CHROMOSOME PHYLOGENIES OF MALARIA MOSQUITOES

© *I. V. Sharakhov*

Department of Entomology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA
e-mail: igor@vt.edu

Malaria mosquitoes often belong to complexes of sibling species, members of which are morphologically and genetically similar to each other. However, members within these complexes can vary significantly in their ecological adaptations and abilities to transmit the malaria parasite. The high degree of genetics similarity among sibling species makes the reconstruction of phylogenetic relationships within species complexes difficult. This paper reviews studies that infer the ancestral — descendant relationships among sibling species using molecular markers and chromosomal inversions. A methodology based on analyzing breakpoints of fixed overlapping inversions is shown to be useful for rooting phylogenies in complexes of sibling species, if the chromosomal arrangements in outgroup species are known. The construction of detailed phylogenies for malaria vectors will help to identify the association of evolutionary genomic changes with the origin of human blood choice and specific ecological adaptations.

Key words: *Anopheles*, species complexes, phylogenetic relationships, chromosomal inversions.

Taxonomic and population complexity is a common feature of malaria mosquitoes (Krzywinski, Besansky, 2003). The rich biodiversity of malaria mosquitoes has the direct epidemiological implications. Of the ~ 500 anopheline species, no more than 30 significantly contribute to the malaria transmission. Understanding the adaptation and speciation in malaria mosquitoes has not only a theoretical interest for evolutionary biology but also practical applications for vector control. Comparative genomic analyses of vector competence and other epidemiologically important traits will be informative if performed within a phylogenetic framework. Inferring ancestral and derived genomic features in anophelines is crucial for identifying the evolutionary changes associated with the origin and loss of human blood choice, ecological and behavioral adaptations, and association with human habitats. Traditionally, reconstructions of the anopheline phylogeny have been done using morphological and molecular markers (Krzywinski, Besansky, 2003). Available data support the monophyly of the six *Anopheles* subgenera, a sister-group relationship between subgenera *Nyssorhynchus* and *Kerteszia*, and a sister-group relationship between subgenera *Cellia* and *Anopheles.*

Complexes of sibling species are common among mosquitoes **(**Krzywinski, Besansky, 2003**)**. Members of such complexes are morphologically similar and partially reproductively isolated from each other. The African *Anopheles gambiae* complex belongs to series Pyretophorus of subgenus *Cellia* and consists of seven sibling malaria mosquito species that remarkably differ in geographic distribution, ecological adaptation, and host-seeking behavior. *An. gambiae* and *An. arabiensis* are the two major vectors of malaria in Africa. *An. merus* and *An. melas* breed in brackish water, and the habitat of *An. bwambae* is restricted to mineral water breeding sites. These three species are relatively minor malaria vectors with narrow geographic distribution (Coluzzi et al., 1979). *A. quadriannulatus* A and *An. quadriannulatus* B are fresh water breeders, zoophilic, and, although to some degree susceptible to *Plasmodium* infections, are not natural vectors of malaria (Coluzzi et al., 2002). Inferring evolutionary history of the *An. gambiae* complex could be important for finding specific genomic changes associated with the origin and loss of human blood choice, shifting in breeding site preference, and variations in vector competence.

A high degree of genetic similarity, caused by shared ancestral polymorphisms and extensive genetic introgression, confounds the ability to determine phylogenetic relationships and the direction of evolution in the *An. gambiae* complex using molecular markers (Besansky et al., 1994, 2003). For example, sequence similarity between the second chromosomes of *An. gambiae* and *An. arabiensis* has been largely attributed to genetic introgression (White et al., 2009). Rooting of molecular phylogenetic trees for the *An. gambiae* complex has been rarely attempted. Using the AT-rich region of the mitochondrial DNA of the most closely related outgroup species *An. christyi* helped to identify sister taxa but could not determine the most basal species in the complex (Caccone et al., 1996). Another attempt to root the tree with the sequence of X chromosomal gene *white* from *An. christyi* failed because of the great divergence of the intron size and sequence between the ingroup and outgroup genes (Besansky et al., 2003). Even the most recent genome-wide transcriptome-based phylogeny reconstruction of multiple Anophelinae species could not unambiguously resolve the relationships among *An. gambiae*, *An. arabiensis*, and *An. quadriannulatus* (Hittinger et al., 2010).

In addition to molecular markers, chromosomal inversions can be used to reconstruct species' phylogeny. The first chromosomal phylogeny was established for wild races of

Drosophila pseudoobscura by using polymorphic overlapping inversions (Sturtevant, Dobzhansky, 1936). The application of overlapping inversions for reconstructing species phylogenies is based on the assumptions that a) an inversion originates from a unique event in evolutionary history; b) inversion fixation occurs once in a lineage with no polymorphism extended across speciation; and c) no introgression of inversions occurs from one species to another (Krimbas, Powell, 1992). Indeed, cytogenetic studies on the *An. gambiae* complex supported the notion that fixed inversions do not introgress across species (Della Torre et al., 1997) and that they are monophyletic in origin (Sharakhov et al., 2006). Polymorphic inversions can potentially carry through speciation events. In this case, phylogenies based on different inversions would contradict each other. An alternative approach to inferring the phylogenetic relationships among species is to analyze the distribution of fixed overlapping inversions (Coluzzi et al., 1979; Stegnii, 1991; Coluzzi et al., 2002). This approach is based on the facts that species-specific inversions do not introgress (Della Torre et al., 1997) and that inversions are predominantly monophyletic, despite rare occurrences of a breakpoint reuse (Gonzalez et al., 2007).

Anopheles quadriannulatus species A and B have standard chromosomal arrangements, which are denoted by a «+» sign followed after the chromosome name: X+, 2R+, 2L+, 3L+, 3R+. These arrangements are considered «standard» because they occupy the central position in the complex by having the minimal inversion distance from arrangements of other species (Coluzzi et al., 2002; Xia et al., 2008). *Anopheles quadriannulatus* A and B had been traditionally considered closest to the basal species of the complex because they have several less specialized traits expected of an ancestral form: large number of hosts, feed on animal blood, tolerance for temperate climates, and disjunctive distribution (Coluzzi et al., 1979; Coluzzi et al., 2002). Phylogenetic status of a chromosomal arrangement can be determined if gene orders across inversion breakpoints are compared between ingroup and multiple outgroup species. Although several outgroup species can have their own apomorphic fixed inversions, their inversions will likely be different from fixed inversions of each other and of ingroup species. Therefore, the genes found across inversion breakpoints in ingroup species are expected to be in their ancestral order in multiple outgroup species. The first attempt to root the chromosomal phylogeny of the *An. gambiae* complex was done by a cytogenetic analysis of the inversion «a» on the left arm of chromosome 2 (denoted as 2La) in outgroup species. This arrangement was found in two outgroup species, both members of the Middle Eastern *An. subpictus* complex (Ayala, Coluzzi, 2005). As a result, *An. arabiensis* had been assumed ancestral because it has the fixed 2La inversion (Ayala, Coluzzi, 2005). Although the molecular analysis of the 2La inversion breakpoints and physical mapping of the sequences adjacent to the breakpoints in the outgroup species *An. stephensi* and *An. nili* confirmed the 2La ancestral state, the multiple origin of this inversion in the complex was ruled out (Sharakhov et al., 2006; Xia et al., 2008; Sharakhova et al., 2011).

In a recent study, the breakpoint sequences of fixed overlapping inversions 2Ro and 2Rp in the *An. merus — An. gambiae* clade and homologous sequences in *An. stephensi*, *Aedes aegypti*, and *Culex quinquefasciatus* were obtained and analyzed (Kamali et al., 2012). This study used two approaches to identify breakpoints of the fixed 2Ro and 2Rp inversions. In the first approach, multiple *An. gambiae* DNA probes derived from the cytological breakpoints to the chromosomes of *An. merus* were physically mapped by fluorescence *in situ* hybridization (FISH). In the second approach, mate-paired sequencing of the *An. merus* genome was performed and reads were mapped to the *An. gambiae* genome assembly, which has all standard arrangements. The study demonstrated that all studied outgroup species had the gene arrangement identical to that in the 2Ro breakpoints of *An. merus* and in the $2R^{+p}$ breakpoints of *An. gambiae.* Thus, sequencing, physical chromosome mapping, and bioinformatic analysis identified the 2Ro and $2R^{+p}$ arrangements in several outgroup species indicating that these arrangements are ancestral. Because 2Ro and $2R^{+p}$ uniquely characterize the *An. gambiae — An. merus* clade, these two species have the least chromosomal differences from the ancestral species of the complex as compared to other members (Kamali et al., 2012). This methodology can be used for rooting chromosomal phylogenies in other complexes of sibling species.

References

Ayala F. J., *Coluzzi M. 2005.* Chromosome speciation: humans, Drosophila, and mosquitoes. Proc. Nat. Acad. Sci. USA. 102 (Suppl. 1) : 6535—6542.

Besansky N. J., *Krzywinski J.*, *Lehmann T.*, *Simard F.*, *Kern M.*, *Mukabayire O.*, *Fontenille D.*, *Toure Y.*, *Sagnon N. 2003.* Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. Proc. Nat. Acad. Sci. USA. 100 : 10 818—10 823.

Besansky N. J., *Powell J. R.*, *Caccone A.*, *Hamm D. M.*, *Scott J. A.*, *Collins F. H. 1994.* Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. Proc. Nat. Acad. Sci. USA. 91 : 6885—6888.

Caccone A., *Garcia B. A.*, *Powell J. R. 1996*. Evolution of the mitochondrial DNA control region in the *Anopheles gambiae* complex. Insect Mol. Biol. 5 : 51—59.

Coluzzi M., *Sabatini A.*, *della Torre A.*, *Di Deco M. A.*, *Petrarca V. 2002.* A polytene chromosome analysis of the *Anopheles gambiae* species complex. Science. 298 : 1415—1418.

Coluzzi M., *Sabatini A.*, *Petrarca V.*, *Di Deco M. A. 1979.* Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. Trans. R. Soc. Trop. Med. Hyg. 73 : 483—497.

Della Torre A., *Merzagora L.*, *Powell J. R.*, *Coluzzi M. 1997.* Selective introgression of paracentric inversions between two sibling species of the *Anopheles gambiae* complex. Genetics. 146 : 239—244.

Gonzalez J., *Casals F.*, *Ruiz A. 2007.* Testing chromosomal phylogenies and inversion breakpoint reuse in *Drosophila*. Genetics. 175 : 167—177.

Hittinger C. T., *Johnston M.*, *Tossberg J. T.*, *Rokas A. 2010.* Leveraging skewed transcript abundance by RNA-Seq to increase the genomic depth of the tree of life. Proc. Nat. Acad. Sci. USA. 107 : 1476—1481.

Kamali M., *Xia A.*, *Tu Z.*, *Sharakhov I. V. 2012.* A new chromosomal phylogeny supports the repeated origin of vectorial capacity in malaria mosquitoes of the *Anopheles gambiae* complex. PLoS Pathog. 8 (10) : e1002960. doi:10.1371/journal.ppat.1002960.

Krimbas C. B., *Powell J. R. 1992. Drosophila* inversion polymorphism. Boca Raton: CRC Press.

Krzywinski J., *Besansky N. J. 2003.* Molecular systematics of *Anopheles*: from subgenera to subpopulations. Annu. Rev. Entomol. 48 : 111—139.

Sharakhov I. V., *White B. J.*, *Sharakhova M. V.*, *Kayondo J.*, *Lobo N. F.*, *Santolamazza F.*, *Della Torre A.*, *Simard F.*, *Collins F. H.*, *Besansky N. J. 2006.* Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2La) in the *Anopheles gambiae* complex. Proc. Nat. Acad. Sci. USA. 103 : 6258—6262.

Sharakhova M. V., *Antonio-Nkondjio C.*, *Xia A.*, *Ndo C.*, *Awono-Ambene P.*, *Simard F.*, *Sharakhov I. V. 2011*. Cytogenetic map for *Anopheles nili*: application for population genetics and comparative physical mapping. Infect. Genet. Evol. 11 : 746—754.

Stegnii V. N. 1991. Population genetics and evolution of malarial mosquitoes. Tomsk: Tomsk State University. (*Стегний В. Н. 1991*. *Популяционная генетика* и *эволюция малярийных комаров*.*Томск*: ТГУ. 136 с.)

Sturtevant A. H., *Dobzhansky T. 1936.* Inversions in the third chromosome of wild races of *Drosophila* pseudoobscura, and their use in the study of the history of the species. Proc. Nat. Acad. Sci. USA. 22 : 448—450.

White B. J., *Cheng C.*, *Sangare D.*, *Lobo N. F.*, *Collins F. H.*, *Besansky N. J. 2009.* The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. Genetics. 183 : 275—288.

Xia A., *Sharakhova M. V.*, *Sharakhov I. V. 2008.* Reconstructing ancestral autosomal arrangements in the *Anopheles gambiae* complex. J. Comput. Biol. 15 : 965—980.

Поступила 26 XI 2012

ХРОМОСОМНЫЕ ФИЛОГЕНИИ МАЛЯРИЙНЫХ КОМАРОВ

И. В. Шарахов

Кафедра энтомологии, Вирджинский политехнический институт и Государственный университет, Блэксбург, штат Вирджиния, США; электронный адрес: *igor@vt.edu*

Малярийные комары часто относятся к комплексам видов-двойников, члены которых морфологически и генетически похожи друг на друга. Однако виды в этих комплексах могут существенно различаться по их экологической адаптации и способности переносить малярию. Высокая степень генетического сходства между видами-двойниками усложняет реконструкцию филогенетических связей внутри комплексов. В настоящей статье рассматриваются исследования, которые выясняют отношения предок—потомок между близнецовыми видами, используя молекулярные маркеры и хромосомные инверсии. Методология, основанная на анализе точек разрыва фиксированных перекрывающихся инверсий, оказалась полезной для укоренения филогений комплексов видов-двойников, если известна хромосомная организация во внешней группе видов. Реконструкция подробного филогенеза для переносчиков малярии поможет выявить ассоциации эволюционных геномных изменений с возникновением предпочтения комарами человеческой крови и специфических экологических адаптаций.

Ключевые слова: *Anopheles*, комплексы видов-двойников, филогенетические связи, хромосомные инверсии.