9

Gene flow among populations of *Anopheles gambiae*: a critical review

Gregory C. Lanzaro[#] *and Frederic Tripet*[#]

Abstract

The success of genetic-control programmes aimed at introducing genes into wild *Anopheles gambiae* populations depends on our understanding of the genetic structure of these populations. Population-genetic studies are required for identifying discrete population groups across Africa, determining their geographical distribution and evaluating the degree to which they may be reproductively isolated. Population studies are also needed to estimate the rate at which genes may spread within and between populations at various spatial scales and to identify biological and physical features of the environment that may interfere with their movement. Studies of the mechanisms of reproductive isolation between molecular/chromosomal forms of *An. gambiae* can be used to validate the results of population-genetic approaches and aid in the development of reproductively competitive laboratory strains. In the following, we review past and recent studies that cover these aspects of *An. gambiae* population genetics and ecology. We critically discuss the validity of the designs and methodologies involved in an attempt to provide a sound basis for future undertakings. We also discuss new directions and priorities in the light of the recent developments toward a genetic-control strategy for *An. gambiae*.

Keywords: *Anopheles gambiae*; mosquito; population structure; gene flow; DNA; microsatellites; transposable element; transgene; dispersal; Africa; divergence; assortative mating; polytene chromosome

Introduction

The movement of **transgenes** from one lineage or population to another depends on mating between an individual carrying the gene and one that does not. This remains true even where transgenes are coupled with drive mechanisms such as **transposable elements** (Ribeiro and Kidwell 1994) or endosymbionts (Turelli and Hoffmann 1999). Although designs for novel approaches to target vector populations of mosquitoes are interesting and potentially useful, the population-genetics component remains poorly understood. Critically, most conceptual models for genetic control assume that the mosquito population into which a transgene system is to be released represents a single, randomly mating unit. There is considerable evidence that some populations of *Anopheles gambiae* (Giles) are subdivided by barriers to

[#] Department of Entomology, University of California, Davis, California 95616, USA. E-mail: gclanzaro@utmb.edu; ftripet@ucdavis.edu

reproduction and that **gene flow** via migration among geographic populations is limited. Field studies designed to estimate levels and patterns of gene flow within and among those structured populations are needed to provide a foundation for predicting the potential utility of new molecular-level approaches, and for designing field trials to evaluate their efficacy under natural conditions in Africa. In many locales that have not been well studied there may exist complexities in population structure that remain to be described. In addition, defining levels of gene exchange among partially isolated subpopulations or "incipient" species is critical in determining to what extent these are reproductively isolated. If some level of gene flow is occurring this information may in fact be used to advantage in the design of a release strategy. The success of any genetic-control programme aimed at controlling or manipulating natural populations through the release of genetically modified mosquitoes will depend on our knowledge of genetic diversity in natural populations, information on how this diversity is distributed in time and space, and on understanding the forces generating and maintaining diversity among populations.

Estimating gene flow among natural populations

Methods for estimating how much gene flow occurs in natural populations have been divided into two broad classes. Direct methods involve estimating **dispersal** distances and reproductive success of individuals that disperse by direct observation. Indirect methods rely on allele frequencies or differences in DNA sequences to estimate levels of gene flow that must have occurred in order to explain observed patterns. Direct methods provide a "snapshot" of contemporary gene flow at the time the observations are made. In studies of gene flow among mosquito populations direct methods have involved mark-release-recapture experiments or studies of the progeny or sperm from field-collected females. Indirect methods rely on determining the standardized variance in allele frequencies among populations ($\mathbf{F_{ST}}$), from which estimates of the number of migrants per generation (N_m) may be inferred by the simple relationship, $F_{ST} \approx (1/4N_m + 1)$ (Wright 1931). These estimates are strongly influenced by patterns of gene flow that may have occurred in the past or are the consequence of rare events. Indirect methods have frequently been applied to comparisons of populations separated by vast distances, but it should be pointed out that studies employing indirect methods on the local scale are more likely to yield usable estimates of contemporary gene flow (see Rousset 2001 for useful discussion). Both methods have limitations. Direct methods are only useful for describing gene flow in mosquito populations separated by short distances, whereas statistical methods for estimating gene flow employing indirect estimates rely on assumptions that may not be appropriate for mosquito populations (Slatkin 1985; Bossart and Prowell 1998; Whitlock and McCauley 1999; however see Bohonak et al. 1998). In addition, genetic methods rely on assumptions concerning the evolutionary behaviour of the genetic markers used to obtain gene frequency data and these may be problematic. A comparison of estimates of gene flow among *An. gambiae* populations in Mali made by both direct and indirect methods yielded consistent results, suggesting that despite their limitations, the methods being applied were reasonably reliable, at least on the scale and in the geographical localities under study (Taylor et al. 2001). The vast majority of studies aimed at describing gene flow among *An gambiae* populations have employed indirect methods.

Patterns of gene flow on a macrogeographic scale

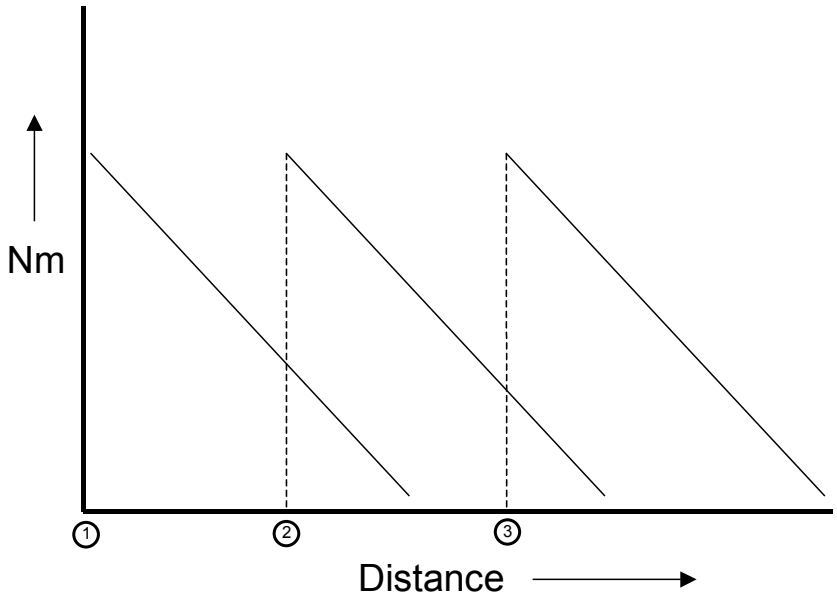
Ultimately malaria-control efforts in Africa will have to be conducted on a large geographical scale. Although the development of a strategic plan for such an undertaking will require co-ordinating efforts along political borders the success of such an effort will depend largely on identifying regions of operation based on biologically meaningful boundaries. Early work based on the distribution of chromosome inversions suggests substantial differentiation between populations from different parts of Africa. More recent studies based on biochemical and molecular genetic markers have found no evidence of **genetic differentiation** among *An. gambiae* populations across Africa. We review these results and show why they are equivocal. We continue with an examination of studies that describe genetic differentiation on a macro-geographical scale within two regions of interest. We first discuss East Africa and the importance of the Great Rift Valley for gene flow, then consider the more complex populations of West and Central Africa. Throughout the text we pay special attention to studies that examined if populations of *An. gambiae* can be described by an isolation by distance model, i.e., if the genetic distance between populations increases with geographical distance. Depending on the scale on which they are conducted these studies may detect unknown genetic complexities within large population groups or tell us about the amount of gene flow among local subpopulations. From a practical point of view these analyses are important because they provide us with an approximation of the geographical scale on which releases will be most effective (see Figure 1).

Gene flow across Africa

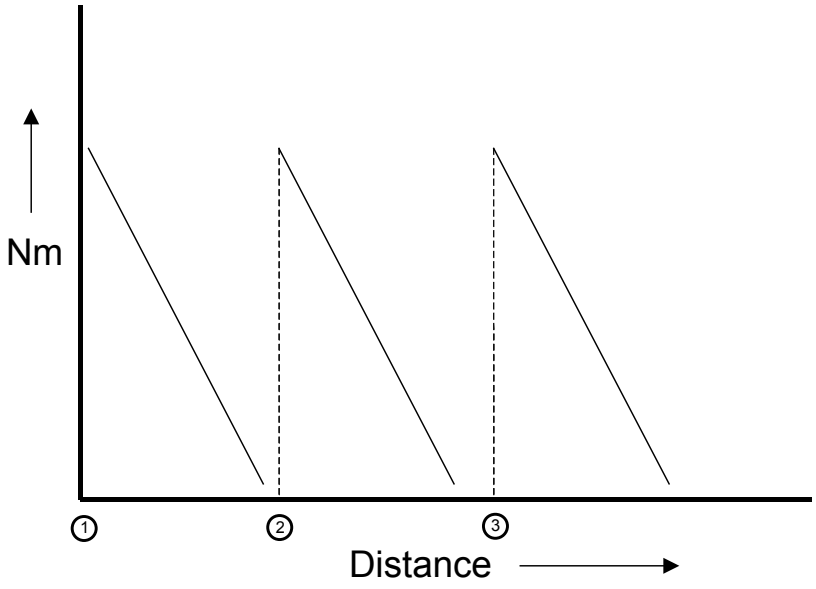
An. gambiae is extremely versatile regarding tolerance to a wide variety of micro- and macro-environmental conditions, as evidenced by its broad geographic distribution and ability to thrive at sites that experience seemingly unsuitable seasonal variation in climate (Coluzzi, Petrarca and Di Deco 1985). This would suggest that individuals are adapted to local conditions and are likely to be genetically distinct from individuals in populations where conditions are different. Evidence for this phenomenon is presented in the work of Coluzzi, Petrarca and Di Deco (1985) and Touré et al. (1998b), who show a strong association of certain chromosome **inversions** with dry or wet habitats. To date there are few studies describing gene flow among populations on a large geographic scale. The early work of Coluzzi and co-workers, as described in detail below, suggests that populations over large distances differ dramatically with respect to the distribution of paracentric chromosome inversions. Later work employing biochemical (isozyme) and DNA markers (mtDNA and **microsatellites**) suggests a very different picture. Lehmann et al. (1996) conducted a study of the distribution of isozyme and microsatellite variation between populations from Kenya and Senegal. They found little differentiation among the populations analysed ($F_{ST} = 0.016$) and estimated remarkably high levels of gene flow between populations ($N_m > 7.7$) separated by more then 6,000 km. Besansky et al. (1997) analysed sequence data for a segment (665 bp) of the mitochondrial ND5 gene from seven villages in Kenya and three villages in Senegal. They also found populations from Kenya and Senegal to be remarkably homogeneous ($F_{ST} = 0.085$); consequently their estimates of gene flow among these populations was also very high ($N_m = 5.4$). They obtained similar results for *An. arabiensis*, where populations separated by up to 7,000 km were relatively homogeneous ($F_{ST} = 0.044$, $N_m = 10.8$).

Figure 1. Schematic representation of three populations that fit an isolation-by-distance model. Solid lines represent independent sets of estimates of Nm in relation to geographical distance calculated for populations (1), (2) and (3).

(A) Introduction of a transgene at location (1) will result in the subsequent introduction of the desired gene into populations (2) and (3) via gene flow. This is because Nm > 0 in all cases. The distance between 1, 2 and 3 is less than the dispersal range, there are no physical barriers separating the three populations and no reproductive barriers to gene flow.



(B) Introduction of a transgene at location (1) will not result in the movement of the desired gene into populations (2) and (3) because of a lack of gene flow among these three populations. In this case Nm ≈ 0. This may be the result of (a) the distance between the three sites exceeding the dispersal ability of individual mosquitoes, (b) the presence of physical barriers separating the sites or (c) reproductive barriers between the three populations.



They found no evidence for "isolation by distance", that is, they observed no relationship between levels of genetic **divergence** (F_{ST}) and geographic distance. Taken together these studies suggest that *An. gambiae* over its range is comprised of populations that exchange individuals at a rate sufficient in magnitude to prevent them from diverging genetically. The authors suggest that the lack of divergence between widely separated populations may reflect their history, being the consequence of recent expansion of *An. gambiae* over the past 2,000-5,000 years (associated with the expansion of human populations in Africa, as suggested by Coluzzi, Petrarca and Di Deco 1985). Alternatively, they suggest that gene flow among populations across continental Africa may be contemporary, resulting from active migration and passive transportation via the activity of man. The authors favour the latter, pointing to the well-known introduction of *An. gambiae* into Brazil via the activities of humans during the 1930's (Soper and Wilson 1943).

Although these studies are interesting we do not feel that they are definitive. The interpretation that these results demonstrate extensive contemporary gene flow over enormous distances is not consistent with what is known about the dispersal capabilities of *An. gambiae.* For example, the maximum flight range of *An. gambiae* estimated by direct observation has been reported to be in the range of 3.6 - 7km (Gillies 1961; Touré et al. 1998a). Indirect estimates suggest a dispersal range of tens or at most hundreds of kilometers (Mclain et al. 1989; Carnahan et al. 2002). Dispersal over long distances via human activity is suggested by the experience in Brazil, as well as by regular cases of airport malaria in areas far from malaria-infested areas (Martens and Hall 2000; Karch et al. 2001). However, notion that *An. gambiae* exists over its range as a single, undifferentiated population is directly challenged by the non-random spatial distribution of chromosome inversions, as reported by numerous authors and discussed in detail below. A careful assessment of the work suggesting extensive gene flow across Africa reveals several shortcomings. There are methodological problems associated with sampling strategies that weaken these reports. In both studies (Lehmann et al. 1996; Besansky et al. 1997) only two regions are included (Kenya and Senegal) that are at the extremes of the range of *An. gambiae*. A more thorough study including many more populations across the range might reveal a pattern in the relationships among them that would suggest alternative explanations for the apparent lack of divergence. For example, the results described by Besansky et al. (1997) for *An. arabiensis* were not supported by the findings of Donnelly and Townson (2000). This was a more thorough study involving nine localities in East Africa over a transect covering 4,500 km, from Sudan to Mozambique. Relationships among populations were assessed using allele frequencies at eight microsatellite loci and sample sizes were large (N = 29-59 per population per locus). These researchers found highly significant differences in genotype frequencies between populations separated by 200 km and in Mozambique, by as little as 25 km. In addition, they report a highly significant positive correlation between F_{ST} and geographic distance, suggesting a good fit to the isolation by distance model of population structure. The study employing mtDNA (Besansky et al. 1997) suffers further by the analysis of too few mosquitoes to adequately assess variation within and among collection sites. As few as four individuals were analysed for some collection sites and not more than ten for any. The authors conducted a re-analysis of the same sites in Kenya, but this time included a substantially larger sample size (Lehmann et al. 2000). When sample sizes were increased (from a total of 37 to a total of 71) their estimate of divergence among the same populations was far greater and

consequently estimates for gene flow were much lower. The authors concluded that this discrepancy was the result of a "small sample size effect" in the original study.

The most extensive analysis to date of *An. gambiae* on a large geographic scale was recently presented by Lehmann et al. (2003). They report the results of analysis of variation at 11 microsatellite loci from 16 sites in 10 countries. A cluster analysis based on F_{ST} values revealed a major subdivision among *An. gambiae* populations in continental Africa: They identified a North-western (NW) population group, containing populations in Senegal, Ghana, Nigeria, Cameroon, Gabon, Democratic Republic of Congo and western Kenya and a South-eastern (SE) group including populations in eastern Kenya, Tanzania, Malawi and Zambia. Differentiation between these two population groups was high, $F_{ST} > 0.1$. Genetic differentiation among populations within the two groups were substantially lower and a significant relationship between genetic distance and geographic distance was observed, consistent with an isolation-by-distance model of population structure. The authors suggest that differentiation between the SE and NW groups may be the consequence of a recent bottleneck in the SE group and physical barriers limiting gene flow between the two groups. On close examination of their data it appears that the population in Zambia occupies a position intermediate between the eastern and western populations and may represent a third group, or equally likely, lie in a zone that represents a bridge, with respect to gene flow, between the eastern and western populations.

In summary, studies conducted for the purpose of describing the genetic structure of *An. gambiae* on the continental scale have yielded conflicting results. The early work based on the distribution of chromosome inversions suggested major differences between East-African and West-African populations. Later work based on isozyme, microsatellite and mitochondrial DNA were interpreted as demonstrating that *An. gambiae* exists as a single, more or less undifferentiated population over its entire range. Recent work, based on microsatellite frequencies among 16 populations spanning the continent, revealed a major division of populations so that two genetically distinct population groups can be recognized. We believe that this latest study most accurately describes the large-scale genetic structure of *An. gambiae*. The earlier work was based on an analysis of populations from only two regions, located 6,000 km apart. The population in Kenya included in these studies was from Asembo Bay, west of the Great Rift Valley which, as discussed in detail below, serves as a major barrier to gene flow. Had these earlier studies included populations closer to the coast of Kenya the results would have been very different. This is an important point because it illustrates nicely the importance of developing an effective sampling scheme when undertaking a population-genetics study, especially on a large spatial scale.

East Africa and the Great Rift Valley

McLain et al. (1989) conducted a study of the distribution of restriction-fragment length polymorphisms (RFLPs) in the IGS region of the ribosomal DNA locus among populations of *An. gambiae* in Kenya. The study included populations from seven villages in western Kenya and an eighth from coastal Kenya. They found that populations located ten or more kilometers apart differed significantly in the frequencies of RFLPs and that the western Kenya populations, located roughly 700 km from the others, contained no RFLPs in common with them. The authors interpreted these results as suggesting that *An. gambiae* populations in this region fit

an isolation-by-distance model with gene flow limited between populations separated by as little as 10 km and that populations no more than a few hundred kilometers apart are, in effect, completely isolated.

Lehmann et al. (1997) conducted a study of populations in the same part of western Kenya. They examined population structure based on five microsatellite loci plus sequence variation in a 648 base-pair fragment of the mitochondrial *ND5* locus. Their analysis was aimed at describing levels of genetic divergence among populations at several spatial scales, including between houses within villages and between villages at distances up to 50 km. Not surprisingly they found no evidence for genetic divergence between houses within villages, but their results at larger scales contradicted the findings of McLain et al. (1989). The distribution of both microsatellite and mtDNA polymorphism suggested no significant differentiation (F_{ST} not significantly different from zero) between villages at distances up to 50 km, leading the authors to conclude that gene flow is extensive among populations up to this distance. These results fit well with their earlier work (described above) in which they found no significant divergence in microsatellite or mtDNA between populations separated by 6,000 km. As in their earlier work they suggest that high rates of contemporary gene flow, relatively rare events of extinction-recolonization and/or historical gene flow may explain their observations. Here they argue that historical gene flow is an unlikely factor because they found no difference in levels of differentiation between rapidly evolving microsatellites compared with slowly evolving isozyme and mtDNA loci.

There are problems with this argument. Divergence estimated by isozymes cannot be compared with microsatellites because the efficiency with which enzyme electrophoresis methods detect polymorphisms is low, with as little as 25% of existing amino acid substitutions detectable (Selander 1976). Furthermore, there are indications that mutation rates for microsatellites in insects may not be as high as once thought (Schug et al. 1998). Donnelly, Licht and Lehmann (2001) have shown that historical gene flow, in the form of range expansion, is the most likely explanation for the high N_m values estimated from the distribution of microsatellite and mtDNA allele frequencies. Kamau et al. (1998) conducted a study similar to that of Lehmann et al. (1997) and in the same region of Kenya. Kamau et al. (1998) included analysis of seven microsatellite loci in populations from seven villages less than 10 km apart in the Asembo Bay area. They likewise found no evidence of divergence among populations at these sites (F_{ST} = 0.0016, N_m = 5.66). Their study also included a comparison between the seven Asembo Bay populations and a population from Kilifi, located about 700 km to the east, on the coast of the Indian Ocean. As in the McLain et al. (1989) study, they found significant divergence between the eastern and western populations (F_{ST} = 0.075, N_m = 1.54). These results suggest that gene flow between populations from the east and west of Kenya is severely restricted and cast doubt on the earlier reports suggesting extensive gene flow among *An. gambiae* populations continent-wide. Kamau et al. (1998) provide the first report that the Great Rift Valley (GRV), which divides east and west Kenya, serves as a barrier to gene flow between *An. gambiae* populations on either side of it.

This phenomenon was further explored in a more detailed study reported by Lehmann et al. (1999). They studied six populations; four west of the GRV, and two on the eastern side. Frequencies at nine microsatellite loci were used to compare populations. The goal of this study was to determine if the high degree of differentiation between populations east and west of the GRV are due to the GRV serving as a barrier to gene flow or due to other factors such as differences in

effective population size (N_e) among the populations analysed or distance alone. Although they detected some difference in N_e between populations on either side of the GRV, neither differences in population size nor distance alone could explain the level of divergence between populations across the GRV. They conclude that the GRV does indeed represent a barrier to gene flow and that this phenomenon explains the observed genetic divergence between populations separated by it. These results are consistent with the earlier reports of McLain et al. (1989) and Kamau et al. (1998), but are in apparent conflict with the work of Besansky et al. (1997), who report no significant differentiation between populations on either side of the GRV. This disagreement was resolved, as described above, when the Besansky et al. study was repeated with an increase in sample size (Lehmann et al. 2000). Taken together these studies present solid evidence for the GRV acting as a significant barrier to gene flow between the populations of *An. gambiae* which it separates. In light of the recent work by Lehmann et al. (2003), in which they described the genetic structure of *An. gambiae* over a large portion of its range (described above), it appears that the GRV, extending from western Kenya to the southern shore of Lake Nyasasa in Mozambique, is a significant barrier to gene flow, forming an east/west division of *An. gambiae* populations. Populations in the southwest of the GRV have not, with the exception of a single population in Zambia, been analysed. Additional sampling in this region (e.g. eastern Zaire and Zambia) may clarify the importance of the southern part of the GRV as barrier to gene flow. The data from the west of Zambia presented in Lehmann et al. (2003) suggests that populations on the southwestern side of the GRV could represent a third group or be intermediate, representing a bridge between the eastern and more western populations.

West and Central Africa

Studies of the distribution of chromosome-inversion polymorphism in Nigeria revealed the existence of two chromosomal forms, the Savanna and Forest forms (Coluzzi et al. 1979; Coluzzi, Petrarca and Di Deco 1985). The relative abundance of the two forms is clinal with the Savanna form predominating in the drier northern part of the country and gradually decreasing so that the Forest form is most abundant in the south. The two forms appear to intergrade, at least to some extent, where they occur together. Onyabe and Conn (2001) conducted a study with the goal of determining the extent of gene flow among populations both within and between ecological zones in Nigeria. Their study was based on the distribution of alleles at ten microsatellite loci and included 39-46 samples per site collected from eight sites over an 833km transect. The microsatellite loci studied were selected so that five were located within inversions and five were located elsewhere in the genome. Overall they found significant levels of genetic differentiation among populations; differences were highest for comparisons between populations from the savanna and forest ecological zones (F_{ST} = 0.028-0.087) as opposed to within zones (F_{ST} = 0.000-0.048 for populations within the savanna zone). An evaluation of the relationship between F_{ST} and distance between sites revealed a highly significant correlation suggesting that population structure fits an isolation-by-distance model. However, on closer examination it was found that the relatively high F_{ST} values were largely due to three loci located within inversions and that, when these were removed from the analysis divergence, they dropped to very low or insignificant levels and the correlation between F_{ST} and distance was no longer significant. They suggest that the higher level of divergence for loci associated with inversions is the consequence of

hitchhiking on genes under selection that are located within the inversions. The hitchhiking phenomenon was also observed by Lanzaro et al. (1998) as described below. Onyabe and Conn (2001) conclude that gene flow among *An. gambiae* populations in Nigeria is extensive and does not appear to be limited by geographical distance. One limitation with their study was their failure to karyotype the material studied. It is possible that at least some of their samples consisted of mixed Savanna and Forest forms. Indeed Coluzzi, Petrarca and Di Deco (1985) found evidence that at some sites in Nigeria the Savanna and Forest forms do exist in sympatry and that this was associated with a deficiency of **heterokaryotypes**. The effect would be to obfuscate any possible correlation between F_{ST} and geographic distance.

A similar study was conducted in Mali by Carnahan et al. (2002). They included samples collected from six villages on a 536 km transect along the Niger River Valley. Populations at some sites were mixed chromosomal forms and individuals were identified to form by cytogenetic analysis, so that the authors identified eleven populations at these six sites. The habitat was uniform over the transect and they report no evidence of obvious barriers to dispersal within the study area. Their analysis included between 5 and 23 microsatellite loci; sample sizes per site ranged between 4 and -190 individuals. Sites with small sample sizes were only included if the number of loci analysed was >20. They found a significant correlation between form F_{ST} and geographic distance over all loci and for chromosomes 3 and X when the analysis was conducted for microsatellites on each chromosome separately. Values for N_m ranged from 64.43 to 1.26 between population pairs along the transect. They conclude that in this part of Africa genetic differentiation at microsatelllite loci is consistent with the isolation-by-distance model. Interestingly N_m reported in this study, $N_m = 1.26$, for two populations separated by a distance of 444 km, was smaller than the N_m reported by Lehmann et al. (1996) for populations separated by 6,000 km ($N_m = 3.4$). There are, however, several shortcomings with this work. Not all loci were the same for each population studied so that if the behaviour of individual microsatellite loci is different this might affect estimates of F_{ST}. In addition, sample sizes per site in some cases were as small as four individuals.

Patterns of gene flow among populations on a local scale

In addition to environmental factors that affect patterns of gene flow at the macro-geographic and regional scales, population structuring within ecological zones and habitats can potentially ruin efforts to drive genes into wild *An. gambiae* populations. In areas where several subspecific forms of the *An. gambiae* complex co-occur theory would predict the existence of strong pre-mating barriers to hybridization (Liou and Price 1994; Butlin 1995). If this proves to be true in field populations, driving a transgene into reproductively isolated subpopulations using a single strain of mass-reared mosquito may be ineffective. Thus as progress is being made in the development of genetically modified mosquitoes, understanding the population structure of *An. gambiae* at a local scale is becoming ever more important. A correct assessment of the amount of reproductive isolation, i.e., the amount of gene flow between *An. gambiae* subpopulations is critical for assessing the feasibility of a mosquito genetic-control programme, developing adequate mass-rearing facilities and designing initial field trials. In the next paragraph we will first briefly overview what is known of the genetic structure of *An. gambiae* on a local scale as well as methods currently employed for identifying and describing complex populations. Thereafter each subsection reviews past and current studies aimed at understanding population

structure and patterns of gene flow by making inferences from genetic analyses based on different markers. A critical limiting factor in choosing optimal genetic markers for describing population complexity is the extent of our understanding of mechanisms of reproductive isolation and speciation, and of their consequences for the evolution of different parts of the genome in this species complex. Behavioural-ecology studies can shed light on processes leading to reproductive isolation and validate interpretations inferred from population-genetic studies. Advances in this area of research will be discussed at the end of this section.

Overview of past and current advances

Population structuring within *An. gambiae sensu stricto* has classically been approached using cytogenetic methods to describe chromosomalinversion polymorphism. While such studies successfully identified the cryptic species within the complex, within *An. gambiae s.s.* they have revealed contrasting levels of polymorphism among different regions in Africa. Broadly speaking, one can distinguish 'East-African populations' with limited amounts of inversion polymorphism, and 'West-African populations' in which a remarkable number of **chromosomal arrangements** have been identified. Cytogenetic studies in West Africa revealed a wide array of inversion karyotypes and led to the definition of several chromosomal forms. The discovery of chromosomal forms that may have recently diverged or are currently undergoing speciation has generated considerable interest and controversy. The apparent absence of post-mating barriers to reproduction between forms and their general lack of genetic differentiation undermined the case for defining additional cryptic species within *An. gambiae*. Recent surveys (Della Torre et al. 2001; 2002) based on sequencing of ribosomal DNA (r-DNA) revealed fixed differences between some populations and led to the definition of two major r-DNA molecular forms, thus far encompassing all populations in continental Africa. At present, neither karyotypes nor r-DNA sequences alone can satisfactorily describe populations within *An. gambiae s.s.* over its entire range, but combining these tools might provide us with adequate resolution for identifying major population groups. Despite these difficulties, considerable advances in our understanding of patterns of current gene flow between karyotypically and/or molecularly defined forms have been made using microsatellite markers. Moreover, the ongoing selective sweep associated with *kdr* resistance (Chandre et al. 1998; Weill et al. 2000) may provide researchers with an ideal marker for estimating the extent of current reproductive isolation between forms (Diabate et al. 2003). A PCR diagnostic based on the r-DNA loci (Favia et al. 1997; 2001) has also simplified the search for hybrids between molecular forms where they occur in sympatry and facilitated the study of mating patterns between forms.

Chromosome inversion polymorphism

Studies of chromosomal rearrangements have been conducted in a number of African countries and have revealed much higher levels of inversion polymorphism in West Africa compared to other areas (Coluzzi, Petrarca and Di Deco 1985; Petrarca and Beier 1992). The frequencies of chromosomal arrangements have been used to test populations for departures from Hardy-Weinberg equilibrium and in many cases provided evidence that *An. gambiae* populations are composed of several discrete units (Bryan et al. 1982; Fonseca et al. 1996; Appawu et al. 1994; Akogbeto and Di

Deco 1995). Coluzzi et al. (Coluzzi, Petrarca and Di Deco 1985) defined 5 forms based on the distribution of the most common chromosomal arrangements (Figure 2 and Table 1): (1) the Forest form, characterized by the typical non-inverted arrangement 2R+/+, 2L+/+, or by a single inversion polymorphism due to inversion 2Rb, 2Rd or 2La; (2) Bissau, characterized by high frequencies of the 2Rd inversion and standard 2L+ arrangement; (3) Savanna, exhibiting high frequencies of 2Rb and 2La inversions as well as polymorphism involving the 2Rcu arrangements and polymorphism in the j, d and the rare k inversion (see Table 1); (4) Bamako, characterized by the fixed 2Rjcu arrangement and polymorphism in the 2Rb inversion; (5) Mopti, showing high frequencies of 2Rbc, 2Ru and nearly fixed for 2La.

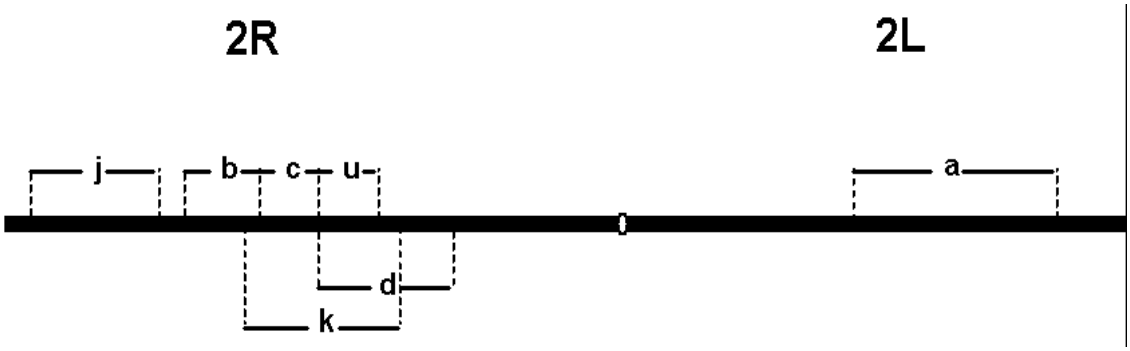


Figure 2. Position of the inversions on the second chromosome of *An. gambiae s.l.*. Five chromosomal forms have been described based on the arrangements of such inversions described from the banding patterns of polytene chromosome.

The best-documented example of population structure based on karyotype frequencies and inferred through testing compliance to the Hardy-Weinberg equilibrium are the studies by Touré et al. (Touré et al. 1994; 1998b). These were conducted in Mali where the Savanna, Bamako and Mopti chromosomal forms occur in sympatry. The authors found hybrid-like karyotypes between the Savanna and Mopti forms and between the Savanna and Bamako forms, but only one Mopti/Bamako heterokaryotype was identified (Touré et al. 1998b; Coluzzi et al. 2002). The three forms were successfully crossed and backcrossed in the laboratory suggesting that reproductive isolation is maintained essentially by pre-mating barriers to reproduction (Di Deco et al. 1980; Persiani, Di Deco and Petrangeli 1986). The interpretation of certain karyotypes as being between-form hybrids was later challenged by genetic analysis of the carriers of such arrangements using r-DNA markers (see next paragraph), and it now appears that these may be caused by polymorphic or floating inversions typical of one form, but occurring rarely in others. This hypothesis is strengthened by reports of "hybrid" karyotypes from areas where only one of the parental forms is known to occur (Touré et al. 1998a). Although it is possible that these "hybrid" individuals migrated into these areas, it seems more likely that these karyotypes represent rare individuals that are members of the indigenous gene pool. On the other hand, when such hybrid-like arrangements are found in areas where multiple forms are sympatric (Touré et al. 1998a), the possibility remains that these are backcrossed hybrids. Generally speaking the occurrence of such atypical floating inversions within the 5 broadly defined forms (Touré et al. 1998a) or the intergradation of forms renders the assignment of certain individuals to a particular form problematic. The Bamako, Savanna and Mopti forms are thought to intergrade with the Forest form where these forms co-occur (Coluzzi, Petrarca and Di Deco

1985; Della Torre et al. 2001). This phenomenon has also been reported between the Bissau and Savanna forms (Bryan et al. 1982).

Table 1. Typical and less frequent **chromosomal arrangements** on the 2R and 2 L arms of the second chromosome in the five chromosomal forms described by Coluzzi, Petrarca and Di Deco (1985). Following the description of the chromosomal arm are the potential inversion arrangements separated by slashes. A '+' describes the standard arrangement or the occurrence of non-inverted standard arrangements among inverted ones. In such cases the arrangements are not fixed, as for example in the various combinations of inversions of the Savanna form. In other instances, inversions can be fixed such as the j inversion in the Bamako form.

Chromosomal forms	Typical arrangements		Less frequent arrangements	
	2R	2L	2R	2L
Forest	2R+	2L+	2Rb/d/+	2La
Bissau	2Rd	2L+	2R+	2La
Savanna	2Rb/+	2La/+	2Rcu/bcu/bd/bcd/d/j/jb/jbd/jbcu/jcu/bk/+	
Bamako	2Rjcu/jbcu	2La		
Mopti	2Rbc/u	2La	2R+	2L+

Paracentric inversions suppress recombination among loci located within or near them, resulting in linkage associations that protect combinations of genes at multiple loci (Sturtevant 1926). These so called co-adapted multi-locus **gene complexes** (Mayr 1963; Dobzhansky 1970) are thought to confer selective advantages in different environmental conditions (Coluzzi 1982; Touré et al. 1994). The adaptive value of certain arrangements in relation to specific ecological zones has been demonstrated by transect studies that described inversion frequencies across or within chromosomal forms and from studies of the seasonal distributions of forms where they occur in sympatry (Coluzzi et al. 1979; Touré et al. 1998b). Clines in the frequency of the 2Rb, bc, d and 2La arrangements suggest that these confer a selective advantage to drier climates and habitats (Coluzzi et al. 1979). For example, changes in frequencies of the Rbc arrangement typical of the Mopti form have been shown to correlate with differences in annual rainfall between localities and with variation in monthly rainfall within a single locality (Touré et al. 1994; 1998b). Because forms intergrade in some localities, it has generally been considered that the genetic determinants of reproductive isolation between forms are not associated with inversions themselves (Coluzzi 1982; Coluzzi, Petrarca and Di Deco 1985). For example, the Mopti and Savanna forms in Mali share overlapping combinations of the same inversion polymorphism, Rbc and Ru in Mopti and Rb and Rcu in Savanna. The co-occurrence of forms such as these that do not differ by a fixed inversion, yet seem to mate assortatively, cannot easily be explained unless genetic differences outside these inversions are involved in mate recognition. The potential role of inversions in the evolution of forms has been discussed by Coluzzi et al. (Coluzzi 1982; Coluzzi, Petrarca and Di Deco 1985; Coluzzi et al. 2002). They consider inversions to be 'chromosomal mechanisms that preserve gene associations arising in temporary isolates subject to flush and crush in geographically and/or ecologically marginal zones' (Coluzzi, Petrarca and Di Deco 1985; Coluzzi et al. 2002). In this context, inversions can be considered important units of selection in a process of polygenic reorganization associated with **peripatric speciation** (Carson 1982).

DNA sequencing and the ribosomal-DNA perspective

A considerable amount of DNA sequencing has been done in attempts to develop chromosomal form-specific diagnostics based on fixed differences between them. These involved sequencing of nuclear genes such as the *white* gene on the X chromosomes (Besansky et al. 1995), the *tryptophan oxygenase* gene (Mukabayire et al. 1996) on 2R, *pKM2* on 2L (Gentile et al. 2001), the *gua introns VIII, V, VI, F72* and *Gambif1* on chromosome 3 (Gentile et al. 2001). None of these single-copy nuclear genes yielded fixed differences between chromosomal forms (Mukabayire et al. 2001; Gentile et al. 2001). Sequencing of the mitochondrial gene *COI/II* likewise failed to provide characteristic form-specific differences (Gentile et al. 2001). In contrast, sequencing of the rapidly evolving non-coding regions of ribosomal DNA, a tandemly arrayed multigene family, proved to be more rewarding. Favia et al. (1997) first found diagnostic RFLPs in this region and identified 10 nucleotide residues that differ between the Mopti and the Savanna or Bamako forms in a 620bp fragment of the Intergenic Spacer (IGS) region (Favia et al. 2001). These findings were critical because they were the first fixed differences found between chromosomal forms; they led to the development of a PCR-based diagnostic to differentiate Mopti individuals carrying the M-form of r-DNA from Bamako and Savanna individuals carrying the S-form of r-DNA. The diagnostic was developed using samples from Mali; among those early samples there were a few equivocal cases where karyotyping did not match the molecular diagnostic (Favia et al. 1997; Della Torre et al. 2001). The diagnostic was also used to identify between-form hybrid-like karyotypes. M/S hybrids produced in the laboratory yielded clearly distinguishable hybrid patterns. Surprisingly, however, field-collected individuals carrying "hybrid" karyotypes did not produce results consistent with their being hybrid, but rather produced either M or S patterns (Favia et al. 1997). This observation supports the notion that certain karyotypes, thought to be fixed in one form or another, are in fact shared, occurring commonly in one form and rarely in another. However, the possibility that these cases could be backcrosses between forms has not been explored. Thus, the discovery of fixed differences between forms validated the rapid concerted evolution of the r-DNA involving gene conversion between r-DNA cistrons. Paradoxically, the r-DNA diagnostic is still considered a reliable tool to identify F_n hybrids.

The Favia diagnostic has since been used to type individuals from other areas of Africa. It has become apparent that the correspondence between the Mopti chromosomal form and the M molecular type does not hold in other areas of West Africa and that the two types can occur within the previously defined chromosomal forms (Della Torre et al. 2002). The M and S types have been found in the Forest and Savanna forms in various parts of West Africa. Thus far, the M form characterizes the Bissau form and the S-form characterizes the Bamako form. The Mopti form is generally associated with the M molecular type except for rare cases (Della Torre et al. 2001). A major advantage of the M/S classification is that it has facilitated processing large numbers of samples and screening them for hybrid-like patterns. Favia et al. (1997) did not report observing F_1 hybrids, but Della Torre et al. (2001) reported 2 hybrids out of 1161 sampled individuals. Tripet et al. (2001) reported 1 hybrid out of a sample of 330 females analysed, Edillo et al. (2002) reported 4 hybrid larvae out of ~350, and Diabate et al. (2003) found 1 hybrid out of 2000 individuals.

Another region of the r-DNA locus, the Intergenic Transcribed Spacers (ITS), ITS 1 and 2, located upstream of the IGS locus has been studied in detail by Gentile et al.

(Gentile et al. 2001; 2002). As in the IGS region, fixed differences were found between chromosomal forms. The two loci (IGS and ITS) were found to be in perfect linkage disequilibrium. Polymorphic sites in the ITS segregated according to M and S forms, thus resulting in the definition of two main ITS types, type I (S form) and type II (M form) (2001; Gentile et al. 2002). They recovered a third ITS type (ITS type III) from an isolated population of S-form individuals on the island of Sao Tomé. ITS type III shares homologies with types I and II. The Gentile et al. (2002) study is particularly interesting in that, in addition to identifying fixed differences between the previously described M and S types, it provides a **parsimony analysis** of the ITS sequences from various populations across Africa. The parsimony analysis was conducted to estimate the minimum number of evolutionary steps, in this case point mutations, between ITS variants. As advocated by different authors (Favia et al. 2001; Gentile et al. 2001) this step should be taken whenever new populations are studied to ensure that the M/S or ITS type I, II and III classifications provide the most parsimonious description of r-DNA distributions.

Evidence from isozymes and microsatellites

Estimates of genetic distances between chromosomal forms have been calculated, first using Wright's F_{ST}'s based on allozyme frequencies (Cianchi et al. 1983) and later using microsatellite loci (Lanzaro et al. 1998; Wang et al. 2001). Estimates based on isozymes yielded values similar to those found between local populations of a single mosquito species (Cianchi et al. 1983). Because isozymes may not have the resolving power to detect differentiation between recently diverged forms, Lanzaro et al. (1998) conducted a study based on 21 microsatellite loci distributed over the genome, examining genetic differentiation between the Bamako and Mopti forms in Mali. The study revealed strong genetic differentiation between *An. gambiae* and *An. arabiensis*, used here as an out-group. Within *An. gambiae s.s.* different patterns of genetic differentiation, depending on the genomic location of the microsatellite loci, were observed. No genetic differentiation was found on the third and X chromosome, whilst strong linkage disequilibrium and low levels of genetic differentiation were found for loci located on the second chromosome (Lanzaro et al. 1998). Another study using microsatellites distributed on all three chromosomes was conducted by Wang et al. (2001), but this time comparing a mixed 'population' of *An. arabiensis* from Mali and Kenya with an M molecular form (equivalent to Mopti in that area) population of *An. gambiae s.s.* and a pooled sample of S form (mixed Savanna and Bamako) from two villages. Because the authors did not provide evidence that the two populations of *An. arabiensis* that were pooled had similar allele frequency distributions, it is difficult to judge if the data can be used for the tests they employ. Similarly, heterogeneity in the S-form pooled samples (2 locations and 2 chromosomal forms) may violate the assumptions of homogeneity required by most methods for estimating genetic divergence between the M and S forms. Despite these flaws, the paper reported results similar to Lanzaro et al. (1998) and, interestingly, showed that two loci located near the r-DNA coding area on the X chromosome exhibited strong differentiation and linkage with the r-DNA M and S types. In a recent paper, Wondji, Simard and Fontenille (2002) found low levels of genetic differentiation (F_{ST} = 0.06) between sympatric populations of the M and S forms of the Forest cytotype in Cameroon. Some microsatellite studies have been interpreted as indicative of incomplete reproductive isolation between chromosomal/molecular forms, with low amounts of gene flow occurring at regions of the genome away from inversions

(Lanzaro et al. 1998; Tripet et al. 2001; Onyabe and Conn 2001). In others, it has been suggested that reproductive barriers between the M and S molecular forms may be complete (Wang et al. 2001; Wondji, Simard and Fontenille 2002).

The *kdr*-resistance genetic sweep

Paradoxically the ongoing spread of *kdr* resistance to pyrethroid insecticides in West Africa has been invaluable in providing researchers with an unambiguous tool to describe the extent of reproductive isolation between *An. gambiae* forms. Early work by Chandre et al. (1999) in Ivory Coast showed *kdr* resistance to be present only in the S molecular form, thus supporting the complete reproductive-isolation hypothesis. Shortly thereafter, however, *kdr* resistance was found in M-form individuals in Benin and molecular comparison of sequence in an upstream intron suggested that it arose through **introgression** rather than as an independent, new mutation (Weill et al. 2000). Sequences from 90 resistant individuals from Benin, Ivory Coast and Burkina Faso also showed a loss of genetic diversity in the intron upstream of the *kdr* locus. These results suggest that areas of the genome proximal to the *kdr* locus may be hitchhiking along with the *kdr* mutation in what is commonly referred to as a **genetic sweep** (Weill et al. 2000). This has important implications in terms of gene flow between molecular forms as it would prove unequivocally that introgression recently occurred between the M and S forms and thus support the hypothesis of residual gene flow in areas of sympatry. In another study of karyotyped and molecular-typed material from Ivory Coast and Benin, Della Torre et al. (Della Torre et al. 2001) showed that *kdr* segregated with the molecular IGS type in Ivory Coast but that in Benin it occurred in both M and S types of the Forest and Savanna chromosomal forms (Fanello, Akogbeto and Della Torre 2000; Della Torre et al. 2001). The *kdr* mutation has since been identified in the M form from Burkina Faso (Diabate et al. 2003) and from M-form individuals from Mali (D. Norris pers. comm.). It is unknown at this point if these instances of *kdr* resistance in M-form populations are caused by independent mutations or if they are again the result of introgression between forms, although the latter seems likely based on the Benin experience. It is noteworthy however that the *kdr* resistance gene has only been reported in populations where the M form co-occurs with S-form Savanna or Forest populations where between-form gene flow is a plausible explanation.

Processes of reproductive isolation

Introgression between chromosomal forms that do not exhibit obvious post-mating reproductive barriers (Di Deco et al. 1980; Persiani, Di Deco and Petrangeli 1986), raises the question of how they maintain genetic identity. If ‘hybrids’ between forms do not suffer fitness costs, recombinational events would ultimately break down pre-mating reproductive barriers. Tripet et al. (2001) used the Favia diagnostic combined with microsatellite genotyping to study mating patterns between the M and S forms in Mali. Genetic analyses of wild-caught *An. gambiae* females and the sperm extracted from their spermatheca revealed strong **assortative mating** within forms. However, a small percentage of matings were between forms (females mated with the wrong males) (Tripet et al. 2001; 2003). If between-form mating, ‘hybrid’ larvae and ‘hybrid’ adults occur in the wild, then one might reasonably expect that selection acts against hybrid-like genotypes. This hypothesis may be tested by comparing MxS hybridization rates at different developmental stages and attempting to detect a

reduction in hybrid survival. Although a formal study of this kind has not yet been conducted, some interesting data exist. In the village of N'Gabacoro Droit in Mali, where the Mopti and Bamako forms predominate during the rainy season, Tripet et al. (2001) found the frequency of cross-mating equal to 0.00839 but the frequency of hybrid adults substantially lower, 0.00303, lending support to the hypothesis that selection acts against hybrids. In the nearby village of Banambani, where the Mopti, Savanna and Bamako forms co-occur, Edillo et al. (2002) found a frequency of hybrid larvae of 0.01127, suggesting higher introgression levels in that population, but, unfortunately, the matching data on adult hybridization rate were not available. Clearly, larger studies examining hybridization rates at different developmental stages within single populations are needed in order to provide the statistical power required to test this hypothesis adequately.

If strong assortative mating occurs in natural populations, then there must be reliable cues allowing chromosomal forms to recognize each other. Finding differences in recognition mechanisms could allow us to map such phenotypic differences to precise areas of the genome. There has been, thus far, little research done on behavioural or physiological differences between forms. In a preliminary study, Milligan et al. (1993) found differences in cuticular hydrocarbons between samples from chromosomal forms living in sympatry, but no study was ever published confirming these results with adequate sample sizes. It has also been suggested that sibling species within the *An. gambiae* complex could recognize each other using flight tones created by their wing-beat frequency. Although recordings of flight tones from laboratory colonies of *An. arabiensis* and *An. gambiae* s.s. seemed to support the wing-beat hypothesis (Brodgon 1998), field data showed significantly different but largely overlapping distributions of wing-beat frequencies (Wekesa et al. 1998). Tripet et al. (unpublished data) used F_1's from field-collected mosquitoes reared in the laboratory and measured under controlled conditions and found no difference in wing-beat frequencies between sympatric populations of *An. arabiensis* and the Mopti and Savanna forms of *An. gambiae s.s.*. The chemical or behavioural cues and mechanisms used by mosquitoes for mate recognition remain unknown.

Conclusions and perspectives

Although discrepancies exist in the literature certain conclusions emerge that provide a useful picture of patterns of gene flow among populations of *An. gambiae*, particularly as they relate to the release of genetically modified individuals for malaria control. This species is genetically diverse over its range, even at the local level. On the continental scale *An. gambiae* does not appear to exist as a single, genetically undifferentiated population, but rather can roughly be subdivided into at least two major population groups, one in the north-western part of its range and a second in the south east. These two groups appear to be separated by the Rift Valley and other inhospitable environments. The importance of physical features that restrict the movement of genes between populations such as the Rift Valley in East Africa has been confirmed by many studies and it is likely that similar, perhaps more subtle physical barriers exist in other places within the range of *An. gambiae*. Distinct, recognizable subpopulations, designated as chromosomal or molecular forms exist on a local scale in West Africa. In some parts of West Africa populations consist of only one form, in others multiple forms may occur in a single village. Although populations in East Africa do contain substantial genetic polymorphism there is no evidence that this is organized into reproductively isolated subpopulations (forms).

Studies aimed at determining the influence of geographic distance on the extent of gene flow between populations within *An. gambiae* forms have led to conflicting results. Some studies yielded results suggesting that gene flow fits an isolation-by-distance model, others not. Populations in which isolation by distance is found can be considered 'problem-free' from the perspective of the release of transgenic strains because the transgenes should spread between sites. Currently, however, detecting isolation-by-distance patterns is too often an issue of using proper methodology rather that a reflection of the mosquito biology. Studies based on indirect estimates of gene flow on the local scale suggest that this species is a good disperser and that gene flow among villages within a radius of tens or even hundreds of kilometers is extensive. Although dispersal is likely to vary depending on the season and the nature of the environment, this suggests that isolation by distance should be detected whenever adequate sample sizes and genetic markers are used and a proper geographical scale is identified. Further research in this area is warranted and could be facilitated by recent methodological improvements. These include development of more powerful statistical procedures for estimating gene flow (e.g. assignment tests) along with improvements in the types of molecular markers available for describing the genetics of populations (e.g. polymorphism in single-copy nuclear genes).

Comparing the relative amounts of gene flow taking place between forms and among populations is the first step towards predicting the trajectory of introduced refractory genes. Given the multiplicity of markers available and their respective pros and cons this poses two major challenges. The first one is to detect genetic complexity itself within often poorly described populations. The second lies in measuring gene flow between those populations. In some areas of Africa, complex populations where current gene flow between forms is known to occur have been identified. Those are populations where M/S form hybrids have been found or hybridization is suspected because the *kdr* resistance introgressed from the S form into the M form. The frequency of 'hybrids' between molecular forms has been estimated at 0.05-0.3% depending on the population under study, a rate adequate to explain the general lack of genetic differentiation among forms that has been observed in numerous studies. If we use a conservative estimate of effective population size (N_e) of 2,000 this yields an estimate of $Nm = 0.003 \times 2{,}000 = 6$, a value large enough to result in the complete introgression of subpopulations into a single, undifferentiated gene pool (Hartl 1980). Decreased fitness of hybrid individuals may provide a mechanism that maintains the integrity of subpopulations, but there are currently no firm data supporting this hypothesis.

Structuring of populations into multiple, reproductively isolated gene pools would present an obvious difficulty to attempts at introducing genes into these populations, but may also present advantages. One difficulty in the stable introduction of a gene into a large population is overcoming an introduction threshold (minimum number of released transgenic mosquitoes required to transform a population) that, among other things, depends on the size of the host population. Thus, the fragmentation of populations into multiple smaller and reproductively isolated subpopulations helps to reduce this threshold effectively and facilitate the successful integration of an introduced gene. This would, however, require the production of multiple engineered strains. There may be additional advantages if, as seems to be the case, reproductive barriers between *An. gambiae* forms, are not complete. In this case a transgene could initially be introduced into a subpopulation of one form. As this gene increases in frequency a higher and higher proportion of the relatively rare matings between forms would involve individuals carrying the gene of interest. If a sufficiently effective drive

mechanism is associated with the desirable gene this introgression would represent a new introduction into the sympatric subpopulation of the other form. If valid, this scenario would not require the production of multiple transgenic strains. Current observations on the spread of the *kdr* gene may be an example of this phenomenon. The *kdr* gene was initially introduced into S-form subpopulations via mutation or migration. Later, the same gene appeared in sympatric M-form subpopulations. In this case *kdr* is being "driven" by exposure of these populations to pyrethroid insecticides, imparting a fitness advantage to those individuals carrying this gene.

There is a strong need for behavioural ecological studies aimed at understanding mating behaviour in habitats where different forms of *An. gambiae* co-exist. For example, the wing-beat hypothesis assumes that the different forms swarm together but such data have never been collected in the field. If such is the case and if, as suggested by wing-beat studies, auditory cues are not involved in form-specific recognition in mating, one can reasonably assume that contact pheromones may be involved, and this needs to be explored. Advances in these areas of research may provide invaluable help in producing transgenic strains that mate competitively in the field. Whatever the genetic-drive system involved, it is likely that a genetically modified mosquito will exhibit fitness costs. These can be associated with the genetic construct itself, linked to the expression of inserted genes or caused by genetic drift and selection processes occurring during their rearing in the laboratory (Huettel 1976; Yan, Severson and Christensen 1997; Catteruccia, Godfray and Crisanti 2003). These costs should have few consequences if released mosquitoes mate randomly with local populations. However, if they mate only with a subset of the target population, selection will drive the evolution of wild populations that avoid mating with released mosquitoes. Under such conditions, transforming local populations may prove much more difficult than expected and may require either much larger releases or complex schemes aiming at regularly refreshing the genetic make-up of transformed strains.

Acknowledgements

We thank Jan Conn, Dia Elnaiem and David Onyabe for their comments on earlier versions of the manuscript.

Glossary

assortative mating – Sexual reproduction involving the non-random pairing of individuals which are more closely alike than the average in respect of one or more traits.

chromosomal arrangement – Structural characteristics of the chromosome with special reference to aberrations such as inversions.

dispersal – Outward spreading of organisms or propagules from their point of origin or release; one-way movement of organisms from one home site to another.

divergence – The acquisition of dissimilar characters or traits by related organisms.

effective population size, Ne – The average number of individuals in a population which are assumed to contribute genes equally to the succeeding generation.

Fst – Correlation between allelic frequency distributions inferred from random gametes within a subpopulation relative to the frequencies calculated for gametes within the entire population.

gene complex - System of interacting genes.

gene flow – The exchange of genes within and between populations by interbreeding or migration.

genetic differentiation - The acquisition of dissimilar genetic characteristics by related organisms.

genetic sweep - The decrease in genetic variation due to hitchhiking found in areas of the genome adjacent to an advantageous mutation after it spread through a population.

heterokaryotype – A genome or individual that is heterozygous for a chromosomal rearrangement such as an inversion.

hitchhiking - Increase in allelic frequency due to the low recombination rates observed for alleles at loci located near an advantageous mutation but themselves neutral with regard to that mutation.

introgression – The spread of genes of one species into the gene pool of another by hybridization and backcrossing.

inversion – A chromosomal aberration in which a segment of the chromosome exhibits the reverse orientation of bases at a particular site, the segment having rotated through 180°.

microsatellite DNA – DNA sequence made up of a single sequence motif, no more than six bases long, that is tandemly repeated without interruption by any other base or motif.

parsimony analysis – Analysis based on the principle of invoking the minimal number of evolutionary changes to infer phylogenetic relationships.

peripatric speciation – The origin of new species by the modification of peripherally isolated founder populations.

transgenes – Genes that are transferred artificially from one species to an unrelated species, in which they are maintained and manifested phenotypically.

transposable element – A class of genes that are capable of spontaneously moving from one chromosome to another or from one position to another in the same chromosome.

References

Akogbeto, M. and Di Deco, M.A., 1995. Répartition des membres du complexe *Anopheles gambiae* et de leurs variants chromosomiques au Bénin et au Togo, Afrique occidentale. *Journal of African Zoology,* 109, 443-454.

Appawu, M.A., Baffoe Wilmot, A., Afari, E.A., et al., 1994. Species composition and inversion polymorphism of the *Anopheles gambiae* complex in some sites of Ghana, West Africa. *Acta Tropica,* 56 (1), 15-23.

Besansky, N.J., Bedell, J.A., Benedict, M.Q., et al., 1995. Cloning and characterization of the white gene from *Anopheles gambiae*. *Insect Molecular Biology,* 4 (4), 217-231.

Besansky, N.J., Lehmann, T., Fahey, G.T., et al., 1997. Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An. arabiensis*, suggest extensive gene flow. *Genetics,* 147 (4), 1817-1828.

Bohonak, A.J., Davies, N., Roderick, G.K., et al., 1998. Is population genetics mired in the past? *Trends in Ecology and Evolution,* 13, 360.

Bossart, J.L. and Prowell, D.P., 1998. Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution,* 13 (5), 202-206.

Brodgon, W.C., 1998. Measurement of flight tonedifferentiates among members of the *Anopheles gambiae* species complex (Diptera: Culicidae). *Journal of Medical Entomology,* 35, 681-684.

Bryan, J.H., Di Deco, M.A., Petrarca, V., et al., 1982. Inversion polymorphism and incipient speciation in *Anopheles gambiae s. str.* in the Gambiae, West Africa. *Genetica,* 59, 167-176.

Butlin, R.G., 1995. Reinforcement: an idea evolving. *Trends in Ecology and Evolution,* 10, 432-434.

Carnahan, J., Zheng, L., Taylor, C.E., et al., 2002. Genetic differentiation of *Anopheles gambiae s.s.* populations in Mali, West Africa, using microsatellite loci. *Journal of Heredity,* 93 (4), 249-253.

Carson, H.L., 1982. Speciation as a major reorganization of polygenic balances. *In:* Barigozzi, C. ed. *Mechanisms of speciation: proceedings from the International Meeting on Mechanisms of Speciation, May 4-8, 1981, Rome, Italy*. Liss, New York, 411-433. Progress in Clinical and Biological Research. Vol. 96.

Catteruccia, F., Godfray, H.C.J. and Crisanti, A., 2003. Impact of genetic manipulation on the fitness of *Anopheles stephensi* mosquitoes. *Science,* 299 (5610), 1225-1227.

Chandre, F., Manguin, S., Brengues, C., et al., 1998. Current distribution of a pyrethroid resistance gene (kdr) in Anopheles gambiae complex from West Africa and further evidence for reproductive isolation of the Mopti form. *The malaria challenge after one hundred years of malariology. Papers from the Malariology Centenary Conference, Accademia Nazionale dei Lincei, Roma,* 41, 1-3.

Chandre, F., Manguin, S., Brengues, C., et al., 1999. Current distribution of a pyrethroid resistance gene (kdr) in *Anopheles gambiae* complex from West Africa and further evidence for reproductive isolation of the Mopti form. *Parassitologia,* 41 (1-3), 319-322.

Cianchi, R., Villani, F., Touré, Y.T., et al., 1983. Electrophoretic study of different chromosomal forms of *Anopheles gambiae s.s. Parassitologia,* 25, 239-241.

Coluzzi, M., 1982. Spatial distribution of chromosomal inversions and speciation in Anopheline mosquitoes. *In:* Barigozzi, C. ed. *Mechanisms of speciation: proceedings from the International Meeting on Mechanisms of Speciation, May 4-8, 1981, Rome, Italy*. Liss, New York, 143-153. Progress in Clinical and Biological Research. Vol. 96.

Coluzzi, M., Petrarca, V. and Di Deco, M.A., 1985. Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Bollettino di Zoologia,* 52, 45-63.

Coluzzi, M., Sabatini, A., Della Torre, A., et al., 2002. A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science,* 298 (5597), 1415-1418.

Coluzzi, M., Sabatini, A., Petrarca, V., et al., 1979. Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 73, 483-497.

Della Torre, A., Costantini, C., Besansky, N.J., et al., 2002. Speciation within *Anopheles gambiae* - the glass is half full. *Science,* 298 (5591), 115-117.

Della Torre, A., Fanello, C., Akogbeto, M., et al., 2001. Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa. *Insect Molecular Biology,* 10 (1), 9-18.

Di Deco, M.A., Petrarca, V., Villani, F., et al., 1980. Polimorfismo cromosomico da inversioni paracentriche ed eccesso degli eterocriotipi in ceppi di *Anopheles* allevati in laboratorio. *Parassitologia,* 22, 304-306.

Diabate, A., Baldet, T., Chandre, C., et al., 2003. Kdr mutation, a genetic marker to assess events of introgression between the molecular M and S forms of *Anopheles gambiae* (Diptera: Culicidae) in the tropical savannah area of West Africa. *Journal of Medical Entomology,* 40 (2), 195-198.

Dobzhansky, T., 1970. *Genetics of the evolutionary process*. Columbia University Press, New York.

Donnelly, M.J., Licht, M.C. and Lehmann, T., 2001. Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Molecular Biology and Evolution,* 18 (7), 1353-1364.

Donnelly, M.J. and Townson, H., 2000. Evidence for extensive genetic differentiation among populations of the malaria vector *Anopheles arabiensis* in eastern Africa. *Insect Molecular Biology,* 9 (4), 357-367.

Edillo, F.E., Touré, Y.T., Lanzaro, G.C., et al., 2002. Spatial and habitat distribution of *Anopheles gambiae* and *Anopheles arabiensis* (Diptera: Culicidae) in Banambani village, Mali. *Journal of Medical Entomology,* 39 (1), 70-77.

Fanello, C., Akogbeto, M. and Della Torre, A., 2000. Distribution of the pyrethroid knockdown resistance gene (kdr) in *Anopheles gambiae s.l.* from Benin. *Transactions of the Royal Society of Tropical Medicine and Hygiene,* 94, 132.

Favia, G., Della Torre, A., Bagayoko, M., et al., 1997. Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Molecular Biology,* 6 (4), 377-383.

Favia, G., Lanfrancotti, A., Spanos, L., et al., 2001. Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae s.s. Insect Molecular Biology,* 10 (1), 19-23.

Fonseca, L.F., diDeco, M.A., Carrara, G.C., et al., 1996. *Anopheles gambiae* complex (Diptera: Culicidae) near Bissau City, Guinea Bissau, West Africa. *Journal of Medical Entomology,* 33 (6), 939-945.

Gentile, G., Della Torre, A., Maegga, B., et al., 2002. Genetic differentiation in the African malaria vector, *Anopheles gambiae s.s.*, and the problem of taxonomic status. *Genetics,* 161 (4), 1561-1578.

Gentile, G., Slotman, M., Ketmaier, V., et al., 2001. Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae s.s. Insect Molecular Biology,* 10 (1), 25-32.

Gillies, M.T., 1961. Studies on the dispersion and survival of *Anopheles gambiae* Giles in East Africa, by means of marking and release experiments. *Bulletin of Entomological Research,* 52, 99-127.

Hartl, D.L., 1980. *Principles of population genetics*. Sinauer, Sunderland, MA.

Huettel, M.D., 1976. Monitoring the quality of laboratory-reared insects: a biological and behavioral perspective. *Environmental Entomology,* 5, 807-814.

Kamau, L., Lehmann, T., Hawley, W.A., et al., 1998. Microgeographic genetic differentiation of *Anopheles gambiae* mosquitoes from Asembo Bay, Western Kenya: a comparison with Kilifi in coastal Kenya. *American Journal of Tropical Medicine and Hygiene,* 58 (1), 64-69.

Karch, S., Dellile, M.F., Guillet, P., et al., 2001. African malaria vectors in European aircraft. *Lancet,* 357 (9251), 235.

Lanzaro, G.C., Touré, Y.T., Carnahan, J., et al., 1998. Complexities in the genetic structure of *Anopheles gambiae* populations in west Africa as revealed by microsatellite DNA analysis. *Proceedings of the National Academy of Sciences of the United States of America,* 95 (24), 14260-14265.

Lehmann, T., Besansky, N.J., Hawley, W.A., et al., 1997. Microgeographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. *Molecular Ecology,* 6 (3), 243-253.

Lehmann, T., Blackston, C.R., Besansky, N.J., et al., 2000. The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya: The mtDNA perspective. *Journal of Heredity,* 91 (2), 165-168.

Lehmann, T., Hawley, W.A., Grebert, H., et al., 1999. The rift valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya. *Journal of Heredity,* 90 (6), 613-621.

Lehmann, T., Hawley, W.A., Kamau, L., et al., 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: Comparison of microsatellite and allozyme loci. *Heredity,* 77 Part 2, 192-200.

Lehmann, T., Licht, M., Elissa, N., et al., 2003. Population structure of *Anopheles gambiae* in Africa. *Journal of Heredity,* 94 (2), 133-147.

Liou, L.W. and Price, T.D., 1994. Speciation by reinforcement of premating isolation. *Evolution,* 48, 1451-1459.

Martens, P. and Hall, L., 2000. Malaria on the move: Human population movement and malaria transmission. *Emerging Infectious Diseases,* 6 (2), 103-109.

Mayr, E., 1963. *Animal species and evolution*. Belknap, Cambridge.

Mclain, D.K., Collins, F.H., Brandling-Bennett, A.D., et al., 1989. Microgeographic variation in rDNA intergenic spacers of *Anopheles gambiae* in western Kenya. *Heredity,* 62 (Pt 2), 257-264.

Milligan, P.J.M., Phillips, A., Broomfield, G., et al., 1993. A study of the use of gas chromatography of cuticular hydrocarbons for identifying members of the *Anopheles gambiae* (Diptera: Culicidae) complex. *Bulletin of Entomological Research,* 83, 613-624.

Mukabayire, O., Caridi, J., Wang, X., et al., 2001. Patterns of DNA sequence variation in chromosomally recognized taxa of *Anopheles gambiae*: evidence from rDNA and single-copy loci. *Insect Molecular Biology,* 10 (1), 33-46.

Mukabayire, O., Cornel, A.J., Dotson, E.M., et al., 1996. The Tryptophan oxygenase gene of *Anopheles gambiae*. *Insect Biochemistry and Molecular Biology,* 26 (6), 525-528.

Onyabe, D.Y. and Conn, J.E., 2001. Genetic differentiation of the malaria vector *Anopheles gambiae* across Nigeria suggests that selection limits gene flow. *Heredity,* 87 Part 6, 647-658.

Persiani, A., Di Deco, M.A. and Petrangeli, G., 1986. Osservzioni di laboratorio su polimorfismi da inversione originati da incroci tra popolazioni diverse di *Anopheles gambiae s.s. Annali dell 'Istituto Superiore di Sanita,* 22 (1), 221-223.

Petrarca, V. and Beier, J.C., 1992. Intraspecific chromosomal polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. *American Journal of Tropical Medicine and Hygiene,* 46, 229-237.

Ribeiro, J.M. and Kidwell, M.G., 1994. Transposable elements as population drive mechanisms: specification of critical parameter values. *Journal of Medical Entomology,* 31 (1), 10-16.

Rousset, F., 2001. Genetic approaches to the estimation of dispersal rates. *In:* Clobert, J., Danchin, E., Dhondt, A. A., et al. eds. *Dispersal*. Oxford University Press, Oxford, 18-28.

Schug, M.D., Hutter, C.M., Wetterstrand, K.A., et al., 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Molecular Biology and Evolution,* 15 (12), 1751-1760.

Selander, R.K., 1976. Genic variation in natural populations. *In:* Ayala, F. J. ed. *Molecular evolution*. Sinauer, Sunderland, 21-45.

Slatkin, M., 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics,* 16, 393-430.

Soper, F.L. and Wilson, D.B., 1943. *Anopheles gambiae in Brazil 1930-1940*. Rockefeller Foundation, New York.

Sturtevant, A.H., 1926. A crossover reducer in *Drosphila melanogaster* due to inversion of a section of the third chromosome. *Biologisches Zentralblatt,* 46, 697-702.

Taylor, C., Touré, Y.T., Carnahan, J., et al., 2001. Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, west Africa. *Genetics,* 157 (2), 743-750.

Touré, Y.T., Dolo, G., Petrarca, V., et al., 1998a. Mark-release-recapture experiments with *Anopheles gambiae s.l.* in Banambani Village, Mali, to determine population size and structure. *Medical and Veterinary Entomology,* 12 (1), 74-83.

Touré, Y.T., Petrarca, V., Traoré, S.F., et al., 1998b. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia,* 40 (4), 477-511.

Touré, Y.T., Petrarca, V., Traoré, S.F., et al., 1994. Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae s. str.* in Mali, West Africa. *Genetica,* 94, 213-223.

Tripet, F., Touré, Y.T., Dolo, G., et al., 2003. Frequency of multiple inseminations in field-collected *Anopheles gambiae* females revealed by DNA analysis of transferred sperm. *American Journal of Tropical Medicine and Hygiene,* 68 (1), 1-5.

Tripet, F., Touré, Y.T., Taylor, C.E., et al., 2001. DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Molecular Ecology,* 10 (7), 1725-1732.

Turelli, M. and Hoffmann, A.A., 1999. Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. *Insect Molecular Biology,* 8 (2), 243-255.

Wang, R., Zheng, L.B., Touré, Y.T., et al., 2001. When genetic distance matters: Measuring genetic differentiation at microsatellite loci in whole-genome scans of recent and incipient mosquito species. *Proceedings of the National Academy of Sciences of the United States of America,* 98 (19), 10769-10774.

Weill, M., Chandre, F., Brengues, C., et al., 2000. The kdr mutation occurs in the Mopti form of *Anopheles gambiae s.s.* through introgression. *Insect Molecular Biology,* 9 (5), 451-455.

Wekesa, J.W., Brogdon, W.G., Hawley, W.A., et al., 1998. Flight tone of field-collected populations of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae). *Physiological Entomology,* 23 (3), 289-294.

Whitlock, M.C. and McCauley, D.E., 1999. Indirect measures of gene flow and migration: F-ST not equal 1/(4Nm+1). *Heredity,* 82 Part 2, 117-125.

Wondji, C., Simard, F. and Fontenille, D., 2002. Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. *Insect Molecular Biology,* 11 (1), 11-19.

Wright, S., 1931. Evolution in Mendelian populations. *Genetics,* 16, 97-159.

Yan, G., Severson, D.W. and Christensen, B.M., 1997. Costs and benefits of mosquito refractoriness to malaria parasites: Implications for genetic variability of mosquitoes and genetic control of malaria. *Evolution,* 51 (2), 441-450.