Population genetics of genes within chromosomal inversions in the *Anopheles gambiae* complex.

A. W. Dickerman, D. Fontenille*, W. A. Hawley†, and M. G. Kidwell.

Department of Ecology and
Evolutionary Biology
University of Arizona
Tucson, Arizona, 85721

*ORSTOM
BP 1386
Dakar, Senegal

†Malaria Branch
Division of Parasitic Diseases
Centers for Disease Control
Nairobi, Kenya

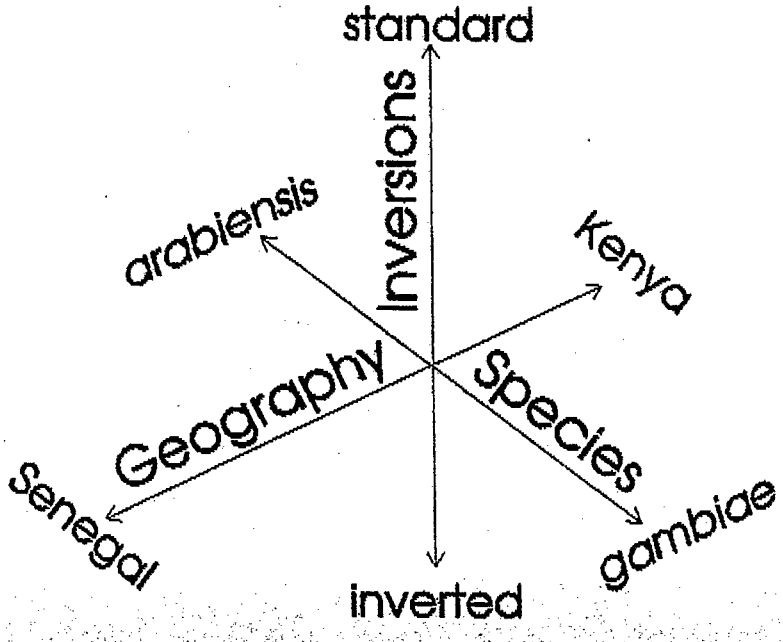
Poster presented at:
Network on the Biology of Parasite Vectors: 1995 Scientific Institute

Inverness Hotel,
Denver, Colorado
June 15-18, 1995

Introduction

We are interested in the phenomenon of polymorphic chromosomal inversions as a potential force structuring variation in natural populations of mosquitoes in the *Anopheles gambiae* complex. Paracentric chromosomal inversions have been studied extensively in the *A. gambiae* complex and are shown to correlate with ecological gradients (Coluzzi, et al., 1978) and even epidemiological factors (Petrarca and Beier, 1992). These studies strongly imply that inversions are maintained by selection in nature and play an active role in evolution at the population level. Inversions can act a barriers to recombination between alleles on chromosomal segments with opposite orientations, giving rise to strongly differentiated haplotypes within a panmictic population. The mechanism for this is inviablility of gametic products of single cross-over events in heterokaryotypes. Mechanisms which could oppose this isolation include double crossing-over events and gene conversion, which would allow recombination between alleles in opposite inversion types.

It remains an open question how strongly genetic variation is structured along lines of inversion karyotypes relative to other ecological and genetic factors. In our investigation of this question we are comparing the effect of inversions to geographic separation across the width of Africa, comparing samples from Senegal to ones from Kenya, and reproductive isolation (i.e., the nominal species categories *gambiae* vs. *arabiensis*) as the major competing forces structuring population diversity. Thus, we are investigating three dimensions of population structure, as depicted in this diagram.

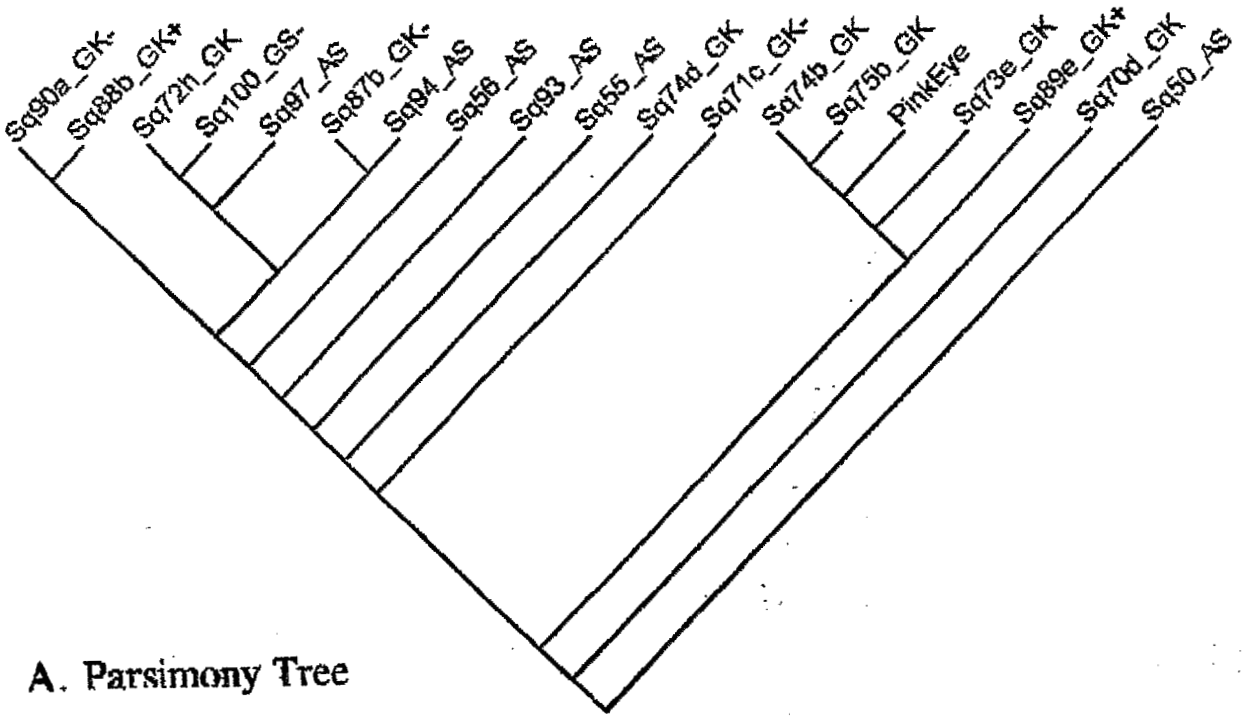


Methods

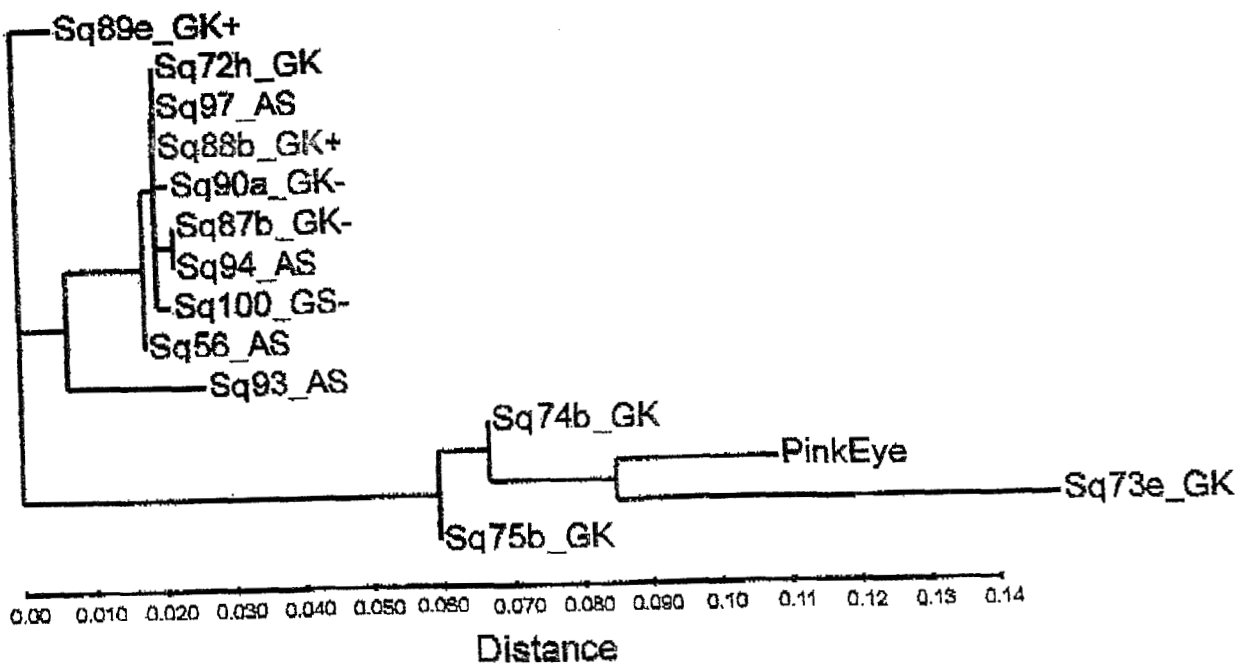
Our samples include mosquitoes collected in Senegal at two localities about 200 km apart, one of which is a study site for hyperendemic malaria (Trape et al, 1994). The Kenyan collections are all from within 30 km of Kisumu, also including a study site for hyperendemic malaria (Githeko et al., 1992). We are sequencing DNA at several genetic loci of known location on the physical map (data provided by Dr. Frank Collins). We have designed sets of nested primers to allow amplification from minute samples using nested PCR. This allows the same mosquitoes to be used for other assays, such as karyotyping, sporozoite ELISA, blood meal ELISA, and other DNA analyses (e.g., with micro-satellite markers). Karyotype analysis is being performed by Ousmane Faye, University of Senegal, Dakar, and by Odette Mukabayire, Malaria Branch, Division of Parasitic Diseases, CDC, Atlanta, both of whom have received training from Dr. Mario Coluzzi.

PCR is performed using *Pfu* DNA polymerase to reduce mis-incorporations. Sequencing is performed by the University of Arizona Macromolecular Structure Facility on an Applied Biosystems model 373a automated sequencer using fluorescent dye chain terminators. PCR products were either sequenced directly or after cloning into pCR-Script (Stratagene). Cloning was required for heterozygotes for major differences, especially insertion-deletion events. So far we have generated data for three variable loci from two different inversions described in Table 1.

Our conceptual approach to analysis of these data is to interpret sequence variants as a pattern of superimposed mutations, or synapomorphies. Estimating population-genetic parameters such as effective population size and gene flow from such data requires two conditions: 1) that one can accurately interpret the pattern of synapomorphies, and 2) that the sequences under study are not under unusual selective forces (purifying or balancing). Our first approximation to estimating the history of mutational changes is to use phylogenetic trees for each locus which are presented here as Figures 1-3. Because the nature of selection at our loci is impossible to know objectively, we restrict ourselves to comparisons within loci, attributing differences in genetic diversity between loci to selection. In the future will apply analyses which accommodate recombination within loci as well as among them.

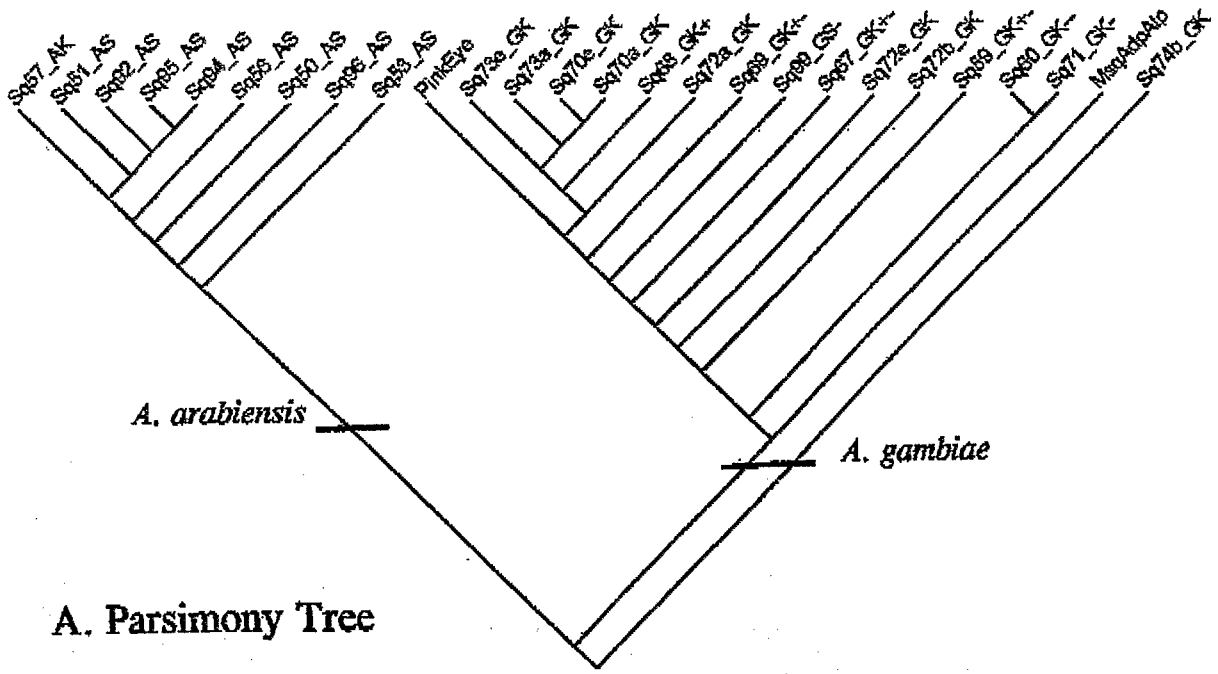


A. Parsimony Tree

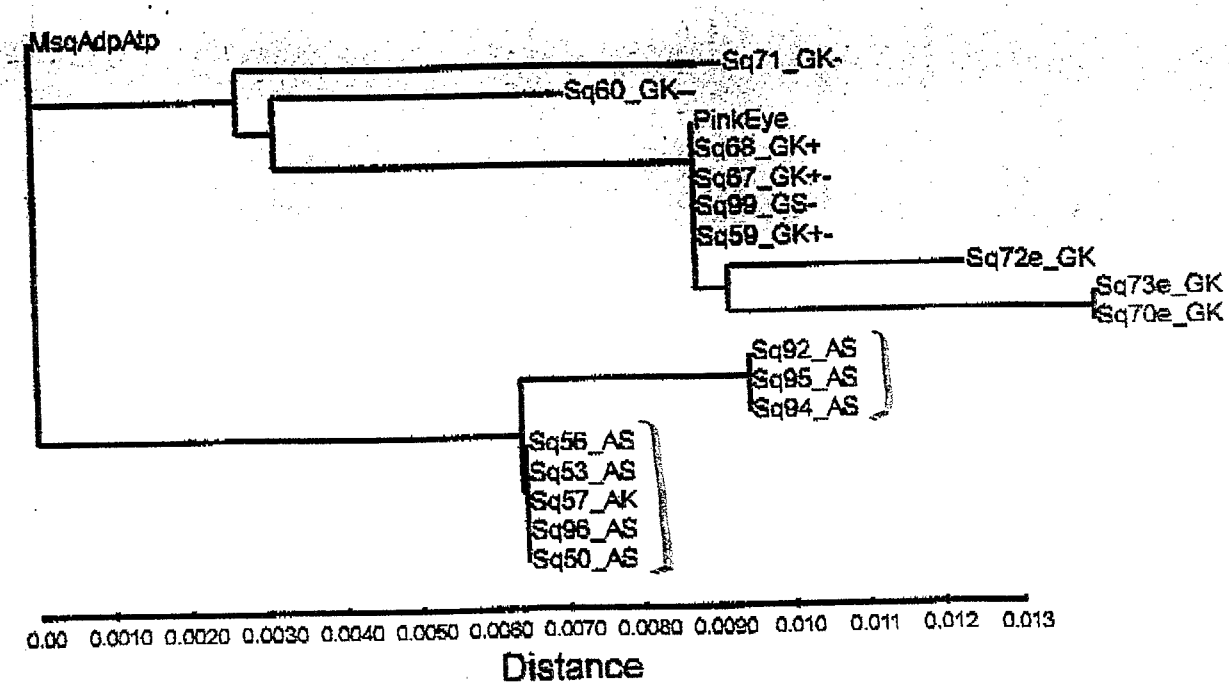


B. Least-squares Distance Tree

Figure 1. Phylogenetic trees of sequence variants at cDNA76 (Transcription Initiation Factor) in inversion 2La. We present both maximum parsimony (A.) and distance analyses (B.). Names refer to individual mosquitoes or clones with indications of species, locality, and inversion type: a "G" indicting *Anopheles gambiae* and "A" indicating *A. arabiensis*; "S" means Senegal, "K" means Kenya. "+" means 2La-Standard, "-" means 2La-Inverted. There is a strong tendency for mixing of all categories on this tree, suggesting either gene flow between species, with little effect of geography or inversion type, or that variants at this locus predate the speciation and inversion events.

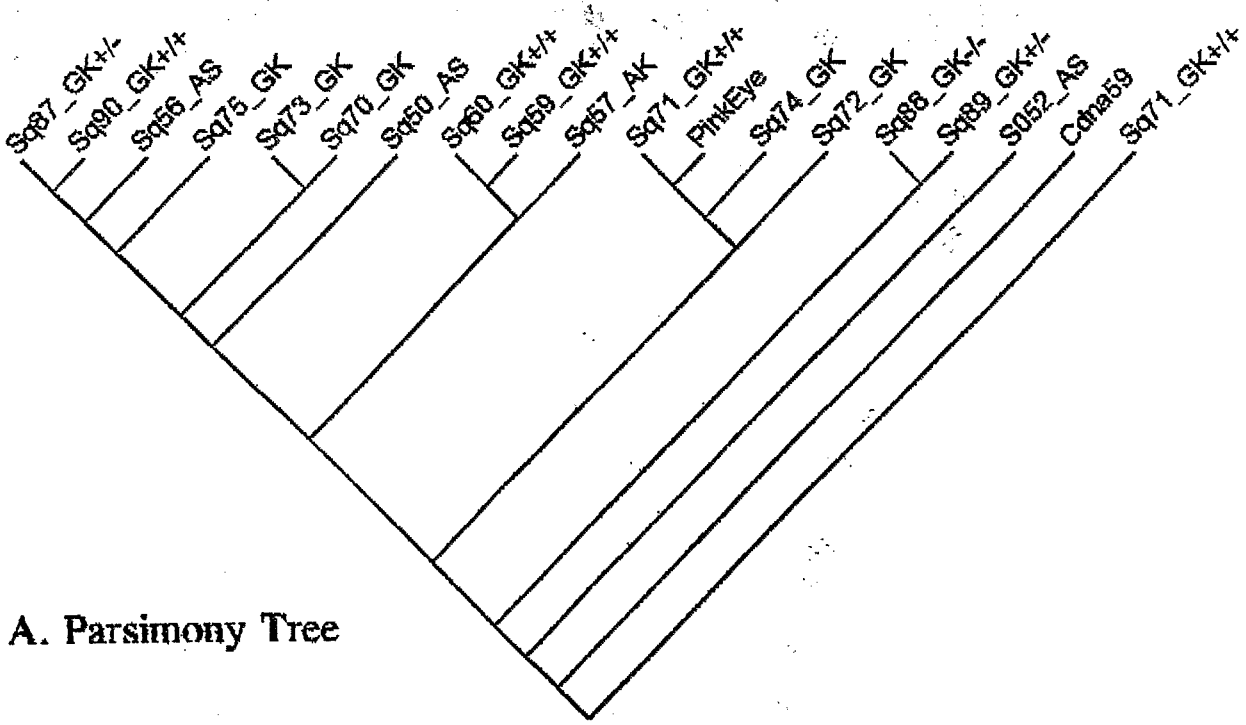


A. Parsimony Tree

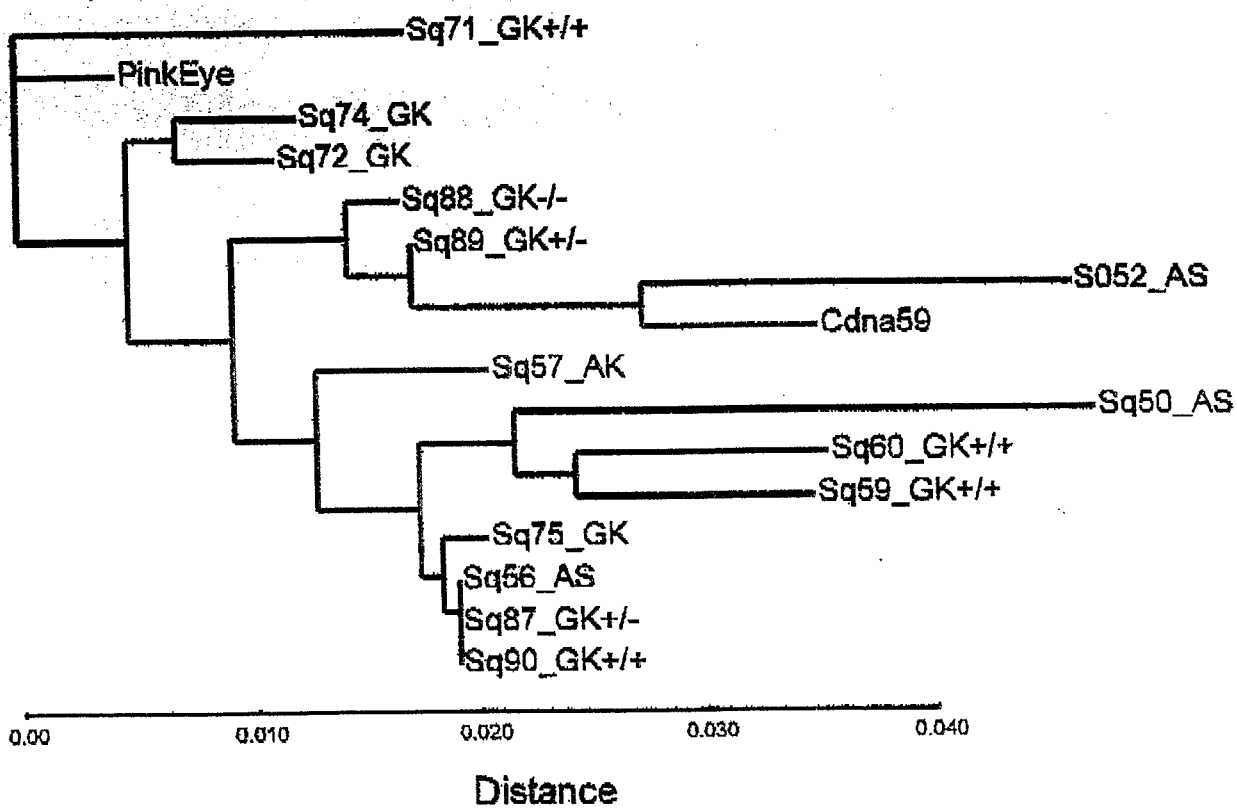


B. Least-squares Distance Tree

Figure 2. Phylogenetic trees of sequence variants at Adp/Atp translocase in inversion 2La. At this locus, both maximum parsimony (A.) and distance analysis (B.) suggest a strong differentiation between *Anopheles gambiae* and *A. arabiensis*. There is an apparent association among all Standard inversion types, with Inverted (including *arabiensis*) being more diverse, suggesting that Inverted is the primitive condition and that inversions do structure variation. Separation by geography is not strongly suggested, although there are still too few samples from Senegal to judge. Abbreviations as in Fig 1



A. Parsimony Tree



B. Least-squares Distance Tree

Figure 3. Phylogenetic trees of sequence variants at cDNA59 in inversion 2Rb. Abbreviations as in Fig. 1 except that "+" and"-" refer to orientations of inversion 2Rb. Here the tendency for mixing between categories is even more pronounced than at locus cDNA76, though as our observations on inversion 2Rb are sparser than on 2La, and more sequences need to be added before any firm conclusions can be made.

Preliminary Interpretations

Although we are in the middle of data-collection, with many individuals of known karyotype still to add to our trees, we believe there are some interpretations which can be derived from the data in hand. The questions framed at the outset of this work seem to have the following tentative answers, subject to change as more data become available:

1) Is variation of DNA sequences in regions affected by inversions structured strongly by inversion orientation? The answer seems to be "yes and no." At the Adp/Atp locus, sequences known to represent the Standard orientation of 2La form a cluster (also including at least one Inverted) to the exclusion of other Inverteds and all of *A. arabiensis* (which is fixed for Inverted at this inversion). This is consistent with Inverted being the primitive condition and the inversion generating Standard being a recent event, with diversity among Standard sequences arising subsequent from that time. If further data bear this up, it could signal a strong effect of inversions on genetic structure, particularly at loci near the inversion break-points, as is the case for Adp/Atp.

2) Is variation structured strongly by classification into the species *gambiae* and *arabiensis*? The tentative answer is "yes and no." At one of three loci, Adp/Atp translocase, is there a strong separation of the two species, while the other two have the two species co-mingling on the various branches. Does this constitute evidence of gene flow between species? This pattern could be explained by variation at the latter two loci may be retained from times before the species separated. The branch-lengths on the distance trees do not support this in that *gambiae* and *arabiensis* are joined by very short tree paths in several cases. We tentatively hypothesize that these loci represent genetic introgression at nuclear loci, a pattern that was not seen by Besansky et al. (1994) at the esterase locus.

3) Is variation structured by geographic separation across the width of Africa? We still do not have enough of the necessary sequences in our data set to answer this question. The few sequences we do have show no strong separation by locality.

Overall, the impression we derive from these data is that [illegible] comprise genetic variation which is only partly structured along the lines one would expect. We suspect the reasons for this include both retention of ancient polymorphisms and recent gene flow.

Literature Cited

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

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

Githeko, A. K., A. D. Bradling-Bennet, M. Beier, F. Atieli, M. Owaga, and F. H. Collins. 1992. The reservoir of *Plasmodium falciparum* malaria in a holoendemic area of western Kenya. Trans. R. Soc. Trop. Med. Hyg., 86: 355-358.

Petrarca, V. and J. Beier. 1992. Intraspecific chromosomal polymoprhism in the *Anopheles gambiae* comples as a factor affecting malaria transmission in the Kisumi aera of Kenya. Am. J. Trop. Med. Hyg. 46: 229-237.

Trape, J.-F., C. Rogier, L. Konate, N. Diagne, H. Bouganali, B. Canque, F. Legros, A. Badhi, G. Ndiaye, P. Ndiaye, K. Brahimi, O. Faye, P. Druilhe, and L. P. Da Silva. 1994. The Dielmo Project: A longitudinal study of natural malaria infection and the mechanisms of protective immunity in a community lining in a holoendemic area of Senegal. Am. J. Trop. Med. Hyg. 51: 123-137.