

Population genetics of *An. gambiae* chromosomal/molecular forms in Mali.

Gregory C. Lanzaro[†], Fredrick Tripet[†] and Guimogo Dolo*

[†] Department of Entomology, University of California, Davis, CA 95616 USA

* Département d' Entomologie, Ecole Nationale de Médecine et de Pharmacie, Bamako B.P. 1805, MALI

In areas where several sub-specific 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* sub-populations is critical for assessing the feasibility of a mosquito genetic control program, developing adequate mass-rearing facilities and designing initial field trials.

Population structuring within *An. gambiae* sensu stricto has classically been approached using cytogenetic methods to describe chromosomal inversion polymorphism. 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. Moreover, the ongoing selective sweep associated with *kdr* resistance (Chandre *et al.* 1999; Weill *et al.* 2000) may provide 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.

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.* (1994; 1998a). 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.* 1998a; 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 *et al.* 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 and it now appears that these may be caused by polymorphic or floating inversions typical of one form, but occurring rarely in others.

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/III* 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 and 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 and among those early samples there were a few equivocal cases where karyotyping did not match the molecular diagnostic (Favia *et al.* 1997; Della Torre 2001). The diagnostic was also used to identify between-form hybrid-like karyotypes. M/S hybrids produced in the laboratory did yield 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.

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 3rd and X-chromosome whilst strong linkage disequilibrium and low levels of genetic differentiation were found for loci located on the 2nd chromosome (Lanzaro *et al.* 1998). Another study using microsatellites distributed on all three chromosomes was conducted by Wang *et al.* 2001. They 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 *et al.* (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).

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 form 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.* (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 *et al.* 2000; della Torre *et al.* 2001). The *kdr* mutation has since been identified in the M-form from Burkina Faso (Diabate *et al.* 2003). 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, 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 do occur in the wild, then one might reasonably expect that selection acts against hybrid-like genotypes. This hypothesis may be tested by comparing M*S hybridization rates at different developmental stages and attempt to detect a reduction in hybrid survival. Although a formal study of this kind has not yet been conducted, some interesting data does 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 was 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 the 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 adequately test this hypothesis.

If strong assortative mating does occur 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 behavioral 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 wingbeat frequency. Although recordings of flight tones from laboratory colonies of *An. arabiensis* and *An. gambiae* s.s. seemed to support the wing-beat hypothesis (Brogdon 1998), field data showed significantly different but largely overlapping distributions of wingbeat frequencies (Wekesa *et al.* 1998). Tripet *et al.* (In press) used F₁'s from field collected mosquitoes reared in the laboratory and measured under controlled conditions and found no difference in wingbeat frequencies between sympatric populations of *An. arabiensis*, and the Mopti and Savanna forms of *An. gambiae* s.s.. The chemical or behavioral cues and mechanisms used by mosquitoes for mate recognition remain unknown.

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 is currently no firm data supporting this hypothesis.

Literature Cited.

- BESANSKY, N.J., LEHMANN, T., FAHEY, G.T., FONTENILLE, D., BRAACK, L.E., HAWLEY, W.A. and COLLINS, F.H., 1997. Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An. arabiensis*, suggest extensive gene flow. *Genetics* 147, 1817-1828.
- BROGDON, W.C., 1998. Measurement of flight tonedifferentiates among members of the *Anopheles gambiae* species complex (Diptera: Culicidae). *Journal of Medical Entomology* 35, 681-684.
- BUTLIN, R.G., 1995. Reinforcement: An idea evolving. *Trends in Ecology and Evolution* 10, 432-434.
- CHANDRE, F., MANGUIN, S., BRENGUES, C., YOVO, J. D., DARRIET, F., DIABATE, A., CARNEVALE, P. and GUILLET, P., 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, 319-322.
- CIANCHI, R., VILLANI, F., TOURÉ, Y. T., PETRARCA, V. and BULLINI, L., 1983. Electrophoretic study of different chromosomal forms of *Anopheles gambiae* s.s. *Parassitologia* 25, 239-241.
- COLUZZI, M., SABATINI, A., DELLA TORRE, A., DI DECO, M.A. and PETRARCA, V., 2002. A polytene chromosome analysis of the *Anopheles gambiae* Complex. *Science* 298, 1415-1418.

- DELLA TORRE, A., FANELLO, C., AKOGBETO, M., DOSSOU-YOVO, J., FAVIA, G., PETRARCA, V. and COLUZZI, M., 2001. Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Molecular Biology* 10, 9-18.
- DELLA TORRE, A., COSTANTINI, C., BESANSKY, N.J., CACCONI, A., PETRARCA, V., POWELL, J.R. and COLUZZI, M., 2002. Speciation within *Anopheles gambiae* the glass is half full. *Science* 298, 115-117.
- DIABATE, A., BALDET, T., CHANDRE, F., DABIRE, K.R., KENGNE, P., GUIGUEMDE, T.R., SIMARD, F., GUILLET, P., HEMINGWAY, J. and HOUGARD, J., 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, 195-198.
- DI DECO, M.A., PETRARCA, V., VILLANI, F. and M. COLUZZI, 1980. Polimorfismo cromosomico da inversioni paracentriche ed eccesso degli eterocriotipi in ceppi di *Anopheles* allevati in laboratorio. *Parassitologia* 22, 304-306.
- EDILLO, F.E., TOURE', Y.T., LANZARO, G.C., DOLO, G. and TAYLOR, C.E., 2002. Spatial and habitat distribution of *Anopheles gambiae* and *Anopheles arabiensis* (Diptera: Culicidae) in Banambani Village, Mali. *Journal of Medical Entomology* 39, 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., BAGYAOKO, M., LANFRANCOTTI, A., SAGNON, N'F., TOURÉ, Y.T. and COLUZZI, M., 1997. Molecular identifications of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Molecular Biology* 6, 377-383
- FAVIA, G., LANFRANCOTTI, A., SPANOS, L., SIDÉN-KIAMOS, I. and LOUIS, C., 2001. Molecular characterization of ribosomal DNA (rDNA) polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. *Insect Molecular Biology* 10, 19-23.
- GENTILE, G., SLOTMAN, M., KETMAIER, V., POWELL, J.R. and CACCONI, A., 2001. Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae* s.s. *Insect Molecular Biology* 10: 25-32.
- HARTL, D.L. 1980. *Principles of population Genetics*. Sinauer associates, Inc., Sunderland, Massachusetts, USA.
- LIU, L.W. and PRICE, T.D., 1994. Speciation by reinforcement of premating isolation. *Evolution* 48: 1451-1459.
- LANZARO, G.C., TOURÉ, Y.T., CARNAHAN, J., ZHENG, L., DOLO, G.T., TRAORÉ, S.F., PETRARCA, V., VERNICK, K.D. and Taylor, C.E., 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 Science USA* 95, 14260-14265.
- MILLIGAN, P.J.M., PHILLIPS, A., BROOMFIELD, G., and MOLYNEUX, D.H., 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., CORNEL, A.J., DOTSON, E.M., COLLINS, F.H. and BESANSKY, N.J., 1996. The tryptophan oxygenase gene of *Anopheles gambiae*. *Insect Biochemichal and Molecular Biology* 26, 525-528.
- MUKABAYIRE, O., CARIDI, J., WANG, X., TOURÉ, Y.T., COLUZZI, M., and BESANSKY, N.J., 2001. Patterns of DNA sequence variation in chromosomally recognized taxa of *Anopheles gambiae*: Evidence from rDNA and single copy loci. *Insect Molecular Biology* 10, 33-46.
- ONYABE, D.Y. and CONN, J. 2001. Genetic differentiation of the malaria vector *Anopheles gambiae* across Nigeria suggests that selection limits gene flow. *Heredity* 87, 647-658.
- PERSIANI A., DIDECO M.A. and PETRANGELI G., 1986. Osservazioni di laboratorio su polimorfismi da inversione originati da incroci tra popolazioni diverse di *Anopheles gambiae* s.s.. *Annali dell 'Istituto Superiore di Sanita* 22, 221-224.
- TOURÉ, Y.T., PETRARCA, V., TRAORÉ, S.F., COULIBALY, A., MAIGA, H.M., SANKARÉ, S.F., SOW, M., DI DECO, M.A. and COLUZZI, M., 1994. Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s. str. in Mali, West Africa. *Genetica* 94, 213-223.

- TOURÉ, Y.T., PETRARCA, V., TRAORÉ, S.F., COULIBALY, A., MAIGA, H.M., SANKARÉ, S.F., SOW, M., DI DECO, M. A. and COLUZZI, M., 1998. The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* 40, 477-511.
- TRIPET, F., TOURÉ, Y.T., TAYLOR, C.E., NORRIS, D.E., DOLO, G. and LANZARO, G.C., 2001. DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of the *Anopheles gambiae* complex. *Molecular Ecology* 10, 1725-1732.
- TRIPET, F., TOURÉ, Y. DOLO, G. and LANZARO, G.C. 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, 1-5.
- TRIPET, F., TRAORÉ, S. DOLO, G. and LANZARO, G.C. 2003. The 'wing-beat' hypothesis of reproductive isolation between members of the *Anopheles gambiae* complex (Diptera: Culicidae) does not fly. *Journal of Medical Entomology*. In press.
- WANG, R., ZHENG, L., TOURÉ, Y., DANDEKAR, T. and KAFATOS, F., 2001. When genetic distance matters: measuring genetic differentiation at microsatellite loci whole genome scans of recent and incipient species. *Proceedings of the National Academy of Science USA* 98, 10769-10774.
- WEILL, M., CHANDRE, F., BRENGUES, C., MANGUIN, S., AKOGBETO, M., PASTEUR, N., GUILLET, P., and RAYMOND, M., 2000. The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae* s.s. through introgression. *Insect Molecular Biology* 9, 451-455
- WEKESA, J.W., BRODGON, W.G., HAWLEY, W.A. and BESANSKY, N.J., 1998. Flight-tone of field populations of *Anopheles gambiae* and *An. arabiensis* (Diptera: Culicidae). *Physiological Entomology* 23, 289-294.
- 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, 11-19.