs) and the molecular hallmarks of aging.* Aging (Albany NY), 2014. **6**(12): p. 992-1009.
102. Miranda, T.B., et al., *DZNep is a global histone methylation inhibitor that reactivates developmental genes not silenced by DNA methylation.* Mol Cancer Ther, 2009. **8**(6): p. 1579-88.
103. Gayacharan and A.J. Joel, *Epigenetic responses to drought stress in rice (Oryza sativa L.).* Physiol Mol Biol Plants, 2013. **19**(3): p. 379-87.
104. Lieberman-Lazarovich, M., et al., *Epigenetic alterations at genomic loci modified by gene targeting in Arabidopsis thaliana.* PLoS One, 2013. **8**(12): p. e85383.
105. Bu, Z., et al., *Regulation of arabidopsis flowering by the histone mark readers MRG1/2 via interaction with CONSTANS to modulate FT expression.* PLoS Genet, 2014. **10**(9): p. e1004617.
106. Perduns, R., I. Horst-Niessen, and C. Peterhansel, *Photosynthetic Genes and Genes Associated with the C4 Trait in Maize Are Characterized by a Unique Class of Highly Regulated Histone Acetylation Peaks on Upstream Promoters.* Plant Physiol, 2015. **168**(4): p. 1378-88.
107. Bonasio, R., *The expanding epigenetic landscape of non-model organisms.* J Exp Biol, 2015. **218**(Pt 1): p. 114-22.
108. Kim, H.J., P. Koedrith, and Y.R. Seo, *Ecotoxicogenomic approaches for understanding molecular mechanisms of environmental chemical toxicity using aquatic invertebrate, Daphnia model organism.* Int J Mol Sci, 2015. **16**(6): p. 12261-87.
109. Chen, S., et al., *Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle.* Nat Genet, 2014. **46**(3): p. 253-60.

110. Robb, S.M. and A. Sanchez Alvarado, *Histone modifications and regeneration in the planarian Schmidtea mediterranea.* Curr Top Dev Biol, 2014. **108**: p. 71-93.
111. Lee, M.C. and A.C. Spradling, *The progenitor state is maintained by lysine-specific demethylase 1-mediated epigenetic plasticity during Drosophila follicle cell development.* Genes Dev, 2014. **28**(24): p. 2739-49.
112. Xiang, H., et al., *Comparative methylomics between domesticated and wild silkworms implies possible epigenetic influences on silkworm domestication.* BMC Genomics, 2013. **14**: p. 646.
113. Ruden, D.M., et al., *Epigenetics as an answer to Darwin's "special difficulty," Part 2: natural selection of metastable epialleles in honeybee castes.* Front Genet, 2015. **6**: p. 60.
114. Herb, B.R., *Epigenetics as an answer to Darwin's "special difficulty".* Front Genet, 2014. **5**: p. 321.
115. Simola, D.F., et al., *Epigenetic (re)programming of caste-specific behavior in the ant Camponotus floridanus.* Science, 2016. **351**(6268).
116. Park, P.J., *ChIP-seq: advantages and challenges of a maturing technology.* Nat Rev Genet, 2009. **10**(10): p. 669-80.
117. Segal, E. and J. Widom, *What controls nucleosome positions?* Trends Genet, 2009. **25**(8): p. 335-43.
118. Farnham, P.J., *Insights from genomic profiling of transcription factors.* Nat Rev Genet, 2009. **10**(9): p. 605-16.
119. Hosomichi, K., et al., *The impact of next-generation sequencing technologies on HLA research.* J Hum Genet, 2015.
120. Flueck, C. and D.A. Baker, *Malaria parasite epigenetics: when virulence and romance collide.* Cell Host Microbe, 2014. **16**(2): p. 148-50.
121. Dennison, N.J., O.J. BenMarzouk-Hidalgo, and G. Dimopoulos, *MicroRNA-regulation of Anopheles gambiae immunity to Plasmodium falciparum infection and midgut microbiota.* Dev Comp Immunol, 2015. **49**(1): p. 170-8.
122. Brancucci, N.M., et al., *Heterochromatin protein 1 secures survival and transmission of malaria parasites.* Cell Host Microbe, 2014. **16**(2): p. 165-76.
123. Karmodiya, K., et al., *A comprehensive epigenome map of Plasmodium falciparum reveals unique mechanisms of transcriptional regulation and identifies H3K36me2 as a global mark of gene suppression.* Epigenetics Chromatin, 2015. **8**: p. 32.
124. Sharma, P., et al., *An epigenetic antimalarial resistance mechanism involving parasite genes linked to nutrient uptake.* J Biol Chem, 2013. **288**(27): p. 19429-40.
125. Lambrechts, L., *Dissecting the genetic architecture of host-pathogen specificity.* PLoS Pathog, 2010. **6**(8): p. e1001019.
126. Martinez-Barnetche, J., et al., *Transcriptome of the adult female malaria mosquito vector Anopheles albimanus.* BMC Genomics, 2012. **13**: p. 207.
127. Gomez-Diaz, E., et al., *Insights into the epigenomic landscape of the human malaria vector Anopheles gambiae.* Front Genet, 2014. **5**: p. 277.
128. Macias, V., et al., *piRNA pathway gene expression in the malaria vector mosquito Anopheles stephensi.* Insect Mol Biol, 2014. **23**(5): p. 579-86.
129. George, P., et al., *Increased production of piRNAs from euchromatic clusters and genes in Anopheles gambiae compared with Drosophila melanogaster.* Epigenetics Chromatin, 2015. **8**: p. 50.
130. Oppold, A., et al., *Epigenetic alterations and decreasing insecticide sensitivity of the Asian tiger mosquito Aedes albopictus.* Ecotoxicol Environ Saf, 2015. **122**: p. 45-53.
131. Jones, P.A. and S.B. Baylin, *The epigenomics of cancer.* Cell, 2007. **128**(4): p. 683-92.
132. Shen, H. and P.W. Laird, *Interplay between the cancer genome and epigenome.* Cell, 2013. **153**(1): p. 38-55.

133. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nat Rev Genet, 2002. **3**(6): p. 415-28.
134. Fahrner, J.A., et al., *Dependence of histone modifications and gene expression on DNA hypermethylation in cancer.* Cancer Res, 2002. **62**(24): p. 7213-8.
135. Ahuja, N., A.R. Sharma, and S.B. Baylin, *Epigenetic Therapeutics: A New Weapon in the War Against Cancer.* Annu Rev Med, 2016. **67**: p. 73-89.
136. Sha, M., et al., *DZNep inhibits the proliferation of colon cancer HCT116 cells by inducing senescence and apoptosis.* Acta Pharm Sin B, 2015. **5**(3): p. 188-93.
137. Cheng, L.L., et al., *TP53 genomic status regulates sensitivity of gastric cancer cells to the histone methylation inhibitor 3-deazaneplanocin A (DZNep).* Clin Cancer Res, 2012. **18**(15): p. 4201-12.
138. Kikuchi, J., et al., *Epigenetic therapy with 3-deazaneplanocin A, an inhibitor of the histone methyltransferase EZH2, inhibits growth of non-small cell lung cancer cells.* Lung Cancer, 2012. **78**(2): p. 138-43.
139. Xia, A., M.V. Sharakhova, and I.V. Sharakhov, *Reconstructing ancestral autosomal arrangements in the Anopheles gambiae complex.* J Comput Biol, 2008. **15**(8): p. 965-80.
140. Holt, R.A., *The Genome Sequence of the Malaria Mosquito Anopheles gambiae.* Science, 2002. **298**(5591): p. 129-149.
141. 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.
142. 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.
143. Chambers, E.W., et al., *Microsatellite isolation and linkage group identification in the yellow fever mosquito Aedes aegypti.* J Hered, 2007. **98**(3): p. 202-10.
144. Nene, V., et al., *Genome sequence of Aedes aegypti, a major arbovirus vector.* Science, 2007. **316**(5832): p. 1718-23.
145. Arensburger, P., et al., *Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics.* Science, 2010. **330**(6000): p. 86-8.
146. 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.
147. Brown, S.E., et al., *Toward a physical map of Aedes aegypti.* Insect Mol Biol, 1995. **4**(3): p. 161-7.
148. Bhatt, S., et al., *The global distribution and burden of dengue.* Nature, 2013. **496**(7446): p. 504-7.
149. Kumar, A. and K.S. Rai, *Chromosomal localization and copy number of 18S + 28S ribosomal RNA genes in evolutionarily diverse mosquitoes (Diptera, Culicidae).* Hereditas, 1990. **113**(3): p. 277-89.
150. Brown, S.E. and D.L. Knudson, *FISH landmarks for Aedes aegypti chromosomes.* Insect Mol Biol, 1997. **6**(2): p. 197-202.
151. Brown, S.E., et al., *Integration of the Aedes aegypti mosquito genetic linkage and physical maps.* Genetics, 2001. **157**(3): p. 1299-305.
152. Wilkins, E.E., P.I. Howell, and M.Q. Benedict, *IMP PCR primers detect single nucleotide polymorphisms for Anopheles gambiae species identification, Mopti and Savanna rDNA types, and resistance to dieldrin in Anopheles arabiensis.* Malar J, 2006. **5**: p. 125.
153. Santolamazza, F., et al., *Insertion polymorphisms of SINE200 retrotransposons within speciation islands of Anopheles gambiae molecular forms.* Malar J, 2008. **7**: p. 163.
154. Gimonneau, G., et al., *Larval habitat segregation between the molecular forms of the mosquito Anopheles gambiae in a rice field area of Burkina Faso, West Africa.* Med Vet Entomol, 2012. **26**(1): p. 9-17.

155. Caputo, B., et al., *Prominent intraspecific genetic divergence within Anopheles gambiae sibling species triggered by habitat discontinuities across a riverine landscape.* Mol Ecol, 2014. **23**(18): p. 4574-89.
156. Diabate, A., et al., *Evidence for divergent selection between the molecular forms of Anopheles gambiae: role of predation.* BMC Evol Biol, 2008. **8**: p. 5.
157. Gimonneau, G., et al., *Behavioural responses of Anopheles gambiae sensu stricto M and S molecular form larvae to an aquatic predator in Burkina Faso.* Parasit Vectors, 2012. **5**: p. 65.
158. Cassone, B.J., et al., *Gene expression divergence between malaria vector sibling species Anopheles gambiae and An. coluzzii from rural and urban Yaounde Cameroon.* Mol Ecol, 2014.
159. Oliveira, E., et al., *High levels of hybridization between molecular forms of Anopheles gambiae from Guinea Bissau.* J Med Entomol, 2008. **45**(6): p. 1057-63.
160. Fontaine, M.C., et al., *Mosquito genomics. Extensive introgression in a malaria vector species complex revealed by phylogenomics.* Science, 2015. **347**(6217): p. 1258524.
161. Smith, C.D., et al., *The Release 5.1 annotation of Drosophila melanogaster heterochromatin.* Science, 2007. **316**(5831): p. 1586-91.
162. Manning, J.E., C.W. Schmid, and N. Davidson, *Interspersion of repetitive and nonrepetitive DNA sequences in the Drosophila melanogaster genome.* Cell, 1975. **4**(2): p. 141-55.
163. Besansky, N.J. and F.H. Collins, *The mosquito genome: organization, evolution and manipulation.* Parasitol Today, 1992. **8**(6): p. 186-92.
164. 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.
165. Neafsey, D.E., et al., *Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes.* Science, 2015. **347**(6217): p. 1258522.
166. Jiang, X., et al., *Genome analysis of a major urban malaria vector mosquito, Anopheles stephensi.* Genome Biol, 2014. **15**(9): p. 459.
167. Timoshevskiy, V.A., et al., *Fluorescent in situ Hybridization on Mitotic Chromosomes of Mosquitoes.* J Vis Exp, 2012(67).
168. 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.
169. Pita, S., et al., *Chromosomal divergence and evolutionary inferences in Rhodniini based on the chromosomal location of ribosomal genes.* Mem Inst Oswaldo Cruz, 2013. **108**(3).
170. Rai, K.S., *A comparative study of mosquito karyotypes.* Ann ent Soc Am, 1963. **56**: p. 160-170.
171. Matsunaga, S., *Junk DNA promotes sex chromosome evolution.* Heredity (Edinb), 2009. **102**(6): p. 525-6.
172. Kejnovsky, E., et al., *The role of repetitive DNA in structure and evolution of sex chromosomes in plants.* Heredity (Edinb), 2009. **102**(6): p. 533-41.
173. Grushko, O.G., et al., *[Localization of repetitive DNA sequences in the pericentromeric heterochromatin of malarial mosquitoes of the "Anopheles maculipennis" complex].* Tsitologiia, 2006. **48**(3): p. 240-5.
174. Naumenko, A.N., et al., *Mitotic-chromosome-based physical mapping of the Culex quinquefasciatus genome.* PLoS One, 2015. **10**(3): p. e0115737.
175. Bull, J.J., *Evolution of sex determining mechanisms*. 1983: The Benjamin/Cummings Publishing Company, Inc.
176. Bachtrog, D., et al., *Sex Determination: Why So Many Ways of Doing It?* PLoS Biology, 2014. **12**(7): p. e1001899.
177. Bachtrog, D., *Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration.* Nat Rev Genet, 2013. **14**(2): p. 113-24.

178. Charlesworth, B. and D. Charlesworth, *The degeneration of Y chromosomes.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2000. **355**(1403): p. 1563-1572.
179. Rice, W.R., *Sex Chromosomes and the Evolution of Sexual Dimorphism.* Evolution, 1984. **38**(4): p. 735.
180. Bellott, D.W., et al., *Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators.* Nature, 2014. **508**(7497): p. 494-499.
181. Cortez, D., et al., *Origins and functional evolution of Y chromosomes across mammals.* Nature, 2014. **508**(7497): p. 488-493.
182. Lemos, B., L.O. Araripe, and D.L. Hartl, *Polymorphic Y Chromosomes Harbor Cryptic Variation with Manifold Functional Consequences.* Science, 2008. **319**(5859): p. 91-93.
183. Lemos, B., A.T. Branco, and D.L. Hartl, *Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict.* Proceedings of the National Academy of Sciences, 2010. **107**(36): p. 15826-15831.
184. Sackton, T.B., et al., *Interspecific Y chromosome introgressions disrupt testis-specific gene expression and male reproductive phenotypes in Drosophila.* Proceedings of the National Academy of Sciences, 2011. **108**(41): p. 17046-17051.
185. Hughes, J.F. and D.C. Page, *The Biology and Evolution of Mammalian Y Chromosomes.* Annu. Rev. Genet., 2015. **49**(1): p. 507-527.
186. Bhatt, S., et al., *The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015.* Nature, 2015. **526**(7572): p. 207-211.
187. Galizi, R., et al., *A synthetic sex ratio distortion system for the control of the human malaria mosquito.* Nature Communications, 2014. **5**.
188. Windbichler, N., P.A. Papathanos, and A. Crisanti, *Targeting the X Chromosome during Spermatogenesis Induces Y Chromosome Transmission Ratio Distortion and Early Dominant Embryo Lethality in Anopheles gambiae.* PLoS Genetics, 2008. **4**(12): p. e1000291.
189. Burt, A., *Heritable strategies for controlling insect vectors of disease.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2014. **369**(1645): p. 20130432-20130432.
190. Deredec, A., H.C.J. Godfray, and A. Burt, *Requirements for effective malaria control with homing endonuclease genes.* Proceedings of the National Academy of Sciences, 2011. **108**(43): p. E874-E880.
191. Sakai, R.K., et al., *Crossing-over in the long arm of the X and Y chromosomes in Anopheles culicifacies.* Chromosoma, 1979. **74**(2): p. 209-218.
192. Redfern, C.P.F., *Satellite DNA of Anopheles stephensi liston (Diptera: Culicidae).* Chromosoma, 1981. **82**(4): p. 561-581.
193. Mitchell, S. and J. Seawright, *Recombination between the X and Y chromosomes in Anopheles quadrimaculatus species A.* The Journal of heredity, 1989. **80**(6): p. 496-499.
194. Marchi, A. and R. Mezzanotte, *Inter- and intraspecific heterochromatin variation detected by restriction endonuclease digestion in two sibling species of the Anopheles maculipennis complex.* Heredity (Edinb), 1990. **65 (Pt 1)**: p. 135-42.
195. White, G., *Academic and applied aspects of mosquito cytogenetics.* Insect Cytogenetics, 1980: p. 245-274.
196. Fraccaro, M., et al., *Karyotype, DNA replication and origin of sex chromosomes in Anopheles atroparvus.* Chromosoma, 1976. **55**(1): p. 27-36.
197. Carvalho, A.B., et al., *Y chromosome and other heterochromatic sequences of the Drosophila melanogaster genome: how far can we go?* Genetica, 2003. **117**(2-3): p. 227-37.
198. Krzywinski, J., *Isolation and Characterization of Y Chromosome Sequences From the African Malaria Mosquito Anopheles gambiae.* Genetics, 2004. **166**(3): p. 1291-1302.

199. Ross, M.G., et al., *Characterizing and measuring bias in sequence data.* Genome Biol, 2013. **14**(5): p. R51.
200. White, B.J., F.H. Collins, and N.J. Besansky, *Evolution of Anopheles gambiae in Relation to Humans and Malaria.* Annu. Rev. Ecol. Evol. Syst., 2011. **42**(1): p. 111-132.
201. Hall, A.B., et al., *Six novel Y chromosome genes in Anopheles mosquitoes discovered by independently sequencing males and females.* BMC Genomics, 2013. **14**(1): p. 273.
202. Giraldo-Calderon, G.I., et al., *VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases.* Nucleic Acids Research, 2014. **43**(D1): p. D707-D713.
203. Eid, J., et al., *Real-Time DNA Sequencing from Single Polymerase Molecules.* Science, 2009. **323**(5910): p. 133-138.
204. Koren, S., et al., *Reducing assembly complexity of microbial genomes with single-molecule sequencing.* Genome Biol, 2013. **14**(9): p. R101.
205. Koren, S., et al., *Hybrid error correction and de novo assembly of single-molecule sequencing reads.* Nat Biotechnol, 2012. **30**(7): p. 693-700.
206. Rohr, C.J., et al., *Structure and evolution of mtanga, a retrotransposon actively expressed on the Y chromosome of the African malaria vector Anopheles gambiae.* Mol Biol Evol, 2002. **19**(2): p. 149-62.
207. Sharma, A., T.K. Wolfgruber, and G.G. Presting, *Tandem repeats derived from centromeric retrotransposons.* BMC Genomics, 2013. **14**(1): p. 142.
208. Cohen, S., et al., *Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells.* Mobile DNA, 2010. **1**(1): p. 11.
209. Ma, J., *Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice.* Genome Research, 2005. **16**(2): p. 251-259.
210. Smith, G., *Evolution of repeated DNA sequences by unequal crossover.* Science, 1976. **191**(4227): p. 528-535.
211. Gatti, M., et al. *Polymorphism of sex chromosome heterochromatin in the Anopheles gambiae complex*. in *Recent developments in the genetics of insect disease vectors: a symposium proceedings/edited by WWM Steiner...[et al.]*. 1982. Champaign: Stipes Publishing Co., c1982.
212. Curtis, C., J. Akiyama, and G. Davidson, *Genetic sexing system in Anopheles gambiae species A.* Mosquito News, 1976.
213. Krzywinski, J., *Satellite DNA From the Y Chromosome of the Malaria Vector Anopheles gambiae.* Genetics, 2004. **169**(1): p. 185-196.
214. Criscione, F., et al., *A unique Y gene in the Asian malaria mosquito Anopheles stephensi encodes a small lysine-rich protein and is transcribed at the onset of embryonic development.* Insect Mol Biol, 2013. **22**(4): p. 433-41.
215. Kamali, M., et al., *Multigene Phylogenetics Reveals Temporal Diversification of Major African Malaria Vectors.* PLoS ONE, 2014. **9**(4): p. e93580.
216. Payseur, B.A., *Y not introgress? Insights into the genetics of speciation in European rabbits.* Molecular Ecology, 2008.
217. Skaletsky, H., et al., *The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes.* Nature, 2003. **423**(6942): p. 825-837.
218. Hughes, J.F., et al., *Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes.* Nature, 2012. **483**(7387): p. 82-86.
219. Koerich, L.B., et al., *Low conservation of gene content in the Drosophila Y chromosome.* Nature, 2008. **456**(7224): p. 949-951.
220. Bachtrog, D., et al., *Genomic degradation of a young Y chromosome in Drosophila miranda.* Genome Biol, 2008. **9**(2): p. R30.

221. Mendez-Lago, M., et al., *A Large Palindrome With Interchromosomal Gene Duplications in the Pericentromeric Region of the D. melanogaster Y Chromosome.* Molecular Biology and Evolution, 2011. **28**(7): p. 1967-1971.
222. Toups, M.A. and M.W. Hahn, *Retrogenes reveal the direction of sex-chromosome evolution in mosquitoes.* Genetics, 2010. **186**(2): p. 763-6.
223. Soh, Y.Q.S., et al., *Sequencing the Mouse Y Chromosome Reveals Convergent Gene Acquisition and Amplification on Both Sex Chromosomes.* Cell, 2014. **159**(4): p. 800-813.
224. Saxena, R., et al., *The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned.* Nature Genetics, 1996. **14**(3): p. 292-299.
225. Carvalho, A.B., et al., *Birth of a new gene on the Y chromosome of Drosophila melanogaster.* Proceedings of the National Academy of Sciences, 2015. **112**(40): p. 12450-12455.
226. Pertile, M.D., et al., *Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability.* Genome Research, 2009. **19**(12): p. 2202-2213.
227. Wei, K.H.C., et al., *Correlated variation and population differentiation in satellite DNA abundance among lines of Drosophila melanogaster.* Proceedings of the National Academy of Sciences, 2014. **111**(52): p. 18793-18798.
228. Repping, S., et al., *High mutation rates have driven extensive structural polymorphism among human Y chromosomes.* Nature Genetics, 2006. **38**(4): p. 463-467.
229. Bernardini, F., et al., *Site-specific genetic engineering of the Anopheles gambiae Y chromosome.* Proceedings of the National Academy of Sciences, 2014. **111**(21): p. 7600-7605.
230. Bayes, J.J. and H.S. Malik, *Altered Heterochromatin Binding by a Hybrid Sterility Protein in Drosophila Sibling Species.* Science, 2009. **326**(5959): p. 1538-1541.
231. Lohe, A.R. and D.L. Brutlag, *Identical satellite DNA sequences in sibling species of Drosophila.* Journal of Molecular Biology, 1987. **194**(2): p. 161-170.
232. Berlin, K., et al., *Assembling large genomes with single-molecule sequencing and locality-sensitive hashing.* Nat Biotechnol, 2015. **33**(6): p. 623-630.
233. Krsticevic, F.J., C.G. Schrago, and A.B. Carvalho, *Long-Read Single Molecule Sequencing to Resolve Tandem Gene Copies: The Mst77Y Region on the Drosophila melanogaster Y Chromosome.* G3: Genes|Genomes|Genetics, 2015. **5**(6): p. 1145-1150.
234. CDC. *CDC Report- 2010*. Available from: http://www.cdc.gov/malaria/.
235. Gatton, M.L., et al., *The importance of mosquito behavioural adaptations to malaria control in Africa.* Evolution, 2013. **67**(4): p. 1218-30.
236. Sternberg, E.D. and M.B. Thomas, *Local adaptation to temperature and the implications for vector-borne diseases.* Trends Parasitol, 2014.
237. Rocca, K.A., et al., *2La chromosomal inversion enhances thermal tolerance of Anopheles gambiae larvae.* Malar J, 2009. **8**: p. 147.
238. 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.
239. Donepudi, S., Mattison, R.J., Kihslinger, J.E., Godley, L.A., ed. *Modulator of DNA Methylation and Histone Acetylation*. Update on Cancer Therapeutics. Vol. 2. 2007: ScienceDirect.
240. Greer, E.L. and Y. Shi, *Histone methylation: a dynamic mark in health, disease and inheritance.* Nat Rev Genet, 2012. **13**(5): p. 343-57.
241. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
242. Clough, E., T. Tedeschi, and T. Hazelrigg, *Epigenetic regulation of oogenesis and germ stem cell maintenance by the Drosophila histone methyltransferase Eggless/dSetDB1.* Dev Biol, 2014. **388**(2): p. 181-91.

243. Conrad, T. and A. Akhtar, *Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription.* Nat Rev Genet, 2011. **13**(2): p. 123-34.
244. Cui, B., et al., *PRIMA-1, a Mutant p53 Reactivator, Restores the Sensitivity of TP53 Mutant-type Thyroid Cancer Cells to the Histone Methylation Inhibitor 3-Deazaneplanocin A (DZNep).* J Clin Endocrinol Metab, 2014: p. jc20133147.
245. Fujiwara, T., et al., *3-Deazaneplanocin A (DZNep), an inhibitor of S-adenosyl-methionine-dependent methyltransferase, promotes erythroid differentiation.* J Biol Chem, 2014.
246. Li, Z., et al., *The polycomb group protein EZH2 is a novel therapeutic target in tongue cancer.* Oncotarget, 2013. **4**(12): p. 2532-49.
247. Nakagawa, S., et al., *Epigenetic therapy with the histone methyltransferase EZH2 inhibitor 3-deazaneplanocin A inhibits the growth of cholangiocarcinoma cells.* Oncol Rep, 2014. **31**(2): p. 983-8.
248. Crea, F., et al., *Pharmacologic disruption of Polycomb Repressive Complex 2 inhibits tumorigenicity and tumor progression in prostate cancer.* Mol Cancer, 2011. **10**: p. 40.
249. Vella, S., et al., *EZH2 down-regulation exacerbates lipid accumulation and inflammation in in vitro and in vivo NAFLD.* Int J Mol Sci, 2013. **14**(12): p. 24154-68.
250. Tan, J., et al., *Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells.* Genes Dev, 2007. **21**(9): p. 1050-63.
251. Chiang, P.K. and G.L. Cantoni, *Perturbation of biochemical transmethylations by 3-deazaadenosine in vivo.* Biochem Pharmacol, 1979. **28**(12): p. 1897-902.
252. Lu, Y.X., D.L. Denlinger, and W.H. Xu, *Polycomb repressive complex 2 (PRC2) protein ESC regulates insect developmental timing by mediating H3K27me3 and activating prothoracicotropic hormone gene expression.* J Biol Chem, 2013. **288**(32): p. 23554-64.