Physical Map of the Malaria Vector *Anopheles gambiae*

A. della Torre, G. Favia, G. Mariotti, M. Coluzzi and K. D. Mathiopoulos

Istituto di Parassitologia, Fondazione "Pasteur Cenci-Bolognetti," Università di Roma "La Sapienza", Rome 00185, Italy

Manuscript received December 18, 1995
Accepted for publication March 16, 1996

ABSTRACT

Random cDNA clones, cosmid clones and RAPD polymorphic fragments have been localized by *in situ* hybridization to the ovarian nurse cell polytene chromosomes of the malaria vector *Anopheles gambiae*. We thus established 85 molecular markers for 110 sites within the whole *A. gambiae* polytene chromosome complement. The cDNA clones analyzed were isolated at random, and their exact localizations were determined by *in situ* hybridization. For 15 of the cDNA clones, a partial nucleotide sequence has been obtained; for nine of them sequence searches in the GenBank database revealed high degrees of similarity with published sequences. The cosmid clones analyzed were obtained as the result of screening with a few of the aforementioned cDNA clones of particular interest, or taken from a small set of randomly isolated cosmid clones. The RAPD clones are polymorphic fragments, potentially diagnostic for the various chromosomal forms of *A. gambiae* that are currently being analyzed.

THE *Anopheles gambiae* complex is constituted of six sibling species, *A. gambiae, A. arabiensis, A. quadriannulatus, A. merus, A. melas* and *A. bwambae*. The first two species represent the most efficient malaria vectors known and are annually responsible for the infection of several million individuals in subsaharan Africa and the death of over a million people, mostly young children (STUERCHEL 1989). In recent decades, the genetic information that has been accumulated on the *A. gambiae* complex is mainly related to insecticide resistance (DAVIDSON 1985) and polytene chromosome studies, including those based on the ovarian nurse cells of half-gravid females (COLUZZI *et al.* 1979). Recently developed molecular techniques can now be coupled with the cytogenetic knowledge and offer a powerful way to obtain information on genomic organization and gene function in this organism. Progress in this respect was achieved with the development of a low-resolution molecular map based on microdissected divisions and subdivisions of the *A. gambiae* chromosomes (ZHENG *et al.* 1991). Furthermore, the use of microsatellite markers and their integration with the banding pattern have led to development of a linkage map for the *A. gambiae X* chromosome (ZHENG *et al.* 1993). Recent information has been added to generate a more complete genetic linkage map based on microsatellites (ZHENG *et al.* 1996) and RAPD markers (DIMOPOULOS *et al.* 1996).

Alternatively, there has also been an effort to construct a physical map by defining the precise chromosomal locations of isolated nucleic acid sequences by *in situ* hybridization to polytene chromosomes. Here we present the first extensive effort to map by *in situ* hybridization a large number of either randomly selected cDNA clones or cosmid clones or cloned RAPD markers, thus establishing 85 molecular markers for 110 sites on the five chromosomal arms characterizing the *A. gambiae* polytene complement. It is also the first report on cosmid markers within the *A. gambiae* genome. Given the relatively small size of this genome (2.6×10^8 bp; BESANSKY and POWELL 1992), these clones taken together constitute a set of markers separated from each other by 3000 kb, on average. The resulting molecular map not only serves as a versatile tool for interrelating molecular, genetic and cytogenetic information, but also opens the possibility to define homologies among chromosomes or discrete chromosomal segments across wide phylogenetic distances. In a similar approach, a set of cloned DNA fragments are also being localized by *in situ* hybridization in the laboratory of Dr. F. COLLINS (A. CORNEL, V. KUMAR and F. COLLINS, unpublished data).

Corresponding author: K. D. Mathiopoulos, Istituto di Parassitologia, Università di Roma "La Sapienza", Piazzale Aldo Moro 5, Roma 00195, Italy. E-mail: kostas@axrma.uniroma1.it

MATERIALS AND METHODS

Mosquito strains: The *A. gambiae* strain used for molecular cloning and for cytogenetic preparations was the homokaryotypic GASUA reference strain (*Xag, 2R, 2La, 3R, 3L*) selected in our insectary in 1989 from a *2Rd-2La* polymorphic colony (*Xag, 2Rd/+, 2La/+, 3R, 3L*) that in turn originated from adult females collected in Liberia in 1986. Other *A. gambiae* strains used for RAPDs-PCR experiments were the following: GASEL-M (*Xag, 2Rbc/u, 2La, 3L, 3R*), a MOPTI strain established in 1987 from adult females collected in Selinkenyi (Mali); GAMOR-M (*Xag, 2Rbc/+, 2La, 3R, 3L*), a MOPTI strain established in 1991 from adult females collected in Moribabougou (Mali); GAMOR-B (*Xag, 2Rjcu,b/+, 2La, 3R, 3L*), a BAMAKO strain established in 1991 from adult females collected in Moribabougou (Mali).

DNA clones: The DNA clones that have been analyzed by *in situ* hybridization were obtained as follows.

cDNA clones: The *A. gambiae* cDNA library construction and the isolation of the cDNA clones was performed as described in MATHIOPOULOS and LANZARO (1995). Briefly, the library

was constructed from RNA from adult GASUA mosquitoes (mixed males and unfed females) in vector pcDNAII (Invitrogen Co.). Several clones were isolated and digested with *Hin*dIII and *Xba*I to release the cloned insert; digests were fractionated by gel electrophoresis and transferred on nylon filters. The filters were hybridized with each of all the available "divisional probes". Divisional probes (microamplified DNA pools) have been generated by microdissection and PCR amplification of 54 divisions and subdivisions of the five arms of *A. gambiae* polytene chromosomes (*X, 2R, 2La, 3R, 3L*) (ZHENG *et al.* 1991). Partial nucleotide sequences for 15 of these clones were obtained and homology searches using the GenBank database have revealed high degrees of similarity ($P < 0.001$) for nine of them (Table 1).

Cosmid clones: Fourteen clones were randomly isolated from a GASUA cosmid library, while 13 more were isolated by screening the same library with selected cDNA clones (K. D. MATHIOPOULOS, unpublished results). Emphasis was placed on the *X* and second chromosomes, as they contain the largest numbers of inversions that are associated with speciation of the six sibling species and incipient speciation of the chromosomal forms in *A. gambiae* (COLUZZI *et al.* 1985). Fourteen other clones were randomly collected from the same cosmid library. Clones cKM13, 42, 52, and 122 were isolated by screening the library with cDNA clones pKM13, 42, 52, and 122, respectively, which mapped close to the inversion breakpoints proximal *2Rd* (of *A. arabiensis*), distal *Xc*, proximal *2Ru* and proximal *2La*, respectively.

RAPD clones: Over 60 commercially available decanucleotide primers (from Operon Technologies, Alameda, CA) as well as a few more custom-made primers (indicated in the Table 1 legend) were used to generate RAPD fingerprints that could yield fragments potentially diagnostic for the three *A. gambiae* chromosomal forms (SAVANNA, MOPTI and BAMAKO) as described in FAVIA *et al.* (1994). RAPD fragments identified as being specific for one of these three *A. gambiae* chromosomal forms were gel-purified after amplification and were either labeled directly, or first subcloned and then labeled for *in situ* hybridization.

***In situ* hybridization:** The polytene chromosome slides were prepared from half-gravid females according to the procedure described by KUMAR and COLLINS (1994) and stored dry. To stably fix the chromosomes on the slides, the slides were first incubated at room temperature in 4% paraformaldehyde in phosphate-buffered saline (PBS) for 2 min, then rinsed in PBS, and finally dehydrated by submersion for 3 min each in 50, 70, 90 and 100% ethanol.

Probes were labeled with the Gibco-BRL Bionick labeling kit according to the instructions supplied by the manufacturer with the following modifications: half of the suggested amount of dNTP and enzyme mix was used without affecting the intensity of the hybridization signal. After ethanol precipitation, probes were resuspended in 400 μl of 2× hybridization buffer (4× SSC, 0.2 M sodium phosphate pH 6.5, 2× Denhardt's solution, 0.1 mg/ml Na azide), and 35 μl of probe were diluted with an equal volume of 20% dextran sulphate (dissolved in water) and hybridized to the chromosomes under a sealed coverslip in a humid chamber overnight at 56°. After hybridization, slides were washed in 0.2× SSC for 20 min once at 56° and once at room temperature. Slides were subsequently blocked with 250 μl blocking solution [50 mg/ml bovine serum albumin (BSA) in 100 mM Tris-HCl, 150 mM NaCl, 0.2 mg/ml Na azide] for 20 min at 43°. Signal detection was achieved as follows: (1) slides were covered with 100 μl of 4 mg/ml Streptavidin-alkaline phosphatase (Dako A/S, Denmark) in conjugated dilution buffer (100 mM Tris-HCl, 150 mM $MgCl_2$, 10 mg/ml BSA, 0.2 mg/ml Na azide) for 20 min at 43°, (2) slides were washed in Tris-buffered saline (100 mM Tris-base, 150 mM NaCl) for 15 min once at 43° and once at room temperature, (3) slides were incubated in alkaline substrate buffer (ASB; 100 mM Tris-base, pH 9.5, 150 mM NaCl, 50 mM $MgCl_2$) for 10 min at room temperature, (4) slides were incubated in NBT/BCIP solution in ASB (5 ml per slide) at 37° until the desired level of signal was achieved (5–60 min). The slides were then washed with distilled water and kept dry. No reduction of signal intensity was observed for as long as at least 1 year after hybridization. The chromosomes were observed by adding a drop of distilled water and a coverslip. Cytogenetic examination was conducted using phase contrast optics with a Leitz Aristoplan compound microscope with a 63× objective. Sites of hybridization were localized to specific bands within numbered and lettered subdivisions of the polytene chromosome map (M. COLUZZI, A. SABATINI, M. A. DI DECO, V. PETRARCA, unpublished data) revised from that of COLUZZI and SABATINI (1967). These sites were determined from a minimum of four nuclei per slide. Representative examples of *in situ* hybridization are presented in Figure 1.

RESULTS

Table 1 shows a list of all the cDNA, cosmid and RAPD clones analyzed with respective sites of hybridization on the *A. gambiae* chromosomes. The exact localization of each signal was carefully determined within a subdivision and can be provided upon request along with the revised *A. gambiae* cytogenetic map. No variation in signal localization was ever detected among all the nuclei examined for a given clone.

cDNA clones: Twenty-nine clones were randomly isolated from an *A. gambiae* (GASUA) adult cDNA library and mapped on four of the five arms of the chromosomes, by the use of divisional probes as described in ZHENG *et al.* (1991). No mapping data were obtained from the fifth arm (*3R*) since no divisional probes were available. Among these clones, 22 were mapped in one division only, four were mapped in two divisions, two were mapped in three divisions and the remaining one in four divisions (*i.e.*, the 29 cDNA clones were mapped in 40 different locations, potentially representing single- as well as multiple-copy genes). Hybridization signals varied in intensity; nine of them were classified as strong, whereas the remaining 31 were classified as weak (data not shown).

Although determining the map position of a certain clone by the use of divisional probes is quick and easy, it only remains indicative of location for two main reasons: first, because it does not provide the exact location and, second, because it is difficult to differentiate between nonspecific hybridization and a weak signal that may be due to uneven amplification during preparation of the divisional probes. As the microdissected DNA is first restricted with *Sau*3A, then ligated with adaptors and subsequently PCR-amplified, one would expect substantial variations in the relative quantities of the PCR products within a given divisional sample. These variations would inevitably result in different intensities of hybridization signals.

Verification of the exact position of each clone was

TABLE 1

Mapping position of cDNA, cosmid and RAPD clones

cDNA	Chromosone	Division	Sequence	Cosmid	Chromosome	Division	RAPD	Chromosome	Division
106	*X*	*3D*	NA[a]	cKM42	*X*	*4A-3D*	W3	*X*	*3B*
42	*X*	*4A-3D*	U50467	K10	*X*	*6*	BS8-3	*2R*	*13B*
107	*2R*	*7A*	U50468	COS12	*2R*	*8B*	N2	*2R*	*15D*
24	*2R*	*8C*	NA	K2	*2R*	*8C*	BS3	*2R*	*16A*
6	*2R*	*8D*	U50469	KM6	*2R*	*8D*	MS3	*2R*	*16B*
18	*2R*	*8D*	NA	K11	*2R*	*8D*	S9	*2R*	*18A*
89	*2R*	*8D*	NA	COS131	*2R*	*9A*	BL2	*2L*	*27C*
65	*2R*	*12A*	U50470	COS5	*2R*	*11B*	MF10	*2L*	*24A*
50	*2R*	*12A-B*	U50471	K12	*2R*	*11C*	SS3	*2L*	*20C*
52	*2R*	*14E'*[b]	U50472	COS4	*2R*	*12B*	B2A	*3R*	*31B*
114	*2R*	*14E"*[b]	U50473	COS13	*2R*	*12C*	MS1	*3L*	*40A*
68	*2R*	*15D*	NA	K3	*2R*	*12C*	SF10	*3L*	*38B*
27	*2R*	*15E*	NA	G11	*2R*	*12D*	ML2	*2R*	*19B*
13	*2R*	*15E-16A*	NA	COS9	*2R*	*13C*		*2L*	*27C*
86	*2R*	*16C*	NA	cKM52	*2R*	*15A*	MS12A	*2L*	*28B*
76	*2R*	*17A*	NA	cKM13	*2R*	*16A*		*3R*	*29D*
11	*2R*	*18A*	NA	K8	*2R*	*17A*		*3L*	*45A*
57	*2L*	*20C*	NA	G16	*2R*	*18C*	Sa	*2R*	*19E*
55	*2L*	*20D*	NA	COS1	*2L*	*28B*		*3R*	*35B*
64	*2L*	*21B*	NA	K5	*2L*	*25B*		*3R*	*37D*
4	*2L*	*21B*	NA	COS3	*2L*	*26C*	C3	*2R*	*13C*
119	*2L*	*22E*[c]	U50474	cKM122	*2L*	*22F*		*2R*	*14A*
97	*2L*	*22F"*	NA	COS10	*3R*	*29A*		*3R*	*31D*
122	*2L*	*22F'*	U51225	G19	*3R*	*29C*		*3L*	*43B*
129	*2L*	*26B*	U50476	COS7	*3L*	*38B*	B11B	*2R*	*11B*
88	*2L*	*24B*	NA	K7	*X*	*6*		*2R*	*16D*
105	*2L*	*24D*	NA		*2R*	*17A*		*3L*	*45B*
2	*2L*	*23C"*	U50477	COS123	*2R*	*22F*		*3L*	*43B*
79	*2L*	*23C'*	U50478		*3R*	*32B*	BS19	*12 sign*	
134	*2L*	*27D*	U50479						
32	*2L*	*28B"*	NA						
21	*2L*	*28B'*	NA						
29	*2L*	*28C*	NA						
3	*3R*	*29B*	U50480						
22	*3R*	*29C*	NA						
19	*3R*	*33A*	U50481						
28	*3R*	*37A*	NA						
14	*3R*	*37D*	NA						
75	*3L*	*39B-C*	NA						
77	*3L*	*40B*	NA						

[a] NA indicates that no sequence is available for the clone. For sequenced clones the GeneBank database Accession Number is provided. Nine of the sequenced clones revealed similarities with other sequences in the GeneBank database ($P < 0.001$). These were as follows: pKM42, guanylate cyclase; pKM107, tubulin alpha chain; pKM6, cuticle protein; pMK50, ribosomal protein S8; pKM52, fatty acid-binding protein; pKM119, trypsin-like precursor; pKM122, hexamerin; pKM79, mastermind protein; pKM134, ribosomal protein S13.

[b] When more than one clone maps in the same division but at different positions, they are presented as ′ and ″, as for example clones 52 and 114 in division 14E.

[c] Clone 119 gave two signals; both localized in division 22E.

determined by direct localization by *in situ* hybridization (Figure 1A). We expanded our analysis to 40 cDNA clones, 29 of which were from the aforementioned set of clones and 11 more were randomly chosen from the GASUA cDNA library. A total of 39 clones gave unique signals, and one (pKM119) gave two distinct signals in division *22E*, producing a total of 41 hybridization sites on the five chromosomal arms of *A. gambiae* (Table 1 and Figure 2).

The comparison between divisional probe analysis and *in situ* hybridization revealed agreement when the divisional-probe signal was strong, with the only exception of clone pKM107. This hybridized to four different divisional probes, giving signals ranging from very strong to very weak. As it turned out, there was concordance of one of the weaker signals with the *in situ* hybridization. With regard to the weaker divisional probe signals, less than a fourth of them was confirmed by *in situ* hybridization. It therefore seems that a strong divisional probe signal generally corresponds to the actual *in situ* localization, whereas a weaker signal should be further confirmed.

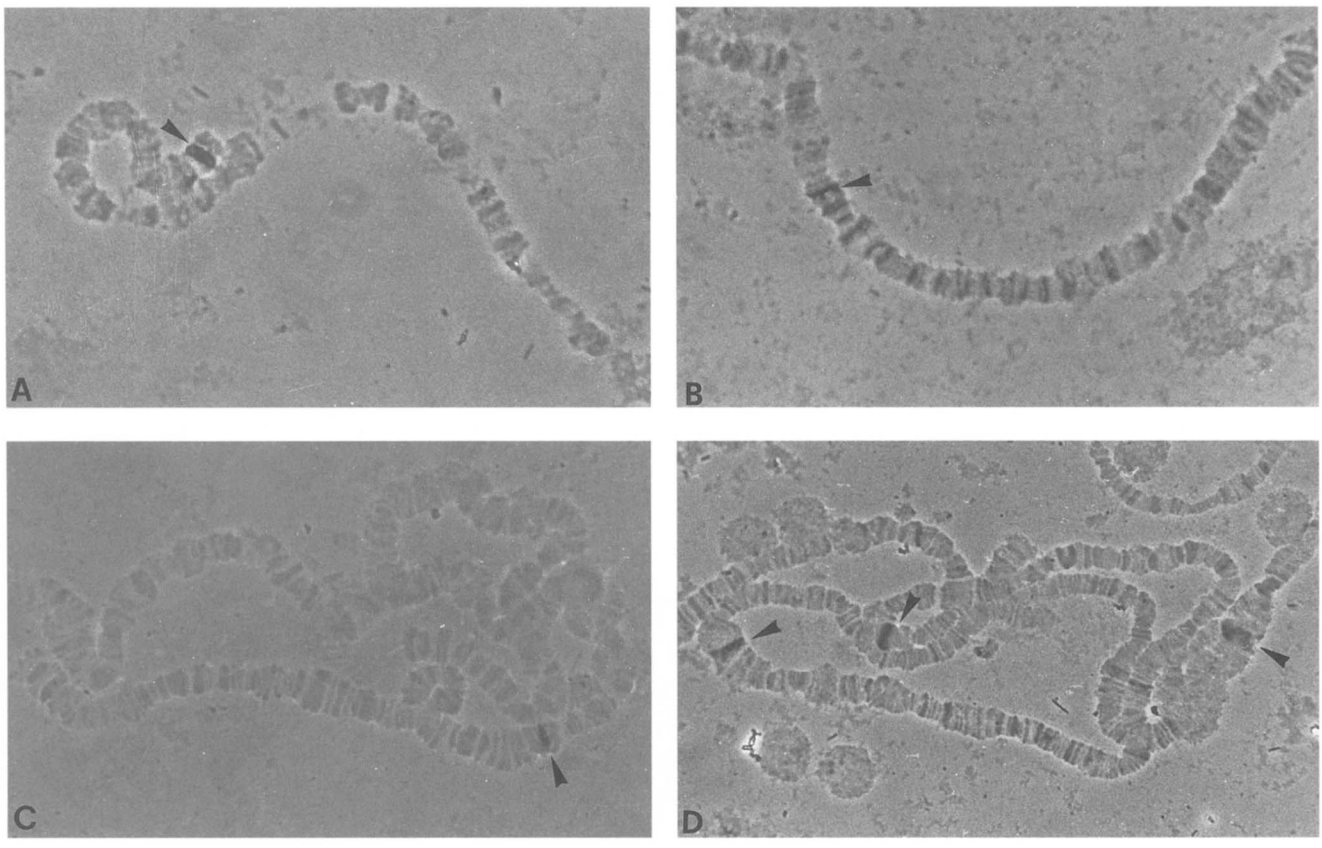

FIGURE 1.—*In situ* hybridization of various clones on *A. gambiae* chromosomes. Arrows indicate positions of the hybridization signals. (A) cDNA clone pKM107 in division *7A*. (B) Cosmid clone K3 in division *12C*. (C) RAPD clone MS1 in division *40A*. (D) RAPD clone Sa with three signals in divisions *19E, 35B* and *37D*.

For 15 of the cDNA clones analyzed, a partial nucleotide sequence has been obtained. Nine of these showed a high degree of similarity ($P < 0.001$) with sequences in the GenBank database (Table 1).

Cosmid clones: Twenty-five out of the 27 cosmids hybridized to a single division by *in situ* hybridization (Figure 1B). Cosmid K7 hybridized in divisions *6* and *17A*. Cosmid cos 123 hybridized in divisions *22F* and *32B*. Partial analysis of the latter cosmid revealed that it was a recombined clone of two *Sau*3A fragments that must have been ligated together during library construction (data not shown). Four cDNA clones (13, 42, 52 and 122) were found to map relatively close to breakpoints *2Rd* (Div. *16A*), *Xc* (Div. *4A-3D*), *2Ru* (Div. *15A*) and *2La* (Div. *22F*), respectively. These cDNA clones were used to screen the cosmid library and identify cosmids (cKM13, cKM42, cKM52 and cKM122) that were further used to perform chromosomal walks. Overlapping clones that span over 100 kb were obtained for cKM42 and cKM122 (K. MATHIOPOULOS, un-

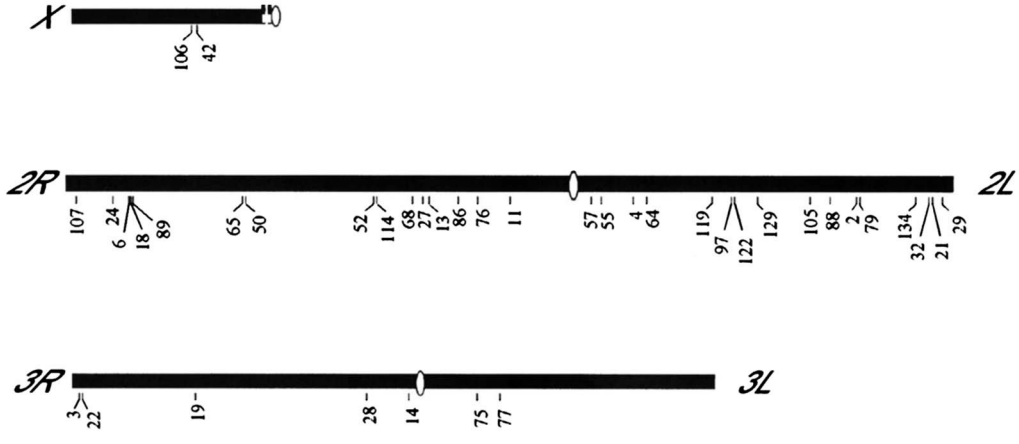

FIGURE 2.—Schematic presentation of the localization of cDNA clones on the *A. gambiae* polytene chromosomes.

published results). This effort was aimed at the characterization of inversion breakpoints, which would open the possibility for their genetic analysis, as well as the analysis of different karyotypes in natural *A. gambiae* populations.

RAPD clones: Twelve of the RAPD markers gave a unique signal after *in situ* hybridization, one of them hybridized to two sites, two to three sites, and two to four sites each, while only one clone hybridized to 12 sites distributed throughout the whole polytene complement (Figure 1, C and D). Further characterization of RAPD clones is in progress.

DISCUSSION

To gain insight into the structure, function, organization and evolution of the *A. gambiae* genome, we set out to isolate, map and partially characterize DNA segments of different origins, *i.e.*, random cDNA clones, cosmid clones, or RAPD markers. In all, 85 markers were localized on the *A. gambiae* chromosomes, yielding a physical map with a mean density of one molecular marker per 3000 kb.

The presence of coding regions among the genome markers (expected to be more conserved across species than RAPDs or other noncoding fragments) opens the possibility for comparisons of chromosomes or discrete chromosomal segments across related Diptera (STEINMANN *et al.* 1984; WHITING *et al.* 1989). Thus, one could identify homologous chromosomal segments across wide phylogenetic distances. In addition, defining conserved syntenic groups of genes could give further insights into the extent of genome conservation, particularly conservation of gene order, that might allow prediction, to some degree, of the location of genes cloned in other species. Incidentally, preliminary results in our laboratory showed that *in situ* hybridization of *A. gambiae* cDNA clones is fully feasible on the *A. stephensi* polytene chromosome complement (data not shown).

The work was partially supported by World Health Organization, Tropical Disease Research grant ID 940411. K.D.M. was financially supported by European Human Capital and Mobility Fellowship No. ERBCHBICT930896. We are grateful to Mr. G. PETRANGELI and Ms. M. CALZETTA for their excellent technical support and to Drs. J. R. POWELL and V. PETRARCA for their comments and suggestions on the manuscript.

LITERATURE CITED

BESANSKY, N. J., and J. R. POWELL, 1992 Reassociation kinetics of *Anopheles gambiae* DNA. J. Med. Entomol. **29:** 125–128.

COLUZZI, M., 1982 Spatial distribution of chromosomal inversions and speciation in Anopheline mosquitoes, pp. 143–153 in *Mechanisms of Speciation*, edited by C. BARIGOZZI. A. R. Liss Inc., New York.

COLUZZI, M., and A. SABATINI, 1967 Cytogenetic observations on species A and B of the *Anopheles gambiae* complex. Parassitologia **9:** 73–88.

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. R. Soc. Trop. Med. Hyg. **73:** 483–497.

COLUZZI, M., V. PETRARCA and M. A. DI DECO, 1985 Chromosomal inversion integradation and incipient speciation in *Anopheles gambiae*. Boll. Zool. **52:** 45–63.

DAVIDSON, G., 1985 Genetics and mosquito control. Parassitologia **27:** 169–178.

DIMOPOULOS, G., L. ZHENG, V. KUMAR, A. DELLA TORRE, F. C. KAFATOS *et al.*, 1996 Integrated genetic map of *Anopheles gambiae:* use of RAPD polymorphisms for genetic, cytogenetic and STS landmarks. Genetics **143:** 953–960.

FAVIA, G., G. DIMOPOULOS, A. DELLA TORRE, Y. T. TOURÉ, M. COLUZZI *et al.*, 1994 Polymorphisms detected by random PCR distinguish between different chromosomal forms of *Anopheles gambiae*. Proc. Natl. Acad. Sci. **91:** 10315–10319.

KUMAR, V., and F. H. COLLINS, 1994 A technique for nucleic acid *in situ* hybridization to polytene chromosomes of mosquitoes in the *Anopheles gambiae* complex. Ins. Mol. Biol. **3:** 41–47.

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

STEINMANN M., W. PINSKER and D. SPERLICH, 1984 Chromosomal homologies within the *Drosophila obscura* group probed by *in situ* hybridization. Chromosoma **91:** 46–53.

STUERCHEL, D., 1989 Malaria, how much is there worldwide? Parasitol. Today **5:** 39–40.

WHITING, J. H. JR, M. D. PLILEY, J. L. FARMER and D. E. JEFFREY, 1989 *In situ* hybridization analysis of chromosomal homologies in *Drosophila melanogaster* and *Drosophila virilis*. Genetics **122:** 99–109.

ZHENG, L., R. D. C. SAUNDERS, D. FORTINI, A. DELLA TORRE, M. COLUZZI *et al.*, 1991 Low-resolution genome map of the malaria mosquito *Anopheles gambiae*. Proc. Nat. Acad. Sci. **88:** 11187–11191.

ZHENG, L., F. H. COLLINS, V. KUMAR and F. C. KAFATOS, 1993 A detailed genetic map for the X chromosome of the malaria vector, *Anopheles gambiae*. Science **261:** 605–608.

ZHENG, L., M. Q. BENEDICT, A. J. CORNEL, F. H. COLLINS and F. C. KAFATOS, 1996 An integrated genetic map of the African human malaria vector mosquito *Anopheles gambiae*. Genetics **143:** 941–952.

Communicating editor: V. G. FINNERTY