Influence of Two Wolbachia Strains on Population Structure of East African *Drosophila simulans*

Matthew D. Dean,* Kirrie J. Ballard,† Anne Glass‡ and J. William O. Ballard*,[1]

**Department of Biological Sciences, University of Iowa, Iowa City, Iowa 52242-1324, †Department of Speech Pathology and Audiology, University of Iowa, Iowa City, Iowa 52242-1324 and ‡Eberhard Karls Universität, Tübingen, Germany 72074*

Manuscript received May 22, 2003
Accepted for publication August 25, 2003

ABSTRACT

Drosophila simulans is hypothesized to have originated in continental East Africa or Madagascar. In this study, we investigated evolutionary forces operating on mitochondrial DNA (mtDNA) in populations of *D. simulans* from Zimbabwe, Malawi, Tanzania, and Kenya. Variation in mtDNA may be affected by positive selection, background selection, demographic history, and/or any maternally inherited factor such as the bacterial symbiont Wolbachia. In East Africa, the *w*Ri and *w*Ma Wolbachia strains associate with the *si*II or *si*III mitochondrial haplogroups, respectively. To ask how polymorphism relates to Wolbachia infection status, we sequenced 1776 bp of mitochondrial DNA and 1029 bp of the X-linked *per* locus from 79 lines. The two southern populations were infected with *w*Ri and exhibited significantly reduced mtDNA variation, while Wolbachia-uninfected *si*II flies from Tanzania and Kenya showed high levels of mtDNA polymorphism. These are the first known populations of *D. simulans* that do not exhibit reduced mtDNA variation. We observed no mitochondrial variation in the *si*III haplogroup regardless of Wolbachia infection status, suggesting positive or background selection. These populations offer a unique opportunity to monitor evolutionary dynamics in ancestral populations that harbor multiple strains of Wolbachia.

A MAJOR goal of this study is to investigate the selective forces that influence population subdivision in the mitochondrial genome of *Drosophila simulans* in East Africa, where the species is thought to have originated (Lachaise *et al.* 1988; Begun and Aquadro 1993; Irvin *et al.* 1998; Hamblin and Veuille 1999; Andolfatto 2001). A common goal of population genetic investigations is to explain the fate of genetic polymorphism within a species. Mechanisms that may shape genetic variation include positive selection (Fay and Wu 2000; Andolfatto and Przeworski 2001), background selection (Charlesworth 1996; Hamblin and Aquadro 1996), or a combination of the two (Kim and Stephan 2000; Jensen *et al.* 2002). Demographic and stochastic processes may also eliminate genetic polymorphism. The mitochondrial genome is especially susceptible to positive and/or background selection because the molecule does not recombine. Positive selection may cause a beneficial mutation arising anywhere in the genome to increase in frequency, carrying with it all linked variants. Similarly, background selection can remove deleterious mutations and all linked variants from populations. In addition, any maternally inherited factor can potentially influence the evolution of mitochondrial DNA (mtDNA). One such factor in many insect species is the maternally inherited bacterial symbiont Wolbachia.

Wolbachia elicits a phenomenon known as cytoplasmic incompatibility in many different insect species (reviewed in Hoffmann and Turelli 1997). In its simplest form, cytoplasmic incompatibility occurs when sperm from an infected male fertilizes ova from an uninfected female, resulting in reduced egg hatch. *D. simulans* is infected with four genetically distinct strains of Wolbachia that average >12% pairwise differences at the nucleotide level of the *wsp* locus (Zhou *et al.* 1998). The nomenclature of the Wolbachia strains and their distribution has recently been reviewed by Ballard (2004). The *w*Ha strain infects flies in Hawaii, Tahiti, New Caledonia, and the Seychelles (O'Neill and Karr 1990; James *et al.* 2002); the *w*Ma strain in New Caledonia, Madagascar, Reunion, and parts of continental East Africa (Merçot and Poinsot 1998; James and Ballard 2000; James *et al.* 2002); the *w*Au strain in Australia, Ecuador, and Florida (Turelli and Hoffmann 1995; Ballard *et al.* 1996); and the widespread *w*Ri strain in North and South America, Europe, Africa, and Asia (Turelli and Hoffmann 1995).

In continental East Africa, two strains of Wolbachia have been identified: *w*Ri in Zimbabwe (Turelli and Hoffmann 1995) and *w*Ma in Tanzania (Merçot and Poinsot 1998). When *w*Ri-infected males fertilize ova from uninfected females, egg hatch is reduced 30–70% in the field (Turelli and Hoffmann 1995). All other crosses produce normal numbers of progeny, granting

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY370193–AY370508.

[1]*Corresponding author:* Department of Biological Sciences, 208 Biology Bldg., University of Iowa, Iowa City, IA 52242-1324.
E-mail: bill-ballard@uiowa.edu

infected females a reproductive advantage since they can successfully reproduce with either infected or uninfected males. As a result, *w*Ri-infected individuals are expected to increase in frequency, although a slight fecundity deficit associated with *w*Ri may counteract this process (Hoffmann *et al.* 1990). A rapid increase in infection frequency has been observed in several natural populations (Turelli and Hoffmann 1991; Turelli *et al.* 1992). In contrast to *w*Ri, the *w*Ma strain exhibits variable levels of intermediate incompatibility in the laboratory (James and Ballard 2000), with little known about the dynamics of this strain in the field.

Wolbachia and mtDNA are maternally inherited in *D. simulans*, although rare cases of paternal leakage for both occur (Hoffmann and Turelli 1988; Satta *et al.* 1988; Kondo *et al.* 1990, 1992; Nigro and Prout 1990). Due to their shared mode of transmission, the mtDNA variant(s) initially associated with a spreading infection is also expected to increase in frequency through a process analogous to genetic hitchhiking. As a result, an infected population is expected to have a high frequency of one mitochondrial type and to lack mitochondrial variation (Turelli *et al.* 1992). If uninfected populations later arise through stochastic loss of infection coupled with founder events, the signature of reduced mtDNA polymorphism will remain. The recovery of mtDNA polymorphism will depend on the scope of and time since the most recent sweep. Wolbachia-induced population sweeps are not expected to affect the nuclear genome, so a nuclear marker may be used as a control comparison to mtDNA variation. In contrast, a population level process may be expected to affect both the mitochondrial and nuclear genomes.

All else equal, strains of Wolbachia inducing strong incompatibility are more likely to induce a population sweep and to reduce host mtDNA polymorphism. Nevertheless, a reduction in mtDNA polymorphism has been observed in every *D. simulans* population surveyed to date (Hale and Hoffmann 1990; Turelli *et al.* 1992; Ballard and Kreitman 1994; Rand *et al.* 1994; Ballard *et al.* 1996; James *et al.* 2002), even though not all strains infecting *D. simulans* induce strong incompatibility (James and Ballard 2000). However, none of these studies have included population samples from East Africa, where *D. simulans* is thought to have arisen (Lachaise *et al.* 1988; Begun and Aquadro 1993; Hamblin and Veuille 1999; Andolfatto 2001).

We sampled *D. simulans* from five populations throughout East Africa to investigate whether the level of mitochondrial variability in uninfected flies is compatible with a neutral model when compared to an autosomal locus. If mtDNA diversity of uninfected lines is significantly reduced relative to an appropriate autosomal locus, we would infer that the mitochondrial genome has been subjected to a recent sweep of genetic variation. In Tanzania and Kenya, we discovered three populations that show a higher amount of mtDNA polymorphism than all previous descriptions of this species. In contrast, Wolbachia-infected populations in Malawi and Zimbabwe to the south show evidence for a reduction in mtDNA variability. The discovery of *D. simulans* populations that have not been subject to a sweep in the detectable past facilitates our understanding of the population level processes that act on mtDNA. More specifically, studying Wolbachia-uninfected populations enhances our understanding of interhaplogroup relationships and the biogeography of *D. simulans* (Ballard 2004).

Another goal of the study was to address a potential conundrum regarding the distribution of distinct Wolbachia strains. Theory suggests that as long as the two strains do not doubly infect single individuals, long-term coexistence is impossible (Rousset *et al.* 1991; Hoffmann and Turelli 1997). The *w*Ri and *w*Ma strains of Wolbachia had previously been reported from this region (Turelli and Hoffmann 1995; Merçot and Poinsot 1998), but it was unknown whether they occurred in sympatry. These strains associate nonrandomly with two mitochondrial haplogroups, *si*II and *si*III, respectively (James and Ballard 2000), which differ by >2% at the nucleotide level (Ballard 2000a). In east Africa we find that the two strains do not stably coexist, but instead are highly subdivided.

MATERIALS AND METHODS

Collecting sites and lines used: Throughout July 2001, single females were placed on instant Drosophila medium within a few hours of collection to establish isofemale lines. Upon returning to the laboratory, male genitalia were examined to confirm species identification (Coyne 1983, 1985).

We report our collection localities from south to north (Figure 1). In Zimbabwe, 48 isofemale lines were established from a citrus orchard on the Victoria Falls Hotel grounds, Victoria Falls, on July 8. In Malawi, 28 isofemale lines were established from tangerines and oranges in Mwanza on July 10. In Tanzania, 23 isofemale lines were established from mixed fruit in the TX Market in the Upanga District of Dar es Salaam on July 13 and 15. In Malindi, Kenya, 60 isofemale lines were established from tomatoes in the New Malindi Market on July 18 and 19. In Nairobi, Kenya, 47 isofemale lines were established from mixed fruit on Forrest Road in the Westlands district on July 23. We also assayed single male flies that were placed immediately in 100% ethanol in the field. From Dar es Salaam, 22 additional males were assayed. From Nairobi, 5 additional males were assayed.

To gain a temporal perspective, we included nine isofemale lines collected 7 years earlier from Harare, Zimbabwe (Hutter *et al.* 1998). The infection status of these lines has not been previously examined.

Cytotype distribution and abundance: We defined cytotype as "the combination of mitochondrial haplogroup/Wolbachia strain." For example, *si*III/*w*Ma denotes an individual carrying *si*III mtDNA and infected with the *w*Ma strain of Wolbachia. We use *w*− to denote uninfected cytotypes, as in *si*III/*w*−.

The distribution and abundance of cytotypes were determined using restriction enzyme digests and allele-specific PCR, with a subset confirmed by sequencing. DNA of adult *D. simulans* was isolated using the fixed tissue protocol from the

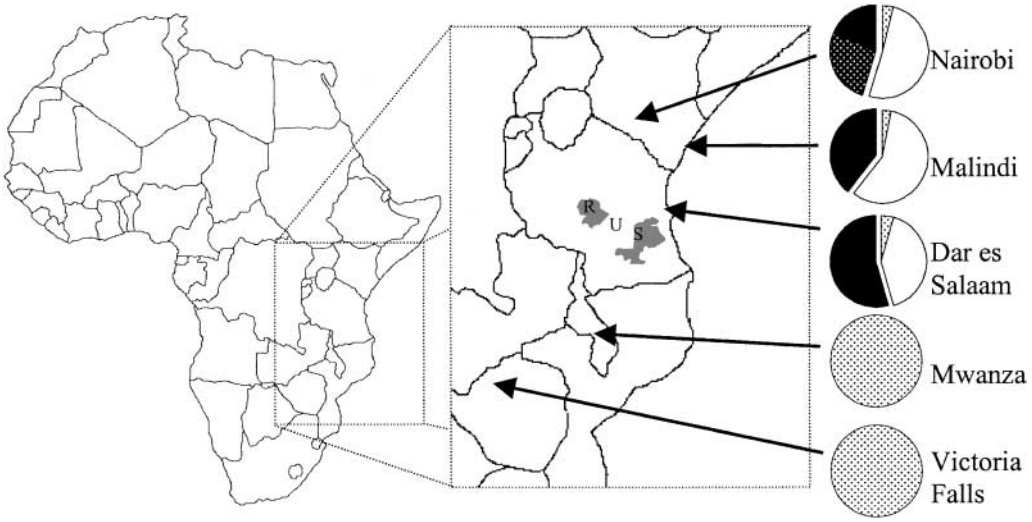

FIGURE 1.—Cytotype distribution and abundance. In pie graphs, an open background represents the *si*II haplogroup, while a solid background represents *si*III. Stippling indicates Wolbachia infection (*w*Ri for the *si*II haplogroup and *w*Ma for the *si*III haplogroup). Arrows indicate approximate locations of Victoria Falls, Zimbabwe; Mwanza, Malawi; Dar es Salaam, Tanzania; Malindi, Kenya; and Nairobi, Kenya. Shaded areas show the parks Ruaha (R) and Selous (S). The Udzungwa Mountains (U) lie between R and S (see DISCUSSION).

Puregene kit (Gentra, Research Triangle Park, NC) with either single male flies or three flies from each isofemale line. From newly established lines, DNA was extracted one to five generations after collection. Primers used in this study are available at http://myweb.uiowa.edu/bballard/eastafrica2001.htm.

Wolbachia typing: An isofemale line was scored as uninfected if two independent DNA extractions tested negative for 16S rDNA amplification (following O'NEILL *et al.* 1992), negative for *wsp* amplification (following ZHOU *et al.* 1998), and positive for amplification of a portion of mtDNA (following JAMES *et al.* 2002). To identify Wolbachia strains from infected lines, we amplified a portion of the *wsp* locus and digested the amplicon with the restriction enzyme *Dnp*II (following JAMES and BALLARD 2000). This digest yields fragments that are diagnostic for *w*Ri, *w*Ma, *w*Au, or *w*Ha using known strains of Wolbachia as positive controls. We sequenced *wsp* from a subset of infected lines to confirm that this assay was accurate (following JAMES and BALLARD 2000).

Distinguishing among siI, siII, and siIII: An allele-specific PCR assay was developed to distinguish between the three distinct *D. simulans* haplogroups (Figure 2A). The 10-μl PCR reactions were carried out with 10 ng genomic DNA, 1.6 pmol of the primer 5983−, 1.7 pmol of 4726+, 0.9 pmol of 5183+, 1.3 pmol of 5545+, 1 μl 35 mM MgCl 10× PCR buffer (Roche), 1 μl 8 mM dNTPs, and 0.05 units Taq polymerase (Roche). The reactions were subjected to 34 cycles of 94° for 15 sec, 54° for 10 sec, and 72° for 60 sec. Negative and positive controls were included in each set of reactions. Amplicons were scored following electrophoresis through a 2% agarose gel.

In a similar assay, we distinguished *si*IIA from *si*IIB by exploiting the fixed difference occurring at position no. 3441 (BALLARD 2000a; Figure 2B). Distinguishing *si*IIA from *si*IIB was important because they have been shown to associate with the *w*Ri or *w*Au strains, respectively, of Wolbachia. Specificity to either *si*IIA or *si*IIB was achieved by aligning the fixed difference at position no. 3441 to the penultimate 3′ base of either primer 3440− or 3442+, with the other primer matching. The 10-μl PCR reactions were carried out with 10 ng genomic DNA, 1.4 pmol each primer, 1 μl 20 mM MgCl 10× PCR buffer, subjected to 32 cycles at 95° for 15 sec, 54° for 15 sec, and 72° for 60 sec, and scored on a 2% agarose gel.

mtDNA sequencing: A total of 1776 bp was sequenced from 91 lines (Table 1). Where possible, we sampled ~10 *D. simulans* isofemale lines of each cytotype from each population. We sequenced only three lines from Victoria Falls because their cytotype was identical to Mwanza and minimal information would be gained by their additional sequencing. *D. melanogaster* sequences from two lines, Z53 and Oregon-R (accession nos. AF200828 and AF200829 from BALLARD 2000b), were used for interspecific comparisons.

Three regions of 599, 601, and 576 bp were amplified and sequenced. These regions included portions of the ND2, COI, COII, ND5, and ND4 genes, the transfer RNAs for Trp, Cys, Tyr, Asp, and His, and four intervening spacer regions. To

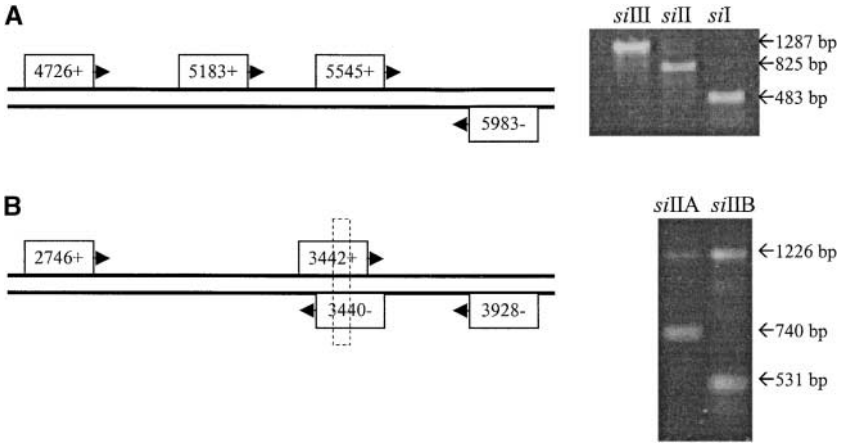

FIGURE 2.—Schematic and empirical results of competitive PCR assays that distinguish (A) *si*I, *si*II, and *si*III mtDNA and (B) *si*IIA and *si*IIB. In A, these primers amplify fragments of 1287, 825, and 483 bp from *si*III, *si*II, or *si*I, respectively. These primers amplify a 1226-bp fragment common to both *si*IIA and *si*IIB and either a 740-bp fragment specific to *si*IIA or a 531-bp fragment specific to *si*IIB. Thick lines indicate DNA strands, boxes indicate primers, arrows indicate direction of elongation, and the dashed box highlights the SNP utilized to distinguish *si*IIA from *si*IIB. Figure not drawn to scale.

amplify each region, 25-μl PCR amplifications were carried out with 35 ng genomic DNA, 4 pmol each primer, and 2.5 μl 30 mM MgCl 10× PCR buffer. Reactions were subjected to 35 cycles of 95° for 15 sec, 52° for 15 sec, and 72° for 60 sec. We visualized 4 μl on an agarose gel and then selectively precipitated the remaining amplicon following a modification of Kreitman (1991). In a 96-well format, we added 10.5 μl of 7.5 M ammonium acetate and 31.5 μl of cold 100% ethanol to each PCR reaction. The mixture was spun for 15 min at 2200 × *g*, caps were removed, and tubes spun inverted for 1 min at 100 × *g*. The pellet was washed by adding 200 μl 70% ethanol, respun for 1 min at 2200 × *g*, dried with a second inverted spin, and resuspended in 25 μl water.

For sequencing, 1–4 μl of the purified PCR reactions was added to a 10-μl reaction containing 4 μl of 1:3 Terminator Ready Reaction mix (Big Dye version 3, Applied Biosystems, Foster City, CA). We added 3 pmol of PCR primer to a 10-μl reaction, which was subjected to 25 cycles of 96° for 10 sec, 50° for 5 sec, and 60° for 2 min, preceded by a 30-sec hold at 96°. Sequencing reactions were precipitated according to the isopropanol precipitation protocol (Applied Biosystems) and then loaded into an ABI 3100 capillary machine. Chromatograms were imported into Sequencher version 4.1, where they were edited manually and aligned against the mtDNA genomes of Ballard (2000a).

Per DNA sequencing: We gathered 1029 continuous base pairs of *per* from 79 *D. simulans* lines, a subset of those from which we had collected mtDNA sequence (Table 1). We did not sequence *per* from the Victoria Falls or Harare populations because their cytotype structure was identical to Mwanza (see below). We sequenced the same region from the *D. melanogaster* lines Z53 and Oregon-R. The region contained one complete exon, two partial exons, and two introns. All *per* sequences were gathered from the same male, ensuring a single copy of the X chromosome. The 50-μl PCR reactions were carried out with 35 ng template DNA, 2.5 pmol of each primer, and 5 μl of 20 mM MgCl PCR buffer and subjected to 35 cycles of 95° for 45 sec, 54°–47° (decreasing 0.2° after each cycle) for 60 sec, and 72° for 75 sec. Reactions were precipitated as described above, and 1–4 μl was employed in sequencing reactions. For sequencing, we used 8 pmol of each primer. The cycling profile was 28 cycles of 96° for 10 sec, 52° for 15 sec, and 60° for 2 min, preceded by a 30-sec hold at 96°. Sequences were edited and aligned against the *per* locus sequenced by Citri *et al.* (1987).

Wolbachia and host variation: We expected populations that have been subjected to a recent Wolbachia-induced sweep to (1) retain significantly fewer segregating sites than populations unaffected by Wolbachia, (2) show a reduction in mtDNA variation relative to *per*, and (3) harbor mtDNA sequences that form a monophyletic assemblage that is not mirrored by *per*. We investigated the first prediction with coalescent simulations and neutrality tests, the second prediction with Hudson-Kreitman-Aguadé (HKA) tests, and the third with network analyses.

Sequence polymorphism: Two estimates of polymorphism within a locality, π (Nei and Li 1979) and θ (Watterson 1975), were calculated from silent and synonymous sites. We compared θ of all sites from both mtDNA and *per* among infected *vs.* uninfected flies by calculating 95% confidence intervals using 50,000 runs of a coalescent simulation (Rozas and Rozas 1997). These simulations assume an infinite sites model and constant population size and can incorporate recombination. Mutations are distributed along genealogies according to a Poisson process.

Neutrality tests: Two heuristic tests were used to evaluate the null hypothesis that silent and synonymous sites evolve in a manner consistent with neutrality. Tajima's *D* (Tajima 1989) tests whether π and θ differ significantly, indicating selection or an expanding population. Fu's *Fs* (Fu 1997) tests whether there is an excess of young mutations. Both tests were carried out on silent/synonymous sites. Significance of both tests was determined using coalescent simulations implemented in DNAsp 3.53 (Rozas and Rozas 1997) so that recombination could be included in the case of *per*.

HKA tests: If a population has recently been subjected to a Wolbachia-induced population sweep, then mtDNA should display reduced polymorphism compared to a nuclear locus, which is not affected by the sweep. The HKA test (Hudson *et al.* 1987) compares polymorphism and divergence at two or more unlinked loci. Differences in effective population sizes among regions of the genome were corrected for by assuming equal numbers of reproducing males and females, in which case the effective population size of mtDNA is approximately one-third that of *per*. Under neutral expectations, the ratio of polymorphism to divergence is equal between the two loci.

Polymorphism and divergence of silent and synonymous sites were calculated using DNAsp 3.53 (Rozas and Rozas 1997). Polymorphism was calculated for both *D. simulans* and *D. melanogaster*. Intervening spacer regions were considered silent sites, but tRNAs were excluded. Our conclusions do not change if we include tRNAs in the calculation but we present the results with tRNAs excluded. Using silent/synonymous sites, HKA tests were calculated for each mtDNA haplogroup from each locality. We did not group flies with distinct mtDNA haplogroups because this would artificially inflate measures of mtDNA polymorphism relative to divergence, which in turn might lead to false acceptance of the null hypothesis. Statistical significance was determined with 10,000 coalescent simulations using the program HKA, written and distributed by Jody Hey (http://lifesci.rutgers.edu/heylab/DistributedProgramsandData.htm#HKA). Coalescent simulations naturally incorporate observations of zero polymorphism, which could otherwise be problematic in traditional χ^2 evaluations of the HKA statistic. Assuming random mating, it could be argued that the effective population size of mtDNA should be further corrected by the proportion of flies with each mtDNA haplogroup at a locality. We present the results using both corrections, but note our conclusions do not change.

Phylogenetic analyses: To visualize genealogical relationships and possible population substructure, we built networks of both mtDNA and *per* with statistical parsimony algorithms (Templeton *et al.* 1992) implemented in TCS 1.13 (Clement *et al.* 2000). Whereas traditional phylogenetic methods represent evolution with bifurcating trees, network analyses account for the persistence of ancestral sequences and recombination by allowing multifurcations (Posada and Crandall 2001). The TCS program calculates the number of mutational steps below which sequences can be joined with 95% confidence. This point is the parsimony limit and no connections can exceed this number (Templeton *et al.* 1992).

RESULTS

Cytotype distribution and abundance: All *si*II flies were of the *si*IIA subtype. Victoria Falls in Zimbabwe and Mwanza in Malawi harbored only *si*II/*w*Ri flies (Table 1, Figure 1). All nine lines from Harare in Zimbabwe, collected in 1994, were also infected with the *w*Ri strain. Dar es Salaam in Tanzania and Malindi in Kenya contained primarily *si*II/*w*− and *si*III/*w*− flies. Nairobi in Kenya contained primarily *si*II/*w*− flies and both *si*III/*w*− and *si*III/*w*Ma flies. The Dar es Salaam, Malindi,

TABLE 1

Cytotype distribution and abundance in the five East African populations

Locality	*si*II		*si*III	
	*w*Ri	*w*−	*w*Ma	*w*−
Victoria Falls	48 (3[a])	—	—	—
Harare[b]	9 (9[a])			
Mwanza	28 (11)	—	—	—
Dar es Salaam	1 (1)	19 (9)	—	25 (10)
Malindi	1 (1)	35 (9)	—	24 (10)
Nairobi	1	27 (11)	14 (9)	10 (8)

Individuals collected from each locality carried *si*II or *si*III mtDNA and were infected with the *w*Ri or *w*Ma strains of Wolbachia or were uninfected (*w*−). Numbers in parentheses indicate the number of lines from which 1776 bp of mtDNA and 1029 bp of *per* were sequenced for further investigations.

[a] Only mtDNA (not *per*) sequenced (see text).

[b] Lines collected in 1994 (Hutter *et al.* 1998).

and Nairobi populations each harbored a single *si*II *w*Ri-infected individual (Table 1, Figure 1).

Above a critical threshold, the frequency of *w*Ri-infected individuals is expected to reach 0.94 (95% "exact" binomial confidence intervals = 0.92–0.96, n = 480; Turelli and Hoffmann 1995). On the basis of this estimate, we considered Victoria Falls (1.00, 0.93–1.00), Mwanza (1.00, 0.88–1.00), and Harare (1.00, 0.66–1.00) to be *w*Ri infected. Dar es Salaam (0.02, 0.00–0.12), Malindi (0.02, 0.00–0.09), and Nairobi (0.02, 0.00–0.10) were considered *w*Ri uninfected even though they contained a single *w*Ri-infected individual.

In nature, the *w*Ma infection attains a much lower equilibrium infection frequency of 0.14 (0.09–0.20, n = 193; combined data of Merçot and Poinsot 1998 and James and Ballard 2000), consistent with the much lower incompatibility that this strain induces (James and Ballard 2000). We considered Nairobi (0.27, 0.16–0.41) to be *w*Ma infected and Victoria Falls (0.00, 0–0.07), Mwanza (0.00, 0.00–0.12), Dar es Salaam (0.00, 0.00–0.08), and Malindi (0.00, 0.00–0.06) to be *w*Ma uninfected. The confidence intervals of Mwanza overlap with the expected infection frequency, but we maintain Mwanza is *w*Ma uninfected because no *w*Ma-infected individuals were collected here and the population was entirely fixed for the *w*Ri infection.

During sequencing, we noted nine *si*III lines that also carried low copy number of the *si*II mtDNA. Three of these lines were from Dar es Salaam, one from Malindi, and five from Nairobi. These lines yielded "clean" *si*III sequence from two amplicons but "clean" *si*II sequence from the third amplicon. This third amplicon had a 2-bp mismatch with *si*III mtDNA, causing selective amplification of the *si*II molecule. The primers from the other two amplicons matched both *si*III and *si*II perfectly. If the *si*II molecule were more abundant, we would not expect to see "clean" sequence from these former two amplicons. We replaced the mismatched primer with one that was an exact match to *si*III mtDNA. Using this new primer, amplification and sequencing from the same DNA extractions yielded "clean" *si*III sequence, supporting the interpretation that *si*II was present in very low copy number and was amplified only because of the mismatch of the original primer. For further analyses these lines were scored as *si*III, using the sequence generated from the new primer. The nine heteroplasmic lines were tested six generations later: only five remained heteroplasmic while four returned to the homoplasmic *si*III state. Heteroplasmy within isofemale lines of *D. simulans* has previously been reported from Reunion (Satta *et al.* 1988; Matsuura *et al.* 1991) and New Caledonia (James *et al.* 2002).

The existence of *si*III backgrounds with low copy number of *si*II molecules gave rise to the question of whether there was directionality to heteroplasmy. We tested 14 *si*II lines from these three localities using the *si*III-matching primer and none were heteroplasmic. This difference was statistically significant (one-tailed Fisher's exact test, $P = 0.04$), supporting previous studies that found heteroplasmy was more likely to arise when the incumbent mitochondrial type was *si*III rather than *si*II (de Stordeur *et al.* 1989; de Stordeur 1997).

Wolbachia and host variation: Among 29 *si*II/*w*− flies, there were 15 unique mtDNA genotypes, counting indels as a "fifth state" (Figure 3). All indels occurred as single sites, so each was counted once. After dividing the 29 *si*II/*w*− mtDNAs into Dar es Salaam, Malindi, or Nairobi, we found no fixed differences and four to six shared polymorphisms in pairwise comparisons. All *si*II/*w*− sequences from this study differed from previously published mitochondrial genomes (Ballard 2000a). Among 25 *si*II/*w*Ri lines sequenced, 24 shared a common genotype regardless of geography (Figure 3). For simplicity, the common mtDNA sequence from the infected *si*II group is hereafter referred to as "*si*II/*w*Ri mtDNA." This sequence is identical to that previously reported for *w*Ri-infected lines DSR and C167 (Ballard 2000a). One *w*Ri-infected line differed from the *si*II/*w*Ri mtDNA by a single substitution (Figure 3), a nonsynonymous G → A transition (in a 5′ → 3′ direction) at position 7925, causing an Asp → Asn amino acid change in the ND5 gene. This single substitution was confirmed with an independent DNA extraction. There were no segregating sites within 37 *si*III lines and all were homosequential with the common *si*III genotype (Ballard 2000a).

Within 1029 bp of *per* there were 105 segregating sites that defined 48 distinct genotypes, including indels as a "fifth state." After dividing the 79 *per* sequences into Mwanza, Dar es Salaam, Malindi, or Nairobi we found no fixed differences and 38 and 52 shared polymorphisms in pairwise comparisons. Consistent with theoretical predictions that Wolbachia should not influence

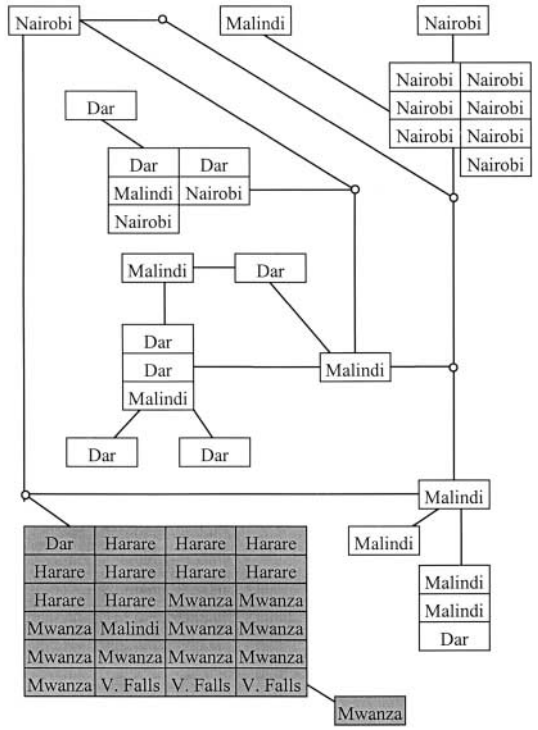

FIGURE 3.—Network of 1776 bp of *si*II mtDNA inferred by statistical parsimony. Adjacent rectangles represent sequences that are homosequential. The *si*III haplogroup exceeds the statistical parsimony limit and cannot be joined to this network with confidence. Sequences are named according to the population from which they were sampled. Branches represent one step and are not drawn to scale. Open nodes are inferred intermediate sequences. Shaded nodes represent individuals infected with *w*Ri. Dar, Dar es Salaam; V. Falls, Victoria Falls; Harare, Harare 1994.

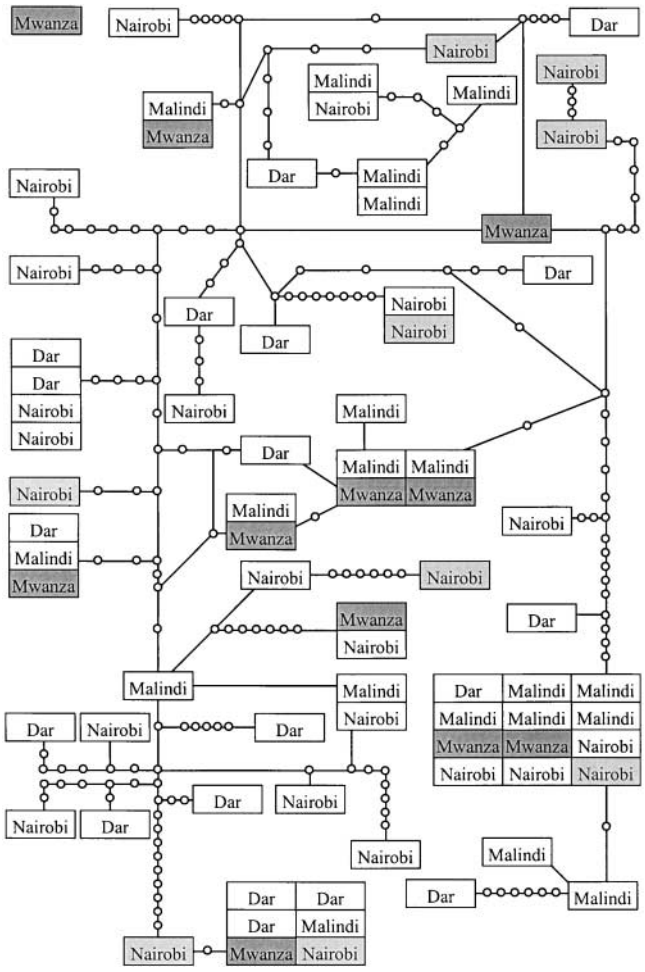

FIGURE 4.—Parsimony network of *per* sequences. Adjacent squares represent sequences that are homosequential. Sequences are named according to the population from which they were sampled. Branches represent one step and are not drawn to scale. Open nodes are inferred intermediate sequences. Dark gray and light gray nodes represent individuals infected with *w*Ri or *w*Ma, respectively. Dar, Dar es Salaam.

nuclear gene flow (CASPARI and WATSON 1959; TURELLI and HOFFMANN 1999), there were no unique *per* sequences associated with infection status or mtDNA haplogroup (Figure 4).

Sequence polymorphism: If we include silent, synonymous, and nonsynonymous sites, estimates of both π and θ for *si*II mtDNA were at least one order of magnitude lower for Mwanza (infected with the *w*Ri Wolbachia strain) compared to Dar es Salaam, Malindi, and Nairobi (not infected with *w*Ri). The estimated θ for *si*II/*w*Ri lines from Mwanza (0.0003, 0.0000–0.0010, $n = 11$) had nonoverlapping confidence intervals compared to the pooled *si*II/*w*− lines from Dar es Salaam, Malindi, and Nairobi (0.0035, 0.0012–0.0070, $n = 29$). This result is conservative because the single substitution found in Mwanza was a nonsynonymous site, so θ for Mwanza would be zero if we analyzed only silent/synonymous sites. As expected, the estimated θ for *per* from the *w*Ri-infected lines (0.0153, 0.0010–0.0213, $n = 11$) had broadly overlapping confidence intervals compared to the pooled uninfected lines (0.0171, 0.0119–0.0226, $n = 29$). For this latter calculation, we employed the empirically determined recombination parameter R (HUDSON 1987) of 71.3. It may be argued that the singleton *si*II/*w*Ri sequences from Dar es Salaam and Malindi should be excluded because they were not selected at random; that is, we deliberately chose to include them because they were infected. If these lines are excluded, the estimates of π changed to 0.0047 and 0.0074 and the estimates of θ change to 0.0052 and 0.0083, respectively, for the two localities.

We cannot compare θ of mtDNA from *si*III/*w*− to *si*III/*w*Ma flies because no polymorphism exists in *si*III. BALLARD (2000a) found only three singleton polymorphisms within 15,034 bp of mtDNA from nine *si*III genomes. For *per*, the estimated θ of *si*III/*w*Ma (0.0169, 0.0101–0.0244) overlapped with the *si*III/*w*− lines (0.0201, 0.0138–0.0263).

Neutrality tests: Tajima's *D* and Fu's *Fs* did not depart from neutral expectations for *si*II mtDNA at each locality (Table 2). Exclusion of *si*II/*w*Ri lines from Dar es Salaam ($D = -0.398$, $Fs = 1.378$) and Malindi ($D = -0.664$, $Fs = -1.298$) did not change these interpreta-

TABLE 2

Measures of mtDNA and *per* polymorphism

	Mwanza	Dar es Salaam	Malindi	Nairobi
N	11/NA/11	10/10/20	10/10/20	11/17/28
π	0/NA/0.0399	0.0060/0/0.0434	0.0079/0/0.0357	0.0061/0/0.0138
θ	0/NA/0.0469	0.0060/0/0.0588	0.0090/0/0.0386	0.0058/0/0.0211
Tajima's *D*	NA/NA/−0.7078	−0.0191/NA/−1.0922	−0.6707/NA/−0.2543	0.1874/NA/−1.3241
Fu's *Fs*	NA/NA/−0.2110	0.7130/NA/−3.6100	−2.0500/NA/−1.158	0.9320/NA/−6.3680

Estimates of π (Nei and Li 1979) and θ (Watterson 1975) are per silent/synonymous site. The numbers in each cell refer to estimates from *si*II/*si*III/*per*. Using coalescent simulations, Tajima's *D* was significantly negative for *per* sampled from Mwanza, Dar es Salaam, and Nairobi (indicated by underlining), and all Fu's *F*'s were significantly negative for *per* at the 2% level, which corresponds to a 5% rejection rate (Fu 1997). NA, not applicable (the mtDNA haplogroup either does not occur at that locality or contains no polymorphism).

tions. It was not possible to calculate Tajima's *D* and Fu's *Fs* for *si*III because there was no polymorphism.

For *per*, we estimated the significance of *D* and *F*'s using coalescent simulations so that recombination could be incorporated. Tajima's *D* was significantly negative for Mwanza, Dar es Salaam, and Nairobi, and Fu's *Fs* was significantly negative for all populations (Table 2). It was not clear whether this results from a demographic effect or selection acting on the *per* locus in East Africa. These results contrast with previous studies, where *per* was found to be consistent with a neutral equilibrium model based on Tajima's *D* and the McDonald-Kreitman test (Kliman and Hey 1993; Kliman *et al.* 2000). A goal of future studies is to investigate this result in more detail.

HKA tests: In this study, *per* was chosen *a priori* because it had been shown to be consistent with a neutral model in *D. simulans* (Kliman and Hey 1993; Kliman *et al.* 2000) and to have high levels of silent and synonymous polymorphism compared to other nuclear loci (Verrelli and Eanes 2000). However, the distribution of variation in East African *D. simulans* violated the neutral model according to Tajima's *D* and Fu's *Fs*. To determine if *per* provided an appropriate reference locus to compare against mtDNA, we compared *per* to the *alcohol dehydrogenase-related* locus (*adhr*; data from Ballard 2000a). *Adhr* has previously been shown to conform to a neutral model of molecular evolution in *D. simulans* (Sumner 1991; Ballard *et al.* 1996). This additional test showed that *per* from East Africa yielded polymorphism and divergence consistent with *adhr* (sum of deviations = 0.7919, $P > 0.05$).

Assuming the effective population size of mtDNA is one-third that of *per*, there was not a significant reduction of mtDNA polymorphism in *si*II/*w*− from Dar es Salaam, Malindi, or Nairobi (Table 3). In contrast, there was a significant reduction of mtDNA polymorphism relative to *per* in the Mwanza population (Table 3). The lack of polymorphism among *si*III lines was also significant, whether or not *si*III lines were pooled. We present the pooled result for simplicity (Table 3). About half the population has each mtDNA type at Dar es Salaam, Malindi, and Nairobi. If we assume random nuclear gene flow between flies harboring distinct mtDNAs, then the effective population size of mtDNA may be closer to one-sixth that of *per* in sympatric populations. This additional correction did not change our conclusions (Table 3). Repeating the above results with the *adhr* data supports these conclusions (sum of deviations = 8.22, 12.25, $P < 0.01$ for Mwanza and pooled *si*III, respectively; sum of deviations = 3.74, 2.57, 3.68, $P > 0.05$ for Dar es Salaam, Malindi, and Nairobi, respectively).

Phylogenetic analyses: According to statistical parsimony, only mtDNA sequences separated by <18 steps were connected at 95% confidence. All *si*II sequences were joined by statistical parsimony (Figure 3). The *si*III haplogroup could not be joined to any *si*II sequence with confidence. The *si*II/*w*Ri mtDNAs formed a distinct group separated by at least two steps from any other sequence in the network. Regardless of geography, infected flies clustered together. From the Harare lines collected in 1994, we can infer that the *si*II/*w*Ri mtDNA has been present for at least 7 years. In contrast, the uninfected *si*II sequences from Malindi, Dar es Salaam, and Nairobi are scattered throughout the parsimony network (Figure 3). These data provide additional evidence that the recent maternal ancestors of uninfected flies were not affected by a sweep of mtDNA variation. No uninfected flies carried the *si*II/*w*Ri mtDNA, arguing against imperfect maternal transmission of the infection.

The parsimony limit for *per* was 14 steps. One *per* sequence from Mwanza could not be joined to the network. Qualitatively, the *per* network did not correlate with mtDNA haplotype, Wolbachia infection status, or geography (Figure 4).

DISCUSSION

East Africa is likely the region of endemism for *D. simulans*. Yet few studies have investigated the population subdivision within these ancestral populations (Begun and Aquadro 1993; Hamblin and Veuille

TABLE 3

HKA test comparing intraspecific polymorphism within or to interspecific divergence among *D. simulans* and *D. melanogaster* using silent and synonymous sites

	*si*II								*si*III:	
	Mwanza		Dar es Salaam		Malindi		Nairobi		(pooled)	
	mtDNA	*per*	mtDNA	*per*	mtDNA	*per*	mtDNA	*per*	mtDNA	*per*
Intraspecific polymorphism[a]										
Sample size	11/2	11/2	9/2	9/2	9/2	9/2	11/2	11/2	37/2	37/2
No. of segregating sites (obs)	0/5	46/5	5/5	39/5	8/5	36/5	6/5	46/5	0/5	87/5
No. of segregating sites (exp)	12.13/2.64	33.87/7.36	12.71/2.89	31.29/7.11	13.76/3.13	30.24/6.87	15.92/3.06	36.08/6.94	18.71/2.15	68.29/7.85
Total no. of sites	352.33	335.95	352.18	335.93	352.33	337.06	352.33	336.90	354.60	332.82
Interspecific divergence										
Mean no. of differences (obs)	52.00	35.86	49.44	36.56	50.56	36.28	50.27	35.68	52.50	36.77
Mean no. of differences (exp)	42.23	45.63	43.84	42.16	46.67	40.17	49.29	43.66	36.64	52.63
Total no. of sites	352.33	335.95	352.18	335.95	352.33	337.06	352.33	336.91	354.60	332.82
Sum of deviations (P)[b]										
MtDNA = 1/3 *per*[c]	7.56 (0.03)		2.80 (0.25)		1.48 (0.48)		3.74 (0.14)		15.40 (0.004)	
MtDNA = 1/6 *per*[d]	NA[e]		2.17 (0.36)		1.19 (0.57)		2.35 (0.33)		12.17 (0.03)	

obs, observed; exp, expected.

[a] Values within each cell are reported for *D. simulans*/*D. melanogaster*, respectively.

[b] P, proportion of simulated values that were greater than the observed sum of deviations.

[c] The effective population size of mtDNA was corrected to be one-third that of *per*, which occurs on the X chromosome.

[d] It may be argued that the effective population size of *D. simulans* mtDNA in populations containing both *si*II and *si*III should be one-sixth that of *per* to accommodate sympatry (see text for details).

[e] NA, not applicable, since Mwanza contained only *si*II flies.

1999; Andolfatto 2001). In this study, we investigated population subdivision in the mtDNA of five East African populations with respect to Wolbachia infections. If there has been no prior selection or hitchhiking in the mitochondrial genome, uninfected isofemale lines are expected to exhibit neutral levels of variation compared to an appropriate nuclear locus. We observed a significant reduction in the mtDNA diversity on the basis of the HKA test in the southern, but not the three northern *si*II populations. Positive or background selection may have been mechanistically involved in the removal of mtDNA polymorphism. Alternatively, a Wolbachia-induced population sweep may have reduced mtDNA polymorphism in the southern *si*II populations. If true, mtDNA polymorphism was removed by sharing a common mode of transmission with Wolbachia.

This study answers two major questions. First, there are populations of *D. simulans si*II in East Africa that do not have a significant reduction in mtDNA variation. This is the first description of such populations in *D. simulans*. Second, the *w*Ri and *w*Ma strains of Wolbachia from East Africa do not exist in sympatry, consistent with theoretical predictions that argue against long-term coexistence (Rousset *et al.* 1991; Hoffmann and Turelli 1997). In the process of answering these questions, we laid the groundwork for future monitoring of a region that could potentially replicate the only other study where a Wolbachia-induced population sweep has been observed, that being in California (Turelli and Hoffmann 1991). We next specifically discuss patterns of variation displayed by *si*II and *si*III flies in East Africa.

***si*II:** We report three populations of *D. simulans* with levels of mtDNA polymorphism significantly higher than all previous descriptions (Solignac *et al.* 1986; Baba-Aïssa *et al.* 1988; Hale and Hoffmann 1990; Ballard and Kreitman 1994; Rand *et al.* 1994; Ballard *et al.* 1996; Ballard 2000a; James *et al.* 2002). In all past studies of this species, mitochondrial variation has apparently been reduced by positive or background selection or by any maternally inherited factor such as the bacterial symbiont Wolbachia. We infer that the maternal ancestors of *si*II/*w*− flies in the three northern populations of Dar es Salaam, Malindi, and Nairobi were not affected by a sweep of mtDNA polymorphism in the detectable past. This inference is supported by (1) their significantly greater number of segregating sites compared to *si*II/*w*Ri flies, (2) their consistency with the neutral predictions of HKA tests, and (3) the scattering of *si*II/*w*− throughout the mtDNA genealogy. Interestingly, infected populations to the south have reduced mtDNA variation, probably the result of a Wolbachia-induced population sweep. It is not clear if the infected *si*II/*w*Ri mtDNA sequence associated with *w*Ri randomly or if there are functional reasons underlying the association between particular mitochondria and Wolbachia strains. However, no *w*Ri-infected flies were found with *si*III mtDNA or *w*Ma-infected flies with *si*II mtDNA.

What has prevented the *w*Ri strain from sweeping northward through these populations? Once the frequency of *w*Ri-infected individuals reaches 8–19%, it should induce a sweep even if *w*Ma is present (Turelli and Hoffmann 1995). This sweep should be accompanied by a reduction in mtDNA polymorphism as the population becomes monomorphic for the infected mtDNA. Temporal data suggest that the *w*Ri infection has been fixed in the southern populations for at least 8 years. The *w*Ri strain was fixed in lines collected in 1994 from Harare. Turelli and Hoffmann (1995) found that 73 of 76 isofemale lines collected in Zimbabwe in 1993 were infected. These latter lines were probably infected with *w*Ri because the only other Wolbachia strain collected here, *w*Ma, does not achieve high infection frequencies (Merçot and Poinsot 1998; James and Ballard 2000). From these previous studies, we assume that the Victoria Falls population has been infected with *w*Ri since at least 1993, a span of 8 years. This assumption is conservative and represents a minimum estimate since we are unaware of collections made before 1993. Turelli and Hoffmann (1991) calculated that *w*Ri infections spread an average of 13.5 km per generation. If we employ the same parameter estimates from their study, we would expect the infection to have moved ~1800 km over 8 years from Victoria Falls. Dar es Salaam is ~1500 km north of Victoria Falls, suggesting that the *w*Ri infection should have reached this population at high frequency. At least four alternative hypotheses exist to explain why *w*Ri has not swept through these northern populations. First, biogeographic barriers such as the Selous and Ruaha game reserves and the Udzungwa Mountains (Figure 1) may inhibit migration of *w*Ri-infected individuals into the three northern populations over time. Second, it is possible that the northern populations are resistant to *w*Ri-induced population sweep, perhaps by harboring unique immunity genes. Third, *w*Ri-infected individuals may be selected against in these northern regions. Fourth, *w*Ri-infected mothers may give rise to uninfected progeny at high frequency in East Africa. This fourth alternative seems least likely, given that we have never observed the *si*II/*w*Ri mtDNA in uninfected individuals.

***si*III:** The *si*III flies found in East Africa have no detectable mtDNA polymorphism. This reduction in variation is likely to be related to positive or background selection rather than to the current Wolbachia infection. Both *w*Ma-infected and uninfected flies have reduced variation, and in the laboratory *w*Ma does not induce strong incompatibility (James and Ballard 2000). In nature, the *w*Ma strain may induce a population sweep if it grants a fitness advantage to infected females and/or has high transmission fidelity that compensates for its inability to induce strong incompatibility (Turelli and Hoffmann 1995). A fitness advantage is

associated with some Wolbachia infections (Hoffmann *et al.* 1998; Dobson *et al.* 2002; Fry and Rand 2002). However, little is known about the wMa infection in nature.

It is more likely that selection has acted directly on a beneficial mutation arising in the *si*III haplotype. We are currently investigating the fitness of *si*III mtDNA with population cage studies and biochemical assays of mitochondrial metabolism. If *si*III mitochondria possess a recently derived beneficial mutation(s), it may be expected that the frequency of *si*III flies increases in regions where they are sympatric with the *si*II/*w*− cytotype, as in Dar es Salaam and Malindi.

Future: Turelli and Hoffmann (1995) modernized the initial theoretical achievements of Caspari and Watson (1959) to describe the dynamics of Wolbachia-induced population sweeps in California populations (Turelli and Hoffmann 1991). The generality and accuracy of these models is uncertain. In East Africa, we have a potential paradox that is not explained by current theoretical models. In spite of migration of *w*Ri-infected individuals, the northern populations have avoided population sweeps. Several hypotheses were proposed here to explain the high amount of mtDNA polymorphism in the three northern populations. Perhaps most interesting among them is the possibility that host genetic factors, such as specific immunity genes, might suppress infection frequencies. Current theoretical models do not take into account host factors. The mitochondrial genome, which is essentially linked to particular Wolbachia strains, should play an important role in the overall fitness of that cytotype and affect population dynamics of the spread of infected cytotypes. Such models could be developed and tested using the populations of East Africa.

Koen Maes and the staff at the National Museum of Kenya provided informative discussions of geography in the area and significant help with collecting. Kingsley Wallani, Ernest Zaranyika, Greenwell Nyirenda, Christine Meena, and Gowele Mtoka assisted with collecting. Chip Aquadro provided the Harare lines and Dana Kurpius assisted with sequencing. Ary Hoffmann, Avis James, and two anonymous reviewers made constructive comments on the manuscript. Josep Comeron, Jody Hey, and Martin Kreitman discussed alternative interpretations of the HKA results. All molecular work was carried out in the Roy J. Carver Center for Comparative Genomics at the University of Iowa. Funding was provided by National Science Foundation grant no. DEB-9702824.

LITERATURE CITED

Andolfatto, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. Mol. Biol. Evol. **18:** 279–290.

Andolfatto, P., and M. Przeworski, 2001 Regions of lower crossing over harbor more rare variants in African populations of *Drosophila melanogaster*. Genetics **158:** 657–665.

Baba-Aïssa, F., M. Solignac, N. Dennebouy and J. R. David, 1988 Mitochondrial DNA variability in *Drosophila simulans*: quasi absence of polymorphism within each of the three cytoplasmic races. Heredity **61:** 419–426.

Ballard, J. W. O., 2000a Comparative genomics of mitochondrial DNA in *Drosophila simulans*. J. Mol. Evol. **51:** 64–75.

Ballard, J. W. O., 2000b Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. J. Mol. Evol. **51:** 48–63.

Ballard, J. W. O., 2004 Sequential evolution of a symbiont inferred from the host: *Wolbachia* and *Drosophila simulans*. Mol. Biol. Evol. (in press).

Ballard, J. W. O., and M. Kreitman, 1994 Unraveling selection in the mitochondrial genome of Drosophila. Genetics **138:** 757–772.

Ballard, J. W. O., J. Hatzidakis, T. L. Karr and M. Kreitman, 1996 Reduced variation in *Drosophila simulans* mitochondrial DNA. Genetics **144:** 1519–1528.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **365:** 548–550.

Caspari, E., and G. S. Watson, 1959 On the evolutionary importance of cytoplasmic sterility in mosquitoes. Evolution **13:** 568–570.

Charlesworth, B., 1996 Background selection and patterns of genetic diversity in *Drosophila melanogaster*. Genet. Res. **68:** 131–149.

Citri, Y., H. V. Colot, A. C. Jacquier, Q. Yu, J. C. Hall *et al.*, 1987 A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene. Nature **326:** 42–47.

Clement, M., D. Posada and K. A. Crandall, 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:** 1657–1659.

Coyne, J. A., 1983 Genetic basis of differences in genital morphology among the three sibling species of *Drosophila*. Evolution **37:** 1101–1117.

Coyne, J. A., 1985 Genetic studies of three sibling species of *Drosophila* with relationship to theories of speciation. Genet. Res. **46:** 169–192.

de Stordeur, E., 1997 Nonrandom partition of mitochondria in heteroplasmic *Drosophila*. Heredity **79:** 615–623.

de Stordeur, E., M. Solignac, M. Monnerot and J. C. Mounolou, 1989 The generation of transplasmic *Drosophila simulans* by cytoplasmic injection: effects of segregation and selection on the perpetuation of mitochondrial DNA heteroplasmy. Mol. Gen. Genet. **220:** 127–132.

Dobson, S. L., E. J. Marsland and W. Rattanadechakul, 2002 Mutualistic Wolbachia infection in *Aedes albopictus*: accelerating cytoplasmic drive. Genetics **160:** 1087–1094.

Fay, J. C., and C. I. Wu, 2000 Hitchhiking under positive Darwinian selection. Genetics **155:** 1405–1413.

Fry, A. J., and D. M. Rand, 2002 *Wolbachia* interactions that determine *Drosophila melanogaster* survival. Evolution **56:** 1976–1981.

Fu, Y. X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147:** 915–925.

Hale, L. R., and A. A. Hoffmann, 1990 Mitochondrial DNA polymorphism and cytoplasmic incompatibility in natural populations of *Drosophila simulans*. Evolution **44:** 1383–1386.

Hamblin, M. T., and C. F. Aquadro, 1996 High nucleotide sequence variation in a region of low recombination in *Drosophila simulans* is consistent with the background selection model. Mol. Biol. Evol. **13:** 1133–1140.

Hamblin, M. T., and M. Veuille, 1999 Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture. Genetics **153:** 305–317.

Hoffmann, A. A., and M. Turelli, 1988 Unidirectional incompatibility in *Drosophila simulans*: inheritance, geographic variation and fitness effects. Genetics **119:** 435–444.

Hoffmann, A. A., and M. Turelli, 1997 Cytoplasmic incompatibility in insects, pp. 42–80 in *Influential Passengers*, edited by S. L. O'Neill, A. A. Hoffmann and J. H. Werren. Oxford University Press, Oxford/New York/Tokyo.

Hoffmann, A. A., M. Turelli and L. G. Harshman, 1990 Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. Genetics **126:** 933–948.

Hoffmann, A. A., M. Hercus and H. Dagher, 1998 Population dynamics of the Wolbachia infection causing cytoplasmic incompatibility in *Drosophila melanogaster*. Genetics **148:** 221–231.

Hudson, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

Hudson, R. R., M. Kreitman and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Hutter, C. M., M. D. Schug and C. F. Aquadro, 1998 Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: a reciprocal test of the ascertainment bias hypothesis. Mol. Biol. Evol. **15:** 1620–1636.

Irvin, S. D., K. A. Wetterstrand, C. M. Hutter and C. F. Aquadro, 1998 Genetic variation and differentiation at microsatellite loci in *Drosophila simulans*: evidence for founder effects in new world populations. Genetics **150:** 777–790.

James, A. C., and J. W. O. Ballard, 2000 Expression of cytoplasmic incompatibility in *Drosophila simulans* and its impact on infection frequencies and distribution of *Wolbachia pipientis*. Evolution **54:** 1661–1672.

James, A. C., M. D. Dean, M. E. McMahon and J. W. O. Ballard, 2002 Dynamics of double and single *Wolbachia* infections in *Drosophila simulans* from New Caledonia. Heredity **88:** 182–189.

Jensen, M. A., B. Charlesworth and M. Kreitman, 2002 Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster* and *D. simulans*. Genetics **160:** 493–507.

Kim, Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics **155:** 1415–1427.

Kliman, R. M., and J. Hey, 1993 DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics **133:** 375–387.

Kliman, R. M., P. Andolfatto, J. A. Coyne, F. Depaulis, M. Kreitman *et al.*, 2000 The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics **156:** 1913–1931.

Kondo, R., Y. Satta, E. T. Matsuura, H. Ishikawa, N. Takahata *et al.*, 1990 Incomplete maternal transmission of mitochondrial DNA in Drosophila. Genetics **126:** 657–663.

Kondo, R., E. T. Matsuura and S. I. Chigusa, 1992 Further observation of paternal transmission of *Drosophila* mitochondrial DNA by PCR selective amplification method. Genet. Res. **59:** 81–84.

Kreitman, M., 1991 Molecular techniques in taxonomy, pp. 357–367 in *NATO ASI Series*, edited by G. M. Hewitt, A. W. B. Johnson and J. P. W. Young. Springer-Verlag, Berlin.

Lachaise, D., M. L. Cariou, J. R. David, F. Lemeunier, L. Tsacas *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22:** 159–225.

Matsuura, E. T., H. Fukuda and S. I. Chigusa, 1991 Mitochondrial DNA heteroplasmy maintained in natural populations of *Drosophila simulans* in Reunion. Genet. Res. **57:** 123–126.

Merçot, H., and D. Poinsot, 1998 …and discovered on Mount Kilimanjaro. Nature **391:** 853.

Nei, M., and W. H. Li, 1979 Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76:** 5269–5273.

Nigro, L., and T. Prout, 1990 Is there selection on RFLP differences in mitochondrial DNA? Genetics **125:** 551–555.

O'Neill, S. L., and T. L. Karr, 1990 Bidirectional incompatibility between conspecific populations of *Drosophila simulans*. Nature **348:** 178–180.

O'Neill, S. L., R. Giordano, A. M. Colbert, T. L. Karr and H. M. Robertson, 1992 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. Proc. Natl. Acad. Sci. USA **89:** 2699–2702.

Posada, D., and K. A. Crandall, 2001 Intraspecific gene genealogies: trees grafting into networks. Trends Ecol. Evol. **16:** 37–45.

Rand, D. M., M. Dorfsman and L. M. Kann, 1994 Neutral and nonneutral evolution of Drosophila mitochondrial DNA. Genetics **138:** 741–756.

Rousset, F., M. Raymond and F. Kjellberg, 1991 Cytoplasmic incompatibilities in the mosquito *Culex pipiens*: How to explain a cytotype polymorphism? J. Evol. Biol. **4:** 69–81.

Rozas, J., and R. Rozas, 1997 DnaSP version 3.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13:** 307–311.

Satta, Y., N. Toyohara, C. Ohtaka, Y. Tatsuno, K. Watanabe *et al.*, 1988 Dubious maternal inheritance of mitochondrial DNA in *D. simulans* and evolution of *D. mauritiana*. Mol. Gen. Genet. **52:** 1–6.

Solignac, M., M. Monnerot and J. C. Mounolou, 1986 Mitochondrial DNA evolution in the melanogaster species subgroup of *Drosophila*. J. Mol. Evol. **23:** 31–40.

Sumner, C. J., 1991 Nucleotide polymorphism in the *alcohol dehyrogenase duplicate* locus of *Drosophila simulans*: implications for the neutral theory. Honors Thesis, Princeton University, Princeton, NJ.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Templeton, A. R., K. A. Crandall and C. F. Sing, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

Turelli, M., and A. A. Hoffmann, 1991 Rapid spread of an inherited incompatibility factor in California *Drosophila*. Nature **353:** 440–442.

Turelli, M., and A. A. Hoffmann, 1995 Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. Genetics **140:** 1319–1338.

Turelli, M., and A. A. Hoffmann, 1999 Microbe-induced cytoplasmic incompatibility as a mechanism for introducing transgenes into arthropod populations. Insect Mol. Biol. **8:** 243–255.

Turelli, M., A. A. Hoffmann and S. W. Mckechnie, 1992 Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. Genetics **132:** 713–723.

Verrelli, B. C., and W. F. Eanes, 2000 Extensive amino acid polymorphism at the *Pgm* locus is consistent with adaptive protein evolution in *Drosophila melanogaster*. Genetics **156:** 1737–1752.

Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 256–276.

Zhou, W., F. Rousset and S. L. O'Neill, 1998 Phylogeny and PCR-based classification of *Wolbachia* strains using *wsp* gene sequences. Proc. R. Soc. Lond. Ser. B: Biol. Sci. **265:** 509–515.

Communicating editor: S. W. Schaeffer