Insect Molecular Biology

Distribution of genetic variation among chromosomal forms of *Anopheles gambiae* s.s.: Introgessive hybridization, adaptive inversions, or recent reproductive isolation?

William C. Black IV\(^1\) and Gregory C. Lanzaro\(^2\)

\(^1\)Department of Microbiology, Colorado State University, Fort Collins, CO 80523-1677, \(^2\)Department of Pathology, University of Texas Medical Branch at Galveston, Galveston, Texas 77555-0609

Correspond with: Dr. William C. Black IV, Department of Microbiology, Colorado State University, Fort Collins, CO 80523-1677, Lab/Office: (970)-491-8530, FAX: (970)-491-1815, E-mail: wcb4@lamar.colo
Abstract

A series of four papers in this issue explores the reproductive status of the five chromosomal forms of *An. gambiae s.s.* using molecular techniques to examine the variation among 12 different genes located throughout the *An. gambiae s.s.* genome. Results of these and previous studies are consistent with a hypothesis of at least partial barriers to gene flow between some chromosomal forms in Cote d'Ivoire and other West African countries to the north and west, but introgression between S and M types in Benin and countries to the east. Collectively these studies indicate the need for broader geographic sampling of *An. gambiae s.s.*, increased research on mechanisms of pre-zygotic reproductive isolation and field-based studies of survival and fecundity in hybrids to test for post-zygotic reproductive isolation.
Introduction

Population genetics underwent a major paradigm shift during the last decade of the 20th century with the discovery that hundreds or thousands of genetic markers of known position in a genome can be analyzed simultaneously in a single organism. Two technologies were critical in developing population genomics in arthropods (Black et al., 2000). The first of these, the Polymerase Chain Reaction (PCR), allowed for the amplification of DNA corresponding to many loci using only nanograms of genomic template DNA. As a result, a variety of highly polymorphic markers were discovered to be abundant in arthropod genomes. These include markers such as microsatellite DNAs, RAPDs (Random Amplified Polymorphic DNA), and AFLPs (Amplified Fragment Length Polymorphisms). A second set of techniques allowed the detection of single nucleotide polymorphisms (SNPs) within genes (e.g. SSCP [Single-Strand Conformation Polymorphism], SNP probes and RFLPs [Restriction Fragment Length Polymorphisms] and in a wide variety of classical genetic markers such as cDNAs and mitochondrial DNA. An important feature of these procedures is that they can be applied to large numbers of individual specimens, thereby facilitating the acquisition of adequate sample sizes required for estimating key parameters in population genetics.

With the advent of genome-wide analysis has come the ability to estimate the sampling distribution of statistics frequently applied in population genetics for individual loci with respect to their position in the genome. These include: genetic distances, population/individual inbreeding coefficients (F<sub>ST</sub> and F<sub>IS</sub>) and
linkage disequilibrium measures ($D_{ST}$ and $D_{IS}$). Within individual genes, similar statistics can be used to estimate the number of segregating and polymorphic sites as well as to identify selection acting upon individual substitutions. *The critical aspect of population genomics is that investigators can, for the first time, begin to examine the sampling distributions of these statistics and identify where estimates derived from individual loci fall within this distribution.* This allows investigators to distinguish effects that act upon the whole genome from those that act on different regions within the genome or upon individual loci. Forces that influence the entire genome include migration, drift, and inbreeding while assortative mating, selection, mutation, and recombination are forces that influence individual loci. The four papers in this series examine the reproductive status of the five chromosomal forms of *An. gambiae s.s.* by applying the principles of population genomics in an examination of variation amongst 12 different genes located throughout the *An. gambiae s.l.* genome.

During laboratory studies on the inheritance of dieldrin resistance, Davidson (1956) demonstrated F$_1$ hybrid male sterility within and among geographic populations of *Anopheles gambiae s.l.* Subsequent studies of the Mendelian inheritance of F$_1$ sterility within and among *An. gambiae s.l.* populations demonstrated the existence of six, reproductively-isolated, cryptic species (Davidson et al., 1967; Coluzzi et al., 1979) that could only be identified on the basis of chromosome inversion patterns. These six species were named *An. gambiae s.s.*, *An. arabiensis*, *An. quadriannulatus*, *An. bwambae*, *An. merus*, and *An. melus*. By way of review, recall that inversion patterns in anopheline
mosquitoes are identified by cytogenetic analysis of polytene chromosomes in the ovarian nurse cells of half-gravid females or in the salivary glands of 4th instar larvae. Recall that endomitosis (chromosome replication without nuclear or cytoplasmic fission) occurs in some insect tissues. When endomitosis occurs in insect orders that have somatic pairing between homologous chromosomes (i.e. Diptera and Collembola), a process known as polytenization occurs in which the chromosomes appear as thick, ropy structures with distinct “bands and puffs” that correspond to different densities and conformations of euchromatin and heterochromatin. The linear arrangements of these landmarks have been used as characters to define cryptic species in a number of anopheline species groups. In insects bearing an inversion on only one homologous chromosome, loop structures are visible when complementary regions align. These loop structures can be seen with compound microscopy and provide a key feature for cytogenetic analyses. During meiosis, recombination amongst loci contained within inversion heterozygotes is reduced or eliminated. As a result, gametes arising from an inversion heterozygous mosquito are orthologous for genes arising from its parents. Recombination will only occur within parents that are homozygous for an inversion.

Analyses of polytene chromosomes in An. gambiae s.s. from West Africa revealed a complex of paracentric (outside the centromere) inversions, with the majority occurring on the right arm of the second chromosome (2R) (Coluzzi and Sabatini, 1967). There are 4 contiguous but non-overlapping inversions called j, b, c, and u, and a fifth, d that overlaps with u. In each case the non-inverted form
is designated wild-type (+). These five inversions are usually present as 12 common karyotypes labeled hereafter as +++++, jb+++ , jbcu+, jb++d, j+cu+, j+++d, +bc++, ++cu+, +bc+d, +bcu+, +b+u+, and +b++d. The spatial distribution of these inversions show a strong association with ecological/climatic zones (Coluzzi et al., 1979, Bryan et al., 1982) and their distributions are not random, even on the microgeographic level (Coluzzi et al. 1977). In addition, where they occur in sympatry, their relative frequencies change seasonally, most likely in response to annual fluctuations in climate (Touré 1991, Touré et al. 1998). These observations are consistent with an hypothesis that multilocus genotypes contained within and maintained by inversions are adaptive, especially with respect to survival under varying degrees of aridity (Coluzzi et al., 1979, Coluzzi, 1982). Extensive field studies of An. gambiae s.s. in areas of sympatry indicate that 2R inversion genotypes are seldom in Hardy-Weinberg equilibrium due to a deficiency or absence of inversion heterozygotes. This result suggests barriers to gene flow among mosquito subpopulations with different 2R inversion genotypes.

On this basis, An. gambiae s.s. was subdivided into five chromosomal inversion forms: Bamako, Bissau, Forest, Mopti, and Savanna (Coluzzi et al. 1979, Bryan et al. 1982, Coluzzi 1984, Coluzzi et al. 1985). Savanna is considered the typical 2R form of An. gambiae s.s. with the broadest geographic distribution, occurring throughout sub-Saharan Africa. Savanna has 2R karyotypes +bc++, ++cu+, +bc+d, +bcu+, +b+u+, and +b++d. Bamako is found in Mali and northern Guinea associated with the upper Niger River and its
tributaries and is distinguished by inversions j+cu+ and jbcu+. Forest is a forest breeding form that is predominantly fixed for +++++ but occasionally contains the b, c, u or d inversions. Forest is thought to represent the ancestral inversion form. Mopti is found in Mali, Guinea, Cote d'Ivoire, and Burkina Faso and because of its association with flooded plains and irrigated fields, it breeds continuously even throughout the dry season. Mopti has 2R inversions +++++ +bc++, and +++u+. Bissau is endemic to The Gambiae and carries inversions +++++ and ++++d.

Population studies have revealed a low frequency of hybrids between some forms (e.g. Mopti and Savanna, Bamako and Savanna) and complete reproductive isolation between others (e.g. Bamako and Mopti) (Touré et al., 1983). However, Bamako x Mopti F₁ hybrids created under laboratory conditions display no post-mating reproductive isolation and in fact exhibit higher fecundity than the parental strains (DiDeco et al., 1980). In addition, estimates of genetic distance between forms, based on allozyme frequencies, yielded a value similar to that typically found between local populations of a single mosquito species (Cianchi, et al., 1983). In contrast, analyses of cuticular hydrocarbons suggest the possibility of pre-zygotic barriers to mating (Milligan et al., 1993). Cuticular hydrocarbon profiles were more differentiated among forms collected sympatrically within Mali and Burkina Faso than among forms collected allopatrically. Reproductive isolation between the Mopti and Savanna forms has been supported by studies of the distribution of the kdr gene, a mutation involved in pyrethroid resistance. Chandre et al. (1999) found the kdr gene at high
frequencies among Savanna populations in Burkina Faso and Côte d'Ivoire, however it was absent in Mopti populations, even in locations where the two forms occurred in sympatry.

Cytogenetic analyses alone do not allow for precise estimates of hybridization amongst the five forms because of “cryptic” heterokaryotypes and because Bamako, Forest, Savanna and Mopti share 2R inversions. Favia et al. (1994 a, b) found distinct differences in RAPD markers between Mopti and among Savanna/Bamako chromosomal forms in samples from Mali and Burkina Faso. However, these were subsequently found to be polymorphic and inconsistent in resolving forms. Favia et al. (1997) analyzed sequence variation in the 5’-end of the intergenic spacer (IGS) regions of the ribosomal DNA and found that Mopti was consistently different from Bamako and Savanna. They developed a diagnostic method based upon PCR-RFLP analysis of the IGS that consistently differentiated Mopti from Bamako and Savanna in samples from Mali and Burkina Faso. There are several incongruities between subpopulations defined as chromosomal forms and those defined by differences in the IGS sequence. The most obvious is the failure of the IGS PCR-RFLP to distinguish the Bamako and Savanna forms. In addition, analysis of field collected individuals with “hybrid” karyotypes failed to produce a hybrid PCR-RFLP pattern. However, Mopti X Bamako hybrids generated in the laboratory did result in a clearly resolved hybrid pattern. Finally, some individuals with Mopti karyotypes yielded a Savanna/Bamako IGS PCR-RFLP pattern. These observations led to a re-definition of the boundaries among forms such that karyotypes earlier
identified as fixed in one form (e.g. +b+++ , +bcu+) are shared among forms (Favia et al. 1997, Touré et al. 1998). The new definitions allow the interpretation that karyotypes identified as hybrids are in fact not hybrids, but are the consequence of low frequency polymorphisms in one or the other taxon. Furthermore, reproductive isolation among the re-defined chromosomal forms may be complete.

Lanzaro et al. (1998) conducted an intensive study of the genetic structure of populations of An. gambiae from two villages in Mali. Populations at each site were composed of the Bamako/Savanna and Mopti forms and the sibling species, An. arabiensis. Karyotypes were determined for each individual mosquito and genotypes were determined at 21 microsatellite loci that had been physically mapped to polytene chromosomes. Genetic divergence amongst chromosome forms was smallest amongst markers on chromosomes 1 and 3 but was large on chromosome 2. They concluded that the majority of observed genetic divergence between chromosomal forms are consistent with a pattern of partial reproductive isolation. In agreement with earlier studies, they also found evidence for low levels of gene flow between An. gambiae and An. arabiensis, similar to estimates based on observed frequencies of hybrid karyotypes in natural populations. While the results of Lanzaro et al. (1998) are by far the most comprehensive and complete to date, the concern over possible size homoplasy with microsatellites remains and as with all previous studies, Lanzaro et al. (1998) was restricted to Mali.
The present and previous studies have sought to resolve these issues using the paradigm of population genomics to determine the magnitude and consistency of genic differentiation among the five chromosomal forms, including analyses of individual loci throughout the genome. Three alternative hypotheses concerning the observed absence or deficiency of inversion homozygotes are implicit with this approach. First, the five taxa may be completely reproductively isolated. If this is true, then the pattern of inversion heterozygote deficiency should also be apparent at most genomic markers regardless of their position with respect to inversions. However, if reproductive isolation is recent then there could be frequent shared polymorphisms among the 12 genes analyzed. Nevertheless, they should still be in Hardy-Weinberg equilibrium within each of the five taxa but exhibit a strong deficiency of heterozygotes when data from all taxa are combined. Where multiple forms occur in sympathy hybrid individuals should never be observed. Second, there may be partial reproductive isolation among the five taxa. Gene flow may occur only in certain geographic locations or in certain seasons. In this scenario, hybrids exist but may be rare. There will again be frequent examples of shared polymorphisms among the five taxa. However, in contrast with hypothesis 1, heterozygote deficiency should be less in areas or seasons of sympathy than among allopatric or seasonally-discontinuous populations. If this pattern is seen, then we are still left to identify pre-zygotic or post-zygotic mechanisms to explain heterozygote deficiency in areas or seasons of sympathy. Third, there is no reproductive isolation among the five taxa. If so, then most genomic markers should be homogeneous among the five taxa and
genotypes should be in Hardy-Weinberg proportions amongst each of the five taxa when in sympaty. If this pattern is seen, then we are still left with the problem of explaining the absence of inversion heterozygotes. If genetic loci are homogeneous among sympatric populations then, under the assumptions of population genomics, the possibility should be considered that inversions are adaptive and subject to rapid directional selection among locations and seasons, a possibility originally suggested by Coluzzi (1982). Only one or a few advantageous alleles are necessary to sweep to fixation an entire set of genes contained within the inversion. Consequently, gene frequencies at loci contained within or linked to inversions would be significantly different between forms, but unlinked loci should be undifferentiated.

The five studies in this series seek to address these three alternative hypotheses by analyzing gene sequences from throughout the *Anopheles gambiae* s.s. genome to determine the magnitude and consistency of genetic differentiation amongst the five chromosomal forms. Della Torre et al. (2000) is a geographically- and conceptually-comprehensive study that seeks to extend the findings of the studies described above to collections made from throughout the geographic range of *An. gambiae* s.s. They classify over 1,000 mosquitoes from 45 collections according to karyotypes and rDNA types S and M. They include an analysis of genotypes at the *kdr* locus (2L) in two populations. Direct evidence of hybridization between rDNA types S and M was found. Out of 1,161 *An. gambiae* s.s. that were karyotyped and analyzed for the rDNA types, either rDNA types S and M were found in 30/42 samples, S and M were sympatric in
12/42 samples and 3/1,161 (0.26%) of mosquitoes were M/S hybrids. Furthermore, in Cote d’Ivoire the \textit{kdr} gene occurs at a frequency of 0.90 in S mosquitoes collected from the savanna and 0.06 in S individuals collected from forest habitats while it is undetected in M individuals. In contrast, in Benin just east of Cote d’Ivoire, the \textit{kdr} occurs at a frequency of 0.90 in S and 0.30 in M mosquitoes. They demonstrate that mosquitoes with rDNA types S and M share the standard 2R arrangement in or near the forest belt.

\textit{Favia \textit{et al}.} (2000) describes the development of chromosomal form-specific PCR using 120 nucleotides in the 3’- end of the 28S gene and \textasciitilde{}2 kb from the 5’- end of the IGS for faster and less expensive diagnosis of chromosome forms. They refer to these IGS genotypes as S (Savanna type) and M (Mopti types). \textit{Gentile \textit{et al}.} (2000), \textit{Favia \textit{et al}.} (2000) and \textit{Mukabayire \textit{et al}.} (2000) examine sequence variation in the adjacent Internal Transcribed Spacer (ITS) and show that two ITS genotypes (ITS I and ITS II) are in complete linkage disequilibrium with the S and M IGS types, respectively. \textit{Gentile \textit{et al}.} (2000) describes a third ITS type III, but its association with the IGS is unknown since the authors did not examine the IGS in Sao Tome. For the remainder of this review we will refer to these as IGS/ITS genotypes as “rDNA types S and M.”

The remaining two studies contribute many additional genes toward a population genomics analysis of the three proposed hypotheses. \textit{Gentile \textit{et al}.} (2000) analyzed the mitochondrial Cytochrome Oxidase I and II genes. On the sex chromosome they examine guanylate cyclase (\textit{gua}) introns VIII and V/VI and the ITS. They analyze genes \textit{pKM122} and \textit{pKM2} on chromosome arm 2L, gene
patterns of DNA sequence variation in the ITS and five unlinked single copy nuclear loci for evidence of reproductive isolation among Savanna, Bamako, Mopti and Forest, as well as *An. arabiensis* and *An. merus*.

These studies collectively support the earlier findings of Lanzaro *et al.* (1998), suggesting that their results cannot be explained by size homoplasy in microsatellites alone. The majority of markers, whether microsatellites, cDNAs or RAPDs indicate that chromosomal forms are not genetically differentiated and therefore are not completely reproductively isolated from one another. The exceptions to this pattern occur in the rDNA IGS/ITS genes in Mali and Burkina Faso and amongst genes within 2R inversions throughout a diversity of geographic regions. The results from the analysis of 2R inversions are expected for the reasons discussed above. Explanation of results from the IGS/ITS analyses await an indication of how much recombination occurs along markers in the heterochromatic arm of the X chromosome or determination of whether the rDNA cistron lies within an inversion. However, it is known from many studies that gene conversion among rDNA cistrons overestimate genetic differences among populations. Thus strong inferences about gene flow cannot be made with this marker.

Collectively, the observations made to date are consistent with an hypothesis of at least partial barriers to gene flow between S and M rDNA types in Cote d'Ivoire and other West African countries to the north and west, but introgression between S and M types in Benin and countries to the east. Mopti,
Savanna and Forest all share the standard 2R arrangements and therefore could intergrade in forested regions. Forest might act as a bridge for gene flow between S and M rDNA types, at least in southern regions of West Africa where all three inversion forms are sympatric. Homosequential individuals could subsequently emigrate to savanna habitats where hybridization would occur with different 2R inversions typical of Savanna/Bamako or could emigrate to flooded plains and irrigated fields in West Africa typically associated with Mopti inversion types. Overall, the results of these studies are consistent with an hypothesis of partial, regional reproductive isolation among the five chromosomal forms.

These results point to the need for intensive research on pre- and post-zygotic mechanisms of genetic isolation. The results of della Torre et al. (2000) and Gentile et al. (2000) demonstrate the importance of broad geographic and comparative research in West African countries where hybridization has and has not been detected. Analysis of pre-zygotic mechanisms is beyond the scope of geneticists and indicates the need for increased funding of behavioral and ecological studies of potential pre-zygotic isolation mechanisms amongst the five chromosomal forms. The results also point to the need for field-based studies of survival and fecundity in hybrids. The growing field of Anopheles genomics may soon allow insect geneticists to apply the principles of population genomics to identify genes within the 2R inversions that correlate with environmental variables.
References


different chromosomal forms of Anopheles gambiae. Proc Natl Acad Sci USA 91: 10315-10319.


hydrocarbons for identifying members of the *Anopheles gambiae* (Diptera: Culicidae) complex. *Bull Entomol Res* 83: 613-624.


