Multiple Origins of Cytologically Identical Chromosome Inversions in the *Anopheles gambiae* Complex

Adalgisa Caccone,*,†,[1] Gi-Sik Min*,[1] and Jeffrey R. Powell*

**Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520-8106 and †Dipartimento di Biologia, II Università di Roma "Tor Vergata," Rome, Italy*

Manuscript received March 15, 1998
Accepted for publication July 13, 1998

ABSTRACT

For more than 60 years, evolutionary cytogeneticists have been using naturally occurring chromosomal inversions to infer phylogenetic histories, especially in insects with polytene chromosomes. The validity of this method is predicated on the assumption that inversions arise only once in the history of a lineage, so that sharing a particular inversion implies shared common ancestry. This assumption of monophyly has been generally validated by independent data. We present the first clear evidence that naturally occurring inversions, identical at the level of light microscopic examination of polytene chromosomes, may not always be monophyletic. The evidence comes from DNA sequence analyses of regions within or very near the breakpoints of an inversion called the $2L^a$ that is found in the *Anopheles gambiae* complex. Two species, *A. merus* and *A. arabiensis*, which are fixed for the "same" inversion, do not cluster with each other in a phylogenetic analysis of the DNA sequences within the $2L^a$. Rather, *A. merus* $2L^a$ is most closely related to strains of *A. gambiae* homozygous for the $2L^+$. *A. gambiae* and *A. merus* are sister taxa, the immediate ancestor was evidently homozygous $2L^+$, and *A. merus* became fixed for an inversion cytologically identical to that in *A. arabiensis*. *A. gambiae* is polymorphic for $2L^a/2L^+$, and the $2L^a$ in this species is nearly identical at the DNA level to that in *A. arabiensis*, consistent with the growing evidence that introgression has or is occurring between these two most important vectors of malaria in the world. The parallel evolution of the "same" inversion may be promoted by the presence of selectively important genes within the breakpoints.

DEDUCING phylogenetic relationships based on chromosomal inversions stems from the classic work of Sturtevant and Dobzhansky (1936) and is a well-accepted methodology that has been applied to a number of species, especially insects with polytene chromosomes. The method relies on the reasonable assumption that inversions are monophyletic in origin; *i.e.*, they occurred once in a single chromosome, and all present-day carriers of the inversion trace their ancestry to this single event. Thus, carriers of the same inversion (gene arrangement) must have a shared common ancestor. Independent phylogenetic data (*e.g.*, morphological and molecular) have been generally congruent with inversion-deduced phylogenies, thus corroborating the validity of the methodology.

Another test of the monophyly of inversions has been to study gene sequences that lie within the breakpoints. This approach is predicated on the assumption that in addition to being monophyletic, inversions effectively suppress recombination between and immediately adjacent to the breakpoints. Thus, the allele captured by the single inversion event will have a phylogenetic history identical to the inversion itself. When this has been attempted, in all cases until now, the gene tree deduced from DNA sequences within breakpoints has been congruent with the inversion tree (Aquadro *et al.* 1991; Popadic and Anderson 1994; Rozas and Aguadé 1994; García *et al.* 1996). Furthermore, in the only case where the actual breakpoints of an inversion have been sequenced from a number of independent copies of the inversion, all have identical breakpoints down to the precise nucleotide (Wesley and Eanes 1994), an extremely unlikely observation in the absence of monophyly.

Here, we provide the first evidence that cytologically identical, naturally occurring inversions may not always be monophyletic. The evidence comes from a mosquito group, the *Anopheles gambiae* complex. This complex consists of six presently described species that are all native to sub-Saharan Africa. Many naturally occurring inversions exist in the group, both as floating polymorphisms and fixed differences among species (Coluzzi *et al.* 1979, 1985). Based on the reasoning described above, Coluzzi and colleagues (1979, 1985) produced a phylogenetic network, a modified version of which is in Figure 1a. There are two clear fixed synapomorphic inversions: the X^{ag}, indicating *Anopheles gambiae* and *A. merus* are sister taxa, and the $3L^a$, indicating that *A. melas* and *A. bwambae* are sister taxa. DNA sequence data of

Corresponding author: Jeffrey R. Powell, Department of Ecology and Evolutionary Biology, 165 Prospect St., Yale University, New Haven, CT 06520-8106. E-mail: jeffrey.powell@yale.edu

[1] These authors contributed equally to this work.

TABLE 1

Details of the genomic regions and numbers of strains sequenced

DNA region[a]	2L div[a]	Acc. no.[b]	Length (bp)[c]			No. of strains sequenced				
			Total	Coding	Noncoding	gam	ara	mer	mel	quad
pkm122	22F′	U51225	2239	2076	163	3	1	2	2	1
pkm2	23″	U50477	727	641	86	3	2	2	2	2
pkm79	23C′	U50478	711	—	—	2	2	2	2	2
pkm129	26B	U50476	816	—	—	2	1	2	1	1

[a] Polytene band designations following della Torre *et al.* (1996).
[b] GenBank database accession numbers of the original cDNA clones.
[c] Lengths are from alignments and include gaps. "Noncoding" includes three putative introns for pkm122 and pkm2. pkm79 and pkm129 are anonymous DNA regions that do not have any open reading frames.

a gene within the X^{ag} confirm its monophyly and, thus, supports the phylogenetic affinity of *A. gambiae* and *A. merus* (García *et al.* 1996). Phylogenetic analysis of mitochondrial DNA sequences confirm the sister status of *A. melas* and *A. bwambae* (Caccone *et al.* 1996), consistent with the monophyly of the $3L^a$. Although there is no inversion indicating the sister status of *A. quadriannulatus* and *A. arabiensis*, accumulating DNA sequence information is converging on this conclusion (our unpublished data).

If the network in Figure 1a accurately reflects the phylogenetic history of these species, it is extremely difficult to explain the phylogenetic distribution of the $2L^a$ inversion. This inversion is fixed in *A. merus* and *A. arabiensis* and is polymorphic along with the $2L^+$ arrangement in *A. gambiae*; the $2L^+$ arrangement is fixed in all other species. To maintain the hypothesis of monophyly of naturally occurring inversions, two hypotheses can be erected. First, the $2L^a/2L^+$ polymorphism may be an ancestral state with the alternative arrangements becoming fixed in the different species while remaining polymorphic in one species. Second, the $2L^a$ may have arisen in one species and may have been transferred via introgression among the species, a phenomenon thought to occur at least between *A. gambiae* and *A. arabiensis* (Besansky *et al.* 1994; Caccone *et al.* 1996; García *et al.* 1996). Hybrids formed among all six species consist of fertile females and sterile males (Davidson *et al.* 1967). Alternatively, the $2L^a$ may not be monophyletic, and what appears to be the identical inversion has arisen independently in different species. We set out to test the monophyly of the $2L^a$ by sequencing four DNA fragments known to lie within or extremely close to the breakpoints of the $2L^a$. The results are inconsistent with the monophyly of this inversion.

MATERIALS AND METHODS

Mosquito strains: The number of strains used in this study are listed in Table 1. Species names are abbreviated as follows: gam, *gambiae*; ara, *arabiensis*; mer, *merus*; mel, *melas*; and quad, *quadriannulatus*. Strain names, with their acronyms used in other publications (Besansky *et al.* 1994; Caccone *et al.* 1996; García *et al.* 1996) and geographic origins in parentheses, are as follows: ara1 (ARZAG, Burkina Faso), ara2 (AJ, Madagascar), ara3 (AB3, Eritrea), mel1 (BAL, Gambia), mel2 (BRE, Gambia), mer1 (V12, Kenya), mer2 (ZULU, Zululand), quad1 (CHIL, Zimbabwe), and quad2 (SQUAD). For the $2L^a$ polymorphic species, *A. gambiae*, we studied one strain homozygous for $2L^+$ (abbreviated gam+/+) and one homozygous for $2L^a$ (gama/a1) derived from a laboratory population from West Africa. We also studied an individual from a recently collected strain (AG, Agbabilame, Benin) that was homozygous for $2L^a$ (gama/a2). All *A. gambiae* strains were provided by Mario Coluzzi and Alessandra della Torre; all others were provided by Nora Besansky and Frank Collins (University of Notre Dame). One to three individuals per strain were sequenced.

DNA sequencing: Total DNA was extracted from frozen or Carnoy preserved specimens using a modified Drosophila extraction protocol (Livak 1984) or the Easy-DNA extraction kit (Invitrogen, San Diego, CA). We chose four DNA fragments from a set of random cDNA clones from *A. gambiae* that had been mapped by hybridization to microdissected divisional probes (Mathiopoulos and Lanzaro 1995) and by *in situ* hybridization to polytene chromosomes (della Torre *et al.* 1996). The polytene chromosome band locations for the four regions are listed in Table 1 and are illustrated in Figure 1b. Table 1 also reports the GenBank accession numbers of the cDNA sequences and the lengths of the fragments studied, with specifics on the occurrence of putative introns and open reading frames. Using the sequence information available from each cDNA clone, we designed several sets of external and internal primers to PCR amplify and sequence the corresponding nuclear DNA in the five species of the *A. gambiae* complex. PCR amplification conditions varied depending on the strain and on the DNA fragment studied. They are available from the authors together with the exact sequence and location of the external and internal primers used. PCR fragments were gel purified by gelase treatment (Epicentre Technologies, Madison, WI) or by glass milk extraction (GeneClean III kit; BIO 101, Vista, CA). The amplified double-stranded DNA fragments were either directly sequenced or cloned using the TA-Cloning kit (Invitrogen). Plasmids were extracted using the Plasmid kit (QIAGEN, Chatsworth, CA). Both direct sequencing of PCR products and sequencing of cloned products were performed on an ABI 373A automated sequencer (Applied Biosystems, Foster City, CA); both strands were sequenced. When using cloned products, to minimize amplification errors, multiple clones were sequenced for each strain.

Data analysis: Sequences from multiple individuals or multiple clones from the same strain were combined into a single

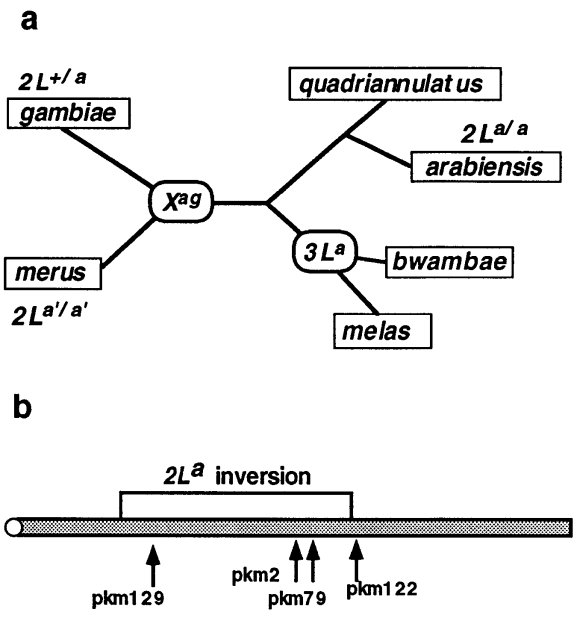

Figure 1.—(a) Inversion phylogeny of the *Anopheles gambiae* complex species, showing the only two inversions (X^{ag} and $3L^{a}$) that are synapomorphies linking sister taxa (at the nodes). The distribution of the $2L^{a}$ inversion is shown; species with no designations are fixed for $2L^{+}$. For *A. merus*, we have designated the inversion $2L^{a'}$ as we present evidence that it is not the same as $2L^{a}$ in the other two species. (b) Schematic map of the left arm of chromosome *2*. The positions of the DNA regions sequenced are noted below; the circle on the left indicates the centromere. pkm122 is just outside the inversion.

consensus sequence for each strain, using polymorphic designations for sites variable among individuals or clones. Sequences were aligned by eye and by using CLUSTAL W (Thompson *et al.* 1994); alignments are available from the authors. Phylogenetic analyses were carried out on each data set and on the combined data set. Because not all strains were sequenced for all four DNA regions, a single consensus sequence for each species was used for the combined analyses, coding variable sites as polymorphic, with the exception of the *A. gambiae* strains with different 2L arrangements. Phylogenetic trees were reconstructed by maximum parsimony (MP, Farris 1970) and neighbor joining (NJ, Saitou and Nei 1987) using PAUP* (version 4.0.0d54; kindly provided by D. Swofford); maximum likelihood (ML, Felsenstein 1981) was performed on the PHYLIP 3.57c program package (Felsenstein 1995). Because of the heterogeneous nature of the DNA regions studied, we compared non-gamma and gamma distances of Tamura and Nei (1993) with the PAUP* program (default gamma parameter $a = 0.5$). Gamma distances allow for heterogeneity of rates among sites. Both methods produced nearly identical distances and did not affect the results; results from using only the non-gamma distances are presented. In the ML analyses, the options "global rearrangement," "empirical base frequency," "jumble" with 10 replicates, and "2.0 transition/transversion ratio" were used. The MP method was carried out using the branch-and-bound option with equal weighting of all nucleotide substitutions. Accelerated transformation (ACCTRAN) was always used for character state optimization. For the two regions (pkm122 and pkm2) with open reading frames and putative introns, MP analyses were performed on coding and noncoding regions separately and on third codon positions only. The robustness of the phylogenetic hypotheses was tested by bootstrapping (Felsenstein 1985), using 1000 pseudoreplicates for MP and NJ. For ML, we used 100 pseudoreplicates for the individual data sets and 500 for the combined analysis. For the combined data set, we evaluated competing phylogenetic hypotheses using two statistical tests. For the MP analysis, we used Templeton's (1983) nonparametric test, using the conservative two-tailed Wilcoxon rank sum test (Larson 1994). For the ML analysis, we used the log likelihood test (Kishino and Hasegawa 1989). The sequence data presented in this article have been submitted to GenBank under accession numbers AF020849–AF020878.

RESULTS

We sequenced a total of 4493 bp for four DNA regions between or very close to the breakpoints of the $2L^{a}$ inversion (Figure 1b) from multiple individuals belonging to several strains of five species of the *A. gambiae* complex. Table 2 summarizes the levels of sequence variation and the potential phylogenetic information of variable sites. A total of 211 variable sites occur across all strains. The variability of the four regions is quite similar, ranging from 3.9 to 5.4% of all sites. For the two regions with open reading frames (pkm122 and pkm2), the majority of variation is in the introns, where ~20% of sites vary. Only ~3% of the sites in the coding regions vary, with silent substitutions being most common. Although all four regions were obtained by using sequence information from cDNA clones, both pkm79 and pkm129 include multiple stop codons in all reading frames in all the species studied and in the original *A. gambiae* clone, precluding any further partitioning of variability among different functional regions. (We suspect that the lack of open reading frames in what were thought to be cDNA clones results from some kind of cloning artifact; the *in situ* mapping was done with the identical clones we used, so there is no doubt about the chromosomal location.) Intraspecific variability was very low, with a maximum of 4 variable sites between conspecific strains for any region. Variable sites occurred mostly as fixed differences between species and between different karyotypes ($2L^{+/+}$ *vs.* $2L^{a/a}$). Of the 169 potentially informative sites, 33 synapomorphies are shared between *A. arabiensis* and *A. gambiae* $2L^{a/a}$ strains, while only one synapomorphy groups *A. merus* with the other $2L^{a/a}$ strains, *i.e.*, makes the $2L^{a}$ monophyletic (Table 2).

Figures 2 and 3 show the results of the phylogenetic analyses as majority rule MP consensus trees after 1000 bootstrap replicates. In Figure 2, the four DNA regions were analyzed separately, combining individuals from the same strains but keeping strains distinct. Figure 3 is the combined analysis with a single consensus sequence for each species except *A. gambiae*. Trees with the same topologies were obtained by MP, NJ, and ML analyses. Further details on the MP analyses are provided in Table 3.

Conspecific strains of all the species (except *A. gambiae*) always cluster together when each DNA fragment is

TABLE 2

Summary of sequence variation and phylogenetic importance of polymorphic sites

Region	Total var[a]	Coding regions				Noncoding regions		
		Var	S/R[b]	Aut.[c]	Syn.[d]	Var	Aut.	Syn.
pkm122	106 (4.7)	75 (3.6)	56/19	7	68 (20, 1)	31 (19.0)	4	27 (4, 0)
pkm2	33 (4.5)	13 (2.0)	13/0	2	11 (4, 0)	20 (23.3)	8	12 (1, 0)
pkm79	28 (3.9)	—	—	—	—	—	4	24 (1, 0)
pkm129	44 (5.4)	—	—	—	—	—	17	27 (3, 0)

[a] Number of variable sites (with percentage in parentheses).
[b] Number of silent (S) *vs.* replacement (R) sites is listed.
[c] Number of autoapomorphic sites.
[d] Number of potentially synapomorphic sites. The two numbers in parentheses are the number of synapomorphic changes that support the *A. gambiae* $2L^{a/a}$-*A. arabiensis* and *A. gambiae* $2L^{a/a}$-*A. arabiensis*-*A. merus* clades, respectively.

analyzed separately, implying that intraspecific variation does not cloud interspecific or interinversion differences (Figure 2). In three out of four DNA regions (pkm122, pkm129, and pkm2), *A. gambiae* individuals homozygous for $2L^{+}$ or the $2L^{a}$ cluster on distant branches of the trees. However, not all the other strains

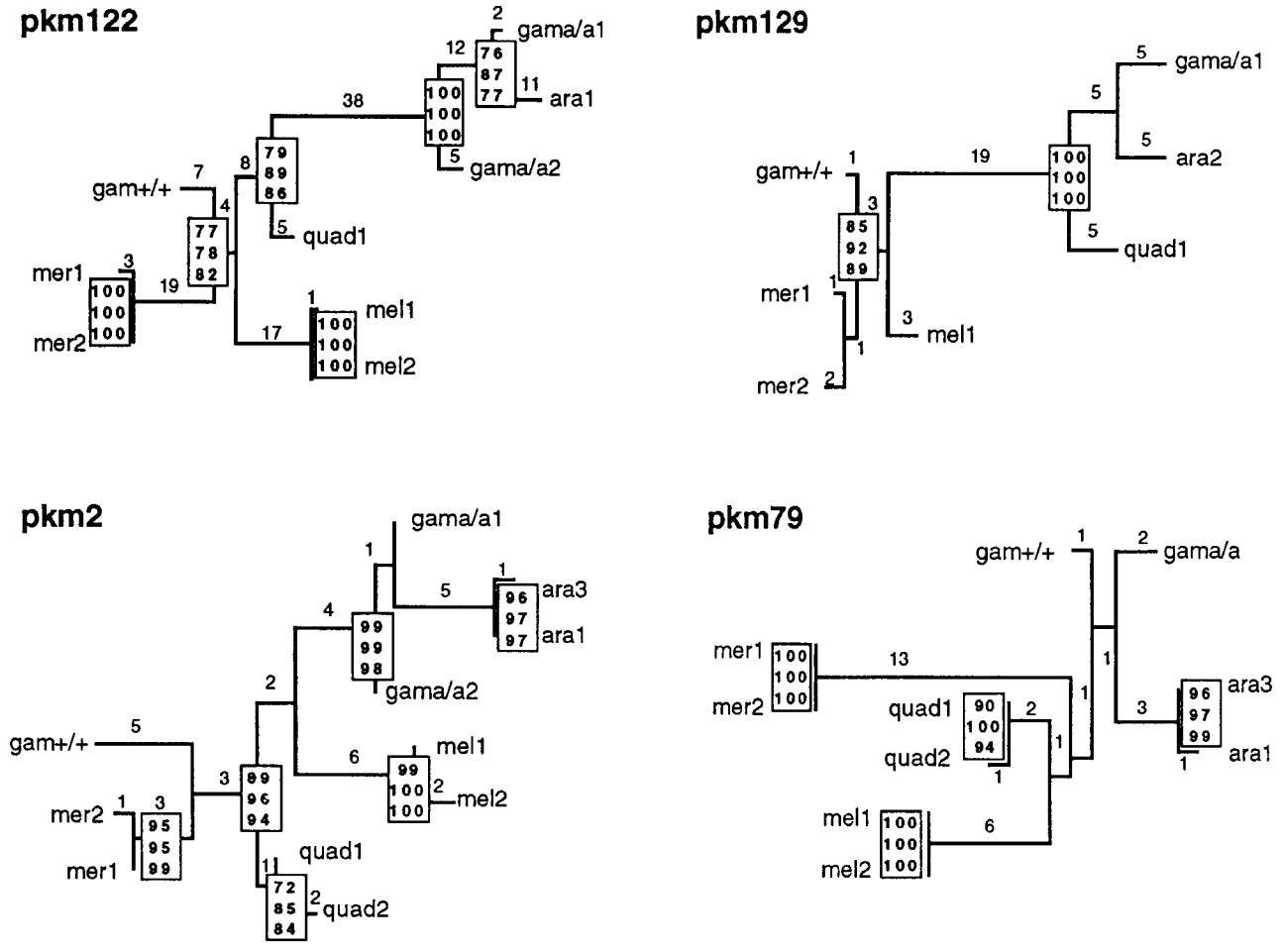

Figure 2.—Majority rule bootstrap consensus trees based on DNA sequences within or close to the $2L^{a}$ inversion in five species of the *A. gambiae* complex. Abbreviations of species names and strains are in materials and methods. Bootstrap values are percentages of 1000 pseudoreplicates for MP and NJ, and 100 for ML (top, middle, and bottom values in the boxes, respectively). Bootstrap values are shown only for nodes for which all three phylogenetic methods had bootstrap support of ≥70%. Numbers on branches are the number of steps separating each node in the MP tree.

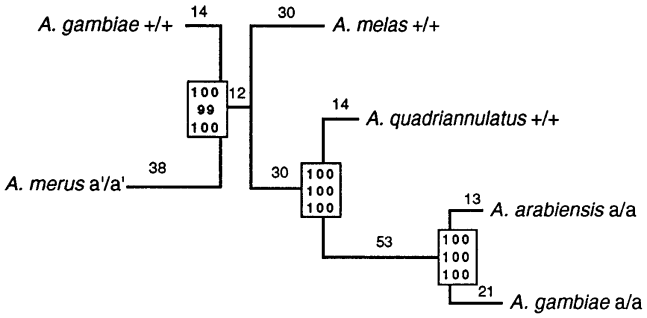

Figure 3.—Majority rule bootstrap consensus tree based on the combined data set (pkm122, pkm2, pkm79, and pkm-129) of DNA sequences within or close to the $2L^a$ inversion in five species of the *A. gambiae* complex. Multiple strains, individuals, or clones sequenced for each species were combined into a single sequence coded for polymorphisms; for *A. gambiae*, the data for $2L^{a/a}$ and $2L^{+/+}$ strains were kept separate. Numbers on branches are the number of steps separating each node in the MP tree. Bootstrap values in boxes are as in Figure 2, except that 500 pseudoreplicates of ML were performed. Symbols after species name indicate the gene arrangement of the 2L.

carrying what is cytologically the same inversion karyotype ($2L^{a/a}$) cluster together. *A. merus* clearly does not cluster with the other $2L^{a/a}$ strains, but it clusters with *A. gambiae* $2L^{+/+}$. This topology is confirmed by the number of steps separating the MP tree from the shortest tree having the $2L^a$ inversion monophyletic (Table 3). The topology of the MP tree for pkm79 is poorly supported (low bootstrap values, short branch lengths, and only one step separating the MP tree from a tree with the taxa carrying the $2L^a$ inversion in a single clade). However, though weak, the pkm79 tree is still not consistent with the monophyly of $2L^a$ because the *A. merus* strains do not cluster with the *A. arabiensis* and *A. gambiae* a/a clade. The clustering among the other taxa is less uniform across the four DNA regions. Two fragments (pkm122 and pkm129) support clustering *A. quadriannulatus* with the *A. arabiensis*- *A. gambiae* a/a clade, while pkm2 favors *A. melas* as sister taxa of the same clade.

The topology of the tree obtained by combining all data confirms the results obtained from the single-region trees (Figure 3). The *arabiensis-gambiae* $2L^{a/a}$ and *merus-gambiae* $2L^{+/+}$ clades are very strongly supported with bootstrap percentages of 99 or 100 for all phylogenetic methods, as well as the relatively long branch lengths that define each clade: 53 for the *arabiensis-gambiae* $2L^{a/a}$ clade and 12 for the *merus-gambiae* $2L^{+/+}$ clade. Both MP and ML trees were statistically significantly better than the shortest trees with *A. merus* in the same clade as *A. arabiensis* and *A. gambiae* a/a, as determined by the Templeton (1983) test on the MP trees ($n = 31$; $T_s = 80$; $P < 0.001$) and the Kishino and Hasegawa (1989) test on the ML trees (difference ln likelihood = −96.32637, SD = 26.9713, $P < 0.001$).

The anomalous phylogenetic position of *A. merus*, which does not cluster with the other $2L^a$ carriers but is the sister taxon of *A. gambiae* $2L^{+/+}$, was also confirmed by genetic distance comparisons (Table 4). The average genetic distances between strains with the same $2L^a$ arrangement (excluding *A. merus*) is 0.014 (SD = 0.004) for the combined data set. The average distance of *A. merus* to gam+/+ is 0.011, whereas the average distance of *A. merus* to the other $2L^{a/a}$ strains is 0.027 (SD = 0.004), which is comparable to the average distance (excluding *A. merus*) between $2L^{+/+}$ and $2L^{a/a}$ strains ($D = 0.024$, SD = 0.004).

TABLE 3

Details of maximum parsimony analyses

DNA region	No. inf. sites[a]	No. trees[b]	Tree length	CI[c]	$2L^{a/a}$ mono[d]
pkm122	95	1	135	0.82	8
Exons	68	1	85	0.92	4
Introns	27	3	43	0.76	10
pkm2	23	4	36	0.92	3
Exons	11	2	14	0.92	0
Introns	12	3	22	0.93	3
pkm79	24	1	32	0.96	1
pkm129	27	1	50	0.84	17
Combined	82	1	260	0.82	21*

[a] Number of informative sites.
[b] Number of equally parsimonious trees.
[c] Consistency index excluding uninformative sites.
[d] Number of additional steps from the MP tree to the shortest tree having the $2L^{a/a}$ inversion monophyletic.
* Significance at $P < 0.001$ for the Templeton (1983) test on the MP tree and for the Kishino and Hasegawa (1989) test on the ML tree.

DISCUSSION

These results are incompatible with the hypothesis of monophyly of the $2L^a$ inversion in this species group. The evidence that the $2L^a$ fixed in *A. merus* and *A. arabiensis* are the "same" inversion is based on the usually accepted criteria: Under light microscopic examination, the banding patterns in polytene chromosomes are identical (Coluzzi and Sabatini 1969). Furthermore, in F_1 hybrids between these species, the second chromosome pairs perfectly (Coluzzi and Sabatini 1969). However, based on our results, the $2L^a$ in *A. merus* and *A. arabiensis* very likely arose independently. It remains to be determined whether, at the nucleotide level, the two are truly identical (perhaps because of a "hot spot" for chromosome breakage), or just that they are sufficiently similar to be indistinguishable at the level of cytological examination and pairing of homologs.

An alternative to polyphyly would be that the ancestral lineage to the whole group was polymorphic for $2L^{a/+}$ and remained polymorphic in *A. gambiae* while becoming fixed in the other species. If so, it is still difficult to

TABLE 4

Average genetic divergences within and between strains fixed for the $2L^a$ or the $2L^+$ inversions

DNA region	Average genetic distances[a]					
	Within[b] a/a	Within[c] +/+	a/a[d] *vs.* +/+	mer[e] *vs.* *a/a*	mer[f] *vs.* +/+	mer[g] *vs.* gam+/+
pkm122	0.009 ± 0.003	0.010 ± 0.005	0.027 ± 0.004	0.031 ± 0.003	0.016 ± 0.003	0.012
pkm2	0.005 ± 0.003	0.013 ± 0.006	0.019 ± 0.005	0.021 ± 0.004	0.014 ± 0.003	0.012
pkm79	0.006 ± 0.004	0.009 ± 0.005	0.013 ± 0.004	0.022 ± 0.003	0.023 ± 0.002	0.019
pkm129	0.013	0.025 ± 0.014	0.030 ± 0.009	0.036 ± 0.004	0.023 ± 0.015	0.004
Combined	0.014 ± 0.006	0.014 ± 0.002	0.024 ± 0.004	0.027 ± 0.004	0.016 ± 0.004	0.011

[a] Average distances based on uncorrected *P* values with associated standard deviations.
[b,c] *D* values within individuals homozygous for the $2L^a$ (within a/a, excluding *A. merus*) and the $2L^+$ (within +/+) inversion.
[d] *D* values between $2L^{a/a}$ and $2L^{+/+}$ individuals.
[e] *D* values of *A. merus* from $2L^{a/a}$ carriers (*A. gambiae* $2L^{a/a}$ and *A. arabiensis*).
[f] *D* values of *A. merus from all* $2L^{+/+}$ carriers.
[g] *D* values of *A. merus* from *A. gambiae* $2L^{+/+}$.

rectify the close affinity of the *A. merus* $2L^a$ to the *A. gambiae* $2L^+$ but not to *A. arabiensis* $2L^a$. If monophyletic, all copies of $2L^a$ should coalesce into a common ancestor before they coalesce with any $2L^+$. Moreover, if the ancestral *gambiae*/*merus* lineage was polymorphic for $2L^{a/+}$, the copies of $2L^a$ in *A. gambiae* and *A. merus* should be more similar to one another than either is to $2L^a$ in *A. arabiensis*, a prediction that is clearly at odds with our data. One could hypothesize that in the ancestral *gambiae*/*merus* lineage, selection acted to bring about convergence of DNA sequences within $2L^a$ and $2L^+$. However, virtually all our analyses are based on noncoding DNA or synonymous substitutions in coding regions. These are the kinds of substitutions thought to be least subject to selection.

Gene conversion is another alternative; *i.e.*, there was gene conversion in the lineage common to *A. gambiae* and *A. merus* such that the $2L^a$ acquired sequences from the $2L^+$. Gene conversion between inversions has been documented (Rozas and Aguadé 1994; Popodic *et al.* 1995). However, in these studies, ≤10% of an inversion has been found to be converted in any single chromosome. Given that all four regions we studied showed a very similar pattern, this would seem to be an unlikely explanation.

What then is the likely history of the $2L^a$ distribution in this group? The most parsimonious scenario is that the lineage common to *A. merus* and *A. gambiae* had the $2L^+$ arrangement. After these species split, *A. merus* independently generated a second chromosome inversion indistinguishable from the *A. arabiensis* $2L^a$, and subsequently, it became fixed in this species. There is increasing evidence that *A. gambiae* and *A. arabiensis* are or recently were undergoing gene exchange, most likely through introgressive hybridization (Besansky *et al.* 1994; Caccone *et al.* 1996; García *et al.* 1996); fertile female hybrids between these species have been found in nature (Coluzzi *et al.* 1979; Touré *et al.* 1998). The *A. arabiensis* $2L^a$ likely passed into *A. gambiae* via this route and remains polymorphic in that species. Several years ago, Coluzzi and colleagues (1985) had already proposed the introgression of $2L^a$ from *A. arabiensis* to *A. gambiae* based solely on biogeographic and ecological considerations. Furthermore, in laboratory populations, it has been shown that the *A. arabiensis* $2L^a$ persists in a freely interbreeding hybrid population formed by backcross of fertile F_1 females to *A. gambiae* males from a strain homozygous for $2L^+$ (della Torre *et al.* 1997). In fact, judging from its increase in frequency and excess of heterozygotes over Hardy-Weinberg expectations in these laboratory populations, this introgressed inversion may actually be forming a stable heterotic polymorphism.

Although it has generally been argued that the arrangements designated + in the *gambiae* complex are the ancestral state (Coluzzi *et al.* 1979, 1985), we cannot formally rule out the possibility that $2L^a$ is the ancestral state and $2L^+$ is the derived state. This would affect our historical interpretation in the previous paragraph, but it would not affect our interpretation of the multiple origin of cytologically identical gene arrangements. Our data would then be interpreted as a second origin of $2L^+$ in *A. gambiae* compared to this arrangement in all other species.

Cases of different inversions sharing one breakpoint are not uncommon, including among these species of mosquitoes (Coluzzi *et al.* 1979), although no clear cases exist where independent inversions share both breakpoints. In Hawaiian Drosophila, cases of parallel evolution of inversions have been observed (Carson 1969) where very similar but not quite identical inversions arose independently for the same region of a chromosome. Carson (1969) attributed this to selective retention of inversions in that particular region caused by the presence of genes that "have some unspecified tendency toward a high fitness when in the heterozygous

state." The $2L^a$ in the *gambiae* complex has been implicated as being selectively important in adaptation to xeric environments and is associated with different behavior patterns, including blood meal preference (Coluzzi *et al.* 1985). The independent origin of the "same" inversion requires both breaks at cytologically indistinguishable locations and the selective advantage of rearrangement (Krimbas and Powell 1992). Thus, chromosomal regions containing blocks of genes that are potentially coadapted in certain combinations would be more prone to independently become fixed or polymorphic for inversions with similar or identical breakpoints.

In addition to the basic interest in the origin and dynamics of inversion polymorphisms in insects, the history of inversions and the issue of introgression have important public health implications for this group of mosquitoes. Members of the *gambiae* complex account for the majority of transmission of malaria in Africa, a continent with >85% of worldwide cases of malaria (Strüchler 1989). Carriers of different gene arrangements vary greatly in their ability to breed in drier or moister habitats, in their propensity to bite humans *vs.* animals, etc. (Coluzzi *et al.* 1977, 1979, 1985). The $2L^a$ in particular has been implicated as directly relevant to malaria transmission; in polymorphic populations of *A. gambiae*, different *2L* karyotypes vary in their prevalence of Plasmodium infection (Petrarca and Beier 1992). Furthermore, in laboratory studies, genetic factors affecting susceptibility/refractoriness to transmission of Plasmodium have been mapped to the $2L^a$ (Vernick *et al.* 1989).

Is multiple origin of cytologically indistinguishable inversions common? Given the general consistency of the hypothesis of monophyly with a variety of data (discussed in the Introduction), it would seem that this is not common. That occasionally an inversion that is cytologically indistinguishable from a preexisting inversion can be generated and become established in a species is not totally surprising, although this is the first documentation of such an event. It does not invalidate the general assumption of monophyly that is the basis for phylogeny reconstruction. However, this example does serve as a caution that the assumption is not universally true.

We thank Etsuko Moriyama for assistance in analyses and comments about the manuscript; Mario Coluzzi, Alessandra della Torre, Nora Besansky, and Frank Collins for mosquitoes; and Kostas Mathiopoulos for the cDNA clones and sequence information before publication. David Swofford kindly made available PAUP* before its formal release. This work was supported by grants from the National Institutes of Health (to J.R.P.) and the European Community, EU INCO project no. ERB3514PL961817 (to A.C.).

LITERATURE CITED

Aquadro, C. F., A. L. Weaver, S. W. Schaeffer and W. W. Anderson, 1991 Molecular evolution of inversions in *Drosophila pseudoobscura:* the amylase gene region. Proc. Natl. Acad. Sci. USA **88:** 305–309.

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

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

Carson, H. L., 1969 Parallel polymorphisms in different species of Hawaiian Drosophila. Am. Nat. **113:** 323–330.

Coluzzi, M., and A. Sabatini, 1969 Cytogenetic observations on the salt water species, *Anopheles merus* and *Anopheles melas* of the *gambiae* complex. Parassitologia **11:** 177–187.

Coluzzi, M., A. Sabatini, V. Petrarca and M. A. Di Deco, 1977 Behavioral divergences between mosquitoes with different inversion karyotypes in polymorphic populations of the *Anopheles gambiae* complex. Nature **266:** 832–833.

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

Coluzzi, M., V. Petrarca and M. A. Di Deco, 1985 Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae.* Boll. Zool. **52:** 45–63.

Davidson, G., H. E. Paterson, M. Coluzzi, G. F. Mason and D. W. Micks, 1967 The *Anopheles gambiae* complex, pp. 211–250 in *Genetics of Insect Vectors of Disease*, edited by J. W. Wright and R. Pal. Elsevier, Amsterdam.

della Torre, A., G. Favia, G. Mariotti, M. Coluzzi and K. D. Mathiopoulos, 1996 First physical map of the malaria vector *Anopheles gambiae* reveals nonrandom distribution of coding regions. Genetics **143:** 1307–1311.

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

Farris, J. S., 1970 Methods for computing Wagner trees. Syst. Zool. **34:** 21–34.

Felsenstein, J., 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:** 368–378.

Felsenstein, J., 1985 Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783–791.

Felsenstein, J., 1995 PHYLIP, version 3.572c, University of Washington, Seattle.

García, B. A., A. Caccone, K. D. Mathiopoulos and J. R. Powell, 1996 Inversion monophyly in African Anopheline malaria vectors. Genetics **143:** 1313–1320.

Kishino, H., and M. Hasegawa, 1989 Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order of Hominoidea. J. Mol. Evol. **29:** 170–179.

Krimbas, C. B., and J. R. Powell (Editors), 1992 *Drosophila Inversion Polymorphism.* CRC Press, Boca Raton, FL.

Larson, A., 1994 The comparison of morphological and molecular data in phylogenetic systematics, pp. 371–390 in *Molecular Ecology and Evolution: Approaches and Applications*, edited by B. Schierwater, B. Streit, G. P. Wagner and R. DeSalle. Birkhäuser, Basel.

Livak, K. J., 1984 Organization and mapping of a sequence on the *Drosophila melanogaster X* and *Y* chromosomes that is transcribed during spermatogenesis. Genetics **107:** 611–634.

Mathiopoulos, K. D., and G. C. Lanzaro, 1995 Distribution of genetic diversity in relation to chromosomal inversions in the malaria mospuito *Anopheles gambiae.* J. Mol. Evol. **40:** 578–584.

Petrarca, V., and J. C. Beier, 1992 Intraspecific chromosomal polymorphism in the *Anopheles gambiae* complex as a factor affecting malaria transmission in the Kisumu area of Kenya. Am. J. Trop. Med. Hyg. **46:** 229–237.

Popadic, A., and W. W. Anderson, 1994 The history of a genetic system. Proc. Natl. Acad. Sci. USA **91:** 6819–6823.

Popadic, A., D. Popadic and W. W. Anderson, 1995 Interchromosomal transfer of genetic information between gene arrangements on the third chromosome of *Drosophila pseudoobscura.* Mol. Biol. Evol. **12:** 938–943.

Rozas, J., and M. Aguadé, 1994 Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila.* Proc. Natl. Acad. Sci. USA **91:** 11517–11521.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Stürchler, D., 1989 How much malaria is there worldwide? Parasitol. Today **5:** 39–40.

Sturtevant, A. H., and Th. Dobzhansky, 1936 Inversions in the third chromosome of wild races of *Drosophila pseudoobscura* and their use in the study of the history of the species. Proc. Natl. Acad. Sci. USA **22:** 448–452.

Tamura, F., and M. Nei, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10:** 512–526.

Templeton, A. R., 1983 Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the humans and apes. Evolution **37:** 221–244.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Touré, Y. T., V. Petrarca, S. F. Traoré, A. Coulibaly, H. M. Maiga *et al.*, 1998 Distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. Parassitologia (in press).

Vernick, K. D., F. H. Collins and R. W. Gwadz, 1989 A general system of resistance to malaria infection in *Anopheles gambiae* controlled by two main genetic loci. Am. J. Trop. Med. Hyg. **40:** 585–592.

Wesley, C. S., and W. F. Eanes, 1994 Isolation and analysis of the breakpoint sequences of chromosome inversion In(3L) Payne in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **91:** 3132–3136.

Communicating editor: W. F. Eanes