Comparative genomics of chromosomal rearrangements in malaria mosquitoes

Ai Xia

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

Dissertation Committee:

Igor V Sharakhov

Zach N Adelman

Jeffrey R Bloomquist

Kevin M Myles

Zhijian Tu

January 25th, 2010

Blacksburg, Virginia

Key words: *Anopheles* mosquitoes, chromosomal inversions, 2La invesion, whole genome amplification, microdissection

Comparative genomics of chromosomal rearrangements in malaria mosquitoes

Ai Xia

Abstract

To better understand the evolutionary dynamics of chromosomal inversions, a physical map for an Asian malaria vector, *Anopheles stephensi*, was created and compared with the maps of the major African malaria vectors *A. gambiae* and *A. funestus* No interchromosomal transposition was observed between *A. gambiae* and *A. stephensi*. Several cases of euchromatin and heterochromatin transitions weridentified between *A. gambiae* and *A. stephensi*. The study of paracentric inversions between lineages in *Anopheles* mosquitoes demonstrated that X chromosome has the fastest rate of inversion fixations and highest density of repetitive elements. Among the autosomes, 2R evolved faster than other autosomes. The slowly evolved autosomes have more M/SARs than rapidly evolving arms. Breakpoint regions are enriched with repetitive elements. The study revealed that fixed inversions are distributed nonrandomly and breakpoint clustering is common in lineages of *A. gambiae* and *A. stephensi*. The parallel association between the density of inversion fixations and polymorphisms suggests that polymorphic inversions can be fixed during evolution.

To understand the direction of evolution in *A. gambiae* complex, the ancestral status of fixed inversions for this complex was identified. The presence of the 2La inversion in outgroups, *A. stephensi* and *A. nili*, confirmed the ancestral status of the 2La inversion. The presences of breakpoint structure of the 2Ro inversion in outgroup species, *A. stephensi,* indicated that the 2Ro is ancestral arrangement. The presence of SINE

elements at the breakpoints of the $2R+^{p}$ in *A. gambiae* PEST strain suggested that the $2R+^{p}$ is a derived arrangement. Therefore, the carrier of 2Rop inversions, *A. merus,* was considered closest to the ancestral species.

We have developed a new protocol for laser microdissection and whole genome amplification of polytene chromosomal fragments to obtain DNA for sequencing and assembly. The chromosomal regions spanning both breakpoints of the 2La in *A. arabiensis* and *A. merus* were laser microdissected from the polytene chromosomes. Subsequently, DNA samples were amplified using Illustra GenomePhi V2 DNA and Whole-pool amplification methods for obtaining amplicons. Successful amplification of our target DNA was confirmed by PCR with specific primers followed by Sanger sequencing.

Key words:

Anopheles stephensi, *A. gambiae*, *A. funestus*, physical map, chromosomal rearrangement, Nadeau & Taylor model, fixed inversion, polymorphic inversion, nuclear architecture, phylogenetic relationship, 2La inversion, whole genome amplification, laser capture microdissection, *plasmodium* resistance island.

Dedication:

Dr. Maria Sharakhova dedicated to partial mapping of *A. stephensi*. Dr. Scotland Leman contributed to the model desighing of chapter three, and Dr. Xinghua Pan to run Wpa amplification reactions. Copy right of Figure 3.1, 3.2, 3.4 and 3.5 belongs to Dr. Scotland Leman. Copy right of Figure 5.4, 5.5, and 5.6 belongs to Dr. Igor Sharakhov, Copy right of Figure 5.22 belongs to Dr. Xinghua Pan.

Acknowledgements

There are many people I want to thank for their kind help, valuable suggestion and guidance as well as strong support in working on these projects. Firstly, I sincerely thank my advisor Dr. Igor Sharakhov. During the study of more than four years, he encouraged and supported me through entire projects. I am truly grateful that he trusted and helped me (especially) during the times when everything didn't progress as we expected. I also like to thank Dr. Maria Sharakova for her help in getting me started in the lab, teaching me Fluorescence *in situ* Hybridization, and helping me with mapping some clones. I like to thank all our collaborators: Dr. Scotland Leman for designing statistic models, analyzing data in Chapter two; Dr. Xinghua Pan for running Wpa amplification reactions, insightful discussions of my 2La sequencing project, and valuable suggestion for developing protocol of whole genome amplification; Dr. Tu for providing *A. stephensi* BAC clones, analyzing transposable elements, his unique insights into my projects and his help throughout my graduate study. Dr. Yogesh Shouche for providing *A. stephensi* cDNA clones. I'd like to thank all committee members: Dr. Adelman, Dr. Bloomquist, Dr. Myles and Dr. Tu, for all the time they've contributed by serving on my committee and contributing to my professional development. I especially thank Dr. Bloomquist for his encouragement and support that gave me tremendous self-confidence. I also want to thank all my lab mates: Phillip George, Maryam Kamali, and Fan Yang for all their help, encouragement. I'd like to thank Randy Saunders with providing mosquitoes for my experiments. I thank Entomology department head Dr. Loke T Kok and Sarah Kenley, Robin Williams, Kathy M Shelor for their help and assistance.

I also want to thank the Malaria Research and Reference Reagent Resource Center (MR4) for providing BAC clones and mosquito species and Virginia Tech, the State Council of Higher Education for Virginia (SCHEV), Fralin Biotechnology Center, The Institute for Biomedical and Public Health Sciences (IBPHS), NIH for providing funding to conduct this research.

Last but not least, I'd like to thank my family and friends for their long support during the time. I especially thank to my parents for all the help and support they have given me. I thank my husband, Daolong Dou for his supporting my drive to seek a PhD, his understanding, encouragement and contributing during the entire time. I thank my daughter Wen Dou and my son Max Dou for the pleasure they've brought to me and their innocent smile inspired and encouraged me to overcome all the difficulties.

Table of Contents

List of tables

List of figures

List of abbreviations

TEs: transposable elements

MY: million years

MYA: million years ago

N-T model: Nadeau-Taylor model

MGR: Multiple Genome Rearrangements

RCIs: rare chromosomal inversions

M/SARs: Matrix/Scaffold Attachment Regions

NE: nuclear envelope

SDs: segmental duplications

SPRING: Reversal and Block InterchanGes

GRIMM: Genome Rearrangements in Man and Mouse

MITE: miniature inverted transposable element

SINE: short interspersed repetitive element

PRI: *Plasmodium falciparum* Resistance Island

Wpa: whole-pool amplification

V2 DNA amplification: GenomePhi V2 DNA amplification

WGA: whole genome amplification

MDA: multiple displacement amplification

Est: esterase

PGD: preimplantation genetics and diagnosis

RNAi: RNA interference

PEP: primer extension preamplification

DOP-PCR: degenerate oligonuceotide primed PCR

TIP: template independent product

CNS: conserved non-coding sequences

CGH: comparative genomic hybridization

EIR: entomological inoculation rate

Chapter one: Introduction to the dissertation

1. *Anopheles* mosquitoes - general information

Malaria is the deadliest infectious disease caused by infection of *Plasmodium* parasites, mainly *P. falciparum* and *P. vivax*. More than 500 million clinical cases of malaria have been reported each year, resulting in 2.7 million deaths, most of them being young children in sub-Saharan Africa (*2*). Malaria is widespread in tropical and subtropical regions, including Africa, South America and Asia. It is transmitted via the bite of infected female mosquitoes. There are about 3500 mosquito species worldwide, which are classified into three subfamilies; *Anophelinae*, *Culicinae* and *Toxorhynchitinae*, but only a small number of *Anophelinae* mosquitoes are genetically competent to transmit malaria parasites. Of *Anopheles* mosquitoes, *A. gambiae,* and *A. funestus* represent the most efficient malaria vectors in sub-Saharan Africa. These species overlap geographically from south of the Sahara Desert to northern South Africa (*3*). Yet *A. funestus* mainly transmits malaria during the dry season, when *A. gambiae* is usually inactive. *A. stephensi* is an efficient malaria vector in Asia and distributed widely in the Middle East and Indian subcontinent (*4*) and the south of Iran (*5*). It is obvious that successful malaria control strategies must take all of these species, as well as others into account. Although efforts to eradicate malaria by vector control have succeed in the United States and Southern Europe in the past, current strategies to limit or prevent transmission of this disease are insufficient due to the lack of effective vaccines, growing insecticide resistance (*6, 7*), and deteriorating socio-economic conditions in many endemic areas. A better understanding of the genetic basis of

vectorial capacity for these major species as well as nonvectors may enable scientists to eventually develop manipulation strategies in order to reduce disease burden.

2. Cytogenetic maps for *Anopheles* mosquitoes

A. gambiae, *A. funestus,* and *A. stephensi* are not only competent malaria vectors, but also are good model organisms for cytogenetics studies. The availability of well-polytenized chromosomes in *Anopheline* mosquitoes provides a great opportunity for those studies (*8*). Polytene chromosomes are gigantic, long rope-like interphase chromosomes and are characterized by distinguishable banding patterns. Polytene chromosomes form when multiple rounds of replication produce many sister chromatids that remain synapsed together. High quality polytene chromosomes have been found in various tissues of mosquitoes including the salivary glands, gut, Malpigian tubules, and ovarian nurse cells of half gravid female mosquitoes (*8*). Nearly all *anopheles* mosquitoes share a mitotic karyotype with two pairs of autosomes and one pair of sex chromosomes. The polytene compliment consists of five chromosomal arms (X or Y, 2R, 2L, 3R and 3L). Several drawn and photo cytogenetic maps with divisions and subdivisions for the *A. gambiae* complex have been developed (*1, 9-13*). Two maps are available for *A. funestus* (*1, 14*). Inversions breakpoints were placed on the former's cytogenetic map (*14*), while the recent high resolution cytogenetic map of *A. funestus* was divided into 46 numbered divisions and 151 lettered subdivisions (*1*). *A. stephensi,* as a malaria vector in Asia, is also a well studied model system with regard to polytene chromosomes. The first drawn and photo map of the polytene chromosomes from the salivary glands of *A. stephensi* was published by Sharma et al in 1969 (*15*). A photo map from ovarian nurse cell chromosomes was developed by Coluzzi et al a year later

(*16*) and in this publication five chromosomal arms were divided into 46 numbered divisions. Then, the first standard drawn map of ovarian nurse cell chromosomes was developed by Mahmood and Sakai (*17*). In 1970, Coluzzi et al further divided the map into 151 lettered subdivisions. But all these published maps are either drawn or without letter subdivisions. Recently, a high resolution image of the ovarian polytene chromosomes with number divisions and letter subdivisions has been developed for *A. stephensi* (*18*).

3. Physical maps for *Anopheles* mosquitoes

Methods for *in situ* hybridization to polytene chromosomes were well developed (*19*) and several physical maps have been developed for *Anopheles* mosquitoes. The first *A. gambiae* physical map included 46 clones microdissected from different divisions of polytene chromosomes (*20*). Later microsatellite markers (*21*), randomly amplified polymorphic DNA (RAPD) markers (*22*), cosmids, and cDNA markers (*23*) have been physically mapped to *A. gambiae* chromosomes. The most detailed physical map for the *A. gambiae* genome project allowed researchers to assign 67 scaffolds of 227 Mbp (megabase pairs) in total length to the chromosomal regions (*24*). An additional 16 scaffolds, from 50kb to 600 kb in length, were localized in pericentromeric regions using cDNA clones (*25*). Another efficient malaria vector, *A. funestus*, has only recently started to receive scientific attention, in part because of the difficulties in colonizing this species. A library prepared from *A. funestus* larval, pupal, and adult mRNA was established. From this library, 157 *A. funestus* complementary DNAs (cDNAs) were isolated and mapped to the polytene chromosomes of this species using fluorescent *in situ* hybridization (*1*). More recently, the same research group also provided an

integrated polytene chromosome map for *A. funestus* which contains 32 polymorphic *A. funestus* microsatellite markers and the breakpoints of all known polymorphic inversions (*26*). However, only eight markers were placed on the cytogenetic map of *A. stephensi* (*18*). Therefore, there is an urgent need to develop a physical map for *A. stephensi*.

Our first objective was to develop a high resolution physical map for *A. stephensi*. Our specific aims were to develop a physical map for *A. stephensi* by fluorescent *in situ* hybridization. Chromosomal rearrangements such as transpositions, duplications and euchromatin-heterochromatin transitions will be analyzed by comparing *A. gambiae* and *A. stephensi*.

4. The patterns and rates of chromosomal rearrangements – general information

During the evolution of the eukaryotic genomes, the order of genes has been reshuffled from species to species via chromosomal rearrangements (*27-30*). The changes in gene order can affect either small or large chromosomal segments (micro- or macro-rearrangements) (*31*). The macro- or 'gross' chromosomal rearrangements (usually several mega bases long) are large-scale changes in chromosomal structure which can be seen at the microscope level, and include deletions, inversions, duplications, and translocations (*32*). Much is known about gross chromosomal rearrangements in *Drosophila* (*30, 31, 33-37*) and *Anopheles* (*1*) because rearrangements are easily detected on the polytene chromosomes. In the genus *Drosophila*, studies of chromosomal evolution by comparative cytogenetic mapping were pioneered more than 40 years ago (*38, 39*); This approach has become an even more powerful and reliable with the development of the *in situ* hybridization technique that allows the localization of DNA sequences to polytene chromosomes (*40*). In the

genus *Anopheles*, studies of paracentric inversions within species (*13, 16, 41*) and among species (*13, 42*) using cytogenetic maps have a long history. Comparative physical mapping allows us to describe and estimate the amount of chromosomal evolution that has occurred during the divergence of species from their common ancestor (*28, 43, 44*). This evolution can be expressed as the number of chromosomal rearrangements separating species' genomes, and as the level of conserved synteny blocks (conserved linear order of contiguous markers). However, comparative physical mapping is not feasible for most organisms without giant polytene chromosomes. The development of linkage maps for a variety of species (*45-48*) and the recent advent of chromosome painting in the early 1990s (*49-52*) provided useful means for studying genomic organization and evolution. For example, linkage maps have been applied to the calculation of the lengths of chromosomal segments conserved since divergence of man and mouse (*43*). In addition, chromosome painting has been most intensively applied to primate genomes (*51, 53*), and has been performed for over 40 mammalian species (*28, 54, 55*). Although both cytogenetic physical mapping and genetic-linkage mapping are extremely powerful for identifying conserved syntenies and estimation of large scale chromosomal rearrangements, they are limited in their inability to discern smaller rearrangements within chromosome segments. The availability of whole-genome sequences from many species and advances in molecular biology enabled the detection of fine-scale changes in chromosomal structure (micro-rearrangements) by direct comparison of complete genome sequences, shifting researchers interest from gross rearrangements to smaller changes, such as small inversions involving very few, or a single gene. Genome-wide comparative studies have been performed in yeast (*29,*

56, 57), nematodes (*58, 59*), mammals (*60-62*) and *Drosophila* (*63, 64*). The above evidence suggests that small inversions or microinversions, which range from a gene to several genes in size, are very common in eukaryotic genomes. The comparative analysis carried out so far with either complete genome sequences, or by the construction of physical or linkage maps have led to three striking conclusions about genome evolution: (1) the eukaryotic genome is unexpectedly malleable to chromosomal rearrangements; (2) the patterns and rates of rearrangements are lineage-specific; (3) chromosomal elements remarkably differ in the rates of evolution.

4.1. Lineage-specific patterns of chromosomal rearrangements

4.1.1. Fungi

Within the fungi, yeasts represent an excellent model for comparative genomic studies because of their small and compact genomes and broad range of phylogenetic distance covered. In the evolution of *Saccharomyces cerevisiae*, an enormous amount of data confirmed the view that complete genome duplication occurred after its divergence from *Kluyveromyces lactis* and before its divergence from *Candida,* followed by massive gene loss (*57, 65-68*). Comparative analysis of whole genomic sequences of *S. cerevisiae* (*69*) and other yeast species also revealed that the genome of *Candida glabrata* was duplicated at some point in its evolutionary past (*57, 70*). In other yeast clades, segmental duplications (a large block of up to 250 continuous gene replications) and tandem gene duplication (the definition was given in the above notion of duplication) have been adopted for extensive genome expansion (*57, 70*). In addition, studying the conservation of synteny between species in yeast revealed considerable gene order reshuffling (*29, 57*). Recently comparative analyses of complete genomes revealed that

genome-wide small scale inversions, less than ten genes long have likewise contributed significantly to chromosomal evolution among ascomycetes yeast (*29*).

4.1.2. Invertebrates

The events of genetic exchange by pericentric inversions (*71*), translocations (*72*), or interchromosomal transpositions (*33, 73-76*) are extremely rare in the genus *Drosophila*, yet they do occur. The results drawn from the comparison between *A. gambiae* and *A. funestus* in the genus *Anopheles* (*1*) agree with this conclusion. However, interchromosomal genetic exchange has been observed between lineages in *Drosophila* and *Anopheles* (*77*). In the genus *Drosophila*, the vast amount of cytogenetic information (*38, 78*) and recent results using DNA markers, as well as *in situ* hybridization mapping (*36, 37, 71, 79-81*) on many *Drosophila* species suggested that the six chromosomal elements (A-F) that constituted the *Drosophila* ancient genome have maintained their integrity in many lineages. The gene content within each chromosomal element is highly conserved during the devolution of most lineages. In the genus *Anopheles*, the gene content of the five chromosomal arms (X, 2R, 2L, 3R and 3L) is also preserved during evolution of two lineages, *A. gambiae* and *A. funestus* (*1*). However, whole arm translocations were found among *A. gambiae, A. funestus* and *A. stephensi* (*1, 18*). The order of genes within each chromosomal arm has been internally rearranged, most often by the paracentric inversions, which are by far the most frequent chromosomal rearrangements, in the genus *Drosophila* (*30, 33, 35, 37, 71, 80*) and other dipterans, including *Anopheles* (*1*). Paracentric inversions are abundant as intraspecific polymorphisms. For example, >500 polymorphic inversions have been discovered in the populations of *D. melanogaster* (*82*), and more than 120 polymorphic

inversions have been described in the *A. gambiae* complex (*42*). Most other *Drosophila* (*42, 83, 84*) and *Anopheles* (*1, 13, 17, 41*) species also harbor many paracentric polymorphic inversions. Paracentric inversions are also commonly fixed between species. In *Drosophila*, comparative mapping analysis has been performed in *D. melanogaster* and *D. replete*, representing the two farthest lineages within the genus (*33*); in *D. subobscura* and *D. pseudoobscura* (closely related species) (*31*); as well as between other lineages (*30, 34-37, 71, 80, 81, 85*). In contrast, only one extensive study on chromosomal evolution was reported between *A. gambiae* and *A. funestus* (*1*).

Little was known about the evolution of nematode chromosomes until 2002 when the genome of *Caenorhabditis elegans* was compared with that of *Caenorhabditis briggsae* (*58*). A total of 517 chromosomal rearrangements, with the ratio of translocations to small fine inversions (most of them are <25 Kb) to transpositions being 1:1:2 were detected. The study reported a four-fold faster rate of genome rearrangement in nematode than in *Drosophila* (*58*).

4.1.3. Vertebrates

Chromosomal rearrangements including fissions, fusions, inversions, translocations, and transpositions are the primary types of chromosomal changes in vertebrates. In mammals, the most striking feature of chromosomal rearrangements is the remarkable prevalence of micro-rearrangements in all genomic regions. Genome-wide comparisons of humans and chimpanzees have indicated that microinversions ranging in size from a few bp to several Mb are astonishingly frequent in all genomic regions examined (*86-88*). Additionally, in recent comparisons of the human and mouse genomes, a few thousand “microarrangements” were found within synteny blocks (*89-92*). Smaller

intrachromosomal rearrangements are relatively frequent in other mammalian genomes (*93*). This is true for even more distant vertebrate comparisons, such as the comparison between the zebrafish and human maps (*94*). In addition to small inversions, segmental duplications (SDs) had a very significant impact on genome plasticity during primate evolution (*95*). For example, in comparisons of human and chimpanzee genomes, SDs were found flanking the sites of the ancestral fusion (*96*), most of the breakpoints of pericentric inversions (*97, 98*), 70-80% of inversions, and 40% of insertion/deletion breakpoints (*86*). Compared with other mammalian genomes (mouse, rat, and dog), the human exhibits a high content of SDs in interchromosomal and intrachromosomal rearrangements (*99, 100*).

In addition to the common features shared by all of the lineages, lineage-specific patterns of chromosomal rearrangements have been identified between vertebrate species since the divergence from their common ancestor. The fusion and pericentric inversions serve to differentiate between human and chimpanzee karyotypes (*101-103*). In comparisons of human and gibbon, enormous fissions and translocations have been revealed (reviewed in (*104*)). The ordered gene maps of the cat and cow genome showed that intrachromosomal changes (inversions) are more prevalent than interchromosomal changes which are in contrast with predictions from human-mouse comparisons (*105*). Whole genome comparative studies between chicken and turkey revealed two interchromosomal changes and three pericentric inversions that suggest that the avian genome has remained relatively stable during evolution as compared to mammalian genomes (*106*).

4.1.4. Plants

The genomes of plants diverged dramatically in terms of genome size, ploidy level, and chromosome number. Polyploidization is defined as whole genome duplication, doubling or tripling the number of chromosomes (*32*). Polyploidy is the most important genetic mechanism in plant evolution; most flowering plant genomes have been thought to undergo polyploidization events during their evolutionary history (*107*). In addition, early evidence based on genetic mapping suggested that ancient segmental duplications exist for *Sorghum* (*108*), *Arabidopsis* (*109*), and *Oryza* (*110*), which has been confirmed by the analysis of the near-complete sequencing of *Arabidopsis* (*111-114*) and *Oryza* (*115, 116*). Despite these significant differences in genome size and chromosome number, it was found by comparative mapping that the linear order of genetic markers and genes (colinearity) is highly conserved between related genomes in plants and such colinearity was remarkable, given differences in genome size of up to 40-fold (*117-121*). Is the conservation of the marker order at the genetic map level retained at the molecular level? Comparisons of DNA sequences of orthologous chromosomal segments from small genomic regions to large genomic regions of many Mb have demonstrated that microcolinearity is retained at the molecular level in most cases (*122*). However, numerous studies revealed significant gene rearrangements within the collinear chromosomal regions including deletions/insertions, translocations, inversions, and duplications. Virtually all of these newly discovered rearrangements were very small, often including one or two genes (*121, 123-128*). For example, several-fold more rearrangements were observed in wheat-rice genome comparison than on compared genetic maps (*127, 129*). Additionally, extensive evidence suggests that small rearrangements are much more common than large rearrangements in the

evolution of plant genomes (*121, 127*).

4.2. Lineage and chromosome specificity in the rates of chromosomal evolution

Remarkable differences in the rates of chromosomal evolution have been observed between distantly related phylogenetic lineages. Among all organisms, the fastest rate of genome rearrangement (0.4-1.0 breakpoints/Mb/Million year) has been observed in the genomes of nematodes (*34, 58*). The high rate of genome evolution in nematodes may be due to the large population size and short generation time. It has been also proposed that chromosomal rearrangements can be less deleterious in nematode genomes because their chromosomes have no centromeres (*130*). Among insects, the highest rates of chromosomal rearrangement have been identified in *Drosophila* (*30, 31*). A detailed comparative study of the largest chromosomal element (Muller's element E) between the species *Drosophila repleta* and *D. melanogaster*, for instance, has revealed about 0.066-0.053 breakpoints/Mb/MY (*30*). Only the genomes of yeast and *Anopheles* seem to exhibit similar or perhaps greater rates of chromosomal rearrangement compared to those of *Drosophila* (*1, 29, 131*). As compared to vertebrate and plant genomes, the *Drosophila* genome evolves two-orders of magnitude higher than that of mammals and at least five-fold faster than the most dynamic plant genomes (~0.03disruptions/Mb/MY) (*30, 132, 133*) . However, all these conclusions must be taken with caution because all of the data were drawn mostly based on the estimation of gross chromosomal rearrangements and the relatively poor resolution of most current comparative maps of mammals and also plants (*33*).

Substantial variation in rearrangement rate is also evident within the genus *Drosophila*. In general, the evolution rate within the *Sophophora* subgenus is faster than

that within the *Drosophila* subgenus (*31, 34, 64*). This pattern is consistent with the distribution of polymorphic inversions in these subgenera (*34, 134, 135*). In the genus of *Drosophila*, the evolutionary rates between species of *D. melanogaster* - *D. pseudoobscura* (0.064-0.142 breakpoints/Mb/MY) probably are the highest currently recorded in *Drosophila* (*63, 64*). This evolution rate was calculated based on the whole genome sequence comparison between two species. The intermediate rates of fixation inversions between *D. melanogaster* and *D. replete*, which representing the two farthest lineages within the genus, have been discovered (*30, 33*). The lowest rate for the entire genus is observed when comparing *D. repleta-D. buzzatii* (*34, 73, 136*). In vertebrates, the rate of mammalian chromosome rearrangements varies dramatically among lineages. For instance, > 15-fold changes in the rate of mammalian chromosome evolution between different vertebrate lineages have been reported (*44, 137*). Some vertebrate lineages (e.g., human, carnivores, and the common shrew) show remarkable conservation, while others (e.g., rodents and primates) show extensive chromosomal rearrangements (*105, 138*). Since the two lineages diverged, rodent lineages evolved far more rapidly than primates (*32*).

Beside the variation of chromosome rearrangements among lineages, different chromosomal arms or chromosomal elements showed uneven evolutionary rates. The most important observation in the genus of *Drosophila* is the difference in the evolutionary rates of rearrangement fixation between the sex (X) chromosome and the autosomes (*33, 37*). For example, the highest rate of rearrangements belongs to element A (X chromosome); the lowest belonging to element B, with intermediate rates for element D and E having been revealed in comparisons between *D. repleta* and *D.*

melanogaster (*33*). In the *virilis* group of species, comparative analysis of three species of the group (*D. virilis*, *D. novamexicana* and *D. Montana*) showed that the X chromosome was four times more rearranged than chromosome 3 (*37*). In addition, element A shows more drastic changes in its chromosomal organization within the *obscura* group and between *obscura* and *D. melanogaster* (*71, 80, 136*). The shortest syntenic blocks of this element between *D. pseudoobscura* and *D. melanogaster* further supported this observation (*63*). In addition, the most recent chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes suggested that chromosome X appeared to be the most rearranged across all species (*64*). This evidence suggested the faster evolution of X-lined rearrangements compared to that for autosomal ones, as was proposed earlier (*139*). In contrast to the genus *Drosophila*, the studies of genome rearrangement in the vertebrate indicate that the X chromosome shows extensive gene order conservation as compared to autosomes (*140, 141*).

5. The distributions of inversion breakpoints

A fundamental question in the study of chromosome evolution is whether the breakpoints of chromosome rearrangements are distributed randomly (uniformly and independently) or if they are happening along evolutionary “faults” (hot spots of rearrangements). To answer this question, Nadeau and Taylor in 1984 introduced the notion of conserved segments (i.e., segments with preserved gene order without disruption by rearrangements) (*43*) when they were studying linkage maps between human and mouse. They provided convincing evidence in favor of the random breakage model of genomic evolution proposed by Ohno (*142*). The main point of the random breakage model is a random distribution of chromosome rearrangement breakpoints

along chromosomes. This model was later further supported by significantly large datasets with increased resolution of maps (*28, 89, 143-145*). In addition, the comparative analysis data obtained from the *Drosophila* species strongly supported the random breakage model (*30, 33*). Therefore, the random breakage model was described as the *de facto* theory of chromosome evolution at that time. However, some data in comparison species in the genus *Drosophila* have revealed that two or more inversion breakpoints clustered together that cannot be explained by the random breakage model (*34, 146-148*). For instance, in the *repleta* species group, 96 breakpoints of the 208 inversions were clustered closely and this rate of coincidence still holds true within subgroup or even within the genus (*34, 147, 148*). Similar observations have been found in the *Anopheles* mosquitoes (*13*). These results are based on the comparative mapping and may suffer from the limited resolution of this method.

With the availability of the draft genome sequences of human and mouse, Pevzner and Tesler in 2003 (*90*) found 281 synteny blocks shared by human and mouse of size at least 1 Mb and the lengths of these synteny blocks still fit the exponential distribution. In addition to these large synteny blocks, they also discovered another 190 short synteny blocks, typically below 1 Mb in length. A similar observation has been made by Kent et al. (*91*) almost at the same time. These microsynteny blocks were never discovered in the comparative mapping studies, and even with the human and mouse sequences available, most of them are hard to find (*149*). The existence of these microsynteny blocks implied that the lengths of synteny blocks do not fit random breakage model. Since these small synteny bloaks are defined by closely located microinversions. The existence of the sites of frequent “breakpoint reuse” in evolution

cannot be explained by the random breakage model (*62*). This led to the formation of fragile breakage, that suggests that chromosomal breakages tend to reoccur at 'fragile sites' or 'hotspots' in mammalian chromosomes (*90*). The traditional definition for synteny blocks or conserved segments (a chromosomal region containing a set of two or more markers in the same order without disruption by rearrangement breakpoints) was challenged, since microinversions were found within these synteny blocks (*56, 58, 89, 150*). Thus a relax criteria was adopted to define synteny blocks. Synteny blocks are segments that can be converted into conserved segments by micro-rearrangements (*90*). In 2004, Sankoff and Trinh argued against the fragile breakage model because Pevzner and Tesler did not provide any proof of the breakpoint reuse phenomenon and they doubted the algorithms Pevzner and Tesler used for the determination of synteny blocks (*60*). Although both of models were supported by enormous evidence, the controversy still continues with the advent of new data. The flaw in the fragile breakage model is that no direct evidence for the fragile regions has been provided (*151*). Despite this flaw, more and more studies provide additional evidence in favor of breakpoint reuse and fragile regions (*61, 93, 152, 153*). In the genus *Drosphila*, for instance, chromosomal rearrangement inferred from the comparison of 12 *Drosophila* genomes (*64*) and the complete sequences of *D. melanogaster* and *D. pseudoobscura* (*63*) revealed that the distributions of inversion fixation breakpoints are not random and a large number of small synteny blocks cannot be explained by the random breakage model (*43*). The random breakage or fragile breakage model can be a matter of resolution (*154*). Two reasons were given to answer the question about the presence of 'hotspots' and 'solid regions' in eukaryotic genomes: (1) the proportion of chromosomal

regions involved in the adaptive selection is probably small because of the fitness cost of eliminating unfit mutations at large loci (*154, 155*); (2) some chromosomal regions, such as regulatory blocks in insects, cannot be changed (*156*) since their functions are vital for the development of all eukaryotes (*154*). It is still unclear where these hotspots are located. What are molecular features of these fragile regions? Why are some evolutionary breakpoints apparently reused?

6. Molecular mechanisms of chromosomal rearrangements

Although evidence suggests that the chromosomal elements evolve at different rates and inversion breakpoints are distributed non-randomly, our knowledge of the molecular mechanisms for generating chromosomal inversions is still very limited. The prevailing view is that chromosomal rearrangements are facilitated by ectopic recombination events between inverted repetitive sequences. This well known ectopic recombination model, also known as illegitimate recombination or non-allelic homology recombination, proposed that if repetitive sequences are located on the same chromosome and in inverted orientation, then the consequence of homologous recombination is a chromosomal inversion (*157, 158*). In addition, transposable elements (TEs) can induce chromosomal rearrangements through aberrant gene conversion and directly by an alternative transposition process (*159*). The role of repetitive elements in the generation of rearrangement breakpoints has been supported by the studies in *Drosophila*, *Anopheles*, nematode, and yeast. To date, seven polymorphic and fixed inversions from *Drosophila* and three from *Anopheles* have been cloned and sequenced. In three of the seven analyses, strong evidence implicating the role of TEs or other repetitive elements in the generation of chromosomal rearrangements was found (*160-162*). In two of the

three analyses of the genus *Anopheles*, transposable and repetitive elements have been found at the breakpoints (*163, 164*). Although several studies have detected the absence of repetitive elements in the breakpoints, their possible involvement in the origin of the inversion cannot be precluded given the time elapsed between the species. After the generation of inversions, copies of TEs flanking the inverted segment of inversion will be brought to the fixation as well; if the inversion succeeds, it will be fixed. These TE insertions may last in the genome several million years, but may be removed by deletions in the long run, since the average time to loss of 50% nonfunctional DNA is 12 MY (*34, 165, 166*). In addition, several lines of indirect evidence (*167-170*) and a study of comparative genome sequencing of *D. melanogaster* and *D. pseudoobscura* suggested that the sequences within the breakpoints are rich in repetitive elements (*63*). The role of repetitive elements in the generation of rearrangement breakpoints has also been reported in nematodes and yeast. In *S. cerevisae*, transposons (Ty elements) or long terminal repeats (LTRs) were found frequently associated with rearrangement breakpoints (*32, 69, 171-173*). In nematodes, there is evidence that the breakpoints of translocation since the divergence of *C. elegans* and *C. briggsae* are associated with repetitive elements, and during the evolution of nematode genomes, most of the rearrangements were generated in repeat-rich chromosome regions (*58, 174*). All of the data suggested that repetitive sequences might be the cause of the majority of inversions in *Drosophila*, *Anopheles*, nematodes and yeast; but is not the only cause.

The mammalian genome is quite different from that of invertebrates; segmental duplications, also known as low-copy repeats (LCRs), are rich in the genomes, representing ~5% of the human genome (*175*) and ~2% of the mouse genome (*176*).

Segmental duplications, either direct or inverted, induce the chromosomal rearrangements by unequal cross over (*177*). A growing body of evidence suggests that segmental duplications play a major role in primate genome evolution. Although the significance of segmental duplications in the generation of inversions has been addressed by enormous amount of data in primates, more than 50% of rearrangements in primate have no SDs. There is a striking association between rearrangement breakpoints and large segmental duplications in comparison of human and mouse genomes (*152, 178-182*). It has been estimated that 25-53% of evolutionary rearrangement breakpoints are associated with segmental duplications between human and mouse (*99, 152, 183*). When comparing human and the great-ape, segmental duplications were found in flanking regions of 70-80% of inversions and ~40% of deletions/duplications (*95, 97, 184-187*). Interestingly, in *A. gambiae*, the 14.6 Kb duplications were found inserted between both breakpoints of the 2Rj polymorphic inversion, which suggested that segmental duplications are also forming inversions in *Anopheles* (*188*).

In a recent paper, the breakpoint regions of the 29 inversions that differentiate the chromosomes of *D. melanogaster* and two closely related species, *D. simulans* and *D. yakuba,* have been analyzed (*189*). The study demonstrated that in 17 out of 29 inversions, duplications of genes or other nonrepetitive sequences are present in opposite orientations at the inversion breakpoints. The data so far are incompatible with the ectopic recombination model. Therefore, Ranz et al 2007 suggests a mechanism of staggered breaks, either isochromatid or chromatid, as the most parsimonious explanation of their origin (*189*). Crucially, the inverted duplications only are present at

the breakpoints of the inversion, the one carrying the rearrangement. In this new model, the duplications at the breakpoints are not the driving force of inversion, but rather the consequence of some of the inversion event (*158*). A similar observation has been found at the breakpoints of *In*(3R)P in *D. melanogaster,* at the breakpoints of the 2La inversion in the *A. gambiae* complex (*164*), at the breakpoints of a pericentric inversion between the homologous chromosome 24 in human and chimpanzee chromosome 10 (*190*), and at the breakpoints of inversion 5g fixed in *D. buzzatii* (*191*). The study of Ranz et al 2007 (*189*) also found that repetitive elements are present in at least one of the co-occurrent breakpoint regions of 62% of the inversions (18 of 29), which cannot be explained by ectopic recombination mechanism. However, the chromosomal rearrangements can also occur by an alternative transposition process of TEs (*159*). Evidence for this possible mechanism has been described for several families of TEs, such as the IS 10/Tn 10 elements in bacteria (*192-194*), Ac/Ds elements in maize and tobacco (*194-196*), Tam 3 in *Antirrhinum majus* (*197-200*), and P elements in *Drosophila (201-203)*. Therefore, all the available data, so far, suggest that the nature and patterns of repetitive DNA such as either TEs or LCRs are the key to understanding the mechanism and dynamics of chromosomal structure among eukaryotic genomes.

7. Polymorphic inversions and fixed inversions in *Drosophila* and *Anopheles*

Although evidence suggested the role of repetitive DNA in the generation of inversions, little is known about the forces responsible for establishing inversions. After the origin of new chromosomal mutations, the majority of them will be lost due to genetic drift (*204, 205*). If the inversion escapes elimination, selective processes may take control of its fate, which can end in maintenance or loss due to directional selection

or genetic drift (*206, 207*). Although both genetic drift and natural selection drive evolution, genetic drift operates randomly, while natural selection functions nonrandomly. In the studies of polymorphic inversions within species and fixed differences between species (*208*), most of data suggested that breakpoints of inversions are distributed nonrandomly (*13, 24, 33, 134, 135, 147, 148, 209*). Interestingly, the most effective malaria vectors such as *A. gambiae*, *A. funestus,* and *A. stephensi* are highly polymorphic in terms of chromosomal inversions. For example, 31 polymorphic inversions were reported for the *A. gambiae* complex (*13*). Additionally, polymorphic inversions are distributed non-randomly among five chromosomal arms in the mosquito genome. In *A. gambiae s.l.*, 18 of the 31 polymorphic inversions (58%) are found on chromosome 2R (*13*). So far, in *A. funestus*, 11 of 17 polymorphic inversions have been identified on the 2R chromosomal arm (*26*). 3L and 2R are the most rearranged arms in *A. stephensi* (*17, 41*). However, the only successful and wide-spread inversion is 2Rb in this species. No polymorphic inversions have been found on the X chromosome in *A. gambiae s.s.*, *A. funestus* or *A. stephensi*. In addition, available data obtained on various organisms strongly suggest that chromosomal polymorphism is a mechanism of rapid adaptation of species to climate change and speciation. Theoretically, the evolutionary importance of inversions might come from their ability to produce genetically isolated populations. Inversions inhibit recombination between standard and inverted karyotype in heterozygotes because single crossover events produce inviable, unbalanced gametes. In *Drosophila*, clinal and seasonal changes are often found associated with the inversion frequency (*209-214*). In *Anopheles* mosquitoes, chromosomal polymorphisms have been proposed to play an important

role in speciation and ecological adaptation (*13, 42, 215*). Of the seven members of the *A. gambiae* complex, *A. arabiensis* and *A. gambiae* s.s., representing the major vectors of human malaria, have highly polymorphic inversions and a continent-wide distribution in the arid sub-Saharan Africa. The other minor or nonvector species with little or no chromosomal polymorphism tend to occupy smaller and wetter regions. Additionally, differential adaptations of mosquitoes to various environments are often associated with dramatic changes in composition and frequency of polymorphic chromosomal inversions. For instance, the 2Rb, 2Rbc, 2Rd, and 2La inversions are frequent in arid Sahel Savanna and almost absent in the humid equatorial Africa, strongly suggesting that these inversions confer adaptive fitness to the drier environment (*13, 42, 216*). Moreover, frequencies of these inversions are higher indoors where a nocturnal saturation deficit is higher than outdoors. Therefore, nonrandom distribution of inversion polymorphisms and their association with adaptation and speciation strongly suggest that these rearrangements are the product of selection. Establishment and maintenance of chromosomal polymorphism has been explained by different mechanisms. The local adaptation model proposes that an inversion will spread if it carries a set of locally adapted alleles. Some alleles will cause it to spread to fixation, while others will lead either to a neutral or a selectively maintained polymorphism (*208*).

Comparative physical mapping has been performed only once between *A. gambiae* and *A. funestus* in *Anopheles* mosquitoes (*1*). Taxonomically, *A. gambiae*, *A. funestus,* and *A. stephensi* belong to different series: *Pyretophorus* (*A. gambiae*), *Myzomyia* (*A. funestus*), and *Neocellia* (*A. stephensi*) of the subgenus *Cellia* (*14, 16*). These distantly related lineages diverged from a common ancestor at least 30 million years ago, and

are good model systems for studying chromosomal rearrangements (*217*). Availability of the genome sequence for *A. gambiae* and physical maps for *A. funestus* enabled a fresh perspective on the relationships between the genomic landscape and evolutionary rates. Our second goal was to identify the patterns of inversion fixations in *Anopheles* mosquitoes and to analyze the molecular features associated with fast and slow chromosomal evolution. Our aims were (1) To idenfity the patterns of inversion fixations in *Anopheles* mosquitoes using the genome sequence of *A. gambiae*, and physical maps for *A. funestus* and *A. stephensi*. (2) To study the distribution of inversion fixations between A. gambiae and A. stephensi by the N - T model. (3) To analyze the repetitive elements using TRF, repeatmasker, and M/SARs with SMARTest associated with fast and slow chromosomal evolutions. (4) To study the relationship between polymorphic and fixed inversions using statistical methods.

8. *A. gambiae* complex – general information

The Afrotropical *Anopheles gambiae* complex is comprised of seven closely related species that are nearly morphologically indistinguishable. Despite almost being morphologically identical, individual species of *A. gambiae* complex have distinct ecological adaptations, geographical distributions, and behaviors which all contribute to differences in vectorial capacity. Among the members of the *A. gambiae* complex, *A. gambiae* sensu stricto (*A. gambiae* s.s.) and *A. arabenesis* represent the most epidemiologically important vectors of human malaria in Africa because of their close association with human habitat, anthropophily, and efficiency in transmitting malaria parasites. These two species are responsible for approximately 90% of global malaria morbidity and mortality in the world. *A. gambiae* s.s. and *A arabiensis* are freshwater

species and are the most widespread, occurring in sympatry throughout most of the sub-Saharan Africa and its off-shore, although *A. gambiae* s.s. is prevailing in the rain forest and *A. arabiensis* in xeric habitats (*42*). Among the members, *A. merus* and *A. melas* are salt-water breeders and are confined to the east and west coast of Africa, respectively. *A. bwambae* is found only in Semliki Forest mineral springs in Uganda (*42, 218*). Both of the salt-water breeders and A. *bwambae* are minor vectors of human malaria. However, the important role of *A. merus* in malaria transmission in Madagascar has been reported (*219*). Historically, *A. quadriannulatus* was considered as a species restricted to southern Africa, south of the Zambezi River. Recent collections from the Jimma in Ethiopia, south-west of Addis Ababa revealed that *A. quadriannulatus* species B is a distinct biological species which is differentiated from the traditional *A. quadriannulatus* species A (*220*). Both *A. quadriannulatus* species A and B strongly prefer to feed on animal blood and do not contribute to the human malaria transmission (*13, 42, 221*). To understand the genetic changes associated with these epidemiologically important phenotypes, the demonstration of the phylogenetic relationships among members of the *A. gambiae* complex is crucial.

9. The molecular and chromosomal phylogeny in the *A. gambiae* complex

DNA sequence data of a gene *within th*e X^{ag} (*227*) and AT-rich control region of the mitochondrial DNA (mtDNA) of six species and the closely related *A. christyi* (*228*), supported the phylogenetic affinity of *A. gambiae* and *A. merus*. However, phylogenetic analysis sequences from mtDNA, X-linked ribosomal, and two chromosomal genes (both outside of the shared inversion) strongly supported a sister taxa relationship of *A. gambiae* and *A. arabiensis,* and suggest gene flow between them (*223, 228-230*). As a

result, phylogenies inferred from individual genes and other DNA markers often conflict with each other, suggesting either *A. arabiensis* or *A. merus* as a sister taxon to *A. gambiae* (*222-225*). Additionally, attempts to reconstruct the phylogenetic history of the *A. gambiae* complex using the sequences of genes within or near the shared breakpoints have failed because of strongly conflicting results (*223*). All these contradictory data may result from the high level of sequence similarity and genetic introgression, along with shared ancestral molecular polymorphism among the seven closely related members of the *A. gambiae* complex. A recent genome-wide analysis of microsatellite markers revealed a mosaic genome architecture in the *A. gambiae* complex and provided direct evidence of introgression in different genomic regions (*231*). Therefore, the phylogeny based on the DNA sequences is not suitable for closely related species.

An alternative approach to infer the phylogeny among species of the *A. gambiae* complex is from ten fixed inversions. Reconstruction of the phylogenetic relationships from chromosomal inversion dates back to the classical work of A. Sturtevant and T. Dobzhansky (*232*). Since then, this methodology has been successfully applied to a number of species, especially insects with polytene chromosomes, such as *Drosophila* (*233, 234*) and *Anopheles* (*13, 42, 221*). A recent study revealed that, in general, chromosomal inversion topology is in completely agreement with DNA sequence topology and it also suggested that inversion data are shown to be more information rich than nucleotide data (*234*). The method relies on the reasonable assumption that inversions are monophyletic in origin; in other words, all the inversions were generated by unique breaks and all present-day carriers of the inversion must have a shared

common ancestor (*235*). Therefore, an ancestral karyotype can be determined if outgroup arrangements are known and inversions are monophyletic (*136, 236, 237*). Although breakpoint reuse occurs in evolution, there is no evidence for multiple origins of the same inversion in the *A. gambiae* complex (*34, 164, 189*). Based on the ten fixed inversions, Coluzzi and colleagues (*42, 238*) produced a phylogenetic network for the *A. gambiae* complex. In this network, the sister group relationships *A. gambiae* + *A. merus* and *A. melas* + *A. bwambae* were proposed. Later, a reconstruction of the *A. gambiae* complex phylogeny has been attempted using fixed inversion and polytene chromosome maps of outgroup species (*226*). Although they failed to identify the ancestral species because of the low resolution of cytogenetic maps, the sister group relationships proposed by Coluzzi have been confirmed (*226*). For a long time, *A. quadriannulatus* (homosequential species A and B) has been considered ancestral because it carries all the “standard” chromosomal rearrangements and its central position to other species in the complex (*13, 42*). Later, however, *A. arabiensis* was assumed an ancestral species and two sources of evidence were given: “it is the only member of the complex present in the Horn of Africa and in the Arabian peninsula”(*221*) and the fixed 2La inversion has been cytologically identified in two outgroup members of the *Anopheles subpictus* complex (*221*). Ayala and Coluzzi also proposed that *A. arabiensis* originated in the Middle East and reached Africa through the arid Arabian peninsula (*221*). Among ten fixed inversions, only both breakpoints of alternative 2La rearrangements have been sequenced. The presence of pseudo gene copies on the standard 2L breakpoint revealed that the 2La inversion is an ancestral chromosome arrangement (*164*). However, no further evidence confirms that *A. arabiensis* is more

likely close to the ancestral species.

Our third goal was to identify ancestral and derived inversion arrangements in the *A. gambiae* complex, and therefore attempt to reconstruct the phylogenetic history of the *A. gambiae* complex. Our specific aims were to identify the ancestral status of the 2La, 2Rop inversions using gene order comparisons between *A. gambiae* and *A. stephensi*, or *A. nili* and computational programs, MGR and SRPING. Then to analyze the molecular organization of 2Rop inversions breakpoints from alternative rearrangments by BLAST and vectorbase search tools.

10. 2La inversion and malaria transmission – general information

Current tools for controlling malaria are multifaceted and include the use of insecticides, antimalaria drug treatment, and vector control. However, current efforts to control this disease are becoming less effective as insecticide resistance grows (*239*). Therefore, there is an urgent need to develop new strategies for malaria control. One novel approach would be to target genes responsible for vector competence. This novel strategy for the control of malaria transmission involves introduction of genes for refractoriness into wild populations of mosquito vectors so that they become unable to support a parasite development. This approach was proposed by Miler and James in 1992 (*240, 241*). Nevertheless, it is in its infancy and much remains to be done before we can evaluate the feasibility of this strategy in nature. First and foremost, it is necessary to identify effective genes responsible for refractoriness and understand the mechanism of the genes action. *A. gambiae* is the most efficient African vector of human malaria, which is caused by *Plasmodium falciparum*. The genome of *A. gambiae* harbors a number of polymorphic inversions within the species. The 2La inversion is the

most well studied polymorphic inversion. Individual *A. gambiae* mosquitoes can bear alternative chromosomal rearrangements: 2La (inverted) and 2L+^{a} (standard) on the left arm of the second chromosome. The inversion is widespread across natural populations of *A. gambiae*. The studies of the 2La inversion in natural populations revealed that it is associated with at least two epidemiologically important phenotypes: the ability of malaria transmission (*242*) and adaptation to the aridity (*13, 42, 216*). The cumulative evidence on laboratory strains of *A. gambiae* suggested that the alternative rearrangement of 2L chromosomes strongly correlate with refractoriness and susceptibility to various *Plasmodium* species (*243-246*). In addition, a recent genetic survey of an *A. gambiae* natural population has identified the strongest *Plasmodium falciparum* resistance locus in a small region of 2La inversion near the proximal breakpoint (*247*).

10.1. Inversions and genetic differentiation

The 2La inversion has been found to be associated with malaria transmission and the degree of aridity, which points to the adaptive value of this polymorphism (*248-251*). However, how an inversion can capture adaptive genes and facilitate their maintenance in nature is still not clear. The main evolutionary importance of chromosomal inversions might come from the fact that they reduce recombination in heterozygotes. Paracentric inversions were first discovered by their effect of inhibiting the genetic change between alternative rearrangements in *Drosophila (252-254)*. This occurs because single crossovers within inversion loops in heterozygotes (heterokaryotypes) result in unbalanced gametes, which are either nonfunctional or produce nonviable zygotes (*255*). If chromosomal rearrangements bind together favorably coadapted gene

complexes (*256*) and/or capture multiple genes which are individually adapted to the local condition (*208*), then natural selection will drive these inversions to high frequency and allow them to establish and spread. Another mechanism involving the effect of the chromosomal inversions is the position effect model. This model proposed that the locations of genes near or inside inversion breakpoints might affect the functions and expression profile of those genes (*257*). Therefore, both mechanisms might work together and allow chromosomal inversions to capture coadapted or locally adapted genes, for instance, the genes within 2La related to the malaria transmission and adaptation to the dry climate, against genetic exchange with migrants from other genetic background (*258*).

10.2. The 2La inversion and adaptation to aridity of *A. gambiae*

A. gambiae is the most competent vector of human malaria and distributed widely throughout most of the sub-Saharan Africa and it's off-shore. The successful invasion and adaptation of *A. gambiae* to most ecosystems on the continent was thought to be facilitated by a number of polymorphic chromosomal inversions (*248-251, 259, 260*). One inversion found strongly linked to the aridity clines in Africa is the 2La inversion in *A. gambiae*. The 2La inversion polymorphism is widely spread in natural populations of *A. gambiae* and is associated with at least two important phenotypes: *Plasmodium* infection rates (*242, 245, 247*), and adaptation to aridity (*13, 42, 216*). Of them, the relationship of the 2La inversion and adaptation to aridity has been extensively studied. In many different locations in Africa, the frequency of the 2La inversion is strongly associated with the degree of aridity in the West African population (*42, 249*). The $2L+^{a}$ arrangement is pervasive in southern Nigeria and southern Cameroon and decreases

progressively to reach a high frequency of 2La fixation in the north of these countries (*42, 249, 261*). Similarly, the seasonal fluctuations and microspatial clines for this inversion have been observed (reviewed in (*216*)). In addition, the 2La inversion was more commonly found in the mosquitoes resting indoors where a nocturnal saturation deficit exists (*216*). The recent studies in the laboratory suggests that 2La *A. gambiae* larvae are better equipped to resist potentially lethal temperatures than those of standard (*259*) and 2La inversion is associated with enhanced desiccation resistance in *A. gambiae (260)*. It was even hypothesized that the 2La inversion has captured a set of locally adapted alleles, which together confer an advantage to its carriers in arid conditions (*216, 258, 260*). Therefore, the 2La inversion may enhance *A. gambiae* survival, may have contributed to the wide distribution of *A. gambiae* in Africa (*13*), And may consequently facilitate malaria transmission in the more xeric parts of its distribution.

10.3. 2La inversion and malaria transmission in *A. gambiae*

The earliest evidence of the relationship between the 2La inversion and malaria transmission was reported by Petrarca and Beier (1992) in Western Kenya (*242*); In the natural population of *A. gambiae*, the $2L+^{a}$ is found to be associated with a two-fold higher rate of *P. falciparum* infection than 2La (*242*). However, the causes of intraspecific variation in *Plasmodium* infection rates were not clear. They hypothesized that the variation of *Plasmodium* infection rates could result from multiple interacting factors, such as differences in longevity, change of behavior, or vectorial capacity (*242*). There are more than 120 *Plasmodium* species which infect birds, reptiles, and diverse mammals including humans. Different *Plasmodium* species utilize different mosquito

vectors for transmission. As a major vector of human malaria, *A. gambiae* only transmits *P. falciparum*. In order to study the molecular mechanisms of the human malaria vector *A. gambiae*'s refractoriness to *P. falciparum* in nature, two genetically selected systems of malaria resistance have been developed using different *Plasmodium* species, leading to two major resistance mechanisms being identified. The genetic factors affecting these two mechanisms have been found to be associated with the 2La inversion (*243, 246*).

10.3.1. Melanotic encapsulation in the hemocoel

Two lines of *A. gambiae*, which are highly refractory (R line) and highly susceptible (S line) against the simian parasite, *P. cynomolgi,* have been genetically selected in the laboratory *(243)*. In the refractory mosquitoes, ookinetes were encapsulated and killed in the wall of the midgut after an effective bloodmeal. Netherless, in the susceptible line, parasites developed normally (*243, 262*). Later studies revealed that the genetic mechanism underlying the phenotypic differences between the lines in *P. cynomolgi* infection by the G3 strain of *A. gambiae* is controlled by an autosomal locus containing the esterase gene (Est) (*244*). Esterase is one of the serine proteases, which is involved the activation of the protease cascade, leading to the encapsulation and melanization of foreign material after infection (*263*). Two alleles, the Est A allele associated with refractoriness and the Est C allele associated with susceptibility, assorted at this locus (*244, 264*). The authors hypothesized that this set of alleles were linked to a polymorphic inversion in *A. gambiae (244)*. In 1993, Crews-Oyen et al. proved that in the *A. gambiae* G3 strain, different arrangements of a polymorphic inversion on the left arm of chromosome two (the 2La inversion) are associated with different alleles of the

esterase locus (*245*). However, the association of the 2La inversion with infection rates in the G3 strain was exactly opposite of what Petrarca and Beier found in Kenya. In this study, mosquitoes carrying 2La inversion had significantly lower *P. falciparum* infection rates (*242*). It is possible that a 2La polymorphic population of G3 strain and wild mosquitoes in Kenya captured the different alleles.

10.3.2. Intracellular ookinete lysis in the midgut epithelial cell

Resistant and susceptible *A. gambiae* lines against the avian parasite, *P. galinaceum* were selected in 1995 by Vernick et al. (*246*). The resistance mechanism was manifested as lysis of ookinetes in midgut epithelial cells (*246, 265*). This refractoriness mechanism does not involve encapsulation of the parasite described previously, but it is associated with the 2La inversion as well. Further analysis of this new intracellular killing mechanism by a genetic crossing experiment demonstrated that the resistance is controlled by a single dominant locus (*246*). But no further evidence demonstrated the genomic location of this dominant locus.

10.3.3. The plasmodium resistance island was mapped to the 2La inversion

Despite the significance of the encapsulation and lyisis mechanisms tested in laboratory strains of *A. gambiae* with simian and avian parasites, these systems do not represent natural vector-parasite combinations. Resistance of wild *A. gambiae* to *P. falciparum* was examined in Mali, West Africa (*266*). Genetic mapping of infection intensity of *A. gambiae* infected with natural *P. falciparum* identified one major resistance locus, which explained almost 90% of the parasite-free mosquitoes in the segregating pedigree. This major QTL (Pfin1) was located on chromosomal arm 2L of *A. gambiae* (*266*). But the exact genomic location of this natural resistance locus in

mosquitoes remained unclear. A more recent genetic survey of the *A. gambiae* natural population in Mali has identified the strongest *P. falciparum* resistance island (PRI) of *A. gambiae,* which controls the majority of naturally segregating variation for *P. falciparum* infection. The genomic position of this PRI was located near the proximal breakpoint of the 2La inversion (*247*). Among the candidate genes in this chromosome region, the role of APL1 encoding a leucine-rich repeat protein in *P. berghei* resistance has been confirmed by RNAi (RNA interference) and further study showed that this gene was located within 1 Mb from the proximal breakpoint of 2La inside inversion (*164, 247*). Therefore, all the evidence from the laboratory strain and wild population of *A. gambiae* against different *Plasmodium* parasites suggests that the 2La inversion is significantly associated with *Plasmodium* infection rates. But the molecular mechanism and alleles responsible for the variation of phenotypes are still not known.

11. Laser Capture Microdissection (LCM)

Scalenghe et al (1981) first developed polytene chromosome microdissection and microcloning technique using *Drosophila melanogaster (267)*. In this microdissection technique, the chromosomes were initially dissected with glass microneedles controlled by a micromanipulator under an inverted microscope (*267*). This technique was then applied to the microdissection of metaphase chromosomes in mouse and 212 microclones from the proximal half of chromosome 17 were obtained (*268*). Later, this method was extended to the human metaphase chromosomes (*269*). At that time, most of studies only focused on the chromosomes easily identified such as the X chromosome (*270*). After the advent of G-banding and PCR techniques, the microdissection and microcloning have been extensively used in the metaphase

chromosomes of human and animal genomics research (*271-276*). The applications of microdissection in plants have been only limited to the chromosomes, which are easy to identify, since chromosome preparation is more difficult in plants than human and animals (*271, 277, 278*). However, mechanical microdissection is strongly operator dependent, time consuming and technically challenging. Even for an expert, it is difficult to dissect and collect a large number of chromosomal fragments from the same regions. Therefore, it is being replaced by laser capture microdissection.

The advent of laser based microdissection technique has marked a new era in microdissection. In 1986, Monajembashi et al. developed a new method to dissect metaphase chromosomes of human lymphocytes using a UV laser microbeam coupled into a microscope (*279*). Unfortunately, this approach didn't find its way into routine biomedical investigation because of the requirement of bulky equipments and susceptibility to difficulties. LCM was initially designed for accurate and efficient selection and isolation of single cells from a heterogeneous tissue samples in order to perform molecular analysis of tumor cells by National Cancer Institute (NCI) (*280*), which was rapidly commercialized by NCI & Arcturus Engineering, California, USA (www.arctur.com) as PixCell System. Other companies subsequently developed new systems for LCM with various characteristics regarding the cell collection and the laser source, etc. There are two general classes of LCM: infrared (IR) capture system and ultraviolet (UV) cutting systems. Today, IR capture system produced by Arcturus Engineering, and UV system by PALM Microlaser Technologies (http://palmmicrolaser.com) and Leica Microsystem (Leica; http://www.leica-microsystems.com) are among the most popular. The principle for each system was

demonstrated in **Figure 1.1**. The pulsed UV laser microdissection allows treating material in a non-contact manner, which minimizes the risk of contamination as compared with IR capture system. The later studies using these systems revealed that almost no damage was done to the structures of the invested materials because of laser's low energy and short pulse duration, however, nucleic acid may result in the partial destruction since the cells are subjected to heat as well as photons from the laser itself (*281, 282*).

12.

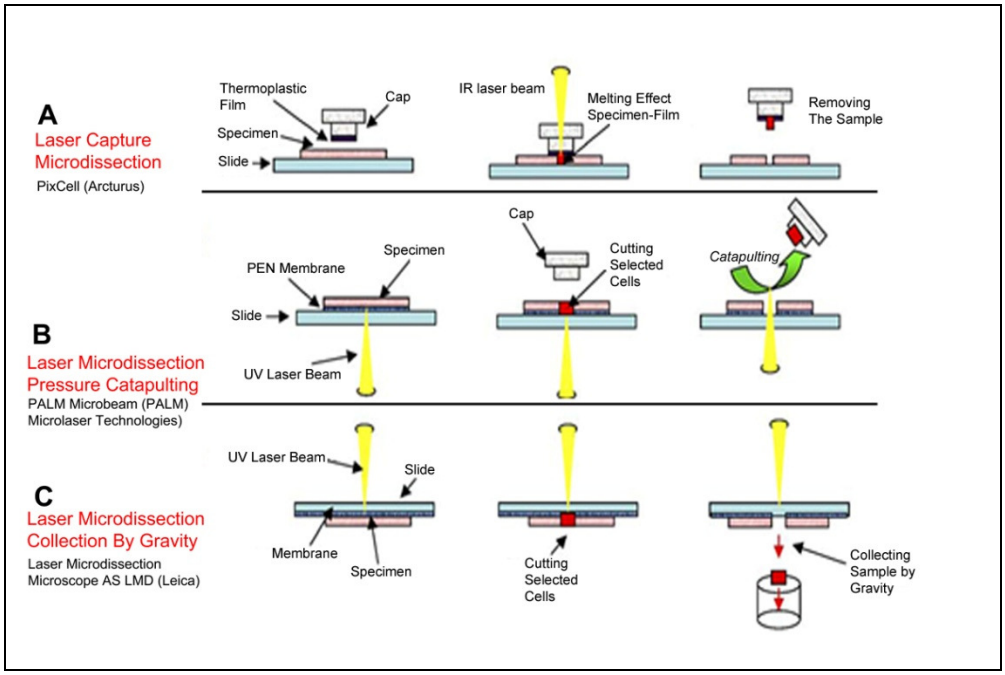


Figure 1.1: Principles of laser assisted microdissection techniques (*282*). (A)The steps involve in LCM. (I) after visualizing of interest under inverted light microscope, a cap with thermoplastic polymer film is placed on the tissue. (II) Laser beam activates the polymer film that melts, expands and surrounds the cells of interest. (III) Subsequently the capture, the cap is removed, effectively microdissecting the cells of interest from the heterogeneous tissue section. (B) The steps of UV laser cutting system. (I) After the tissue, cells or chromosomes are mounted on the membrane slide (PEN or PET

membrane slide), the samples can be visualized under an inverted microscope. (II) The cutting material can be marked using computer software; UV laser performs the cutting of the marked region. A cap with adhesive lid or the cap with Catapult Buffer is then placed on the selected area. (III) Subsequently, the dissected material is catapulting into the adhesive cap or the cap with Catapult Buffer of a tube positioned above the slide. (C) The principle of the Leica AS LMD UV laser beam system. (I) Tissue is spread on a thin transparent membrane which is mounted on a glass slide. The tissue is visualized under an upright microscope. (II) The UV laser performs the cutting of the selected cells. (III) The microdissected sample falls into a collection tube positioned below the slide by gravity.

From glass microneedle to laser microbeam, microdissection methods have undergone change from manual operation to computer driven manipulation. As a result, both the rate and precision of microdissection are improved and the applications have been extended from the microcloning of DNA markers from polytene chromosome and metaphase chromosomes to isolation of single cells, specific tissue population within a microenvironment, even sub-cellular components, such as nuclei, nucleus free cytoplasm, and chromosomes (*283-286*). At present, three classes of biomolecules: DNA, RNA and proteins can be recovered from microdissected materials (*287*). Due to the limitation of sample preservation, tissue microdissection is currently more widely employed to analyze DNA than RNA or protein. Since the latter two biomolecules are more sensitive to degradation and fixation (*286*).

The advent of LCB represents an utmost important and interesting technique in molecular pathology and creates a link between histology and molecular analysis. The

most current studies focus on the characterization of genetic alternation in pathologic conditions such as Chronis pancreatitis (*288*), motor neuron disease (*289*) and various pre-malignmant (*290*), and malignmant tumors (*290-292*). In addition, LCM also made a major contribution to studies aimed at understanding the gene expression features (*293-297*) and proteins in defined cell population. In cytogenetics, chromosomal microdissection is an extremely valuable tool for isolating DNA from any cytogenetically recognizable region of a chromosome. The isolated DNA can be used for genetic linkage and physical map construction, generation of chromosomal painting probes and generation of chromosome specific expressed sequence tags libraries (*298*). Traditionally, the probes for fluorescence *in situ* hybridization (FISH) usually generated from the clones of cosmid library, yeast artificial library (YAC), bacterial artificial library or PCR products. The microdissected DNA was also found to be suitable for use as a labeled probe for *in situ* hybridization. The new technique which combines chromosomal microdissection and chromosomal painting is named micro-FISH (*298-301*), but has been applied to polytene chromosomes only twice (*302, 303*). The profound impact of microdissection on biomedical research and disease management as well as other molecular analysis has penetrated in every area and any laboratory, and LCM is becoming a routine technology.

12. Whole genome amplification

Current genomic sequencing methods including the SOLID™ system, Illumina and 454 require micrograms of DNA template. Even highly sensitive analytical methods such as PCR are often constrained by a limited amount of DNA template for instance to perform DNA analysis in a single cell (*304*). At the same time, in microbial ecology, low

biomass and the predominance of a few abundant community members have impeded the amplifications of molecular techniques such as PCR and microarray (*305*). Detection of the status of an embryo from a single cell biopsied 2-3 days after fertilization in preimplantation genetics and diagnosis (PGD) can also be limited by insufficient amount of DNA (*306*). Precious clinical samples painstakingly collected in population and epidemiological studies presents a number of complications that are rarely faced in routine molecular diagnostic laboratories. In cytogenetics, to obtain enough DNA from particular chromosome structure has become an important technique for analyzing the unknown DNA sequences such as chromosomal inversions. All these examples strongly suggest that a new technique is urgently required for acquiring large amounts of DNA from limiting samples.

The development of whole genome amplification (WGA) has recently made it possible to obtain microgram DNA from single bacterial cells (*307*). Several strategies have been utilized to amplify the existing DNA samples. The earlier amplification methods are PCR based amplification using Taq polymerase and there are two most relevant principle methods (PEP and DOP-PCR).

12.1. Primer extension preamplification (PEP)

Primer extension preamplification (PEP) was first invented by Zhang et al for the haplotyping of a single sperm cell (*308*). This method uses Taq polymerase and a random mixture of 15-base oligonucleotides as primers. For each 50 cycles, the template is denatured at 92°C. Then the primers are annealing to a low temperature (37°C), and are gradually heated to 55°C followed by a 4 minute elongation at 55°C (*304, 309*). This PCR-based WGA has been widely used for genetic disease in PGD

and prenatal diagnosis for instance the single-cell analysis of sex-linked sequences (*310*), Tay-Sachs disease (*311*), cystic fibrosis (CF) (*312, 313*) and so on (*309, 314, 315*). The analysis for single cells revealed that 228/250 (91%) of loci were successfully amplified (*316*). PEP was also applied on DNA amplification from ethanol-fixed paraffin embedded tissues (*317*) and formalin fixed paraffin embedded tissues (*318*).

12.2. Degenerate oligonuceotide primed PCR (DOP-PCR)

Degenerate oligonucleotide primed PCR is another well established and widely accepted WGA method invented by Telenius et al 1992 (*319*). Unlike PEP, DPO-PCR uses partially degenerated oligonucleotides as primers, which are capable of annealing to many sites at low temperature. Then, the annealing temperature increases to a higher temperature to allow specific fragment amplification (*304, 309*). DOP-PCR in junction with microdisssection has played an important role in chromosomal analysis in PGD. In prenatal diagnosis, comparative genomic hybridization (CGH) allows a full analysis of DNA copy number change of whole genome between normal person and patient. This new method has become a valuable and reliable tool for detection complex chromosome aberrations including aneuploidy, translocations deletions, chromosome mosaicim, and the breakage (*320-322*), and characterization of cytogenetically unclassified aberrations (*323-328*). Apart from the significance of DOP-PCR in cytogenetic analysis, it is also widely used in tumor research such as breast cancer, prostatic carcinoma, and pancreas tumors (*329-333*), single nucleotide polymorphism (SNP) genotyping (*334, 335*) and microsatellite genotyping (*336-338*).

Both of PEP and DOP-PCR have been further developed and refined to increase the fidelity and the length of the amplification products (*338, 339*). A commercial kit

(GenomePlex single cell whole genome amplification kit) was developed by Sigma-Aldrich and was used to amplify single human (*340*) and mouse cells (*341*). Other subgroup of WGA were also developed either based on restriction enzyme cleavage (*342-344*), random shearing of genomic DNA (*345*) or nick translation (*346*). However, all the PCR-based WGA have several limitations factored by Taq polymerase. (1) Short amplification fragments: the length is usually less than 3 Kb. For instance, the amplification fragments of PEP are only 450 bp. The average size of DOP-PCR fragments is 500bp (*304*). (2) Uneven coverage of genome. It should be expected that significant alleles drop out and biased amplification when whole genome is amplified especially when scarce amount of DNA is available or when the DNA template is degraded or partially degraded (*309*). (3) High possibility of mutation introduction into the products.

12.3. Multiple displacement amplification (MDA)

A new method for WGA named multiple displacement amplification (MDA) has been introduced since the discovery of phi29 DNA polymerase derived from the *Bacilllus subtilis* bacteriophage φ 29 (*347*). This enzyme has a extremely high processivity, which shows that the average number of nucleotides added to the 3'-terminus before the polymerase dissociates from the DNA is 70,000bp (*348*). Unlike PCR-based amplification methods, MDA uses phi29 DNA polymerase and random hexamer primers (*306, 349*) and principle is illustrated in **Figure 1.2**. There are several advantages for this new method. (1) It generates fragment sizes > 10 kb because of the high processivity of phi29 DNA polymerase. (2) It results in lower error rate because phi29 enzyme has better proofreading activity (*309*). The error rate is 3 in 10,000 for the

native Taq DNA polymerase, (*350*) and 1 in 10^6-10^7 for phi29 (*351*). (3) MDA has the lowest amplification bias of any WGA methods reported to data (*306, 352*). After the introduction of MDA as a new principle method of WGA, two commercial kits (GenomiPhi V2 DNA amplification kit, GE Healthcare and REFLI-g kit, Qiagen Inc) have been developed.

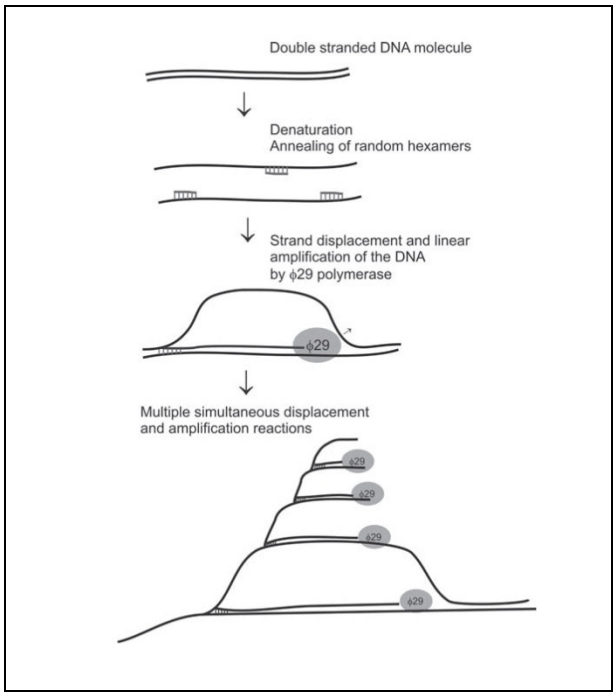


Figure 1.2: Principle of multiple displacement amplification (*353*)

MDA has replaced earlier PCR-based amplification methods and is applied to the medical, cytogenetic and molecular analysis. The combination of laser microdissection followed by WGA has recently emerged and was used for large scale genomic analyses of pure populations of cells. The starting material for WGA ranged from 50-1000 cells (*288-292, 354*). A more recent study revealed that WGA can be performed on microdissected single mouse cell (*355*). However, this combinational approach has been applied on the metaphase chromosome only once (*356*) and the amount of

amplified DNA is still not enough for sequencing. Despite the low amplification bias and dropout rate and high yields of amplicons, MDA results in the loss of some genomic sequences, particularly repetitive and telomeric sequences (*352*). Significant alleles dropout and locus bias can be observed even with large input (10-100 ng) (*306, 307, 349, 357-359*) or with a recently modified protocol (*360*). Alleles dropout rate (ADO) defined as "failure of amplification of one out of two alleles in a heterozygote locus" from laser microdissected single cell is 71.9±13.9 (fresh tissue) and 85.3±3.8 (presorted in 80°C for up to several weeks) (*355*). Another serious limitation which reduces the quality of MDA is the abundant template independent product (TIP). The TIP is derived from primer dimer formation or exogenous DNA contamination (*357, 358, 361-364*). When the starting material is very limited, such as single cells or a subnanogram amount of template DNA, TIPs often represent more than 70% of the total yield (*362, 365-367*). Several efforts have been made to eliminate the TIPs including strict control of experimental procedure to avoid exogenous DNA contamination (*358*), and minimization of reaction volume (*361, 368*) and reaction time (*367, 369*). Although these strategies have successfully reduced the TIPs, they don't completely eliminate TIPs. Therefore, there is an urgent need for developing a new technique to cover evenly on the target template DNA without TIPs. In 2008, Pan et al developed a new method using trehalose and optimized amplification reaction which can amplify microgram amplicons from 0.5-2.5ng of template DNA without TIPs. This new technique results in >99.7% accuracy compared with results on unamplified DNA (*367*). However, how this technique can be used for amplification of laser microdissected chromosomes was unclear.

Comparative genomic study of the *Plasmodium* resistance island (PRI) in members of the *A. gambiae* complex can yield important insights into the mechanism of refractoriness to a parasite. Our fourth goal was to develop a protocol for obtaining the DNA sequence from both breakpoints of the 2La inversion of *A. arabiensis*. Our aims were (1) To determine the chromosomal regions for microdissection by FISH results and the Vectorbase search tool. (2) To isolate chromosomal fragments spanning 2La breakpoints by laser microdissection. (3) To amplify DNA from microdissected chromosomal fragments using V2 DNA amplification and Wpa methods. (4) To test specificity of amplified target DNA by PCR and Sanger sequencing analysis.

Chapter two: A physical map for *Anopheles stephensi*

1. Abstract

A total of 422 BAC and cDNA clones from *A. gambiae*, *A. funestus,* and *A. stephensi* were mapped to the polytene chromosomes of *A. stephensi*. No interchromosomal transposition event was identified between *A. gambiae* and *A. stephensi* which is consistent with the rare occurrence of transpositions in mosquitoes. Of 422 markers, 363 probes were hybridized to single chromosomal sites, while 59 clones yielded multiple signals. Among these multiple located markers, two BAC clones from *A. gambiae*, 141A14 and 146D17 were confirmed to be located in the breakpoints of fixed inversions in the *A. gambiae* complex. Most of the multiple located markers belong to gene families, pseudogenes, or DNA fragments sharing a certain sequence homology. Additionally, several cases of euchromatin and heterochromatin transitions between *A. gambiae* and *A. stephensi* suggest that heterochromatic sequences evolved rapidly. A physical map for *A. stephensi* will facilitate the genome sequence assembly of this species.

2. Introduction

A. gambiae, *A. funestus,* and *A. stephensi* represent the competent malaria vectors, and also are good model organisms for cytogenetics studies. Among these species, several physical maps have been developed for *A. gambiae (20-22, 24)*. 157 markers were mapped to the polytene chromosomes of *A. funestus* (*1*). However, only eight probes were placed on the cytogenetic map of *A. stephensi* (*18*). Therefore, a physical map for *A. stephensi* could facilitate the assembly of DNA scaffolds of this species into chromosomes. It also provides a suitable basis for determining the extent of

chromosome conservation and rearrangements between this species and others, therefore shedding light on mechanisms of chromosomal inversions and on phylogenetic relationships among species (*33, 189*).

For those purposes, the Indian wild-type strain of *A. stephensi* was used to develop a physical map for *A. stephensi*. The BAC and cDNA clones from *A. gambiae*, *A. funestus* and *A. stephensi* were hybridized to the polytnene chromosomes of *A. stephensi*. All the signal locations were placed on the standard cytogenetic photomap of *A. stephensi* carefully. Meanwhile, the chromosomal rearrangements between *A. gambiae* and *A. stephensi* such as transpositions, duplications, and euchromatin/heterochromatin transitions, were also studied in this chapter.

3. Materials and methods

3.1. Mosquito strains and chromosome preparations

The Indian wild-type strain of *A. stephensi*, a standard laboratory strain, was used in this study. Ovaries from half-gravid females (25 hours after blood feeding at 26^{o}C and 83% humidity) were dissected and fixed in Carnoy's solution (Methonal: Acetic Acid Glacial = 3:1). Polytene chromosomes were prepared from fixed ovaries according to the procedure described by (*18*).

3.2. DNA clones and Probe preparations for *in situ* hybridization

3.2.1. BAC clone and *A. gambiae* cDNA

A. gambiae cDNAs of A.Gam.ad.cDNA1 and A.Gam.ad.cDNA.blood1 libraries (*24*) and *A. gambiae* BAC clones of NotreDame1 (*24*) and ND-TAM (*370*) libraries were obtained from the Malaria Research and Reference Reagent Resource Center (MR4)

(www.mr4.org). *A. funestus* BAC clones and *A. stephensi* BAC clones were acquired from Dr. Collins' lab, Dr. Tu's lab and Dr. Sehouche's lab respectively.

Recombinant cDNA and BAC clones were isolated using PhasePrep^{TM} BAC DNA Kit (Sigma). The isolated DNA was labeled with Cy5-AP3-dUTP and Cy3-AP3-dUTP(GE Healthcare UK Ltd, Buckinghamshire, England) using a modified Nick translation labeling protocol or labeled with Biotin-16-dUTP by a modified Nick Translation Mix protocol (Roache Applied science).

3.2.2. cDNA clones

A. funestus cDNAs were genomic inserts from the *A. funestus* SMART Library (*1*). The DNA was PCR amplified using Amplimer primers. The PCR conditions with the Amplimer primers were 95℃ for 5 min; 30 cycles of 94℃ for 30 s, 70℃ for 2 min; 68℃ for 3 min.

Genomic inserts from *A. stephensi* cDNA library were obtained from Dr. Shouche's lab and the cDNA clones were amplified using M13 forward (-20) and reverse primers. The PCR conditions with the M13 forward and reverse primers were 94℃ for 5 min; 45 cycles of 94℃ for 45 s, 50℃ for 45 s, 72℃ for 45 s; extra extension at 72℃ for 5 min.

The sequences of the additional 12 *A. stephensi* cDNA sequenes were obtained from the GenBank (http://ncbi.nih.gov/Genbank/). For each cDNA sample, one pair of primers was designed using Primer 3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). For PCR amplification of 12 *A. stephensi* cDNAs, template DNA was extracted from single mosquito using the Wizard SV Genomic Purification System (Promega Corporation, Madison, WI, USA). The PCR conditions were following: 95℃ for 4 min; 35 cycles of 94℃ for 45 s, 54℃ for 45 s, 72℃ for 45 s;

72℃ for 5 min. And then all the PCR products were purified from 1% agarose gel using the GENECLEAN III kit (MP Biomedicals). Finally the purified DNA was labeled with Cy5-AP3-dUTP and Cy3-AP3-dUTP (GE Healthcare UK Ltd, Buckinghamshire, England) using modified Random Primers DNA Labeling System (Invitrogen Corporation, Carlsbad, CA, USA)

3.2.3. Nick translation labeling

A reaction mixture was prepared in a PCR tube by mixing 2.5ul of 10 × reaction buffer from Fermentas, 1.25µl of 1.0mM 3dNTP Mix (without a labeled dNTP), 0.5µl of U* fluorescence, 1µl of DNase 1 from Fermentas, 1µl of Polymerase I from Fermentas, and 1µg of template DNA. Then water was added into the reaction mixture to 25µl. Finally, the sample was incubated at 15°C for 3 hours.

3.2.4. Random primer labeling

First, a reaction mixture with 1µl DNA (10ng), 10µl 2.5 × Random Primer Solution (Invitrogen Corporation, Carlsbad, CA, USA), and 2.5µl sterile water were made in a PCR tube. Then, the sample was denatured at 95°C for 5 min. After the denaturizing, the tubes were transfered on ice immediately. 10µl dNTP mixture (dNTP mixture: T=1: A=3.5; G=3.5; C=3.5 from Nick Translation Kit) and 1µl Klenow Fragment (Invitrogen Corporation, Carlsbad, CA, USA) were added into the cooled PCR tube. 0.5µl of U* fluorescence dye was added into the reaction, and mixed by inverting the tube. After adding the U* fluorescence, the samples were covered immediately to protect from light. Finally, all the samples were incubated at 37°C for 1.5 hour.

3.2.5. Biotin labeling

First, a Biotin-dNTP Mixture (10ul) was prepared by mixing 3.5µl of Biotin (Roahe Applied science), 1µl of 10mM dCTP, 1µl of 10mM dATP, 1µl of 10mM dGTP, 0.65µl of 10mM dTTP and 2.85µl of H_2O. Then a reaction mixture was prepared with 2µl of template DNA, 2.5µl of Biotin-dNTP Mix, 5µl of Nick Transmix (Roahe Applied science), and 15.5µl H_2O. After a sample was mixed well by inverting the tubes, it was incubated at 15 °C for 3h.

3.3. Precipitation of labeled probes

All labeled probes were precipitated by following steps: 2.5µl (1/10) of 3M NaAC and 62.5µl (2.5-3 volumes) of 100% Ethanol were added into the labeled probes, and then mixed by inverting the tubes, stored at –80°C or –20°C for long-term storage.

3.4. *In situ* hybridization

3.4.1. Prehybridization

If slides are more than two months old, they were first prefixed in 1:3 glacial acetic acid: methanol at RT for 10 min, and air-dried. Then, slides were dehydrated in 100% ethanol for 10 min and air dried again. Fresh slides or prehybridized slides were immersed in 1 × PBS for 20min at RT, and were fixed in 4% paraformaldehyde for 1 min. Finally, the slides were dehydrated through an ethanol series of 50%, 70%, 90%, and 100% (2 times) for 5 min each at RT, and then air-dried.

3.4.2. Preparation of 4% paraformaldehyde in PBS

2g of paraformaldehyde was dissolved in 50 ml of 1 × PBS at 65°C. The solution was cooled at RT.

3.4.3. *In situ* Hybridization with Cy3 and Cy5 labeled probes and with biotin labeled probes

First, the tubes of labeled probes were centrifuged at 14,000rcf for 10 min. After supernatant was carefully removed from tube, it was vacuumfuged for 10 min to dry pellets. 40µl Hybridization buffer was added into the tube to dissolve DNA. 20µl of red probe and 20µl of blue probe were mixed together in another clean tube. After 40µl of 20% Dextran Sulfate (Prewarmed at 39°C before use) was added into the tube, it was vortexed vigoriously, and then centrifuged at low speed briefly. 80µl of prepared solution was transferred on chromosome preparation slide (pre-warm at 90°C for 3 sec) and covered with 22×22 mm coverslip. Any large air bubble was removed by gentle pressure. After the slide was denatured on PCR machine at 90°C for 10 min, the edges of coverslipe were sealed with rubber cement. Slide was then placed in the pre-warmed humid chambers with 4 × SSC, and incubated at 39°C for interspecies or 42°C for intraspecies for 3-18h (usually overnight).

3.4.4. Washing and signal detection

3.4.4.1. Washing Cy3 and Cy5 labeled probes

After carefully removing rubber cement with forceps and coverslip, the slide was washed with 1 × SSC (interspecies) or 0.2 × SSC (intraspecies) at 39°C (interspecies) or 42°C (intraspecies) for 20 min. Then the slide was washed again at RT for 20 min. After washing, the slide was dipped in 1 × PBS, and then a mixture of 10µl 100 × diluted YoYo and 90µl 1 × PBS was placed on the slide, covered with parafilm. The sample was incubated in humid chamber at RT for 10 min, and then was dipped in 1 × PBS again. 10µl DABCO was placed on the sildes and covered it with coverslip. Finally, bubble on the slide was blotted out by gentle pressing. All the hybridized slided were kept in the slide box at 4°C.

3.4.4.2. Washing biotin labeled probes

Rubber cement was carefully removed from slides with forceps and coverslip, and then the slides were washed two times in SSC for 20 minutes.

3.4.4.3. Detection of biotin labeled probes

After the slide was placed in the humid chamber, 100µl Blocking Solution (must be without pellets) was placed on the sample. The slide was incubated at RT for 15 min in the humid chamber. Then, a working conjugate solution was prepared by mixing 10µl of streptavidin-alkaline phosphatase conjugate with 90µl of conjugate dilution buffer for each slide. After incubation, the blocking solution was removed from each slide by touching absorbent paper to the edge of the slide, and then 100µl of working conjugate solution was placed on the sample followed by a incubation in the humid chamber at RT for 15 min. The slide was washed twice in TBS for 15 min each at RT in a coplin jar, and once in ASB at RT for 10 min. Humid chamber was prewarmed at 37°C. After washing, 100µl BCIP/NBT solution was added on each slide and covered with parafilm. Slide was incubated in the BCIP/NBT solution at 37°C until the desired level of signal was achieved (from 10 min to 2 h). The color development was checked periodically by removing a slide from BCIP/NBT solution, covering the sample and residual solution with a coverslip, and observing the sample under the microscope. To stop the color development, slide was rinsed in several changes of deionized water, and air dried at RT. After one drop of deionized water was added on the slide, it was covered with 22×22mm coverslip, and then analyzed using OLYMPUS CX 41 microscope.

3.4.4.4. Signal detection and mapping

Fluorescent signals were detected and recorded using a Zeiss LSM 510 Laser Scanning Microscope (Carl Zeiss MicroImaging, Inc., Thornwood, NY, USA). Localization of a signal was accomplished using a standard cytogenetic map for *A. stephensi* (*18*).

4. Results

A total of 422 cDNA and BAC clones including of 23 (8 +15) previous published markers (*18, 371*) from *A. stephensi*, *A. gambiae*, and *A. funestus* were hybridized to 379 chromosomal sites on five chromosomal arms of *A. stephensi* by *in situ* hybridization. The exact localization of each signal was carefully determined within a subdivision and no variation in single localization was ever detected among all the nuclei examined for a given clone. The localizations of 422 clones are shown in **Appendix 2.1-2.5**. For the five chromosome arms, the markers on each chromosome arm are not distributed evenly. The summary of *in situ* hybridization results for five chromosomal arms is present in **Table 2.1**. 49 probes yield 44 signals on X chromosome of *A. stephensi* which are listed in **Appendix 2.6**; 154 markers on 2R chromosome are given in **Appendix 2.7**; the localization of the hybridization signals on 2L, 3R and 3L is presented as a list in **Appendix 2.8-2.10** respectively.

Table 2.1: Summary of *in situ* hybridization results on the five chromosome arms of *A. stephensi*

Chromosome	# of markers	# of signals	Resolution(# of signals per Mb)	# of multiple located clones	# of unique located clones
X	49	44	1.8	15	34
2R	154	117	1.9	21	132
2L	76	63	1.3	28	68
3R	74	69	1.3	20	54
3L	115	68	1.6	33	82

4.1. Interchromosomal transpositions

Given the differentiation undergone by the banding pattern and morphology of the polytene chromosomes of so distantly related species as *A. gambiae* and *A. stephensi*, the only transpositions that we can safely detect with our mapping procedure are those taking place between different chromosomal arms. Transpositions within the same chromosomal arms are probably overlooked, although they do exist (*73*). In our study, out of 422 probes, no transposition event has been identified between *A. gambiae* and *A. stephensi*. This indicates that transposition is rare in *Anopheles* mosquitoes.

4.2. Multiple located clones

Of 422 probes, 363 clones were mapped to single chromosome locations on the *A. stephensi* cytogenetic map. While 59 markers yielded multiple signals (50 of which are given in **Table 2.2** and 9 are given in **Table 2.3**). The multiple locations of these 59 clones may result from repetitive elements, inversions, duplications or transpositions. We found that eight *A. gambiae* BAC clones: (10E06, 109B13, 127F13, 141A14, 146D17, 155I2, , 25D14, 31H07) gave two distinct signals on the same chromosome arm in *A. stephensi*. However, all 8 clones have unique BLAST hits in the *A. gambiae* genome with significant e-value. These clones have a large DNA insert, usually up to 300 kb, which contains several genes. Therefore, our results suggest that these eight clones contain rearrangement breakpoints fixed between *A. gambiae* and *A. stephensi*. It is also possible that the *A. gambiae* BACs contained some sequences that in *A. stephensi* either had been transposed to a new location in the same chromosomal element or were repetitive. In addition, when we subcloned several genes from 141A14 and 146D17 and hybridized those probes to *A. stephensi* chromosomes, our data

showed that clone 141A14 spans the proximal breakpoint of inversion “O” on 2R chromosome in *A. gambiae*. This breakpoint is located between cDNAs: AGAP002934 and AGAO002935 (which is discussed in chapter 3). Our results also indicate that another BAC clone, 146D17, contains one inversion breakpoint fixed on the 2L chromosome in *A. gambiae.* These results provide a basis for our interpretation of multiple signals as the result of the presence of fixed breakpoints in these genomic clones.

Due to the importance of duplication in chromosomal evolution, in this study, we also analyzed the duplication events that occurred in *A. gambiae* and *A. stephensi*. 9 of 59 multiple located markers listed from 9-17 in **Table 2.2** were mapped to the multiple locations on *A. stephensi* cytogenetic map. When we used the blast tool from vectorbase (http://www.vectorbase.org/Tools/BLAST/) to analyze the sequences of these probes, BLASTN results against the *A. gambiae* genome suggest that all those clones yield multiple hits in the *A. gambiae* genome with an e-value greater than 1e-8. This result may indicate that in *A. stephensi* and *A. gambiae*, the mulitple signals of these clones are most likely due to the presence of other copies of the same gene family, pseudogenes, or DNA fragments sharing a certain sequence similarity with the cloned gene. In addition to these markers, 40 clones were localized to multiple chromosome sites in *A. stephensi* while they gave unique blast hits in *A. gambiae*. It is possible that those 40 probes contain some sequences that in *A. stephensi* either resulted from gene duplication or segmental duplication. The above data demonstrated that gene duplication is very common in the *Anopheles* genomes. The high frequency of duplication events suggests their important role in genome evolution.

4.3. *A. stephensi* species-specific markers

When the partial sequences of 24 *A. stephensi* cDNAs were used to BLAST against the database of all the available organisms (http://blast.ncbi.nlm.nih.gov/), no hit can be found for all of them. The blast results may suggest that these 24 *A. stephensi* cDNAs are *A. stephensi* species specific clones. The localizations of 24 clones on *A. stephensi* cytogenetic map were listed in **Table 2.3**. *In situ* hybridization results of these 24 cDNAs on *A. stephensi* chromosomes showed that 8 clones (33%) were hybridized on the X chromosome and 9 clones (38%) had multiple signals on chromosomal arms of *A. stephensi*. Due to the limited information of those gene annotations, the functions of these 24 *A. stephensi* species-specific clones are still not clear.

Table 2.2: The summary of multiple located clones in *A. stephensi*

	Clone	Accession #	*A. stephensi*	*A. gambiae*	e-value	Clone type
1	10E06	AL145314	2R:11A(2)	2R:13D		*gam* BAC
			2R:10D(2)		0	
2	109B13	BH385043	3L:42B(3)	2L:25D		*gam* BAC
			3L: 44B(1)		0	
3	127F13	BH387145	2R:11D(3)	2R:13B		*gam* BAC
			2R:15A(2)		0	
4	141A14	BH367876	2R:11A(3)	2R:13D		*gam* BAC
			2R:15B(2)		0	
5	146D17	BH400736	3L:40A(3)	2L:23A		*gam* BAC
			3L:44C(1)		1e-144	
6	155I2	BH374558	2L:24A(1)	3L:43A		*gam* BAC
			2L:24B(3)		0	
7	25D14	AL610688	X:3A(1)	X:4C		*gam* BAC
			X:4C		0	
8	31H07	AL156465	2L:26A	3L:41A		*gam* BAC
			2L:21A(3)		0	
9	04B15	AL145409	X:4B(3)	X:5D	0	*gam* BAC
			2R: 11D(2)	3R:33C	0	
			2R: 15A(1)	2L:21D	0	
				multiple hit on others		
10	139K20	BH402428	2L:20A_B(het)(1)	3L:38C	0	*gam* BAC
			3R: 37B(5)	3R:37D	3e-89	
				multiple hit on others		
11	20K19	AL609825	X: 2C(1)	X:2B	0	*gam* BAC
			X: 2C(3)	X:2C	1e-47	
				multiple hit on others		
12	25G06	AL610709	X: 3A(3)	X:3B	0	*gam* BAC
			X: 6B (het)(5)	3R:37C	1e-104	
			3L: 44A (het)(5)	X:5A	1e-99	

				multiple hit on others		
13	AST028O6	Tu lab	2R: 16C(3)	2R:18D*	NA	*st* BAC
			3L: 46D(1)	2L:27A	NA	
14	671280	BM657097	2R: 11C(3)	2R:7B	0	*gam* cDNA
		BM657097	3L: 46C(4)	2L:27D	0	
15	729481	BM577567	2L: 22B(1)	3L:38C	0	
			2L: 25C	3R:31A	1e-146	
				2L:22A	3e-52	*gam* cDNA
16	11_C03	BU038902	2R: 13B(4)	2R:9C	4e-73	*fun* cDNA
			3L: 42A(2)	UNKN	1e-08	
17	212F07	Sehouche lab	2L: 21A(1)	3R:35C	7e-83	*st* cDNA
			3R: 35B(4)	3R:30B	1.9e-48	
18	627837	BM647307	2R: 7B(1)	2R:7B	0	
			2L: 20A_B(het)(2)			*gam* cDNA
19	669234	BM655755	2L: 21B_20A(1)	3L:44B	0	
			3R: 37D(2)			*gam* cDNA
20	702140	BM594831	2L: 21B_20A(2)	3L:44B	0	*gam* cDNA
	702140	BM594831	3L: 45A(1)			*gam* cDNA
21	211E11	Sehouche lab	3R: 29B(1)	3R:32C	1e-19	*st* cDNA
			3R:34B(2)			
22	05_D12	BU038881	2L: 26B(1)	3L:45C	0	
			2L: 26C(1)			*fun* cDNA
23	06_E11	BU038887	X: 3A(1)	X:4C	1e-92	
			2R: 14B(1)			*fun* cDNA
24	105F8	BH392724	2L: 25B(7)	3L:40B	0	
			3R: 37B(3)			*gam* BAC
25	106D14	BH399330	2R: 18D(1)	2R:17C	0	
			2L: 20C(2)			*gam* BAC
26	145G13	BH370252	2L: 24A(2)	3L:43A	0	
			3R: 37B(4)			*gam* BAC
27	16_G10	BU038933	2R: 7A(5)	2L:28C	1e-72	
			3L: 41B(1)			*fun* cDNA
			3L: 46B(2)			
28	180K21	BH367855	2L: 25B(3)	3L:39C		
			3R: 36C(1)		3e-88	
			2L: 23B(1)			*gam* BAC
29	211G05	Sehouche lab	2L: 25B(6)	3L:40B	1.2e-13	*st* cDNA
			2L: 28A(2)			
30	211G09	Sehouche lab	2L: 20C(4)	3L:38B	9.7e-43	*st* cDNA
			3L: 43A_B			
31	21G01	AL150712	X: 3A(4)	X:3B	0	
			X: 5C			*gam* BAC
32	23_D08	BU038965	2L: 25A(1)	2L:22B	5e-20	
			3L: 41B(2)			*fun* cDNA
33	23I16	AL610348	X: 6B (het)(6)	X:2B	0	
			2R: 11D(2)			
			2R: 12B(1)			
			2L: 20A_B(het)(2)			
			2L: 27C(2)			*gam* BAC
34	24J01	AL152501	2R: 11D(1)	X:4A		
			X: 4A(2)		1e-151	
			3R: 37B(5)			*gam* BAC
35	30P20	AL156193	X: 6A(het)(4)	3R:35C	0	
			3R: 37B(3)			*gam* BAC
36	31B09	AL156246	2L: 25B(7)	3R:35C	0	
			3R: 30B(2)			
			3R: 35A			*gam* BAC
37	178A3	BH398965	X: 6A(het)(4)	3R:30A		
			2L: 20A_B(het)(2)		0	*gam* BAC

			3R: 29E			
38	AsK3MO	AY065662	2R: 15A(2) 2R: 17C(2) 3R: 35B(3) 3R: 37C(1) 3L: 44B(2)	2L:23C	0	*st* cDNA
39	Ag1972	AGAP001972	2R: 10A(10) 2R: 13B(1) 2R: 13C(1) 2R:18A(1)	2R:9C	0	*gam* cDNA
40	Ag1985	AGAP001985	2R: 10A(7) 2R: 14A(1)	2R:10A	0	*gam* cDNA
41	Ag3325	AGAP003325	2R: 17B(2) 3L: 40D(2)	2R:15A	0	*gam* cDNA
42	Ag3327	AGAP003327	2R: 16A(1) 2R: 17B(1)	2R:15A	0	*gam* cDNA
43	Ag5789	AGAP005789	2R: 13B(2) 3L: 45C(1)	2L:23A	0	*gam* cDNA
44	Ag6898	AGAP006898	3R: 37B(5) 3L: 43C(4)	2L:26B	0	*gam* cDNA
45	Ag6968	AGAP006968	3L: 43A(3) 3L: 44A(2)	2L:26C	0	*gam* cDNA
46	Ag7006	AGAP007006	3L: 41C 3L: 45C(1) 3L: 45C(6)	2L:26C	0	*gam* cDNA
47	Ag7051	AGAP007051	2R: 11D(1) 2L: 20A_B(het)(2) 2L: 27C(3) 3R: 37B(5) 3L: 44C(5)	2L:26D	0	*gam* cDNA
48	Ag7075	AGAP007075	X: 6A(het)(3) 2L: 20A_B(het)(2) 2L: 27C(3) 3L: 40B(1)	2L:26D	0	*gam* cDNA
49	Ag7086	AGAP007086	2R: 15A(3) 3L: 40A(1)	2L:26D	0	*gam* cDNA
50	AST041P6	Tu lab	2L: 20A_B(het)(3) 2R: 18A(1)	no good hit		*st* BAC

gam: *A. gambiae*; *st*: *A. stephensi*; *fun*: *A. funestus*

Table 2.3: *A. stephensi* species specific clones

	Clone	**Source**	**Arm**	**Division and subdivision**	**Clone type**	
1	211E12'	Sehouche lab	3R	31B(3)	*stephensi* cDNA	*st* specific
			3R	34C		
2	212E05'	Sehouche lab	X	1B_C	*stephensi* cDNA	*st* specific
			2L	22A(1)		
3	211B09'	Sehouche lab	2R	12B(3)	*stephensi* cDNA	*st* specific
			3L	38F(1)		
4	211D07'	Sehouche lab	2L	26C(2)	*stephensi* cDNA	*st* specific
			3L	39A(1)		
5	211F08'	Sehouche lab	2L	25B(6)	*stephensi* cDNA	*st* specific
			3L	38B(1)		
6	212C07'	Sehouche lab	X	1C(1)	*stephensi* cDNA	*st* specific
			3L	46B(3)		
7	212D07'	Sehouche lab	X	1C(2)	*stephensi* cDNA	*st* specific

			2L	21A(1)		
			3L	46B(3)		
			3L	39B(2)		
8	211E02'	Sehouche lab	2L	23B(2)	*stephensi* cDNA	*st* specific
			3L	39B(2)		
			3L	43B		
9	211D03'	Sehouche lab	X	1C(2)	*stephensi* cDNA	No good hit in *A. gambiae*
			2L	23B(1)		
10	211A12	Sehouche lab	X	6A	*stephensi* cDNA	*st* specific
11	211B07	Sehouche lab	X	4B	*stephensi* cDNA	*st* specific
12	211C05	Sehouche lab	2R	11B	*stephensi* cDNA	*st* specific
13	211E04	Sehouche lab	2L	23C	*stephensi* cDNA	*st* specific
14	211F03	Sehouche lab	3L	46B	*stephensi* cDNA	*st* specific
15	211F07	Sehouche lab	X	2A	*stephensi* cDNA	*st* specific
16	212A02	Sehouche lab	3R	29A	*stephensi* cDNA	*st* specific
17	212D05	Sehouche lab	2R	11C	*stephensi* cDNA	*st* specific
18	212D06	Sehouche lab	3R	29C	*stephensi* cDNA	*st* specific
19	212D10	Sehouche lab	3L	40C	*stephensi* cDNA	*st* specific
20	212E01	Sehouche lab	2L	22A	*stephensi* cDNA	*st* specific
21	212E10	Sehouche lab	2L	25A	*stephensi* cDNA	*st* specific
22	212F09	Sehouche lab	2R	10D	*stephensi* cDNA	*st* specific
23	212H03	Sehouche lab	2R	9C	*stephensi* cDNA	*st* specific
24	211E05	Sehouche lab	X	6A	*stephensi* cDNA	*st* specific

st: *A. stephensi*; ' indicates multiple signals for the clone.

4.4. Clones located in heterochromatic regions and euchromatin-heterochromatin transitions between species

So far, we have identified two *A. stephensi* species-specific cDNAs, 211A12 and 211E05, and one *A. stephensi* BAC clone, AST041P6, that were localized to heterochromatin regions in *A. stephensi*. Both of 211A12 and 211E05 were from the *A. stephensi* EST library and AST041P6 has no good hit in *A. gambiae* genome. Our results suggest that the above three clones are *A. stephensi* species-specific heterochromatin markers (**Table 2.4**).

In malaria mosquitoes, a euchromatin region of one species can correspond to a heterochromatin region in another species. In our analysis, three clones (AST018J14, 12G16 and 211A02) (**Table 2.4**) were localized to euchromatin regions in *A. gambiae* but to heterochromatin regions in *A. stephensi*. Of them, one *A. gambiae* BAC clone,

12G16, and one *A. stephensi* cDNA, 211A02, were localized to unique locations of the euchromatic region 2L: 23C in *A. gambiae,* and the distance between these two clones was 796,743 bp. These two clones were hybridized to heterochromatic regions: 44A of 3L chromosomal arm in *A. stephensi* (**Figure 2.1**).

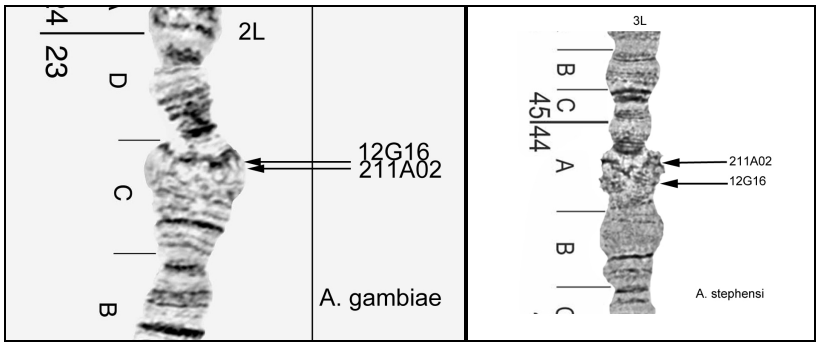


Figure 2.1: the localizations of 12G12 and 211A02 in *A. gambiae* and *A. stephensi*

One BAC clone from *A. gambiae*, 139K20, mapped to a heterochromatic region (3L:38C) in *A. gambiae* was localized to a heterochromatic region 20AB of 2L chromosomal arm in *A. stephensi,* and it also produced one minor signal on euchromatic region 37B of 3R (**Table 2.4**). Our data suggest that 139K20 is a heterochromatin marker shared by both species.

Table 2.4: Clones located in heterochromatic regions

Clone		***A. gambiae***	***A. stephensi***	**Type**
211A12	Species-specific	No hit	X:6A(het)	*stephensi* cDNA
211E02	Species-specific	No hit	X:6A(het)	*stephensi* cDNA
AST041P6	Species-specific	No hit	3R:37D(het),2L:20A_B(het)	*stephensi* BAC
AST018J14		3L:38C(euch),X:5 D(euch)	2R:centromere(het)	*stephensi* BAC
12G16		2L:23C(euch)	3L:44A(het)	*gambiae* BAC
211A02		2L:23C(euch)	3L:44A(het)	*stephensi* cDNA
139K20		3L:38C(het)	2L:20A_B(het)*,3R:37B	*gambiae* BAC

5. Discussion

5.1. A physical map for *A. stephensi*

The karyotype of *A. stephensi* consists of 3 pairs of chromosomes, one pair of sex chromosome (X and Y in male, XX in female) and two pairs of autosomes: 2 and 3. Each autosome was divided by centromere into left and right arms. High-quality polytene chromosomes (X, 2R, 2L, 3R, and 3L) are present in salivary glands and ovarian nurse cells of *A. stephensi*. Polytene chromosomes from ovarian nurse cells in *A. stephensi* are more favorable for cytogenetic studies than those from salivary glands because of easier preparation, as well as constant and clear banding patterns (*16*). A high-resolution cytogenetic map of *A. stephensi* was developed in our lab (*18*). By using this standard cytogenetic map, a total of 422 probes have been mapped to the polytene chromosome of *A. stephensi*. 363 clones yielded unique signals, and 59 markers were localized to multiple locations. If the size of the genome of *A. stephensi* was assumed to be the same as that of *A. gambiae* (230.5 Mb, as evident from the *A. gambiae* mapped assembly), the resolution of our current *A. stephensi* map is about 608 Kb on the average, which make the density of the *A. stephensi* physical map only second to the *A. gambiae* physical map among malaria mosquitoes.

5.2. Interchromosomal arm transpositions

In species with polytene chromosomes (like those in the genera *Anopheles* and *Drosophila*), *in situ* hybridization has facilitated the comparative mapping of multiple markers. This approach can detect chromosome change due to macro-rearrangement such as paracentric inversions, translocation or transposition between elements. In the genera *Drosophila* and *Anopheles*, paracentric inversions are considered the chief mechanism of chromosome rearrangement (*1, 30, 256*). Although translocations are

scarce in the genus *Drosophila* (*372*), whole arm translocation has been found in *A. gambiae* and *A. funestus* (*1*). Gene transposition is the movement of a gene or small group of genes from one genomic location to another(*32*). Detection of gene transposition can be achieved by comparing the chromosomal location of genes between different species. In the present study, no interchromosomal transposition event has been identified between *A. gambiae* and *A. stephensi* using *in situ* hybridization results. This agrees well with the case in *Drosophila*. Transposition is very rare in *Drosophila* (*73*), and no cases of gene transposition were detected from 26 P1 phage (*373*) and 154 clones (*30*). In other work, two possible transposition events out of a total of 328 clones hybridized to the *D. repleta* chromosomes(*33*). The rate of gene transposition in the *Drosophila* genus is 4.9×10^{-5} transpositions/gene/myr (million years) by combining the results of previously obtained (*33*). This rate, however, doesn't include tandemly repeated genes such as histone or rRNA genes, which are often show transposition (*374, 375*). It also does not include intra-chromosomal transpositions.

5.3. Gene duplications in *Anopheles* mosquitoes

Of 59 multiple located clones, the locations of two BAC markers, 141A14 and 146D17 have been confirmed to be within the breakpoints of fixed inversions in the *A. gambiae* complex. Eight clones were located in the euchromatin and also yielded signals on the heterochromatin. Since repepititive elements were found to be concentrated on the heterochromatic regions (*207*), we hypothesized that these eight probes contained copies of repetitive elements. The other 49 clones might result from gene duplication events. In eukaryotes, the estimation of the rate of gene duplication is about 1 gene per 100 MY (*376*). In Drosophila melanogaster, 41% of the total number of

genes was duplicated (*377*). Our evidence suggested that 49 multiple located markers were caused by gene duplication in *A. gambiae* and *A. stephensi*. These duplications might provide new genetic material for mutation in *Anopheles* mosquitoes.

5.4. Clones located in heterochromatic regions and euchromatin-heterochromatin transitions between species

A total of seven cases of euchromatin – heterochromatin transition events have been identified between *A. gambiae* and *A. stephensi*. A very recent study has shown that genomic DNA of heterochromatin is extremely differentiated among populations of *D. melanogaster* (*378*). Analysis of insertion patterns of three transposable elements (TEs) has determined 20 M-form- or S-form specific insertion sites in *A. gambiae*, of which seven were found to be integrated within repeated sequences, and three were located in the heterochromatin (*379*). A genome-wide microsatellite study of members of the *A. gambiae* complex has determined a high level of genetic introgression among species. However, the *A. gambiae* microsatellites at six heterochromatic loci of X, 3L, and 3R could not be amplified in all sibling species, indicating significant sequence divergence from the major malaria vector (*380*). Heterochromatin constitutes about one-third of the *Anopheles* genome (*25, 381*). Therefore, genome projects should give attention to heterochromatin in order to characterize its structural and functional organization. A pioneering effort in *D. melanogaster* heterochromatin provided a detailed computational and manual annotation of 24 megabases of heterochromatic sequence (*382*).

6. Conclusions

1) A of 608-Kb-resolution physical map has been developed for the Asian human malaria vector, *A. stephensi*. This resolution is only second to the resolution of the *A. gambiae* map.
2) Interchromosomal transpositions are rare between *A. gambiae* and *A. stephensi.*
3) Most multiple located markers result from gene duplication events in the *Anopheles* genome.
4) Cases of euchromatin and heterochromatin transition between *A. gambiae* and *A. stephensi* suggest that heterochromatic sequences evolve rapidly.

Chapter three: Comparative analysis of inversion fixations in chromosomal arms of malaria mosquitoes

1. Abstract

We used 231 markers physically mapped to the polytene chromosomes of *A. stephensi* and 127 markers previously mapped to the *A. funestus* chromosomes for comparative analysis of chromosomal rearrangements. The comparative mappings between *A. gambiae* and *A. funestus*, *A. gambiae* and *A. stephensi* revealed that the sex (X) chromosome has the highest rate of chromosomal evolution despite the paucity of polymorphic inversions on X. The analysis of molecular features on five chromosomal arms suggested that the accumulation of transposable elements, microsatellites, minisatellites and satellites as well as inverted repeats on the sex chromosome, contributed to the rapid generation of inversion fixation breakpoints on X chromosome. Among the autosomes, 2R harbored smaller synteny blocks and accumulated more disrupted blocks per unit length than other chromosomal arms indicating that 2R evolved faster than other autosomes. This observation is consistent with the highest number of inversion polymorphisms on the 2R chromosome in *Anopheles* mosquitoes. The positive association between evolutionary rate and the level of chromosomal polymorphism on autosomes suggests that local adaptation can drive the polymorphic inversions into fixation. To understand the origin of fixed inversions in *Anopheles* mosquitoes, the molecular features in synteny blocks and breakpoint regions were analyzed. We demonstrated that transposable elements, AT repeats, and inverted repeats play important roles in the origin of inversions in malaria mosquitoes. In addition, the analysis of M/SARs revealed that nuclear architecture might play a role in

determining chromosome specificity of rearrangement rates. We also provided evidence that the breakpoints of fixed inversions are distributed nonrandomly and breakpoint clustering is common in *A. gambiae* and *A. stephensi*.

2. Introduction

In *Anopheles* mosquitoes, the most popular type of chromosomal rearrangements is paracentric inversions (*1, 13*). Although enormous comparative analysis revealed that the patterns and rates of rearrangements are lineage-specific, and chromosomal elements remarkably differ in the rates of evolution (*118, 136, 139, 152*). The comparative mapping has only been performed between *A. gambiae* and *A. funestus* in the genus of *Anopheles* (*1*). The patterns and rates of inversion fixation in *Anopheles* mosquitoes are not clear due to the low resolution of a physical map of *A. funestus*. Additionally, the studies of fixed inversions in *A. gambiae* complex suggested that the sex chromosome harbored most of inversions. While among autosomes, 2R chromosome has more inversions than other autosomes (*13*). If the pattern of inversion fixations identified in *A. gambiae* complex is general, the fastest rate of chromosomal evolution should be discovered on the X chromosome among three species. For this purpose, we studied the correspondence of chromosomal elements between three malaria vectors; *A. gambiae*, *A. funestus*, and *A. stephensi*, different members of the same subgenus *Cellia*: *Pyretophorus* (*A. gambiae*), *Myzomyia* (*A. funestus*), and *Neocellia* (*A. stephensi*) (*14, 16*). These distantly related lineages diverged from a common ancestor at least 30 million years ago, and are good model systems for studying chromosomal rearrangements (*217*). The *A. stephensi* physical genome map has been developed and compared with the existing genome maps of *A. funestus* and *A.*

gambiae. Three-way analysis of rearrangements has allowed us to assign rearrangement events to one of the three lineages. A computer algorithm, GRIMM, and the N-T model were used to infer the number of rearrangements fixed between the species. Rates of chromosomal evolution were calculated subsequently.

A fundamental question in the study of chromosome evolution is the distribution of inversion breakpoints. There exist two most polular models, the N-T model and fragile breakage model. Random breakage model suggested that the inversion breakpoints are distributed randomly along chromosomes (*43*), and was supported by enormous comparative mapping data (*29, 90, 242-244*). Fragile breakage model proposed that chromosomal breakages tend to reoccur at 'fragile sites' or 'hotspots' (*91*). Several subsequently studies were in favor of this model (*57, 59, 90, 249*). However, the analysis of the breakpoints of paracentric inversions between *Anopheles* lineages is limited due to the unavailable genome sequences, physical and linkage maps for most of species. For the first time, the lengths of synteny blocks between *A. gambiae* and *A. stephensi* were fitted into the Nadeau & Taylor model to analyze the distribution of inversion fixations in *Anopheles* mosquitoes.

Another question in the study of chromosomal evolution is the molecular mechanisms for generating chromosomal inversions. The prevailing view is that chromosomal rearrangements are generated by ectopic recombination events between inverted repetitive sequences such as transposable elements in *Drosophila* (*158*) and *Anopheles* mosquitoes(*164*), as well as segmental duplications in mammals (*177*). Additionally, other repetitive elements such as microsatellites, minisatellites, satellites, inverted, and AT repeats have beem implicated in the origin of chromosomal rearrangements (*160-*

162). However, what is the major evolutional force responsible for the fast chromosomal evolution in *Anopheles* mosquitoes is uncertain. Therefore, the molecular features associated with fast and slow chromosomal evolution, and with synteny blocks and breakpoint regions were analyzed to illustrate the molecular mechanism question of inversion fixation in *Anopheles* mosquitoes.

In the studies of polymorphic inversions within species and fixed differences between species (*208*), most of data suggested that breakpoints of inversions are distributed nonrandomly (*13, 24, 33, 134, 135, 147, 148, 209*). In *A. gambiae*, *A. funestus,* and *A. stephensi*, no polymorphic inversion has been identified on X chromosome, and 2R exhibits the majority of inversion polymorphisms (*13, 17, 26*). The local adaptation model proposes that an inversion will spread if it carries a set of locally adapted alleles. Some alleles will cause it to spread to fixation, while others will lead either to a neutral or a selectively maintained polymorphism (*208*). If this model is true, the positive association between the content of polymorphic and fixed inversions should be expected. In this chapter, we also infer the evolutionary forces responsible for the establishing of fixed inversions in *Anopheles* mosquitoes.

3. Material and methods

3.1. A physical map for *A. stephensi*

A total of 231 markers were hybridized to the polytene chromosomes of *A. stephensi* (**Appendix 3.1**). The *A. gambiae* BAC clones of NotreDame1 (*24*) and ND-TAM (*370*) libraries were obtained from the Malaria Research and Reference Reagent Resource Center (MR4) (www.mr4.org). *A. funestus* cDNAs were derived from the *A. funestus* SMART Library (*1*). *A. funestus* BAC, *A. stephensi* BAC and cDNA were kindly provided

by Dr. Collins, Dr. Shouche and Dr. Tu, respectively. The probe preparation and *in situ* hybridization of these clones were described in detail in the material and methods of chapter one.

3.2. The distributions of markers on five chromosomal arms

The locations of 231 markers on *A. stephensi* chromosomes and previously published 145 clones (*24, 164*) of *A. funestus* were compared with the coordinates of homologous sequences in the *A. gambiae* genome (**Appendix 3.1**)**.** Genomic locations of the homologous sequences in the *A. gambiae* genome were determined by BLASTN and TBLASTX with default parameters using the VectorBase website (http://www.vectorbase.org/Tools/BLAST/#). Chromosomal locations of the *A. gambiae* BAC clones were found using the search option on the VectorBase website (http://www.vectorbase.org/Search/Keyword/). The distribution of 231 markers on *A. stephensi* chromosomes and *A. gambiae* (**Appendix 3.2**) as well as 145 probes on *A. funestus* (**Appendix 3.3**) were analyzed by statistically methods. In order to determine if the markers along each chromosome arm are distributed uniformly along the entire length of each chromosome, χ^2 test was used, which is computed as $X^2 = \frac{(\sum_{i=1}^{N}(Oi-Ei)^2)}{Ei}$ where N denotes a number of equally spaced bins. The null hypothesis was that markers are distributed uniformly on chromosomes. Ei is the expected number of observations and Oi is the observed number. Some of the chromosomes exhibit low marker counts (especially the X chromosome), hence simulated p-values, and based on bootstrap (100,000) replications, were also provided. For our analysis, degrees of freedom (N -1) were determined by placing uniformly distributed bins such that the expected counts are greater than 5.

3.3. Nadeau-Taylor model (N-T model)

The random breakage model assumes the random (i.e., uniform and independent) distribution of chromosomal inversion breakpoints. Under this model, the observation that the lengths of synteny blocks shared between *A. gambiae* and *A. stephensi* (**Appendix 3.7**) should be well fitted by the expected synteny block lengths. The theoretical distribution of lengths of synteny blocks is an exponential distribution with density function

f(x) = $1/L.\, e^{-x/L}$ where L is the average length of all synteny blocks. Nadeau-Taylor (N-T) model can estimate L from the observation of already discovered synteny blocks.

3.4. The calculation of the rates of chromosomal evolution

The calculation of inversion distances among species of the *A. gambiae*, *A. funestus* and *A. stephensi* were performed using the Genome Rearrangements in Man and Mouse (GRIMM) program which is available at http://grimm.ucsd.edu/GRIMM/ (*383*). This algorithm also finds optimal scenarios for the transformation of one genome into another. The number of inversion fixations between *A. gambiae* and *A. stephensi* were also analyzed by the N-T model (*43*). This model estimates the average lengths of synteny blocks for each chromosomal arm and the number of inversion fixations was calculated by the average length of synteny blocks divided by the total chromosomal length.

3.5. Analysis of the rates of syntenic block disruptions

Before endeavoring into the mathematical process describing how these blocks are conserved between the three species, we first considered the evolutionary insights that best describe this process. At some point in time of the evolutionary history, these three

separated species were descendants of some common species, and so each of the three groups had one full block (representing its entire genome) conserved simultaneously with respect to the other two species. We denoted this time by t0 = 0. After these three species speciated, each of their respective genomes evolved at different rates, which resulted in sets of disruptions between the species groups. For discrete time points (t0 < t1 < ··· < tc) (tc is the current state of time), these disruptions appeared, resulting in a lower level of conserved blocks on each chromosomal arm. This is illustrated in **Figure 3.1**.

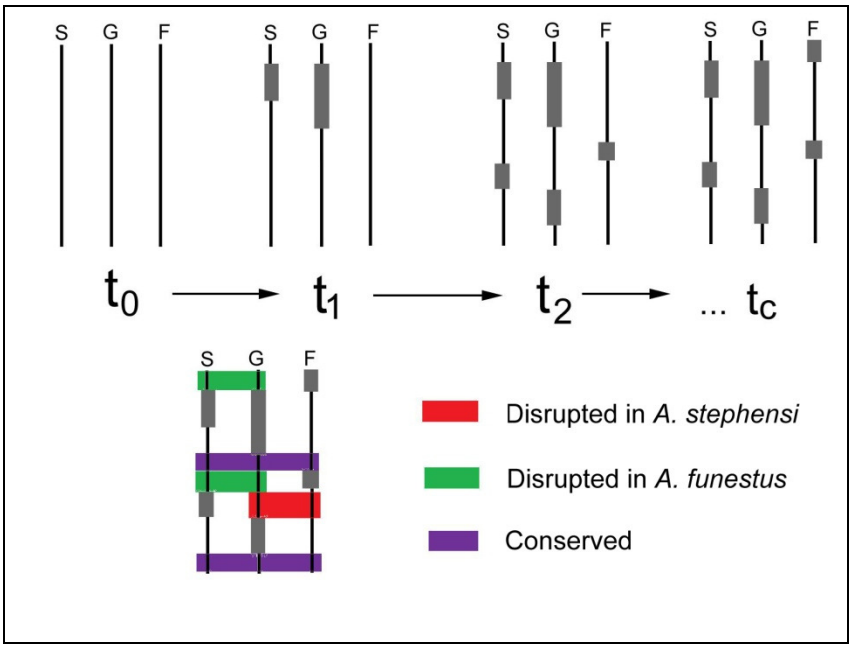


Figure 3.1: Schematic illustration of disrupted and conserved blocks.

From **Figure 3.1**, we observed that through time, the length of the conserved blocks decreased; however, the number of them may be increasing. While we do not observe this progression through time, we do observe the level of conservation at time tc. If this process were allowed to continue for an infinite amount of time, we would imagine that all blocks would eventually be disrupted. We should note that our schematic only

illustrates the basic idea, since recombination and other factors are able to alter the location of the conserved blocks.

In this analysis, we were only conserned with how this process changes for each of the chromosome arms: (*18 3L*). Since the disrupted blocks are accumulating through time, we can envision through a backwards time perspective, the conserved blocks are accumulating through reversed-time. Using this framework, we accounted for both the number of conserved blocks and the length of each of these. We modeled this using a compound Poisson process, where the number of conserved blocks follows a Poisson process, and the length of the chromosomal arm scales the rate of the process; a separate process governs the length of each conserved block. Formally, for each arm j ∈ {2R, 2L, 3R, 3L}, with total chromosome length Lj , we model this process by Rj = $\sum_{i=1}^{N(Lj)} \mathrm{bi,j}$ where N(Lj) = Nj follows a Poisson process and each conserved block (bi,j) follows some i.i.d. distribution.

Parameter interpretation: $(r_j{}^{-1})^{(diff)}$ indicates the rate of accumulation of disruption, as compared to the conserved blocks. On the other hand, highly negative values associate with the inference that the disrupted rates are less than the conserved, and the positive direction suggests higher rates for the disrupted blocks.

3.6. Molecular features in synteny blocks and breakpoint regions

In this analysis, we determined different groups: synteny blocks and breakpoint regions impact on each molecular feature. We denoted that each molecular feature counts, in segment i, as C_i. Since each of the regions (either a break or block group) has different lengths, we expected that, on average, larger segments contain more of the feature. For region i, we denoted the length as Li. Our analysis was based on a

Poisson regression model, which is generally appropriate for modeling discrete count type data. Under such a model, the probability of observing the feature count Ci,j , for the i^{th} region on chromosome j ∈ {X, 2R, 2L, 3R, 3L}, is Prob(Ci,j) = $e^{-\lambda i,j} \frac{\lambda_{i,j}^{Ci,j}}{Ci,j!}$ Where, λi,j is the mean count for observation i, on chromosome j. This mean form is generalizable to account for different sources of variability found in the data; and in our case, we must account for the variability specific to each chromosome arm j ∈ {X, 2R, 2L, 3R, 3L}, the length of each region (Li,j), and the group (G = {break, block}) effect. We used the canonical log-link for representing this mean (this is the standard canonical approach), which we write as log(λi,j) = α1δG + ξj + log(Li,j); where ξj relates to the chromosome specific level of the feature, and

δG = $\begin{cases} 1 \; break \; group \\ 0 \; block \; group \end{cases}$

Parameter interpretation: Since log (λi, j) models the logged expectation of counts for each molecular feature type, we can interpret the estimated parameters by noting the relationship: log(λi,j=Li,j) = α1δG + ξj: From this it is clear that ξj models the average density of molecular features on chromosome j, and α11 is the increase attributed to breaks over block groups. Also, we note that $\frac{\exp(\alpha 1 + \xi j + \log(Li,j))}{\exp(\xi j + \log(Li,j))}$ = exp(α1) is the mean, region length controlled, increase of counts in break groups (over block groups). From the posterior distribution for the unknown parameters, we are able to assess the probability that break groups are different from block groups.

4. Results

4.1. The distributions of markers on five chromosomal arms

A total of 231 unique and multiple located markers have been used for comparative mapping of *A. gambiae* and *A. stephensi*. The localizations of the markers in *A. stephensi* have been determined by *in situ* hybridization to the chromosomes of *A. stephensi* and the coordinates in *A. gambiae* were obtained from the VectorBase website (http://www.vectorbase.org) (**Appendix 3.1**). If the size of the mapped *A. gambiae* genome assembly is 230.5 Mb, the resolution of our current comparative map between *A. gambiae* and *A. stephensi* is about 1 Mb. We first examined the distribution of markers along the five chromosomal arms of *A. gambiae* and *A. stephensi*. The data are present in **Appendix 3.2**. Our null hypothesis is that the distribution of markers is uniform. The analysis of distributional fit is often based on the p-value, where the hypothesis is rejected when the p-value is under some predetermined threshold (0.05 for this study). Our analysis reveals that the markers on X in *A. gambiae*, on 2R, 2L, 3R and 3L in *A.gambiae* and in *A. stephensi* are distributed uniformly. However, the markers on the X chromosome of *A. stephensi* (0.0339) show some deviation from uniformity (**Table 3.1**). This deviation of the X chromosome of *A. stephensi* could result from the large heterochromatin region near the centromere where no marker has been placed. Therefore, the total chromosome length of *A. stephensi* has been modified and the new result (0.6596) shows a uniform distribution of the marker on euchromatin of *A. stephensi* X chromosome.

A total of 127 probes previously mapped to the polytene chromosomes of *A. funestus* (*24*) (listed in **Appendix 3.1**) were used to estimate the number of fixed inversions between *A. gambiae* and *A. funestus*. The coordinates of markers in *A. gambiae* and the locations in *A. funestus* are present as a list in **Appendix 3.3**. The data analysis

suggests that all the markers on five chromosomal arms of *A. funestus* are distributed uniformly (**Table 3.1**).

Table 3.1: Uniform distribution of markers in *A. gambiae*, *A. stephensi* and *A. funestus*

Chromosome arm	***A.gambiae*** **(p-value)**	***A.stephensi*** **(p-value)**	***A. funestus*** **(p-value)**
X	0.1468	0.0339/0.6596	0.8750
2R	0.7141	0.7798	0.9294
2L	0.9333	0.7645	0.2480
3R	0.8258	0.2531	0.3751
3L	0.7798	0.7922	0.4753

4.2. Length distributions of conserved synteny blocks in *A. gambiae* and *A. stephensi*

The gene orders and conserved synteny blocks between *A. gambiae* and *A. stephensi* are shown in **Figures 3.2-3.3** and **Appendix 3.4-3.6**. Synteny blocks are defined here as a set of markers that are consecutive (show the same relative distance from each other and same order) and share similar banding pattern along polytene chromosome. A total of 55 synteny blocks have been identified between *A. gambiae* and *A. stephensi* at a 1 Mb resolution. However, the size of the conserved segments varied significantly among five different chromosomal arms. For the X chromosome, gene order reshuffled dramatically along the whole arm and only four small synteny blocks (<0.5 Mb) can be identified. For the autosomes, the sizes of conserved segments also vary dramatically as the following: 2R<2L<3R<3L. A total of 19 synteny blocks can be detected on 2R and the largest one is less than 2 Mb while the largest conserved blocks (up to 8 Mb) can be found on 3R and 3L of *A. gambaie.* Therefore, our results suggest that the gene orders on the sex chromosome are less conserved than those on autosomes.

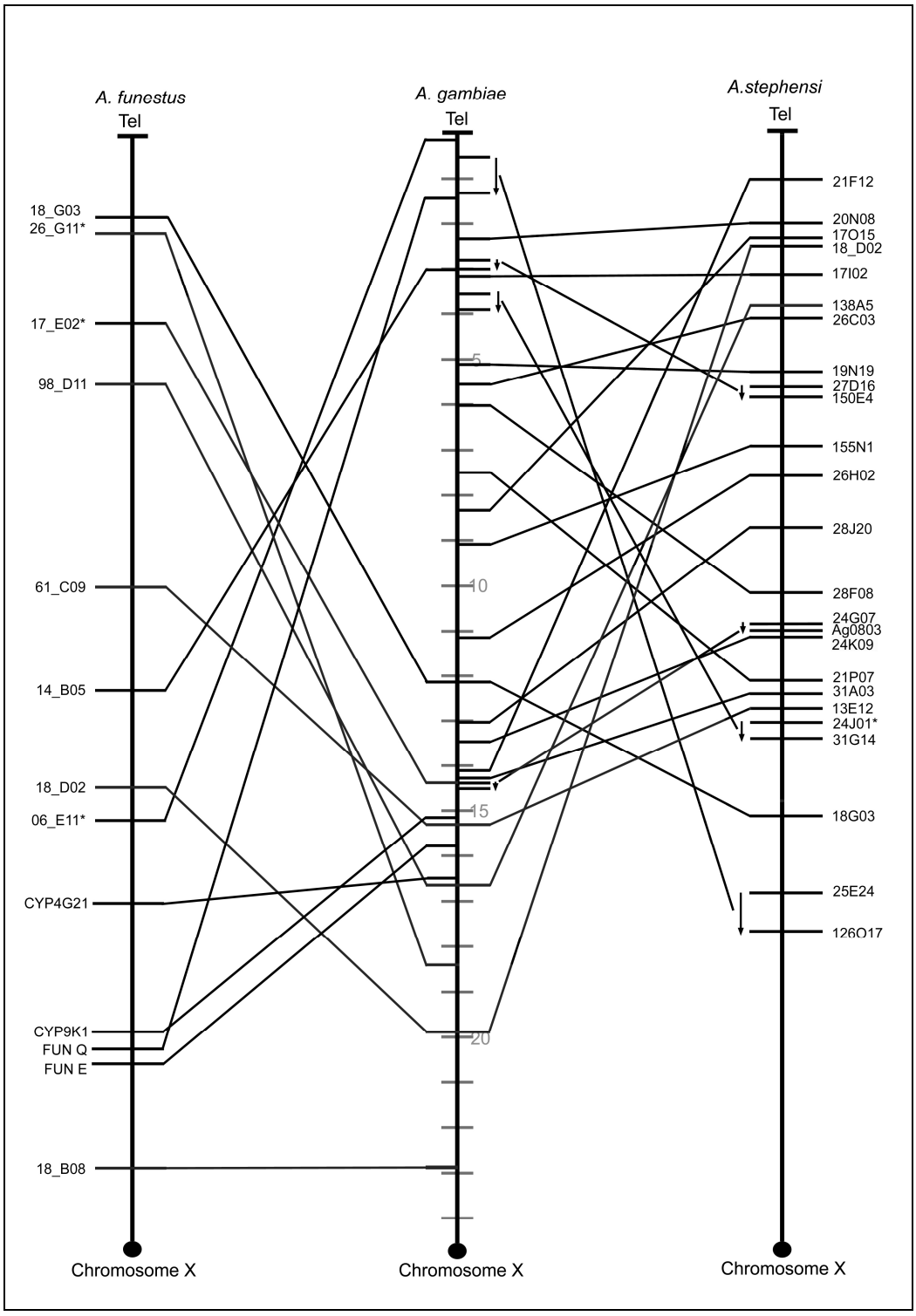


Figure 3.2: Comparison between *A. gambiae* with *A. funestus* and between *A. gambiae* with *A. stephen*si on X chromosomal arm.

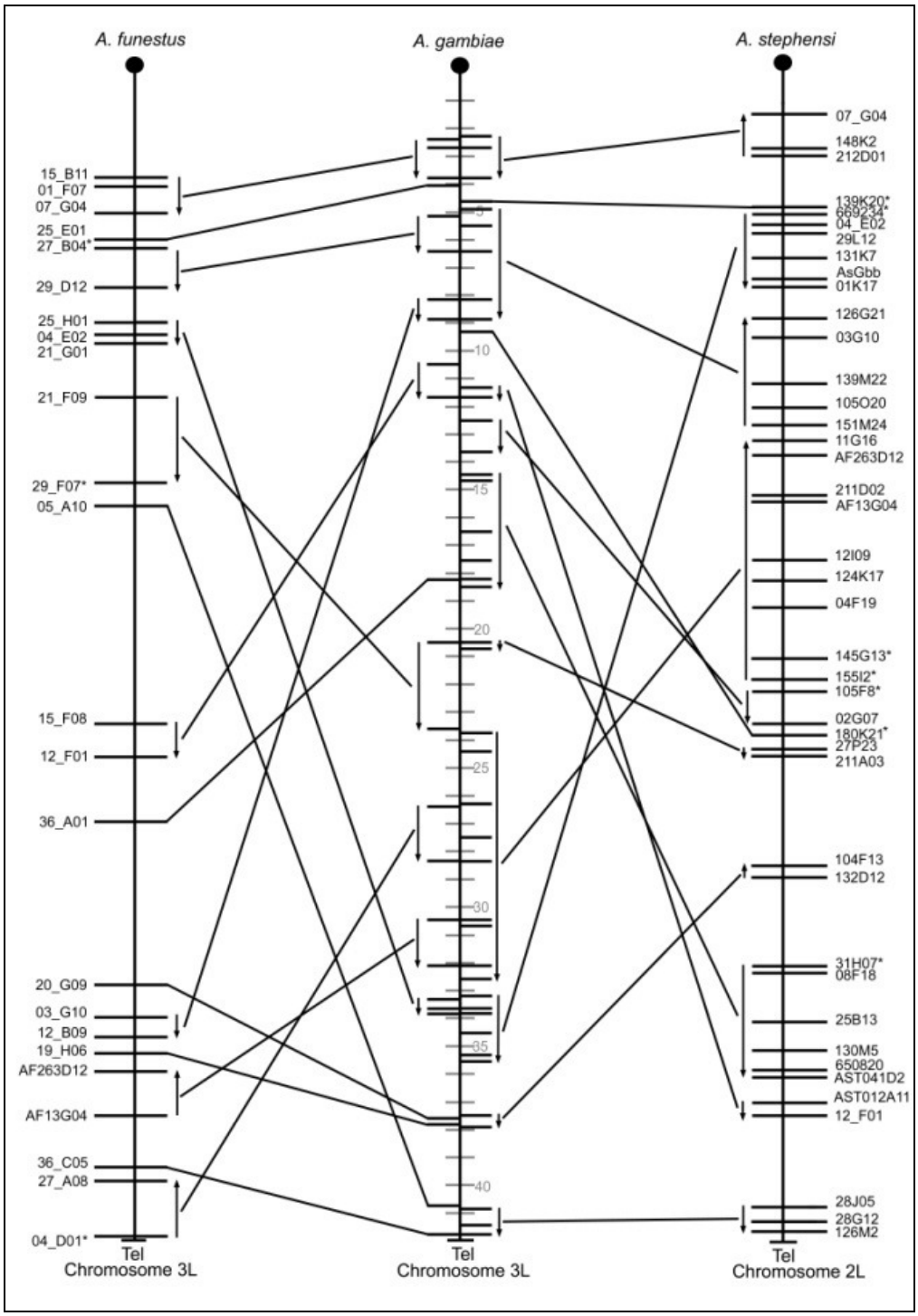


Figure 3.3: Comparison Comparison between *A. gambiae* with *A. funestus* and between *A. gambiae* with *A. stephen*si on 3L chromosomal arm.

In order to test the distribution of the lengths of synteny blocks, the Nadeau & Taylor model was used to fit our data listed in **Appendix 3.7**. **Figure 3.4** compares the cumulative distribution and density function from the N-T model (L= 1.95 Mb) to the empirical distribution of the data. Using the Nadeau & Taylor model, we noted that the CDF region defined between E[L] ± 2SE should (asymptotically) correspond to a 95% confidence region. **Figure 3.5** demonstrates that ≈ 29.6 % of the data fall outside of the E[L] ± 2SE, which suggest the poor predictivity of the N-T model for our data. In reality, there are more short synteny blocks than would be expected under N-T model. Our results agree well with the comparison results between human and mouse genome sequences (*62, 90, 91*). In comparing human and mouse, the lengths of large synteny blocks with the size of at least 1Mb still fitted the exponential distribution, but a large number of short synteny blocks in length < 1Mb cannot be explained by the N-T model. Similar observations have been found in *Drosophila* (*63, 64*). In this study, although the complete genome sequence of *A. stephensi* is not available, several short synteny blocks several Kb in length have been found in *A. gambiae* and *A. stephensi* genomes. The existence of these small synteny blocks allows our data against random breakage model.

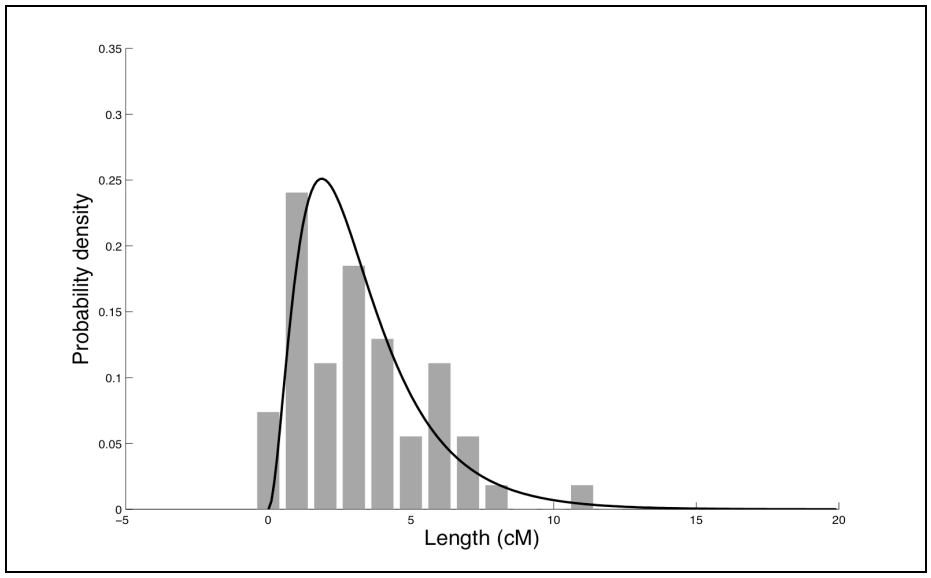


Figure 3.4: Comparison of data (histogram) to the N-T density function.

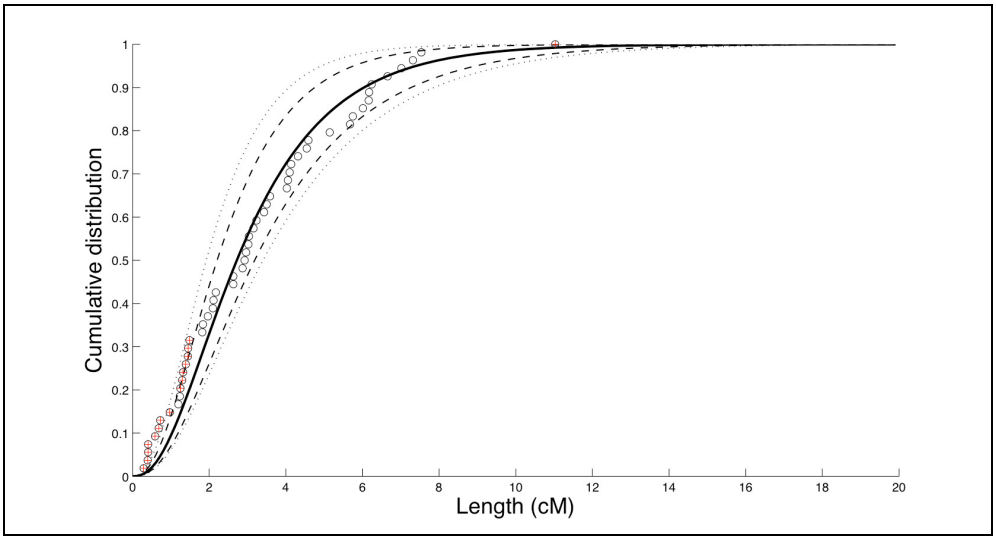


Figure 3.5: Error rates of fitting data to the density function of N-T: circles with (red) pertain to data outside side of the CDF curves, computed under E[L] ±2SE.

4.3. Chromosomal arms evolve at different rates: X is the fastest

The Genome Rearrangements in Man and Mouse (GRIMM) program was used to estimate the minimum number of rearrangement events and find an optimal scenario transformation from *A. gambiae* to *A. stephensi*. A total of 78 fixed inversions have been identified between *A. stephensi* and *A. gambiae* (**Table 3.2**). Our results revealed that

the X chromosome has the highest rate of inversion fixation whereas autosomes vary in the inversion density: 2R>2L>3R>3L. Moreover, statistical analysis results suggest that the densities of inversion fixation on the five different chromosomal arms vary significantly ($\chi^2 = 13.6241$, $p = 0.0086$). The mean lengths of synteny blocks on each arm have been estimated by the N-T model and the mean lengths (X, 0.600050 megabases (Mb); 2R, 1.315324 Mb; 2L, 1.712251 Mb; 3R, 3.755731 Mb; and 3L, 2.412339 Mb) were used to infer the number of fixed inversions between *A. gambiae* and *A. stephensi.* If each inversion requires two disruption events, then *n* inversions result in 2*n* - 1 conserved segments. The number of inversions can be calculated by dividing the total length of the arm by the mean length CSBs, which was given in **Table 3.2**. The correlation coefficient between the two methods was 0.92. Both of these methods revealed that the highest number of fixed inversions was found on the X chromosome which is consistent with the shortest synteny blocks. While for the autosomes, the 2R chromosomal arm has more fixed inversions and shorter synteny blocks than other chromosomes. Our results strongly suggest that the sex chromosome evolves faster than autosomes. A similar observation has been found in some *Drosophila* lineages (*33, 37, 63, 64, 71, 80, 136*), which is in contrast with primate genome evolution (*140, 141*).

The minimum number of inversion fixations required for transformation from *A. gambiae* to *A. funestus* was also estimated using the GRIMM program based on the 1.8 Mb resolution of physical mapping which is presented in **Table 3.3**. Our results suggest that the fastest evolution of the X chromosome was observed between *A. gambiae* and *A. funestus*. Among the autosomes, the density of fixed inversions on 2R is higher than

the others. Therefore, the fast evolution of sex chromosome may be an important feature of chromosomal rearrangements in *Anopheles* mosquitoes.

Table 3.2: The number of inversion fixations calculated by the N-T model and GRIMM program between *A. gambiae* and *A. stephensi*

Chromosome arm (*A. gambiae*)	**Chromosome length (Mb)**	**N-T analysis**		**GRIMM analysis**	
		Number of inversions	**Number of inversions per Mb**	**Number of inversions**	**Number of inversions per Mb**
X	24.4	19.826	0.813	15	0.61
2R	61.5	22.895	0.372	29	0.47
2L	49.4	13.915	0.282	16	0.32
3R	53.3	6.583	0.124	11	0.21
3L	42	8.198	0.195	7	0.17
Total	230.5	71.417	0.31	78	0.34

Table 3.3: The number of inversion fixation calculated by GRIMM program between *A. gambiae* and *A. funestus*

Chromosome arm (A. gambiae)	**Chromosome length (Mb)**	**GRIMM analysis**	
		Number of inversions	**Number of inversions per Mb**
X	24.4	8	0.33
2R	61.5	16	0.26
2L	49.4	10	0.2
3R	53.3	6	0.11
3L	42	7	0.17
Total	230.5	47	0.204

4.4. Breakpoints clusters and the distributions of fixed inversions between *A. gambiae* and *A. stephensi*

The studies of chromosomal evolution in *Drosophila* species demonstrated the non-random distribution of breakpoints. There are many chromosomal regions with two or more inversion breakpoints clustered together (*34, 37, 146-148*). A similar observation has been made in *Anopheles* mosquitoes (*13*). The extensive studies of genome evolution suggested that breakpoint reuse is an outstanding feature of mammal genome evolution (*62, 93, 384*). In this study, GRIMM software was used to estimate the minimal number of rearrangement events between *A. gambiae* and *A. stephensi*, and this program also found an optimal scenario for the transformation from *A. gambiae* to *A. stephensi* for each chromosomal arm (**Figure 3.6-3.7** and **Appendix 3.8-3.10**). Fifteen paracentric inversions are required to transform the X chromosome of *A. gambiae* to that of *A. stephensi*. **Figure 3.6** shows that the breakpoints of the fixed inversions are distributed randomly from telomere to centromere and the sizes of inversions range from two markers to 22 markers. This suggests that the breakage can occur at any X chromosome site. The proximal breakpoint of inversion 1 is located very close to the proximal breakpoint of inversion 2. The proximal breakpoint of fixed inversion 7 clusters together with the distal breakpoint of inversion 8. The proximal breakpoint of inversion 5 overlaps with the distal breakpoint of inversion 6. Although it took only 7 inversions to transform the *A. gambiae* 3L to the *A. stephensi* 2L, there were two breakpoints (distal breakpoints of inversion 4 and 5) clustering together. The closely located breakpoints were also observed in other autosomes (**Table 3.4**). Therefore, breakpoint clustering is

a common feature in the genome evolution between *A. gambiae* and *A. stephensi*

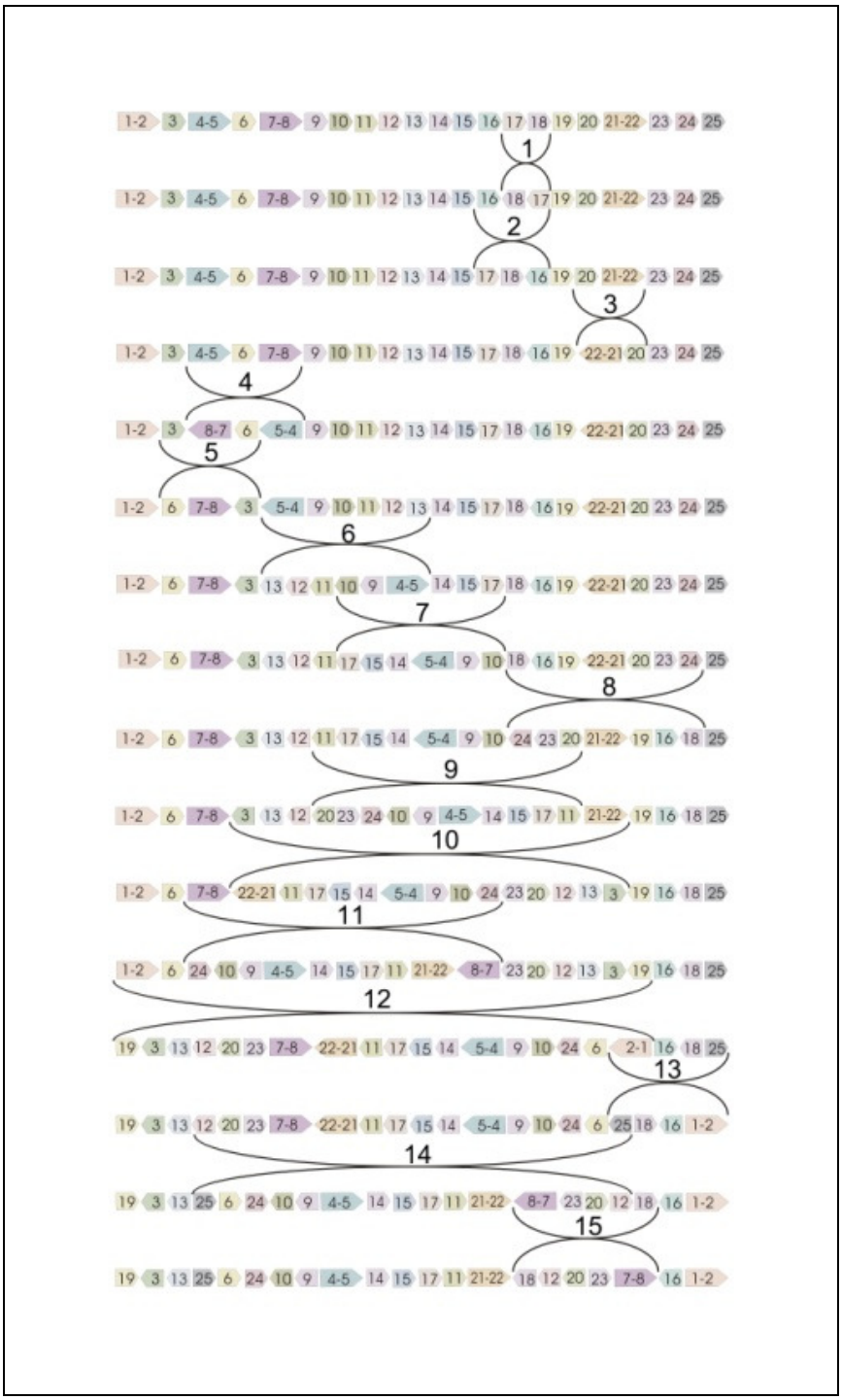


Figure 3.6: The scenario of transformation from *A. gambiae* X to *A. stephensi* X chromosome. Conserved synteny blocks are numbered consecutively (from telomere to Centromere) in *A. gambiae*.

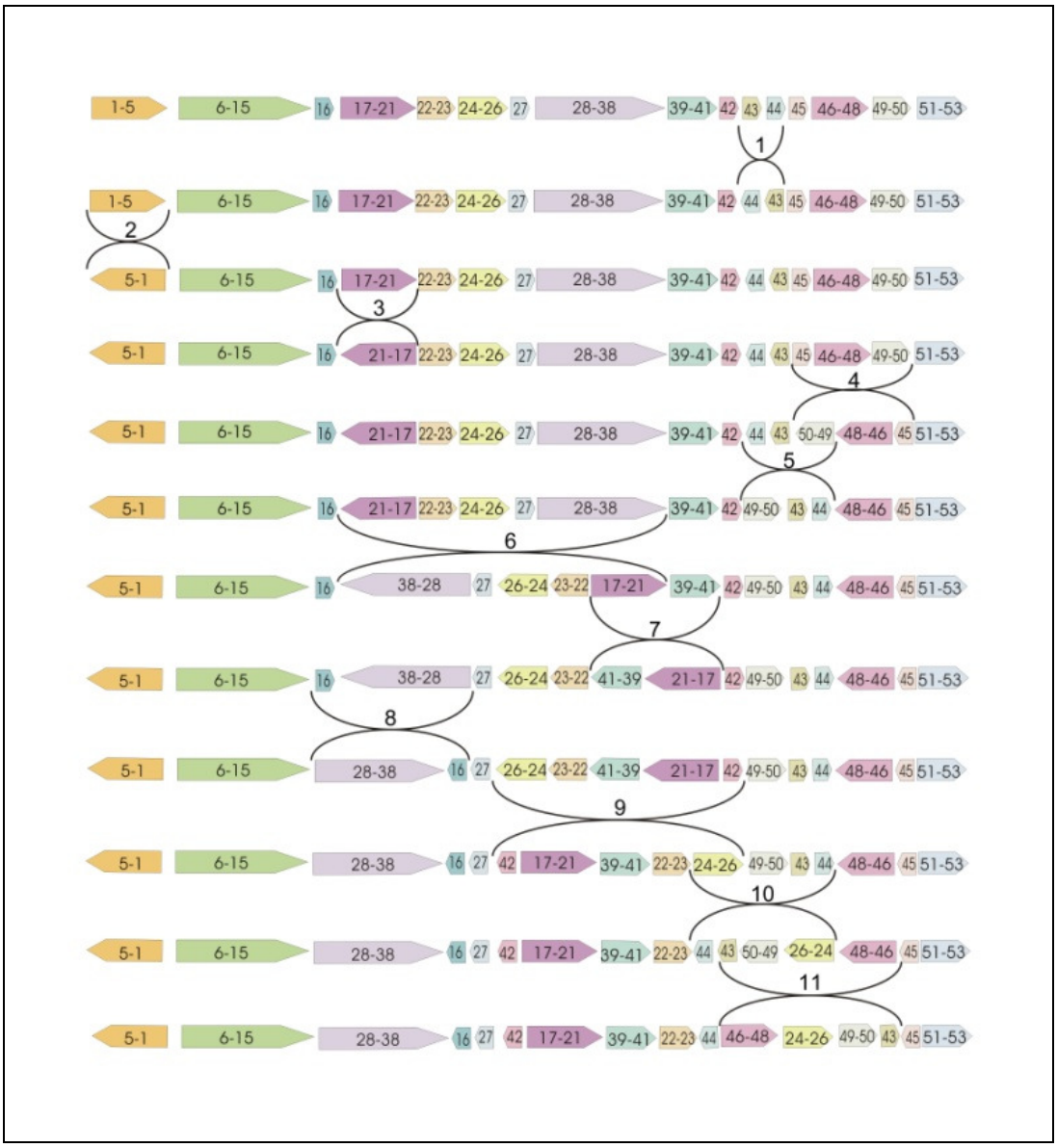


Figure 3.7: The scenario of transformation from *A. gambiae* 3R to *A. stephensi* 3R chromosome.

Table 3.4: The inversion breakpoint clusters in *A. gambiae* and *A. stephensi*

Chromosomal arm in *A. gambiae*	Fixed inversion breakpoint cluster
X	(1) Proximal breakpoints of inversion 1 and 2 (2) Proximal breakpoint of 5 and distal breakpoint of 6 (3) Proximal breakpoint of 7 and distal breakpoint of 8
2R	(1) Proximal breakpoints of inversion 1 and 5 (2) Proximal breakpoint of 2 and distal breakpoint of 4 (3) Distal breakpoints of 6 and 8 (4) Proximal breakpoints of inversion 14 and 16 (5) Proximal breakpoints of inversion 13 and 17 (6) Proximal breakpoints of inversion 25 and 17 (7) Distal breakpoints of 24 and 27 (8) Distal breakpoints of 20 and 21
2L	(1) Distal breakpoints of 8 and 9 (2) Distal breakpoints of 15 and 16
3R	(1) Distal breakpoints of 1 and 5 (2) Distal breakpoints of 3 and 6
3L	(1) Distal breakpoints of 4 and 5
Total	16 breakpoint clusters

4.5. Polymorphic and fixed inversions in *Anopheles* mosquitoes

The five chromosomal arms differ not only in their rates of inversion fixation, but also in the extent of chromosomal polymorphisms in malaria mosquitoes. The difference in the distribution of polymorphic inversions may contribute to the variation of the fixation rate among the five chromosomal arms. According to the Kirkpatrick and Barton model, the local adaptation mechanism can cause the establishment of the new inversion as polymorphic, and drive it to fixation (*208*). If this model is correct, then parallelism between the rates of polymorphic and fixed inversions should be expected.

The polymorphic inversions in *A. gambiae* and *A. stephensi* have been described in previous publications (*13, 17, 41*) and listed in **Table 3.5**. In *A. gambiae* s.s, seven common polymorphic inversions were identified, and these inversions were considered as markers of local ecological adaptation (*207*). This view is based on the observations showing the associations of alternative rearrangement with adaptation to the contrasting habitats in Africa (*13, 42, 216, 250, 385*). In *A. stephensi*, the common polymorphic inversion 2Rb, is the most frequent and widespread. This inversion was also found associated with adaptation to the urban environment (*17, 41, 386*). The absence of polymorphic inversions on the X chromosome in *A. gambiae* and *A. stephensi* conflicts with the fastest accumulation of fixed inversions on the sex chromosome. This suggests that the sex chromosome may contribute to speciation through rapid generation and fixation of new inversions without maintaining them as polymorphic. When only autosomes are considered, chromosome 2R, with the highest fixation rate of inversion fixation exhibits the highest level of polymorphism, followed by chromosome 2L with lower fixation rate and polymorphism. Chromosome 3R and 3L with no polymorphic inversions harbored the least number of inversion fixations. Therefore, there is a good correspondence between evolutionary rate and the level of chromosomal polymorphism on autosomes. When common polymorphic inversions in *A. gambiae* and *A. stephensi* were jointly analyzed, a significant correlation in the number of fixed and polymorphic inversions ($r = 0.94$ for GRIMM and $r = 0.88$ for N-T) had been found (**Figure 3.8**). The relationship between polymorphic inversion and the level of inversion fixation was also analyzed between *A. gambiae* and *A. funestus*. The data regarding polymorphic inversions have been described in the paper of Sharakhov et al (*24*) and are present in

Table 3.6. When common polymorphic inversions in *A. gambiae* and *A. funestus* were jointly analyzed, a significant correlation in the number of inversion fixations and polymorphic inversions (r = 0.95) had been found on autosomes (**Figure 3.9**).

Table 3.5: The polymorphic inversions and fixed inversions between *A. gambiae* and *A. stephensi*

Arm	Fixed inversions/10 Mb (GRIMM)	Fixed inversions /10Mb (N-T)	Polymorphic inversions/10Mb in *A. gambiae*	Polymorphic inversions/10Mb in *A. stephensi*	Common polymorphic inversions/10Mb (joint)
X	6.148	8.125	0.000	0.000	0.000
2R	4.715	3.723	0.976	0.163	1.139
2L	3.239	2.817	0.202	0.000	0.202
3R	2.068	1.237	0.000	0.000	0
3L	1.667	1.952	0.000	0.000	0

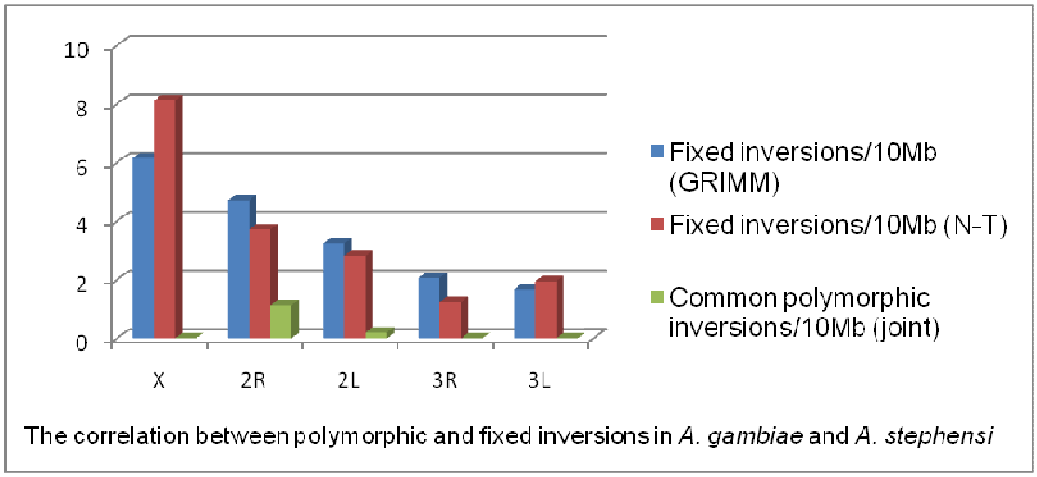


Figure 3.8: The fastest evolution of the X chromosome and parallelism between the extent of inversion polymorphism and inversion fixation rates on the autosomes in *A. gambiae* and *A. stephensi*.

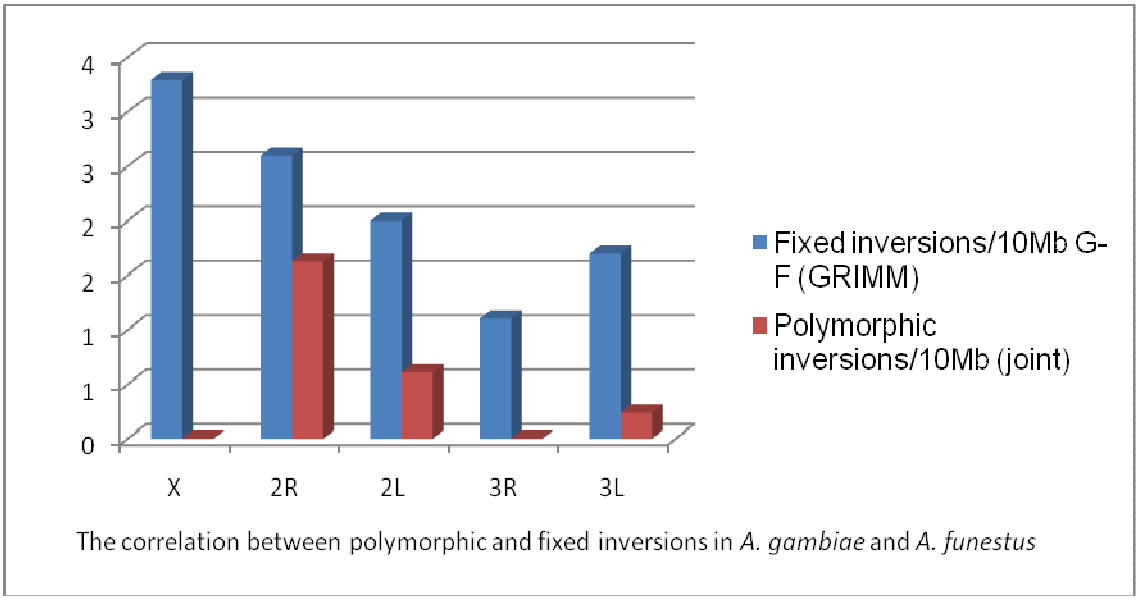


Figure 3.9: The fastest evolution of the X chromosome and parallelism between the extent of inversion polymorphism and inversion fixation rates on the autosomes in *A. gambiae* and *A. funestus*.

Table 3.6: The polymorphic inversions and fixed inversions between *A. gambiae* and *A. funestus*

Arm	**Fixed inversions/10Mb *G-F* (GRIMM)**	**Polymorphic inversions/10Mb in *A. gambiae***	**Polymorphic inversions/10Mb in *A. stephensi***	**Polymorphic inversions/10Mb (joint)**
X	3.3	0.000	0	0.000
2R	2.6	0.976	0.65	1.626
2L	2	0.202	0.405	0.607
3R	1.1	0.000	0	0
3L	1.7	0.000	0.238	0.238

4.6. The conserved and disrupted synteny blocks in *A. gambiae*, *A. funestus* and *A. stephensi*

231 *A. stephensi* and 145 *A. funestus* physically mapped probes (**Figure 3.1-3.2** and **Appendix 3.4-3.6**) were used to analyze whether synteny blocks were conserved among three species or disrupted in one of the species. The results (**Table 3.7**) show that most of the synteny blocks were conserved among *A. gambiae, A. stephensi,* and *A. funestus*. The number of shared synteny blocks is significantly higher than the number of disrupted synteny blocks ($\chi 2 = 14.1697$, $p= 0.0002$). This suggests the existence of functional gene clusters that constrain chromosomal breakage. The number of synteny blocks shared between *A. gambiae* and *A. stephensi* is significantly different than the number of synteny blocks shared between *A. gambiae* and *A. funestus* ($\chi 2 = 56.5537$, $p<0.0001$). This suggests that *A. gambiae* is more closely related to *A. stephensi* than to *A. funestus*, which agrees with molecular analysis of phylogenetic relationships (*387, 388*).

Since the disrupted blocks were accumulating through time, we can envision through a backwards time perspective, the conserved blocks are accumulating through reversed-time. Using this framework, we accounted for both the number of conserved blocks and the length of each of these. We modeled this using a compound Poisson process, where the number of conserved blocks follows a Poisson process, where the length of the chromosomal arm scales the rate of the process; a separate process governs the length of each conserved block. We found that 2R has the highest rate of accumulation of disrupted blocks per unit length $\lambda_j^{(dif)}$ with probability equal 0.905 (**Table 3.8**). In contrast, 3R had the lowest $\lambda_j^{(dif)}$ value with probability 0.5 because no block disruption was detected for this arm.

Table 3.7: Conserved synteny blocks between species and among species.

Chromosomal arm	**Conserved synteny blocks among three species**	**Synteny blocks shared by *funestus* and *gambiae*, but disrupted between *gambiae* and *stephensi***	**Synteny blocks shared by *gambiae* and *stephensi*, but disrupted between *gambiae* and *funestus***
X	0	0	0
2R	9	3	3
2L	5	0	3
3R	6	0	0
3L	8	0	3
Total	28	3	9

Table 3.8: The rate of accumulation of disrupted blocks per unit length in autosomes

Parameter	**2R**	**2L**	**3R**	**3L**
$(r_j^{-1})^{(diff)}$	0.939	0.470	0.500	0.662

4.7. Molecular features associated with fast and slow chromosomal evolution

Although our data suggests that the chromosomal elements evolved at different rates and a significant correlation between polymorphic and fixed inversions in *A. gambiae* and *A. stephensi* was found, the molecular mechanisms for generating chromosomal inversion breakpoints are still not clear. In this research, we analyzed the distributions of molecular features on five chromosomal arms.

Three-dimensional (3D) organization of chromosomes in the nuclear space can affect inter-chromosomal interactions by facilitating or hindering rearrangements. Therefore, Matrix/Scaffold Attachment Regions (M/SARs) can potentially mediate an interaction of specific chromosome sites by binding to the nuclear envelope (NE). We identified M/SARs in the *A. gambiae* genome sequence using the SMARTest bioinformatic tool (**Appendix 3.11**). The analysis of the *A. gambaie* genome revealed a significant

negative correlation between the number of fixed inversions and M/SARs (r = -0.927) (**Figure 3.10**). Our result suggests a role of nuclear architecture in determining chromosome arm specificity of rearrangement rates. However, further work should be done to understand how M/SAR facilitates the generation of inversion breakpoints.

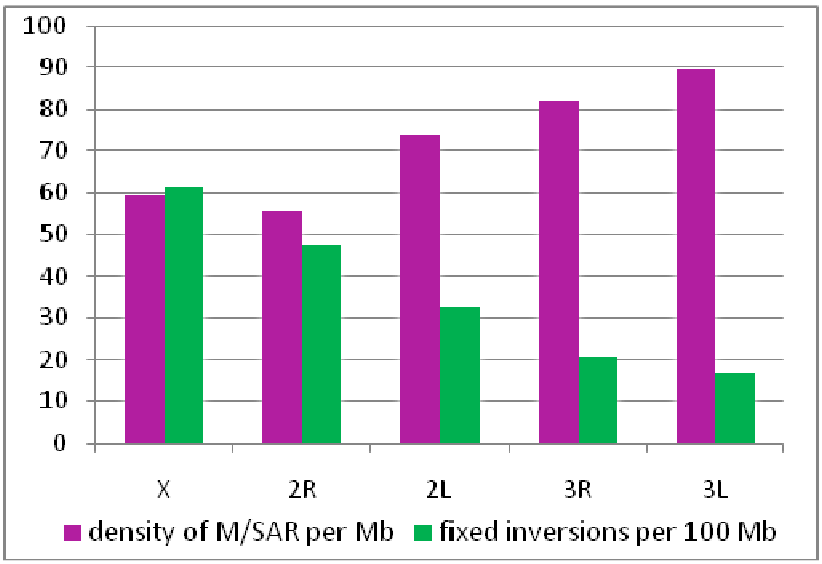


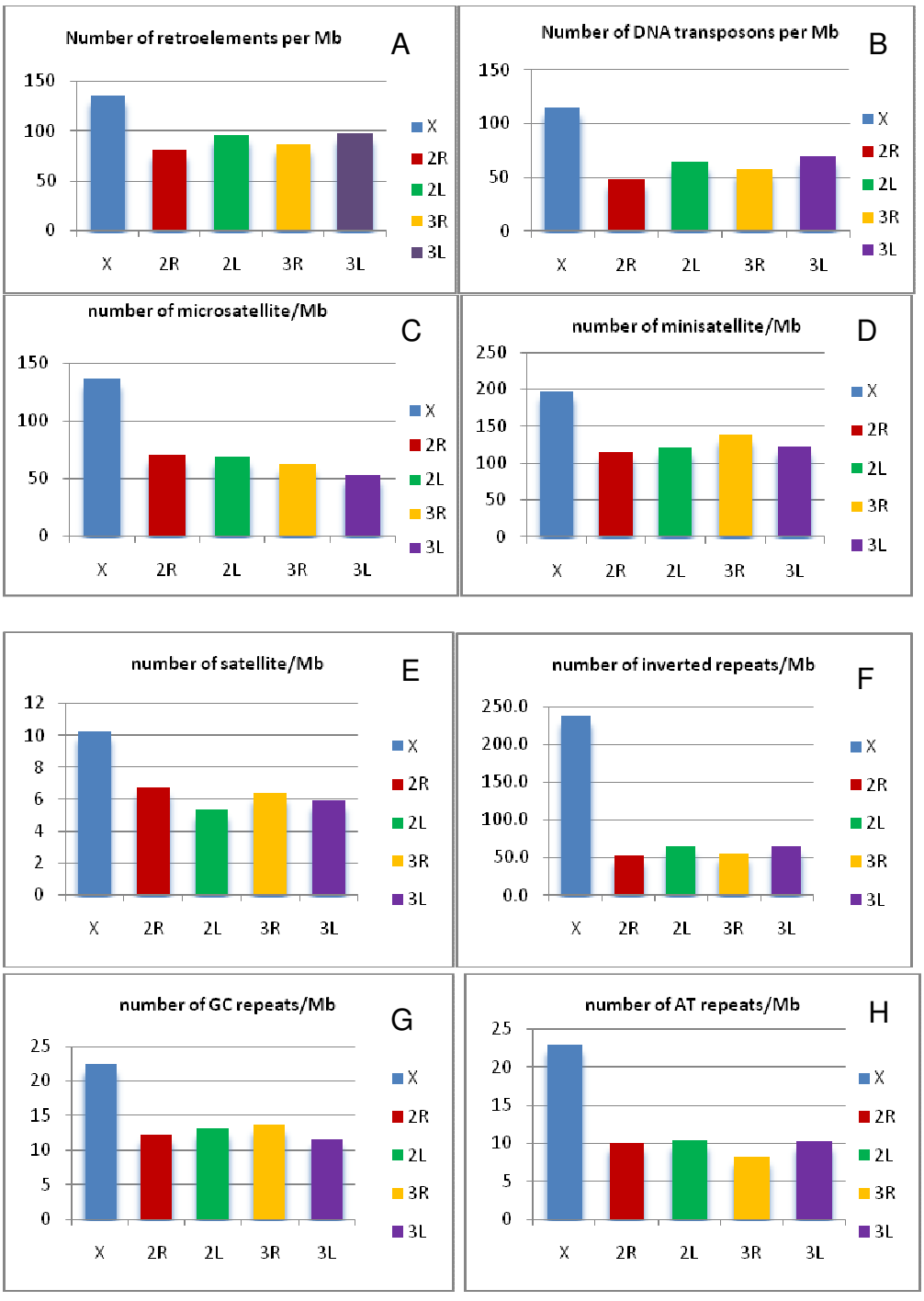
Figure 3.10: Correlation between fixed inversions and M/SARs

Although the molecular mechanism underlying the formation of chromosome rearrangement is not clear, there is strong evidence that TEs are the major force of chromosomal inversions in *Drosophila* (*160-162*) and *Anopheles* (*163, 164*), whose breakpoints have been characterized at the molecular level. Comparative sequence data also indicated that interspecific paracentric inversion breakpoints are enriched in TEs in *Diptera* species (*63, 389*). In order to understand the association of TEs with the variation in evolution rate among the *Anopheles* five chromosomal arms, the DNA transposons and retroelements in *A. gambiae* have been analyzed by the RepeatMasker program (www.repeatmasker.org). The distributions of TEs on chromosomes are present in **Figure 3.11**. **Figures 3.11 A** and **B** show that the density of transposable elements is the highest on the X chromosome. However, the 2R

chromosomal arm, which exhibites a faster evolution rate than other autosomes, has the lowest density of TEs. We have to take into account that there can be local difference in the densities of TEs within chromosomal arms. Indeed, TEs are not distributed randomly on chromosomal arms. There is evidence that TEs are concentrated in centromeric heterochromatin and centromere-proximal euchromatin (*390*).

Other studies suggest that simple repeats (microsatellites) such as AT- and GC-rich micro- and minisatellites as well as inverted repeats can generate unstable secondary structures that could induce the chromosomal rearrangements (*61, 391, 392*). For this purpose, simple repeats were analyzed by Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.submit.options.html). Inverted Repeats Finder (http://tandem.bu.edu/cgi-bin/irdb/irdb.exe) was used to predict the number of inverted repeats. AT and GC repeats were calculated using ATcontent program from Dr. Tu's program. The densities of each molecular feature on the five chromosomal arms were calculated using the total counts of molecular features divided by the length of the chromosome (**Appendix 3.11**). **Figure 3.11** shows that the X chromosome harbored more microsatellites, minisatellites and satellites than autosomes. AT and GC repeat densities on the X chromosomes are at least twice that of any of the autosomes. The density of inverted repeats is about five times higher on sex chromosome than on autosomes. All these factors could contribute to the fast generation rate of fixed inversions on the X chromosome. However, the contribution of repetitive elements to the variation of fixation rate of autosomes seems difficult to explain.

Another type of repeat, the segmental duplication (SD), was extensively studied in primate genomes and has been implicated in the generation of chromosomal rearrangements. In this study, SD in the *A. gambiae* genome has been analyzed by Dr. Bailey. **Figure 3.11 I** revealed that the highest density of SDs was observed on the 2R chromosome which might contribute to the fragility of the 2R arm. However, the association of SDs with the level of inversion fixation rate on other autosomes is unclear. In addition to the repetitive sequences, we also analyzed the gene density on five arms. The number of genes was counted by Biomart. The X chromosome has the least number of genes as compared to autosomes. This may suggest that the chromosomal arm with a low gene density might have more fixed inversions. Characterization of inversion breakpoints revealed that fixed inversions more likely have breakpoint between genes (*162, 393*) than within transcription units (*394*). In the latter case, chromosomal inversions would have a strong deleterious effect and could be removed by natural selection. But no apparent correlation between gene density and evolution rate was identified on autosomes.


Number of retroelements per Mb
A
150
100
50
0
X
2R
2L
3R
3L
Number of DNA transposons per Mb
B
150
100
50
0
X
2R
2L
3R
3L
number of microsatellite/Mb
C
150
100
50
0
X
2R
2L
3R
3L
number of minisatellite/Mb
D
250
200
150
100
50
0
X
2R
2L
3R
3L
number of satellite/Mb
E
12
10
8
6
4
2
0
X
2R
2L
3R
3L
number of inverted repeats/Mb
F
250.0
200.0
150.0
100.0
50.0
0.0
X
2R
2L
3R
3L
number of GC repeats/Mb
G
25
20
15
10
5
0
X
2R
2L
3R
3L
number of AT repeats/Mb
H
25
20
15
10
5
0
X
2R
2L
3R
3L

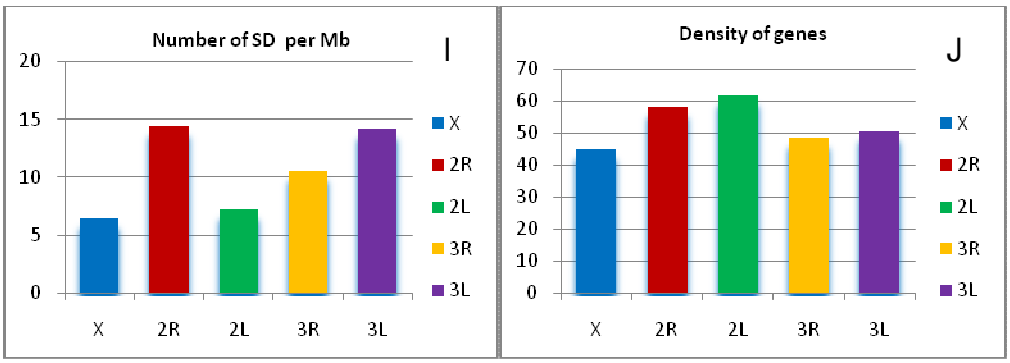


Figure 3.11: The density of molecular features in chromosomal arms of *An. gambiae*. (A) Number of retroelements per 1 Mb. (B) Number of DNA transposons per 1 Mb. (C) Number of microsatellites. (D) Counts of minisatellites. (E) Counts of satellites. (F) Number of inverted repeats. (G) Number of GC repeats. (H) Number of AT repeats. (I) Number of segmental duplications. (J) Gene density per Mb.

4.8. The comparison of molecular features in breakpoint regions and synteny blocks

The analysis of molecular features on whole chromosomal arm level has revealed that the highest density of repetitive elements could be the driving force of the fastest genome evolution on X chromosome. However, these data failed to explain the content of fixed inversions on autosomes. The local differences in molecular features were also studied in this chapter. In order to analyze the association of molecular features in breakpoint regions (<1 Mb), which is the chromosomal regions between the synteny blocks and synteny blocks (indicated by arrows in **Figure 3.1- 3.2** and **Appendix 3.4-3.6)**, we collected all the data from breakpoint regions and synteny blocks (listed in **Appendix 3.12-3.13**) and analyzed using a regression model. In **Table 3.9**, e^{θ} is interpreted as the expected rate increase in the number of molecular elements that would be expected by being in a synteny breakpoint group (as opposed to a synteny block group). For example, when the molecular features on the five chromosomal arms

were analyzed, we found that there is approximately 3.8651 times more AT repeats in breakpoint regions than in synteny blocks for the same length. **Table 3.9** shows that breakpoint regions are rich with AT repeats, transposable elements, inverted repeats, minisatellites, unclassified repetitive elements, and M/SARs. When we excluded X and 2R chromosomal arms, our data show that the value of parameters are much higher. This could be explained by the high number of fixed inversions on the X and 2R chromosomes. Our results reveal that the presence of a higher number of TEs, AT repeats, inverted repeats, and other unknown repeats, as well as M/SARs in breakpoint regions may facilitate the generation of inversion fixation breakpoints in malaria mosquitoes.

Table 3.9: The estimated parameter values of molecular elements

Element	**e^{θ} (include all five chromosomal arms)**	**e^{θ}(exclude of X and 2R)**
AT repeats	3.8651	8.2796
Transposable elements	2.6548	5.4236
DNA transposons	2.0434	3.4388
Retroelements	3.1362	7.0584
Inverted repeats	3.102	6.6781
Microsatellite	0.9153	2.5954
Genes	0.7211	0.8517
M/SAR	1.8937	3.68
Simple repeats	0.7033	0.5951
Minisatellite	1.3313	2.9275
Satellite	0.5740	0.7428
Unclassified	3.5574	10.0645
Low complexity	1.4478	1.9869

5. Discussion

5.1. Rates of chromosomal evolution in *Anopheles* and *Drosophila*

231 markers uniformly distributed on the five chromosomal arms of *A. stephensi* and 127 previously published markers with even coverage on chromosomes of *A. funestus* have been used for a detailed comparison of gene orders with those of the homologous *A. gambiae* chromosomal arms. About 71 to 78 inversions have been fixed between *A. gambiae* and *A. stephensi,* (**Table 3.2**) and 47 paracentric inversions have been identified between *A. gambiae* and *A. funestus* after the divergence of the two lineages (**Table 3.3**). Taxonomically, *A. gambiae*, *A. funestus,* and *A. stephensi* belong to different series: *Pyretophorus* (*A. gambiae*), *Myzomyia* (*A. funestus*), and *Neocellia* (*A. stephensi*) of the subgenus *Cellia* (*14, 16*). Further molecular analysis of the complete mitochondrial DNA from *A. funestus* and *A. gambiae* reveals that *A. gambiae* and *A. funestus* lineages diverged from a common ancestor at least 36 million years ago (*217*). Additional evidence of phylogenetic relationships based on the mitochondrial DNA and ribosomal DNA within the subgenus *Cellia* estimate that *A. gambiae* is more closely related to *A. stephensi* than to *A. funestus* (*387, 388*). If we assume that *A. gambiae* and *A. stephensi* diverged from common ancestor about 30 million years ago (MYA), the evolutionary rate of fixed inversions can be estimated. These estimates have been normalized as the rates of chromosome evolution, which are the numbers of disruptions per megabase per million years and then compared with other previously published data in the *Drosophila* species (**Table 3.10**). Lineage specific patterns of chromosomal rearrangements have been observed in the genus *Drosophila*. So far, the fastest rates of chromosomal inversions can be observed within the *Sophophora* subgenera. The intermediate rates were found between *Drosophila* and *Sophophora* subgenera and the

lowest rate is within *Drosophila* subgenera. Our data agree well with previous studies (*31, 34*). The rates of chromosomal rearrangements in *Anopheles* are similar to those in subgenus *Drosophila*. Moreover, the rates of fixed inversions differ among chromosomal arms. The available evidence from *Drosophila* (*33, 37*) and our current comparative analysis in *Anopheles* suggest that the sex (X) chromosome evolve faster than autosomes. For the autosomes, the rates of chromosomal rearrangements vary in different lineages. In the genus *Anopheles*, the 2R chromosome has the highest level of inversion polymorphism and fixation. This common feature can be extended to other *Anopheles* species.

Table 3.10: The comparison of the evolutionary rates in *Anopheles* mosquitoes and *Drosophila* species

Subgenera comparison	**Species comparison (divergence time)**	**Chromosome element**	**Rate of rearrangement (breakpoints/Mb/MY)**	**Reference**
Sophophora - Sophophora	*D.subobscura-D. pseudoobscura* (8MYA)	B C	0.08 0.17	2006 (*31*)
	D.melanogaster-D. pseudoobscura (30 MYA)	B C	0.138 0.126	
	D.melanogaster-D. pseudoobscura (30MYA)	A B C D E	0.14 0.08 0.1725 0.0925 0.117	Calculated based on 2005 (*63*)
Sophophora - Drosophila	*D.melaogaster-D. repleta* (62MYA)	A D	0.087 0.037	2002 (*33*)
		E	0.066-0.053	2001 (*30*)
		B	0.021	2000 (*373*)
	D. melanogaster-D. virilis (40MYA)	B	0.016	1996 (*395*)
Drosophila - Drosophila	*D.virilis-D.Montana* (9MYA)	A C	0.027 0.005	Calculated based on 1997 (*37*)
	D.novamexicana-D. montata (9MYA)	A C	0.027 0.0075	
	D.viriis-D. novamexicana (3.8MYA)	A C	0.015 0.0055	

	D.repleta-D.buzzatii (15MYA)	B C E	0.003 0.003 0.020	2003 (*73*)
		A and B	0.004	2007 (*34*)
Anopheles	*A.gambiae-A.funestus* (36MYA)	X 2R 2L 3R 3L	0.0032 0.0079 0.0030 0.0029 0.0061	Recalculated based on 2002 (*24*)
		X 2R 2L 3R 3L Average	0.0092 0.0072 0.0056 0.0031 0.0047 0.0057	This study
	A.gambiae-A.stephensi (30MYA)	X 2R 2L 3R 3L Average	0.0271(0.0204) 0.0124(0.0156) 0.0094(0.0107) 0.0041(0.007) 0.0065(0.0057) 0.010(0.0113)	This study

5.2. Nonrandom distribution of inversion breakpoints

The random breakage model implies that the inversion breakpoints are distributed randomly on the chromosome and there are no evolutionary “fragile regions” in genomes (*43*). Therefore, this model rules out the reuse of breakpoints on chromosome sites. However, the discovery of the breakpoint clustering from the comparison of human and mouse genome sequences strongly suggests that chromosomal breakage tends to reoccur at “fragile sites” or “hotspots” in mammalian genome (*62*). Later, more studies provided additional evidence in favor of the fragile breakage model (*61, 93, 152, 153, 176*). For instance, the analysis of genome reorganization of *Drosophila* using comparative mapping revealed enormous breakpoints clustered together (*35, 37, 63, 64, 146, 148*). A similar observation has been made in *Anopheles* mosquitoes (*13*). For the first time in the *Anopheles* species, our fairly dense physical map of *A. stephensi*

allowed a detailed comparison of gene rearrangements with those of *A. gambiae*. A total of 55 synteny blocks have been identified between *A. gambiae* and *A. stephensi* since their divergence from the common ancestor. The length of these 55 synteny blocks has been used to fit the random breakage model. Our results revealed that 29.6% of our data can't be explained by the N-T model (**Figure 3.5**). Further analysis of these data demonstrated that small synteny blocks (<1Mb) caused deviation from the N-T model. This is in good agreement with the observations from mammalian (*62*) and *Drosophila* genomes (*63, 64*). Thus, our data support the fragile breakage model. Computational software GRIMM has been used to analyze the minimum number of inversions required for transformation from the *A. gambiae* to the *A. stephensi* genome, and the scenarios suggested that there are 16 sites of breakpoints clustering. Although our data failed to confirm that these breakpoints are reused or are just coincidently close together, our results strongly suggest that some chromosomal regions harbored more breakpoints than the others.

5.3. The sex chromosome has the highest rate of rearrangements and is enriched with repetitive DNA

The studies of gene orders between *A. gambiae* and *A. stephensi*, *A. gambiae* and *A. funestus* revealed that the sex (X) chromosome has the highest inversion rates among the five chromosomal arms. Only four synteny blocks have been identified on the X chromosome between *A. gambiae* and *A. stephensi* at a 1 Mb resolution of physical mapping. The gene orders have been extensively reshuffled from telomere to centromere along the chromosome. The GRIMM scenario for the X chromosome revealed that fixed inversions contain from two to 22 markers and breakage can occur

at any chromosome sites. However, in mammals, the X chromosome exhibits the extensive conservation of synteny blocks (*145, 396*), which is in contrast to the fast evolution of the X chromosome in *Drosophila* species (*33, 37*). The unequal evolution rates between the sex chromosome and autosomes in *Anopheles* and *Drosophila* could be explained by the following factors. Rice, in 1984, pointed out that X-linked chromosomal inversions with antagonistic effects in the two sexes will invade the population under a wide range of conditions (*397*). Later, Charlesworth et al. (1987) showed that the sex chromosome should evolve faster than autosomes because of the high fixation rate of underdominant and advantageous partially recessive mutations (*139*). An additional reason for the higher rearrangement rate could be because of the lesser functional constrains (*145*). For these reasons, the X chromosome has probably played an important role in speciation. The evidence from *Drosophila* suggests that many hybrid sterility genes are X-linked (*398, 399*). Interestingly, in mice, many genes on the X chromosome are involved in the sperm formation (*399, 400*).

Despite the faster evolution of the X chromosome, the molecular mechanism in the generation of inversion fixation breakpoints on the X chromosome and driving forces responsible for the establishment and maintenance of these inversions are still not clear. We demonstrated that X chromosome harbored more TEs, microsatellites, minisatellites and satellites than autosomes (**Figure 3.11**). AT, GC, and inverted repeat densities on X chromosomes are at least twice higher than on the autosomes. TEs have been implicated in the formation of inversion breakpoints in *Drosophila* (*160, 162, 401, 402*). Additionally, the sequencing of inversion breakpoints in *Anopheles* has confirmed that TEs are present at the inversion junction (*163, 164*). Moreover, it has been proposed

that microsatellite sequences and inverted repeats can generate unstable secondary structures (*403, 404*), which are capable of forming chromosomal inversions (*392, 405, 406*). In *Drosophila*, the microsatellite density of the chromosomes parallels their evolution rates (*33*). All these data suggest that repetitive elements might play a major role in the origin of fixed inversions on the X chromosome in *Anopheles* species. In contrast to the high inversion fixation rate on the X chromosomes, polymorphic inversions are rare on the sex chromosome in *Anopheles* mosquitoes.

5.4. Molecular determinants of autosomal evolution

In addition to the unequal rates of rearrangements between the sex chromosome and autosomes, the densities of breakpoints also vary among autosomes. Our comparative analysis shows that the 2R chromosome exhibits the highest density of inversion fixation in *Anopheles* mosquitoes; the intermediate rate is exhibited by 2L, and the lowest density is on 3R and 3L. The GRIMM scenarios reveal that the gene orders on 2R are extensively scrambled from telomere to centromere (**Appendix 3.10**). The comparative gene order image (**Appendix 3.6**) also shows that 2R harbored 14 small synteny blocks (< 1Mb) and the largest one is about 1.2 Mb. Additional common and disrupted synteny blocks in *A. gambiae*, *A. stephensi* and *A. funestus* suggested that 2R has the highest rate of accumulation of disrupted blocks per unit length (**Table 3.8**). All our current evidence suggests that the 2R chromosome is more prone to breakage than any other autosome. The intermediate evolution rate among autosomes has been identified on the 2L chromosome in *A. gambiae* which can be explained by the intermediate rate of accumulation of disrupted blocks per unit length and middle sized

synteny blocks. The 3R chromosome has large synteny blocks that are not tolerant to disruption suggesting the presence of functional constrains to breakage.

We demonstrate that the densities of M/SARs have a negative correlation with the extents of inversion fixations on the five chromosome arms in *Anopheles* mosquitoes (**Figure 3.10**). The systematic analyses of three-dimensional nuclear organization have been extensively studied in *Drosophila* using polytene chromosomes (*407-410*). Other studies suggest that M/SARs can potentially mediate interactions of specific chromosome sites with the NE (*411-414*). In addition, the studies of nuclear architecture in human revealed that the spatial organization of chromosomes in the nucleus might facilitate or hinder chromosomal rearrangements by affecting chromosome interactions (*415-420*). We provided the first clue that nuclear architecture plays a role in determining chromosome specificity of rearrangement rates.

Our results revealed that breakpoint regions have more transposable elements (DNA transposons and retroelements) than synteny blocks (**Table 3.9**). This is in good agreement with the studies of TEs in *Drosophila* (*160, 162, 401, 402*) and *Anopheles* (*163, 164*). Although the studies of microsatellites or simple repeats failed to establish a relationship to inversion fixation, we found that some types of microsatellites such as AT repeats, significantly contribute to the formation of inversion fixation. We also revealed that inverted repeats, minisatellites, M/SARs, and unknown repetitive elements might play an important role in the origin of inversion breakpoints in *Anopheles* mosquitoes. Our data are consistent with the previous overview of these factors in the generation of chromosomal rearrangements (*392, 405, 406*). Therefore, we cautiously conclude that

several causes discussed above contribute to the variation of evolution rates on autosomes of *Anopheles* mosquitoes.

We also studied the driving force for the establishment and maintenance of chromosomal inversions. In the present studies, a good correspondence between evolutionary rates and the extent of chromosomal polymorphism has been identified in *A. gambiae* and *A. stephensi* or *A. funestus* (**Figure 3.8 – 3.9**). Our results could be explained by the local adaptation model (*208*). Similar observations have been made in the *Drosophila* species (*31, 33*). Our results indicate that autosomes may play an important role in ecological adaptation through rapid fixation of polymorphic inversions. In the studies of chromosomal evolution in the *Drosophila* species, Gonzalez et al. suggested that the different extent of polymorphic inversions may contribute to the difference in the inversion fixation rate between elements (*33*). This view was confirmed by the studies of chromosomal evolution in element B and C of the *Sophophora* subgenus of *Drosophila* (*31*). In addition, the major difference in the rate of rearrangement between *Sophophora* and *Drosophila* may result from the differences in the polymorphism levels within species of these subgenera (*64, 134*). All these observations can be explained by the local adaptation model (*208*). In this model, the conditions in which an inversion can be spread were studied and they suggest that if an inversion carries a set of locally adapted alleles, the local adaptation mechanism will cause it to spread to fixation. The association of inversion polymorphisms with adaptive intraspecific variation in the *Anopheles* group has been found (*13, 421, 422*). Therefore, the local adaptation mechanism may be the major evolutionary forces for driving the polymorphism into fixation on autosomes of *Anopheles* mosquitoes.

6. **Conclusions**

1) A pattern of inversion fixation in *Anopheles* mosquitoes has been identified. The sex chromosome has the fastest rate of inversion fixation. Among autosomes, 2R evolved faster than other chromosomal arms.
2) The inversion breakpoints are not distributed randomly in *A. gambiae* and *A. stephensi*. The breakpoint clustering is common in *Anopheles* mosquitoes.
3) Repetitive elements such as TEs, AT and inverted repeats play an important role in the origin of inversions in the lineages of *A. gambiae* and *A. stephensi*.
4) We demonstrated that nuclear architectures contributed to the patterns of inversion fixation during the evolution of *A. gambiae* and *A. stephensi*.
5) Nonrandomly distributed polymorphic inversions can be driven to fixation in *Anopheles* mosquitoes.

Chapter four: The ancestral status of chromosomal inversions in the *Anopheles gambiae* complex

1. Abstract

The *A. gambiae* complex is comprised of seven closely related species which can be differentiated by ten fixed inversions. To understand the genetic changes associated with the distinct ecological adaptation, geographical distribution, and malaria transmission, the demonstration of the phylogenetic relationships among members of the *A. gambiae* complex is crucial. In this study, our computational and experimental analysis confirmed the ancestral status of the 2La inversion. The distal and proximal breakpoints of the 2La inversion have been mapped to chromosome 3L of *A. stephensi* and *A. nili*. Determination of ancestral status of 2Rop inversions, which differentiate *A. gambiae* and *A. merus,* was also attempted. Distal and proximal breakpoints of the alternative rearrangements have been determined in *A. gambiae* and *A. merus* using *in situ* hybridization results. The molecular features near the breakpoints of $2R+^{op}$ in the *A. gambiae* PEST strain were also analyzed. The computational analysis suggested that 2Rop typical for *A. merus* are ancestral inversions. Experimental evidence demonstrated that the breakpoint structure of 2Ro arrangement is present in outgroup species *A. stephensi*. The sequence analysis near breakpoints revealed that $2R+^{p}$ of *A. gambiae* is derived. Therefore, the carrier of 2Rop inversions, *A. merus*, was considered closest to the ancestral species.

2. Introduction

The Afrotropical *Anopheles gambiae* complex includes seven closely related species. Individual members of the *A. gambiae* complex have distinct ecological adaptations,

geographical distributions, and behaviors. The demonstration of the phylogenetic relationships among members of the *A. gambiae* complex is a crucial step toward understanding the genomic changes associated with the origin and loss of human blood choice, ecological and behavioral adaptation, and ability to support the development of a malaria parasite. Although reconstruction of the *A. gambiae* complex phylogeny using molecular markers (*222-225*) and fixed inversions, as well as polytene chromosome maps of outgroup species (*226*) have been attempted, the phylogenetic relationships among the members remain unsolved.

Despite the low level of genetic divergence, members of the *A. gambiae* complex can be differentiated by ten fixed inversions and molecular markers. The karyotype of *Anopheles* mosquitoes is comprised of five chromosomal arms: one pair of sex chromosomes: X (X and Y in male), and four autosomes: 2R, 2L, 3R and 3L. The notation for the standard karyotype is X+, 2R+, 2L+, 3R+, 3L+. *A. quadriannulatus* A and *A. quadriannulatus* B carry the standard chromosomal arrangements (*14*), and the other members of the complex have fixed inversions on various chromosomal arms. The 2La inversion is fixed in *A. arabiensis* and *A. merus*, but is polymorphic in *A. gambiae* (*14*) (*150*). *A. merus* and *A. gambiae s.s*. share the Xag inversion, while *A. arabiensis* has the Xbcd inversion (inverted b, c, d arrangements on X chromosome). Additionally, *A. merus* and *A. gambiae* differ from each other by two overlapping inversions on 2R, “o” and “p”. *A. bwambae* and *A. melas* share the 3La arrangement, while *A. melas* carries a 2Rm inversion (*14*).

Our study attempted to use physical maps of outgroup species *A. stephensi* and *A. funestus*, *A. nili*, as well as *A. moucheti* for inferring ancestral status of fixed inversions

in the *A. gambiae* complex. *A. moucheti* Evans, and *A. nili* Theobald are the major malaria vectors in Africa. *A. moucheti* is restricted to equatorial Africa, spreading from Guinea to Uganda and the south of Sudan. This species is a very efficient malaria vector: sporozoite rates are up to 4%, and the annual entomological inoculation rate (EIR) reaches 300 (*423*). *A. nili* has a wide geographic distribution, spreading across most of West and Central Africa, mainly in humid savannas areas. Sporozoite rates in *A. nili* can reach 3%, and the annual EIR can be over 100 (*424*). For example, in a village in Eastern Senegal, *A. nili* was responsible for 56 infected bites per human per year (*425*). The calculation of inversion distances among species was performed using computational programs. Inversion breakpoints in the outgroup species and the *A. gambiae* complex were determined by *in situ* hybridization results. The molecular organization of the 2Rop breakpoint regions in the standard and inverted arrangement were also analyzed and results shed light on the mechanisms of the origin of these inversions. Further study of the sequences near the breakpoints of ten fixed inversions will help to discover the important alleles responsible for epidemiologically important phenotypes.

3. Material and methods

3.1. Mosquito strains

A. stephensi, the Indian wild type laboratory strain, was used in this research. The female mosquitoes of *A. nili* and *A. moucheti* were collected from Africa by Dr. Antonio Nkondjio and Dr. Frederic Simard. The mosquitoes of *A. merus* were obtained from the Malaria Research and Reference Reagent Resource Center (MR4) (www.mr4.org).

3.2. Probe preparations and *in situ* hybridization

Thirty six conserved *A. gambiae* BAC and cDNA clones, as well as *A. funestus* cDNA clones were mapped to *A. stephensi* polytene chromosomes 2R and 3L using FISH (Fluorescent *In Situ* Hybridization) (**Table 4.2 and 4.4**). The probe preparation and *in situ* hybridization methods for these 36 markers were described in detail in Chapter one (Material and Methods). The locations of 36 previously mapped *A. funestus* cDNA and microsatellite markers on chromosomes 2R and 3R of *A. funestus* were obtained from earlier studies (*1, 26*). The sequences of *A. gambiae* cDNA clones were downloaded from Vectorbase (http://www.vectorbase.org/index.php). The primers were designed using Primer 3 software available at http://frodo.wi.mit.edu/primer3/ for each marker. Genomic DNA was isolated from single *A. gambiae Sua* strain mosquito or from fixed ovaries of *A. nili* or *A. moucheti* using Qiagen DNeasy Blood & Tissue Kit. The PCR conditions were the following: 95℃ for 4 min; 35 cycles of 94℃ for 30 s, 55℃ for 30 s, 72℃ for 30 s; 72℃ for 5 min. And then all the PCR products were purified from agarose gel using the GENECLEAN III kit (MP Biomedicals). Purified DNA was labeled with Cy5-AP3-dUTP and Cy3-AP3-dUTP (GE Healthcare UK Ltd, Buckinghamshire, England) using a modified Random Primers DNA Labeling System (Invitrogen Corporation, Carlsbad, CA, USA). *In situ* hybridization was performed as described in Chapter one.

3.3. Computational methods and analysis

Analysis included 36 probes located near the 2La and 2Rop inversion breakpoints in the *A. gambiae* genome and on the *A. stephensi* and *A. funestus* chromosomes (*1, 26*). Locations of DNA sequences in the *A. gambiae* genome were determined by BLASTN with default parameters using the VectorBase website

(http://www.vectorbase.org/Tools/BLAST/#). We considered only hits with e-values less than e^{-25} and alignments longer than 100 nt for ESTs and with e-values no more than e^{-5} and alignments longer than 40 nt for noncoding sequences. Chromosomal locations of the *A. gambiae* BAC clones were found using the search option on the VectorBase website (http://www.vectorbase.org/Search/Keyword/).

The calculation of inversion distances among species of the *A. gambiae* complex, as well as those for *A. funestus* and *A. stephensi,* were performed using the Multiple Genome Rearrangements (MGR) and Sorting Permutation by Reversals and Block-InterchanGes (SPRING) programs. The MGR program is available at www.cs.ucsd.edu/groups/bioinformatics/MGR. We used the signed option of the MGR program when the gene directions were known. This program implements an algorithm which seeks a tree that minimizes the sum of the rearrangements over all the edges of the tree (*383*). The SPRING program is available at http://algorithm.cs.nthu.edu.tw/tools/SPRING/index.php. SPRING computes both the breakpoint and rearrangement distances between any pair of two chromosomes (*426*). It also shows phylogenetic trees that are reconstructed based on the rearrangement and breakpoint distance matrixes. The algorithms of MGR and SPRING are different. MGR uses heuristic strategies to reconstruct a phylogenetic tree of input species. SPRING uses the Neighbor-Joining method to reconstruct a tree. We have chosen these two methods for our analysis because they use gene order information as opposed to nucleotide sequences. For our analyses, we used trees produced by MGR and SPRING based on rearrangement distances.

4. Results

4.1. Inversion distance in the *A. gambiae* complex

There are ten fixed inversions in the *A. gambiae* complex (*13*). Although the exact location of each breakpoint is still not clear, the positions of ten fixed inversions on the five chromosome arms were indicated by Coluzzi et al (*13*) (**Figure 4.1**). Members of the *A. gambiae* complex can be differentiated by ten fixed inversions, and the phylogenetic relationships based on the chromosomal inversions among the species were shown in **Figure 4.2**. In order to test the computational programs, MGR and SPRING, the inversion distances among the members were calculated using these programs and then the results were compared to **Figure 4.2**. According to previous publications, the banding patterns of polytene chromosomes can be used as genetic markers, in which bands and interbands are considered as alleles (*427, 428*). Based on the information on position of the inversion breakpoints provided by Coluzzi et al (*13*), 30 cytogenetic markers were selected to analyze the inversion distance in the *A. gambiae* complex using the MGR and SPRING programs. These markers are the patterns of chromosome bands and interbands associated with breakpoints. The following gene orders were input into the MGR and SPRING program.

>*A. gambiae* 2R+ 2La 3L+ Xag
1 2 3 4 5 6 7 8 9 10 11 12 $
13 14 $
15 16 $
-28 -27 -26 -25 -24 -23 -22 -21 -20 -19 17 18 29 30 $
>*A. merus* 2Rop 2La 3L+ Xag
1 -7 -6 -5 -4 -9 -8 2 3 10 11 12 $
13 14 $
15 16 $
-28 -27 -26 -25 -24 -23 -22 -21 -20 -19 17 18 29 30 $
>*A. arabiensis* 2R+ 2La 3L+ Xbcd

1 2 3 4 5 6 7 8 9 10 11 12 $
13 14 $
15 16 $
17 18 19 20 -26 -25 -30 -29 -28 -27 21 22 -24 -23 $
>*A. melas* 2Rm 2L+ 3La X+
1 2 3 4 5 -11 -10 -9 -8 -7 -6 12 $
13 -14 $
15 -16 $
17 18 19 20 21 22 23 24 25 26 27 28 29 30 $
>*A. quadriannulatus* (species A and B) 2R+ 2L+ 3L+ X+
1 2 3 4 5 6 7 8 9 10 11 12 $
13 -14 $
15 16 $
17 18 19 20 21 22 23 24 25 26 27 28 29 30 $
>*A. bwambae* 2R+ 2L+ 3La X+
1 2 3 4 5 6 7 8 9 10 11 12 $
13 -14 $
15 -16 $
17 18 19 20 21 22 23 24 25 26 27 28 29 30 $

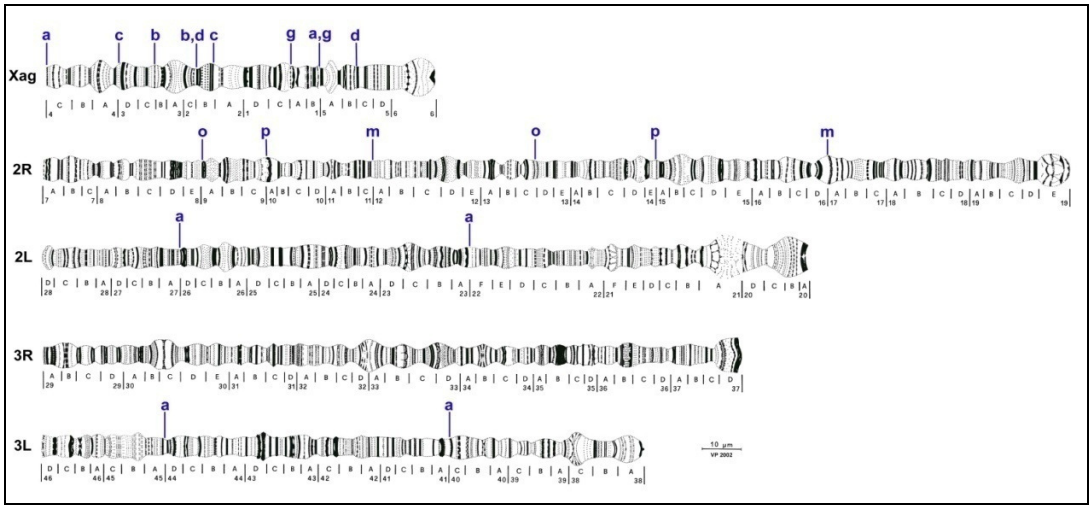


Figure 4.1: Chromosome distribution of the 10 fixed inversions in the *A. gambiae* complex. This figure is taken from (*13*).

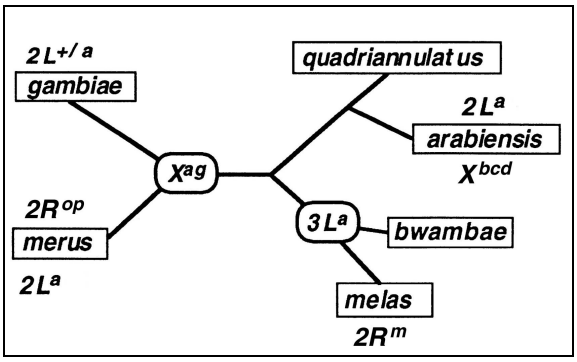


Figure 4.2: Inversion phylogeny of the *A. gambiae* complex species showing the ten fixed inversions. The phylogeny assumes an introgression of the 2La inversion from *A. arabiensis* to *A. gambiae*. The figure is modified from (*235*).

Total distances among the species of the *A. gambiae* complex calculated by MGR and SPRING were given in **Table 4.1**. These data show that *A. quadriannulatus* has the smallest sum of total distances to other species (15) and it has the central position in the complex as related to other species. For this reason, this species was regarded as the closest to the ancestral species (*13, 42*). The phylogenetic trees generated by MGR and SPRING in **Figure 4.3** are in agreement with **Figure 4.2**. And both MGR and SPRING programs successfully recovered all 10 fixed inversions. In addition, inversion distances among all species were identified by both of computational programs (**Figure 4.3**). Therefore, the phylogenetic trees produced by these programs can serve as working hypotheses for determining phylogenetic relations in the complex.

Table 4.1: An MGR pairwise distance matrix of the input genomes showing the sums of total distances among species **(*371*)**.

	A. gambiae	***A. arabiensis***	***A. merus***	***A. melas***	***A. quadriannulatus***	***A. bwambae***
A. gambiae	0	5	2	5	3	4
A. arabiensis	5	0	7	6	4	5
A. merus	2	7	0	7	5	6
A. melas	5	6	7	0	2	1
A. quadriannulatus	3	4	5	2	0	1
A. bwambae	4	5	6	1	1	0

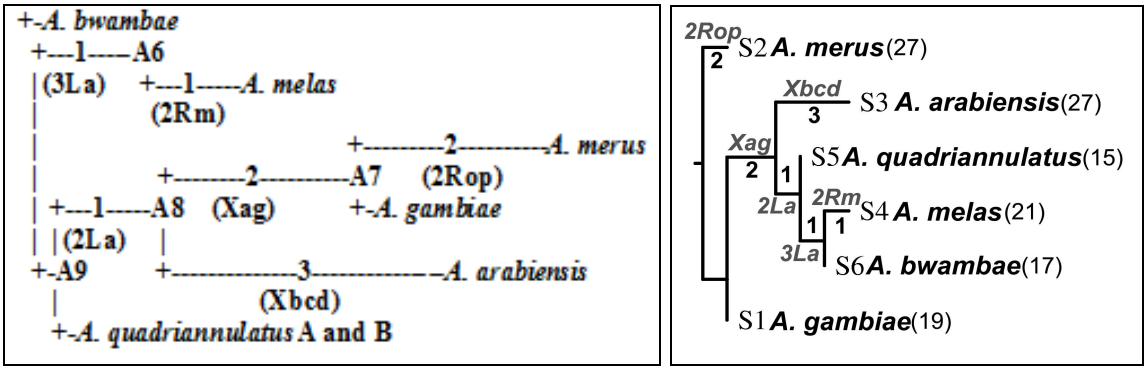


Figure 4.3 (*371*): (A) An unrooted tree of the *A. gambiae* complex recovered by MGR program. The number of rearrangements that occurred on each edge is shown. The names of fixed inversions are shown in parentheses. (B) The SPRING phylogenetic tree corresponding to the rearrangement distance matrix. The names and numbers of fixed inversions are shown at the branches. A6, A7, A8 and A9 are putative ancestral species.

The SPRING program forced the tree to be a rooted one. The strategy is based on an assumption that a species that is far from others, i.e., has a bigger sum of total distances to other species, can be regarded as a root. However, the members of the same complex are too close to each other for this assumption to be applicable.

4.2. Ancestral status of the 2La inversion

To confirm the ancestral status of 2La inversion in the *A. gambiae* complex, outgroup

species, *A. stephensi* and *A. funestus,* were used to reconstruct the phylogenetic relationship among members of the *A. gambiae* complex using MGR and SPRING program. The *A. gambiae* and *A. funestus* lineages diverged from a common ancestor at least 36 million years ago (*217*) and *A. gambiae* is more closely related to *A. stephensi* than to *A. funestus* (*387, 388*). Additional *in situ* hybridization data from *A. stephensi* and *A. nili* were also used to compare the gene orders with *A. gambiae*. Our results strongly suggest that the both of 2La inversion breakpoints are present in the genomes of *A. stephensi* and *A. nili*.

4.2.1. Computational analysis for the ancestral status of the 2La inversion

To reconstruct the phylogenetic relationships within the *A. gambiae* complex using outgroup species *A. stephensi* and *A. funestus*, a physical map for *A. stephensi* has been developed and the chromosomal locations of markers on the polytene chromosomes of *A. funestus* were acquired from previous publications (*1, 26*). In subgenus *Cellia*, because of the reciprocal whole arm translocation, the 2L arm in the *A. gambiae* complex corresponds to the 3R arm of *A. funestus* and the 3L arm of *A. stephensi* (*1, 18*). Fourteen conserved *A. gambiae* and *A. funestus* cDNAs and *A. gambiae* BAC clones were *in situ* hybridized to the polytene chromosomes of *A. stephensi*. Total 14 *A. funestus* cDNA clones and microsatellite markers were mapped to 3R chromosome of *A. funestus* before (*1, 26*). The coordinates in the *A. gambiae* genome sequence were obtained by BLASTN with default parameters using the VectorBase website (http://www.vectorbase.org/Tools/BLAST/). **Table 4.2** shows a list of DNA probes mapped to the chromosomes of *A. gambiae*, *A. funestus* and *A. stephensi*.

Table 4.2: Genomic and cytological locations of DNA probes mapped to chromosomes 2L in *A. gambiae*, 3R in *A. funestus*, and 3L in *A. stephensi*.

Probes on 2L of *A. gambiae*					**Chromosomal location**		
#	**Probe**	**Accession**	**Ensembl Gene**	**e-value**	***A. gambiae***	***A. funestus***	***A. stephensi***
1	28_C07	BU038985	ENSANGG00000018809	4e-67	2L:20C	3R:35C	NA
	101C3	BH388218			2L:20C	NA	3L:38B
2	04_D07	BU038878	ENSANGESTG00000001599	8e-30	2L:21A	3R:35A	NA
	02A19	AL140406			2L:21A	NA	3L:39D
3	95_H01	BU039015	ENSANGESTG00000004105	4e-27	2L:22D	3R:31D	NA
	27O10	AL154432				NA	3L:42C
4	AFND19	AF171049	ENSANGG00000003474	2e-62	2L:22F	3R:34A	NA
	131F22	BH390198				NA	3L:44C
5	11_F09	BU038906	ENSANGESTG00000005786	1e-100	2L:26B	3R:35F	3L:45C
6	09_C11	BU038897	ENSANGESTG00000002349	1e-63	2L:26A	3R:36C	3L:43C
7	16_F07	BU038931	ENSANGG00000023866 ENSANGG00000009052	1e-129 4e-57	2L:25D-26A* 2L:25A	3R:36E	3L:42A
8	21_E03	BU038955	ENSANGESTG00000007040	2e-35	2L:24C	3R:35F	NA
	140N16	BH384642			2L:24C	NA	3L:39C
9	61_E02	BU039003	ENSANGESTG00000000416	9e-86	2L:23D	3R:30C	3L:45A
10	36_A10	BU038993	ENSANGESTG00000000773	8e-25	2L:23C	3R:35F	NA
	150F12	BH385494			2L:23C	NA	3L:44A
11	66_E11	BU038987	ENSANGESTG00000002884	1e-134	2L:23A	3R:33D	NA
	716320	BM606621	ENSANGESTG00000002884		2L:23A	NA	3L:40A
12	08_B09	BU038894	ENSANGESTG00000002208	1e-48	2L:26D	3R:30C	NA
	04C08	AL607764			2L:26D	NA	3L:45A
13	06_G08	BU038889	ENSANGESTG00000006614	5e-56	2L:27A	3R:30C	3L:46D
14	18_G01	BU038941	ENSANGESTG00000008141	7e-69	2L:28C	3R:29B	3L:46A

NA–not available. Asterisks denote a major signal. #–number of markers.

These probes were determined by their chromosomal locations to the inversion breakpoints of 2La. Only markers in close proximity to the inversion breakpoints in the *A. gambiae* genome and those yielded unique or major signal in *A. funestus* and *A. stephensi* were used in this study. We assumed that ancestral arrangements should be preserved in outgroup species and thus gene orders around the breakpoints are conserved among these species. For each breakpoint, one or several probes at each

side were chosen. If a chromosomal site had at least two probes it was possible to determine its sign or orientation.

The inversion distances were calculated by the MGR and SPRING programs among the *A. merus* 2La chromosome, the *A. quadriannulatus* 2L+ arm, the *A. funestus* 3R, and the *A. stephensi* 3L chromosome. Fourteen uniquely located clones were used and the following gene orders were input.

>*A. bwambae, A. melas, A. quadriannulatus* 2L+
1 2 3 4 $
>*A. arabiensis, A. gambiae, A. merus* 2La
1 -3 -2 4 $
>*A. stephensi* 3L
-2 4 ±1 3 $
>*A. funestus* 3R
±3 -1 -2 4 $

Figure 4.4 shows that both MGR pairwise distance matrix and SPRING rearrangement matrix determined that the inversion distance between 2La and outgroup chromosomes is one step shorter than the inversion distance between 2L+ and the outgroup chromosomes. Therefore, computational result reveals that 2La inversion is ancestral inversion which is good consistent with previously published work (*164, 221*).

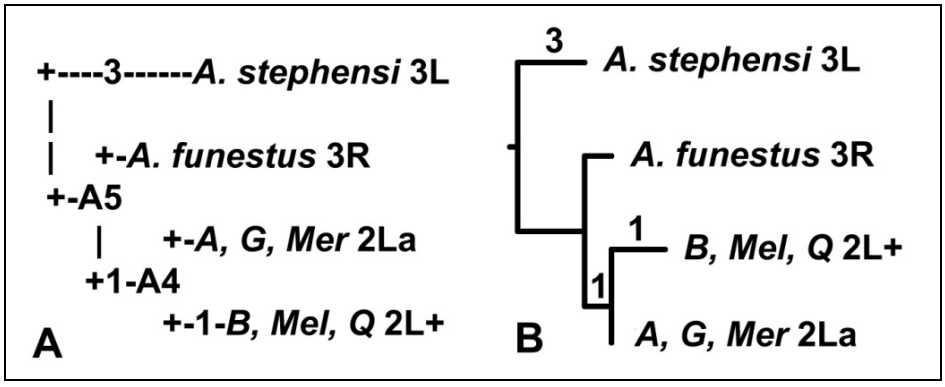


Figure 4.4 (from (*371*): Trees recovered by MGR (A) and SPRING (B) programs showing that the 2La has smaller distance to outgroup chromosomes than 2L+. The

number of rearrangements that occurred on each edge is shown. A4 and A5 are putative ancestral species.

4.2.2. Experimental evidence for the ancestral status of the 2La inversion

Hybridization results of BAC clones, and cDNAs from *A. gambiae* confirmed that the inverted 2La rearrangement is present in *A. stephensi*. The 146D17 BAC clone was found spanning the 2L+/+ proximal breakpoint in *A. gambiae* (*164*) and was hybridized to two locations, 40A and 44C, on the chromosome 3L of *A. stephensi*. The additional BAC clone: 131F22 (**Table 4.3**) that partly overlaps with 146D17 at one side of the breakpoint was mapped to a single location, 44C on *A. stephensi* 3L chromosome (**Figure 4.5** and **Table 4.3**). Follow-up experiments involving fragment: the SuaPh6_1.8EcoRI fragment (**Table 4.3**), which is homologous to another side of the breakpoint (within 146D17) yielded only single site: the subdivision 40A to *A. stephensi* chromosome. Therefore, these results indicate that the breakpoint structure of the 2La arrangement is present in the outgroup species *A. stephensi* and, therefore, is more likely ancestral.

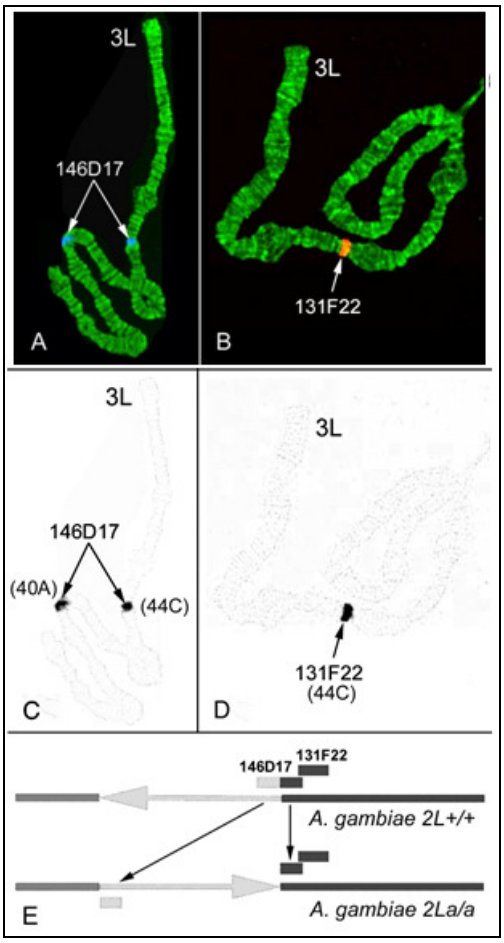


Figure 4.5 (from (*371*)): FISH of 146D17 labeled with Cy5 (A, C) and 131F22 labeled with Cy3 (B, D) performed on the chromosomes of *A. stephensi*. Arrows point at the hybridization signals. A and B show the banding pattern of the chromosomes counterstained with the fluorophore YOYO-1. C and D show fluorescence due to hybridization. The images are inverted. (E) The scheme showing the location of the BAC clones on the *A. gambiae* 2L+/+ and 2La/a maps. The centromeres are on the right side.

Table 4.3: The localizations of cDNA and BAC clones near the 2La breakpoints in *A. stephensi* and the coordinated in *A. gambiae*

#	Probe	Accession	Coordinates in *A. gambiae*	*A. gambiae*	*A. stephensi*
1	131F22	BH390198	20,364,135 - 20,459,325	2L:22F	3L:44C
2	146D17	BH400736	20,451,607 - 20,591,927	2L:22F	3L:40A,44C
3	SuaPh6_1.8EcoRI	NA	20,535,740 - 20,538,254	2L:23A	3L:40A

NA: not available.

To determine the distal and proximal breakpoints of 2La inversion in *A. stephensi* genome and the presence of 2La in *A. nili*, 52 markers were hybridized to chromosome 3L of *A. stephensi* and *A. nili*. The results were summarized in **Table 5.1 of Chapter five**. The gene order comparison was shown in **Figure 5.5 of Chapter five**. In *A. gambiae* PEST strain (standard 2L), distal breakpoint of 2L is located between Ag5778 and Ag5779, proximal is located in Ag7068-Ag7069. The gene orders span and near breakpoints in *A. stephensi* are same as those of 2La in *A. gambiae* which shows that 2La inversion is present in *A. stephensi*. *A. nili* shared the same distal breakpoint structure of 2La inversion but gene orders near proximal breakpoints are different from 2La inversion in *A. gambiae*. This can be explained by the occurrence of two more inversions after divergence of *A. gambiae* and *A. nili* **(Appendix 4.1)**. If we restore the phylogenetic history back to the status of *A. nili* by reinverting two rearrangements after it diverged from ancestral species, the gene orders near both of breakpoints in *A. nili* are the same as those of 2La in *A. gambiae* (**Appendix 4.2**). This result suggests that *A. nili* and *A. stephensi* carry 2La breakpoint structures. Therefore, our data confirmed that 2La is an ancestral inversion and the carriers of 2La inversion are considered cloestest to the ancestral species.

4.3. The ancestral state of the 2Rop inversion

4.3.1. Computational analysis for the ancestral status of the 2Rop inversions

A. merus and *A. gambiae* differ from each other by two overlapping inversions "o" and "p" on 2R arm. Twenty two uniquely located markers that were common for the *A. funestus* and *A. stephensi* maps were used to run MGR and SPRING programs. Of them, the location of 11_D03 in *A. stephensi* was established in our previous study (*18*). The locations of 22 markers in *A. funestus* were obtained from previous data (*1, 26*).The genome coordinates in the *A. gambiae* genome sequence were obtained by BLASTN with default parameters using the vectorbase website (http://www.vectorbase.org/Tools/BLAST/). **Table 4.4** shows a list of DNA probes mapped to the chromosome 2R in *A. gambiae*, *A. funestus* and *A. stephensi*.

Table 4.4: Genomic and cytological locations of DNA probes mapped to 2R chromosome in *A. gambiae*, *A. funestus*, and *A. stephensi*.

Probes on 2R of *A. gambiae*					**Chromosomal location**		
	Probe	**Accession**	**Ensembl Gene**	**e-value**	***A. gambiae***	***A. funestus***	***A. stephensi***
1	21_F03	BU038956	ENSANGG00000027869	1e-23	2R:7A	2R:7B	NA
	04L11	AL141975			2R:7A	NA	2R:7A
2	01_H04	BU038873	ENSANGESTG00000007874	1e-165	2R:7B	2R:7A	2R:7B
3	21_F12	BU038958	ENSANGESTG00000001934	2e-69	2R:8A	2R:12D	2R:13A
4	36_B06	BU038996	ENSANGESTG00000008316	3e-80	2R:8D	2R:8E	NA
	105H10	BH368219			2R:8D	NA	2R:9C
5	12_G10	BU038913	ENSANGESTG00000009208	1e-105	2R:8E	2R:15C	2R:8A
6	AFND5	AF171035	ENSANGESTG00000008727	3e-18	2R:9B	2R:15B	NA
	25P09	AL153306			2R:9B	NA	2R:10D
7	FUN O	AY116019		3e-05	2R:9C	2R:18A	NA
	11A13	AL145719			2R:9C	NA	2R:14B
8	11_D03	BU038903	ENSANGESTG00000003457	1e-60	2R:10A	2R:9A	2R:10A
9	04_D06	BU038877	ENSANGESTG00000008689	6e-35	2R:11A	2R:10C	2R:16AB
10	08_E06	BU038895	ENSANGESTG00000007439	4e-61	2R:11C	2R:16A	2R:10D
11	25_E09	BU038972	ENSANGESTG00000008987	3e-83	2R:12B	2R:12B	2R:18B
12	13_F11	BU038919	ENSANGG00000027321	6e-57	2R:12B	2R:12B	NA
	129M18	BH377340			2R:12B	NA	2R:18B

13	15_F10	BU038925	ENSANGESTG00000008970	9e-65	2R:12B	2R:12B	NA
	626240	BM655548	ENSANGESTG00000008970	1e-136	2R:12B	NA	2R:18B
14	06_B01	BU038882	ENSANGESTG00000004771	1e-90	2R:12D	2R:14D	2R:9A
15	03_D09	BU038874	ENSANGESTG00000007166	5e-27	2R:13E	2R:17C	NA
	31M01	AL611707			2R:13E	NA	2R:8C
16	12_G11	BU038914	ENSANGESTG00000005173	3e-40	2R:15B	2R:18C	2R:14C
17	12_H09	BU038915	ENSANGG00000024830	5e-82	2R:15D	2R:18D	2R:11C
18	29_F03	BU038988	ENSANGESTG00000000136	5e-48	2R:15D	2R:11B	NA
	169F11	BH369697			2R:15D	NA	2R:19A
19	11_E07	BU038905	ENSANGG00000017799	1e-131	2R:16A	2R:14C	2R:19A
20	13_C03	BU038918	ENSANGG00000011859	3e-55	2R:17C	2R:13C	NA
	08O05	AL144514				NA	2R:17A
21	18_D12	BU038940	ENSANGESTG00000004138	7e-71	2R:18C	2R:14B	2R:12C
22	11_B04	BU038900	ENSANGESTG00000009273	1e-132	2R:19C	2R:19C	2R:19BC

NA–not available. Asterisks denote a major signal. #–number of markers.

The following gene orders for these arms were input into the GRIMM and SPRING programs.

>*A. gambiae* 2R+
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 $
>*A. merus* 2Rop
1 2 3 4 5 -14 -13 -12 -11 -10 -9 -8 -15 6 7 16 17 18 19 20 21 22 $
>*A. funestus* 2R
2 1 4 8 9 18 11 12 13 3 20 21 19 14 6 5 10 15 7 16 17 22 $
>*A. stephensi* 2R
1 2 5 15 14 4 8 10 6 17 21 3 7 16 9 20 11 12 13 18 19 22 $

Both MGR pairwise distance matrix and SPRING rearrangement matrix determined that the inversion distance between 2Rop and the outgroup chromosome is two steps shorter than the distance between 2R+ and the outgroup 2R (**Figure 4.6**). Therefore, computational programs suggest that the 2Rop are ancestral inversions.

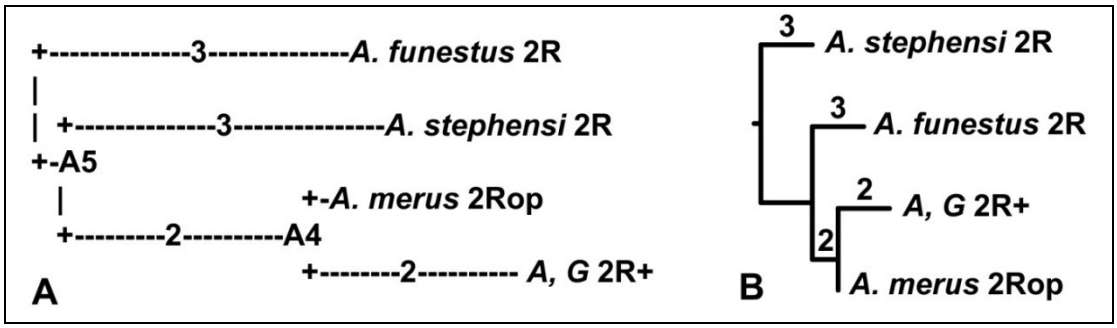


Figure 4.6: Trees recovered by MGR (A) and SPRING (B) programs showing that the 2Rop has smaller distance to *A. stephensi* and *A. funestus* 2R than 2R+. The number of rearrangements that occurred on each edge is shown. A4 and A5 are putative ancestral species **(*371*)**.

4.3.2. Experimental evidence for the ancestral status of the 2Rop inversions

4.3.2.1. The determination of the 2Rop breakpoints in *A. gambiae* and *A. merus*

Thirty four cDNA clones from *A. gambiae* were hybridized to the chromosomes of *A. merus.* The summary of hybridization results were listed in **Appendix 4.4** and the chromosomal localizations of markers were shown in **Appendix 4.3**.

The gene order comparison between *A. gambiae* and *A. merus* (**Figure 4.7**) shows that the distal 2R+^{o} breakpoint is located in between cDNA markers, Ag1759 and Ag1763, the proximal 2R+^{o} breakpoint is in between Ag2934 and Ag2935. The distal and proximal breakpoints of 2R+^{p} in *A. gambiae* are located in between Ag1983 and Ag1984 and between Ag3327 and Ag3328.

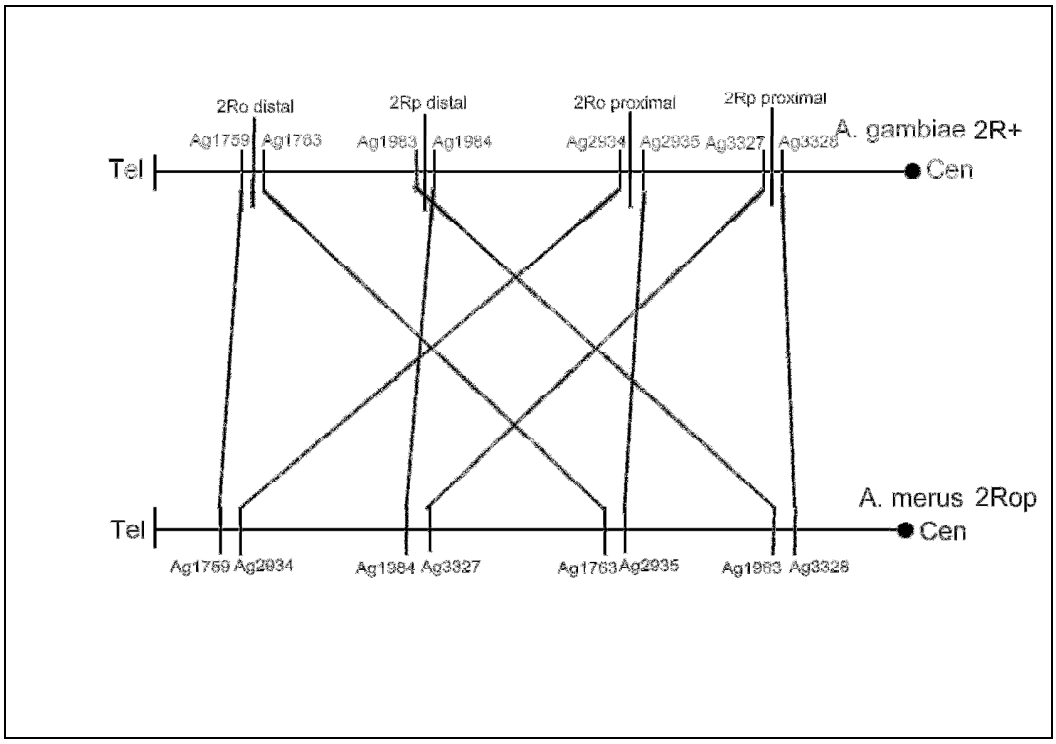


Figure 4.7: Gene orders in *A. gambiae* and *A. merus*. Short black lines indicate the breakpoints.

4.3.2.2. The ancestral status of the 2Ro inversion in the *A. gambiae* complex

Among the BAC clones used for the construction of a physical map of *A. stephensis*, one BAC clone: *A. gambiae* BAC 141A14 yielded one single hybridization singnal for *A. gambiae* on 13D which is located near the proximal breakpoint of Ro inversion and was hybridized to 11A and 12B in *A. stephensi* and to 9A (**Figure 4.8**). 141A14 were also hybridized to the chromosome sites: 8E and 9A on *A. merus* 2R arm (**Figure 4.10**). Therefore, 141A14 might locate on the chromosome site which may span the proximal breakpoint of 2R+^{o} inversion and *A. stephensi* carries the 2Ro inversion.

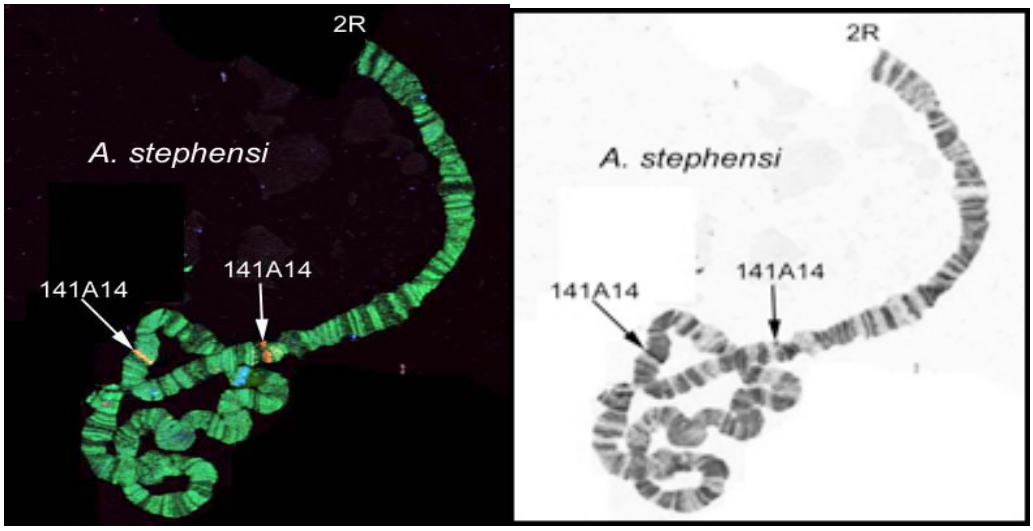


Figure 4.8: Fluorescent i*n situ* hybridization of 141A14 labeled with with Cy3 performed on the chromosomes of *A. stephensi*. Arrows point at the hybridization signals. Left panel: colored image; Right panel: Gray inverted.

Hybridization of cDNA markers within the BAC clone revealed that two closely located clones, Ag2934 and Ag2935, in *A. gambiae* were localized to different chromosomal sites in *A. stephensi* (**Figure 4.9**) and *A. merus* (**Figure 4.10-4.12**). Ag2933 was mapped to both chromosomal sites which indicate that one pseudogene copy is in the proximal breakpoint of 2Ro in *A. merus*. More markers near both breakpoints of 2Ro were hybridized to the *A. stephensi* 2R arm (**Table 4.5**). **Figure 4.13** show that *A. stephensi* shares 2Ro breakpoints with *A. merus*.

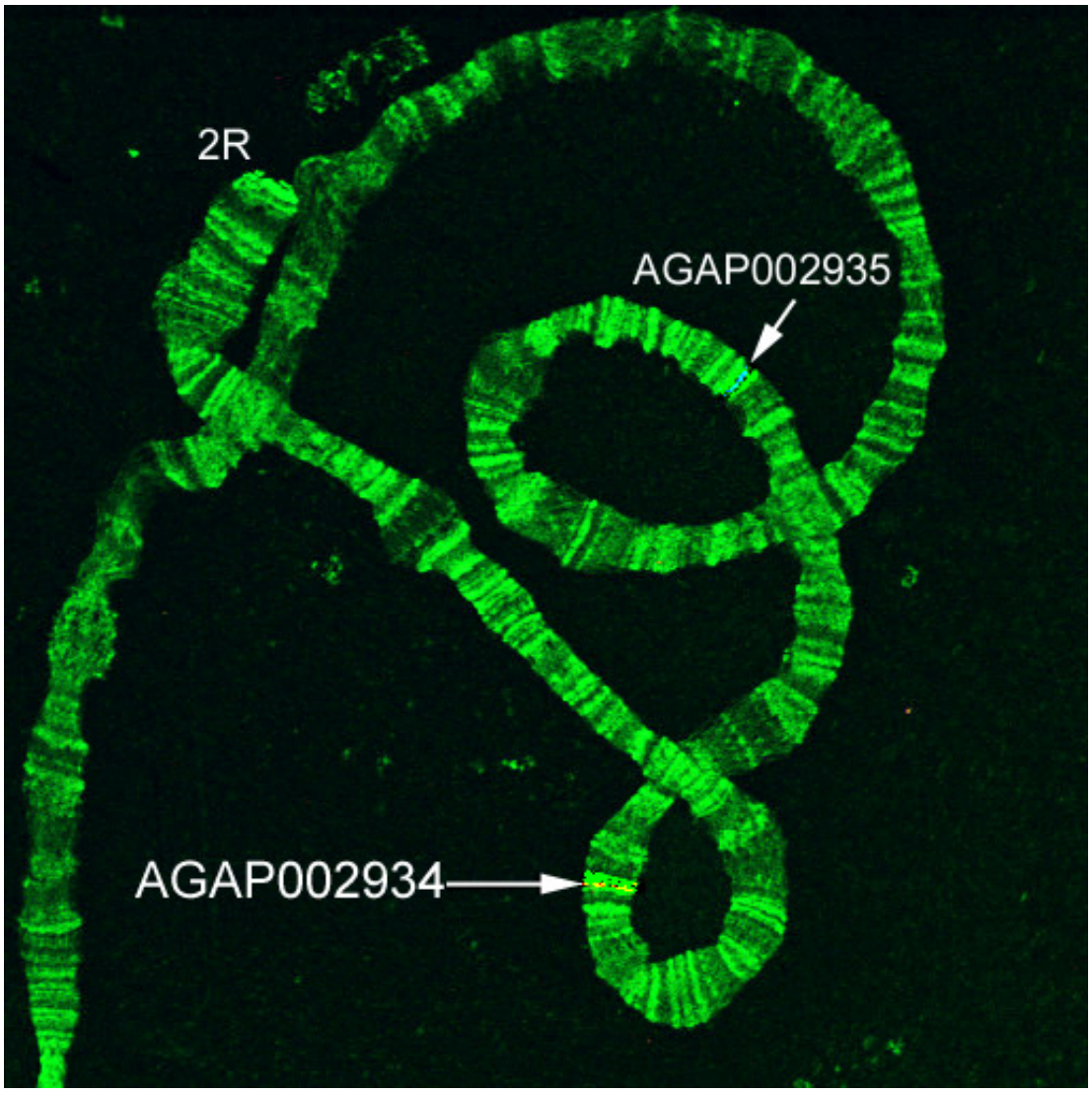


Figure 4.9: Fluorescent i*n situ* hybridization of AGAP002934 labeled with Cy3 and AGAP002935 labeled with performed on the chromosomes of *A. stephensi*. Arrows point at the hybridization signals.

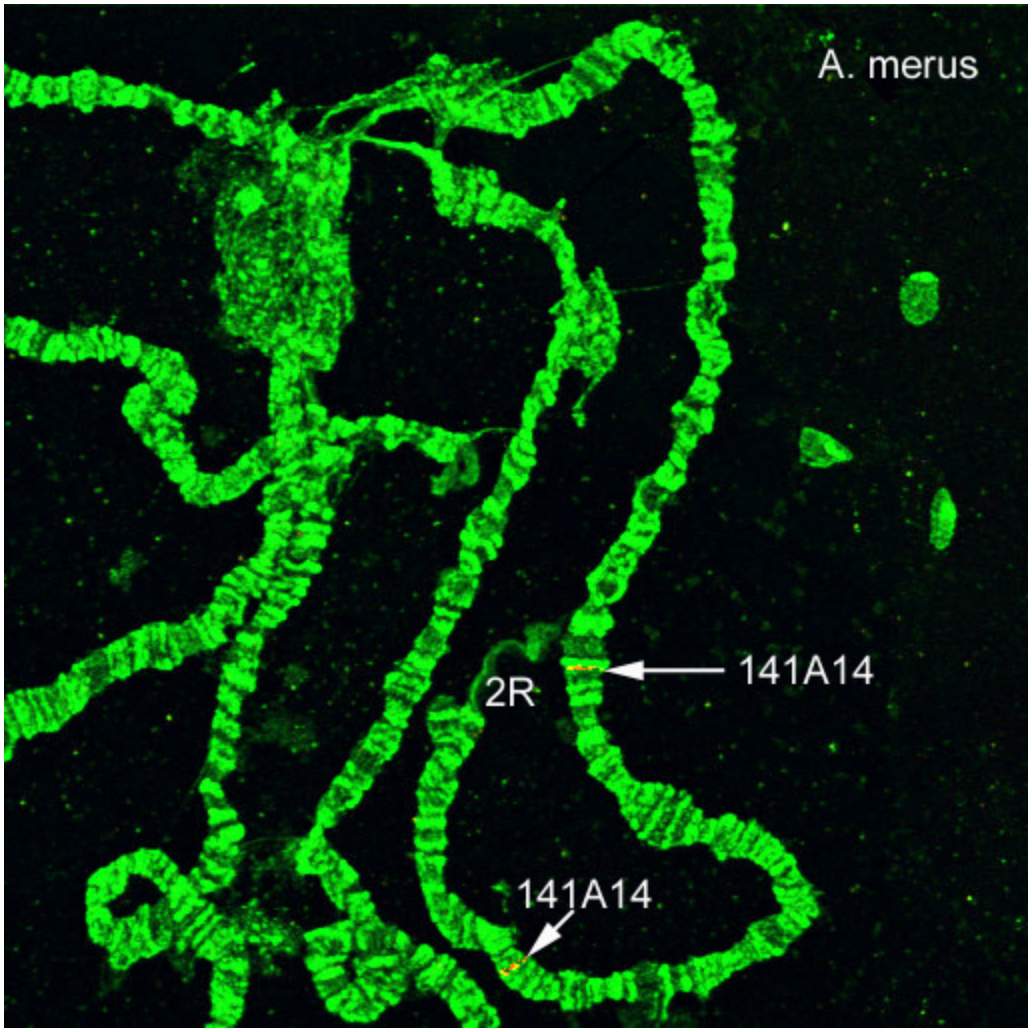


Figure 4.10: Fluorescent i*n situ* hybridization of 141A14 labeled with with Cy3 performed on the chromosomes of *A.merus.*

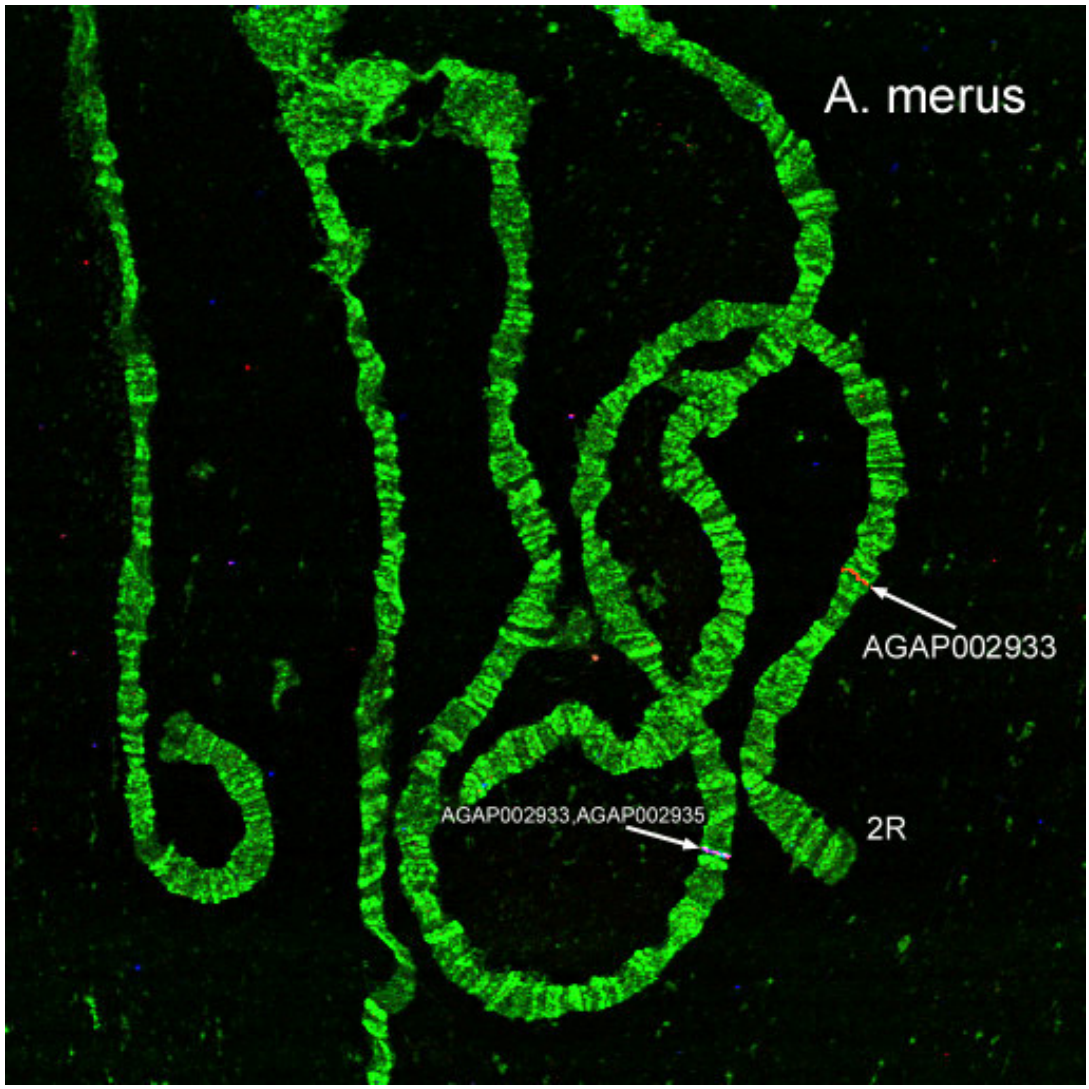


Figure 4.11: Fluorescent i*n situ* hybridization of AGAP002933 labeled with Cy3 and AGAP002935 labeled with Cy5 performed on the chromosomes of *A. merus*.

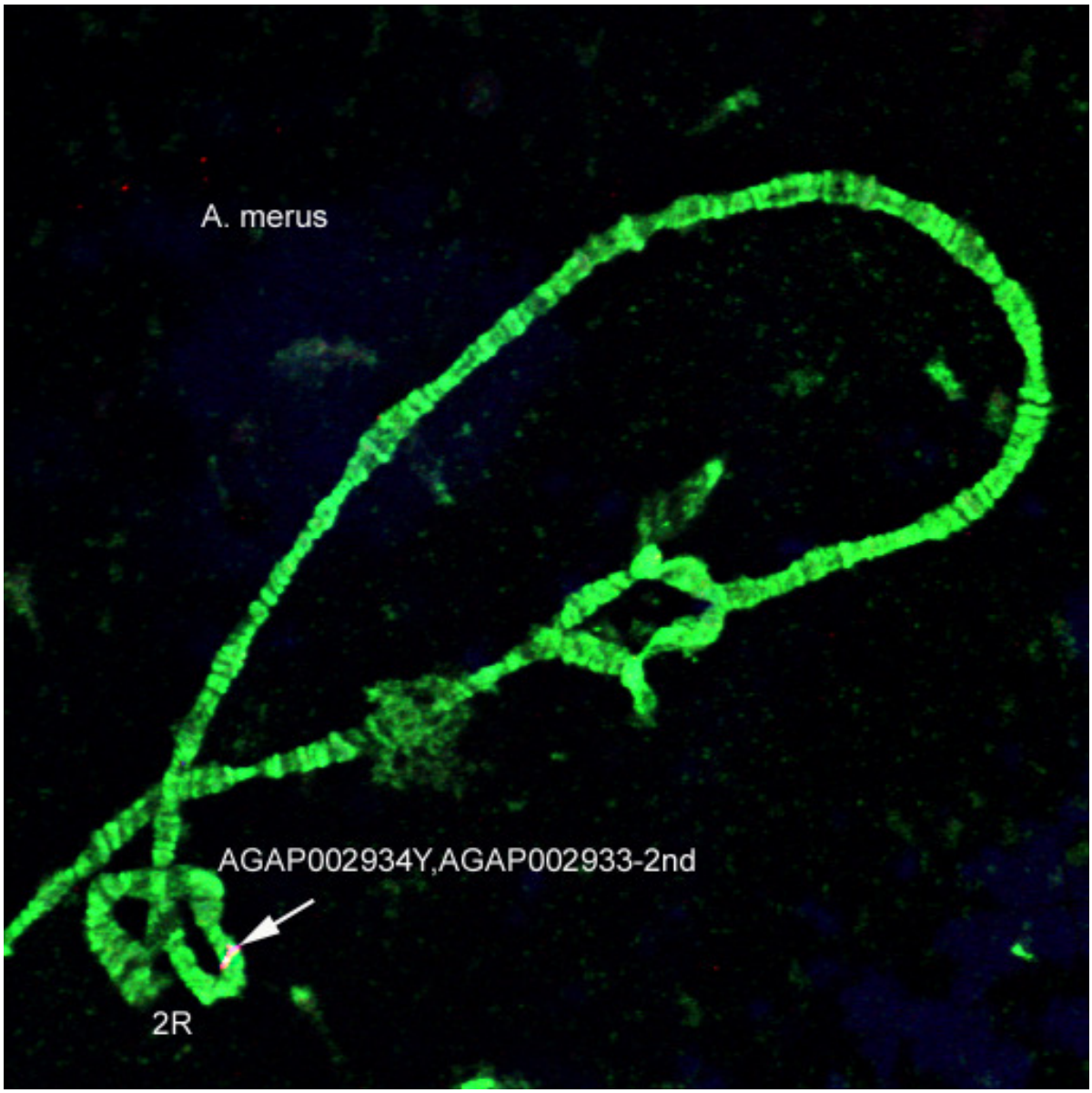


Figure 4.12: Fluorescent i*n situ* hybridization of AGAP002933-2nd labeled with Cy3 and AGAP002934 labeled with Cy5 performed on the chromosomes of *A. merus*. Arrows indicated the hybridization signals.

Table 4.5: The localizations of probes near the 2Ro breakpoints in A. *stephensi*, *A. merus* and the coordinates in *A. gambiae*

	clone	Accession	Genomic location in *A. gambiae*	e value	*A. gambiae*	*A. merus*	*A. stephensi*
1	04A11	AL141561	9,422,797–9,524,853	0	2R:8E	nd	2R:11A
2	Ag1759	AGAP001759	9,479,549-9,483,291	0	2R:9A	2R:8E	2R:11A
3	Ag1763	AGAP001763	9,523,856-9,528,904	0	2R:9A	2R:9A	2R:16A
4	10E06	AL145314	29,490,903-29,598,621	0	2R:13D	nd	2R:11A,10D
5	AsRPS6	AY237124	29,609,284–29,611,282	0	2R:13C	nd	2R:11A
6	Ag2934	AGAP002934	29,835,569–29,836,999	0	2R:13C	2R:8E	2R:11A
7	Ag2935	AGAP002935	29,839,388–29,840,621	0	2R: 13C	2R:9A	2R:16A
8	27I24	AL154218	30,150,936-30,271,431	0	2R:13D	nd	2R:16A

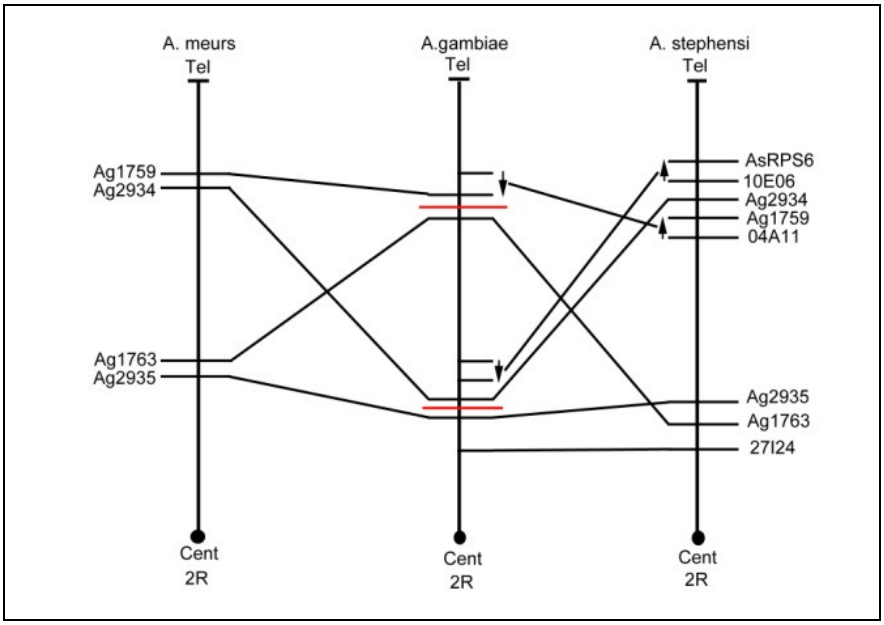


Figure 4.13: Gene orders in *A. gambiae*, *A. stephensi* and *A. merus*. Red lines indicate the possible 2R+^{o} inversion breakpoints.

In situ hybridization of cDNA marker also shows that Ag2934Y and Ag2935, Ag1759 and Ag1763 were hybridized to the different chromosomal sites on *A. moucheti* polytene chromosomes (**Figure 4.14**). However the lack of good polytene chromosomes in *A. moucheti* hinders determination of the gene orders for these markers.

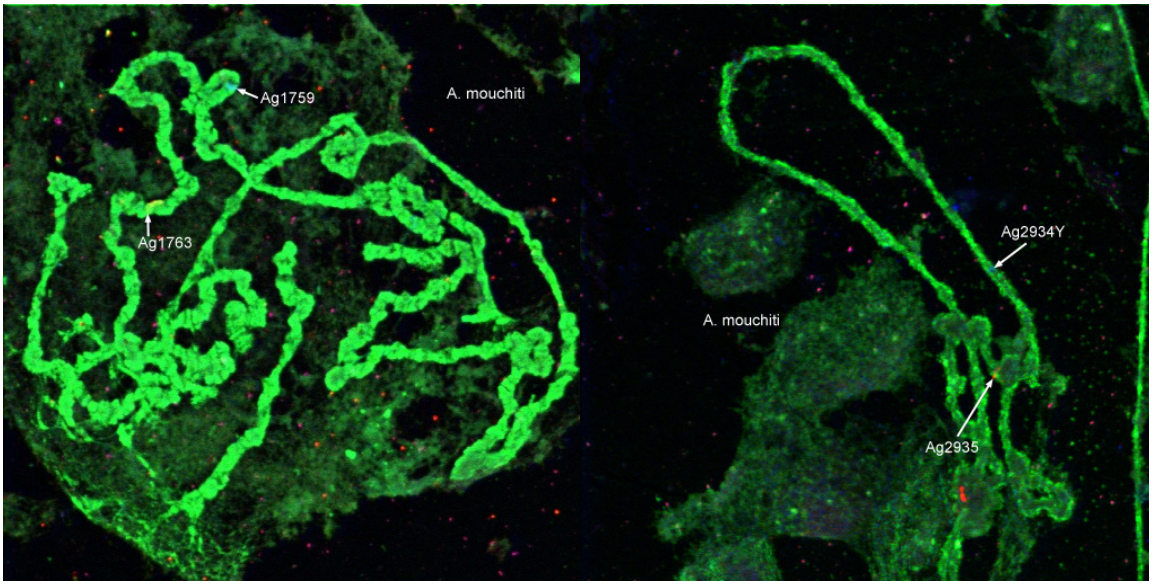


Figure 4.14: (Left) In situ hybridization of Ag1759 and Ag1763 on *A. moucheti.* (Right) In situ hybridization of Ag2934Y and Ag2935 on *A. moucheti*

4.3.2.3. The ancestral status of the 2Rp inversion

The comparative gene orders between *A. gambiae* and *A. merus* revealed that the proximal breakpoint of $2R+^{p}$ is located in between Ag3327 and Ag3328, distal is between Ag1983 and Ag1984 in *A. gambiae*. 12 markers near the breakpoints were hybridized to the polytene chromosome 2R of *A. stephensi* (**Table 4.6**). The gene order comparison between *A. gambiae, A. merus* and *A. stephensi* (**Figure 4.15**) shows that Ag1983 and Ag1984 are localized to a synteny block (Ag1970-Ag2009) in *A. stephensi*, which suggest that the distal breakpoint structure of 2R rearrangement in *A. stephensi* is identical to the $2R+^{p}$ of *A. gambiae*. There is a breakpoint located between Ag3326 and Ag3328 in *A. gambiae* and *A. stephensi*.

Table 4.6: The localizations of probes near the 2Rp breakpoints in A. *stephensi*, *A. merus* and the coordinates of *A. gambiae*

#	Clone	Accession	Genomic location in *A. gambiae* genome	e - value	*A. gambiae*	*A. merus*	*A. stephensi*
1	Ag1763	AGAP001763	9,523,856-9,528,904	0	2R:9A	2R:9A	2R:16A
2	Ag1970	AGAP001970	12,992,452-12,993,976	0	2R:9C	2R:9C	2R:10A
3	Ag1983	AGAP001983	13133424 to 13135252	0	2R:10A	2R:9C	2R:10A
4	Ag1984	AGAP001984	13150829 to 13154837	0	2R:10A	2R:14E	2R:10A
5	Ag2009	AGAP002009	13,876,353-13,889,539	0	2R:10A	2R:14E	2R:10A
6	Ag2015	AGAP002015	13,936,706-13,955,790	0	2R:10A	2R:14E	2R:9D
7	Ag2030	AGAP002030	14,084,885-14,096,582	0	2R:10B	2R:14E	2R:10A
8	Ag3315	AGAP003315	35,837,690-35,839,243	0	2R:14E	nd	2R:17B
9	Ag3326	AGAP003326	35998659 to 35999124	0	2R:15A	2R:14E	2R:17B
10	Ag3328	AGAP003328	36027604 to 36028480	0	2R:15A	2R:9C	2R:18A
11	Ag3342	AGAP003342	36,307,756-36,311,720	0	2R:15A	2R:15A	2R:17C
12	Ag3366	AGAP003366	36,878,347-36,890,038	0	2R:15B	2R:15A	2R:14A

Ag1983 and Ag1984 were also mapped to the polytene chromosome of *A. nili* and *A. moucheti*. Hybridization results suggest that both clones of AGAP001983 and AGAP001984 were localized to the same chromosome site in *A. nili* and *A. moucheti* (**Figure 4.16**), which is consistent with *in situ* hybridization result in *A. stephensi*.

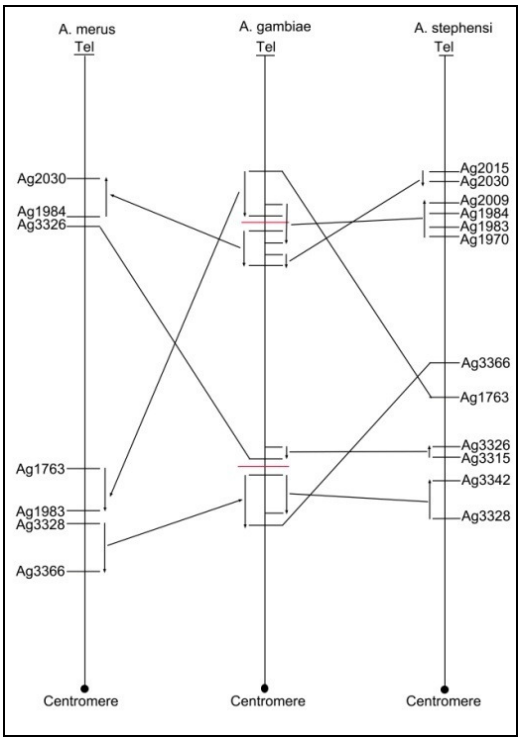


Figure 4.15: Gene orders of 12 cDNA markers near the breakpoints of 2R+^{p} on *A. gambiae*, *A. merus* and *A. stephensi*. Red lines indicate the both breakpoints of 2R+^{p} inversion in *A. gambiae*.

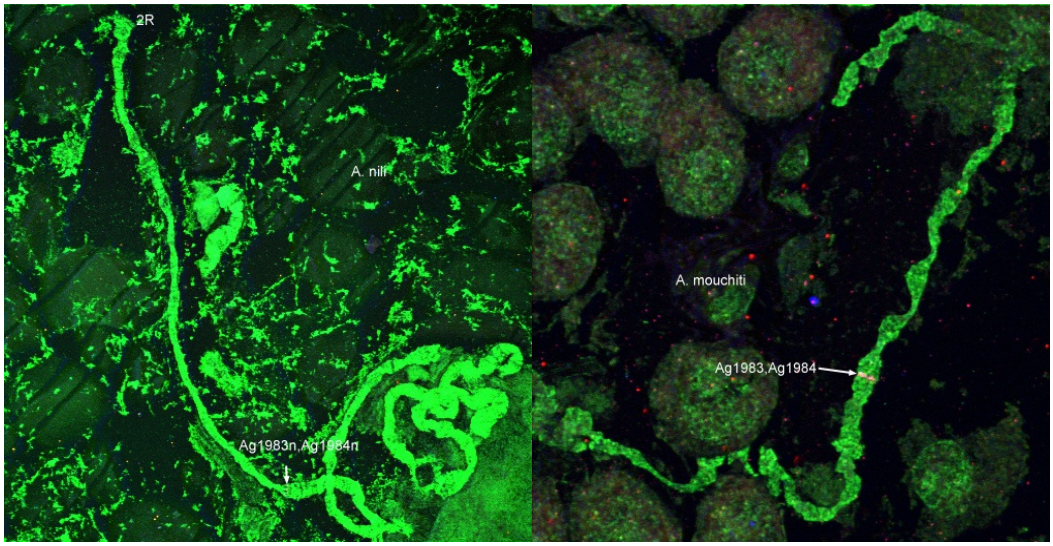


Figure 4.16: (left) In situ hybridization of Ag1983 and Ag1984 on *A. nili.* (Right) In situ hybridization of Ag1983 and Ag1984 on *A. moucheti*

4.3.2.4. The molecular analysis of the 2Rop breakpoint regions in *A. gambiae* PEST strain and *A. stephensi*

In situ hybridization results revealed that breakpoints of 2R+^{op} in *A. gambiae* complex are located between transcription units, not within the coding sequences. DNA transposons, retroelements, microsatellite, minisatellite, satellite and inverted repeats in all breakpoint regions were analyzed using http://www.repeatmasker.org/, http://tandem.bu.edu/trf/trf.html, http://tandem.bu.edu/cgi-bin/irdb/irdb.exe and the data are present in **Table 4.7**. The sequences of four markers, 1759, 1763, 2934 and 2933 were obtained from database of *A. gambiae* PEST strain (http://www.vectorbase.org/index.php) and used to BLAST against the shotgun genome sequences of *A. stephensi* (Zhijian Tu, unpublished). Two scaffolds (03514 and 09371) have been found containing both breakpoints of 2Ro in *A. stephensi.* BLAST these scaffolds against the sequences of markers near the breakpoints in *A. gambiae* PEST strain using BLAST 2 sequences (http://blast.ncbi.nlm.nih.gov/) revealed the molecular organization of the 2Ro breakpoint regions in the standard and inverted arrangements (**Figure 4.17**). Due to distantly related phylogenetic relationship between *A. gambiae* and *A. stephensi (217, 388, 429)*, those genomes sequences have been highly diverged even within genes. Almost no similarity could be traced for the noncoding regions in *A. gambiae* and *A. stephensi*. The presence of transposable elements in both breakpoint regions of 2R+^{op} in *A. gambiae* suggest that TEs involve in the formation of these inversion breakages. Further analysis of TEs in 2R+^{p} demonstrates that one Short interspersed repetitive element (SINE) is located in distal and proximal breakpoints of 2R+^{p} in opposite orientations in *A. gambiae*. The discovery of several microsatellites

and minisatellites in the distal breakpoint regions of 2R+^{op} suggests the susceptibility of these breakpoint regions to breakage. Additionally, *in situ* hybridization results revealed one pseudo gene in the proximal breakpoint of 2Ro in *A. merus* (**Figure 4.11 - 4.12)** which yield weaker signal than original copy located in distal breakpoint. The distance of this pseudo gene to the breakpoint of 2Rp and the orientation to the original copy is unclear.

Table 4.7: The analysis of molecular features in breakpoint regions of 2R+^{op} of *A. gambiae*

	Region	Length (bp)	Inverted repeats	Microsatellite	Minisatellite	Satellite	Retroelement	DNA transposon
2R+^{o} in *A. gambiae*								
Distal	1759-1763	40565	2	5	17	0	1	2
Proximal	2934-2935	2388	2	0	0	0	1 MITE	0
2R+^{p} in *A. gambiae*								
Distal	1983-1984	15577	2	5	3	0	4	0
Proximal	3327-3328	8211	1	0	3	0	3	1

MITE: miniature inverted-repeat transposable element identified by Z. Tu.

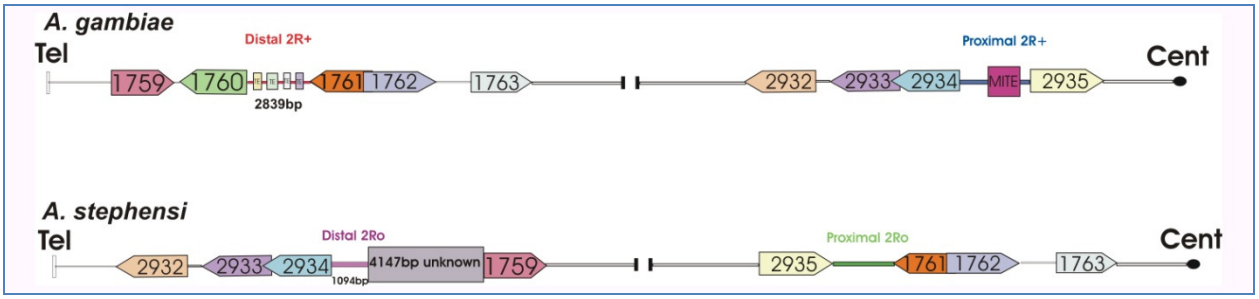


Figure 4.17: Sequence organization of alternative 2Ro inversions in *A. gambiae* and *A. stephensi*.

5. Discussion

This project provided a physical mapping framework for inferring ancestral chromosome arrangements and polarizing the evolutionary history of the *A. gambiae*

species complex. This knowledge is crucial in identifying the evolutionary genomic changes associated with the origin and loss of human blood choice, ecological and behavioral adaptations, and association with human habitats. Comparative genomic analysis performed within a phylogenetic framework is a powerful tool for finding genes underlying rapidly evolving “vector traits.” These genes can be targeted and manipulated so that malaria transmission can be reduced and eventually eliminated.

5.1. The ancestral status of the 2La inversion in *A. gambiae* complex

For a long time, the 2L+ was considered ancestral arrangement while 2La was derived (*13, 42*). In addition, another study proposed the multiple origin of 2La inversion and suggested that 2La in *A. merus* and *A. arabiensis* arose independently (*235*). However these views were questioned by the discovery of the presence of 2La arrangement in the Oriental *A. subpictus* complex (*221*). The more recent sequencing of 2La breakpoints rejected the view of the multiple origins of 2La in *A. gambiae* complex (*164*). The molecular organization of the 2La breakpoints in all three species: *A. gambiae Sua* (2La), *A. arabiensis* and *A. merus* is identical (*164*). The full-length genes at the breakpoints of 2La and their pseudogene copies only at breakpoints of the 2L+ arrangement indicated that 2La inversion is the ancestral arrangement (*164*). Based on the monophyletic origin of 2La inversion assumption, if 2La inversion is ancestral arrangement thus the outgroup species: *A. funestus* and *A. stephensi* should carry this inversion. Since the low resolution of the physical map of *A. funestus*, the identification of 2La inversion breakpoints in *A. funestus* has been failed in this project. Computational analysis using physical maps of outgroup species: *A. stephensi* and *A. funestus* supported the ancestral status of the inverted arrangement. The presence of

2La inversion breakpoint structures in *A. stephensi* and *A. nili* confirmed the ancestral status of 2La inversion. Therefore all the data suggested that the molecular organizations of the breakpoints of 2La inversion are present in the genomes of *A. stephensi* and *A. nili*. Previous direct sequencing of 2La breakpoints together with our current results confirmed that 2La inversion is ancestral while the standard 2L+ is derived. A hypothetical phylogenetic tree of the *A. gambiae* complex has been developed based on the ancestry of the 2La arrangement (**Figure 4.18**).The derived nature of the 2L+ arrangement and the ancestral status of 2La suggest that anyone of *A. arabiensis*, *A. gambiae*, or *A. merus* can be considered the closest to the ancestral species.

5.2. The ancestral status of the 2Rop inversions in *A. gambiae* complex

Because of the ancestral status of the 2La inversion, *A. merus*, *A. gambiae* or *A. arabiensis* can be considered as close to the ancestral species in the *A. gambiae* complex (**Figure 4.18**). *A. merus* and *A. gambiae* share the Xag inversion, *A. arabiensis* carries the Xbcd inversion, and the other members of the complex have the “standard” X arrangements. Unfortunately, the ancestral status of five inversions on X chromosome is difficult to determine because of two reasons. First, accumulated data in *Drosophila* and *Anopheles* suggest that the X chromosome evolves faster than autosomes (*1, 33, 37*).This results in very small conserved synteny blocks on the X chromosome between *A. gambiae* and *A. stephensi.* Second, in our studies, a large number of cDNA and BAC clones obtained from *A. gambiae* X chromosome failed to map on the chromosome X of *A. stephensi* because of high sequence divergence. Therefore, the low resolution of the physical maps of *A. funestus* and *A. stephensi* provide insufficient markers to cover all

five inversions. But *A. gambiae* and *A. merus* can be differentiated by two inversions, "0" and "p" on 2R chromosomal arm. *A. merus* carries the inverted arrangement and *A. gambiae* has the standard. To identify the ancestral status for 2Rop inversions have been attamped in this study.

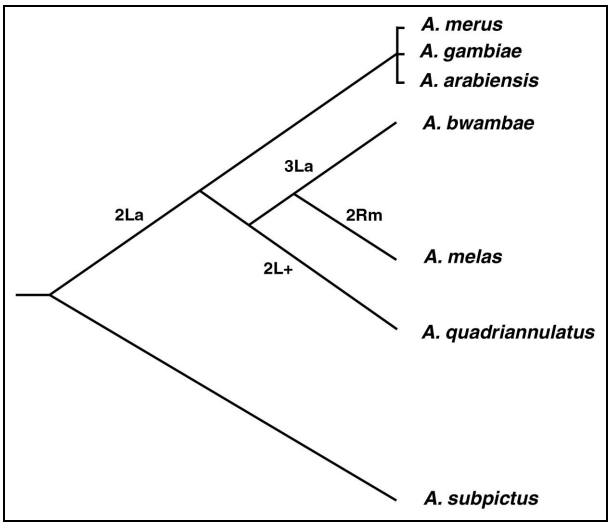


Figure 4.18: A hypothetical phylogenetic tree of the *A. gambiae* complex based on the ancestry of the 2La arrangement. The known chromosomal arrangements that support this tree are shown **(*371*)**.

Several TEs in the 2R+^{o} distal and one MITE in proximal breakpoint region in *A. gambiae* suggested the role of these TEs in the generation of the 2Ro inversion, however, MITE has no sequence similarity with the other TEs which indicate that this inversion was not generated by ectopic recombination mechanism. This mechanism requires the presence of homologous sequences, such as TE or segmental duplications, in opposite orientations at two sites in the parental chromosome (*430*). The study of breakpoint regions of 2R+^{p} in *A. gambiae* strongly support ectopic recombination model. The presence of one SINE on both breakpoints in inverted orientation suggests that 2R+^{p} in *A. gambiae* is a derived arrangement.

Phylogenetic trees generated by computational programs suggested that 2Rop inversions are closer to outgroup species, *A. stephensi* and *A. funestus*. Gene order of markers near the breakpoints revealed that 2Ro inversion is present in *A. stephensi. However, A. stephensi*, *A. nili* and *A. moucheti* shared the standard 2Rp ($2R+^{p}$) distal breakpoint region. Therefore, our current evidence suggested that 2Rop are ancestral rearrangements in *A. gambiae* complex and the carrier of 2Rop inversion, *A. merus* is cloestest to ancestral species. The contradiction of 2Rp between the molecular and *in situ* hybridization analyses might result from ancestral polymorphism of 2Rp being fixed after speciation independently in different lineages or by breakpoint reuse.

6. Conclusions

1) Physical mapping of ingroup and outgroup species can be used for identifying ancestral arrangements within species complexes, if the inversions are monophyletic.
2) The *A. gambiae* complex shares the 2La and 2Ro arrangements with *A. stephensi, A. nili*.
3) The presence for TEs at the 2Rp+ breakpoints of *A. gambiae* suggests the ancestral status of the inverted arrangement.
4) The inconsistency between the molecular and *in situ* hybridization analyses could be explained by ancestral polymorphism or reuse of the distal 2Rp breakpoint in evolution.

Chapter five: Comparative genomics of the *Plasmodium* resistance island in malaria mosquitoes

1. Abstract

The study of *A. gambiae* natural populations has identified a *Plasmodium falciparum* resistance island (PRI) in a small region of the 2La inversion near the proximal breakpoint in *A. gambiae*. Using the cross-species mapping, we identified syntenic regions within PRI in *A stephensi* and *A. nili*. We also developed a protocol for sequential performing laser microdissection and multiple displacement amplification to obtain DNA from both breakpoints of the 2La for sequencing. The chromosomal regions spanning both breakpoints of the 2La in *A. arabiensis* and *A. merus* were laser microdissected from the polytene chromosomes. Subsequently, the DNA samples were amplified using Illustra GenomePhi V2 DNA and Whole-pool amplification methods. The successful amplification of our target DNA was confirmed by PCR with specific primers followed by Sanger sequencing. The sequence data from alternative 2La arrangements will shed light on the molecular mechanism of the 2La inversion responsible for the variation of malaria transmission.

2. Introduction

2La inversion is widespread across natural populations of *A. gambiae*, and also the most well studied polymorphic inversion. The studies of 2La inversion in natural populations and laboratory strains of *A. gambiae* suggested that the alternative rearrangements of 2L chromosomes strongly correlated with refractoriness and susceptibility to various *Plasmodium* species (*242-246*). A recent genetic survey of an *A. gambiae* natural population has identified the strongest *Plasmodium falciparum*

resistance locus in a small region of 2La inversion near the proximal breakpoint (*247*). Therefore, all evidence revealed that chromosomal regions spanning both breakpoints from alternative rearrangements captured different alleles responsible for parasite infection and resistance. To obtain DNA sequences from both breakpoints of 2La inversion is an essential step to analyze the alleles difference with the sequences from standard 2L+^{a}, therefore shedding light on the molecular mechanism of mosquitoes resistance against different *Plasmodium*.

The advent of Laser capture microdissection represents an utmost important and interesting technique in molecular biology. LCM can isolate chromosomal fragments from polytene chromosomes and metaphase chromosomes, microdissect single cells, and specific tissue population within a microenvironment (*356-359*). Another technology named whole genome amplification makes it possible to obtain amplicons from above isolated chromosal regions, single cell or specific tissue (*379, 380, 422, 430-432*). Therefore, obtaining DNA amplicons using WGA from microdissected chromosomal regions or whole chromosomal arm is a powerful approach for characterization of inversion breakpoints and genomic analysis of specific chromosomal regions. However, WGA was applied to the microdissected metaphase chromosomes only once (*429*). We applied, for the first time, WGA and sequencing to laser micro-dissected chromosomal regions from isolated polytene chromosomes. A breakthrough protocol has been developed for performing the two procedures. The distal and proximal breakpoints of 2La inversion from *A. arabiensis* and *A. merus* laboratory strains were microdissected using Laser Microdissection Microscope. The pg amount of DNA material was amplified by Wpa (Whole-pool amplification) and GenomiPhi V2 respectively. The confirmation of

target amplicons was performed by conventional PCR procedure with specific primers followed by sequencing. The success of amplification rate was estimated.

3. Material and methods

3.1. Physical mapping

The physical map of *A. stephensi* was developed in the chapter one of this dissertation. The physical map of *A. nili* was developed for this project. The mosquitoes of *A. nili* were collected from Africa and *A. gambiae* Sua strain was obtained from Malaria Research and Reference Reagent Resource Center (http://www.mr4.org/). Chromosomal preparations for these two species were performed using the method of chapter one. The sequences of all the markers mapped to the polytene chromosome 2L of *A. gambiae* Sua strain and 3L of *A. nili* were acquired from the Vectorbase using search tool (http://www.vectorbase.org/Search/Keyword/). Primer 3 (http://frodo.wi.mit.edu/primer3/) was used to design primers for all the probes. The PCR products were amplified using genomic DNA of *A. gambiae* Sua strain or *A. nili* isolated using Qiagen DNeasy Blood & Tissue Kit as template and by following PCR conditions: 95℃ for 4 min; 35 cycles of 94℃ for 30 s, 55℃ for 30 s, 72℃ for 30 s; 72℃ for 5 min. And then all the PCR products were purified from agarose gel using the GENECLEAN III kit (MP Biomedicals). Finally the purified DNA was labeled with Cy5-AP3-dUTP and Cy3-AP3-dUTP (GE Healthcare UK Ltd, Buckinghamshire, England) using modified Random Primers DNA Labeling System (Invitrogen Corporation, Carlsbad, CA, USA) and hybridized to the polytene chromosomes of *A. gambiae* and *A. nili* based on the *in situ* hybridization method described in Chapter one.

3.2. Chromosomal preparation for laser capture microdissection

Malaria Research and Reference Reagent Resource Center (http://www.mr4.org/) kindly provided *A. arabiensis* and *A. merus* eggs and these colonies were kept in the Insectary of Fralin Biotechnology Center (http://www.biotech.vt.edu/). The ovaries of *A. gambiae* and *A. arabiensis* as well as *A. merus* mosquitoes were dissected from half-gravid female mosquitoes under dissection microscope and fixed in Carnoy's solution (Methanol: Acetic Acid Glacial = 3:1). The following protocol was developed to make chromosomal preparations.

One ovary was removed from the vials with a pair of forceps and placed into a drop of Carnoy's solution (Ethanol: Acetic Acid Glacial = 3:1) on a UV treated regular slide. Remains of abdomen and/or blood meal, gut, etc were dissected and discard with dissecting needles. Then one ovary was divided into four pieces, and one piece was placed on one 1.0 mm PET-membrane slide purchased from ZEISS (Before use, treat the membrane slides with UV light 254nm for 30 minutes, which facilitates adherence of paraffin or frozen sections); One drop of 50% Propionic acid was added on the membrane slide and left for about 5 min until follicles are cleared and swollen to about twice their original size. Under a dissecting microscope, after the follicles were carefully separated from each other, one autoclaved coverslip was placed on the top. (Siliconizing cover glass was critical to keep chromosomes sticking on slides after removal the coverslip in the following steps, but decontamination of a coverslip was crucial to avoid the contamination.)

To prepare siliconized cover-glasses, they were first fully immersed indiviually in Sigmacote (Sigma) in a beaker or other container made of glass (repel silane solutions

from other companies can be used as well). Then the beaker was covered tightly with foil, and kept in a fume hood for overnight. The next day, each coverslip was lifted by its corner with forceps and dipped into a beaker of water and then 100% ethanol to clean. Each cover-glass was placed on a lint free surface to dry, and thereafter stored them in a box until required (http://www.mr4.org/).

Then, the coverslip was covered with a filter paper and held it with two fingers and tapped gently with the eraser end of a pencil to break the cell membranes and released chromosomes. The membrane slide with chromosomes was immersed in 50% ethanol in a slide jar for 5 min at room temperature, and coverslip was removed gently using razor blade (UV treatment or autoclaved). After membrane slide was dehydrated in 70%, 90% and 100% Ethanol for 5min each at RT, it was air-dred in a UV treated slide box and kept at RT until needed.

All the procedures were performed at clean and decontiminated surface wiped by 70% ethanol and RNase AWAY (MβP, Catalog # 7000). All the tools such as forceps, dissection needle, slides, paper towel, slide boxes, slide jars were UV treated in a Crosslinker (254nm for 30 minutes). Coverslides were autoclaved using gravid cycle for 20 minute. The membrane slides were dipped for a few seconds into RNase AWAY, followed by two separate washing steps in RNase-free distilled water and drying at 37°C for 30 minutes up to 2 hours. Subsequently, the membrane was irradiated with UV light at 254nm for 30 minutes. During the whole process of chromosomal preparation, membrane slides were always kept in a UV treated slide box to avoid exposing them to the open air.

3.3. Laser capture microdissection

The following protocol was developed to perform microdissection.

To prepare catapult buffer, 2 µl of 0.5M EDTA PH 8.0, 20µl of 1M Tris PH 8.0, 5 µl of Igepal CA-630(Sigma # I-3021), and 10 µl of Proteinase K 20mg/ml were mixed up, and then filled up with water (RNase, DNAse and Protease free water, PCR grade) to 1ml. .

First under 5 × objective of dissection microscope, we visualized the polytene chromosomes, and then focused on them using 40× objective. After we found chromosome 2L based on the banding patterns, the breakpoints of 2La was localized on the slide. The chromosomal regions needed to cut were marked with software, and parameters were set up for energy at 42 and focuses at 36. Laser microdissection was performed around the distal or proximal breakpoint of 2La inversion. After the laser cut, 10µl of Catapult buffer was pipetted in the middle of the cap of the regular tube (keep upside down all the time). The regular tube with Catapult buffer or adhesive tube was placed directly into the cap/tube holder of PALM laser microdissection system (Carl Zeiss Microimaging, Inc). After the cap was positioned on above the microdissected material, chromosomal fragments were captured into the cap of tube using proper energy and focus. Then, the tube was removed from tube holder, and closed with attached cap. The regular tube was centrifuged at 16000 rcf for 10 minutes tocollect the DNA samples. 10 µl of Catapult Buffer was added into the tube, and digested at 55^{o}C for 18 hours followed by a heating step at 95^{o}C for 5 minutes to inactivate Proteinase K.

50 µl Catapult Buffer without Igepal was added into the cap of the adhesive tube, and incubated upside for 18 hours. After digestion, adhesive tube was centrifuged at low

speed for 5 minutes to collect solution. A heating step at 95 ^{o}C for 5 minutes w required to inactivate Proteinase K.

3.4. DNA purification and mass amplification

After the digestion of microdissected chromosomal materials, the released DNAs were divided into different groups for downstream amplification. The purification of DNA samples was performed either with QIAEXII (Qiagen) or precipitated by 100% Ethanol. Purified DNAs were amplified either by Illustra GenomiPhi V2 DNA Amplification Kit following their protocol (GE Healthcare Life Science, Catalog #: 25-6600-31) or by the protocol of Wpa described in the publication (*367*).

3.4.1. The protocol for GenomiPhi V2 DNA amplification

First, DNA samples were denatured by heating or chemical. For heat denaturation of template, 1 μl of template DNA (at least 10ng) was mixed with 9 μl of sample buffer, and then the sample was heated to 95°C for 3 min followed by cooling to 4°C on ice. For chemical denaturation of template DNA, 1 μl template DNA was mixed with 1 μl of Denaturation Solution (400 mM KOH, 10 mM EDTA), and then the samples was incubated at room temperature for 3 minutes. 1 μl of Neutralization Buffer (400 mM HCl, 600 mM Tris-HCl, PH 7.5) was added into the tube, and stored on ice.

Then, amplification reaction was prepared. For the heat denatured sample, 9 μl of reaction buffer was combined with 1 μl of enzyme mix on ice, and then added to the cooled sample. For the chemical denatured sample, 7 μl of Sample Buffer and 9 μl of Reaction Buffer were combined together on ice, and then added 1 μl of Enzyme Mix. The mixture was added into a cooled sample.

Finally, the reaction was incubated at 30°C for 2 hour, followed by heating the sample to 65°C for 10 min to inactivate the reaction.

3.4.2. The protocol of Wpa amplification (*367*)

First, 1μl of freshly prepared 8-fold water diluted buffer A was placed on the bottom of a precooled 200μl thin-wall PCR tube, then add 1 μl of DNA (1–5 ng). After 8 min denaturation with diluted buffer A, add 1 μl of 8 fold water diluted buffer N (freshly prepared) and mixed. 5 μl of 1.2M trehalose was immediately added, and the solution was left on ice.

Then, a master mixture was made on ice with the following components for one sample: 19.8μl of distilled water, 10μl of10×RXNbuffer [500 mM Tris.HCl (pH 7.5), 100 mM NaCl, 100 mM MgCl, 100 mM (NH4)2SO_4, 50 mM DTT], a 1.5-μl mixture of 4×dNTPs (each 25 mM), 1.2 μl of 1 nmol/μl N9, 1 μl of 100 ng/μl BSA (NEB), and 57.5 μl of 1.2 M Trehalose . These components were stirred well and spun down before 1 μl of phi29 DNA polymerase (1 μg/μl; Amersham Pharmacia) was added. A 92-μl aliquot of the master mixture was delivered to each sample, stirred well, spun down again, and incubated at 30℃ for 16 h. Finally, the tubes were heated at 70℃ for 20 min to stop the reaction.

After amplification, the amplicons then were purified with QIAEXII (Qiagen) with additional wash, which is optimal for the purification of product > 4Kb. The quantification of amplified DNA concentration was measured by UV absorbance with a Nano-Drop ND-2000 (Nano Technologies).

To perform DNA purification and amplification as well as above laser capture microdissection, all the reagents, tips, and tubes and their cups were autoclaved and

before use, UV light treatment is necessary for decontamination. All the water used must be collected from purification system and then autoclaved.

3.5. PCR confirmation of amplified DNA samples

Amplification of cDNA markers near the distal and proximal breakpoints of inverted 2L in *A. merus* was performed using available primers for physical mapping in *A. merus* and *A. arabiensis* under following conditions: 95℃ for 10 min; 45 cycles of 95℃ for 30 s, 55℃ for 30 s, 72°C for 1 min; 72°C for 5 min. After the amplification, the PCR products were load on 1% of agarose gel and the bands around 500bp were cut from gel and purified using Qiaquick Gel Extraction Kit (Qiagen). The sequencing was performed by VBI (Virginia Bioinformatics Institute) Core Lab on 10µl of 100ng PCR product for each sample using Sanger sequencing method.

4. Results

4.1. Physical mapping of *Plasmodium* resistance island in *A. nili* and *A. stephensi*

A survey of *A. gambiae* population in Western Africa Mali for naturally occurring genetic loci that control mosquito infection with the human malaria parasite, *P. falciparum* identified a genomic *Plasmodium*-resistance island (PRI) (*247*). This PRI included four loci: Pfin1, Pfin4, Pfin5 (*Plasmodium* infection intensity short for Pfin) and Pfmel2 (*P. falceparum* melanization 2) in a small chromosome region of 2La in *A. gambiae* (**Figure 5.1**) and each locus explained at least 89% of parasite-free mosquitoes in independent pedigrees. Together, this PRI explained most of the genetic variations of malaria parasite infection of mosquitoes in nature. The analysis of the chromosomal location of PRI indicated that this cluster is located near the proximal

breakpoint of 2La in *A. gambiae*. Among the candidate genes in this PRI, APL1 (*Anopheles Plasmodium*-responsive leucine-rich repeat 1) has been confirmed by RNA interference to play a significant role of protection from *P. berghei* infection (*247*). The later study suggested the APL1 gene is located within 1Mb from the proximal breakpoint of 2La inversion (*164*). Manual reannotation revealed that APL1 is a family of at least 3 independently transcribed genes (APL1A, APL1B, APL1C) and APL1 functions within the Rel1-Cactus immune signaling pathway (*431*). In other human malaria vectors for instance, *A. nili* and *A. stephensi*, the preservation of the gene orders in the small chromosomal region (PRI) was unclear. *A. nili,* an important vector of human malaria in Africa, has a wide geographic distribution, spreading across most of West and Central Africa, mainly in humid savannas areas. *A. stephensi* is the major Asian malaria vector. The gene orders between *A. gambiae* and *A. stephensi*, *A. gambiae* and *A. nili* were compared. The comparison results provide valuable information on the evolution of this PRI and APL1 gene family between malaria mosquito lineages.

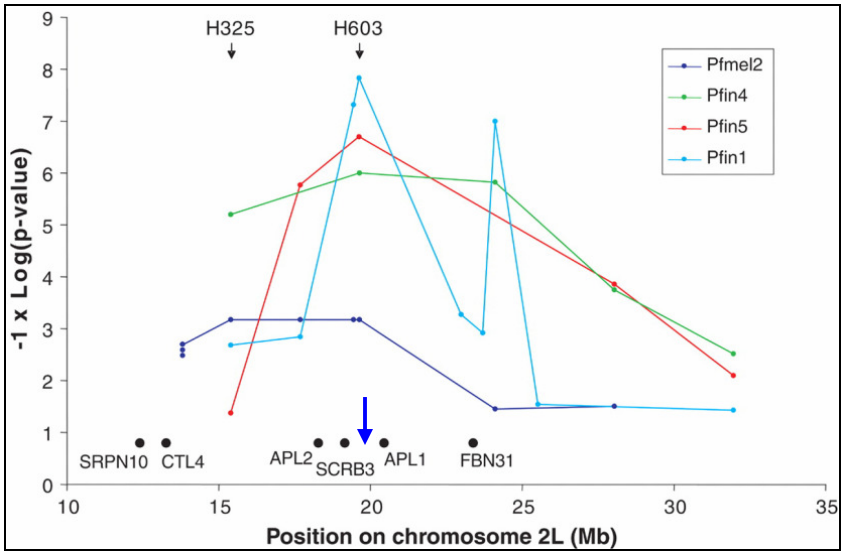


Figure 5.1: Fine mapping of the Plasmodium resistance island. The locations of the markers yielding significant linkage signals in the initial 9-cM scan, H325 and H603, are

indicated by arrows. Predicted functions or functional domains: SRPN10, serpin; CTL4, c-type lectin; SCRB3, scavenger receptor; and FBN31, fibrinogen. The blue arrow indicates the proximal breakpoint of the 2La inversion (*247*).

4.1.1. A physical map for *A. nili*

A physical map for *A. nili* has been developed and 27 clones have been localized to the polytene chromosomes of *A. nili*. All the probes and *in situ* hybridization results are listed in **Table 5.1** and the locations of all markers on polytene chromosomes of *A. nili* were indicated on **Figure 5.2**.

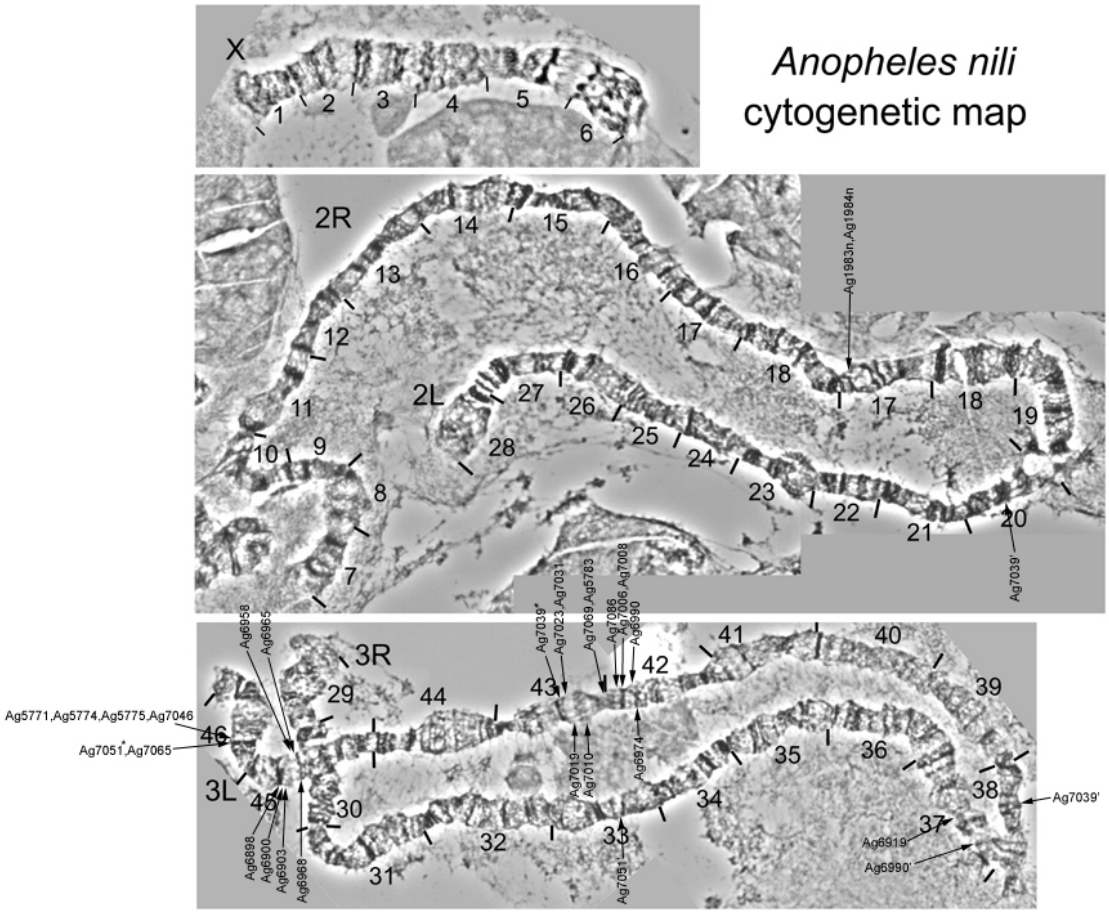


Figure 5.2: The localizations of 27 clones on the polytene chromosomes of *A. nili.*

Asterisks or apostrophes indicate the major signal or multiple signals, respectively, on the polytene chromosomes of *A. nili*.

4.1.2. The gene order comparison between *A. gambiae* and *A. nili*, *A. gambiae* and *A. stephensi*

The chromosomal locations and coordinates of probes of *A. gambiae* were obtained from Vectorbase using search tool which is available at http://www.vectorbase.org/Search/Keyword/. The chromosomal locations of markers in *A. stephensi* were obtained from Chapter one. All the locations of probes on the chromosome of *A. stephensi* and *A. nili* as well as *A. gambiae* were present in a list of **Table 5.1**. The gene orders among these three species were compared and shown in **Figure 5.3**. The size of PRI region on 2La chromosome of *A. gambiae* is in about 15 Mb interval from around 15 - 30 Mb (*247*). All the probes near the proximal breakpoint of 2La inversion are within the PRI. The comparative analysis reveals that despite most of genes are located inside the synteny blocks between *A. gambiae* and *A. nili* or *A. stephensi*, but several fixed inversions have occurred in the PRI region between lineages. One large with size of more than 2 Mb, and three small synteny blocks were identified between *A. gambiae* and *A. stephensi* genomes. Five synteny blocks were found between *A. gambiae* and *A. nili*. Despite the chromosomal rearrangements between lineages, the APL1 genes (**Red in Table 5.1**) were still located in the synteny block. This might suggest that the same resistance mechanism was shared among three species. When we analyze the gene orders among three species, we also found *A. stephensi* carries the breakpoint structure of 2La inversion, which is consistent with the ancestral status of 2La inversion (Chapter three). However, the proximal breakpoint of

2La inversion was reused after the divergence of *A. gambiae* and *A. nili* from common ancestor.

Table 5.1: The chromosomal locations of the markers on *A. nili*, *A. stephensi* and the coordinates of *A. gambiae*

#	Probe	Coordinates in *A. gambiae*	*A. gambiae*	*A. stephensi*	*A. nili*
1	Ag5711	19,260,915 - 19,275,432	2L:22E	3L:44B	no signal
2	Ag5761	20,336,662 - 20,340,021	2L:22F	3L:44C	no signal
3	131F22	20,364,135 - 20,459,325	2L:22F	3L:44C	no this clone
4	Ag5765	20,357,930 - 20,373,012	2L:22F	3L:44C	no PCR product
5	131F22	20,364,135 - 20,459,325	2L:22F	3L:44C	no this clone
	AFND19	20,396,064 - 20,405,726			
6	Ag5771	20,415,195 - 20,425,711	2L:22F	3L:44C	3L:46
7	Ag5774	20,455,251 - 20,464,457	2L:22F	3L:44C	3L:46
8	Ag5775	20,498,195 - 20,501,158	2L:22F	3L:44C	3L:46
9	Ag5778	20,521,764 - 20,523,605	2L:22F	3L:44C	no signal
Proximal 2La breakpoint					
#	**Probe**	**Coordinates in *A. gambiae***	***A. gambiae***	***A. stephensi***	***A. nili***
10	Ag7068	42,163,506 - 42,164,602	2L:26D	3L:44C	no signal
11	Ag7065	42,135,463 - 42,147,851	2L:26D	3L:44C	3L:46
12	Ag7063	42,126,193 - 42,127,751	2L:26D	3L:44C	no PCR product
13	Ag7051	41,835,037 - 41,874,573	2L:26D	3L:44C	3L:46
14	Ag7046	41,624,489 - 41,644,582	2L:26D	3L:44C	3L:46
15	Ag7039	41,292,153 - 41,294,040	2L:26D	3L:44C	3L:43
16	**Ag7036, APL1A**	41,271,509 - 41,272,901	2L:26D	no signal	no signal
17	**Ag7035, APL1B**	41,266,619 - 41,268,364	2L:26D	no signal	no signal
18	**Ag7033, APL1C**	41,257,877 - 41,260,194	2L:26D	no signal	no signal
19	Ag7031	41,165,090 - 41,222,640	2L:26D	3L:44C	3L:43
20	Ag7023	41,013,266 - 41,025,452	2L:26D	3L:44C	3L:43
21	Ag7022	41,006,603 - 41,009,476	2L:26D	3L:44C	no signal
22	Ag7019	40,995,291 - 40,999,389	2L:26D	3L:43A	3L:43
23	Ag7014	40,974,685 - 40,976,342	2L:26D	3L:45C	no PCR product
24	Ag7008	40,792,493 - 40,836,990	2L:26D	3L:45C	3L:42
25	Ag7007	40,774,625 - 40,776,740	2L:26D	3L:45C	no signal
26	Ag7006	40,730,489 - 40,759,895	2L:26D	3L:45C	3L:42
27	11_F09	40,536,371 - 40,538,248	2L:26D	3L:45C	no this clone
28	Ag6990	40,485,842 - 40,495,867	2L:26C	3L:45C	3L:42
29	AF262H10	40,442,200 - 40,442,600	2L:26C	3L:45C	no this clone
30	Ag6974	40,434,580 - 40,440,489	2L:26C	3L:45C	3L:42
31	Ag6968	40,404,294 - 40,407,881	2L:26C	3L:43A	3L:45C
32	Ag6965	40,281,222 - 40,285,900	2L:26C	3L:43A	3L:45C
33	Ag6958	40,256,207 - 40,258,932	2L:26C	3L:43A	3L:45C
34	Ag6945	40,195,614 - 40,198,230	2L:26C	3L:43A	no signal
35	09_C11	39,995,907 - 39,997,107	2L:26B	3L:43C	no this clone
36	Ag6903	39,809,854 - 39,812,317	2L:26B	3L:43C	3L:45C
37	Ag6900	39,634,588 - 39,672,322	2L:26B	3L:43C	3L:45C
38	Ag6898	39,574,075 - 39,578,295	2L:26B	3L:43C	3L:45C
39	211H03	39,285,040 - 39,287,251	2L:26B	3L:42A	no this clone
40	16_F07	39,215,802 - 39,216,191	2L:26B	3L:42A	no this clone

	12G16	24,626,085 - 24,724,371	2L:23C	3L:44A	no this clone
41	29_E12	24,671,856 - 24,674,367			
42	716320	22,321,985 - 22,327,760	2L:23B	3L:40A	no this clone
43	AsSP53.7	21,865,434 -21,867,120	2L:23B	3L:40A	no this clone
44	Ag5783	20,572,013 - 20,578,767	2L:23A	3L:40A	3L:43
45	SuaPh6_1.8EcoRI	20,535,740 - 20,538,254	2L:23A	3L:40A	no this clone
46	Ag5779	20,528,559 - 20,529,407	2L:23A	3L:40A	no PCR product
Distal 2La breakpoint					
#	**Probe**	**Coordinates in *A. gambiae***	***A. gambiae***	***A. stephensi***	***A. nili***
47	Ag7069	42,165,841 - 42,176,356	2L:27A	3L:40A	3L:43
48	Ag7070	42,178,250 - 42,181,793	2L:27A	3L:40A	no signal
49	31L22	42,206,288 - 42,309,500	2L:27A	3L:40A	no this clone
50	Ag7077	42,212,968 - 42,215,566	2L:27A	3L:40A	no signal
51	Ag7086	42,327,399 - 42,406,342	2L:27A	3L:40A	3L:42
52	04C08	43,540,182 - 43,634,335	2L:27A	3L:45A	no this clone

Highlight indicates the synteny block in different species.

No PCR product: no PCR product can be amplified from the genomic DNA of *A. nili*.

No signal: using amplified DNA from *A. nili* genomic DNA, *in situ* hybridization failed to yield signal on the polytene chromosome of *A. nili*.

No this clone: no clone available.

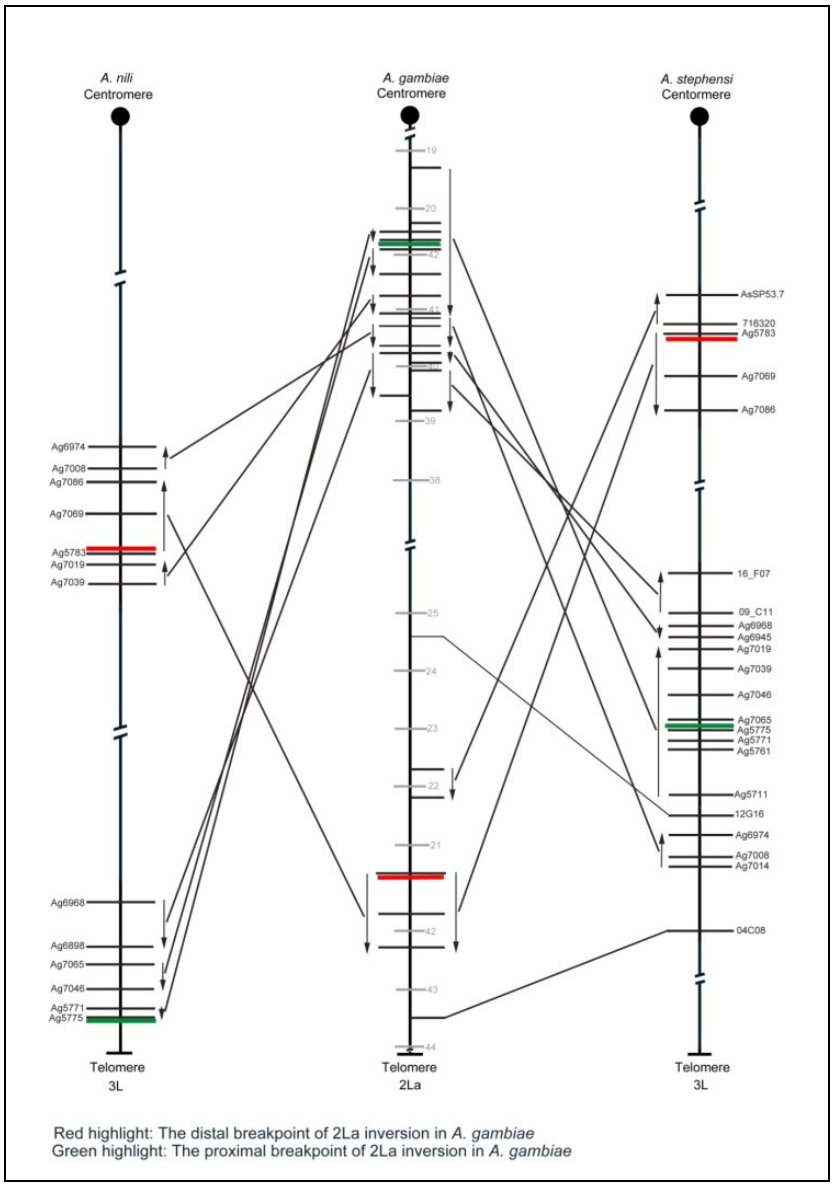


Figure 5.3: The gene orders comparison between *A. gambiae* and *A. nili*, *A. gambiae* and *A. stephensi* near the breakpoints of 2La. Arrow indicates the synteny block which contains at least two continuous markers. The locations of probes near the breakpoints in *A. gambiae* were drawn using the coordinates of *A. gambiae* 2L+[a], however for the rest chromosomal regions were shown not in scale. The breakpoints are shown by red and green lines.

4.2. The determination of the chromosomal regions near breakpoints of 2La and 2L+^{a} for sequencing

Several lines of evidence strongly suggested that infection rates of *A. gambiae* to different malaria parasite species are associated with the 2La inversion (*242, 245, 247*). Therefore, we hypothesized that the alternative 2La inversions captured distinct alleles which are responsible for the different phenotypic affect and adaptive values. The genomic differentiation between 2La and 2L+^{a} might be detected in the chromosomal region near the breakpoints due to the suppressed recombination. When we align the 12 Kb sequence adjacent to the distal breakpoint of the inverted arrangement of Sua strain (*164*) with the standard arrangement of PEST strain sequence which is available on Vectorbase using the AVID pair-wise alignment (*432*), we have found significant structural divergence between alternative rearrangements (**Figure 5.4**). Three large indels in the conserved non-coding sequences (CNS) with low similarity are identified and further analysis of three indels demonstrates that these three indels are insertions of blocks of repetitive DNA (**Figure 5.5**). The analysis of the ZNF 183 (nearest gene to the proximal breakpoint) sequences reveals that greater nucleotide difference between alternative rearrangements (Sua and PEST) than between different mosquito strains which carry 2La rearrangement (Bamako and Sua) (**Figure 5.6**). Therefore, all these data suggest that chromosomal inversion affects the structure and function of genes located in the area of breakpoints where recombination is absent or reduced. However, how large is the area of suppressed recombination around the breakpoints? In *Drosophila*, suppressed recombination effects can extend as far as 1 Mb from the breakpoints with almost no effect on the central regions (*433*). The recombination map

length of *D. melanogaster* is 294.9 cM and the length in *A. gambiae* is about 226.7 cM. Because of low recombination rates and large genomic size in *Anopheles* than in *Drosophila*, we expect that the area of suppressed recombination extend beyond 1 Mb from breakpoints in *A. gambiae*. More recently, a complete map of differentiation between the 2La and its alternative 2L+^{a} has been developed across the *A. gambiae* genome by hybridizing genomic DNA from individual wild caught *A. gambiae* mosquito to the oligonucleotide microarrays (*258*) (**Figure 5.7**). In the 22 Mb included within alternative arrangements, two chromosomal regions near to the breakpoints were identified as being significantly diverged. One is located near the distal breakpoint range from 20 to 22 Mb and another is a region from 39 to 42 Mb near the proximal breakpoint. Based on the genomic locations of these two chromosomal regions, four markers (**Table 5.2**) were labeled and hybridized to the polytene chromosome 2L in *A. gambiae* (**Figure 5.8**).

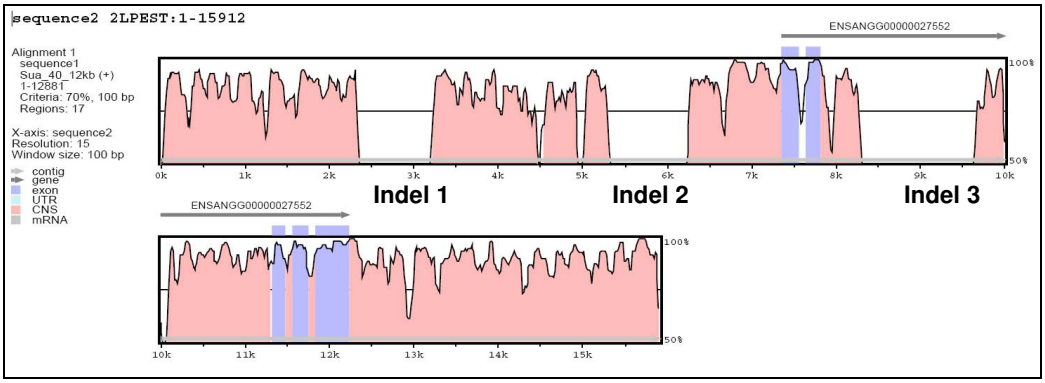


Figure 5.4: Alignment of the 2La SUA genomic sequence adjacent to the distal breakpoint to the 2L+^{a} PEST genome. X-axis shows the relative position of the PEST DNA, which is used as reference. Y-axis shows the % identity between SUA and PEST sequences.

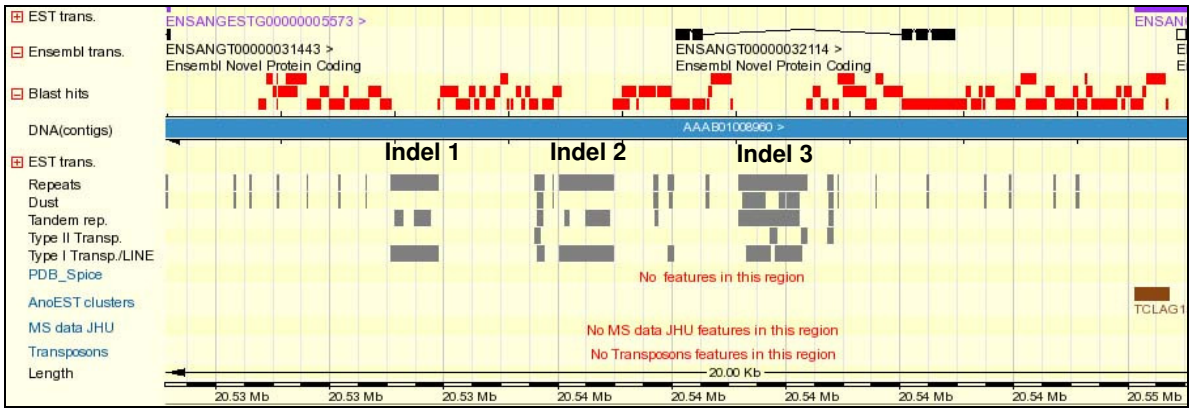


Figure 5.5: BLASTn of the 12 kb 2La SUA genomic sequence adjacent to the distal breakpoint against the 2L+[a] PEST genome.

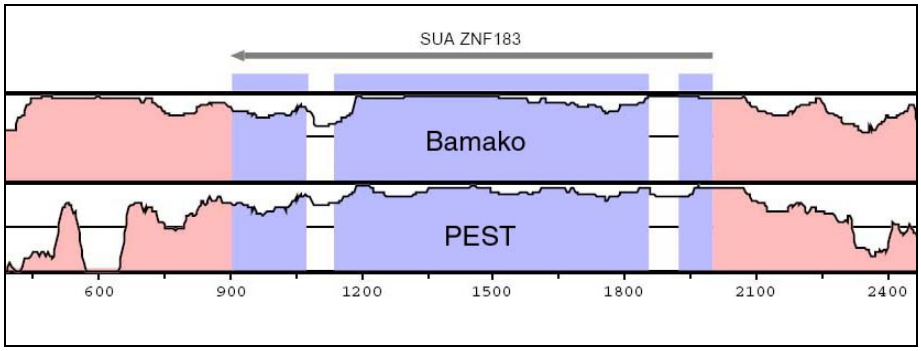


Figure 5.6: LAGAN multiple alignment of Bamako and PEST gene encoding ZNF183 to the SUA sequence (*434, 435*) **The** X-axis shows the relative position of the SUA DNA, which is used as reference. The Y-axis shows the % identity between Bamako, PEST, and SUA sequences. The cut-off for sequence conservation is 70% identity over 100 bp.

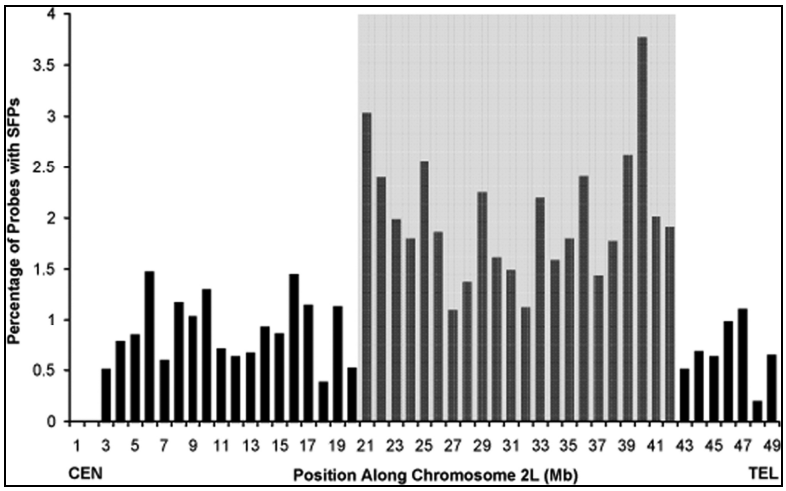


Figure 5.7 (*258*): Divergence across chromosome 2L between homozygous carriers of 2La and 2L+^{a} arrangements. Shaded area represents the rearranged region (*258*). CEN, centromere; TEL, telomere

Table 5.2: The chromosomal locations and the coordinates of four markers near the breakpoints of alternative 2L rearrangements in *A. gambiae*

	Marker	**Coordinate**	**Chromosome location**	**Length (bp)**
2L+Distal breakpoint	Ag7077	42212968 to 42215566	2L:27A	3,640,437
	Ag6797	38575129 to 38577166	2L:26A	
2L+Proximal breakpoint	Ag5831	22279212 to 22283130	2L:23B	1,921,282
	Ag5765	20357930 to 20373012	2L:22F	
2LaDistal breakpoint	Ag7077	42212968 to 42215566	2L:27A	1,798,696
	Ag5831	22279212 to 22283130	2L:23B	
2LaProximal breakpoint	Ag6797	38575129 to 38577166	2L:26A	3,766,941
	Ag5765	20357930 to 20373012	2L:22F	

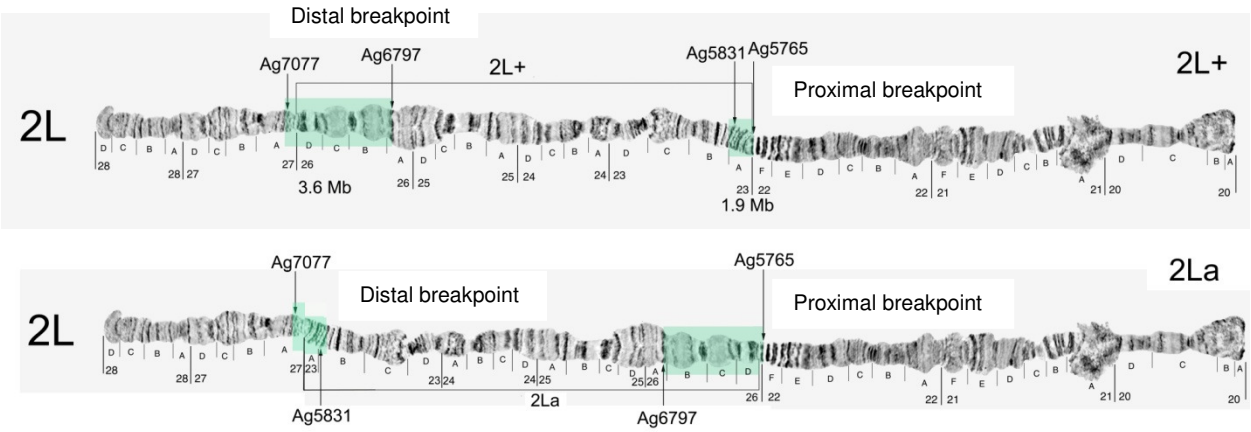


Figure 5.8: *In situ* hybridization results of markers on the polytene chromosome 2L of *A. gambiae*. Green highlight indicates the chromosomal regions for the following microdissection, amplification and sequencing.

4.3. The laser capture microdissection and mass amplification

4.3.1. The chromosome preparation for the microdissection

The mosquitoes of *A. merus* and *A. arabiensis* were grown at 26°C and 83% humidity in the insectary room of Fralin Biotechnology Center. After feeding three or more days of the emerged adult mosquitoes with 10% sugar water, mosquitoes were blood fed with live Guinea pigs. After 2nd or 3rd time blood feeding, let mosquitoes to develop ovaries for 25 hours and then dissected them. The ovaries were kept at Carnoy's solution (Methanol: Glacial Acetic acid = 3:1) at room temperature for 24 hours and then moved to the freezer (-20°C) for long term storage. The chromosome preparations were made on the membrane slides. Carl Zeiss provided four different membrane slides: 1.0 mm PEN-membrane covered (Cat #: 415101-4401-000), 1.0 mm PET-membrane covered (415101-4401-050), 0.17 mm PEN-membrane covered (415101-4401-500) and 0.17 mm PET-membrane covered (415101-4401-052). We used all of them to make chromosome preparations. The thin membrane slides were ideal for the microdissection

of small sized chromosome pieces, for instance one band under 100× objective, but the membranes were very fragile so that we had to handle them carefully. When we compared PEN and PET membrane slides, we found that PET-membrane slide can hold chromosomes more efficient than PEN slides because a lot of chromosomes were washed away from PEN slides. Therefore, for our purpose, we used the 1.0 mm PET membrane slides for our following procedures.

We also determined the effect of Carnoy's solution on our target DNAs, for instance DNA shearing or other damages. We isolated the genomic DNAs from one fixed ovary of *A. gambiae* Sua strain (Fixed in Carnoy's solution for more than half year) and one fresh live mosquito using Qiagen DNeasy Blood & Tissue Kit. RNase was added to remove RNA during the isolation. 10ul of genomic DNA from each sample was loaded on the 1% agarose gel and the image (**Figure 5.9**) shows that Carnoy's solution does not shear genomic DNA. But how much DNA was lost during the fixation was unclear. The DNA fragments were still large and therefore it was suitable for amplifying our target DNA directly from the polytene chromosome.

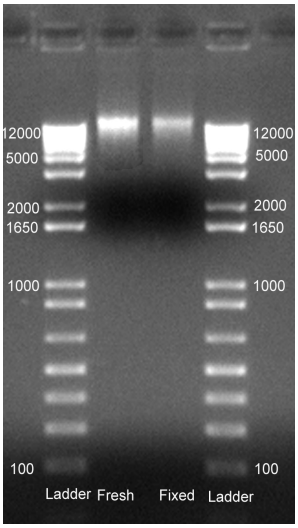


Figure 5.9: The genomic DNAs isolated from mosquito and fixed ovary from *A. gambiae Sua* on the agarose gel. Fresh indicates the genomic DNA isolated from one live mosquito. Fixed is the genomic DNA from one fixed ovary of *A. gambiae Sua*.

4.3.2. The laser capture microdissection

4.3.2.1. The determination of the effect of Giemsa staining on the PCR reaction

After making the chromosome preparation, in order to visualize the chromosomes and banding patterns on membrane slides, the polytene chromosome of *A. gambiae Sua* were stained with Giemsa and compared with the chromosomes without staining. **Figure 5.10** shows the difference of banding patterns of chromosomes stained and without staining. After the staining chromosome with Giemsa, we could find chromosomes on membrane slides easily but the banding patterns were not as clear as the chromosome without staining. And additionally, we needed to further confirm the effect of Giemsa on the following application. To do so, 30 chromosome pieces spanning the distal and proximal breakpoints of 2La/+ (heterozygous 2L) were microdissected using PALM UV laser microdissection system (**Figure 5.11**) from the membrane slides for each group (stained and without staining), respectively. After the

microdissection, the chromosome pieces were collected in the tube cap with Catapult buffer. After centrifuge at 14000 rcf for 15 minutes, the tube was digested at 55°C overnight and then inactivated at 90 °C for 10 min. After digestion, all the material was taken to do PCR with 2La diagnostic primers (23A2, 27A2 and Dpcross5) using the following condition: 95℃ for 7 min; 35 cycles of 94℃ for 30 s, 60℃ for 30 s, 72℃ for 30 s; 72℃ for 10 min. Two bands with 200 and 500 bp can be obtained from successful amplification of 2La/+. **Figure 5.12** show that Giemsa inhibits PCR reaction. Therefore, Giemsa staining polytene chromosomes is not feasible in present study since it might inhibit the following mass amplification.

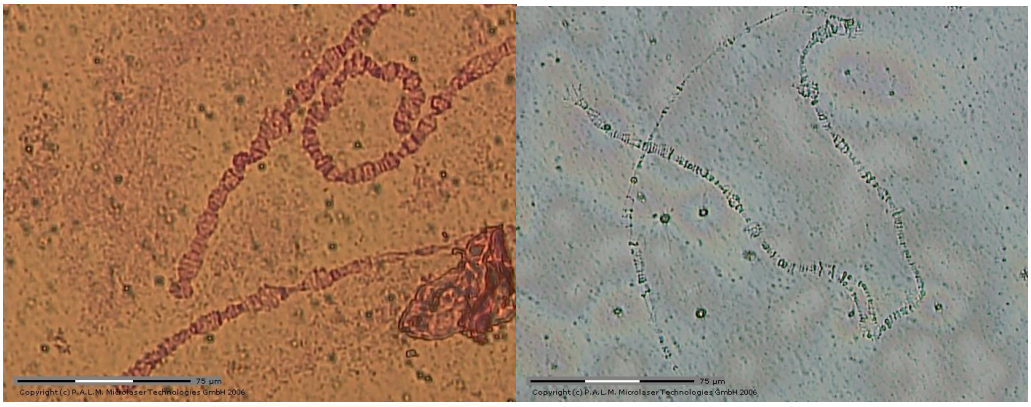


Figure 5.10: The 2La chromosome of *A. gambiae Sua* stained with Giemsa and 2La from *A. merus* without staining. Left image shows that 2La chromosome stained with Giemsa. Right image indicates the 2La chromosome without Giemsa staining.

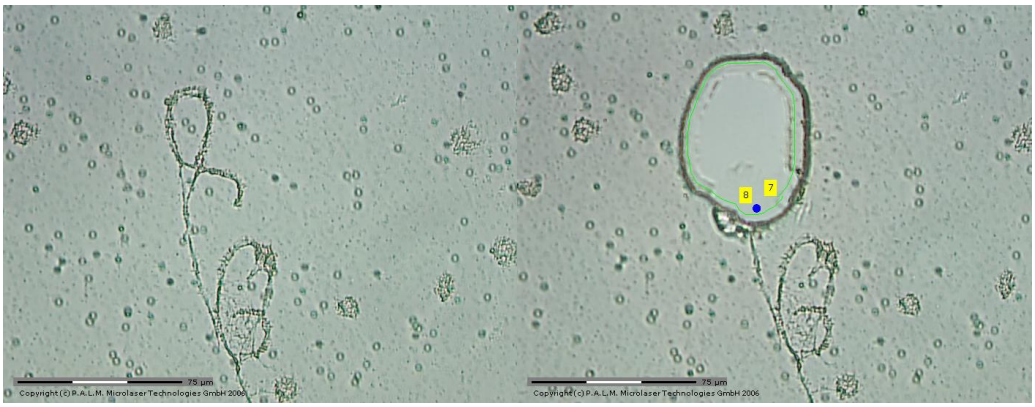


Figure 5.11: The microdissection of chromosome fragments from *A. gambiae* 2La/+

(2L heterozygous). Left panel: the chromosome preparation before microdissection. Right panel: the chromosome preparation after laser microdissection.

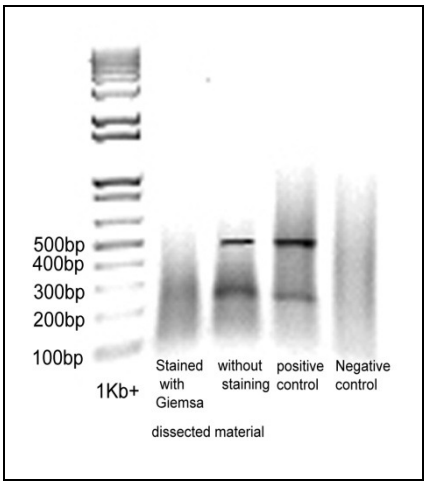


Figure 5.12: The amplification of microdissected DNA material from *A. gambiae Sua* using 2La diagnostic PCR primers. Positive control is the PCR product of genomic DNA from one ovary of *A. gambiae Sua*. Negative control used water

4.3.2.2. Laser capture microdissection of chromosomal regions spanning the breakpoints

Freshly made chromosome preparations without staining were used for laser microdissection and the specific chromosomal regions cover two breakpoints of 2La arrangement were cut and captured in the caps of tubes. Microdissection was carried out using 40 × objectives to obtain optimal banding patterns and the length of the chromosomal regions microdissected in field. **Figure 5.13** and **5.14** showed two examples of laser microdissection of distal and proximal breakpoints of 2La in *A. merus*. All the samples were performed consistently as these two breakpoints. After microdissection, chromosomal pieces were spun down into the tube at 14000 rcf for 15 minutes. All microdissected DNA material were then digested in Catapult buffer at 55^{o}C

overnight and then inactivated at 90 °C for 10 min.

Figure 5.13: The laser microdissection of distal breakpoint of 2La inversion in *A. merus*. Left panel shows the breakpoint region was marked base on the banding pattern. Middle panel indicates the performance of laser microdissection. Right panel revealed the chromosomal piece was captured into the cap of tube.

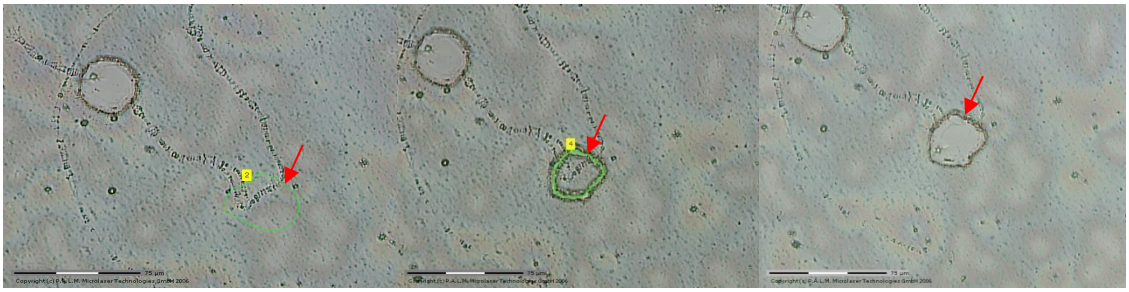


Figure 5.14: The laser microdissection of proximal breakpoint of 2La inversion in *A. merus*. Left panel shows the breakpoint region was marked base on the banding pattern. Middle panel indicates the performance of laser microdissection. Right panel revealed the chromosomal piece was captured into the cap of tube.

4.3.3. The mass amplification using GenomiPhi V2 method

4.3.3.1. The determination of the effect of Catapult buffer components on the mass amplification

In order to determine how the components of Catapult Buffer affect the following amplification. A set of different Catapult Buffers were made and 2 µl of different Catapult Buffers were added to the amplification reactions. The template DNA is the genomic DNA isolated from one *A. gambiae* Sua mosquito. The 5 µl of 22 µl amplified DNAs

were loaded on the 1% agarose gel. **Figure 5.15** shows Igepal CA-630 inhibit the amplification reaction. Igepal is nonionic detergent which might inhibit the activity of Phi 29 enzyme.

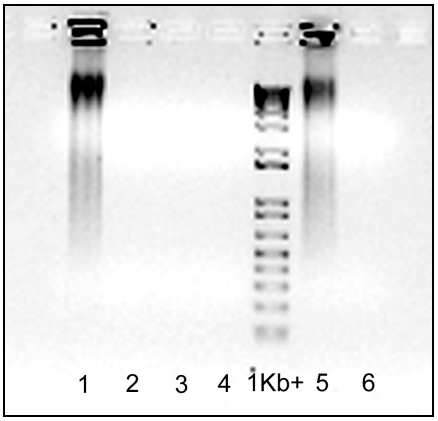


Figure 5.15: The effect of the different components of Catapult Buffer on the V2 DNA amplification. **1:** 10ng of genomic DNA of *A. gambiae Sua* was used as template and no Catapult Buffer added. **2:** 2 µl of Catapult Buffer without Proteinase K was added into amplification (0.5 M EDTA PH 8.0, 1 M Tris PH 8.0, Igepal CA-630). **3:** 2 µl of Catapult Buffer without 0.5 M EDTA PH 8.0 was added into amplification (1 M Tris PH 8.0, Igepal CA-630, Proteinase K 20 mg/ml). **4:** 2 µl of Catapult Buffer without 1 M Tris PH 8.0 added (0.5 M EDTA PH 8.0, Igepal CA-630, Proteinase K 20 mg/ml). **5:** 2 µl of Catapult Buffer without Igepal CA-630 added (0.5 M EDTA PH 8.0, 1 M Tris PH 8.0, Proteinase K 20 mg/ml). **6:** Negative control. Pure water was used as template and no Catapult Buffer. For the samples 3, 4 and 5, Proteinase K was inactivated at 65°C for 10 minutes before adding to the amplification reaction.

4.3.3.2. The comparison of heat and chemical denaturation in V2 DNA amplification

To denature the DNA samples, two methods: heat and chemical were provided in the protocol of the amplification kit. According to our experience, it is likely that chemical denaturation is better than heat denaturation for our DNA samples since it can avoid the damage on our DNA. The comparison of effect of two different denaturation methods on

the yield of V2 DNA amplification has been attempted. Using the 10 ng of genomic DNA from A. *gambiae* Sua mosquito as template DNAs, amplification with two methods were performed and 5 µl of 20 µl of amplicons was loaded on the 1% agarose gel. **Figure 5.16** shows that heat denaturation yielded more amplification products than chemical denaturation. However, chemical denaturation will be recommended for the amplification of small amount of DNAs to minimize the DNA shearing and damage.

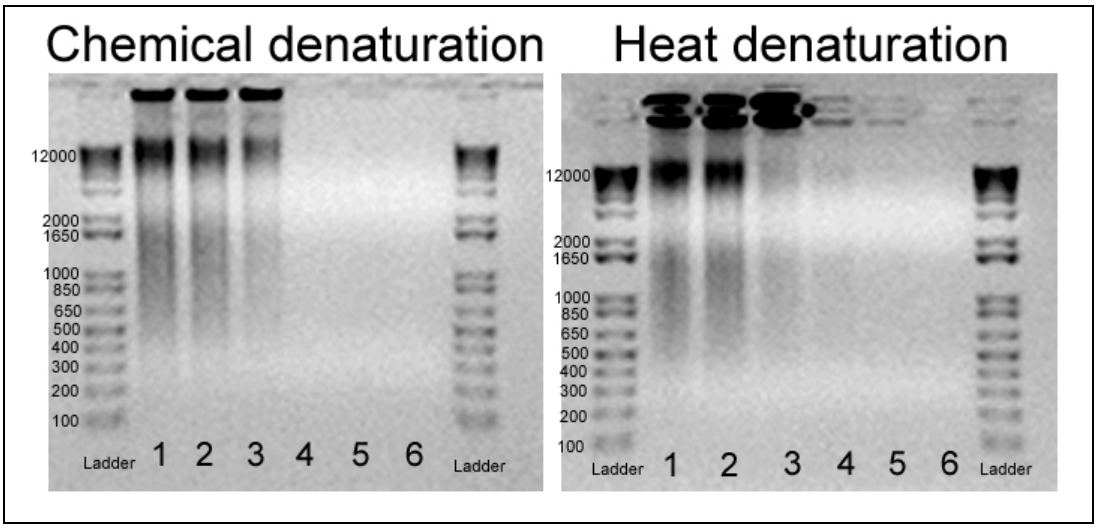


Figure 5.16: The amplified DNA samples with different denaturation methods and treated with Trehalose. Left image: the amplified DNAs using chemical denaturation method of V2 DNA amplification. Right image: the amplified DNAs with heat denaturation method of V2 DNA amplification.1: The amplification of 10ng control DNA from kit. 2: The amplification of 10ng *A. gambiae* genomic DNA. 3: The amplification of 15 µl of pure PCR grade water. 4: Amplification of 10ng control DNA from kit treated with 13µl of 2 M Trehalose (0.74 M final concentration). 5: The amplification of 10ng *A. gambiae Sua* genomic DNA treated with 13µl of 2 M Trehalose (0.74 M final concentration).

4.3.3.3. The amplification of DNA samples spanning breakpoints of 2La

arrangement using V2 DNA amplification kit

Nineteen distal and twenty proximal breakpoints of 2La inversion of *A. arabiensis* were laser microdissected and catapult captured into the adhesive caps of microdissection tubes. Under the microscope, the microdissected chromosome pieces were visible in the adhesive cap (**Figure 5.17**). 50 µl of Catapult Buffer without Igepal was added into the upside down adhesive cap for each sample and incubate at 55°C for 9 hours and then centrifuge at low speed briefly (High speed is not applicable for adhesive tube). The solution in the tube was transferred to one autoclaved PCR tube. Another 50 µl of Catapult Buffer without Igepal was added into the upside down adhesive cap again and incubate at 55°C for another 9 hours. Low speed centrifuge to spin down the solution and combine it to the previously collected sample. The collected samples were heated to 95°C for 5 minutes to inactivate Proteinase K. For each total 100µl of digested DNA sample, 10µl of 3M NaAC and 250µl of 100% Ethanol were added to the tube, and then mixed by inverting tubes. The sample was centrifuged at 14,000rcf for 5 minutes to collect pellet. Supernatant was removed and 500µl of 70% Ethanol was added for additional washing. The pellet was dried in a decontaminated container until turning into white powder. 2µl of water was added to dissolve DNA which was used as template DNA for amplification using GenomiPhi V2 DNA amplification methods. After amplification, the amplicons were purified using QIAEXII kit (Qiagen), which is optimal for the purification of product > 4Kb and the concentration of amplicons was measured by NanaDrop 2000 (**Table 5.3**).

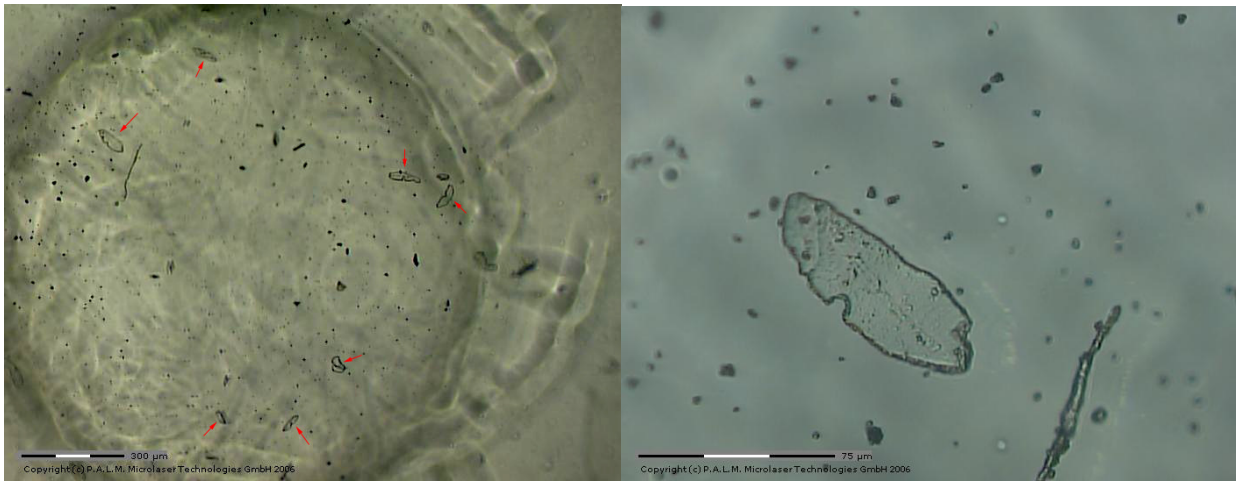


Figure 5.17: Visualization of laser microdissected chromosome pieces under microdissection microscope. Left panel is the image under 5 × objectives. Red arrow indicates the microdissected proximal breakpoints of *A. arabiensis*. Right panel is the image under 40 × objectives.

Table 5.3: The concentrations of amplicons of breakpoint DNAs from *A. arabiensis*

Sample ID	**Concentration (ng/µl)**	**Total volume (µl)**	**Total amount of DNA (ng)**
Amplicons of 19 distal 2La breakpoint DNA (DA19)	2.5	50	125
Amplicons of 20 proximal 2La breakpoint DNA(PA20)	3.5	50	175
Amplicons of positive control (10ng DNA)	3.1	50	155
Negative control (H_2O) (Negative)	2.8	50	140

Table 5.3 shows that DNA digested from the chromosomal regions span both of breakpoints in *A. arabiensis* were amplified. But V2 DNA amplification kit results in more than 90% of TIPs compared with result on positive control. It also suggests that to obtain at least 1 microgram of DNA for sequencing, about 154 chromosomal pieces span the distal breakpoint of 2La inversion, 114 for proximal breakpoints are required if V2 DNA amplification method is applied.

To determine the dropout rate and confirm that our target DNA was amplified, 2ng of amplicons by V2 DNA amplification kit was used as template DNA and 12 primers were used to do PCR amplification. There are five primers from distal breakpoint (7073, 7077, 7078, 7086 and 5779), five from proximal breakpoint (5774, 5778, 6898, 7019, and 7068) of 2la inversion in *A. arabiensis* as well as two primers (3322 and 1981) on 2R chromosome. After the PCR amplification, PCR product for each sample was loaded on 1% agarose gel. **Figure 5.18** and **5.19** show that cDNA: 7068 and 5779 were amplified from amplified DNA of distal breakpoint of 2La and 5778, 6898, 7019 and 7068 were amplified from amplified DNA of 2La proximal breakpoint since right PCR product size are between 500-650bp while other nonspecific markers yielded no PCR product. PCR bands of 7068, 5778, 6898, 7019 and 7068 were cut from gel and purified using Qiaquick Gel Extraction Kit (Qiagen). The concentration of purified DNA samples were measured by NanoDrop 2000 and listed in **Table 5.4**. 100ng of purified DNA for each sample was used for sequencing. Due to the low amount of DNA for 6898+PA20, this sample failed for sequencing. The analysis of sequences of the rest four DNA samples was present in **Table 5.5** which suggest that cDNA markers were successfully and correctly amplified however the amplification using V2 DNA amplification kit strongly implies that some loci showed preferential amplification (*367*). Dropout rate which was defined as the proportion of unsuccessful amplification of markers in total markers is 40% (4/10).

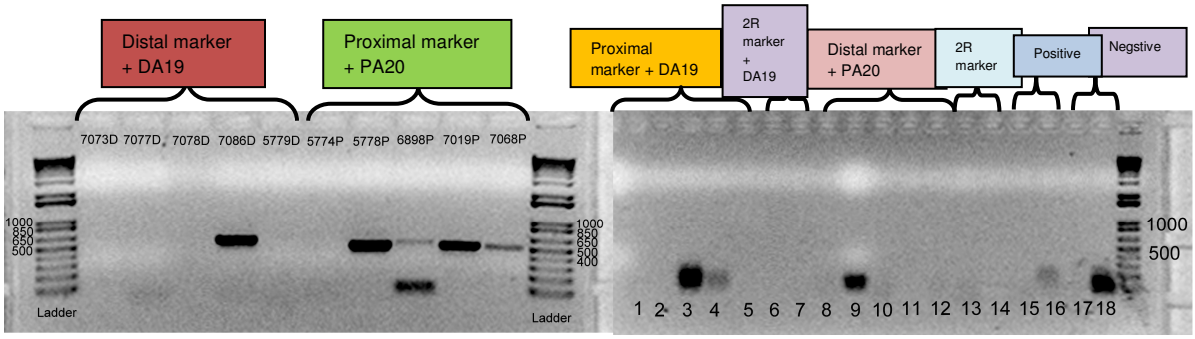


Figure 5.18: The amplification results with specific primers and DA19 and PA20 as template DNA. Left panel: (1) 7073+DA19 (2) 7077+DA19 (3)7078+DA19 (4) 7086+DA19 (5) 5779+DA19 (6) 5774+PA20 (7) 5778+PA20 (8) 6898+PA20 (9) 7019+PA20 (10) 7068+PA20 Right Panel: (1) 5774+DA19 (2) 5778+DA19 (3) 6898+DA19 (4) 7019+DA19 (5) 7068+DA19 (6) 3322+DA19 (7) 1981+DA19 (8) 7073+PA20 (9) 7077 +PA20 (10) 7078+PA20 (11) 7086+PA20 (12) 5779+PA20 (13) 3322+PA20 (14) 1981+PA20 (15) 5774+ *gambiae* gDNA (16) 7073+*gambiae* gDNA (17) 5774+Negative control (18) 7073+Negative control

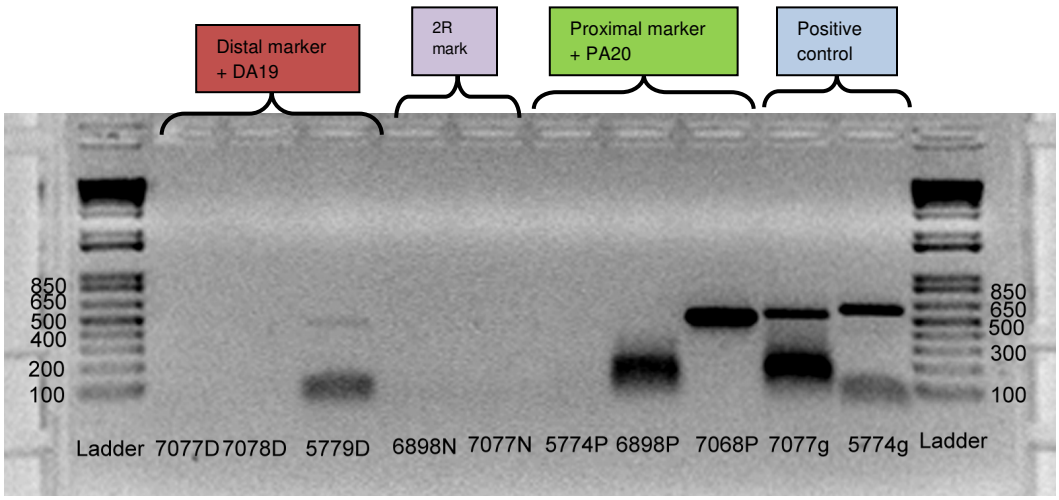


Figure 5.19: The amplification results with specific primers and DA19 and PA20 as template DNA (continued).

(1) 7077+DA19 (2) 7078+DA19 (3) 5779+DA19 (4) 6898+Negative control (5) 7077+Negative control (6) 5774+PA20 (7) 6898+PA20 (8) 7068+PA20 (9) 7077+*gambiae* gDNA (10) 5574+*gambiae* gDNA.

Table 5.4: The concentrations of purified PCR products

Sample ID	Concentration (ng/µl)	Total volume (µl)	Total amount of DNA (ng)
7086+DA19	18.3	50	915
5778+PA20	14.7	50	735
6898+PA20	2.2	50	110
7019+PA20	9.4	50	470
7068+PA20	2.9	50	145

Table 5.5: The sequence analysis of the PCR products of cDNA markers

Sample ID	Genomic location in *A. gambiae* PEST strain		Sequence Blast result in *A. gambiae* PEST strain				
	Coordinate in *A. gambiae*	Chromosomal location	Coordinate in *A. gambiae*	Chromosomal location	e-value	Length (bp)	Identity (%)
5778PA20	20,521,764 - 20,523,605	2L:22F	20,522,147-20,522,590	2L:22F	0	451	98.67
6898PA20	39,574,075 - 39,578,295	2L:26B	39,576,168-39,576,670	2L:26B	0	511	98.04
			39,574,997-39,575,073	2L:26B	5e-07	77	84.42
			39,574,835-39,574,906	2L:26B	2e-06	72	84.72
7019PA20	40,995,291 - 40,999,389	2L:26D	40,997,301-40,997,764	2L:26D	0	464	97.2
7086DA19	42,327,399 - 42,406,342	2L:27A	42,397,370-42,397,879	2L:27A	0	510	98.24

4.3.3.4. Modified V2 DNA amplification method

A recent study suggested that trehalose with appropriate concentration (0.54–0.84 M final concentration) can suppress formation of TIPs (*367*). We hypothesized that the addition of trehalose into GenomiPhi V2 DNA amplification could reduce the TIPs. 13µl of 2M trehalose (0.74 M final concentration) was applied to the amplification reaction. **Figure 5.16** reveals that incubation of amplification reaction treated with trehalose for 2 hours yields small amount of amplicons and almost invisible on the 1% agarose gel. To

modify the V2 DNA amplification protocol, incubation period was extended to 16 hours. A series of different amplification reactions were set up and 5µl of amplified DNAs were loaded on the gel. **Figure 5.20** reveals that if we extend the incubation time to 16 hours, all our DNA samples were amplified even treated with appropriate concentration of trehalose. However, TIPs are still abundant in these DNA samples. Addition of Trehalose to the amplification reactions does not suppress the TIPs as we expected. In addition, our results also indicate that no significant amplicon difference could be observed by using heat and chemical denaturation.

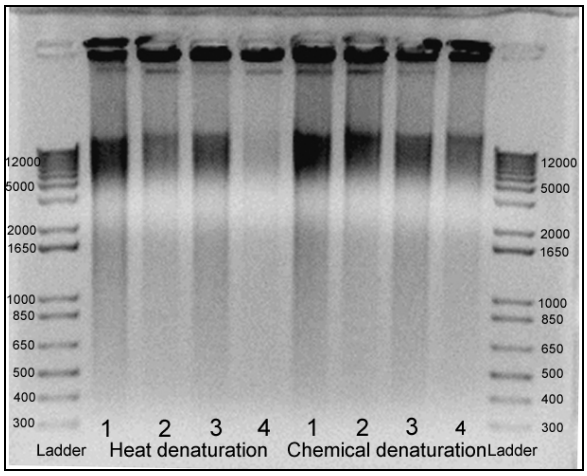


Figure 5.20: The amplified DNA from different samples using modified V2 DNA amplification method (incubation for 16 hours and trehalose was applied). The left four DNA samples were denatured by heat and the right four by chemical. **1:** 10ng *A. gambiae Sua* genomic DNA. **2:** 14µl DNase, RNase and Protease free water. **3:** 10ng genomic DNA + 13µl Trehalose. **4:** 1µl DNase, RNase and Protease free water + 13µl Trehalose

In addition to the genomic DNA, microdissected DNA samples were also amplified using modified V2 DNA amplification kit. 34 distal and 24 proximal breakpoint chromosome pieces from *A. merus* were applied to modified V2 DNA amplification

reactions and the concentration of purified amplicons were determined by NanoDrop 2000 and listed in **Table 5.6**. Our result suggest that the addition of trehalose into the amplification reaction can dramatically reduce TIPs from 90% (2.8/3.1 Table 4) to 33% (29.3/89.2) if start template DNA is 10ng. However, if the template DNA is picrogram amount or less, TIPs are still abundant. Despite abundant of TIPs, however the amplicons of PM24 can reach microgram (1.015ug) level which is feasible for sequencing by Illumina method. Additionally, to further demonstrate whether modified method can suppress amplification bias. PCR procedures were performed for 29 cDNA markers located near the proximal breakpoint of 2La in *A. merus* using 3-4ng of PM24 as template. **Figure 5.21** reveals that 10 of 29 markers were amplified and dropout rate is about 65.5%. Thus, significant locus and alleles biases reduce the quality of V2 amplicons (*367*).

Table 5.6: The concentrations of amplicons of breakpoint DNAs from *A. merus*

Sample ID	**Concentration (ng/µl)**	**Total volume (µl)**	**Total amount of DNA (ng)**
Amplicons of 34 distal 2La breakpoint (DM34)	1.5	50	75
Amplicons of 24 proximal 2La breakpoint (PM24)	20.3	50	1015
Amplicons of 10ng DNA (Positive)	89.2	50	4460
Amplicons of negative control (H_2O) (Negative)	29.3	50	1465

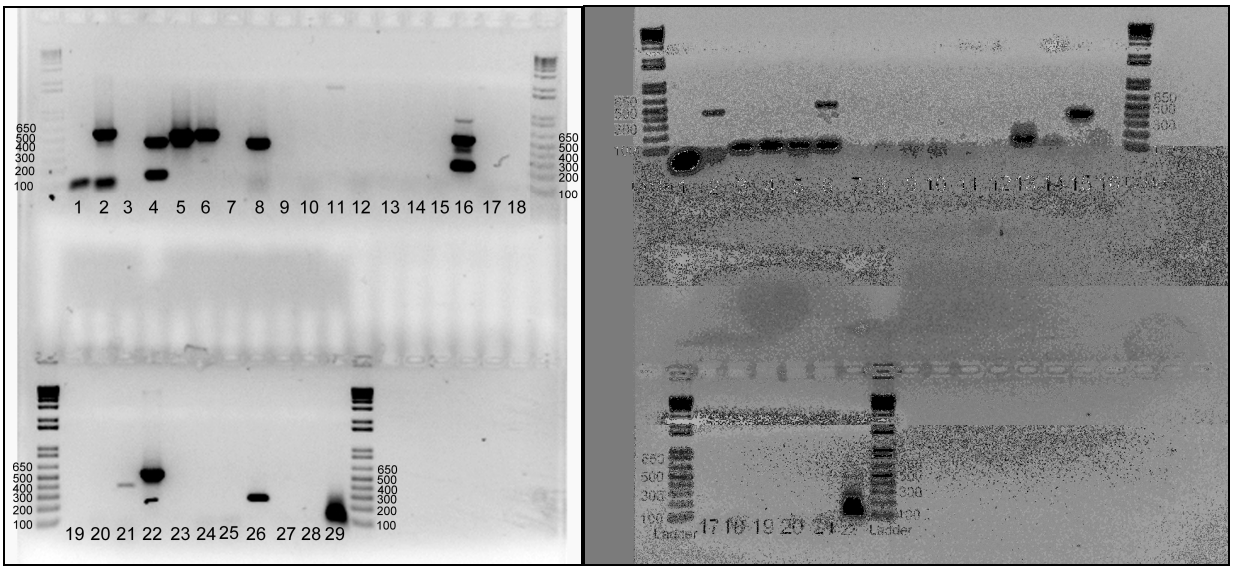


Figure 5.21: PCR amplification of 29 cDNA markers near proximal breakpoint of 2La in *A. merus* using PM24 as template. Left panel: (1) 5765 (2) 5771 (3) 5774 (4) 5775 (5) 5778 (6) 7068 (7) 7065 (8) 7063 (9) 7051 (10) 7046 (11) 7039 (12) 7031 (13) 7030 (14) 7023 (15) 7022 (16) 7019 (17) 7014 (18) 7008 (19) 7007 (20) 7006 (21) 6990 (22) 6974 (23) 6968 (24) 6965 (25) 6958 (26) 6945 (27) 6903 (28) 6900 (29) 6898
Right panel: (1) 5765 (2) 5774 (3) 7065 (4) 7051 (5) 7046 (6) 7039 (7) 7031 (8) 7030 (9) 7023 (10) 7022 (11) 7014 (12) 7008 (13) 7007 (14) 7006 (15) 6990 (16) 6968 (17) 6965 (18) 6958 (19) 6945 (20) 6903 (21) 6900 (22) 6898 Red highlight indicates that this marker is amplified by PCR protocol.

4.3.4. The amplification of DNA samples using Wpa (Whole-pool amplification) method

A new whole genome amplification method labeled as Wpa was developed in 2008 by Pan et al (*367*). This new approach provides highly specific (without TIP), unbiased and hypersensitive amplification of small amounts of DNA (0.5-2.5ng) or entire genome. 4 microdissected samples, 23 distal breakpoints in *A. merus* (DM23), 21 proximal of *A. merus* breakpoints (PM21), 6 distal *A. arabiensis* breakpoints (DA6), 6 proximal *A. arabiensis* breakpoints (PA6), were laser captured in adhesive tubes, digested and

precipitated using 100% Ethanol as above protocol. The purified amplicons of DA19 and PA20 with V2 DNA amplification and four samples were used as template for Wpa reactions and incubate at 30°C for 16 hours. 5 of 100µl reaction was load on gel and **Figure 5.22 (top)** show that despite the strong signal of 0.1 ng of human genomic DNA, however the amplicons of six samples were invisible on the gel as negative control. These samples were repeated under the same condition with extended 5 incubation hours (total 21 incubation hours). **Figure 5.22 (middle and bottom)** reveal that DA19 and PA20 yield strong signals but other samples have too less DNA to detect on the gel. The amplicons of five samples; B#2DA19, B#3PA20, B#5PM21, C#2DA19 and C#3PA20 were purified with QIAEXII kit (Qiagen). The concentration was determined by NanoDrop 2000 and **Table 5.7** show that although the amplicons of B#5PM21 is invisible on the gel image; total 610 ng of DNA was obtained.

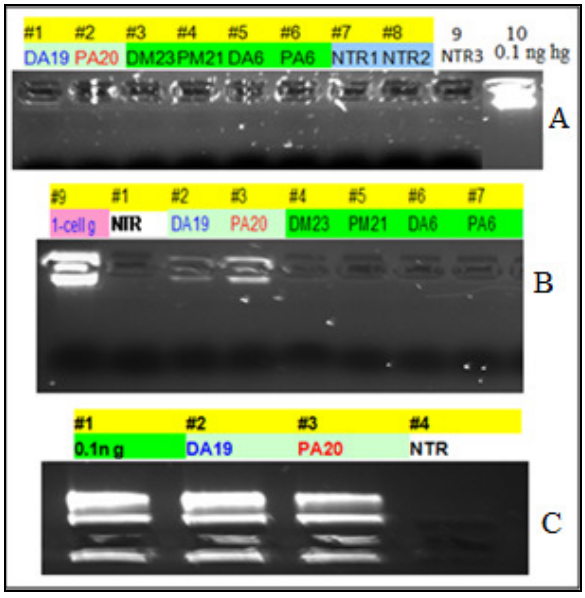


Figure 5.22: The visualization of amplified DNA samples with Wpa method on the gel. NTR: negative control; hg: human genomic DNA; 1-cell g: one fresh human cell DNA. The bottom 0.1ng indicates 0.1ng of human genomic DNA. **Top:** Amplification of

samples using Wpa method incubated at 30°C for 16 hours. **Middle:** Amplification of samples using Wpa method incubated at 30°C for 21 hours. **Bottom:** Amplification of samples using Wpa method incubated at 30°C for 21 hours.

Table 5.7: The concentration of amplicons of microdissected DNA samples with Wpa method

Sample ID	Concentration (ng/µl)	A260/280	Total Volume	Total amount of DNA (ng)
B#2DA19	56.7	1.83	100 µl	5670
B#3PA20	61.6	1.82	100 µl	6160
B#5PM21	6.1	1.82	100 µl	610
C#2DA19	86.2	1.86	100 µl	8620
C#3PA20	80.2	1.84	100 µl	8020

Figure 5.22 B also shows that about 10 µg of amplicons can be obtained from one fresh human cell which is equal to 6 pg of starting DNA. However, 21 chromosomal pieces spanning proximal breakpoint of *A. arabiensis* containing about 75 pg of DNA only yield 610 ng of amplicons. Therefore, we assumed that the DNA from microdissected polytene chromosomal regions has been seriously degraded or damaged. To improve the quality of our DNA samples, polytene chromosomal preparations were made from fresh dissected ovaries without pre-fixation with Carnoy's solution and microdissection was performed immediately. Two samples, 15 distal breakpoints and 12 proximal breakpoints of *A. arabiensis* (DA15 and PA12) were amplified using Wpa methods for two round runs. The amplicons were purified and the concentration was determined by NanoDrop. **Table 5.8** reveals that the DNA samples

prepared from fresh chromosomal preparation with all the procedures performed immediately can significantly increase the yield of amplification. Wpa method yield about 58% of TIPs (3.2/5.5) which is lower than that with V2 DNA amplification kit (>90%). After the primary amplification, 5µl of amplicons from each sample were used as input DNA for second round amplification together. **Table 5.8** shows that more than 10ug of amplicons were obtained after second amplification reactions from microdissected chromosomal regions, in the meanwhile, Wpa also yield abundant TIPs.

Table 5.8: The determination of total amount of amplicons from DA15 and PA12 using Wpa method by NanoDrop

Amplification	Cycle	Sample ID	Concentration (ng/µl)	Total Volume(µl)	Amount of DNA (ng)	Total(ng)
Primary amplification	First run	DA15	5.5	50	275	605
	Second run	DA15	6.6	50	330	
	First run	PA12	5.5	50	275	550
	Second run	PA12	5.5	50	275	
	First run	NTR	3.2	50	160	160
	Second run	NTR	NA			
Second amplification	First run	DA15	85.82	100	8582	15151
	Second run	DA15	65.69	100	6569	
	First run	PA12	94.46	100	9446	17109
	Second run	PA12	76.63	100	7663	
	First run	NTR	61.33	100	6133	6133

PCR reactions were performed using 2 ng of B#5PM21, B#2DA19 (DA192) and B#3PA20 (PA202) as template DNA. **Figure 5.23** shows that 3 (two proximal and one distal) of 10 markers were amplified from B#5PM21 but no DNA can be amplified for 2R marker. This suggests that distal and proximal DNA samples were cross contaminated for some reasons. The estimation of dropout rate is 7/10 (70%). **Figure 5.24** reveals

that in addition to the correctly amplified four markers, we also found that 5771 and 5778 near proximal breakpoints, 7086 near distal breakpoint of 2La are amplified from DA192 and PA202. This might indicate that our two samples (DA192 and PA202) were cross-contaminated. All the markers (3322 and 1981) on 2R chromosome yield no PCR products. To confirm that PCR products are our target DNA markers, seven samples were cut from gel, purified and sequenced. The sequence analysis (**Table 5.9)** confirmed that DNA samples spanning distal and proximal breakpoints of 2La in *A. arabiensis* were cross-contaminated. To further confirm that DA192 and PA202 don't contain DNA from other chromosomal arms, amplification of markers from other chromosomal arms (803 and 205 on X, 1980 and 3434 on 2R, 6263 on 3L, 7780 on 3R, 10387 and 10404 on 3L) using same template DNAs were repeated. **Figure 5.25** shows that no markers from 2R, 2L, 3R and X arm were amplified. Also if we considered DA192 and PA202 as one sample, 7 of 14 (50%) markers (distal markers: 7070, 7073, 7075, 7077, 7078, 7086, 5779; Proximal markers: 5765, 5771, 5774, 5775, 5778, 7065, 7068) near 2La breakpoints were amplified with specific primers.

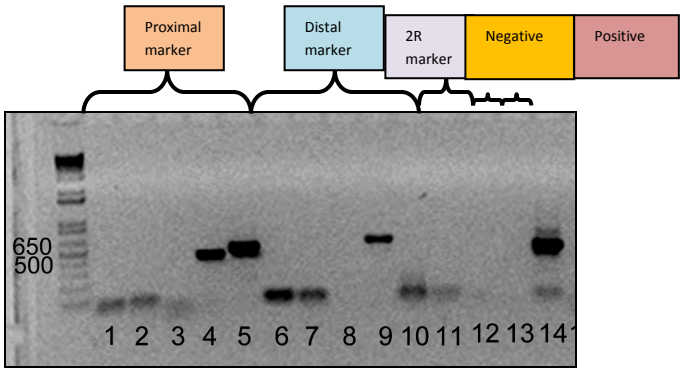


Figure 5.23: PCR amplification of cDNA markers near the proximal 2La breakpoint in *A. merus* using amplicons of PM21 as template DNA. (1) 5765 (2) 5771 (3) 5774 (4) 5775 (5) 5778 (6) 7073 (7) 7077 (8) 7078 (9) 7086 (10) 5779 (11) 3322 (12) 1981 (13) 5778+H_2O (14) 5778+Sua gDNA

Red highlight indicates that those markers are successfully amplified from amplicons of B#5PM21.

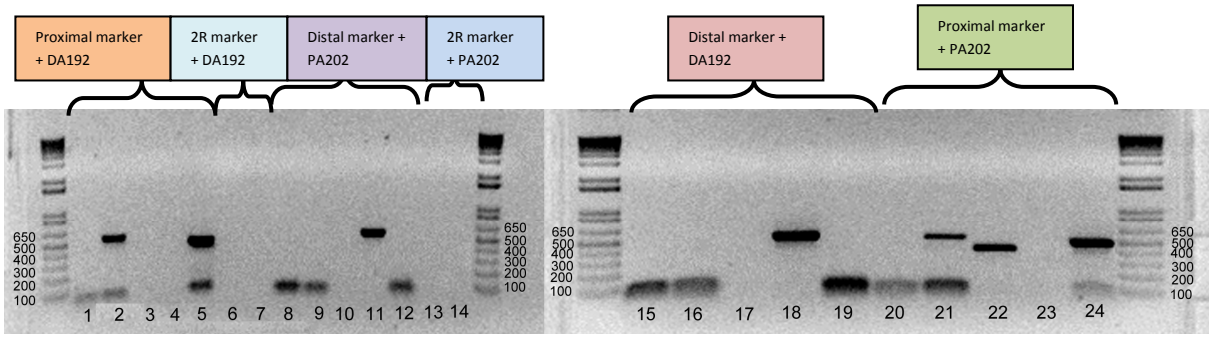


Figure 5.24: PCR amplification of cDNA markers near 2La breakpoints in *A. arabiensis* using DA192 and PA202 as template DNA. (1) 5765+DA192 (2) 5771+DA192 (3) 5774+DA192 (4) 5775+DA192 (5) 5778+DA192 (6) 3322+DA192 (7) 1981+DA192 (8) 7073+PA202 (9) 7077+PA202 (10) 7078+PA202 (11) 7086+PA202 (12) 5779+PA202 (13) 3322+PA202 (14) 1981+PA202 (15) 7073+DA192 (16) 7077+DA192 (17) 7078+DA192 (18) 7086+DA192 (19) 5779+DA192 (20) 5765+PA202 (21) 5771+PA202

(22) 5774+PA202 (23) 5775+PA202 (24) 5778+PA202 Red highlight indicates that these markers are amplified.

Table 5.9: Sequence analysis of PCR products with specific primers and DA192 and PA202 as template DNAs

Sample ID	Genomic location in *A. gambiae* PEST strain		Sequence Blast result in *A. gambiae* PEST strain				
	Coordinate in *A. gambiae*	Chromosomal location	Coordinate in *A. gambiae*	Chromosomal location	e-value	Length (bp)	Identity (%)
5774PA202	20,455,211-20,464,457	2L:22F	20,461,427-20,461,796	2L:22F	1e-172	598	94.86
5778+PA202	20,521,764-20,523,605	2L:22F	20,522,147-20,522,597	2L:22F	0	558	98.45
7086+DA192	42,327,399-42,406,342	2L:27A	42,397,370-42,397,868	2L:27A	0	539	98.2
5778+DA192	20,521,764-20,523,605	2L:22F	20,522,136-20,522,597	2L:22F	0	527	98.05
5771+DA192	20,415,195-20,425,711	2L:22F	20,422,904-20,423,419	2L:22F	0	526	98.84
7086+PA202	42,327,399-42,406,342	2L:27A	42,397,370-42,397,873	2L:27A	0	537	98.21

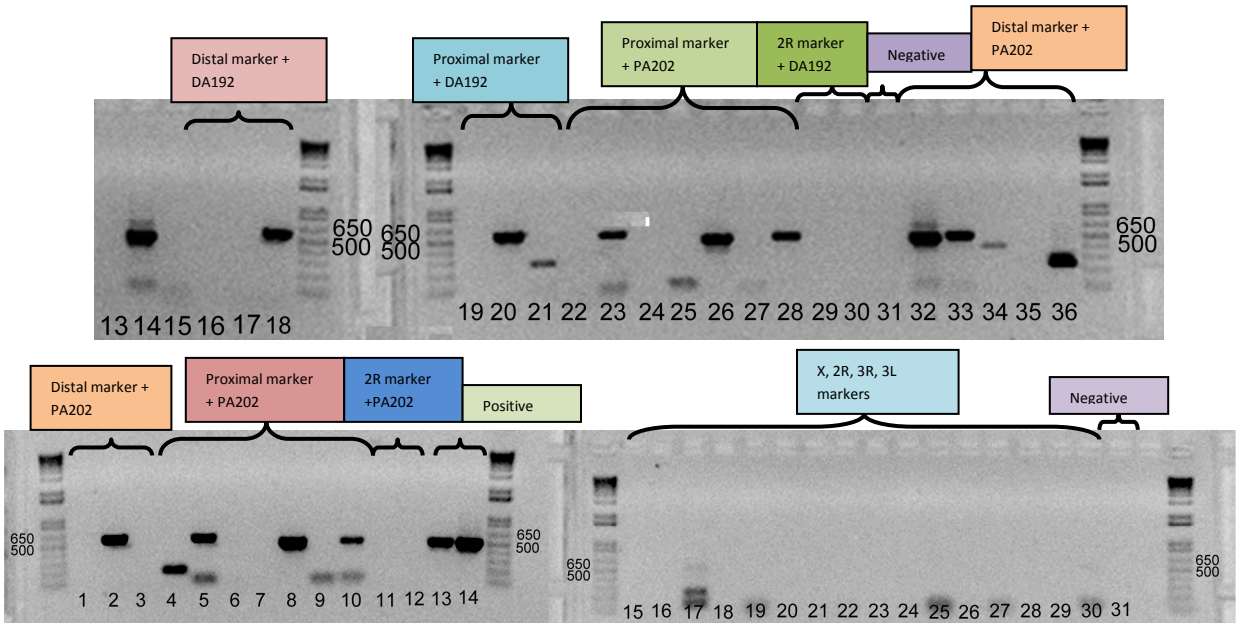


Figure 5.25: PCR amplification of markers near 2La breakpoints and markers on 2R, 2L, 3R, 3L and X using DA192 and PA202 as template DNAs.

Top panel: (13) 5778+H_2O (14) 5778+Sua gDNA (15)7070+DA192 (16)7073+DA192 (17)7075+DA192 (18)7077+DA192 (19)7078+DA192 (20) 7086+DA192 (21)5779+DA192 (22)5765+DA192 (23)5771+DA192 (24)5774+DA192 (25)5775+DA192 (26)5778+DA192 (27) 7065+DA192 (28) 7068+DA192 (29) 3322+DA192 (30)1981+DA192 (31) 5778+H_2O (32) 5778+Sua gDNA (33) 7070+PA202 (34) 7073+PA202 (35) 7075+PA202 (36) 7077+PA202
Bottom panel: (1) 7078+PA202 (2) 7086+PA202 (3) 5779+PA202 (4) 5765+PA202 (5) 5771+PA202 (6) 5774+PA202 (7) 5775+PA202 (8)5778+PA202 (9)7065+PA202 (10)7068+PA202 (11)3322+PA202 (12) 1981+PA202 (13) 7077+Sua gDNA (14) 7068+Sua gDNA (15) 803+DA192 (16) 205+DA192 (17) 1980+DA192 (18) 3434+DA192 (19) 6263+DA192 (20) 7780+DA192 (21) 10387+DA192 (22) 10404+DA192 (23) 803+PA202 (24) 205+PA202 (25) 1980+PA202 (26) 3434+PA202 (27) 6263+PA202 (28) 7780+PA202 (29)10387+PA202 (30) 10404+PA202 (31) 5778+H_2O.Red highlight indicates that the marker was amplified by PCR.

Primary amplicons of DA15 (DA151) and PA12 (PA121), and the secondary amplicons of DA15 (DA152) and PA12 (PA122) were also used as template DNA to do PCR. **Figure 5.26** shows that markers from distal and proximal breakpoints were amplified from both distal and proximal DNA samples but no marker on other chromosomal arms was amplified. Total 9 of 14 markers were amplified from DA151 and PA121 (64%). Amplification results with DA152 and PA122 are consistent with these of DA192 and PA202 (gel image not shown). PCR results double confirmed that no any markers from X, 2R, 3R, and 3L chromosomes can be amplified from DA152 and PA122. And total 7 of 14 markers were amplified from both samples (50%).

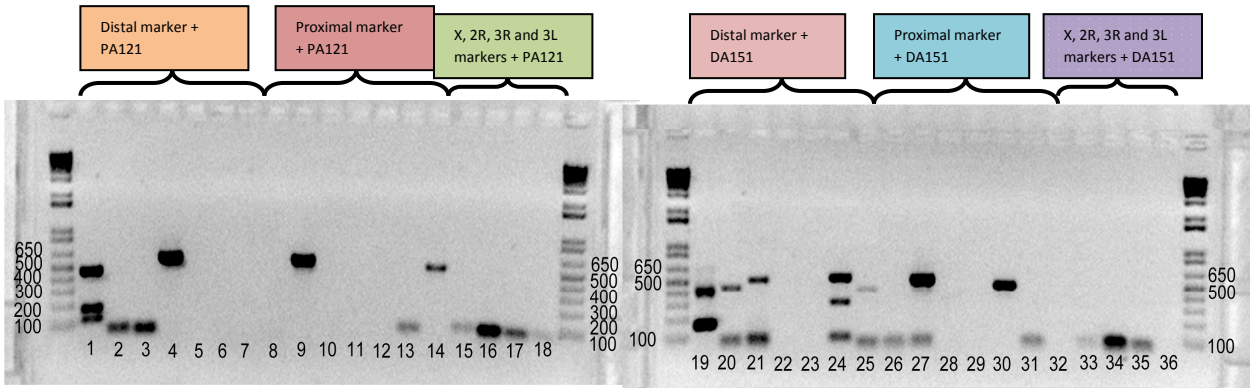


Figure 5.26: PCR amplification results using DA151 and PA121 as template DNAs. (1) 7070+PA121 (2) 7073+PA121 (3) 7075+PA121 (4) 7077+PA121 (5) 7078+PA121 (6) 7086+PA121 (7) 5779+PA121 (8) 5765+PA121 (9) 5771+PA121 (10) 5774+PA121 (11) 5775+PA121 (12) 5778+PA121 (13) 7065+PA121 (14) 7068+PA121 (15) 803+PA121 (16) 1980+PA121 (17) 6263+PA121 (18) 7780+PA121 (19) 7070+DA151 (20) 7073+DA151 (21) 7075+DA151 (22) 7077+DA151 (23) 7078+DA151 (24) 7086+DA151 (25) 5779+DA151 (26) 5765+DA151 (27) 5771+DA151 (28) 5774+DA151 (29) 5775+DA151 (30) 5778+DA151 (31) 7065+DA151 (32) 7068+DA151 (33) 803+DA151 (34) 1980+DA151 (35) 6263+DA151 (36) 7780+DA151 Red highlight indicated that these markers were amplified.

5. Discussion

The cross-mapping of PRI identified APL1-syntenic regions in the other species. The lack of APL1 signal against *A. stephensi* indicates that these genes are highly variable and polymorphic. The presence of a block of contiguous synteny on either side of APL1 in *A. stephensi* and *A. nili*, which is strong evidence that it is the correctly identified candidate region in *A. gambiae*. This also suggests functionality of this locus in other species. The *A. nili* results are completely new because the species is almost

untouched by molecular or genetic research. Future research should extract candidate genes and specifically test them by RNAi knockdown and *P. falciparum* infection.

Although, obtaining micrograms DNA for sequencing using WGA from microdissected polytene chromosomal regions or whole chromosomal arm is a great challenge, this is a powerful approach for characterization of inversion breakpoints and genomic analysis of specific chromosomal regions. The microdissected polytene and metaphase chromosomes have been used as template DNA for conventional PCR (*301, 302, 436, 437*). However, the first application of whole genome amplification methods to microdissected metaphase chromosomes was described in 2009 by Hockner et al (*356*). In this study, 10 metaphase chromosomes which is equal about one fourth of one human cell (1.5 pg) were microdissected and amplified by four different amplification kits including GenomiPhi DNA amplification kit (Amersham Biosciences), REPL1-g Midi kit (Qiagen), GenomePlex Single Cell kit(Sigma-Aldrich) and Improved Primer Extension Preamplification PCR (*356*). Their results revealed that the negative control was a smear with GenomePhi kit (*356*) which is consistent with about 70-90% of TIPs of total yield in our study. The WGA drop-out rate was 87.5% which is higher than 40%-65.5% in present project. The evidence suggests that when microdissected DNA samples from metaphase or polytene chromosomal regions were used as template of WGA, TIPs are abundant and it yield poor genotyping performance. This observation is well agreed with the results of WGA from single cell (*307, 355, 365, 366, 369, 438*). To date, no sequencing has been performed on microdissected and whole-genome-amplified samples.

In several reported single-cell research, MDA allows successful amplification of micrograms (1-40 μg) of DNA from a single cell (*355, 369, 438, 439*). The variation was suggested due to the different commercial kits (*369*). Although there are large amounts of TIPs and DNA yields from these studies will not reflect the specific amplification from the template, our research and Hockner et al still failed to obtain several micrograms DNA from a microdissected chromosome if amplification reaction was performed once. Even the most updated Wpa method which can acquire about 10 microgram DNA from one fresh single cell only yields about 500-600 ng of amplicons from one amplification reaction (**Figure 5.22**). If our microdissected DNA samples are long and intact, we should expect about 40-90 μg of amplicons using Wpa method from 20 microdissected 1.8-3.7 Mb chromosomal regions. The failure to amplify this amount DNA from our samples might be caused by DNA degradation and damage resulting from several reasons. (1) Carnoy' solution routinely used for ovary fixation may damage DNA. (2) During storage, nuclease activity might increase due to temperature change or other unidentified technical problem. It is believed that DNA is expected to be stable for years in the absence of nuclease (*355*). Frumkin et al reported that pre-stored cell (freeze at -80°C for several weeks prior to WGA) yield poorer amplification results than fresh cell (*355*). (3) Chromosomes are damaged during the process of chromosomal preparation. (4) Nucleic acid may result in the partial destruction during laser microdissection since the chromosomes are subjected to heat as well as photons from the laser itself (*281, 282*). Therefore, in order to reduce the DNA degradation and damage, we used fresh chromosomal preparation from fresh ovaries instead of fixed ovaries in Carnoy's solution and to perform all the procedures in shortest time. Our results demonstrate that

freshly microdissected chromosomal preparation yield as almost twice of DNA amplicons as those from fixed ovaries and the number of markers amplified by PCR increased from 30% to 64%. This suggested that using fresh experimental material and to perform all the procedure in the shortest time significantly improved the quality of our DNA samples.

In present study, one commercial kit, GenomePhi V2 DNA amplification kit was compared with the most updated Wpa method. Our evidence suggests that Wpa method yield less TIPs and more amplicons than V2 amplification method. However, Wpa did not improved amplification bias. Using primarily amplified DNA as input for the second round amplification, more than 10 μg of DNA was obtained from each microdissected DNA sample. This amount is suitable for direct sequencing.

6. Conclusions

1) The cross-mapping of PRI identified APL1-syntenic regions in *A. stephensi* and *A. nili.*
2) A protocol for multiple displacement amplification from laser capture microdissected polytene chromosomes of *Anopheles* mosquitoes was developed.
3) The successful amplification of our target DNA was confirmed by PCR with specific primers followed by sequencing analysis.
4) Wpa method yields less TIPs from laser microdissected polytene chromosomes than GenomiPhi V2 DNA amplification method. Also, the quality of DNA sample from freshly made chromosomal preparation is better than that from fixed ovaries.
5) By using the second round amplification, more than 10ug of DNA was obtained from microdissected polytene chromosomes.

Chapter six: Summary of the results and future directions

1. Summary of the results

1): A relative high resolution physical map has been developed for *A. stephensi*, a major Asian human malaria vector. Total 422 probes have been mapped to the standard cytogenetic map of *A. stephensi*. Therefore, the resolution of our current physical map is about 600 kb per marker which is only second to it of *A. gambiae*. The development of physical map for *A. stephensi* will help the genome assembly of this species and to study chromosomal evolution as well as evolutionary forces responsible for inversion generation and fixation.

2): The gene orders among species have been reshuffled by chromosomal rearrangements such as transposition, translocation, and inversions during the evolution. The study of the chromosomal rearrangements revealed that no interchromosomal transposition has been found in the lineages of *A. gambiae* and *A. stephensi*. This is consistent with the rare occurrence of transposition events in *Anopheles* mosquitoes and *Drosophila*. Seven cases of euchromatin-heterochromatin transitions have been identified between lineages which suggested that heterochromatin evolved faster than euchromatin regions.

3): The most popular type of chromosomal rearrangements in *Anopheles* mosquitoes is the paracentric inversions. The gene orders comparisons among three lineages of *A. gambiae*, *A. funestus* and *A. stephensi* demonstrated that sex chromosome evolve faster than any other autosomes. The identification of molecular features revealed that X chromosome is enriched with repetitive elements and has fewer genes. This

suggested that sex chromosome is more fragile than other autosome. The presence of high density of repetitive elements contributes to the fast evolution of this chromosome.

4): Among autosomes, the pattern of inversion fixation is positively correlated the content of polymorphic inversions. Our evidence suggested that polymorphic inversions can be driven to high frequency and to be fixed by local adaptation.

5): The faster evolved 2R among autosomes has higher density of microsatellite, satellite, segmental duplication and fewer M/SARs. These molecular factors might contribute together to the fragility of this chromosomal arm. The slow evolved autosomes such as 3R, 2L and 3L have higher density of M/SARs than 2R and X which hindered them to generate inversion breakpoints.

6): *A. gambiae* complex is a good model system to study the genetic change associated with the distinct geographic distribution, ecological adaptation and ability of malaria transmission. Our results revealed that 2La inversion is present in outgroup species, *A. stephensi* and *A. nili*. The carrier of 2La inversion, *A. gambiae*, *A. arabiensis* and A. *merus* can be considered closest to the ancestral species. The cumulated studies of 2La inversions demonstrated that 2La inversion is associated with the *Plasmodium* resistance and adaptation to the arid climate. Therefore, we hypothesized that Asian human malaria vector, *A. stephensi*, shared the same molecular mechanism with African mosquitoes. It is likely that the African mosquitoes acquired the *plasmodium* resistance and ability of adaptation to aridity from Asian malaria vector, such as *A. stephensi*. If this hypothesis is true, it will provide the value information to understand the molecular mechanism of the *plasmodium* resistance and ecological adaptation.

7): *A. gambiae* and *A. merus* can be differentiated by two inversions, o and p on 2R chromosome. Determination of the status of 2Rop has been attempted. The presence of 2Ro inversion in outgroup, *A. stephensi,* indicated that 2Ro inversion is ancestral. The analysis of molecular organization of both breakpoints of $2R+^{p}$ in *A. gambiae* revealed that the $2R+^{p}$ is the derived inversion. Therefore, 2Rp is an ancestral inversion. The ancestral status of 2Rop inversions lead us to consider that *A. merus* is closest to the ancestral species in *A. gambiae* complex. We hypothesize that *A. gambiae* and *A. arabiensis* were originated from *A. merus* by fixation of 2Rop and Xbcd inversions. Since both of species carried 2La inversion which is associated with adaptation to arid climates, the 2La inversion may enhance *A. gambiae* and *A. arabiensis* survival and have contributed to the wide distribution of both species in Africa. *A. gambiae* gradually acquired vectoral capacity by fixation of $2L+^{a}$ inversion, and *A. arabiensis* obtained this ability by gene flow from *A. gambiae* because of overlapped distribution of both species. Other members of *A. gambiae* complex were originated either from *A. gambiae* or *A. arabiensis* by fixation of $2L+^{a}$ and other inversions. Due to low tolerance of $2L+^{a}$ to aridity, these species only can occupy small areas in Africa. Although $2L+^{a}$ is strongly associated with malaria transmission, these species had low fitness to ecological adaptation, therefore they gradually lost vectorial capacity.

8): Accumulated evidence suggested that 2La inversion is associated with *Plasmodium* resistance. To understand the molecular mechanism of this resistance, we have developed a new protocol for obtaining the DNA sequence from the polytene chromosomal fragments spanning the 2La breakpoints. The chromosomal fragments were isolated by laser capture microdissection and then amplified by GenomePhi V2

DNA amplification and Wpa methods. The amplification of our target DNA was confirmed by conventional PCR procedure with specific primers followed by Sanger sequencing. Wpa method yielded significant more amount of amplicons, less TIPs than V2 DNA amplification kit. Using primary amplified DNA as input DNA, secondary amplification yields enough DNA for routine molecular analysis and sequencing. After amplification, two samples were sent for 454 sequencing. Our PCR results suggested that about 50% of cDNA markers were amplified from two samples, we hypothesized that our sequence will cover 50% of both breakpoints sequences of 2La inversion in *A. arabiensis*. To fill in another 50% of sequences gaps can be done by other methods such as long range PCR. The protocol developed in this study can be used to obtain DNA from any polytnene chromosomal fragments and also can be applied to the metaphase chromosomal fragments.

2. Future directions

1): To develop a high resolution physical maps for other *Anopheles* species and use them for genome sequence assembly. The current physical map for *A. stephensi* was successfully used for studying the patterns of inversion fixations and shedding light on the molecular mechanisms in the origin of inversion breakpoints in *Anopheles* mosquitoes. However, few maps are not enough for studying the molecular mechanism of generation of inversion breakpoints and identifying the hot spots of chromosomal inversions. Another direction could be to improve the annotation of gene sequences. To discover what genes were captured in synteny blocks and why it is so important to keep these genes clustered.

2): To sequence the breakpoint regions of alternative 2Rop arrangement in *A. gambiae* complex. Because *A. gambiae* complex was diverged recently and transposable elements should not be wiped out during evolution, the sequences on the breakpoint regions will help to discover the molecular mechanism in generating these inversions and this data will determine the ancestral status of 2Rop inversions. There is a large unknown sequence on the distal breakpoint of 2Ro in *A. stephensi*. To develop a protocol for obtaining this sequence is necessary and crucial to understand the driving force in the origin of 2Ro inversion. After identification of ancestral status of 2Rop inversions in *A. gambiae* complex, it is also important to locate the inversion breakpoints of five inversions (a, g, b, c, and d) on X chromosome and to sequence these breakpoints. The demonstration of ancestral status of five inversions on X and 2Rop inversions will determine the ancestral species or the closest ancestral species in *A. gambiae*. And then evolution direction will be clear in the *A. gambiae* complex.

3): To combine several amplification methods for obtaining DNA from both 2La breakpoints. Several studies suggested that the drop-out rates of amplification using GenomePlex Single Cell kit (Sigma-Aldrich) were less than using other amplification methods. Therefore, this method and GenoPhi V2 DNA amplification as well as Wpa can be combined to amplify our DNA samples. After sequencing, the gaps produced by Wpa method can be filled by other methods. Another direction could be to improve the quality of DNA samples microdissected from polytene chromosomal fragments. It is not clear whether our DNA before amplification were sheared and degraded or not. If our DNA was sheared into small fragments, a ligation will be needed before amplification.

Reference

1. I. V. Sharakhov *et al.*, *Science* **298**, 182 (Oct 4, 2002).
2. J. G. Breman, A. Egan, G. T. Keusch, *Am J Trop Med Hyg* **64**, iv (Jan-Feb, 2001).
3. M. T. Gillies, *Johannesburg: The South African Institute for Medical Research*, (1968).
4. HKrishnan, in *National Society of India for malaria and other mosquito-borne disease*. (1961).
5. A. V. Manouchehri, E. Javadian, N. Eshighy, M. Motabar, *Trop Geogr Med* **28**, 228 (Sep, 1976).
6. A. A. Enayati, H. Vatandoost, H. Ladonni, H. Townson, J. Hemingway, *Med Vet Entomol* **17**, 138 (Jun, 2003).
7. V. Gayathri, P. B. Murthy, *J Am Mosq Control Assoc* **22**, 678 (Dec, 2006).
8. I. F. Zhimulev, *Adv Genet* **34**, 1 (1996).
9. M. Coluzzi, and A. Sabatini, *Parassitologia* **9**, 5 (1967).
10. M. Coluzzi, and A. Sabatini, *Parassitologia* **10**, 9 (1968).
11. M. Coluzzi, and A. Sabatini, *Parassitologia* **11**, 11 (1969).
12. C. A. Green, *Ann Trop Med Parasitol* **66**, 143 (Mar, 1972).
13. M. Coluzzi, A. Sabatini, A. della Torre, M. A. Di Deco, V. Petrarca, *Science* **298**, 1415 (Nov 15, 2002).
14. C. Green, R. Hunt, *Genetica* **51**, 187 (1980).
15. G. P. Sharma, R. Parshad, S. L. Narang, J. B. Kitzmiller, *J Med Entomol* **6**, 68 (Jan, 1969).
16. M. Coluzzi, G. Cancrini, M. Di Deco, *Parassitologia* **12**, 101 (1970).
17. F. Mahmood, R. K. Sakai, *Can J Genet Cytol* **26**, 538 (Oct, 1984).
18. M. V. Sharakhova, A. Xia, S. I. McAlister, I. V. Sharakhov, *J Med Entomol* **43**, 861 (Sep, 2006).
19. R. D. Saunders, *Methods Mol Biol* **123**, 103 (2000).
20. L. Zheng *et al.*, *Proc Natl Acad Sci U S A* **88**, 11187 (Dec 15, 1991).
21. L. Zheng, M. Q. Benedict, A. J. Cornel, F. H. Collins, F. C. Kafatos, *Genetics* **143**, 941 (Jun, 1996).
22. G. Dimopoulos *et al.*, *Genetics* **143**, 953 (Jun, 1996).
23. A. della Torre, G. Favia, G. Mariotti, M. Coluzzi, K. D. Mathiopoulos, *Genetics* **143**, 1307 (Jul, 1996).
24. R. A. Holt *et al.*, *Science* **298**, 129 (Oct 4, 2002).
25. M. V. Sharakhova *et al.*, *Genome Biol* **8**, R5 (Jan 8, 2007).
26. I. Sharakhov *et al.*, *J Hered* **95**, 29 (Jan-Feb, 2004).
27. M. D. Gale, K. M. Devos, *Science* **282**, 656 (Oct 23, 1998).
28. S. J. O'Brien *et al.*, *Science* **286**, 458 (Oct 15, 1999).
29. C. Seoighe *et al.*, *Proc Natl Acad Sci U S A* **97**, 14433 (Dec 19, 2000).
30. J. M. Ranz, F. Casals, A. Ruiz, *Genome Res* **11**, 230 (Feb, 2001).
31. M. Papaceit, M. Aguade, C. Segarra, *Evolution* **60**, 768 (Apr, 2006).
32. A. Coghlan, E. E. Eichler, S. G. Oliver, A. H. Paterson, L. Stein, *Trends Genet* **21**, 673 (Dec, 2005).
33. J. Gonzalez, J. M. Ranz, A. Ruiz, *Genetics* **161**, 1137 (Jul, 2002).
34. J. Gonzalez, F. Casals, A. Ruiz, *Genetics* **175**, 167 (Jan, 2007).
35. J. M. Ranz, C. Segarra, A. Ruiz, *Genetics* **145**, 281 (Feb, 1997).
36. J. Vieira, C. P. Vieira, D. L. Hartl, E. R. Lozovskaya, *Chromosoma* **106**, 99 (Jul, 1997).
37. J. Vieira, C. P. Vieira, D. L. Hartl, E. R. Lozovskaya, *Genetics* **147**, 223 (Sep, 1997).
38. A. H. Sturtevant, E. Novitski, *Genetics* **26**, 517 (Sep, 1941).
39. J. T. Patterson, and W. S. Stone, *evolution in the genus Drosophila*. (Macmillan, New York, 1952).
40. M. L. Pardu, S. A. Gerbi, R. A. Eckhardt, J. G. Gall, *Chromosoma* **29**, 268 (1970).
41. M. Coluzzi, M. Di Deco, G. Cancrini, *Parassitologia* **15**, 129 (Apr-Aug, 1973).
42. M. Coluzzi, A. Sabatini, V. Petrarca, M. A. Di Deco, *Trans R Soc Trop Med Hyg* **73**, 483 (1979).
43. J. H. Nadeau, B. A. Taylor, *Proc Natl Acad Sci U S A* **81**, 814 (Feb, 1984).

44. J. H. Nadeau, D. Sankoff, *Trends Genet* **14**, 495 (Dec, 1998).
45. M. T. Davisson, P. A. Poorman, T. H. Roderick, M. J. Moses, *Cytogenet Cell Genet* **30**, 70 (1981).
46. B. J. B. Keats, *Linkage and chromosome mapping in man*. (University of Hawaii Press, Honolulu,HI, 1981).
47. P. L. Pearson, T. H. Roderick, M. T. Davisson, P. A. Lalley, S. J. O'Brien, *Cytogenet Cell Genet* **32**, 208 (1982).
48. J. E. Womack, *Isozymes Curr Top Biol Med Res* **6**, 207 (1982).
49. J. Wienberg, A. Jauch, R. Stanyon, T. Cremer, *Genomics* **8**, 347 (Oct, 1990).
50. H. Scherthan *et al.*, *Nat Genet* **6**, 342 (Apr, 1994).
51. J. Wienberg, R. Stanyon, *Curr Opin Genet Dev* **5**, 792 (Dec, 1995).
52. R. Glas, J. A. Marshall Graves, R. Toder, M. Ferguson-Smith, P. C. O'Brien, *Mamm Genome* **10**, 1115 (Nov, 1999).
53. A. Jauch *et al.*, *Proc Natl Acad Sci U S A* **89**, 8611 (Sep 15, 1992).
54. S. J. O'Brien *et al.*, *Science* **286**, 463 (Oct 15, 1999).
55. B. P. Chowdhary, T. Raudsepp, L. Fronicke, H. Scherthan, *Genome Res* **8**, 577 (Jun, 1998).
56. S. Wong, G. Butler, K. H. Wolfe, *Proc Natl Acad Sci U S A* **99**, 9272 (Jul 9, 2002).
57. B. Dujon *et al.*, *Nature* **430**, 35 (Jul 1, 2004).
58. A. Coghlan, K. H. Wolfe, *Genome Res* **12**, 857 (Jun, 2002).
59. A. Coghlan, *WormBook*, 1 (2005).
60. D. Sankoff, P. Trinh, *J Comput Biol* **12**, 812 (Jul-Aug, 2005).
61. A. Ruiz-Herrera, J. Castresana, T. J. Robinson, *Genome Biol* **7**, R115 (2006).
62. P. Pevzner, G. Tesler, *Proc Natl Acad Sci U S A* **100**, 7672 (Jun 24, 2003).
63. S. Richards *et al.*, *Genome Res* **15**, 1 (Jan, 2005).
64. A. Bhutkar *et al.*, *Genetics* **179**, 1657 (Jul, 2008).
65. M. Kellis, B. W. Birren, E. S. Lander, *Nature* **428**, 617 (Apr 8, 2004).
66. K. H. Wolfe, D. C. Shields, *Nature* **387**, 708 (Jun 12, 1997).
67. D. Lalo, S. Stettler, S. Mariotte, P. P. Slonimski, P. Thuriaux, *C R Acad Sci III* **316**, 367 (1993).
68. F. S. Dietrich *et al.*, *Science* **304**, 304 (Apr 9, 2004).
69. A. Goffeau *et al.*, *Science* **274**, 546 (Oct 25, 1996).
70. A. Goffeau, *Nature* **430**, 25 (Jul 1, 2004).
71. C. Segarra, M. Aguade, *Genetics* **130**, 513 (Mar, 1992).
72. M. Papaceit, E. Juan, *Chromosome Res* **6**, 49 (Jan, 1998).
73. J. M. Ranz, J. Gonzalez, F. Casals, A. Ruiz, *Evolution* **57**, 1325 (Jun, 2003).
74. H. W. Brock, D. B. Roberts, *Genetics* **103**, 75 (Jan, 1983).
75. T. P. Neufeld, R. W. Carthew, G. M. Rubin, *Proc Natl Acad Sci U S A* **88**, 10203 (Nov 15, 1991).
76. S. Yi, B. Charlesworth, *Genetics* **156**, 1753 (Dec, 2000).
77. E. M. Zdobnov *et al.*, *Science* **298**, 149 (Oct 4, 2002).
78. J. H. Muller, *Bearings of the Drosophila work on systematics in New systematic* J. Huxley, Ed., (Clarendon Press, Oxford, 1940).
79. J. H. Whiting, Jr., M. D. Pliley, J. L. Farmer, D. E. Jeffery, *Genetics* **122**, 99 (May, 1989).
80. C. Segarra, E. R. Lozovskaya, G. Ribo, M. Aguade, D. L. Hartl, *Chromosoma* **104**, 129 (Nov, 1995).
81. C. Segarra, G. Ribo, M. Aguade, *Genetics* **144**, 139 (Sep, 1996).
82. S. Aulard, L. Monti, N. Chaminade, F. Lemeunier, *Genetica* **120**, 137 (Mar, 2004).
83. M. J. D. White, *Animal cytology and evolution*. (Cambridge University Press, 1973).
84. W. S. Stone, W. C. Guest, F. D. Wilson, *Proc Natl Acad Sci U S A* **46**, 350 (Mar, 1960).
85. R. Morales-Hojas, S. Paallysaho, C. P. Vieira, A. Hoikkala, J. Vieira, *Chromosoma* **116**, 21 (Feb, 2007).
86. T. L. Newman *et al.*, *Genome Res* **15**, 1344 (Oct, 2005).

87. L. Feuk *et al.*, *PLoS Genet* **1**, e56 (Oct, 2005).
88. J. M. Szamalek *et al.*, *Hum Genet* **119**, 103 (Mar, 2006).
89. R. J. Mural *et al.*, *Science* **296**, 1661 (May 31, 2002).
90. P. Pevzner, G. Tesler, *Genome Res* **13**, 37 (Jan, 2003).
91. W. J. Kent, R. Baertsch, A. Hinrichs, W. Miller, D. Haussler, *Proc Natl Acad Sci U S A* **100**, 11484 (Sep 30, 2003).
92. H. Kehrer-Sawatzki, D. N. Cooper, *Hum Mutat* **28**, 99 (Feb, 2007).
93. S. Zhao *et al.*, *Genome Res* **14**, 1851 (Oct, 2004).
94. J. H. Postlethwait *et al.*, *Genome Res* **10**, 1890 (Dec, 2000).
95. H. Kehrer-Sawatzki, D. N. Cooper, *Chromosome Res* **16**, 41 (2008).
96. Y. Fan, E. Linardopoulou, C. Friedman, E. Williams, B. J. Trask, *Genome Res* **12**, 1651 (Nov, 2002).
97. D. P. Locke *et al.*, *Genome Biol* **4**, R50 (2003).
98. H. Kehrer-Sawatzki, J. M. Szamalek, S. Tanzer, M. Platzer, H. Hameister, *Genomics* **85**, 542 (May, 2005).
99. J. A. Bailey, E. E. Eichler, *Nat Rev Genet* **7**, 552 (Jul, 2006).
100. X. She *et al.*, *Nature* **430**, 857 (Aug 19, 2004).
101. B. Dutrillaux, *Hum Genet* **48**, 251 (May 10, 1979).
102. J. Lejeune, B. Dutrillaux, M. O. Rethore, M. Prieur, *Chromosoma* **43**, 423 (1973).
103. J. J. Yunis, O. Prakash, *Science* **215**, 1525 (Mar 19, 1982).
104. L. Froenicke, *Cytogenet Genome Res* **108**, 122 (2005).
105. J. Ehrlich, D. Sankoff, J. H. Nadeau, *Genetics* **147**, 289 (Sep, 1997).
106. D. K. Griffin *et al.*, *BMC Genomics* **9**, 168 (2008).
107. G. L. Stebbins, *Science* **152**, 1463 (Jun 10, 1966).
108. A. H. Paterson *et al.*, *Nat Genet* **14**, 380 (Dec, 1996).
109. S. P. Kowalski, T. H. Lan, K. A. Feldmann, A. H. Paterson, *Genetics* **138**, 499 (Oct, 1994).
110. N. e. a. Kishimoto, *Theor. Appl. Genet* **88**, 4 (1994).
111. G. Blanc, A. Barakat, R. Guyot, R. Cooke, M. Delseny, *Plant Cell* **12**, 1093 (Jul, 2000).
112. T. J. Vision, D. G. Brown, S. D. Tanksley, *Science* **290**, 2114 (Dec 15, 2000).
113. A. H. Paterson *et al.*, *Plant Cell* **12**, 1523 (Sep, 2000).
114. K. Vandepoele, Y. Saeys, C. Simillion, J. Raes, Y. Van De Peer, *Genome Res* **12**, 1792 (Nov, 2002).
115. K. Vandepoele, C. Simillion, Y. Van de Peer, *Plant Cell* **15**, 2192 (Sep, 2003).
116. A. H. Paterson, J. E. Bowers, D. G. Peterson, J. C. Estill, B. A. Chapman, *Curr Opin Genet Dev* **13**, 644 (Dec, 2003).
117. M. D. Gale, K. M. Devos, *Proc Natl Acad Sci U S A* **95**, 1971 (Mar 3, 1998).
118. J. L. Bennetzen, *Plant Cell* **12**, 1021 (Jul, 2000).
119. K. M. Devos, M. D. Gale, *Plant Cell* **12**, 637 (May, 2000).
120. B. Keller, C. Feuillet, *Trends Plant Sci* **5**, 246 (Jun, 2000).
121. C. Feuillet, B. Keller, *Ann Bot (Lond)* **89**, 3 (Jan, 2002).
122. M. Chen *et al.*, *Proc Natl Acad Sci U S A* **94**, 3431 (Apr 1, 1997).
123. M. E. Sorrells *et al.*, *Genome Res* **13**, 1818 (Aug, 2003).
124. A. Kilian, J. Chen, F. Han, B. Steffenson, A. Kleinhofs, *Plant Mol Biol* **35**, 187 (Sep, 1997).
125. R. Tarchini, P. Biddle, R. Wineland, S. Tingey, A. Rafalski, *Plant Cell* **12**, 381 (Mar, 2000).
126. A. P. Tikhonov *et al.*, *Proc Natl Acad Sci U S A* **96**, 7409 (Jun 22, 1999).
127. J. L. Bennetzen, *Curr Opin Plant Biol* **10**, 176 (Apr, 2007).
128. J. L. Bennetzen, J. Ma, *Curr Opin Plant Biol* **6**, 128 (Apr, 2003).
129. G. Moore, K. M. Devos, Z. Wang, M. D. Gale, *Curr Biol* **5**, 737 (Jul 1, 1995).
130. A. F. Dernburg, *J Cell Biol* **153**, F33 (Jun 11, 2001).
131. B. Llorente *et al.*, *FEBS Lett* **487**, 101 (Dec 22, 2000).

132. U. Lagercrantz, *Genetics* **150**, 1217 (Nov, 1998).
133. K. Yogeeswaran *et al.*, *Genome Res* **15**, 505 (Apr, 2005).
134. D. Sperlich, and P. Pfriem, *Chromosomal polymorphism in natural and experimental populations*. L. C. a. J. N. T. M. Ashburner, JR, Ed., The Genetics and Biology of Drosophila (Academis Press, London, 1986), vol. 3a, pp. 257-309.
135. Powell. J. R., *Progress and Prospects in Evolutionary Biology: The Drosophila model*. (Oxford University Press, New York, 1997).
136. C. Bartolome, B. Charlesworth, *Genetics* **173**, 779 (Jun, 2006).
137. W. J. Murphy, R. Stanyon, S. J. O'Brien, *Genome Biol* **2**, REVIEWS0005 (2001).
138. D. W. Burt *et al.*, *Nature* **402**, 411 (Nov 25, 1999).
139. B. Charlesworth, J.A.Coyne and N. Barton, *Am. Nat* **130**, 113 (1987).
140. F. Grutzner *et al.*, *Genome Res* **12**, 1316 (Sep, 2002).
141. *Nature* **437**, 69 (Sep 1, 2005).
142. S. Ohno, *Nature* **244**, 259 (1973).
143. N. G. Copeland *et al.*, *Science* **262**, 57 (Oct 1, 1993).
144. R. W. DeBry, M. F. Seldin, *Genomics* **33**, 337 (May 1, 1996).
145. E. S. Lander *et al.*, *Nature* **409**, 860 (Feb 15, 2001).
146. J. Tonzetich, T. W. Lyttle, H. L. Carson, *Proc Natl Acad Sci U S A* **85**, 1717 (Mar, 1988).
147. M. Wasserman, *Cytological evolution of the Drosophila repleta species group* C. B. K. a. J. R. Powell, Ed., Drosophila inversion polymorphism (CRC press, Boca Raton FL, 1992), pp. 455-552.
148. M. Caceres, A. Barbadilla and A. Ruiz, *Evolution* **51**, 1149 (1997).
149. V. B. a. P. Pevzner, *SLAM J. Computing* **25**, 272 (1996).
150. S. Cai.
151. P. Trinh, A. McLysaght, D. Sankoff, *Bioinformatics* **20 Suppl 1**, i318 (Aug 4, 2004).
152. J. A. Bailey, R. Baertsch, W. J. Kent, D. Haussler, E. E. Eichler, *Genome Biol* **5**, R23 (2004).
153. H. Hinsch, S. Hannenhalli, *BMC Evol Biol* **6**, 90 (2006).
154. T. S. Becker, B. Lenhard, *Mol Genet Genomics* **278**, 487 (Nov, 2007).
155. S. Ohno, *Nature* **244**, 259 (1973).
156. P. G. Engstrom, S. J. Ho Sui, O. Drivenes, T. S. Becker, B. Lenhard, *Genome Res* **17**, 1898 (Dec, 2007).
157. D. J. Finnegan, *Trends Genet* **5**, 103 (Apr, 1989).
158. F. Casals, A. Navarro, *Heredity* **99**, 479 (Nov, 2007).
159. Y. H. Gray, *Trends Genet* **16**, 461 (Oct, 2000).
160. P. Andolfatto, J. D. Wall, M. Kreitman, *Genetics* **153**, 1297 (Nov, 1999).
161. F. Casals, M. Caceres, A. Ruiz, *Mol Biol Evol* **20**, 674 (May, 2003).
162. M. Caceres, J. M. Ranz, A. Barbadilla, M. Long, A. Ruiz, *Science* **285**, 415 (Jul 16, 1999).
163. K. D. Mathiopoulos, A. della Torre, V. Predazzi, V. Petrarca, M. Coluzzi, *Proc Natl Acad Sci U S A* **95**, 12444 (Oct 13, 1998).
164. I. V. Sharakhov *et al.*, *Proc Natl Acad Sci U S A* **103**, 6258 (Apr 18, 2006).
165. D. A. Petrov, E. R. Lozovskaya, D. L. Hartl, *Nature* **384**, 346 (Nov 28, 1996).
166. N. D. Singh, D. A. Petrov, *Mol Biol Evol* **21**, 670 (Apr, 2004).
167. W. R. Engels, C. R. Preston, *Genetics* **107**, 657 (Aug, 1984).
168. M. B. Evgen'ev *et al.*, *Proc Natl Acad Sci U S A* **97**, 11337 (Oct 10, 2000).
169. T. W. Lyttle, D. S. Haymer, *Genetica* **86**, 113 (1992).
170. V. Ladeveze, S. Aulard, N. Chaminade, G. Periquet, F. Lemeunier, *Proc Biol Sci* **265**, 1157 (Jul 7, 1998).
171. G. Fischer, S. A. James, I. N. Roberts, S. G. Oliver, E. J. Louis, *Nature* **405**, 451 (May 25, 2000).
172. M. J. Dunham *et al.*, *Proc Natl Acad Sci U S A* **99**, 16144 (Dec 10, 2002).

173. M. Kellis, N. Patterson, M. Endrizzi, B. Birren, E. S. Lander, *Nature* **423**, 241 (May 15, 2003).
174. L. D. Stein *et al.*, *PLoS Biol* **1**, E45 (Nov, 2003).
175. J. A. Bailey *et al.*, *Science* **297**, 1003 (Aug 9, 2002).
176. J. A. Bailey, D. M. Church, M. Ventura, M. Rocchi, E. E. Eichler, *Genome Res* **14**, 789 (May, 2004).
177. L. G. Shaffer, J. R. Lupski, *Annu Rev Genet* **34**, 297 (2000).
178. M. C. Valero, O. de Luis, J. Cruces, L. A. Perez Jurado, *Genomics* **69**, 1 (Oct 1, 2000).
179. W. Bi *et al.*, *Genome Res* **12**, 713 (May, 2002).
180. G. Gimelli *et al.*, *Hum Mol Genet* **12**, 849 (Apr 15, 2003).
181. U. DeSilva *et al.*, *Genome Res* **12**, 3 (Jan, 2002).
182. P. Dehal *et al.*, *Science* **293**, 104 (Jul 6, 2001).
183. L. Armengol, M. A. Pujana, J. Cheung, S. W. Scherer, X. Estivill, *Hum Mol Genet* **12**, 2201 (Sep 1, 2003).
184. D. P. Mortlock, M. E. Portnoy, R. L. Chandler, E. D. Green, *Genomics* **84**, 814 (Nov, 2004).
185. V. Eder *et al.*, *Mol Biol Evol* **20**, 1506 (Sep, 2003).
186. P. Stankiewicz, S. S. Park, K. Inoue, J. R. Lupski, *Genome Res* **11**, 1205 (Jul, 2001).
187. V. Goidts, J. M. Szamalek, H. Hameister, H. Kehrer-Sawatzki, *Hum Genet* **115**, 116 (Jul, 2004).
188. M. B. Coulibaly *et al.*, *PLoS One* **2**, e849 (2007).
189. J. M. Ranz *et al.*, *PLoS Biol* **5**, e152 (Jun, 2007).
190. H. Kehrer-Sawatzki, C. A. Sandig, V. Goidts, H. Hameister, *Cytogenet Genome Res* **108**, 91 (2005).
191. O. Prazeres da Costa, J. Gonzalez, A. Ruiz, *Chromosoma* **118**, 349 (Jun, 2009).
192. D. E. Roberts, D. Ascherman, N. Kleckner, *Genetics* **128**, 37 (May, 1991).
193. R. M. Chalmers, N. Kleckner, *EMBO J* **15**, 5112 (Sep 16, 1996).
194. C. F. Weil, S. R. Wessler, *Plant Cell* **5**, 515 (May, 1993).
195. J. English, K. Harrison, J. D. Jones, *Plant Cell* **5**, 501 (May, 1993).
196. J. J. English, K. Harrison, J. Jones, *Plant Cell* **7**, 1235 (Aug, 1995).
197. C. Lister, C. Martin, *Genetics* **123**, 417 (Oct, 1989).
198. C. Lister, D. Jackson, C. Martin, *Plant Cell* **5**, 1541 (Nov, 1993).
199. C. Martin, S. Mackay, R. Carpenter, *Genetics* **119**, 171 (May, 1988).
200. T. P. Robbins, R. Carpenter, E. S. Coen, *EMBO J* **8**, 5 (Jan, 1989).
201. Y. H. Svoboda, M. K. Robson, J. A. Sved, *Genetics* **139**, 1601 (Apr, 1995).
202. Y. H. Gray, M. M. Tanaka, J. A. Sved, *Genetics* **144**, 1601 (Dec, 1996).
203. C. R. Preston, J. A. Sved, W. R. Engels, *Genetics* **144**, 1623 (Dec, 1996).
204. R. Fisher, *The Genetical Theory of Natural Selection*. (Carendon Press, Oxford, 1930).
205. C. C. Li, *Population Genetics*. (University of Chicago Press, Chicago, 1955).
206. C. B. Krimbas, Powell, J. R.,, *Drosophila Inversion Polymophism*. C. B. Krimbas, Powell, J. R.,, Ed., (CRC Press, 1992), pp. 1-52.
207. M. Pombi *et al.*, *BMC Evol Biol* **8**, 309 (2008).
208. M. Kirkpatrick, N. Barton, *Genetics* **173**, 419 (May, 2006).
209. A. Brehm, C. B. Krimbas, *J Hered* **82**, 110 (Mar-Apr, 1991).
210. E. Sole, J. Balanya, D. Sperlich, L. Serra, *Evolution* **56**, 830 (Apr, 2002).
211. A. Prevosti *et al.*, *Proc Natl Acad Sci U S A* **85**, 5597 (Aug, 1988).
212. J. Balanya *et al.*, *Evolution* **57**, 1837 (Aug, 2003).
213. P. J. Fernandez Iriarte, F. M. Norry, E. R. Hasson, *Heredity* **91**, 51 (Jul, 2003).
214. J. Van 't Land, W. F. Van Putten, H. Villarroel, A. Kamping, W. Van Delden, *Evolution* **54**, 201 (Feb, 2000).
215. M. Coluzzi, *Prog Clin Biol Res* **96**, 143 (1982).
216. J. R. Powell, V. Petrarca, A. della Torre, A. Caccone, M. Coluzzi, *Parassitologia* **41**, 101 (Sep, 1999).
217. J. Krzywinski, O. G. Grushko, N. J. Besansky, *Mol Phylogenet Evol* **39**, 417 (May, 2006).

218. M. Coetzee, M. Craig, D. le Sueur, *Parasitol Today* **16**, 74 (Feb, 2000).
219. J. M. Pock Tsy *et al.*, *Malar J* **2**, 33 (Oct 8, 2003).
220. R. H. Hunt, M. Coetzee, M. Fettene, *Trans R Soc Trop Med Hyg* **92**, 231 (Mar-Apr, 1998).
221. F. J. Ayala, M. Coluzzi, *Proc Natl Acad Sci U S A* **102 Suppl 1**, 6535 (May 3, 2005).
222. N. J. Besansky *et al.*, *Proc Natl Acad Sci U S A* **91**, 6885 (Jul 19, 1994).
223. N. J. Besansky *et al.*, *Proc Natl Acad Sci U S A* **100**, 10818 (Sep 16, 2003).
224. M. J. Donnelly, J. Pinto, R. Girod, N. J. Besansky, T. Lehmann, *Heredity* **92**, 61 (Feb, 2004).
225. J. Krzywinski, D. Sangare, N. J. Besansky, *Genetics* **169**, 185 (Jan, 2005).
226. T. Pape, *Mosq. Syst.* **24**, 1 (March, 1992).
227. B. A. Garcia, A. Caccone, K. D. Mathiopoulos, J. R. Powell, *Genetics* **143**, 1313 (Jul, 1996).
228. A. Caccone, B. A. Garcia, J. R. Powell, *Insect Mol Biol* **5**, 51 (Feb, 1996).
229. K. D. Mathiopoulos, J. D. Powell, T. F. McCutchan, *Mol Biol Evol* **12**, 103 (Jan, 1995).
230. N. J. Besansky *et al.*, *Genetics* **147**, 1817 (Dec, 1997).
231. R. Wang-Sattler *et al.*, *PLoS One* **2**, e1249 (2007).
232. A. H. Sturtevant, T. Dobzhansky, *Proc Natl Acad Sci U S A* **22**, 448 (Jul, 1936).
233. H. L. Carson, *Science* **168**, 1414 (Jun 19, 1970).
234. P. M. O'Grady, R. H. Baker, C. M. Durando, W. J. Etges, R. DeSalle, *BMC Evol Biol* **1**, 6 (2001).
235. A. Caccone, G. S. Min, J. R. Powell, *Genetics* **150**, 807 (Oct, 1998).
236. C. A. Green, *J Hered* **73**, 2 (Jan-Feb, 1982).
237. A. G. Clark *et al.*, *Nature* **450**, 203 (Nov 8, 2007).
238. M. Coluzzi, V.Petrarca and M. A. Di Deco, *Boll. Zool* **52**, 19 (1985).
239. J. Hemingway, H. Ranson, *Annu Rev Entomol* **45**, 371 (2000).
240. L. H. Miller, *Science* **257**, 36 (Jul 3, 1992).
241. A. A. James, *Science* **257**, 37 (Jul 3, 1992).
242. V. Petrarca, J. C. Beier, *Am J Trop Med Hyg* **46**, 229 (Feb, 1992).
243. F. H. Collins *et al.*, *Science* **234**, 607 (Oct 31, 1986).
244. K. D. Vernick, F. H. Collins, *Am J Trop Med Hyg* **40**, 593 (Jun, 1989).
245. A. E. Crews-Oyen, V. Kumar, F. H. Collins, *Am J Trop Med Hyg* **49**, 341 (Sep, 1993).
246. K. D. Vernick *et al.*, *Exp Parasitol* **80**, 583 (Jun, 1995).
247. M. M. Riehle *et al.*, *Science* **312**, 577 (Apr 28, 2006).
248. V. Petrarca, G. Sabatinelli, M. A. Di Deco, M. Papakay, *Parassitologia* **32**, 371 (Dec, 1990).
249. M. Coluzzi, *Parasitol Today* **8**, 113 (Apr, 1992).
250. C. Wondji *et al.*, *J Med Entomol* **42**, 998 (Nov, 2005).
251. J. H. Bryan, *Genetica* **59**, 167 (1982).
252. A. H. Sturtevant, *Proc Natl Acad Sci U S A* **3**, 555 (Sep, 1917).
253. A. H. Sturtevant, *Proc Natl Acad Sci U S A* **7**, 235 (Aug, 1921).
254. A. H. Sturtevant, *Biol* **46**, (1926).
255. J. R. Powell, *Drosophila inversion polymrphism*. C. B. Krimbas, Ed., (1992).
256. T. H. Dobzhansky, *genetics and the origin of species*. (Columbia university, New York, ed. 1st edition, 1937).
257. C. S. Wesley, W. F. Eanes, *Proc Natl Acad Sci U S A* **91**, 3132 (Apr 12, 1994).
258. B. J. White *et al.*, *PLoS Genet* **3**, e217 (Dec, 2007).
259. K. A. Rocca, E. M. Gray, C. Costantini, N. J. Besansky, *Malar J* **8**, 147 (2009).
260. E. M. Gray, K. A. Rocca, C. Costantini, N. J. Besansky, *Malar J* **8**, 215 (2009).
261. F. Simard *et al.*, *BMC Ecol* **9**, 17 (2009).
262. S. M. Paskewitz, M. R. Brown, A. O. Lea, F. H. Collins, *J Parasitol* **74**, 432 (Jun, 1988).
263. A. M. Lackie, *Parasitol Today* **4**, 98 (Apr, 1988).
264. K. D. Vernick, F. H. Collins, R. W. Gwadz, *Am J Trop Med Hyg* **40**, 585 (Jun, 1989).

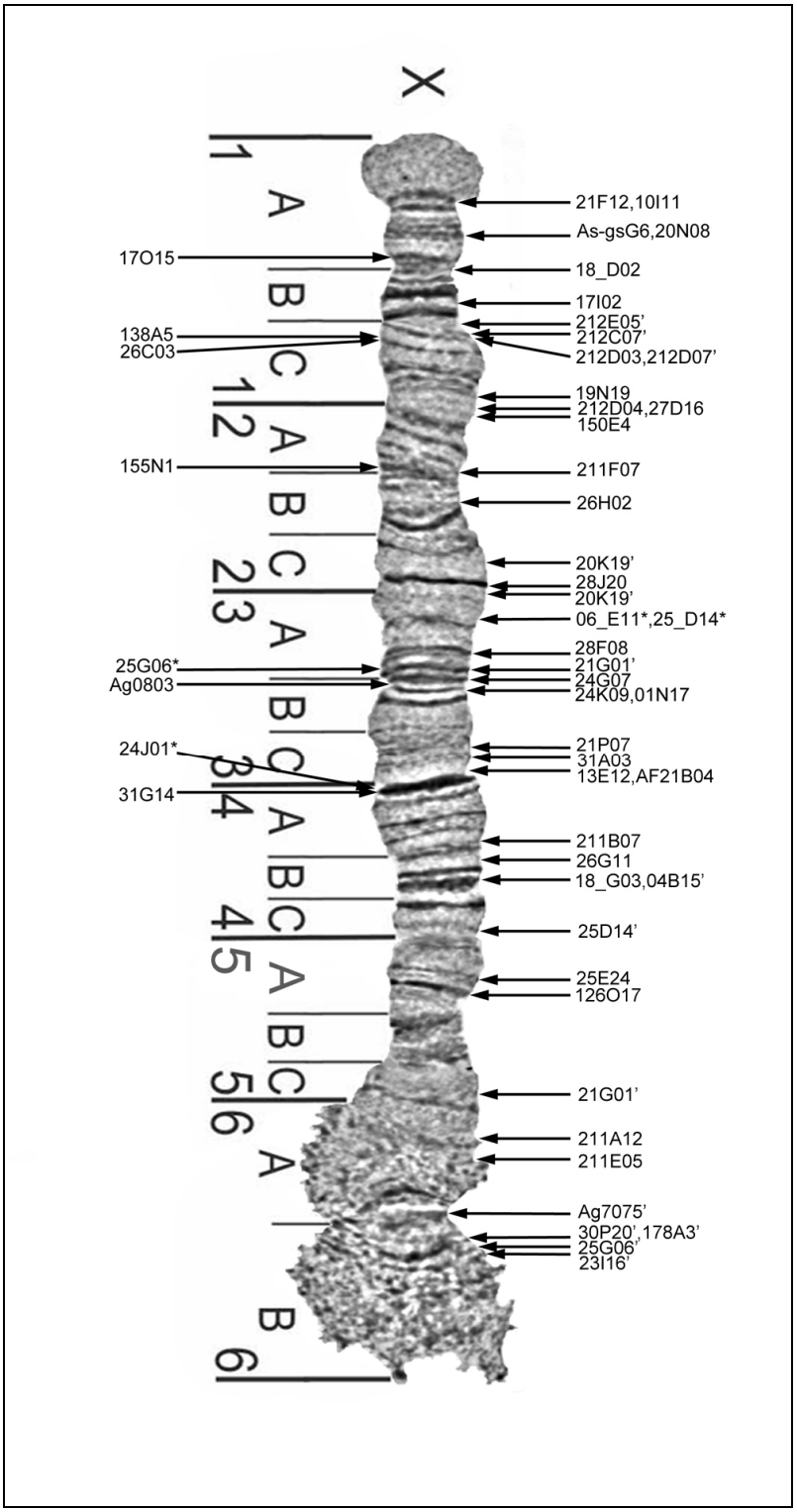
265. K. D. Vernick *et al.*, *Curr Top Microbiol Immunol* **295**, 383 (2005).
266. O. Niare *et al.*, *Science* **298**, 213 (Oct 4, 2002).
267. F. Scalenghe, E. Turco, J. E. Edstrom, V. Pirrotta, M. Melli, *Chromosoma* **82**, 205 (1981).
268. D. Rohme *et al.*, *Cell* **36**, 783 (Mar, 1984).
269. G. P. Bates, B. J. Wainwright, R. Williamson, S. D. Brown, *Mol Cell Biol* **6**, 3826 (Nov, 1986).
270. N. Brockdorff, E. M. Fisher, J. S. Cavanna, M. F. Lyon, S. D. Brown, *EMBO J* **6**, 3291 (Nov, 1987).
271. F. T. Kao, J. Yu, J. Qi, S. Tong, M. Muenke, *Cytogenet Cell Genet* **68**, 17 (1995).
272. H. J. Ludecke, G. Senger, U. Claussen, B. Horsthemke, *Nature* **338**, 348 (Mar 23, 1989).
273. H. J. Ludecke, G. Senger, U. Claussen, B. Horsthemke, *Hum Genet* **84**, 512 (May, 1990).
274. X. Y. Guan, P. S. Meltzer, J. Cao, J. M. Trent, *Genomics* **14**, 680 (Nov, 1992).
275. J. Yu, S. Tong, T. Yang-Feng, F. T. Kao, *Genomics* **14**, 769 (Nov, 1992).
276. J. Yu *et al.*, *Am J Hum Genet* **51**, 263 (Aug, 1992).
277. S. M. Jiang *et al.*, *Genome* **47**, 1114 (Dec, 2004).
278. C. Jung, U. Claussen, B. Horsthemke, F. Fischer, R. G. Herrmann, *Plant Mol Biol* **20**, 503 (Nov, 1992).
279. S. Monajembashi, C. Cremer, T. Cremer, J. Wolfrum, K. O. Greulich, *Exp Cell Res* **167**, 262 (Nov, 1986).
280. M. R. Emmert-Buck *et al.*, *Science* **274**, 998 (Nov 8, 1996).
281. H. S. Willenberg, R. Walters, S. R. Bornstein, *Methods Enzymol* **356**, 216 (2002).
282. P. Pinzani, C. Orlando, M. Pazzagli, *Mol Aspects Med* **27**, 140 (Apr-Jun, 2006).
283. V. Espina, M. Heiby, M. Pierobon, L. A. Liotta, *Expert Rev Mol Diagn* **7**, 647 (Sep, 2007).
284. V. Espina *et al.*, *Nat Protoc* **1**, 586 (2006).
285. A. F. Okuducu, J. C. Hahne, A. Von Deimling, N. Wernert, *Int J Mol Med* **15**, 763 (May, 2005).
286. G. Esposito, *Adv Exp Med Biol* **593**, 54 (2007).
287. D. G. Standaert, *Arch Neurol* **62**, 203 (Feb, 2005).
288. E. Heinmoller *et al.*, *Pathol Res Pract* **199**, 363 (2003).
289. R. Pamphlett, P. R. Heath, H. Holden, P. G. Ince, P. J. Shaw, *J Neurosci Methods* **147**, 65 (Aug 30, 2005).
290. N. S. Agar *et al.*, *Proc Natl Acad Sci U S A* **101**, 4954 (Apr 6, 2004).
291. V. W. Wang, D. A. Bell, R. S. Berkowitz, S. C. Mok, *Cancer Res* **61**, 4169 (May 15, 2001).
292. M. S. Rook, S. M. Delach, G. Deyneko, A. Worlock, J. L. Wolfe, *Am J Pathol* **164**, 23 (Jan, 2004).
293. R. Burgemeister, R. Gangnus, B. Haar, K. Schutze, U. Sauer, *Pathol Res Pract* **199**, 431 (2003).
294. K. M. Keays, G. P. Owens, A. M. Ritchie, D. H. Gilden, M. P. Burgoon, *J Immunol Methods* **302**, 90 (Jul, 2005).
295. K. Kobayashi *et al.*, *J Vet Med Sci* **65**, 917 (Aug, 2003).
296. Y. Niyaz *et al.*, *Methods Mol Med* **114**, 1 (2005).
297. R. Parlato *et al.*, *Anal Biochem* **300**, 139 (Jan 15, 2002).
298. R. N. Zhou, Z. M. Hu, *Curr Genomics* **8**, 67 (Mar, 2007).
299. H. X. Deng *et al.*, *Hum Genet* **89**, 13 (Apr, 1992).
300. X. Y. Guan, J. M. Trent, P. S. Meltzer, *Hum Mol Genet* **2**, 1117 (Aug, 1993).
301. P. S. Meltzer, X. Y. Guan, A. Burgess, J. M. Trent, *Nat Genet* **1**, 24 (Apr, 1992).
302. C. S. Wesley, M. Ben, M. Kreitman, N. Hagag, W. F. Eanes, *Nucleic Acids Res* **18**, 599 (Feb 11, 1990).
303. N. Ponelies, E. K. Bautz, S. Monajembashi, J. Wolfrum, K. O. Greulich, *Chromosoma* **98**, 351 (Nov, 1989).
304. W. Peng, H. Takabayashi, K. Ikawa, *Eur J Obstet Gynecol Reprod Biol* **131**, 13 (Mar, 2007).
305. E. K. Binga, R. S. Lasken, J. D. Neufeld, *ISME J* **2**, 233 (Mar, 2008).
306. R. S. Lasken, M. Egholm, *Trends Biotechnol* **21**, 531 (Dec, 2003).

307. A. Raghunathan *et al.*, *Appl Environ Microbiol* **71**, 3342 (Jun, 2005).
308. L. Zhang *et al.*, *Proc Natl Acad Sci U S A* **89**, 5847 (Jul 1, 1992).
309. S. Coskun, O. Alsmadi, *Prenat Diagn* **27**, 297 (Apr, 2007).
310. A. Sekizawa *et al.*, *Hum Genet* **102**, 393 (Apr, 1998).
311. W. E. Gibbons, S. A. Gitlin, S. E. Lanzendorf, *Am J Obstet Gynecol* **172**, 1088 (Apr, 1995).
312. K. Xu, Y. Tang, J. A. Grifo, Z. Rosenwaks, J. Cohen, *Hum Reprod* **8**, 2206 (Dec, 1993).
313. F. Schaaff, H. Wedemann, E. Schwinger, *Hum Genet* **98**, 158 (Aug, 1996).
314. K. Kristjansson *et al.*, *Nat Genet* **6**, 19 (Jan, 1994).
315. Z. Jiao *et al.*, *Prenat Diagn* **23**, 646 (Aug, 2003).
316. D. Wells, J. K. Sherlock, A. H. Handyside, J. D. Delhanty, *Nucleic Acids Res* **27**, 1214 (Feb 15, 1999).
317. E. Heinmoller *et al.*, *Lab Invest* **82**, 443 (Apr, 2002).
318. F. Bataille *et al.*, *Mol Pathol* **56**, 286 (Oct, 2003).
319. H. Telenius *et al.*, *Genomics* **13**, 718 (Jul, 1992).
320. H. Malmgren *et al.*, *Mol Hum Reprod* **8**, 502 (May, 2002).
321. D. Wells, J. D. Delhanty, *Mol Hum Reprod* **6**, 1055 (Nov, 2000).
322. Z. Y. He *et al.*, *Fertil Steril* **72**, 341 (Aug, 1999).
323. L. Voullaire, L. Wilton, H. Slater, R. Williamson, *Prenat Diagn* **19**, 846 (Sep, 1999).
324. J. J. Engelen *et al.*, *Genet Couns* **11**, 13 (2000).
325. G. Senger *et al.*, *Prenat Diagn* **17**, 369 (Apr, 1997).
326. J. Muller-Navia, A. Nebel, E. Schleiermacher, *Hum Genet* **96**, 661 (Dec, 1995).
327. J. Muller-Navia *et al.*, *Prenat Diagn* **16**, 915 (Oct, 1996).
328. C. A. Brandt, T. Lyngbye, S. Pedersen, L. Bolund, U. Friedrich, *J Med Genet* **31**, 234 (Mar, 1994).
329. Y. Daigo *et al.*, *Am J Pathol* **158**, 1623 (May, 2001).
330. H. Zitzelsberger *et al.*, *Virchows Arch* **433**, 297 (Oct, 1998).
331. A. Marchio *et al.*, *Mol Pathol* **54**, 270 (Aug, 2001).
332. T. Harada *et al.*, *Cancer Res* **62**, 835 (Feb 1, 2002).
333. J. Y. Lee, Y. N. Park, K. O. Uhm, S. Y. Park, S. H. Park, *J Korean Med Sci* **19**, 682 (Oct, 2004).
334. S. Barbaux, O. Poirier, F. Cambien, *J Mol Med* **79**, 329 (Jun, 2001).
335. S. F. Grant *et al.*, *Nucleic Acids Res* **30**, e125 (Nov 15, 2002).
336. V. G. Cheung, S. F. Nelson, *Proc Natl Acad Sci U S A* **93**, 14676 (Dec 10, 1996).
337. S. H. Kim, T. Godfrey, R. H. Jensen, *J Urol* **162**, 1512 (Oct, 1999).
338. R. Kittler, M. Stoneking, M. Kayser, *Anal Biochem* **300**, 237 (Jan 15, 2002).
339. W. Dietmaier *et al.*, *Am J Pathol* **154**, 83 (Jan, 1999).
340. H. Fiegler *et al.*, *Nucleic Acids Res* **35**, e15 (2007).
341. A. Wasserstrom *et al.*, *PLoS One* **3**, e1939 (2008).
342. C. A. Klein *et al.*, *Proc Natl Acad Sci U S A* **96**, 4494 (Apr 13, 1999).
343. R. Lucito *et al.*, *Proc Natl Acad Sci U S A* **95**, 4487 (Apr 14, 1998).
344. P. Vos *et al.*, *Nucleic Acids Res* **23**, 4407 (Nov 11, 1995).
345. C. Tanabe *et al.*, *Genes Chromosomes Cancer* **38**, 168 (Oct, 2003).
346. J. P. Langmore, *Pharmacogenomics* **3**, 557 (Jul, 2002).
347. L. Blanco, M. Salas, *Proc Natl Acad Sci U S A* **81**, 5325 (Sep, 1984).
348. L. Blanco *et al.*, *J Biol Chem* **264**, 8935 (May 25, 1989).
349. F. B. Dean *et al.*, *Proc Natl Acad Sci U S A* **99**, 5261 (Apr 16, 2002).
350. K. A. Eckert, T. A. Kunkel, *PCR Methods Appl* **1**, 17 (Aug, 1991).
351. J. A. Esteban, M. Salas, L. Blanco, *J Biol Chem* **268**, 2719 (Feb 5, 1993).
352. S. Hosono *et al.*, *Genome Res* **13**, 954 (May, 2003).
353. a. S. Silander K., J, *Methods in Molecular Biology*. M. S. a. R. Elaswarrapu, Ed., Genomic Protocols: second edition (Human Press Inc, NJ).

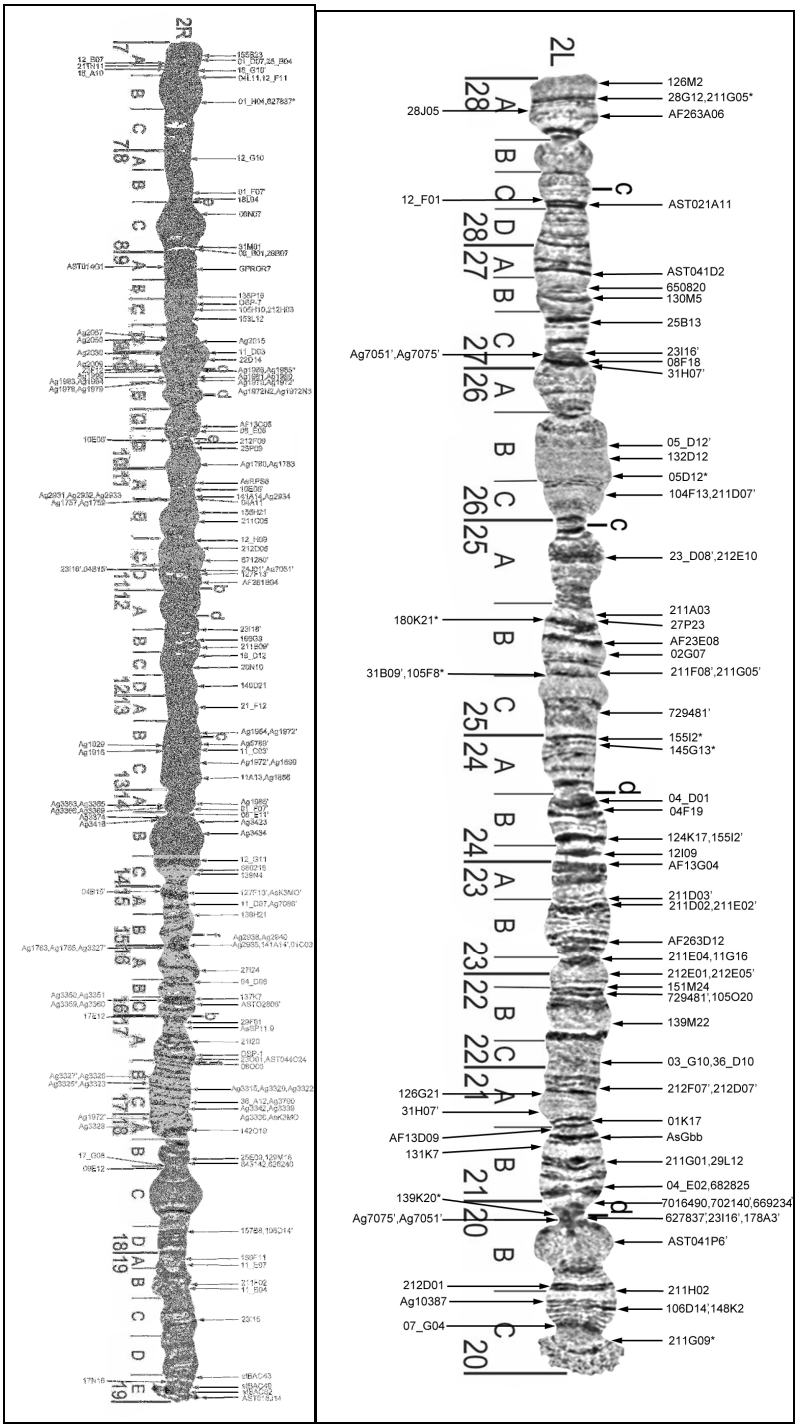
354. S. Hughes *et al.*, *BMC Genomics* **7**, 65 (2006).
355. D. Frumkin *et al.*, *BMC Biotechnol* **8**, 17 (2008).
356. M. Hockner, M. Erdel, A. Spreiz, G. Utermann, D. Kotzot, *Cytogenet Genome Res* **125**, 98 (2009).
357. J. F. Holbrook, D. Stabley, K. Sol-Church, *J Biomol Tech* **16**, 125 (Jun, 2005).
358. K. Zhang *et al.*, *Nat Biotechnol* **24**, 680 (Jun, 2006).
359. A. W. Bergen, Y. Qi, K. A. Haque, R. A. Welch, S. J. Chanock, *BMC Biotechnol* **5**, 24 (2005).
360. K. Iwamoto *et al.*, *PLoS One* **2**, e1306 (2007).
361. C. A. Hutchison, 3rd, H. O. Smith, C. Pfannkoch, J. C. Venter, *Proc Natl Acad Sci U S A* **102**, 17332 (Nov 29, 2005).
362. C. Spits *et al.*, *Hum Mutat* **27**, 496 (May, 2006).
363. F. B. Dean, J. R. Nelson, T. L. Giesler, R. S. Lasken, *Genome Res* **11**, 1095 (Jun, 2001).
364. J. M. Lage *et al.*, *Genome Res* **13**, 294 (Feb, 2003).
365. Z. Jiang, X. Zhang, R. Deka, L. Jin, *Nucleic Acids Res* **33**, e91 (2005).
366. C. Le Caignec *et al.*, *Nucleic Acids Res* **34**, e68 (2006).
367. X. Pan *et al.*, *Proc Natl Acad Sci U S A* **105**, 15499 (Oct 7, 2008).
368. Y. Marcy *et al.*, *PLoS Genet* **3**, 1702 (Sep, 2007).
369. C. Spits *et al.*, *Nat Protoc* **1**, 1965 (2006).
370. J. Krzywinski, D. R. Nusskern, M. K. Kern, N. J. Besansky, *Genetics* **166**, 1291 (Mar, 2004).
371. A. Xia, M. V. Sharakhova, I. V. Sharakhov, *J Comput Biol* **15**, 965 (Oct, 2008).
372. J. R. Powell, *progress and prospects in evolutionary biology. The Drosophila model*. (Oxford university press, New York, 1997).
373. J. Gonzalez, E. Betran, M. Ashburner, A. Ruiz, *Chromosome Res* **8**, 375 (2000).
374. C. Alonso, H. D. Berendes, *Chromosoma* **51**, 347 (Aug 11, 1975).
375. D. H. Fitch, L. D. Strausbaugh, V. Barrett, *Chromosoma* **99**, 118 (Apr, 1990).
376. M. Lynch, J. S. Conery, *Science* **290**, 1151 (Nov 10, 2000).
377. G. M. Rubin *et al.*, *Science* **287**, 2204 (Mar 24, 2000).
378. T. L. Turner, M. T. Levine, M. L. Eckert, D. J. Begun, *Genetics* **179**, 455 (May, 2008).
379. C. Esnault *et al.*, *PLoS ONE* **3**, e1968 (2008).
380. R. Wang-Sattler *et al.*, *PLoS ONE* **2**, e1249 (2007).
381. N. J. Besansky, J. R. Powell, *J Med Entomol* **29**, 125 (Jan, 1992).
382. C. D. Smith, S. Shu, C. J. Mungall, G. H. Karpen, *Science* **316**, 1586 (Jun 15, 2007).
383. G. Bourque, P. A. Pevzner, *Genome Res* **12**, 26 (Jan, 2002).
384. W. J. Murphy *et al.*, *Science* **309**, 613 (Jul 22, 2005).
385. Y. T. Toure *et al.*, *Parassitologia* **40**, 477 (Dec, 1998).
386. B. N. Nagpal, A. Srivastava, N. L. Kalra, S. K. Subbarao, *J Med Entomol* **40**, 747 (Nov, 2003).
387. R. E. Harbach, I. J. Kitching, *Systematics and Biodiversity* **3**, 345 (Dec, 2005).
388. J. C. Marshall, J. R. Powell, A. Caccone, *Am J Trop Med Hyg* **73**, 749 (Oct, 2005).
389. C. M. Bergman *et al.*, *Genome Biol* **3**, RESEARCH0086 (2002).
390. O. Mukabayire, N. J. Besansky, *Chromosoma* **104**, 585 (Jun, 1996).
391. N. Chuzhanova, S. S. Abeysinghe, M. Krawczak, D. N. Cooper, *Hum Mutat* **22**, 245 (Sep, 2003).
392. K. S. Lobachev, A. Rattray, V. Narayanan, *Front Biosci* **12**, 4208 (2007).
393. S. Cirera, J. M. Martin-Campos, C. Segarra, M. Aguade, *Genetics* **139**, 321 (Jan, 1995).
394. S. Schneuwly, A. Kuroiwa, W. J. Gehring, *EMBO J* **6**, 201 (Jan, 1987).
395. D. I. Nurminsky, E. N. Moriyama, E. R. Lozovskaya, D. L. Hartl, *Mol Biol Evol* **13**, 132 (Jan, 1996).
396. S. Ohno, *Sex chromosomes and sex-linked genes*. (Springer-Verlag, New York, 1967).
397. W. R. Rice, *Evolution* **38**, 735 (1984).
398. H. A. Orr, *Annu. Rev. Ecol. Syst* **28**, 195 (1997).
399. L. D. Hurst, *Nature* **411**, 149 (May 10, 2001).

400. P. J. Wang, J. R. McCarrey, F. Yang, D. C. Page, *Nat Genet* **27**, 422 (Apr, 2001).
401. E. A. Montgomery, S. M. Huang, C. H. Langley, B. H. Judd, *Genetics* **129**, 1085 (Dec, 1991).
402. J. K. Lim, M. J. Simmons, *Bioessays* **16**, 269 (Apr, 1994).
403. M. Mitas, *Nucleic Acids Res* **25**, 2245 (Jun 15, 1997).
404. H. Moore, P. W. Greenwell, C. P. Liu, N. Arnheim, T. D. Petes, *Proc Natl Acad Sci U S A* **96**, 1504 (Feb 16, 1999).
405. M. T. Pletcher *et al.*, *Genome Res* **10**, 1463 (Oct, 2000).
406. R. Puttagunta *et al.*, *Genome Res* **10**, 1369 (Sep, 2000).
407. M. Hochstrasser, D. Mathog, Y. Gruenbaum, H. Saumweber, J. W. Sedat, *J Cell Biol* **102**, 112 (Jan, 1986).
408. Y. Urata, S. J. Parmelee, D. A. Agard, J. W. Sedat, *J Cell Biol* **131**, 279 (Oct, 1995).
409. D. Mathog, J. W. Sedat, *Genetics* **121**, 293 (Feb, 1989).
410. D. Mathog, M. Hochstrasser, Y. Gruenbaum, H. Saumweber, J. Sedat, *Nature* **308**, 414 (Mar 29-Apr 4, 1984).
411. E. A. Baricheva *et al.*, *Gene* **171**, 171 (Jun 1, 1996).
412. R. Rzepecki, S. S. Bogachev, E. Kokoza, N. Stuurman, P. A. Fisher, *J Cell Sci* **111 (Pt 1)**, 121 (Jan, 1998).
413. V. Stierle *et al.*, *Biochemistry* **42**, 4819 (May 6, 2003).
414. T. Dechat *et al.*, *Genes Dev* **22**, 832 (Apr 1, 2008).
415. W. F. Marshall, *Mech Dev* **120**, 1217 (Nov, 2003).
416. W. F. Marshall, *Curr Biol* **12**, R185 (Mar 5, 2002).
417. M. Gandhi, M. Medvedovic, J. R. Stringer, Y. E. Nikiforov, *Oncogene* **25**, 2360 (Apr 13, 2006).
418. E. Bartova, A. Harnicarova, J. Pachernik, S. Kozubek, *Leuk Res* **29**, 901 (Aug, 2005).
419. H. Neves, C. Ramos, M. G. da Silva, A. Parreira, L. Parreira, *Blood* **93**, 1197 (Feb 15, 1999).
420. M. Gandhi, V. Evdokimova, Y. E. Nikiforov, *Mol Cell Endocrinol*, (Sep 18, 2009).
421. A. Cohuet, I. Dia, F. Simard, M. Raymond, D. Fontenille, *Insect Mol Biol* **13**, 251 (Jun, 2004).
422. F. Tripet, G. Dolo, G. C. Lanzaro, *Genetics* **169**, 313 (Jan, 2005).
423. D. Fontenille, F. Simard, *Comp Immunol Microbiol Infect Dis* **27**, 357 (Sep, 2004).
424. P. Carnevale, G. Le Goff, J. Toto, R. V, *Med Vet Entomol* **6**, 135 (1992).
425. I. Dia, T. Diop, I. Rakotoarivony, P. Kengne, D. Fontenille, *J Med Entomol* **40**, 279 (May, 2003).
426. Y. C. Lin, C. L. Lu, Y. C. Liu, C. Y. Tang, *Nucleic Acids Res* **34**, W696 (Jul 1, 2006).
427. J. R. Powell, *Progress and prospects in evolutionary biology: The Drosophila model*. (Oxford University Press, Oxford, 1997).
428. C. B. Krimbas, Powell,J.R., *Drosophila inversion polymorphism*. (CRC press, 1992).
429. R. E. Harbach, I. J. Kitching, *Systematics and Biodiversity* **3**, 30 (2005).
430. T. D. Petes, C. W. Hill, *Annu Rev Genet* **22**, 147 (1988).
431. M. M. Riehle *et al.*, *PLoS One* **3**, e3672 (2008).
432. N. Bray, L. Pachter, *Nucleic Acids Res* **31**, 3525 (Jul 1, 2003).
433. P. Andolfatto, M. Przeworski, *Genetics* **158**, 657 (Jun, 2001).
434. M. Brudno *et al.*, *Genome Res* **13**, 721 (Apr, 2003).
435. V. Petrarca *et al.*, *Med Vet Entomol* **14**, 149 (2000).
436. R. J. Post, Kruger, A., and Somiari, S. B, *Molecular Ecology Notes* **6**, 634 (2006).
437. X. Y. Guan *et al.*, *Cancer Genet Cytogenet* **134**, 65 (Apr 1, 2002).
438. A. H. Handyside *et al.*, *Mol Hum Reprod* **10**, 767 (Oct, 2004).
439. A. Hellani *et al.*, *Mol Hum Reprod* **10**, 847 (Nov, 2004).
440. D. Lawson *et al.*, *Nucleic Acids Res* **37**, D583 (Jan, 2009).
441. C. S. Wondji *et al.*, *BMC Genomics* **8**, 34 (2007).
442. T. J. Hubbard *et al.*, *Nucleic Acids Res* **37**, D690 (Jan, 2009).

Appendices

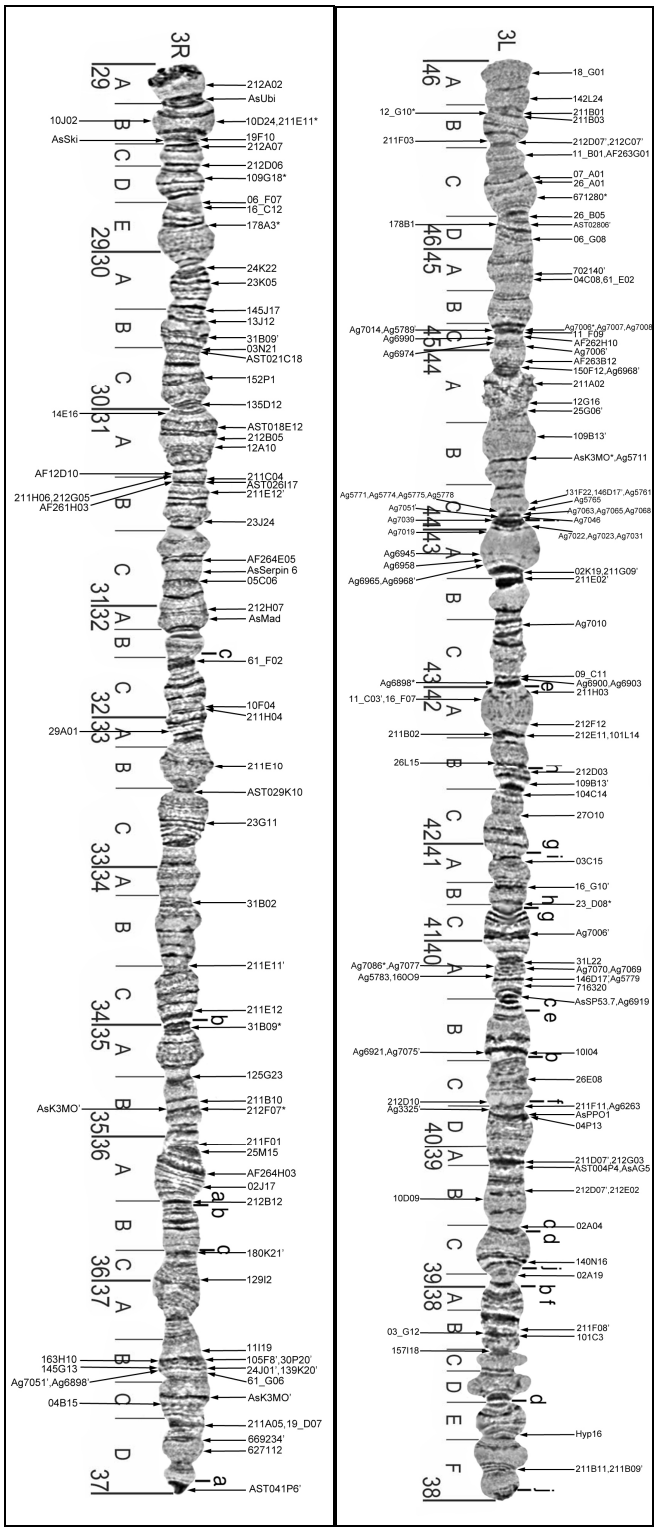


Appendix 2.1: *In situ* hybridization locations of markers on X chromosome of *A. stephensi*.



Appendix 2.2 (left): *In situ* hybridization locations of markers on 2R chromosome of *A. stephensi*.

Appendix 2.3 (right): *In situ* hybridization locations of markers on 2L chromosome of *A. stephensi*.



Appendix 2.4 (left): *In situ* hybridization locations of markers on 3R chromosome of *A. stephensi*.

Appendix 2.5 (right): *In situ* hybridization locations of markers on 3L chromosome of *A. stephensi*.

Appendix 2.6: The locations of in situ hybridization probes on *A. stephensi* X chromosome

#	Signal #	Clone	Acession #	Chromosome	Division	Clone type	
1	1	21F12	AL150696	X	1A(1)	*gambiae* BAC	
2	1	10I11	AL145448	X	1A(1)	*gambiae* BAC	
3	2	As_gsG6	AY226456	X	1A(2)	*stephensi* cDNA from GENBank	
4	2	20N08	AL150407	X	1A(2)	*gambiae* BAC	
5	3	17O15	AL609326	X	1A(3)	*gambiae* BAC	
6	4	18_D02	BU038939	X	1B(1)	*funestus* cDNA	
7	5	17I02	AL148591	X	1B(2)	*gambiae* BAC	
8	6	212E05'	Sehouche lab	X	1B_C	*stephensi* cDNA from cDNA library	multiple
9	7	212C07'	Sehouche lab	X	1C(1)	*stephensi* cDNA from cDNA library	multiple
10	8	212D07'	Sehouche lab	X	1C(2)	*stephensi* cDNA from cDNA library	multiple
11	8	211D03'	Sehouche lab	X	1C(2)	*stephensi* cDNA from cDNA library	multiple
12	9	138A5	BH380601	X	1C(3)	*gambiae* BAC	
13	10	26C03	AL153390	X	1C(4)	*gambiae* BAC	
14	11	19N19	AL609649	X	1C(5)	*gambiae* BAC	
15	12	212D04	Sehouche lab	X	1C(6)	*stephensi* cDNA from cDNA library	
16	12	27D16	AL154011	X	1C(6)	*gambiae* BAC	
17	13	150E4	BH388828	X	1C(7)	*gambiae* BAC	
18	14	211F07	Sehouche lab	X	2A	*stephensi* cDNA from cDNA library	
19	15	155N1	BH384248	X	2A_B	*gambiae* BAC	
20	16	26H02	AL610921	X	2B	*gambiae* BAC	
21	17	20K19'	AL609825	X	2C(1)	*gambiae* BAC	multiple
22	18	28J20	AL154814	X	2C(2)	*gambiae* BAC	
	19	20K19'	AL609825	X	2C(3)	*gambiae* BAC	multiple
23	20	06_E11*	BU038887	X	3A(1)	*funestus* cDNA	multiple
24	20	25D14*	AL610688	X	3A(1)	*gambiae* BAC	multiple
25	21	28F08	AL154661	X	3A(2)	*gambiae* BAC	
26	22	25G06*	AL610709	X	3A(3)	*gambiae* BAC	multiple
27	23	21G01'	AL150712	X	3A(4)	*gambiae* BAC	multiple
28	24	24G07	AL152397	X	3B(1)	*gambiae* BAC	
29	25	Ag0803	AGAP000803	X	3B(2)	*gambiae* cDNA	
30	26	24K09	AL152559	X	3B(3)	*gambiae* BAC	
31	26	01N17	AL607334	X	3B(3)	*gambiae* BAC	
32	27	21P07	AL151067	X	3C(1)	*gambiae* BAC	
33	28	31A03	AL611577	X	3C(2)	*gambiae* BAC	
34	29	AF21B04	Collins lab	X	4A(1)	*funestus* microsatellite	
35	29	13E12	AL146915	X	4A(1)	*gambiae* BAC	
36	30	24J01*	AL152501	X	4A(2)	*gambiae* BAC	multiple
37	31	31G14	AL156439	X	4A(3)	*gambiae* BAC	
38	32	211B07	Sehouche lab	X	4B(1)	*stephensi* cDNA from cDNA library	
39	33	26_G11	BU038979	X	4B(2)	*funestus* cDNA	
40	34	18_G03	BU038942	X	4B(3)	*funestus* cDNA	
41	34	04B15'	AL145409	X	4B(3)	*gambiae* BAC	multiple
	35	25D14'	AL610688	X	4C	*gambiae* BAC	multiple
42	36	25E24	AL610701	X	5A(1)	*gambiae* BAC	
43	37	126O17	BH404578	X	5A(2)	*gambiae* BAC	
44	38	21G01'	AL150712	X	5C	*gambiae* BAC	multiple
	39	211A12	Sehouche lab	X	6A(het)(1)	*stephensi* cDNA	

						from cDNA library	
45	40	211E05'	Sehouche lab	X	6A(het)(2)	*stephensi* cDNA from cDNA library	
46	41	Ag7075'	AGAP007075	X	6A(het)(3)	*gambiae* cDNA	multiple
47	41	30P20'	AL156193	X	6A(het)(4)	*gambiae* BAC	multiple
48	42	178A3'	BH398965	X	6A(het)(4)	*gambiae* BAC	multiple
	43	25G06'	AL610709	X	6B (het)(5)	*gambiae* BAC	multiple
49	44	23I16'	AL610348	X	6B (het)(6)	*gambiae* BAC	multiple

Appendix 2.7: The locations of in situ hybridization probes on *A. stephensi* 2R chromosome

#	Signal #	clone	Accession #	Chromosome	Division	Clone type	
1	1	155B23	BH368691	2R	7A(1)	*gambiae* BAC	
2	2	01_D07	BU038871	2R	7A(2)	*funestus* cDNA	
3	2	25_B04	BU038970	2R	7A(2)	*funestus* cDNA	
4	3	12_B07	BU038908	2R	7A(3)	*funestus* cDNA	
5	4	211H11	Sehouche lab	2R	7A(4)	*stephensi* cDNA from cDNA library	
6	5	16_G10'	BU038933	2R	7A(5)	*funestus* cDNA	multiple
7	6	16_A10	BU038927	2R	7A(6)	*funestus* cDNA	
8	7	12_F11	BU038912	2R	7A(7)	*funestus* cDNA	
9	7	04L11	AL141975	2R	7A(7)	*gambiae* BAC	
10	8	01_H04	BU038873	2R	7B(1)	*funestus* cDNA	
11	8	627837*	BM647307	2R	7B(1)	*gambiae* cDNA	multiple
12	9	12_G10	BU038913	2R	8A	*funestus* cDNA	
13	10	18L04	AL149296	2R	8B(2)	*gambiae* BAC	
14	11	09N07	AL145079	2R	8C(1)	*gambiae* BAC	
15	12	31M01	AL611707	2R	8C(2)	*gambiae* BAC	
16	13	06_B01	BU038882	2R	9A(1)	*funestus* cDNA	
17	13	29B07	AL155076	2R	9A(1)	*gambiae* BAC	
18	14	AST014G1	Tu lab	2R	9A(2)	*stephensi* BAC	
19	15	GPROR7	Tu lab	2R	9A(3)	*stephensi*	
20	16	135P16	BH387168	2R	9B	*gambiae* BAC	
21	17	OBP-7	Tu lab	2R	9C(1)		
22	18	212H03	Sehouche lab	2R	9C(2)	*stephensi* cDNA	
23	18	105H10	BH368219	2R	9C(2)	*gambiae* BAC	
24	19	153L12	BH380684	2R	9C(3)	*gambiae* BAC	
25	20	Ag2067	AGAP002067	2R	9D(2)	*gambiae* cDNA	
26	21	Ag2050	AGAP002050	2R	9D(3)	*gambiae* cDNA	
27	22	Ag2015	AGAP002015	2R	9D(4)	*gambiae* cDNA	
28	23	Ag2030	AGAP002030	2R	10A(1)	*gambiae* cDNA	
29	24	11_D03	BU038903	2R	10A(2)	*funestus* cDNA	
30	25	22D14	AL151204	2R	10A(3)	*gambiae* BAC	
31	26	Ag2009	AGAP002009	2R	10A(4)	*gambiae* cDNA	
32	27	23F12	AL151850	2R	10A(5)	*gambiae* BAC	
33	28	Ag1999	AGAP001999	2R	10A(6)	*gambiae* cDNA	
34	29	Ag1986	AGAP001986	2R	10A(7)	*gambiae* cDNA	
35	29	Ag1985*	AGAP001985	2R	10A(7)	*gambiae* cDNA	multiple
36	30	Ag1984	AGAP001984	2R	10A(8)	*gambiae* cDNA	
37	30	Ag1983	AGAP001983	2R	10A(8)	*gambiae* cDNA	
38	31	Ag1981	AGAP001981	2R	10A(9)	*gambiae* cDNA	
39	31	Ag1980	AGAP001980	2R	10A(9)	*gambiae* cDNA	
40	31	Ag1978	AGAP001978	2R	10A(9)	*gambiae* cDNA	
41	31	Ag1979	AGAP001979	2R	10A(9)	*gambiae* cDNA	
42	32	Ag1970	AGAP001970	2R	10A(10)	*gambiae* cDNA	
43	32	Ag1972'	AGAP001972	2R	10A(10)	*gambiae* cDNA	multiple
	32	Ag1972N2	AGAP001972	2R	10A(10)	*gambiae* cDNA	

	32	Ag1972N3	AGAP001972	2R	10A(10)	*gambiae* cDNA	
44	33	AF13C05	Collins lab	2R	10C	*funestus* BAC	
45	34	08_E06	BU038895	2R	10D(1)	*funestus* cDNA	
46	35	10E06'	AL145314	2R	10D(2)	*gambiae* BAC	multiple
47	36	212F09	Sehouche lab	2R	10D(3)	*stephensi* cDNA from cDNA library	
48	37	25P09	AL153306	2R	10D(4)	*gambiae* BAC	
49	38	Ag1780	AGAP001780	2R	10D(5)	*gambiae* cDNA	
50	38	Ag1783	AGAP001783	2R	10D(5)	*gambiae* cDNA	
51	39	AsRPS6	AY237124	2R	11A(1)	*stephensi* cDNA from GENBank	
	40	10E06'	AL145314	2R	11A(2)	*gambiae* BAC	multiple
52	41	141A14'	BH367876	2R	11A(3)	*gambiae* BAC	multiple
53		Ag2931	AGAP002931	2R	11A(3)	*gambiae* cDNA	
54		Ag2932	AGAP002932	2R	11A(3)	*gambiae* cDNA	
55		Ag2933	AGAP002933	2R	11A(3)	*gambiae* cDNA	
56	41	Ag2934	AGAP002934	2R	11A(3)	*gambiae* cDNA	
57	41	Ag1757	AGAP001757	2R	11A(3)	*gambiae* cDNA	
58	41	Ag1759	AGAP001759	2R	11A(3)	*gambiae* cDNA	
59	42	04A11	AL141561	2R	11A(4)	*gambiae* BAC	
60	43	155H21	BH398459	2R	11B(1)	*gambiae* BAC	
61	44	211C05	Sehouche lab	2R	11B(2)	*stephensi* cDNA from cDNA library	
62	45	12_H09	BU038915	2R	11C(1)	*funestus* cDNA	
63	46	212D05	Sehouche lab	2R	11C(2)	*stephensi* cDNA	
64	47	671280'	BM657097	2R	11C(3)	*gambiae* cDNA	multiple
65	48	24J01'	AL152501	2R	11D(1)	*gambiae* BAC	multiple
66	48	Ag7051'	AGAP007051	2R	11D(1)	*gambiae* cDNA	multiple
67	49	04B15'	AL145409	2R	11D(2)	*gambiae* BAC	multiple
68	49	23I16'	AL610348	2R	11D(2)	*gambiae* BAC	multiple
69	50	127F13'	BH387145	2R	11D(3)	*gambiae* BAC	multiple
70	51	AF261B04	Collins lab	2R	11D(4)	*funestus* BAC	
	52	23I16'	AL610348	2R	12B(1)	*gambiae* BAC	multiple
71	53	166G9	BH383888	2R	12B(2)	*gambiae* BAC	
72	54	211B09'	Sehouche lab	2R	12B(3)	*stephensi* cDNA	multiple
73	55	18_D12	BU038940	2R	12C(1)	*funestus* cDNA	
74	56	20N10	AL150410	2R	12C(2)	*gambiae* BAC	
75	57	140D21	BH370864	2R	12D	*gambiae* BAC	
76	58	21_F12	BU038958	2R	13A	*funestus* cDNA	
77	59	Ag1954	AGAP001954	2R	13B(1)	*stephensi* cDNA	
	59	Ag1972'	AGAP001972	2R	13B(1)	*gambiae* cDNA	multiple
78	60	Ag5789'	AGAP005789	2R	13B(2)	*gambiae* cDNA	multiple
79	61	Ag1929	AGAP001929	2R	13B(3)	*gambiae* cDNA	
80	62	11_C03'	BU038902	2R	13B(4)	*funestus* cDNA	multiple
81	63	Ag1916	AGAP001916	2R	13B(5)	*gambiae* cDNA	
	64	Ag1972'	AGAP001972	2R	13C(1)	*gambiae* cDNA	multiple
82	64	Ag1899	AGAP001899	2R	13C(1)	*gambiae* cDNA	
83	65	11A13	AL145719	2R	13C(2)	*gambiae* BAC	
84	65	Ag1866	AGAP001866	2R	13C(2)	*gambiae* cDNA	
	66	Ag1985'	AGAP001985	2R	14A(1)	*gambiae* cDNA	multiple
85	67	Ag3363	AGAP003363	2R	14A(3)	*gambiae* cDNA	
86	67	Ag3365	AGAP003365	2R	14A(3)	*gambiae* cDNA	
87	68	Ag3366	AGAP003366	2R	14A(4)	*gambiae* cDNA	
88	68	Ag3369	AGAP003369	2R	14A(4)	*gambiae* cDNA	
89	69	06_E11'	BU038887	2R	14B(1)	*funestus* cDNA	multiple
90	70	Ag3374	AGAP003374	2R	14B(2)	*gambiae* cDNA	
91	71	Ag3416	AGAP003416	2R	14B(3)	*gambiae* cDNA	

92	72	Ag3423	AGAP003423	2R	14B(4)	*gambiae* cDNA	
93	73	Ag3434	AGAP003434	2R	14B(5)	*gambiae* cDNA	
94	74	12_G11	BU038914	2R	14C(1)	*funestus* cDNA	
95	75	660216	BM641646	2R	14C(2)	*gambiae* cDNA	
96	76	139N4	BH379254	2R	14C(3)	*gambiae* BAC	
	77	04B15'	AL145409	2R	15A(1)	*gambiae* BAC	multiple
97	78	AsK3MO'	AY065662	2R	15A(2)	*stephensi* cDNA	multiple
	78	127F13'	BH387145	2R	15A(2)	*gambiae* BAC	multiple
98	79	11_D07	BU038904	2R	15A(3)	*funestus* cDNA	
99	79	Ag7086'	AGAP007086	2R	15A(3)	*gambiae* cDNA	multiple
100	80	138H21	BH381119	2R	15B(1)	*gambiae* BAC	
101	81	Ag2935	AGAP002935	2R	15B(2)	*gambiae* cDNA	
	81	141A14'	BH367876	2R	15B(2)	*gambiae* BAC	multiple
102		Ag2938	AGAP002938	2R	15B(2)	*gambiae* cDNA	
103		Ag2940	AGAP002940	2R	15B(2)	*gambiae* cDNA	
104	81	01C03	AL139911	2R	15B(2)	*gambiae* BAC	
105	82	Ag3327'	AGAP003327	2R	16A(1)	*gambiae* cDNA	multiple
106	82	Ag1763	AGAP001763	2R	16A(1)	*gambiae* cDNA	
107	82	Ag1765	AGAP001765	2R	16A(1)	*gambiae* cDNA	
108	83	27I24	AL154218	2R	16A(2)	*gambiae* BAC	
109	84	04_D06	BU038877	2R	16A_B	*funestus* cDNA	
110	85	137K7	BH371689	2R	16C(1)	*gambiae* BAC	
111	86	Ag3360	AGAP003360	2R	16C(2)	*gambiae* cDNA	
112	86	Ag3359	AGAP003359	2R	16C(2)	*gambiae* cDNA	
113	86	Ag3351	AGAP003351	2R	16C(2)	*gambiae* cDNA	
114	86	Ag3350	AGAP003350	2R	16C(2)	*gambiae* cDNA	
115	87	AST028O6'	Tu lab	2R	16C(3)	*stephensi* BAC	multiple
116	88	17E12	AL148468	2R	17A(1)	*gambiae* BAC	
117	89	29F01	AL155229	2R	17A(2)	*gambiae* BAC	
118	90	AsSP11.9	AY162245	2R	17A(3)	*stephensi* cDNA from GENBank	
119	91	21I20	AL609973	2R	17A(4)	*gambiae* BAC	
120	92	OBP-1	Tu lab	2R	17A(5)		
121	93	23O01	AL152140	2R	17A(6)	*gambiae* BAC	
122	93	AST044C24	Tu lab	2R	17A(6)	*stephensi* BAC	
123	94	08O05	AL144514	2R	17A(7)	*gambiae* BAC	
	95	Ag3327'	AGAP003327	2R	17B(1)	*gambiae* cDNA	multiple
124	95	Ag3326	AGAP003326	2R	17B(1)	*gambiae* cDNA	
125	96	Ag3323	AGAP003323	2R	17B(2)	*gambiae* cDNA	
126	96	Ag3325*	AGAP003325	2R	17B(2)	*gambiae* cDNA	multiple
127	97	Ag3322	AGAP003322	2R	17B(3)	*gambiae* cDNA	
128	97	Ag3320	AGAP003320	2R	17B(3)	*gambiae* cDNA	
129	97	Ag3315	AGAP003315	2R	17B(3)	*gambiae* cDNA	
130	98	36_A12	BU038994	2R	17C(1)	*funestus* cDNA	
131	98	Ag3790	AGAP003790	2R	17C(1)	*gambiae* cDNA	
132	99	Ag3336	AGAP003336	2R	17C(2)	*gambiae* cDNA	
133	99	Ag3339	AGAP003339	2R	17C(2)	*gambiae* cDNA	
134	99	Ag3342	AGAP003342	2R	17C(2)	gambiae cDNA	
	99	AsK3MO'	AY065662	2R	17C(2)	*stephensi* cDNA from GENBank	multiple
	100	Ag1972'	AGAP001972	2R	18A(1)	*gambiae* cDNA	multiple
135	101	Ag3328	AGAP003328	2R	18A(2)	*gambiae* cDNA	
136	102	142O19	BH368703	2R	18A(3)	*gambiae* BAC	
137	103	25_E09	BU038972	2R	18B(1)	*funestus* cDNA	
138	103	129M18	BH377340	2R	18B(1)	*gambiae* BAC	
139	104	626240	BM655548	2R	18B(2)	*gambiae* cDNA	
140	104	643142	BM624660	2R	18B(2)	*gambiae* cDNA	
141	105	17_G08	BU038935	2R	18B(3)	*gambiae* BAC	
142	106	09E12	AL144757	2R	18C	*funestus* cDNA	

143	107	157B8	BH384608	2R	18D(1)	*gambiae* BAC	
144	107	106D14'	BH399330	2R	18D(1)	*gambiae* BAC	multiple
145	108	169F11	BH369697	2R	19A(1)	*gambiae* BAC	
146	109	11_E07	BU038905	2R	19A(2)	*funestus* cDNA	
147	110	211F02	Sehouche lab	2R	19B	*stephensi* cDNA from cDNA library	
148	111	11_B04	BU038900	2R	19B_C	*funestus* cDNA	
149	112	23I15	AL151968	2R	19C	*gambiae* BAC	
150	113	stBAC43	Sehouche lab	2R	19E(1)	*stephensi* BAC	
151	114	17N16	AL148799	2R	19E(2)	*gambiae* BAC	
152	115	stBAC49	Sehouche lab	2R	19E(3)	*stephensi* BAC	
153	116	stBAC62	Sehouche lab	2R	19E(4)	*stephensi* BAC	
154	117	AST018J14	Tu lab	2R	19E(het)(5)	*stephensi* BAC	

Appendix 2.8: The locations of in situ hybridization probes on *A. stephensi* 2L chromosome

#	Signal #	Clone	Accession #	Chromosome	Division	Clone type	
1	1	7016490	BM603740	2L	21B_20A(1)	*gambiae* cDNA	
2	1	669234'	BM655755	2L	21B_20A(1)	*gambiae* cDNA	multiple
3	2	702140'	BM594831	2L	21B_20A(2)	*gambiae* cDNA	multiple
4	3	139K20*	BH402428	2L	20A_B(het)(1)	*gambiae* BAC	multiple
5	4	23I16'	AL610348	2L	20A_B(het)(2)	*gambiae* BAC	multiple
6	4	627837'	BM647307	2L	20A_B(het)(2)	*gambiae* cDNA	multiple
7	4	Ag7051'	AGAP007051	2L	20A_B(het)(2)	*gambiae* cDNA	multiple
8	4	Ag7075'	AGAP007075	2L	20A_B(het)(2)	*gambiae* cDNA	multiple
9	4	178A3'	BH398965	2L	20A_B(het)(2)	*gambiae* BAC	multiple
10	5	AST041P6'	Tu lab	2L	20A_B(het)(3)	*stephensi* BAC	multiple
11	6	212D01	Sehouche lab	2L	20B(1)	*stephensi* cDNA from cDNA library	
12	7	211H02	Sehouche lab	2L	20B(2)	*stephensi* cDNA from cDNA library	
13	8	Ag10387	AGAP010387	2L	20C(1)	*gambiae* cDNA	
14	9	148K2	BH396659	2L	20C(2)	*gambiae* BAC	
15	9	106D14'	BH399330	2L	20C(2)	*gambiae* BAC	multiple
16	10	07_G04	BU038892	2L	20C(3)	*funestus* cDNA	
17	11	211G09*	Sehouche lab	2L	20C(4)	*stephensi* cDNA from cDNA library	multiple
18	12	212F07'	Sehouche lab	2L	21A(1)	*stephensi* cDNA from cDNA library	multiple
19	12	212D07'	Sehouche lab	2L	21A(1)	*stephensi* cDNA from cDNA library	multiple
20	13	126G21	BH375705	2L	21A(2)	*gambiae* BAC	
21	14	31H07'	AL156465	2L	21A(3)	*gambiae* BAC	multiple
22	15	01K17	AL140205	2L	21A(4)	*gambiae* BAC	
23	16	AF13D09	Collins lab	2L	21B(1)	*funestus* BAC	
24	17	AsGbb	AY578815	2L	21B(2)	*stephensi* cDNA from GENBank	
25	18	131K7	BH378526	2L	21B(3)	*gambiae* BAC	
26	19	211G01	Sehouche lab	2L	21B(4)	*stephensi* cDNA from cDNA library	
27	19	29L12	AL155478	2L	21B(4)	*gambiae* BAC	

28	20	04_E02	BU038879	2L	21B(5)	*funestus* cDNA	
29	20	682825	BM613652	2L	21B(5)	*gambiae* cDNA	
30	21	212E01	Sehouche lab	2L	22A(1)	*stephensi* cDNA from cDNA library	
31	21	212E05'	Sehouche lab	2L	22A(1)	*stephensi* cDNA from cDNA library	multiple
32	22	151M24	BH399147	2L	22A(2)	*gambiae* BAC	
33	23	729481'	BM577567	2L	22B(1)	*gambiae* cDNA	multiple
34	23	105O20	BH379197	2L	22B(1)	*gambiae* BAC	
35	24	139M22	BH387916	2L	22B(2)	*gambiae* BAC	
36	25	03_G10	BU038875	2L	22C(1)	*funestus* cDNA	
37	25	36_D10	BU038999	2L	22C(1)	*funestus* cDNA	
38	26	AF13G04	Collins lab	2L	23A	*funestus* microsatellite	
39	27	211D03'	Sehouche lab	2L	23B(1)	*stephensi* cDNA from cDNA library	multiple
40	28	211D02	Sehouche lab	2L	23B(2)	*stephensi* cDNA from cDNA library	
41	28	211E02'	Sehouche lab	2L	23B(2)	*stephensi* cDNA from cDNA library	multiple
42	29	AF263D12	Collins lab	2L	23B(3)	*funestus* BAC	
43	30	11G16	AL145956	2L	23C(1)	*gambiae* BAC	
44	30	211E04	Sehouche lab	2L	23C(1)	*stephensi* cDNA from cDNA library	
45	31	155I2*	BH374558	2L	24A(1)	*gambiae* BAC	multiple
46	32	145G13*	BH370252	2L	24A(2)	*gambiae* BAC	multiple
47	33	04_D01	BU038876	2L	24B(1)	*funestus* cDNA	
48	34	04F19	AL141759	2L	24B(2)	*gambiae* BAC	
49	35	124K17	BH372801	2L	24B(3)	*gambiae* BAC	
	35	155I2'	BH374558	2L	24B(3)	*gambiae* BAC	multiple
50	36	12I09	AL146524	2L	24C	*gambiae* BAC	
51	37	23_D08'	BU038965	2L	25A(1)	*funestus* cDNA	multiple
52	37	212E10	Sehouche lab	2L	25A(1)	*stephensi* cDNA from cDNA library	
53	38	211A03	Sehouche lab	2L	25B(1)	*stephensi* cDNA from cDNA library	
54	39	27P23	AL154485	2L	25B(2)	*gambiae* BAC	
55	40	180K21*	BH367855	2L	25B(3)	*gambiae* BAC	multiple
56	41	AF23E08	Collins lab	2L	25B(4)	*funestus* microsatellite	
57	42	02G07	AL140620	2L	25B(5)	*gambiae* BAC	
58	43	211G05'	Sehouche lab	2L	25B(6)	*gambiae* BAC	multiple
59	43	211F08'	Sehouche lab	2L	25B(6)	*stephensi* cDNA from cDNA library	multiple
60	44	31B09'	AL156246	2L	25B(7)	*gambiae* BAC	multiple
61	44	105F8*	BH392724	2L	25B(7)	*gambiae* BAC	multiple
	45	729481'	BM577567	2L	25C	*gambiae* cDNA	multiple
	46	31H07'	AL156465	2L	26A	*gambiae* BAC	multiple
62	47	05_D12'	BU038881	2L	26B(1)	*funestus* cDNA	multiple
63	48	132D12	BH373552	2L	26B(2)	*gambiae* BAC	
	49	05_D12*	BU038881	2L	26C(1)	*funestus* cDNA	multiple

64	50	104F13	BH377291	2L	26C(2)	*gambiae* BAC	
65	50	211D07'	Sehouche lab	2L	26C(2)	*stephensi* cDNA from cDNA library	multiple
66	51	AST041D2	Tu lab	2L	27A	*stephensi* BAC	
67	52	650820	BM650357	2L	27B(1)	*gambiae* cDNA	
68	53	130M5	BH384886	2L	27B(2)	*gambiae* cDNA	
69	54	25B13	AL152825	2L	27C(1)	*gambiae* BAC	
	55	23I16'	AL610348	2L	27C2)	*gambiae* BAC	multiple
	56	Ag7051'	AGAP007051	2L	27C(3)	*gambiae* cDNA	multiple
	56	Ag7075'	AGAP007075	2L	27C(3)	*gambiae* cDNA	multiple
70	57	08F18	AL144178	2L	27C(4)	*gambiae* BAC	
71	58	126M2	BH386745	2L	28A(1)	*gambiae* BAC	
	59	211G05*	Sehouche lab	2L	28A(2)	*stephensi* cDNA from cDNA library	multiple
72	59	28G12	AL154806	2L	28A(2)	*gambiae* BAC	
73	60	28J05	AL154796	2L	28A(3)	*gambiae* BAC	
74	61	AF263A06	Collins lab	2L	28A(4)	*funestus* BAC	
75	62	12_F01	BU038911	2L	28C(1)	*funestus* cDNA	
76	63	AST012A11	Tu lab	2L	28C(2)	*stephensi* BAC	

Appendix 2.9: The locations of in situ hybridization probes on *A. stephensi* 3R chromosome

#	Signal #	Clone	Accession#	Chromosome	Division	Clone type	
1	1	212A02	Sehouche lab	3R	29A	*stephensi* cDNA from cDNA library	
2	2	AsUbi	AJ415521	3R	29A_B	*stephensi* cDNA from GENBank	
3	3	10J02	AL608684	3R	29B(1)	*gambiae* BAC	
4	3	211E11*	Sehouche lab	3R	29B(1)	*stephensi* cDNA from cDNA library	multiple
5	3	10D24	AL145308	3R	29B(1)	*gambiae* BAC	
6	4	19_F10	BU038947	3R	29B(2)	*funestus* cDNA	
7	5	AsSki	AY578814	3R	29B(3)	*stephensi* cDNA from GENBank	
8	6	212A07	Sehouche lab	3R	29C(1)	*stephensi* cDNA from cDNA library	
9	7	212D06	Sehouche lab	3R	29C(2)	*stephensi* cDNA from cDNA library	
10	8	109G18*	BH368579	3R	29D(1)	*gambiae* BAC	
11	9	06_F07	BU038888	3R	29D(2)	*funestus* cDNA	
12	10	16_C12	BU038929	3R	29D(3)	*funestus* cDNA	
13	11	178A3*	BH398965	3R	29E	*gambiae* BAC	multiple
14	12	24K22	AL152579	3R	30A(1)	*gambiae* BAC	
15	13	23K05	AL610371	3R	30A(2)	*gambiae* BAC	
16	14	145J17	BH373436	3R	30A(3)	*gambiae* BAC	
17	15	13J12	AL147066	3R	30B(1)	*gambiae* BAC	
18	16	31B09'	AL156246	3R	30B(2)	*gambiae* BAC	multiple
19	17	03N21	AL141483	3R	30B(3)	*gambiae* BAC	
20	18	AST021C18	Tu lab	3R	30C(1)	*stephensi* BAC	
21	19	152P1	BH375235	3R	30C(2)	*gambiae* BAC	
22	20	135D12	BH371101	3R	30C(3)	*gambiae* BAC	
23	21	14E16	AL147454	3R	31A(1)	*gambiae* BAC	
24	22	AST018E12	Tu lab	3R	31A(2)	*stephensi* BAC	
25	23	212B05	Sehouche lab	3R	31A(3)	*stephensi* cDNA from cDNA library	
26	24	12A10	AL146243	3R	31A(4)	*gambiae* BAC	
27	25	AF12D10	Collins lab	3R	31A(5)	*funestus* microsatellite	

28	26	211H06	Sehouche lab	3R	31A(6)	*stephensi* cDNA from cDNA library	
29	26	212G05	Sehouche lab	3R	31A(6)	*stephensi* cDNA from cDNA library	
30	27	211C04	Sehouche lab	3R	31B(1)	*stephensi* cDNA from cDNA library	
31	28	AST026I17	Tu lab	3R	31B(2)	*stephensi* BAC	
32	28	AF261H03	Collins lab	3R	31B(2)	*funestus* BAC	
33	29	211E12'	Sehouche lab	3R	31B(3)	*stephensi* cDNA from cDNA library	multiple
34	30	23J24	AL152013	3R	31B(4)	*gambiae* BAC	
35	31	AF264E05	Collins lab	3R	31C(1)	*funestus* BAC	
36	32	AsSerpin6	Eapen	3R	31C(2)	*stephensi* cDNA from GENBank	
37	33	05C06	AL607943	3R	31C(3)	*gambiae* BAC	
38	34	212H07	Sehouche lab	3R	32A(1)	*stephensi* cDNA from cDNA library	
39	35	AsMad	AY578813	3R	32A(2)	*stephensi* cDNA from GENBank	
40	36	61_F02	BU039004	3R	32C(1)	*funestus* cDNA	
41	37	10F04	AL145343	3R	32C(2)	*gambiae* BAC	
42	38	211H04	Sehouche lab	3R	32C(3)	*stephensi* cDNA from cDNA library	
43	39	29A01	AL611251	3R	33A	*gambiae* BAC	
44	40	211E10	Sehouche lab	3R	33B(1)	*stephensi* cDNA from cDNA library	
45	41	AST029K10	Tu lab	3R	33B(2)	*stephensi* BAC	
46	42	23G11	AL151884	3R	33C	*gambiae* BAC	
47	43	31B02	AL146601	3R	34B(1)	*gambiae* BAC	
	44	211E11'	Sehouche lab	3R	34B(2)	*stephensi* cDNA from cDNA library	multiple
	45	211E12'	Sehouche lab	3R	34C	*stephensi* cDNA from cDNA library	multiple
	46	31B09*	AL156246	3R	35A	*gambiae* BAC	multiple
48	47	125G23	BH389922	3R	35B(1)	*gambiae* BAC	
49	48	211B10	Sehouche lab	3R	35B(2)	*stephensi* cDNA from cDNA library	
50	49	AsK3MO'	AY065662	3R	35B(3)	*stephensi* cDNA from GENBank	multiple
51	50	212F07*	Sehouche lab	3R	35B(4)	*stephensi* cDNA from cDNA library	multiple
52	51	211F01	Sehouche lab	3R	36A(1)	*stephensi* cDNA from cDNA library	
53	52	25M15	AL153203	3R	36A(2)	*gambiae* BAC	
54	53	AF264H03	Collins lab	3R	36A(3)	*funestus* BAC	
55	54	02J17	AL140752	3R	36A(4)	*gambiae* BAC	
56	55	212B12	Sehouche lab	3R	36B	*stephensi* cDNA	
57	56	180K21'	BH367855	3R	36C(1)	*gambiae* BAC	multiple
58	57	129I2	BH372692	3R	36C(2)	*gambiae* BAC	
59	58	11I19	AL146017	3R	37B(1)	*gambiae* BAC	
60	59	163H10	BH385794	3R	37B(2)	*gambiae* BAC	
61	60	30P20'	AL156193	3R	37B(3)	*gambiae* BAC	multiple
62	60	105F8'	BH392724	3R	37B(3)	*gambiae* BAC	multiple
63	61	145G13'	BH370252	3R	37B(4)	*gambiae* BAC	multiple
64	62	24J01'	AL152501	3R	37B(5)	*gambiae* BAC	multiple
65	62	139K20'	BH402428	3R	37B(5)	*gambiae* BAC	multiple
66		Ag7051'	AGAP007051	3R	37B(5)	*gambiae* cDNA	multiple
67		Ag6898'	AGAP006898	3R	37B(5)	*gambiae* cDNA	multiple
68	63	61_G06	BU039005	3R	37B(6)	*funestus* cDNA	
	64	AsK3MO'	AY065662	3R	37C(1)	*stephensi* cDNA from	multiple

						GENBank	
69	65	04B15'	AL145409	3R	37C(2)	*gambiae* BAC	multiple
70	66	211A05	Sehouche lab	3R	37D(1)	*stephensi* cDNA from cDNA library	
71	66	19_D07	BU038946	3R	37D(1)	*funestus* cDNA	
72	67	669234'	BM655755	3R	37D(2)	*gambiae* cDNA	multiple
73	68	627112	BM636978	3R	37D(3)	*gambiae* cDNA	
74	69	AST041P6'	Tu lab	3R	37D (het)(4)	*stephensi* BAC	multiple

Appendix 2.10: The locations of in situ hybridization probes on *A. stephensi* 3L chromosome

#	Signal #	Clone	Accession #	Chromosome	Division	Clone type	
1	1	211F08'	Sehouche lab	3L	38B(1)	*stephensi* cDNA from cDNA library	multiple
2	2	03_G12	AL141218	3L	38B(2)	*gambiae* BAC	
3	3	101C3	BH388218	3L	38B(3)	*gambiae* BAC	
4	4	157I18	BH367786	3L	38C	*gambiae* BAC	
5	5	AsHyp16	AY162228	3L	38E	*stephensi* cDNA	
6	6	211B11	Sehouche lab	3L	38F(1)	*stephensi* cDNA from cDNA library	
7	6	211B09'	Sehouche lab	3L	38F(1)	*stephensi* cDNA from cDNA library	multiple
8	7	211D07'	Sehouche lab	3L	39A(1)	*stephensi* cDNA from cDNA library	multiple
9	8	212G03	Sehouche lab	3L	39A(2)	*stephensi* cDNA from cDNA library	
10	9	AsAG5	AY162227	3L	39B(1)	*stephensi* cDNA from GENBank	
11	9	AST004P4	Dr. Tu	3L	39B(1)	*stephensi* BAC	
12	10	212D07'	Sehouche lab	3L	39B(2)	*stephensi* cDNA from cDNA library	multiple
13	10	211E02'	Sehouche lab	3L	39B(2)	*stephensi* cDNA from cDNA library	multiple
14	11	10D09	AL145285	3L	39B(3)	*gambiae* BAC	
15	12	02A04	AL140380	3L	39C(1)	*gambiae* BAC	
16	13	140N16	BH384642	3L	39C(2)	*gambiae* BAC	
17	14	02A19	AL140406	3L	39D	*gambiae* BAC	
18	15	31L22	AL156623	3L	40A(1)	*gambiae* BAC	
19	15	Ag7069	AGAP007069	3L	40A(1)	*gambiae* cDNA	
20	15	Ag7070	AGAP007070	3L	40A(1)	*gambiae* cDNA	
21	15	Ag7077	AGAP007077	3L	40A(1)	*gambiae* cDNA	
22	15	Ag7086*	AGAP007086	3L	40A(1)	*gambiae* cDNA	multiple
23	16	Ag5783	AGAP005783	3L	40A(2)	*gambiae* cDNA	
24	16	160O9	BH367937	3L	40A(2)	*gambiae* BAC	
25	17	146D17'	BH400736	3L	40A(3)	*gambiae* BAC	multiple
26	17	Ag5779	AGAP005779	3L	40A(3)	*gambiae* cDNA	
27	18	716320	BM606621	3L	40A(4)	*gambiae* cDNA	
28	19	AsSP53.7	AY162233	3L	40A(5)	*stephensi* cDNA from GENBank	
29	19	Ag6919	AGAP006919	3L	40A(5)	*gambiae* cDNA	
30	20	10I04	AL145436	3L	40B(1)	*gambiae* BAC	
31	20	Ag6921	AGAP006921	3L	40B(1)	*gambiae* cDNA	
32	20	Ag7075'	AGAP007075	3L	40B(1)	*gambiae* cDNA	multiple
33	21	26E08	AL610881	3L	40C(1)	*gambiae* BAC	
34	22	212D10	Sehouche lab	3L	40C(2)	*stephensi* cDNA from cDNA library	
35	23	211F11	Sehouche lab	3L	40D(1)	*stephensi* cDNA from cDNA library	
36	23	Ag6263	AGAP006263	3L	40D(1)	*gambiae* cDNA	

37	24	Ag3325'	AGAP003325	3L	40D(2)	*gambiae* cDNA	multiple
38	25	AsPPO1	AY559300	3L	40D(3)	*stephensi* cDNA from GENBank	
39	26	04P13	AL142126	3L	40D(4)	*gambiae* BAC	
40	27	03C15	AL141092	3L	41A	*gambiae* BAC	
41	28	16_G10'	BU038933	3L	41B(1)	*funestus* cDNA	multiple
42	29	23_D08*	BU038965	3L	41B(2)	*funestus* cDNA	multiple
43	30	Ag7006'	AGAP007006	3L	41C	*gambiae* cDNA	multiple
44	31	211H03	Sehouche lab	3L	42A(1)	*stephensi* cDNA from cDNA library	
45	32	11_C03'	BU038902	3L	42A(2)	*funestus* cDNA	multiple
46	32	16_F07	BU038931	3L	42A(2)	*funestus* cDNA	
47	33	212F12	Sehouche lab	3L	42A(3)	*stephensi* cDNA from cDNA library	
48	34	212E11	Sehouche lab	3L	42A(4)	*stephensi* cDNA from cDNA library	
49	34	211B02	Sehouche lab	3L	42A(4)	*stephensi* cDNA from cDNA library	
50	34	101L14	BH382930	3L	42A(4)	*gambiae* BAC	
51	35	26L15	AL153718	3L	42B(1)	*gambiae* BAC	
52	36	212D03	Sehouche lab	3L	42B(2)	*stephensi* cDNA	
53	37	109B13'	Sehouche lab	3L	42B(3)	*gambiae* BAC	multiple
54	38	104C14	BH391906	3L	42C(1)	*gambiae* BAC	
55	39	27O10	AL154432	3L	42C(2)	*gambiae* BAC	
56	40	Ag6945	AGAP006945	3L	43A(1)	*gambiae* cDNA	
57	41	Ag6958	AGAP006958	3L	43A(2)	*gambiae* cDNA	
58	42	Ag6965	AGAP006965	3L	43A(3)	*gambiae* cDNA	
59	42	Ag6968'	AGAP006968	3L	43A(3)	*gambiae* cDNA	multiple
60	43	02K19	AL140790	3L	43A_B	*gambiae* BAC	
61	43	211G09'	Sehouche lab	3L	43A_B	*stephensi* cDNA from cDNA library	multiple
	44	211E02'	Sehouche lab	3L	43B	*stephensi* cDNA from cDNA library	multiple
62	45	Ag7010	AGAP007010	3L	43C(1)	*gambiae* cDNA	
63	46	09_C11	BU038897	3L	43C(2)	*funestus* cDNA	
64	47	Ag6900	AGAP006900	3L	43C(3)	*gambiae* cDNA	
65	47	Ag6903	AGAP006903	3L	43C(3)	*gambiae* cDNA	
66	48	Ag6898*	AGAP006898	3L	43C(4)	*gambiae* cDNA	multiple
67	49	AF263B12	Collins lab	3L	44A(1)	*funestus* BAC	
	50	Ag6968'	AGAP006968	3L	44A(2)	*gambiae* cDNA	multiple
68	50	150F12	BH385494	3L	44A(2)	*gambiae* BAC	
69	51	211A02	Sehouche lab	3L	44A(het)(3)	*stephensi* cDNA from cDNA library	
70	52	12G16	AL146467	3L	44A(het)(4)	*gambiae* BAC	
71	53	25G06'	AL610709	3L	44A (het)(5)	*gambiae* BAC	multiple
	54	109B13'	BH385043	3L	44B(1)	*gambiae* BAC	multiple
72	55	AsK3MO*	AY065662	3L	44B(2)	*stephensi* cDNA from GENBank	multiple
73	55	Ag5711	AGAP005711	3L	44B(2)	*gambiae* cDNA	
74	56	131F22	BH390198	3L	44C(1)	*gambiae* BAC	
	56	146D17'	BH400736	3L	44C(1)	*gambiae* BAC	multiple
75	56	Ag5761	AGAP005761	3L	44C(1)	*gambiae* cDNA	
76	57	Ag5765	AGAP005765	3L	44C(2)	*gambiae* cDNA	
77	58	A5771	AGAP005771	3L	44C(3)	*gambiae* cDNA	
78	58	Ag5774	AGAP005774	3L	44C(3)	*gambiae* cDNA	
79	58	Ag5775	AGAP005775	3L	44C(3)	*gambiae* cDNA	
80	58	Ag5778	AGAP005778	3L	44C(3)	*gambiae* cDNA	
81	59	Ag7063	AGAP007063	3L	44C(4)	*gambiae* cDNA	
82	59	Ag7065	AGAP007065	3L	44C(4)	*gambiae* cDNA	
83	59	Ag7068	AGAP007068	3L	44C(4)	*gambiae* cDNA	

84	60	Ag7051'	AGAP007051	3L	44C(5)	*gambiae* cDNA	multiple
85	61	Ag7046	AGAP007046	3L	44C(6)	*gambiae* cDNA	
86	62	Ag7039	AGAP007039	3L	44C(7)	*gambiae* cDNA	
87	63	Ag7031	AGAP007031	3L	44C(8)	*gambiae* cDNA	
88	63	Ag7023	AGAP007023	3L	44C(8)	*gambiae* cDNA	
89	63	Ag7022	AGAP007022	3L	44C(8)	*gambiae* cDNA	
90	64	Ag7019	AGAP007019	3L	44C(9)	*gambiae* cDNA	
91	65	702140'	BM594831	3L	45A(1)	*gambiae* cDNA	multiple
92	66	04C08	AL607764	3L	45A(2)	*gambiae* BAC	
93	66	61_E02	BU039003	3L	45A(2)	*funestus* cDNA	
	67	Ag7006*	AGAP007006	3L	45C(1)	*gambiae* cDNA	multiple
94	67	Ag7007	AGAP007007	3L	45C(1)	*gambiae* cDNA	
95	67	Ag7008	AGAP007008	3L	45C(1)	*gambiae* cDNA	
96	67	Ag7014	AGAP007014	3L	45C(1)	*gambiae* cDNA	
97	67	Ag5789'	AGAP005789	3L	45C(1)	*gambiae* cDNA	multiple
98	68	11_F09	BU038906	3L	45C(2)	*funestus* cDNA	
99	69	AF262H10	Collins lab	3L	45C(3)	*funestus* BAC	
100	70	Ag6990	AGAP006990	3L	45C(4)	*gambiae* cDNA	
101	71	Ag6974	AGAP006974	3L	45C(5)	*gambiae* cDNA	
	72	Ag7006'	AGAP007006	3L	45C(6)	*gambiae* cDNA	multiple
102	73	18_G01	BU038941	3L	46A(1)	*funestus* cDNA	
103	74	142L24	BH399793	3L	46A(2)	*gambiae* BAC	
104	75	211B01	Sehouche lab	3L	46A(3)	*stephensi* cDNA from cDNA library	
105	76	211B03	Sehouche lab	3L	46B(1)	*stephensi* cDNA from cDNA library	
	77	16_G10*	BU038933	3L	46B(2)	*stephensi* cDNA	multiple
	78	212D07'	Sehouche lab	3L	46B(3)	*stephensi* cDNA from cDNA library	multiple
106	78	211F03	Sehouche lab	3L	46B(3)	*stephensi* cDNA from cDNA library	
107	78	212C07'	Sehouche lab	3L	46B(3)	*stephensi* cDNA from cDNA library	multiple
108	79	11_B01	BU038899	3L	46C(1)	*funestus* cDNA	
109	80	07_A01	BU038890	3L	46C(2)	*funestus* cDNA	
110	81	26_A01	BU038977	3L	46C(3)	*funestus* cDNA	
111	82	671280*	BM657097	3L	46C(4)	*gambiae* cDNA	multiple
112	83	26_B05	BU038978	3L	46C_D	*funestus* cDNA	
113	84	AST028O6'	Tu lab	3L	46D(1)	*stephensi* BAC	multiple
114	85	178B1	BH372501	3L	46D(2)	*gambiae* BAC	
115	86	06_G08	BU038889	3L	46D(3)	*funestus* cDNA	

Appendix 3.1: The chromosomal locations of markers in *A. stephensi* and *A. funestus* and the coordinates in *A. gambiae* genome.

Markers on the X chromosome									
Marker in *A. stephensi*	**Markers in *A. funestus***	**Clone name**	**Accession**	**VectorBase Gene ID**	**Genomic Location in *A. gambiae*[1]**	**e-value**	**Chromosomal location[2]**		
							A. gambiae	***A. funestus***	***A .stephensi***
	1	06_E11	BU038887	AGAP000007	84,525-88,380	1e-92	X:4C	X:3D*; 2R:18C	X:3A*; 2R:14B
1		25E24	AL610701		459,415-587,625	0	X:4C	nd	X:5A
2		126O17	BH404578		1,397,208-1,512,610	0	X:4C	nd	X:5A
	2	FUN Q	AY116021	AGAP000087	1,451,178-1,454,861	7e-03	X:4C	X:5CD	nd
3		20N08	AL150407		2,363,189-2,453,423	0	X:4A	nd	X:1A
4		27D16	AL154011		2,898,136-3,033,190	0	X:4A	nd	X:1C
5		150E4	BH388828		2,986,103-3,033,400	0	X:4A	nd	X:1C
	3	14_B05	BU038921	AGAP000180	3,000,117-3,001,769	1e-101	X:4A	X:3A	
6		17I02	AL148591		3,147,527-3,274,341	0	X:4A	nd	X:1B
7		24J01	AL152501		3,632,054-3,771,323	0	X:3D	nd	X:4A*; 2R:11D; 3R:37B
8		31G14	AL156439		3,944,341-4,029,848	0	X:3D	nd	X:4A
9		19N19	AL609649		5,014,419-5,147,712	0	X:3C	nd	X:1C
10		26C03	AL153390		5,566,114-5,672,612	0	X:3C	nd	X:1C
11		28F08	AL154661		6,001,831-6,002,087	0	X:3B	nd	X:3A
12		21P07	AL151067		7,585,473-7,691,519	0	X:2C	nd	X:3C
13		17O15	AL609326		8,365,624-8,366,297	0	X:2B	nd	X:1A
14		155N1	BH384248		9,048,469-9,170,197	0	X:2B	nd	X:2A
15		26H02	AL610921		11,106,310-11,234,911	0	X:1D	nd	
16	4	18_G03	BU038942	AGAP000679	12,117,478-12,119,325	3e-65	X:1C	X:1B*; 2R:12C; 3L:41C	X:4B
17		28J20	AL154814		13,022,001-13,149,832	0	X:1C	nd	X:2C
18		24K09	AL152559		13,406,025-13,503,738	0	X:1A	nd	X:3B
19		21F12	AL150696		14,068,002-14,190,611	0	X:1B	nd	X:1A

20		31A03	AL611577		14,227,297-14,355,594	0	X:1B	nd	X:3C
	5	17_E02	BU038934	AGAP000794 AGAP006891	14,442,142-14,443,554 39,426,797-39,428,212	1e-137 2e-75	X:1B*; 2L:26A	X:1C	nd
21		24G07	AL152397		14,548,696-14,659,022	0	X:1B	nd	X:3B
22		Ag0803	XM_311343	AGAP000803	14,738,909-14,744,472	0	X:5A	nd	X:3B
	6	CYP9K1	AY987362	AGAP000818	15,240,763-15,242,715	1e-09	X:5A	X:5B	nd
23		13E12	AL146915		15,322,088-15,435,208	0	X:5A	nd	X:4A
	7	61_C09	BU039001	AGAP000824	15,336,624-15,339,506	1e-33	X:5A	X:2B	nd
	8	FUN E	AY116009		15,736,855-15,736,883 23,418,653-23,418,681	9e-02	X:5A	X:5D	nd
	9	CYP4G21	AY648704	AGAP000877	16,619,141-16,621,043	4e-32	X:5AB	X:4C	nd
24		138A5	BH380601		16,677,135-16,799,196	0	X:5B	nd	X:1C
	10	98_D11	BU039017	AGAP000886	16,737,992-16,741,889	1e-14	X:5B	X:1C	nd
	11	26_G11	BU038979	AGAP000953 AGAP002395	18,339,741-18,342,792 20,904,493-20,905,615	1e-102 7e-78	X:5C*; 2R:12B	X:1B*; 2L:28C	X:4B
25	12	18_D02	BU038939	AGAP001036	19,936,008-19,937,213	2e-87	X:6	X:3D	X:1B
	13	18_B08	BU038937	AGAP001076	22,937,938-22,947,129	1e-60	X:6	X:6	nd

Markers on 2R

Markers in *A. stephensi*	**Markers in *A. funestus***	**Clone name**	**Accession**	**VectorBase Gene ID**	**Genomic Location in *A. gambiae*[1]**	**e-value**	**Chromosomal location**		
							***A. gambiae*[2]**	***A. funestus*[3]**	***A .stephensi***
1		155B23	BH368691		528,525-648,904	0	2R:7A	nd	2R:7A
		23_G01	BU038967	AGAP001141	582,418-585,357	2e-34	2R:7A	2R:10C5*,17 B	nd
		01_D07	BU038871	AGAP001164	673,798-674,927	1e-174	2R:7A	2R:10C*,7C	nd
2		04L11	AL141975		1,599,882-1,708,302	0	2R:7B	nd	2R:7A
	1	21_F03	BU038956	AGAP001215	1,608,204-1,609,027	1e-23	2R:7B	2R:7B	nd
3	2	01_H04	BU038873	AGAP001306	2,802,248-2,803,419	1e-165	2R:7B	2R:7A	2R:7B
4	3	21_F12	BU038958	AGAP001380	4,050,701-4,051,820	2e-69	2R:8A	2R:12D	2R:13A
		07_E10	BU038891	AGAP001394	4,153,647-4,167,690	3e-49	2R:8A	2R:12D	nd
5		140D21	BH370864		4,677,078-4,755,258	0	2R:8A	nd	2R:12D
	4	06_E01	BU038885	AGAP001420	4,739,226-4,740,361	1e-162	2R:8A	2R:12E	nd
6		AsOBP-7	EU816361	AGAP001556	6,152,011-6,168,838	1e-38	2R:8C	nd	2R:9C
		36_B06	BU038996	AGAP001588 AGAP012818	6,491,921-6,495,261 31,394,884-31,395,776	3e-80 1e-82	2R:8C; UNKN	2R:8E	nd

7		153L12	BH380684		6,813,959-6,894,483	0	2R:8D	nd	2R:9C
	5	36_E01	BU039000	AGAP001617	6,827,869-6,831,707	1e-75	2R:8D	2R:8D	nd
8	6	12_G10	BU038913	AGAP001721	8,925,685-8,926,882	1e-105	2R:8E	2R:15C	2R:8A
		16_A10	BU038927	AGAP001734	9,093,431-9,095,845	2e-22	2R:8E	2R:15C*; 2L:25D	2R:7A
9		Ag1759	XM_321324	AGAP001759	9,479,549-9,483,291	0	2R:8E	nd	2R:11A
10		Ag1763	XM_321320	AGAP001763	9,523,856-9,528,904	0	2R:9A	nd	2R:16A
11		Ag1780	XM_321295	AGAP001780	10,216,796-10,230,335	0	2R:9A	nd	2R:16A
12		25P09	AL153306		10,530,256-10,650,334	0	2R:9A	nd	2R:10D
	7	AFND5	AF171035	AGAP001797	10,545,856-10,583,613	5.5e-18	2R:9A	15B	nd
13		11A13	AL145719		11,505,616-11,604,069	0	2R:9B	nd	2R:14B
		FUN O	AY116019		11,565,685-11,565,728	3e-05	2R:9B	2R:18A	nd
	8	23_B02	BU038961	AGAP001903	11,975,829-11,977,395	1e-111	2R:9B	2R:18B	nd
14		20N10	AL150410		12,874,429-12,976,648	0	2R:9C	nd	2R:12C
15		Ag1980	XM_321082	AGAP001980	13,087,584-13,091,963	0	2R:9C	nd	2R:10A
	9	36_B02	BU038995	AGAP001983	13,133,425-13,135,252	1e-48	2R:9C	2R:9B	nd
16		Ag2015	XM_321040	AGAP002015	13,936,706-13,955,790	0	2R:10A	nd	2R:9D
	10	11_D03	BU038903	AGAP002020 AGAP012731	14,012,136-14,013,179 25,798,127-25,799,580	4.5e-118 5.2e-59	2R:10A UNKN	2R:9A	2R:10A
17		22D14	AL151203		14034511-14034752	0	2R:10A	nd	2R:10A
18		135P16	BH387168		14,892,477-14,952,998	0	2R:10B	nd	2R:9B
19	11	04_D06	BU038877	AGAP002166	16,812,718-16,815,833	6e-35	2R:11A	2R:10C	2R:16AB
20		137K7	BH371689		17,597,143-17,688,592	0	2R:11A	nd	2R:16C
	12	61_D05	BU039002	AGAP002213	17,648,143-17,653,542	3e-91	2R:11A	2R:10C	nd
21	13	08_E06	BU038895	AGAP002317	19,444,433-19,447,708	4e-61	2R:11C	2R:16A	2R:10D
22		AF13C05	F.H. Collins lab		19,883,222-19,883,271[5]	4e-17	2R:11C	2R:16A	2R:10C
	14	06_D06	BU038884		19,946,249-19,946,745	1e-42	2R:11C	2R:16B	nd
23		29F01	AL155230		20,468,607-20,555,519	0	2R:11C	nd	2R:17A
	15	15_G03	BU038926	AGAP002413	21,027,264-21,030,999	6e-75	2R:12A	2R:12A	nd
24		25_E09	BU038972	AGAP002440	21,425,189-21,426,085	3e-83	2R:12B	2R:12B	2R:18B
		129M18	BH377340		21,521,726-21,621,799	0	2R:12B		2R:18B
	16	13_F11	BU038919	AGAP002457	21,573,636-21,587,145	6e-57	2R:12B	2R:12B2	
		626240	BM655548	AGAP002465	21,825,430-21,827,282	1e-136	2R:12B	nd	2R:18B
	17	15_F10	BU038925	AGAP002465	21,825,430-21,827,282	9e-65	2R:12B	2R:12B	nd
25	18	17_G08	BU038935	AGAP002468	21,835,973-21,836,716	1e-153	2R: 12B	2R:12B;2L:23C	2R:18C
26		AsGPROR7	Z. Tu lab	AGAP002560	22,849,252-22,858,650	1e-159	2R: 12C	nd	2R:9A

27	19	06_B01	BU038882	AGAP002606	23,835,495-23,836,568	1e-90	2R:12C	2R:14D	2R:9A
		20_E04	BU038952	AGAP002608	23,842,548-23,842,828	1e-27	2R:12C	2R:14D	nd
28		09N07	AL145079		25,146,360-25,266,886	0	2R:12D	nd	2R:8C
29		18L04	AL149296		26,187,553-26,313,071	0	2R:12E	nd	2R:8B
30		155H21	BH398459		27,025,144-27,125,121	0	2R:13A		2R:11B
	20	AFND32	AY291371		27,025,145-27,125,121	0.040	2R:13A	2R:15E	nd
		AFND37	AY291373	AGAP002790	27,262,932-27,311,956	1e-24	2R:13A	2R:15E	nd
31		138H21	BH381119		28,310,572-28,311,109	0	2R:13C	nd	2R:15B
32		AsRPS6	AY237124	AGAP002919	29,609,284-29,611,282	0	2R:13C	nd	2R:11A
33		Ag2934	XM_001237408	AGAP002934	29,835,568-29,836,999	0	2R: 13C	nd	2R:11A
34		Ag2935	XM_311967	AGAP002935	29,839,387-29,840,621	0	2R:13C	nd	2R:15B
	21	25_H11	BU038976	AGAP002935	29,839,388-29,840,621	8e-49	2R: 13C	2R:10B	nd
		01C03	AL139911		29,974,464-30,077,977	0	2R:13D	nd	2R:15B
35		27I24	AL154218		30,150,936-30,271,431	0	2R:13D	nd	2R:16A
36		139N4	BH379254		30,694,330-30,796,747	0	2R:13D	nd	2R:14C
		66_G12	BU039012	AGAP002994	30,716,522-30,717,395	2e-75	2R, 13D	2R, 17C*; 3R, 34A	nd
37		31M01	AL611707		31,181,535-31,317,988	0	2R:13E	nd	2R:8C
	22	03_D09	BU038874	AGAP003024	31,230,560-31,232,558	5e-27	2R:13E	2R:17C	nd
38		09E12	AL144757		33,575,891-33,693,463	0	2R:14C	nd	2R:18C
		23_C09	BU038963	AGAP003184	33,620,457-33,622,184	1e-82	2R, 14C	X, 4C 2R, 12B*	nd
39	23	11_D07	BU038904	AGAP003209	33,903,940-33,905,170	1e-143	2R:14C	2R:13A	2R:15A[4]
40		21I20	AL609973		34,742,360-34,873,652	0	2R:14D	nd	2R:17A
41		AsOBP1	Z. Tu lab	AGAP003309	35,643,035-35,644,609	1e-89	2R:14D	nd	2R:17A
42		23O01	AL152140		35,770,783-35,790,690	0	2R:14E	nd	2R:17A
		22_H10	BU038960	AGAP003312	35,775,371-35,775,597	1e-51	2R:14E	2R:13B	nd
43		Ag3315	XM_319546	AGAP003315	35,837,690-35,839,243	0	2R:15A	nd	2R:17B
44		Ag3342	XM_314239	AGAP003342	36,307,756-36,311,720	0	2R:15A	nd	2R:17C
45		Ag3351	XM_001688498	AGAP003351	36478446-36480531	0	2R:15A	nd	2R:16C
46		Ag3363	XM_314265	AGAP003363	36,850,789-36,853,077	0	2R:15A	nd	2R:14A
	24	27_E05	BU038983	AGAP003384	37,194,861-37,195,629	1e-54	2R:15B	2R:18C	nd
	25	12_G11	BU038914	AGAP003416	37,466,896-37,468,119	3e-40	2R:15BC	2R:18C	nd
47		Ag3434	XM_311720	AGAP003434	37,703,243-37,711,020	0	2R:15BC	nd	2R:14B
48	26	12_H09	BU038915	AGAP003500	38,738,669-38,739,898	5e-82	2R:15C	2R:18D	2R:11C
	27	66_E07	BU039009	AGAP003553	39,452,027-39,453,046	7e-44	2R:15D	2R:16C	nd
49	28	AF261B04	F.H. Collins lab		39,751,219-39,751,269	2e-15	2R:15D	2R:16C	2R:11D

		66_A02	BU039006	AGAP003625	40,683,416-40,683,582	1e-37	2R:15E	2R:12C	nd
50		169F11	BH369697		40,944,146-41,044,921	0	2R:15E	nd	2R:19A
	29	29_F03	BU038988	AGAP003650 AGAP012611	41,002,243-41,003,465 19,943,725-19,944,865	5e-48 1e-45	2R:15E; UNKN	2R:11B	nd
51	30	11_E07	BU038905	AGAP003664	41,338,854-41,360,919	1e-131	2R:16A	2R:14C	2R:19A
		36_A12	BU038994	AGAP003790	43,342,500-43,352,336	0	2R:16B	2R:8C*; 3R:35E	nd
52		142O19	BH368703		45,428,742-45,560,792	0	2R:16D	nd	2R:18A
	31	13_A06	BU038916	TCLAG15867 1	45,474,529-45,474,606 20,332,357-20,332,644	2e-31 2e-31	2R:16D UNKN	2R:13D	nd
53		08O05	AL144514		47,277,152-47,354,855	0	2R:17B	nd	2R:17A
	32	13_C03	BU038918	AGAP003971	47,284,629-47,308,582	3e-55	2R:17B	2R:13C	nd
54		157B8	BH384608		48,013,258-48,137,831	0	2R:17C	nd	2R:18D
		20_A10	BU038950	AGAP004085	49,526,117-49,527,342	5e-45	2R:17C	2R:16C*, 18C1, 18C4; 2L:23D, 26A; 3R: 36E; 3L:41A, 42B	nd
55		166G9	BH383888		52,013,039-52,095,531	0	2R:18B	nd	2R:12B
56	33	18_D12	BU038940	AGAP004247	53,207,213-53,209,347	7e-71	2R:18C	2R:14B	2R:12C
57		AsSP11.9	AY162245	AGAP004316	54,393,078-54,393,473	2e-28	2R:18D	nd	2R:17A
58		211F02	EX227558	AGAP004422	55,869,598-55,870,844	0	2R:19A	nd	19B
59		23I15	AL151968		56,667,003-56,767,697	0	2R:19B	nd	2R:19C
60	34	11_B04	BU038900	AGAP004552	57,542,706-57,543,865	1e-132	2R:19C	2R:19C	2R:19BC
61		17N16	AL148800		58,537,340-58,618,071	0	2R:19D	nd	2R:19E
62		StBAC62	Shouche lab	AGAP004662	60,236,329-60,253,235	1e-89	2R:19D	nd	2R:19E
Markers on 2L									
Markers in *A. stephensi*	**Markers in *A. funestus***	**Clone name**	**Accession**	**VectorBase Gene ID**	**Genomic Location in *A. gambiae*[1]**	**e-value**	**Chromosomal location**		
							***A. gambiae*[2]**	***A. funestus*[3]**	***A .stephensi***
1		AsHyp16	AY162228	AGAP004799	3,720,848-3,721,878	1e-103	2L:20D	nd	3L:38E
2		101C3	BH388218		5,708,924-5,858,418	0	2L:21A	nd	3L:38B
	1	28_C07	BU038985	AGAP004904	5,771,985-5,780,988	4e-67	2L:21A	3R:35C	nd
3		03G12	AL141218		6,081,630-6,182,049	0	2L:21A	nd	3L:38B
	2	29_H01	BU038990	AGAP004929	6,114,439-6,115,700	2e-22 2e-22	2L:21A UNKN	3R:35B	nd
	3	FUN D	AY116008		6,734,718-6,734,834	1e-34	2L:21A	3R:35B	nd
4		02A19	AL140406		8,667,456 - 8,791,152	0	2L:21C	nd	3L:39D
	4	04_D07	BU038878	AGAP005037	8,668,402-8,700,721	8e-30	2L:21C	3R:35A	nd
5		02A04	AL140380		9,902,459- 9,981,236	0	2L:21D	nd	3L:39C

	5	25_E12	BU038973	AGAP005117	10,241,196-10,244,380	1e-97	2L:21D	3R:36F	nd
6		10I04	AL145436		11,534,512-11,595,103	0	2L:21E	nd	3L:40B
7		157I18	BH367786		13,513,086-13,564,568	0	2L:21F	nd	3L:38C
8		26E08	AL610881		14,601,776-14,728,697		2L:22A	nd	3L:40C
9		23_D08[8]	BU038965	AGAP005410	15,211,390-15,213,382	5e-20	2L:22B	3R:35C*, 34A; 2R:16B,19A	3L:41B*; 2L:25A
10		03C15	AL141092		16,516,545-16,614,317		2L:22C	nd	3L:41A
11		27O10	AL154432		17,882,981-17,974,647	0	2L:22D	nd	3L:42C
	6	95_H01	BU039015	AGAP005618	17,920,080-17,921,140	4e-27	2L:22D	3R:31D	nd
12		104C14	BH391906		18,512,073-18,616,664		2L:22E		3L:42C
	7	20_D11	BU038951	AGAP005712	19,277,265-19,282,704	1e-58	2L:22E	3R:33C	nd
13		131F22	BH390198		20,364,135-20,459,325	0	2L:22F	nd	3L:44C
	8	AFND19	AF171049	AGAP005770	20,396,064-20,405,726	9.9e-84	2L:22F	3R:34A	nd
14		SuaPh6_1.8E coRI	NA	AGAP005780	20,535,740-20,538,254 20,538,082-20,543,536	0	2L: 23A	nd	3L:40A
	9	30_G04	BU038991	AGAP005862	21,199,038-21,200,529	1e-133	2L: 23A	3R:33D	nd
15		AsSP53.7	AY162233	AGAP005822	21,865,434-21,867,120	7e-86	2L:23B	nd	3L:40A
16		716320	BM606621	AGAP005838	22,321,985-22,327,760	0	2L:23B	nd	3L:40A
	10	66_E11	BU038987	AGAP005838	22,321,985-22,327,760	1e-134	2L:23B	3R:33D	nd
		AsK3MO	AY065662	AGAP005948	23,985,166-23,988,714	0	2L:23C	nd	3L:44B*; 2R:15A,17C; 3R:35B, 37C
17		12G16	AL146467		24,626,085-24,724,371	0	2L:23C	nd	3L:44A
	11	29_E12	BU039020	AGAP006015	24,671,856-24,674,367	4e-36	2L:23C	3R:33A	nd
18		211A02	EX227513	AGAP006037	25,521,114-25,523,716	1e-105	2L:23D	nd	3L:44A
19		150F12	BH385494		25,924,194-26,087,837	0	2L:23D	nd	3L:44A
	12	36_A10	BU038993	AGAP006071	26,055,839-26,056,556	8e-25	2L:23D	3R:35F	nd
20	13	61_E02	BU039003	AGAP006148	27,158,330-27,159,179	9e-86	2L:24A	3R:30C	3L:45A
21		AsPPO1	AY559300	AGAP006258	28,702,474-28,705,612	1e-103	2L:24B	nd	3L:40D[4]
	14	18_G09	BU038943	AGAP006263	28,771,405-28,774,168	1e-115	2L:24B	3R:35F	nd
22		04P13	AL142126		29,574,218-29,659,442	0	2L:24B	nd	3L:40D
23		140N16	BH384642		30,990,832-31,090,638	0	2L:24D	nd	3L:39C
	15	21_E03	BU038955	AGAP006389	31,048,592-31,050,091	2e-35	2L:24D	3R:35F	nd
24		AsAG5	AY162227	AGAP006421	31,693,742-31,694,830	2e-66	2L:24D	nd	3L:39B
25		AST004P4	Z. Tu lab		31,715,321-31,760,529	0	2L:24D	nd	3L:39A

26		212G03	EX227637	AGAP006442	32,055,094-32,055,857	8E-39	2L:24D	nd	3L:39A
27		10D09	AL145285		32,640,047-32,736,529	0	2L:25A	nd	3L:39B
28		02K19	AL140790		35,622,251-35,720,581	0	2L:25C	nd	3L:43A
		109B13	BH385033		36,603,004-36,704,974	0	2L:25D	nd	3L:42B, 44B
	16	95_D09	BU039013	AGAP006677	36,653,143-36,654,063	2e-48	2L:25D	3R:33C	nd
29		212D03	EX227607	AGAP006709	37,118,475-37,119,488	4e-30	2L:25D	nd	3L:42B
30		26L15	AL153718		37,497,029-37,625,484	0	2L:25D	nd	3L:42B
31		101L14	BH382930		38,518,994-38,594,734	0	2L:26A	nd	3L:42A
	17	06_C09	BU038883	AGAP006795	38,567,686-38,568,743	13e-14	2L:26A	3R:32B	nd
32	18	16_F07	BU038931	AGAP006861	39,215,802-39,216,191	1e-129	2L:26B	3R:36E	3L:42A
33		211H03	EX227578	AGAP006871	39,285,040-39,287,251	1e-71	2L:26B	nd	3L:42A
34	19	09_C11	BU038897	AGAP006918	39,995,907-39,997,107	1e-63	2L:26C	3R:36D	3L:43C
35	20	AF262H10	F.H. Collins lab	AGAP006975	40,442,200-40,442,600	8e-55	2L:26C	3R:35F	3L:45C
36	21	11_F09	BU038906	AGAP006996	40,536,371-40,538,248	1e-100	2L:26D	3R:35F	3L:45C
37		Ag7063	XM_308701	AGAP007063	42,126,193-42,127,751	0	2L:26D	nd	3L:44C
38		Ag7070	XM_001688008	AGAP007070	42,178,250-42,181,793	0	2L:27A	nd	3L:40A
39		31L22	AL156623		42,206,288-42,309,500	0	2L:27A	nd	3L:40A
		AFND33	AY291372	AGAP007078	42,216,309-42,216,467	1e-12	2L:27A	3R:33C	nd
40		04C08	AL607764		43,540,182-43,634,335	0	2L:27A	nd	3L:45A
	22	08_B09	BU038894	AGAP007160	43,603,779-43,607,408	1e-48	2L:27A	3R:30C	nd
41	23	06_G08	BU038889	AGAP007249	44,638,197-44,642,288	5e-56	2L:27C	3R:30C	3L:46D
42		178B1	BH372501		45,026,766-45,130,891	0	2L:27C	nd	3L:46D
		25_D11	BU039019	AGAP007297	45,066,291-45,067,164	1e-116	2L:27C	3R:30B	nd
43		26_B05	BU038978	AGAP007309	45,256,867-45,259,724	4e-21	2L:27C	3R:31C	3L:46CD
	24	23_E09	BU038966	AGAP007347	45,981,005-45,981,721	9e-09	2L:27D	3R:30A	nd
44		26_A01	BU038977	AGAP007362	46,062,101-46,079,300	1e-76	2L:27D	3R:30A	3L:46C[4]
45	25	07_A01	BU038890	AGAP007406	46,337,490-46,341,061	8e-58	2L:27D	3R:29D	3L:46C
46	26	11_B01	BU038899	AGAP007508	46,995,122-46,995,883	2e-96	2L:28A	3R:29C	3L:46C
		14_C12	BU038922	AGAP007558 AGAP007414	47,556,815-47,558,944 46,395,202-46,395,819	0 2e-16	2L:28B* 2L:28A	3R:29C	nd
47		211B01	EX227521	AGAP007618	48,340,253-48,341,468	3e-48	2L:28C	nd	3L:46A
48	27	18_G01	BU038941	AGAP007643	48,608,644-48,628,119	7e-69	2L:28C	3R:29B	3L:46A
49		142L24	BH399793		49,243,380-49,333,455	0	2L:28D	nd	3L:46A

Markers on 3R

Markers in *A. stephensi*	Markers in *A.*	Clone name	Accession	VectorBase Gene ID	Genomic Location in *A. gambiae*[1]	e-value	Chromosomal location		
							A. gambiae[2]	*A. funestus*[3]	*A .stephensi*

	funestus								
1		212A07	EX227590	AGAP007747	217,370-218,365	4e-22	3R:29A	nd	3R:29C
2		AsSki	AY578814	AGAP007776	620,123-621,199	1e-110	3R:29A	nd	3R:29B
3	1	19_F10	BU038947	AGAP007786	838,035-840,304	1e-167	3R:29A	2L:27D	3R:29B[4]
4		10J02	AL608684		1,421,006-1,514,782	0	3R:29A	nd	3R:29B
	2	11_H04	BU038907	AGAP007827	1,483,634-1,486,754	0	3R:29A	2L:27D	nd
5		AsUbi[8]	AJ415521	AGAP007927 AGAP001971 AGAP008001	2,920,166-2,921,050 12,998,558-12,999,247 3,991,227-3,993,208	1e-100 1e-69 4e-66	3R:29C 2R:9C 3R:29D	nd	3R:29AB
6		109G18	BH368579		3,963,987-4,060,834	0	3R:29CD	nd	3R:29D
	3	01_C07	BU038870	AGAP008001	3,991,227-3,993,208	6e-84 1e-29	3R:29CD 2R:29D	2L:27C	nd
7	4	06_F07	BU038888	AGAP008053	4,869,662-4,871,563	1e-179	3R:29D	2L:27B	3R:29D
8	5	16_C12	BU038929	AGAP008054	4,880,891-4,881,271	1e-162	3R:29D	2L:27B	3R:29D
9		178A3	BH398965		5,630,291-5,756,885	0	3R:29D	nd	3R:29E*; X:6A(het); 2L:20AB(het)
10		24K22	AL152579		6,541,606-6,644,135	0	3R:30A	nd	3R:30A
	6	CYP6Z3	AY193727	AGAP008217	6,971,669-6,973,217	0	3R:30A	2L:26D	nd
	7	CYP6Z1	AF487535	AGAP008219	6,976,539-6,978,081	0	3R:30A	2L:26D	nd
11		23K05	AL610371		7,477,602-7,608,857	0	3R:30AB	nd	3R:30A
	8	21_D06	BU038954	AGAP008233	7,543,137-7,543,741	2e-25	3R:30AB	2L:26D	nd
	9	AFUB10	AY029717	AGAP008241	7,587,587-7,590,544	2e-32	3R:30B	2L:26CD	nd
12		145J17	BH373436		8,596,001-8,683,047	0	3R:30B	nd	3R:30A
	10	06_E04	BU038886	AGAP008294	8,630,719-8,631,546	2e-22	3R:30B	2R, 18C 2L, 26C*	nd
13		13J12	AL147066		8,716,522-8,842,512	0	3R:30B	nd	3R:30B
	11	Fun P	AY116020	AGAP008304	8,793,109-8,802,070	1e-40	3R:30B	2L:26C	nd
14		03N21	AL141483		10,019,888-10,142,236	0	3R:30C	nd	3R:30B
	12	13_H04	BU038920	AGAP008369	10,088,721-10,093,543	4e-64	3R:30C	2L:26A	nd
15		AST021C18	Z. Tu lab		10,123,759-10,129,187	4e-78	3R:30C	nd	3R:30C
16	13	AF264E05	F.H. Collins lab		10,774,634-10,774,684	9e-05 (35 bp)	3R:30D	2L:24B	3R:31C
17		AsMad	AY578813	AGAP008551	12,545,417-12,547,013.	0	3R:30E	nd	3R:32A
18	14	61_F02	BU039004	AGAP008647	13,939,170-13,940,623	4e-55	3R:31A	2L:22B	3R:32C
19		10F04	AL145343		16,038,336-16,118,290	0	3R:31B	nd	3R: 32C
	15	13_C02	BU038917	AGAP008725	16,069,773-16,071,472	8e-75	3R:31B	2L:22D	nd

20		211H04	EX227579	AGAP008727	16,158,263-16,159,216	1E-96	3R:31BC	nd	3R:32C
21		29A01	AL611251		17,147,606-17,221,188	0	3R:31C	nd	3R:33A
	16	23_B09	BU038962	AGAP008762	17,144,470-17,150,005	1e-57	3R:31C	2L:23A	nd
	17	AFUB2	AY029709		17,154,074-17,154,296	5e-39	3R:31C	2L:23A	nd
22		31B02	AL146601		19,607,644-19,703,771	0	3R:32A	nd	3R:34B
23		211E11[8]	EX227555	AGAP008923	21,042,156-21,046,135	3e-14	3R:32C	nd	3R:34B, 3R:29B*
24		25M15	AL153203		25,546,570-25,694,004	0	3R:33B	nd	3R:36A
	18	98_F05	BU039018	AGAP009096	25,578,233-25,580,290	4e-61	3R:33B	2L:24D	nd
25	19	AF264H03	F.H. Collins lab		25,983,063-25,983,530	2.0e-7	3R:33B	2L:24C	3R:36A
	20	AFND18	AF171048		26,560,353-26,560,412	7e-12	3R:33C	2L:24C	nd
	21	FUN L	AY116016		26,816,555-26,816,623	8.0e-18	3R:33C	2L:24C	nd
26		02J17	AL140752		27,508,676-27,597,800	0	3R:33C		3R:36A
27		AsSerpin6	M. Jacobs-Lorena lab	AGAP009212	28,811,997-28,818,217	1e-128	3R:33C	nd	3R:31C
28		152P1	BH375235		30,494,707-30,573,515	0	3R:33D	nd	3R:30C
29		135D12	BH371101		31,230,315-31,384,115	0	3R:33D	nd	3R:30C
	22	25_H10	BU038975	AGAP009324	31,268,756-31,270,480	7e-36	3R:33D	2L:28C	nd
30		14E16	AL147453		31,471,236-31,630,052		3R:33D	nd	3R:31A
31		AST018E12	Z. Tu lab		32,213,137-32,216,768	7e-12	3R:34A	nd	3R:31A
32		212B05	EX227597	AGAP009441	33,496,836-33,501,776	5e-66	3R:34A	nd	3R:31A
33		12A10	AL146243		33,562,651-33,640,581	0	3R:34B	nd	3R:31A
34	23	AF12D10	F.H. Collins lab		34,898,584-34,898,654	3.0e-14	3R:34B	2L:28B	3R:31A
34		211H06	EX227580	AGAP009508	34,969,675-34,970,901	1e-113	3R:34B	nd	3R:31A
36		211C04	EX227533	AGAP009515	35,054,613-35,076,625	4e-33	3R:34BC	nd	3R:31A
37		AST026I17	Z. Tu lab		35,131,422-35,136,776	2e-18	3R:34C	nd	3R:31B
	24	66_E10	BU039010	AGAP009537	35,605,568-35,607,573	1e-130	3R:34C	2L:28A	nd
38		23J24	AL152013		36,477,003-36,591,879	0	3R:34C	nd	3R:31B
	25	AFND23	AY291367		36,927,785-36,927,839	3e-10	3R:34D	2L:24A	nd
39		211E10	EX227554	AGAP009610	37,001,849-37,003,632	7e-22	3R:34D	nd	3R:32C
40		AST029K10	Z. Tu lab		37,408,194-37,420,606	2e-12	3R:34D	nd	3R:33B
41		23G11	AL151883		38,599,533-38,718,166	0	3R:35B	nd	3R:33C
42		05C06	AL607943		39,502,006-39,614,515	0	3R:35B	nd	3R:31C
43		163H10	BH385794		40,492,051-40,582,084	0	3R:35B	nd	3R:37B
44		31B09	AL156246		41,547,299-41,646,263	0	3R:35B	nd	3R:35A*,30B;

									2L:25B
45		30P20	AL156192		42,487,765-42,579,439	0	3R:35C	nd	3R:37B*, X:6A(het)
46		125G23	BH389922		43,532,477-43,646,278	0	3R:35C	nd	3R:35B
47		211B10	EX227528	AGAP009839	44,086,477-44,088,125	5e-24	3R:35D	nd	3R:35B
48		211F01	EX227557	AGAP009944	45,804,678-45,806,458	1e-39	3R:36AB	nd	3R:36A
49		129I2	BH372692		46,515,634-46,627,757	0	3R:36B	nd	3R:
50		11I19	AL146017		48,498,702-48,595,676	0	3R:36D	nd	37B
51	26	61_G06	BU039005	AGAP010142	49,370,093-49,372,300	2e-59	3R:37A	2L:20D	3R:37B
	27	11_C01	BU038901	AGAP010207	50,715,566-50,720,049	3e-45	3R:37C	2L:20C	nd
52		211A05	EX227515	AGAP010216	50,847,429-50,848,247	7e-43	3R:37C	nd	3R:37C
	28	19_D07	BU038946	AGAP010216	50,847,429-50,848,247	3e-89	3R:37C	2L:20C*; 3L:44D,46A	3R:37C
53		627112	BM636978	AGAP010252	51,662,576-51,663,611	0	3R:37D	nd	3R:37D
	29	66_A04	BU039008	AGAP010252	51,662,576-51,663,611	6e-33	3R:37D	2L:20B	nd
Markers on 3L									
Markers in *A. stephensi*	**Markers in *A. funestus***	**Clone name**	**Accession**	**VectorBase Gene ID**	**Genomic Location in *A. gambiae*[1]**	**e-value**	**Chromosomal location**		
							***A. gambiae*[2]**	***A. funestus*[3]**	***A .stephensi***
	1	15_B11	BU038923	AGAP10387	2,444,111-2,447,808	4e-95	3L:38B	3L:38C	nd
1		212D01		AGAP010364	2,246,808-2,247,865	9e-09	3L:38B	nd	2L:20C
2		148K2	BH396659		2,709,606-2,840,913	0	3L:38B	nd	2L:20C
	2	01_F07	BU038872	AGAP010404	2,779,188-2,784,335	7e-69	3L:38B	3L:38C	nd
3	3	07_G04	BU038892	AGAP010445	3,752,432-3,757,036	3e-73	3L:38C	3L:39A	2L:20C
	4	25_E01	BU038971	AGAP10469	4,015,796-4,031,373	1e-42	3L:38C	3L:39A	nd
4		139K20	BH402428		4,607,484-4,702,647	0	3L:38C	nd	2L:20A*
5		151M24	BH399147		4,982,245-5,081,675	0	3L:38C	nd	2L:22A
	5	27_B04	BU038981	AGAP010500	5,056,173-5,058,317	3e-41	3L:38C	3L:39A*,46A	nd
6		105O20	BH379197		5,545,686-5,611,702	0	3L:39A	nd	2L:22B
7		139M22	BH387916		6,371,454-6,495,502	0	3L:39A	nd	2L:22B
	6	29_D12	BU038986	AGAP010565	6,421,493-6,424,472	1e-102	3L:39A	3L:39B	nd
8	7	03_G10	BU038875	AGAP010657	8,102,799-8,116,056	9e-61	3L:39B	3L:41A	2L:22C[4]
		36_D10	BU038999	AGAP010657	8,102,799-8,116,056	1e-152	3L:39B	3L:41A	2L:22C
9		126G21	BH375705		8,852,435-8,989,815	0	3L:39C	nd	2L:21A
	8	12_B09	BU038909	AGAP010716	8,934,066-8,937,311	1e-60	3L:39C	3L:41A	nd
10		180K21	BH367855		9,407,395-9,639,192	0	3L:39C	nd	2L:25B*; 3R:36C
	9	15_F08	BU038924	AGAP010792	10,412,154-10,413,607	1e-171	3L:40A	3L:43A	nd
11		AST012A11	Z. Tu lab		11,244,143-11,260,272	8e-19	3L:40B	nd	2L:28C

12	10	12_F01	BU038911		11,726,542-11,727,596	1e-91	3L:40B	3L:43B	2L:28C
13		105F8	BH392724		12,527,507-12,636,192		3L:40B	nd	3R:37B; 2L:25B*
14		02G07	AL140620		13,725,402-13,907,560	0	3L:40C	nd	2L:25B
15		31H07	AL156465		14,419,173-14,419,723	0	3L:41A	nd	2L:21A,26A*
16		08F18	AL144178		14,578,788-14,693,933	0	3L:41A	nd	2L:27C
17		25B13	AL152825		16,548,366-16,667,678	0	3L:41B	nd	2L:27C
18		130M5	BH384886		17,484,256-17,609,908	0	3L:41C	nd	2L:27B
19		650820	BM650357	AGAP011160	18,112,740-18,113,828	0	3L:41C	nd	2L:27C
	11	36_A01	BU038992	AGAP011160	18,112,740-18,113,828	1e-162	3L:41C	3L:42B	nd
20		AST041D2	Z. Tu lab		18,440,434-18,526,417	5e-29	3L:41D	nd	2L:27A
21		27P23	AL154485		20,554,497-20,635,434	0	3L:42B	nd	2L:25B
	12	21_F09	BU038957	AGAP011291	20,578,594-20,579,662	1e-79	3L:42B	3L:44D	nd
22		211A03	EX227514	AGAP011298	20,688,570-20,689,982	7e-36	3L:42B	nd	2L:25B
	13	29_F07	BU038989	AGAP011402	23,869,077-23,869,479	3e-74	3L:42C	3L:44A*, 2R: 17B	nd
23		155I2	BH374558		23,877,430-23,986,176		3L:43A	nd	2L:24B,24A*
24		145G13	BH370252		24,356,682-24,417,969		3L:43A	nd	2L:24A*; 3R:37B
25		04F19	AL141759		26,250,362-26,366,741	0	3L:43B	nd	2L:24B
	14	04_D01	BU038876	AGAP011514	26,294,061-26,295,679	1e-36	3L:43B-gene	3L:46D	2L:24B
26		124K17	BH372801		27,566,602-27,673,312		3L:43C	nd	2L:24B
27		12I09	AL146524		28,283,793-28,363,998	0	3L:43C	nd	2L:24C
	15	27_A08	BU038980	AGAP011581	28,373,084-28,373,665	2e-87	3L:43C	3L:46B*; X:4B,3	nd
28	16	AF13G04	F.H. Collins lab		30,431,704-30,431,754	1e-10	3L:43D	3L:45C	2L:23A
29		211D02	EX227540	AGAP011644	30,654,707-30,670,454	5.3e-66	3L:43D	nd	2L:23B
30	17	AF263D12	F.H. Collins lab		32,057,098-32,057,136	1.1e-6[7]	3L:44A	3L:45B	2L:23B
31		11G16	AL145956		32,599,213-32,701,072	0	3L:44B	nd	2L:23C
32		669234	BM655755	AGAP011788	33,271,412-33,273,378	6e-44	3L:44B	3L:40A	2L:21B-20A*[8] 3R:37D
	18	25_H01	BU038974						
33	19	04_E02	BU038879	AGAP011828	33,872,056-33,877,105	1e-176	3L:44C	3L:40B	2L:21B
34		29L12	AL155478		33,874,267-33,982,382	0	3L:44C	nd	2L:21B
	20	21_G01	BU038959	AGAP011839	33,925,153-33,927,790	5e-42	3L:44C	3L:40B	nd
35		131K7	BH378526		34,547,460-34,646,125	0	3L:44D	nd	2L:21B
36		AsGbb	AY578815	AGAP011934	35,258,637-35,260,057	2e-88	3L:44D	nd	2L: 21B
37		01K17	AL140205		35,505,997-35,582,996	0	3L:44D	nd	2L:21A

38		132D12	BH373552		37,599,013-37,651,873	0	3L:45C	nd	2L:26B
	21	20_G09	BU038953	AGAP012096	37,631,554-37,633,381	1e-101	3L:45C	3L:41B*, 3R:29C	nd
	22	19_H06	BU038949	AGAP012131	37,960,351-37,965,533	2e-42	3L:45C	3L:40C	nd
39		104F13	BH377291		37,971,196-38,080,288	0	3L:45C	nd	2L:26C
	23	05_A10	BU038880	AGAP012334	40,732,639-40,738,864	7e-47	3L:46C	3L:44A	nd
40		28J05	AL154795		40,982,673-41,158,235	0	3L:46C	nd	2L:28A
41		28G12	AL154707		41,538,346-41,662,654	0	4L:46D	nd	2L:28A
42		126M2	BH386745		41,792,338-41,892,789	0	3L:46D	nd	2L:28A
	24	36_C05	BU038998	AGAP012418	41,834,603-41,835,368	1e-44	3L:46D	3L:46B	nd

[1)] Coordinates are given for *An. gambiae* genes and BAC clones available at VectorBase.org (*440*).
[2)] *An. gambiae* and *An. funestus* divisions and subdivisions are taken from VectorBase.org (*440*) and from (*1, 441*), respectively.
[3)] Asterisks indicate primary BLAST hits and hybridization signals.
[4)] Data taken from (*18*).
[5)] Coordinates of the BLAST hits.
[7)] BLAST with distant homology option in Ensembl.org (*442*).
[8)] Location supported by the homology of banding pattern

Appendix 3.2: The chromosomal locations of 231 probes on the polytene chromosomes of *A. stephensi* and the coordinates in *A. gambiae* for analyzing the distribution of markers.

Chromosome X				
Marker#	**Clone name**	**Accession**	**Locations in *gambiae*[1]**	**Locations in *stephensi*[2]**
1	21F12	AL150696	14.068	1.3
2	20N08	AL150407	2.363	2
3	17O15	AL609326	8.365	2.4
4	18_D02	BU038939	19.936	2.6
5	17I02	AL148591	3.147	3.36
6	138A5	BH380601	16.7	4
7	26C03	AL153390	5.566	4.2
8	19N19	AL609649	5.014	5
9	27D16	AL154011	2.898	5.3
10	150E4	BH388828	2.986	5.5
11	155N1	BH384248	9.048	6.55
12	26H02	AL610921	11.106	7.06
13	28J20	AL154814	13.022	8.5
14	28F08	AL154661	6.001	9.95
15	24G07	AL152397	14.548	10.2
16	Ag0803	XM_311343	14.738	10.5
17	24K09	AL152559	13.406	10.7
18	21P07	AL151067	7.585	11.8
19	31A03	AL611577	14.227	11.95
20	13E12	AL146915	15.322	12.08
21	24J01	AL152501	3.632	12.2
22	31G14	AL156439	3.944	12.3
23	18_G03	BU038942	12.117	14.38
24	25E24	AL610701	0.459	16.05
25	126O17	BH404578	1.397	16.45
Total			24.4Mb	23.4 centimere
Chromosome 2R				
Marker#	**Clone name**	**Accession**	**Locations in *gambiae*[1]**	**Locations in *stephensi*[2]**
1	155B23	BH368691	0.5	0.65
2	01_D07	BU038871	0.673	0.9
3	211H11	Sehouche lab	1.275	1.35
4	04L11	AL141975	1.6	1.85
5	01_H04	BU038873	2.802	3.2
6	12_G10	BU038913	8.925	6.2
7	18L04	AL149296	26.187	8.5
8	09N07	AL145079	25.146	9.45
9	31M01	AL611707	31.2	11.1
10	06_B01	BU038882	23.835	11.3
11	GPROR7	Tu lab	22.849	12.4
12	135P16	BH387168	14.892	14.1
13	OBP-7	Tu lab	6.152	14.45
14	105H10	BH368219	6.4	14.85
15	153L12	BH380684	6.8	15.3
16	22D14	AL151203	14.034	17.55
17	23F12	AL151849	13.422	17.8
18	Ag1980	XM_321082	13.087	18.6
19	AF13C05	Collins lab	19.883	21.18
20	08_E06	BU038895	19.444	21.5
21	25P09	AL153306	10.5	22.55
22	Ag1783	XM_321284	10.326	23.25
23	AsRPS6	AY237124	29.609	24.25
24	Ag2934	XM_001237408	29.835	24.95

25	04A11	AL141561	9.422	25.2
26	155H21	BH398459	27.025	25.9
27	12_H09	BU038915	38.738	27.5
28	AF261B04	Collins lab	39.751	29.9
29	166G9	BH383888	52.013	33.3
30	20N10	AL150410	12.874	33.8
31	18_D12	BU038940	53.207	34.2
32	140D21	BH370864	4.7	35.25
33	21_F12	BU038958	4.05	36.5
34	11A13	AL145719	11.5	42.8
35	12_G11	BU038914	37.466	44.9
36	139N4	BH379254	30.69	45.7
37	11_D07	BU038904	33.903	47.4
38	138H21	BH381119	28.31	47.6
39	Ag2935	XM_311967	29.839	49.8
40	27I24	AL154218	30.15	51
41	04_D06	BU038877	16.812	51.7
42	137K7	BH371689	17.5	52.65
43	Ag3351	XM_001688498	36.478	52.8
44	29F01	AL155230	20.468	53.9
45	AsSP11.9	AY162245	54.393	54.25
46	21I20	AL609973	34.742	55
47	AsOBP1	Tu lab	35.643	55.75
48	23O01	AL152140	35.77	55.95
49	08O05	AL144514	47.28	56.25
50	36_A12	BU038994	43.342	58.5
51	142O19	BH368703	45.428	60
52	25_E09	BU038972	21.425	61
53	17_G08	BU038935	21.835	61.3
54	09E12	AL144757	33.575	61.65
55	157B8	BH384608	48.013	66.15
56	169F11	BH369697	41	66.6
57	11_E07	BU038905	41.338	67
58	211F02	Sehouche lab	55.869	68
59	11_B04	BU038900	57.542	68.3
60	23I15	AL151968	56.667	70.1
61	17N16	AL148800	58.537	73.5
62	StBAC62	Sehouche lab	60.236	74
Total			61.5Mb	74.4 centimere
Chromosome 2L				
Marker#	**Clone name**	**Accession**	**Locations in *gambiae*[1]**	**Locations in *stephensi*[2]**
1	AsHyp16	AY162228	3.72	2.05
2	AF262H10	Collins lab	40.442	3
3	11_F09	BU038906	40.536	3.4
4	157I18	BH367786	13.513	4.35
5	101C3	BH388218	5.7	4.9
6	03G12	AL141218	6.081	5.05
7	02A19	AL140406	8.667	6.6
8	140N16	BH384642	30.99	7.05
9	02A04	AL140380	9.902	8
10	10D09	AL145285	32.64	8.7
11	AsAG5	AY162227	31.693	9.8
12	AST004P4	Tu lab	31.715	9.9
13	212G03	Sehouche lab	32.055	10
14	04P13	AL142126	29.574	11.2
15	AsPPO1	AY559300	28.702	11.4
16	26E08	AL610881	14.601	12.45
17	10I04	AL145436	11.534	13.1
18	AsSP53.7	AY162233	21.865	14.75

19	716320	BM606621	22.321	14.9
20	SuaPh6_1.8EcoRI	NA	20.535	15.2
21	Ag7070	XM_001688008	42.178	15.3
22	31L22	AL156623	42.206	15.55
23	23_D08[8]	BU038965	15.211	17.4
24	03C15	AL141092	16.516	18.5
25	27O10	AL154432	17.882	19.8
26	104C14	BH391906	18.512	20.5
27	212D03	Sehouche lab	37.118	21.1
28	26L15	AL153718	37.497	21.4
29	101L14	BH382930	38.518	22.1
30	16_F07	BU038931	39.215	23.2
31	211H03	Sehouche lab	39.285	23.45
32	09_C11	BU038897	39.995	24
33	02K19	AL140790	35.622	26.8
34	Ag7063	XM_308701	42.126	28.6
35	131F22	BH390198	20.364	28.85
36	12G16	AL146467	24.626	31.65
37	211A02	Sehouche lab	25.521	32.3
38	150F12	BH385494	25.924	32.8
39	61_E02	BU039003	27.158	35.1
40	04C08	AL607764	43.54	35.4
41	06_G08	BU038889	44.638	36.4
42	178B1	BH372501	45.026	36.7
43	26_B05	BU038978	45.256	37.1
44	26_A01	BU038977	46.062	38
45	07_A01	BU038890	46.337	38.25
46	11_B01	BU038899	46.995	38.85
47	211B01	Sehouche lab	48.34	40.2
48	142L24	BH399793	49.243	40.58
49	18_G01	BU038941	48.608	41.3
Total			49.4Mb	41.5 centimere
Chromosome 3R				
Marker#	**Clone name**	**Accession**	**Locations in *gambiae*[1]**	**Locations in *stephensi*[2]**
1	AsUbi8	AJ415521	2.92	1.1
2	10J02	AL608684	1.421	1.6
3	19_F10	BU038947	0.838	2.35
4	AsSki	AY578814	0.62	2.5
5	212A07	Sehouche lab	0.217	3.3
6	109G18	BH368579	3,963	3.6
7	06_F07	BU038888	4.869	4.25
8	16_C12	BU038929	4.88	4.45
9	178A3	BH398965	5.63	5.1
10	24K22	AL152579	6.541	6.3
11	23K05	AL610371	7.477	7.1
12	145J17	BH373436	8.596	7.7
13	13J12	AL147066	8.716	8.15
14	03N21	AL141483	10.019	9
15	AST021C18	Tu lab	10.123	9.2
16	152P1	BH375235	30.494	10
17	135D12	BH371101	31.23	10.85
18	14E16	AL147453	31.471	11
19	AST018E12	Tu lab	32.213	11.5
20	212B05	Sehouche lab	33.496	11.9
21	12A10	AL146243	33.562	12.3
22	AF12D10	Collins lab	34.898	13
23	211H06	Sehouche lab	34.969	13.4
24	211C04	Sehouche lab	35.054	13.65
25	AST026I17	Tu lab	35.131	13.8

26	23J24	AL152013	36.477	14.5
27	AF264E05	Collins lab	10.774	15.65
28	AsSerpin6	Jacobs-Lorena lab	28.811	16.15
29	05C06	AL607943	39.502	16.55
30	AsMad	AY578813	12.545	17.75
31	61_F02	BU039004	13.939	19.05
32	10F04	AL145343	16.038	20.4
33	211H04	Sehouche lab	16.158	20.6
34	29A01	AL611251	17.147	21
35	211E10	Sehouche lab	37.001	22.4
36	AST029K10	Tu lab	37.408	23.05
37	23G11	AL151883	38.599	24.25
38	31B02	AL146601	19.607	26.75
39	211E11	Sehouche lab	21.042	28.7
40	31B09	AL156246	41.547	30.7
41	125G23	BH389922	43.532	32.4
42	211B10	Sehouche lab	44.086	33.05
43	211F01	Sehouche lab	45.804	34.35
44	25M15	AL153203	25.546	34.65
45	AF264H03	Collins lab	25.983	35.3
46	02J17	AL140752	27.508	36.05
47	129I2	BH372692	46.515	38.8
48	11I19	AL146017	48.498	41.1
49	163H10	BH385794	40.492	41.3
50	30P20	AL156192	42.487	41.6
51	61_G06	BU039005	49.37	41.9
52	211A05	Sehouche lab	50.847	43.25
53	627112	BM636978	51.662	44.1
Total			53.2Mb	45.5 centimere
Chromosome 3L				
Marker#	**Clone name**	**Accession**	**Locations in *gambiae*[1]**	**Locations in *stephensi*[2]**
1	07_G04	BU038892	3.752	1.4
2	148K2	BH396659	2.709	2.25
3	212D01		2.246	2.75
4	139K20	BH402428	4.607	5
5	6692348	BU038974	33.271	5.3
6	04_E02	BU038879	33.872	5.7
7	29L12	AL155478	33.85	6.05
8	131K7	BH378526	34.547	6.75
9	AsGbb	AY578815	35.258	7.15
10	01K17	AL140205	35.505	7.6
11	126G21	BH375705	8.852	8.35
12	03_G10	BU038875	8.102	9.35
13	139M22	BH387916	6.371	10.5
14	105O20	BH379197	5.545	11.4
15	151M24	BH399147	4.982	11.7
16	11G16	AL145956	32.599	12.5
17	AF263D12	Collins lab	32.057	13
18	211D02	Sehouche lab	30.654	14.1
19	AF13G04	Collins lab	30.431	15.3
20	12I09	AL146524	28.3	15.7
21	124K17	BH372801	27.566	16.1
22	04F19	AL141759	26.25	16.9
23	145G13	BH370252	24.356	18.8
24	155I2	BH374558	23.877	19.2
25	105F8	BH392724	12.527	20.8
26	02G07	AL140620	13.725	21.5
27	180K21	BH367855	9.407	22.3
28	27P23	AL154485	20.55	22.65

29	211A03	Sehouche lab	20.688	22.85
30	104F13	BH377291	37.971	26.25
31	132D12	BH373552	37.6	27.4
32	31H07	AL156465	14.419	30
33	08F18	AL144178	14.578	30.3
34	25B13	AL152825	16.548	31.5
35	130M5	BH384886	17.484	32.2
36	650820	BM650357	18.112	32.55
37	AST041D2	Tu lab	18.44	33
38	AST012A11	Tu lab	11.244	35
39	12_F01	BU038911	11.726	35.3
40	28J05	AL154795	40.982	37.7
41	28G12	AL154707	41.538	38.25
42	126M2	BH386745	41.8	38.6
Total			42Mb	38.8 centimere

[1)] The coordinates of *A. gambiae* in Mb. The coordinates for *A. gambiae* cDNAs and BAC clones were available at VectorBase.org (*440*). For the coordinates of other clones were the blast hits using the sequences blastn against *A. gambiae* genome (http://www.vectorbase.org/Tools/BLAST/).
[2)] The locations of probes in *A. stephensi* were obtained from the standard photomap of *A. stephensi*.
NA indicates not available.

Appendix 3.3: The chromosomal locations of 127 probes on the polytene chromosomes of *A. funestis* and the coordinates in *A. gambiae* for analyzing the distribution of markers.

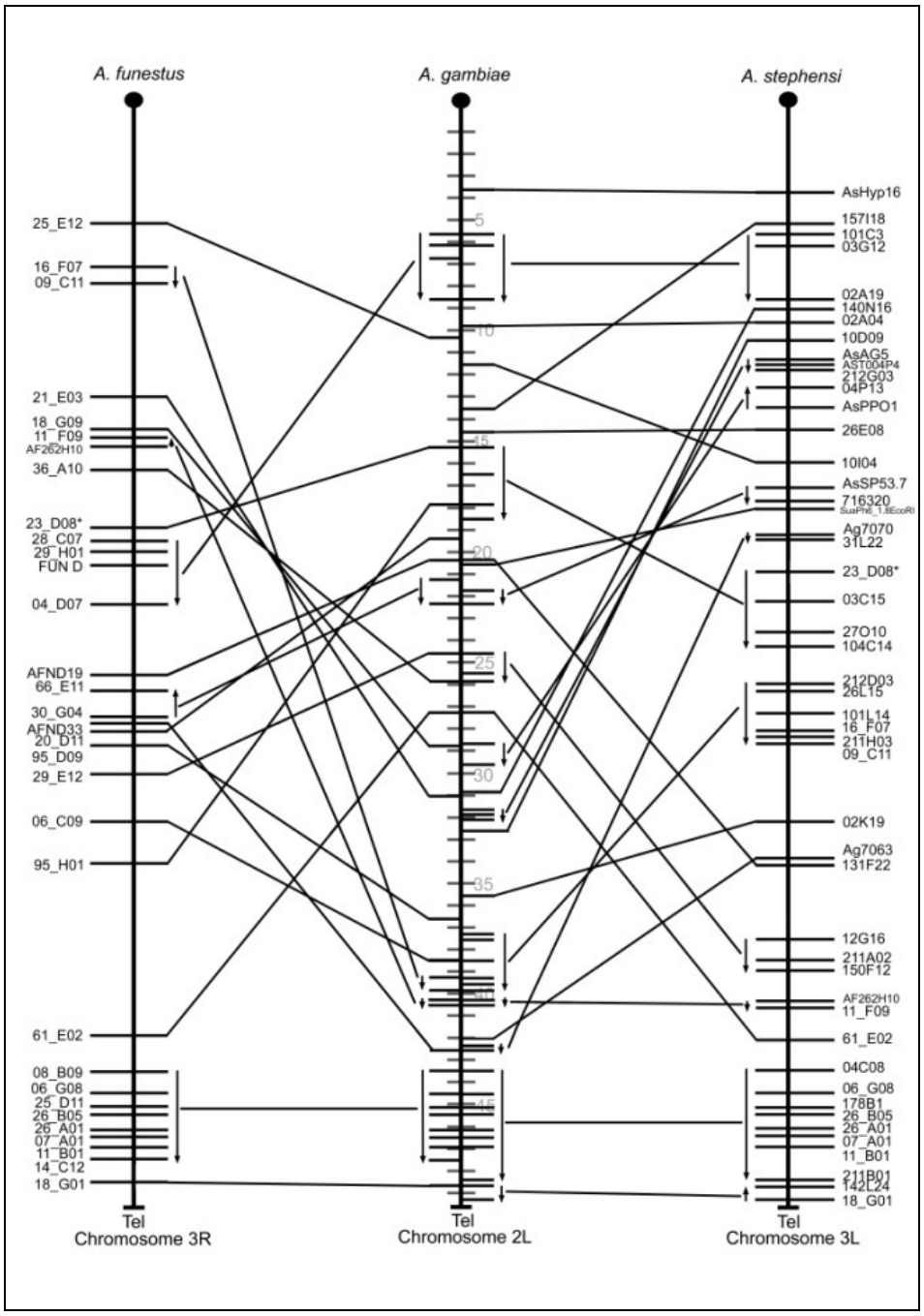
# in *A. gambiae* X	**Marker**	**Position on *A. gambiae* chromosome[1]**	**Pisition on *A. funestus* chromosme[2]**
1	06_E11	84,525	145
2	FUN Q	1,451,178	187
3	14_B05	3,000,117	119
4	18_G03	12,117,478	24
5	17_E02	14,442,142	42
6	CYP9K1	15,240,763	180
7	61_C09	15,336,624	95
8	FUN E	15,736,855	190
9	CYP4G21	16,619,141	164
10	98_D11	16,737,992	52
11	26_G11	18,339,741	28
12	18_D02	19,936,008	138
13	18_B08	22,937,938	225
Total length of X		**24,393,108**	**240**
# in *A. gambiae* 2R	**Marker**	**Position on *A. gambiae* chromosome[1]**	**Pisition on *A. funestus* chromosme[2]**
1	21_F03	1,608,204	16
2	01_H04	2,802,248	0
3	21_F12	4,050,701	258
4	06_E01	4,739,226	264
5	36_E01	6,827,869	49
6	12_G10	8,925,685	433
7	AFND5	10,545,856	426
8	23_B02	11,975,829	625
9	36_B02	13,133,425	75
10	11_D03	14,012,136	66
11	04_D06	16,812,718	112
12	61_D05	17,648,143	118

13	08_E06	19,444,433	499
14	06_D06	19,946,249	504
15	15_G03	21,027,264	191
16	13_F11	21,573,636	203
17	15_F10	21,825,430	206
18	17_G08	21,835,973	208
19	06_B01	23,835,495	383
20	AFND32	27,025,145	473
21	25_H11	29,839,388	97
22	03_D09	31,230,560	580
23	11_D07	33,903,940	287
24	27_E05	37,194,861	644
25	12_G11	37,466,896	660
26	12_H09	38,738,669	672
27	66_E07	39,452,027	527
28	AF261B04	39,751,219	530
29	29_F03	41,002,243	174
30	11_E07	41,338,854	351
31	13_A06	45,474,529	321
32	13_C03	47,284,629	302
33	18_D12	53,207,213	342
34	11_B04	57,542,706	725
Total length of 2R		**61,545,105**	**745**
# in *A. gambiae* 2L	**Marker**	**Position on *A. gambiae* chromosome[1]**	**Pisition on *A. funestus* chromosme[2]**
1	28_C07	5,771,985	163
2	29_H01	6,114,439	165
3	FUN D	6,734,718	172
4	04_D07	8,668,402	183
5	25_E12	10,241,196	42
6	95_H01	17,920,080	285
7	20_D11	19,277,265	230
8	AFND19	20,396,064	206
9	30_G04	21,199,038	218
10	66_E11	22,321,985	214
11	29_E12	24,671,856	245
12	36_A10	26,055,839	119
13	61_E02	27,158,330	332
14	18_G09	28,771,405	108
15	21_E03	31,048,592	96
16	95_D09	36,653,143	233
17	06_C09	38,567,686	264
18	16_F07	39,215,802	53
19	09_C11	39,995,907	60
20	AF262H10	40,442,200	113
21	11_F09	40,536,371	111
22	08_B09	43,603,779	338
23	06_G08	44,638,197	346
24	23_E09	45,981,005	362
25	07_A01	46,337,490	375
26	11_B01	46,995,122	384
27	18_G01	48,608,644	405
Total length of 2L		**49,364,325**	**414**
# in *A. gambiae* 3R	**Marker**	**Position on *A. gambiae* chromosome[1]**	**Pisition on *A. funestus* chromosme[2]**
1	19_F10	838,035	98
2	11_H04	1,483,634	83
3	01_C07	3,991,227	118
4	06_F07	4,869,662	135

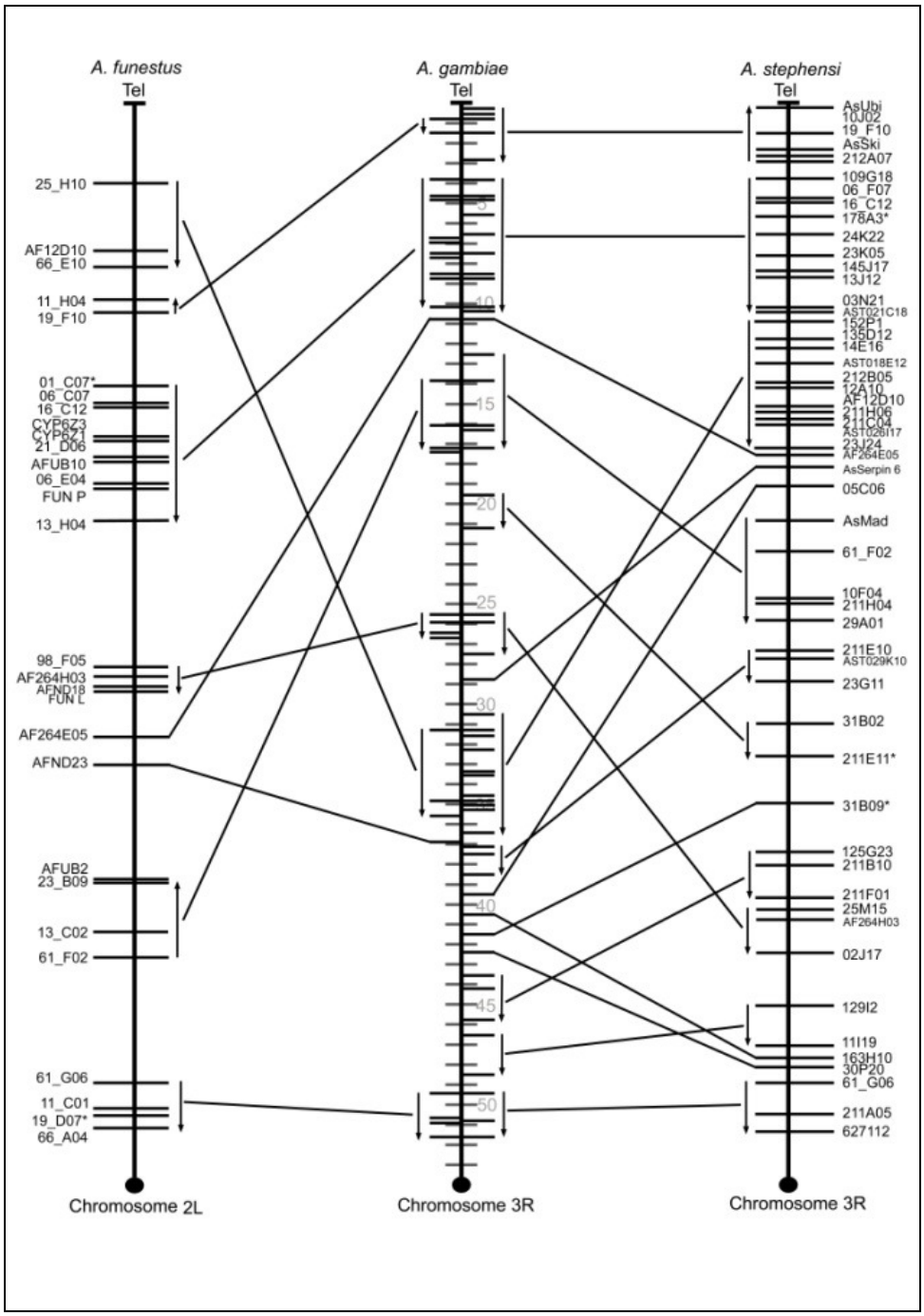
5	16_C12	4,880,891	142
6	CYP6Z3	6,971,669	155
7	CYP6Z1	6,976,539	160
8	21_D06	7,543,137	163
9	AFUB10	7,587,587	165
10	06_E04	8,630,719	172
11	Fun P	8,793,109	175
12	13_H04	10,088,721	185
13	AF264E05	10,774,634	276
14	61_F02	13,939,170	354
15	13_C02	16,069,773	338
16	23_B09	17,144,470	332
17	AFUB2	17,154,074	330
18	98_F05	25,578,233	241
19	AF264H03	25,983,063	252
20	AFND18	26,560,353	255
21	FUN L	26,816,555	255
22	25_H10	31,268,756	31
23	AF12D10	34,898,584	53
24	66_E10	35,605,568	62
25	AFND23	36,927,785	286
26	61_G06	49,370,093	423
27	11_C01	50,715,566	435
28	19_D07	50,847,429	437
29	66_A04	51,662,576	444
Total length of 3R		**53,200,684**	**452**
# in *A. gambiae* 3L		**Position on *A. gambiae* chromosome[1]**	**Pisition on *A. funestus* chromosme[2]**
1	15_B11	2,444,111	19
2	01_F07	2,779,188	29
3	07_G04	3,752,432	32
4	25_E01	4,015,796	48
5	27_B04	5,056,173	52
6	29_D12	6,421,493	70
7	03_G10	8,102,799	302
8	12_B09	8,934,066	312
9	15_F08	10,412,154	231
10	12_F01	11,726,542	243
11	36_A01	18,112,740	249
12	21_F09	20,578,594	156
13	29_F07	23,869,077	185
14	04_D01	26,294,061	390
15	27_A08	28,373,084	368
16	AF13G04	30,431,704	346
17	AF263D12	32,057,098	335
18	25_H01	33,271,412	97
19	04_E02	33,872,056	114
20	21_G01	33,925,153	116
21	20_G09	37,631,554	296
22	19_H06	37,960,351	325
23	05_A10	40,732,639	189
24	36_C05	41,834,603	363
Total length of 3L		**41,963,435**	**390**

[1)] The coordinates of *A. gambiae* in Mb. The coordinates for *A. gambiae* cDNAs and BAC clones were available at VectorBase.org (*440*). For the coordinates of other clones were the blast hits using the sequences blastn against *A. gambiae* genome (http://www.vectorbase.org/Tools/BLAST/).

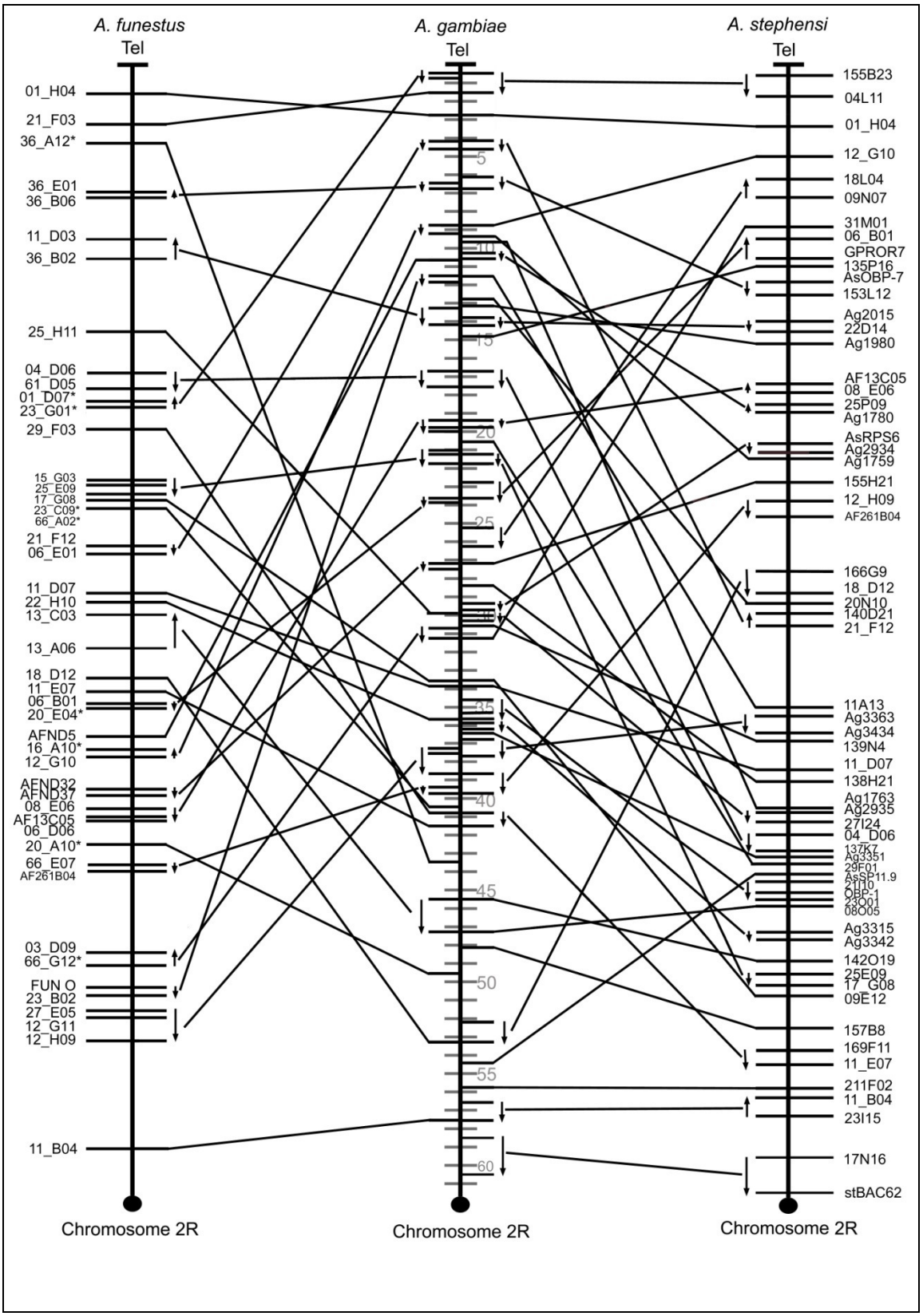
[2)] The locations of probes in *A. funestus* were obtained from the standard photomap of *A. funestus*.



Appendix 3.4: The gene order omparison between *A. gambiae* and *A. funestus*, between *A. gambiae* and *A. stephensi* on 2L chromosomal arm. The *A. gambiae* 2L arm correspondences to the 3R in *A. funestus* and 3L of *A. stephensi*.



Appendix 3.5: The gene order omparison between *A. gambiae* and *A. funestus*, between *A. gambiae* and *A. stephensi* on 3R chromosomal arm. The *A. gambiae* 3R arm correspondences to the 2L in *A. funestus* and 3R of *A. stephensi*.



Appendix 3.6: The gene order omparison between *A. gambiae* and *A. funestus*, between *A. gambiae* and *A. stephensi* on 2R chromosomal arm.

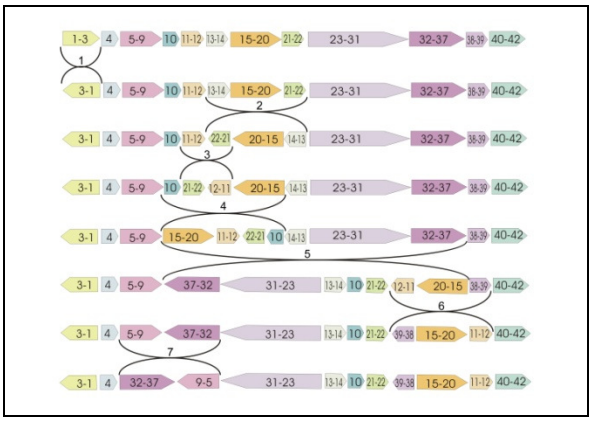
Appendix 3.7: The lengths of conserved synteny blocks between *A. gambiae* and *A. stephensi*.

X in *A. gambiae*	Synteny block	Start	End	Length of synteny block	# of markers
1	25E24-126O17	459,415	1,512,610	1,053,195	2
2	27D16-150E4	2,898,136	3,033,400	135,264	2
3	24J01-31G14	3,632,054	4,029,848	397,794	2
4	24G07-Ag0803	14,548,696	14,744,472	195,776	2
Total					8
Total chromosome length of X = 24,400,000bp					
2R in *A. gambiae*	**Synteny block**	**Start**	**End**	**Length of synteny block**	**# of markers**
1	155B23-04L11	528,525	1,708,302	1,179,777	2
2	21_F12-140D21	4,050,701	4,755,258	704,557	2
3	AsOBP-7-153L12	6,152,011	6,894,483	742,472	2
4	Ag1780-25P09	10,216,796	10,650,334	433,538	2
5	Ag2015-22D14	13,936,706	14,034,752	98,046	2
6	04_D06-137K7	16,812,718	17,688,592	875,874	2
7	08_E06-AF13C05	19,444,433	19,883,271	438,838	2
8	25E09-17_G08	21,425,189	21,836,716	411,527	2
9	GPROR7-06_B01	22,849,252	23,836,568	987,316	2
10	09N07-18L04	25,146,360	26,313,071	1,166,711	2
11	AsRPS6-Ag2934	29,609,284	29,836,999	227,715	2
12	Ag2935-27I24	29,839,387	30,271,431	432,044	2
13	21I10-23O01	34,742,360	35,790,690	1,048,330	3
14	Ag3315-Ag3342	35,837,690	36,311,720	474,030	2
15	Ag3363-Ag3434	36,850,789	37,711,020	860,231	2
16	12_H09-AF261B04	38,738,669	39,751,269	1,012,600	2
17	169F11-11_E07	40,944,146	41,360,919	416,773	2
18	166G9-18_D12	52,013,039	53,209,347	1,196,308	2
19	23I15-11_B04	56,667,003	57,543,865	876,862	2
20	17N16-stBAC62	58,537,340	60,253,235	1,715,895	2
Total					41
Total chromosome length of 2R = 61,500,000bp					
2L in *A. gambiae*	**Synteny block**	**Start**	**End**	**Length of synteny block**	**# of markers**
1	101C3-02A19	5,708,924	8,791,152	3,082,228	3
2	23_D08-104C14	15,211,390	18,616,664	3,405,274	4
3	AsSP53.7-716320	21,865,434	22,327,760	462,326	2
4	12G16-150F12	24,626,085	26,087,837	1,461,752	3
5	AsPPO1-04P13	28,702,474	29,659,442	956,968	2
6	AsAG5-212G03	31,693,742	32,055,857	362,115	3
7	212D03-09_C11	37,118,475	39,997,107	2,878,632	6
8	AF262H10-11_F09	40,442,200	40,538,248	96,048	2
9	04_C08-211B01	43,540,182	48,341,468	4,801,286	8
10	18_G01-142L24	48,608,644	49,333,455	724,811	2
11	Ag7070-31L22	42,178,250	42,309,500	131,250	2
Total					37
Total chromosome length of 2L = 49,400,000					
3R in *A. gambiae*	**Synteny block**	**Start**	**End**	**Length of synteny block**	**# of markers**
1	212A07-AsUbi	217,370	2,921,050	2,703,680	5
2	109G18-AST021C18	3,963,987	10,129,187	6,165,200	10
3	AsMad-29A01	12,545,417	17,221,188	4,675,771	5
4	31B02-211E11	19,607,644	21,046,135	1,438,491	2
5	25M15-02J17	25,546,570	27,597,800	2,051,230	3
6	152P1-23J24	30,494,707	36,591,879	6,097,172	11
7	211E10-23G11	37,001,849	38,718,166	1,716,317	3

8	125G23-211F01	43,532,477	45,806,458	2,273,981	3
9	129I2-11I19	46,515,634	48,595,676	2,080,042	2
10	61_G06-627112	49,370,093	51,663,611	2,293,518	3
Total					47
Total chromosome length of 3R = 53,200,000					
3L in *A. gambiae*	**Synteny block**	**Start**	**End**	**Length of synteny block**	**# of markers**
1	212D01-07_G04	2,246,808	3,757,036	1,510,228	3
2	151M24-126G21	4,982,245	8,989,815	4,007,570	5
3	AST012A11-12_F01	11,244,143	11,727,596	483,453	2
4	105F8-02G07	12,527,507	13,907,560	1,380,053	2
5	31H07-AST041D2	14,419,173	18,526,417	4,107,244	6
6	27P23-211A03	20,554,497	20,689,982	135,485	2
7	155I2-11G16	23,877,430	32,701,072	8,823,642	9
8	669234-01K17	33,271,412	35,582,996	2,311,584	6
9	132D12-104F13	37,599,013	38,080,288	481,275	2
10	28J05-126M2	40,982,673	41,892,789	910,116	3
Total					40
Total chromosome length of 3L = 42,000,000					
Total number of markers = 173					

The genomic locations of all the markers were acquired from Table S1.

Appendix 3.8: The scenario of chromosomal transformation from *A. gambiae* 2L to *A. stephensi* 3L chromosome.



Appendix 3.9: The scenario of chromosomal transformation from *A. gambiae* 3L to *A. stephensi* 2L chromosome.

Appendix 3.10: The scenario of chromosomal transformation from *A. gambiae* 2R to *A. stephensi* 2R chromosome.

Appendix 3.11: The density of fixed inversions and molecular features on five chromosomal arms of *A. gambiae*

Molecular feature	**X**	**2R**	**2L**	**3R**	**3L**
Density of fixed inversion (Number per Mb)	6.148	4.715	3.239	2.068	3.384
M/SARs(Count per Mb)	59	55	74	82	90
Retroelements (Number per Mb)	135.20	80.60	95.92	86.64	96.81
DNA transposons(Number per Mb)	115.08	48.11	64.10	58.05	69.74
Microsatellite(Number per Mb)	136.41	70.26	68.70	62.16	53.23
Minisatellite(Number per Mb)	197.96	114.98	120.75	137.81	122.81
Satellite(Number per Mb)	10.20	6.73	5.31	6.34	5.90
Inverted repeats(Number per Mb)	236.84	52.98	65.43	54.87	64.12
GC repeats(Number per Mb)	22.38	12.20	13.18	13.55	11.52
AT repeats(Number per Mb)	22.91	10.11	10.40	8.20	10.21
SD(Number per Mb)	6.52	14.44	7.27	10.51	14.24
Genes(Number per Mb)	44.84	58.24	61.82	48.48	50.43

Appendix 3.12: The molecular features in breakpoint regions and synteny blocks between *A. gambiae* and *A. stephensi*

X	Breakpoint region	Start	End	Length	Genes	Inverted repeats	Microsatellite	Minisatellite	Satellite
1	150E4-17I02	3033400	3147527	114127	6	8	20	15	0
2	21F12-31A03	14190611	14227297	36686	0	2	5	2	1
Total				**150813**	**6**	**10**	**25**	**17**	**1**
X	**Synteny block**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	25E24-126O17	459415	1512610	1053195	54	328	298	283	8
2	27D16-150E4	2898136	3033400	135264	17	15	10	13	4
3	24J01-31G14	3632054	4029848	397794	2	199	128	65	5
4	24G07-Ag0803	14548696	14744472	195776	7	64	45	30	2
Total				**1782029**	**80**	**606**	**481**	**391**	**19**
2R	**Breakpoint region**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	Ag1759-Ag1763	9483291	9523856	40565	4	2	30	83	1
2	20N10-Ag1980	12976648	13087584	110936	15	20	36	30	0
3	Ag2934-Ag2935	29836999	29839387	2388	2	1	0	0	0
4	09E12-11_D07	33693463	33903940	210477	21	13	33	25	2
5	23O01-Ag3315	35790690	35837690	47000	2	0	48	77	7
6	Ag3342-Ag3351	36311720	36478446	166726	9	9	51	80	3
Total				**578092**					
2R	**Synteny block**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	155B23-04L11	528525	1708302	1179777	100	26	53	101	1
2	21_F12-140D21	4050701	4755258	704557	46	49	86	52	1
3	AsOBP-7-153L12	6152011	6894483	742472	62	45	63	84	1
4	Ag1780-25P09	10216796	10650334	433538	26	29	37	46	1
5	Ag2015-22D14	13936706	14034752	98046	8	3	54	61	0
6	04_D06-137K7	16812718	17688592	875874	51	35	59	64	5
7	08_E06-AF13C05	19444433	19883271	438838	10	23	46	37	4
8	25E09-17_G08	21425189	21836716	411527	29	24	27	31	1
9	GPROR7-06_B01	22849252	23836568	987316	44	16	81	68	1
10	09N07-18L04	25146360	26313071	1166711	72	60	115	113	1
11	AsRPS6-Ag2934	29609284	29836999	227715	16	5	24	44	1
12	Ag2935-27I24	29839387	30271431	432044	37	43	71	43	12
13	21I10-23O01	34742360	35790690	1048330	41	39	64	95	0
14	Ag3315-Ag3342	35837690	36311720	474030	28	15	39	58	2
15	Ag3363-Ag3434	36850789	37711020	860231	70	8	48	102	0
16	12_H09-AF261B04	38738669	39751269	1012600	68	19	62	72	0
17	169F11-11_E07	40944146	41360919	416773	20	9	15	51	0
18	166G9-18_D12	52013039	53209347	1196308	28	45	55	107	2
19	23I15-11_B04	56667003	57543865	876862	97	41	48	81	0
20	17N16-stBAC62	58537340	60253235	1715895	44	309	121	306	17
Total				**15299444**	**897**	**843**	**1168**	**1616**	**50**
2L	**Breakpoint region**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	26E08-23_D08	14728697	15211390	482693	34	29	18	65	2
2	131F22-SuaPh6	20459325	20535740	76415	7	5	3	11	0
3	140N16-AsAG5	31090638	31693742	603104	22	23	46	83	1
4	212G03-10D09	32055857	32640047	584190	32	42	30	71	2
5	09_C11-AF262H10	39997107	40442200	445093	57	30	29	50	1

6	Ag7063-Ag7070	42127751	42178250	50499	7	4	5	3	0
7	211B01-18_G01	48341468	48608644	267176	53	4	13	25	0
8	142L24-end	49333455	49364325	30870	5	2	0	20	2
Total				**2540040**	**217**	**139**	**144**	**328**	**8**
2L	**Synteny block**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	101C3-02A19	5708924	8791152	3082228	144	141	134	330	7
2	23_D08-104C14	15211390	18616664	3405274	265	216	213	462	16
3	AsSP53.7-716320	21865434	22327760	462326	17	24	44	25	0
4	12G16-150F12	24626085	26087837	1461752	72	111	175	178	7
5	AsPPO1-04P13	28702474	29659442	956968	69	79	81	124	7
6	AsAG5-212G03	31693742	32055857	362115	22	21	14	68	4
7	212D03-09_C11	37118475	39997107	2878632	210	232	248	277	6
8	AF262H10-11_F09	40442200	40538248	96048	22	6	9	24	0
9	04_C08-211B01	43540182	48341468	4801286	482	139	354	368	8
10	18_G01-142L24	48608644	49333455	724811	84	15	44	40	3
11	Ag7070-31L22	42178250	42309500	131250	16	6	10	11	0
Total				**18362690**	**1403**	**990**	**1326**	**1907**	**58**
3R	**Breakpoint region**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	0-212A07	0	217370	217370	17	21	11	19	1
2	AST021C18-AF264E05	10129187	10774634	645447	59	28	48	50	1
3	23J24-211E10	36591879	37001849	409970	28	21	42	76	11
4	23G11-05C06	38718166	39502006	783840	22	75	42	98	1
5	05C06-163H10	39614515	40492051	877536	6	60	40	143	4
6	163H10-31B09	40582084	41547299	965215	5	101	72	136	4
7	31B09-30P20	41646263	42487765	841502	41	45	43	80	0
8	30P20-125G23	42579439	43532477	953038	31	79	58	116	2
9	211F01-129I2	45806458	46515634	709176	14	35	32	86	1
10	11I19-61_G06	48595676	49370093	774417	74	49	43	137	1
Total				**7177511**	**297**	**514**	**431**	**941**	**26**
3R	**Synteny block**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	212A07-AsUbi	217370	2921050	2703680	181	45	167	161	2
2	109G18-AST021C18	3963987	10129187	6165200	384	127	360	409	17
3	AsMad-29A01	12545417	17221188	4675771	215	320	295	842	25
4	31B02-211E11	19607644	21046135	1438491	68	63	78	203	4
5	25M15-02J17	25546570	27597800	2051230	58	99	180	178	1
6	152P1-23J24	30494707	36591879	6097172	309	330	397	1043	80
7	211E10-23G11	37001849	38718166	1716317	80	129	107	295	14
8	125G23-211F01	43532477	45806458	2273981	147	116	124	233	4
9	129I2-11I19	46515634	48595676	2080042	112	103	95	217	8
10	61_G06-627112	49370093	51663611	2293518	111	131	101	258	11
Total				**31495402**	**1665**	**1463**	**1904**	**3839**	**166**
3L	**Breakpoint region**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	07_G04-139K20	3757036	4607484	850448	40	123	36	190	3
2	139K20-151M24	4702647	4982245	279598	4	27	9	43	2
3	126G21-180K21	8989815	9407395	417580	21	12	21	48	1
4	12_F01-105F8	11727596	12527507	799911	50	44	34	114	9
5	02G07-31H07	13907560	14419173	511613	14	25	16	67	1
6	11G16-669234	32701072	33271412	570340	52	38	30	112	3

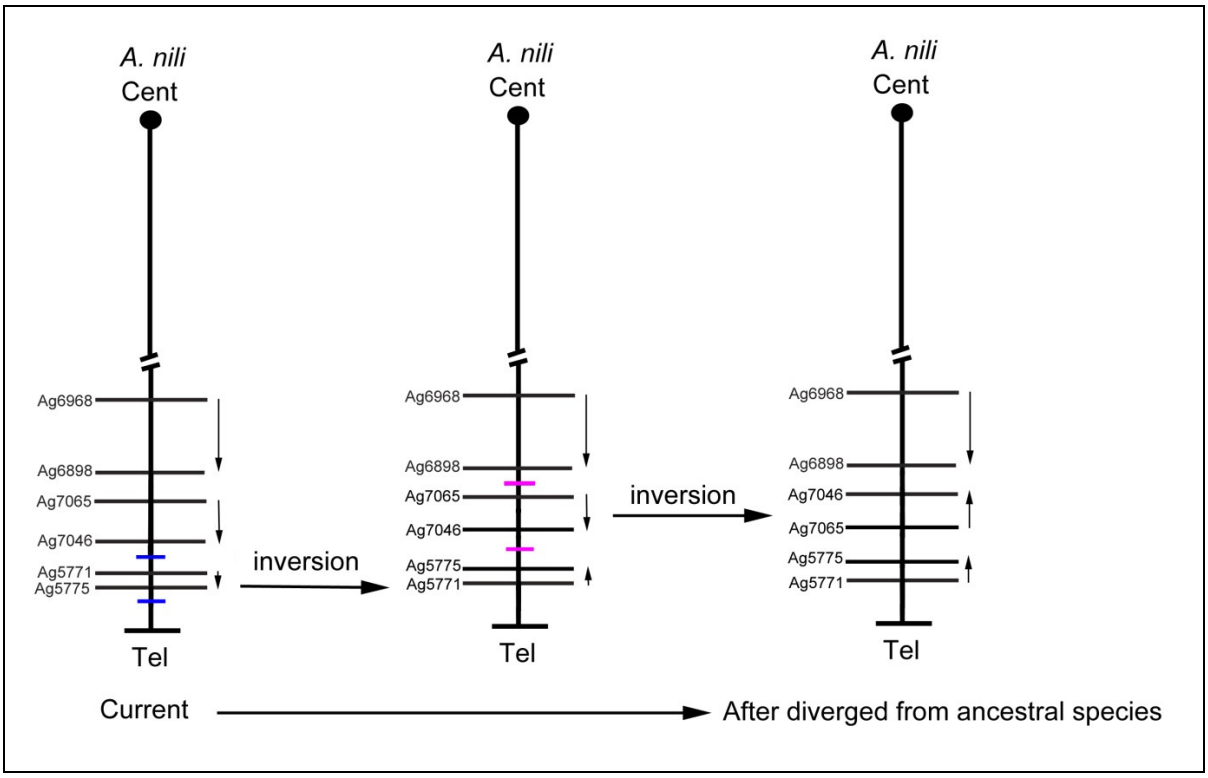
Total				**3429490**	**181**	**269**	**146**	**574**	**19**
3L	**Synteny block**	**Start**	**End**	**Length**	**Genes**	**Inverted repeats**	**Microsatellite**	**Minisatellite**	**Satellite**
1	212D01-07_G04	2246808	3757036	1510228	82	109	65	195	6
2	151M24-126G21	4982245	8989815	4007570	234	299	193	452	23
3	AST012A11-12_F01	11244143	11727596	483453	20	18	26	51	1
4	105F8-02G07	12527507	13907560	1380053	77	66	78	230	5
5	31H07-AST041D2	14419173	18526417	4107244	210	142	313	387	10
6	27P23-211A03	20554497	20689982	135485	15	11	8	20	1
7	155I2-11G16	23877430	32701072	8823642	333	471	523	1401	69
8	669234-01K17	33271412	35582996	2311584	171	101	158	241	10
9	132D12-104F13	37599013	38080288	481275	58	2	23	70	5
10	28J05-126M2	40982673	41892789	910116	70	34	66	84	3
Total				**24150650**	**1270**	**1253**	**1453**	**3131**	**133**

Appendix 3.13: The molecular features in breakpoint regions and synteny blocks between *A. gambiae* and *A. stphensi*.

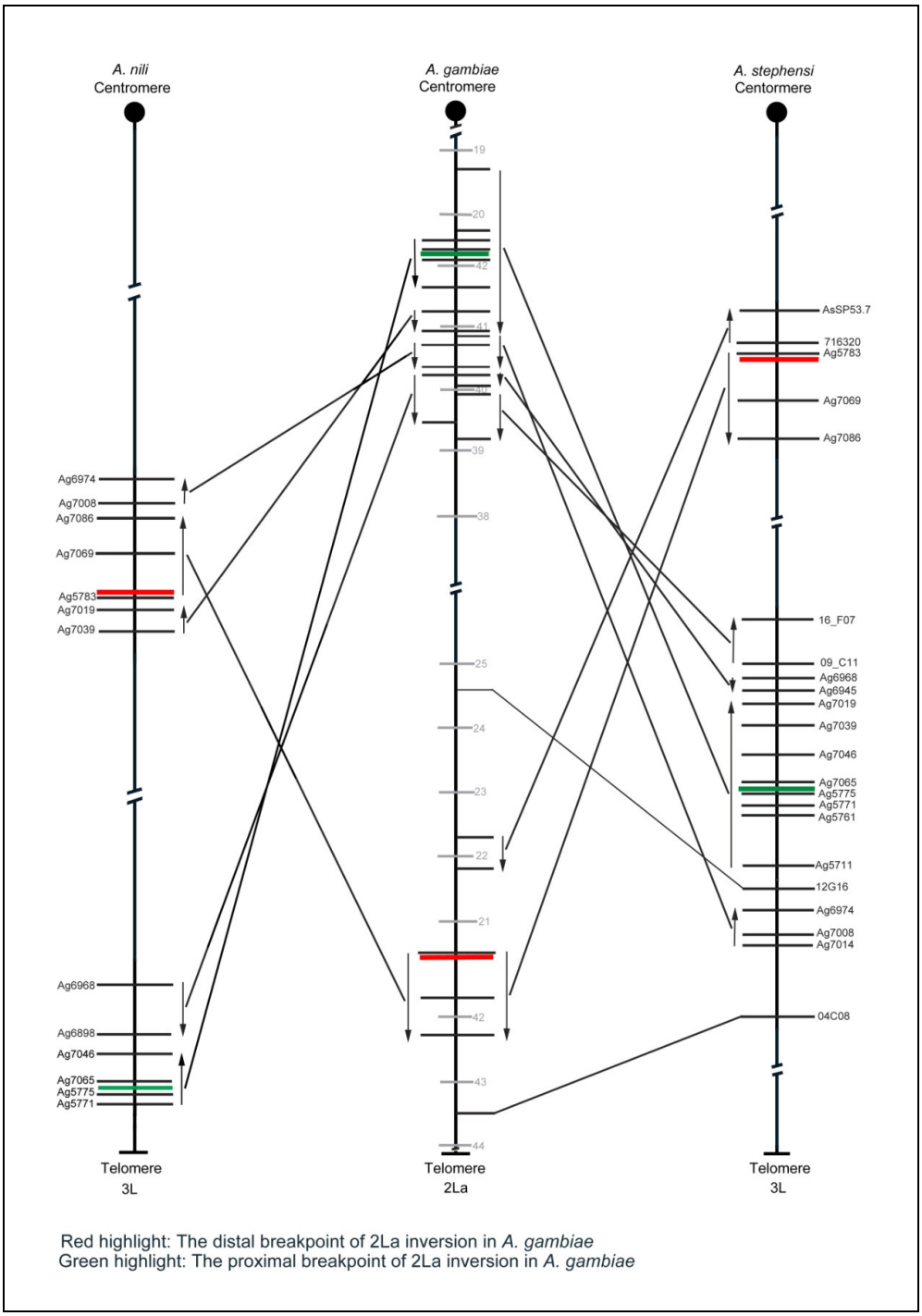
X	**Breakpoint region**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**M/SAR**	**Low complexity**
1	150E4-17I02	114,127	4	7	11	2	3	5	22
2	21F12-31A03	36,686	4	2	6	0	1	1	6
Total		**150813**	**8**	**9**	**17**	**2**	**4**	**6**	**28**
X	**Synteny block**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**M/SAR**	**Low complexity**
1	25E24-126O17	1,053,195	54	44	98	0	38	25	238
2	27D16-150E4	135,264	5	7	12	3	1	5	23
3	24J01-31G14	397,794	15	23	38	1	5	12	79
4	24G07-Ag0803	195,776	17	12	29	1	4	29	43
Total		**1782029**	**91**	**86**	**177**	**5**	**48**	**71**	**383**
2R	**Breakpoint region**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**M/SAR**	**Low complexity**
1	Ag1759-Ag1763	40565	1	2	3	0	1	3	6
2	20N10-Ag1980	110936	1	2	3	0	1	0	7
3	Ag2934-Ag2935	2388	0	0	0	0	0.2	0	0
4	09E12-11_D07	210477	17	7	24	0	4	8	24
5	23O01-Ag3315	47000	1	3	4	0	1	5	7
6	Ag3342-Ag3351	166726	3	7	10	2	1	9	16
Total		**578092**	**23**	**21**	**44**	**2**	**8.2**	**25**	**60**
2R	**Synteny block**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**M/SAR**	**Low complexity**
1	155B23-04L11	1179777	34	3	37	0	2	9	84
2	21_F12-140D21	704557	16	18	34	1	6	9	71
3	AsOBP-7-153L12	742472	25	11	26	1	3	8	77
4	Ag1780-25P09	433538	18	6	24	0	2	8	49
5	Ag2015-22D14	98046	3	2	5	0	0	3	7
6	04_D06-137K7	875874	42	18	60	0	8	40	95
7	08_E06-AF13C05	438838	25	13	38	4	1	16	41
8	25E09-17_G08	411527	35	17	52	0	5	16	36
9	GPROR7-06_B01	987316	23	25	48	1	7	28	103

10	09N07-18L04	1166711	57	43	100	2	13	38	122
11	AsRPS6-Ag2934	227715	11	5	16	0	5	7	19
12	Ag2935-27I24	432044	18	12	30	0	5	14	50
13	21I10-23O01	1048330	82	40	122	2	13	65	128
14	Ag3315-Ag3342	474030	37	17	54	1	6	22	64
15	Ag3363-Ag3434	860231	53	27	80	2	4	41	86
16	12_H09-AF261B04	1012600	64	34	98	1	7	54	113
17	169F11-11_E07	416773	40	19	59	1	1	32	58
18	166G9-18_D12	1196308	124	61	185	1	13	100	150
19	23I15-11_B04	876862	97	74	171	4	9	70	93
20	17N16-stBAC62	1715895	496	353	849	72	83	333	222
Total		**15299444**	**1300**	**798**	**2088**	**93**	**193**	**913**	**1668**
2L	**Breakpoint region**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**M/SAR**	**Low complexity**
1	26E08-23_D08	482693	55	38	93	2	2	53	70
2	131F22-SuaPh6	76415	9	9	18	0	2	2	113
3	140N16-AsAG5	603104	58	34	82	1	6	57	86
4	212G03-10D09	584190	52	41	93	5	6	49	81
5	09_C11-AF262H10	445093	25	22	47	1	2	24	44
6	Ag7063-Ag7070	50499	8	2	10	0	0	0	6
7	211B01-18_G01	267176	12	4	16	0	2	1	11
8	142L24-end	30870	2	1	3	0	1	5	5
Total		**2540040**	**221**	**151**	**362**	**9**	**21**	**191**	**416**
2L	**Synteny block**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**MSAR**	**Low complexity**
1	101C3-02A19	3082228	446	252	698	20	38	388	404
2	23_D08-104C14	3405274	269	181	450	15	28	243	411
3	AsSP53.7-716320	462326	16	7	23	0	2	15	39
4	12G16-150F12	1461752	78	60	138	6	10	66	179
5	AsPPO1-04P13	956968	63	48	111	5	6	100	124
6	AsAG5-212G03	362115	20	14	34	1	2	20	46
7	212D03-09_C11	2878632	215	111	326	10	18	151	360
8	AF262H10-11_F09	96048	7	4	11	1	1	6	7
9	04_C08-211B01	4801286	180	82	262	8	27	59	352
10	18_G01-142L24	724811	30	11	41	0	6	18	46
11	Ag7070-31L22	131250	16	0	16	4	0	3	18
Total		**18362690**	**1340**	**770**	**2110**	**70**	**138**	**1069**	**1986**
3R	**Breakpoint region**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**M/SAR**	**Low complexity**
1	0-212A07	217370	15	31	46	5	4	11	17
2	AST021C18-AF264E05	645447	20	11	41	3	10	19	72
3	23J24-211E10	409970	27	26	53	1	3	35	62
4	23G11-05C06	783840	148	64	212	5	8	100	112
5	05C06-163H10	877536	192	93	285	6	11	99	114
6	163H10-31B09	965215	178	78	256	9	23.5	131	121
7	31B09-30P20	841502	43	41	84	3	8	53	111
8	30P20-125G23	953038	148	92	240	5	2	98	133

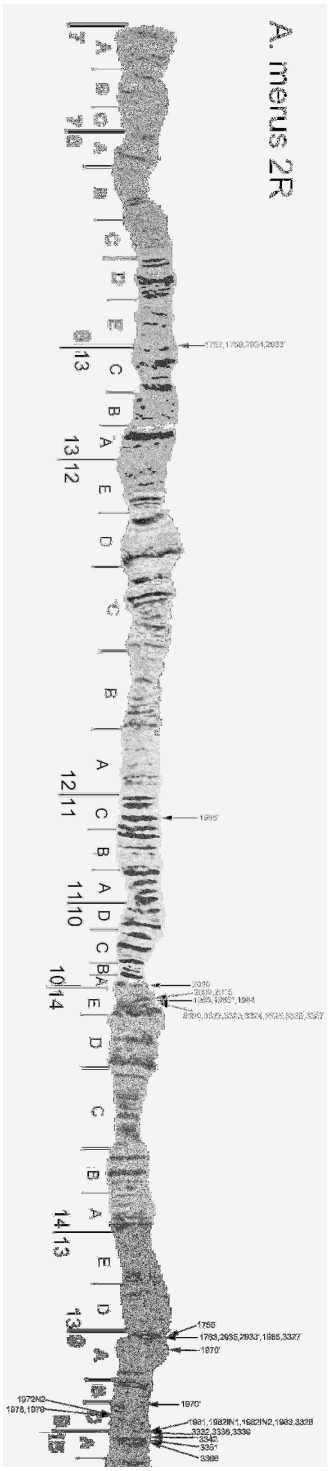
9	211F01-129I2	709176	117	45	162	3	4	112	102
10	11I19-61_G06	774417	141	75	216	5	21	92	111
Total		**7177511**	**1029**	**556**	**1595**	**45**	**94.5**	**750**	**955**
3R	**Synteny block**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**M/SAR**	**Low complexity**
1	212A07-AsUbi	2703680	69	23	92	2	14	33	175
2	109G18-AST021C18	6165200	166	64	230	3	33	128	461
3	AsMad-29A01	4675771	411	343	754	14	28	594	690
4	31B02-211E11	1438491	118	107	215	13	9	162	204
5	25M15-02J17	2051230	149	92	241	7	9	155	248
6	152P1-23J24	6097172	301	255	556	13	44	323	638
7	211E10-23G11	1,716,317	201	132	333	6	20	183	260
8	125G23-211F01	2,273,981	222	151	373	5	14	234	300
9	129I2-11I19	2,080,042	285	177	462	13	18	240	223
10	61_G06-627112	2,293,518	336	215	551	15	16	269	288
Total		**31495402**	**2258**	**1559**	**3807**	**91**	**205**	**2321**	**3487**
3L	**Breakpoint region**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified**	**AT repeats**	**M/SAR**	**Low complexity**
1	07_G04-139K20	850,448	235	159	394	17	36	99	103
2	139K20-151M24	279,598	81	67	148	2	4	42	38
3	126G21-180K21	417,580	81	33	114	0	6	62	62
4	12_F01-105F8	799,911	76	47	123	3	7	78	167
5	02G07-31H07	511,613	26	31	57	2	5	56	72
6	11G16-669234	570,340	49	38	87	0	3	40	64
Total		**3429490**	**548**	**375**	**923**	**24**	**61**	**377**	**506**
3L	**Synteny block**	**Length**	**Retroelements**	**DNA transposons**	**TEs**	**Unclassified q**	**AT repeats**	**M/SAR**	**Low complexity**
1	212D01-07_G04	1,510,228	220	148	368	14	20	233	208
2	151M24-126G21	4,007,570	661	346	1007	34	56	571	539
3	AST012A11-12_F01	483,453	46	41	87	6	0	44	59
4	105F8-02G07	1,380,053	144	98	242	6	14	162	228
5	31H07-AST041D2	4,107,244	222	164	386	7	21	212	463
6	27P23-211A03	135,485	9	11	20	0	1	14	19
7	155I2-11G16	8,823,642	521	426	947	30	56	890	1167
8	669234-01K17	2,311,584	169	112	281	2	12	163	284
9	132D12-104F13	481,275	19	8	27	0	7	7	45
10	28J05-126M2	910,116	32	40	72	6	12	34	64
Total		**24150650**	**2043**	**1394**	**3437**	**105**	**199**	**2330**	**3076**



Appendix 4.1: Restore phylogenetic history in *A. nili*



Appendix 4.2: The gene orders comparison between *A. gambiae* and *A. nili* after diverged from ancestral species, *A. gambiae* and *A. stephensi* near the breakpoints of 2La. Arrow indicates the synteny block which contains at least two continuous markers. The locations of probes near the breakpoints in A. gambiae were drawn using the coordinates of A. gambiae 2L+a, however for the rest chromosomal regions were shown not on scale.



Appendix 4.3: The physical map of *A. merus* 2R chromosome; 2R chromosome of *A. merus* was modified from (*1*) based on the locations of 2Ro and 2Rp breakpoints and the banding patterns were compared with the chromosome preparations of *A. merus*.

Appendix 4.4: The locations of in situ hybridization probes on *A. merus* 2R chromosome

#	clone	Accession	chromosome	division	
1	1757	AGAP001757	2R	8E	
2	1759	AGAP001759	2R	8E	
3	1763	AGAP001763	2R	9A(1)	
4	1765	AGAP001765	2R	9A(1)	
			2R	9A(2)	
5	1970	AGAP001970	2R	9C(1)	multiple
6	1972N2	AGAP001972	2R	9C(2)	
7	1978	AGAP001978	2R	9C(3)	
8	1979	AGAP001979	2R	9C(3)	
9	1981	AGAP001981	2R	9C(4)	
	1982N1	AGAP001982	2R	9C(4)	
10	1982N2	AGAP001982	2R	9C(4)	
11	1983	AGAP001983	2R	9C(4)	
12	1984	AGAP001984	2R	14E(3)	
			2R	14E(3)*	
	1985		2R	11C	
13		AGAP001985	2R	9A(1)	multiple
14	1986	AGAP001986	2R	14E(3)	
15	2009	AGAP002009	2R	14E(2)	
16	2015	AGAP002015	2R	14E(2)	
17	2030	AGAP002030	2R	14E(1)	
			2R	8E	
18	2933	AGAP002933	2R	9A(1)	multiple
19	2934	AGAP002934	2R	8E	
20	2935	AGAP002935	2R	9A(1)	
21	3320	AGAP003320	2R	14E(4)	
22	3322	AGAP003322	2R	14E(4)	
23	3323	AGAP003323	2R	14E(4)	
24	3324	AGAP003324	2R	14E(4)	
25	3325	AGAP003325	2R	14E(4)	
26	3326	AGAP003326	2R	14E(4)	
27	3327	AGAP003327	2R	14E(4)	
28	3328	AGAP003328	2R	9C(4)	
29	3332	AGAP003332	2R	15A(1)	
30	3336	AGAP003336	2R	15A(1)	
31	3339	AGAP003339	2R	15A(1)	
32	3342	AGAP003342	2R	15A(2)	
33	3351	AGAP003351	2R	15A(3)	
34	3366	AGAP003366	2R	15A(4)	

* Indicated as the major signal.