

2011

Analysis of the evolutionary forces shaping mitochondrial genomes of a Neotropical malaria vector complex

Jaroslav Krzywinski

Vector Group, Liverpool School of Tropical Medicine, Liverpool, jarek@liv.ac.uk

Cong Li

Division of Entomology, Walter Reed Army Institute of Research, Silver Spring, MD

Marion Morris

Vector Group, Liverpool School of Tropical Medicine, Liverpool

Jan E. Conn

Wadsworth Center, Griffin Laboratory, New York State Department of Health

José B. Lima

Department of Entomology, Oswaldo Cruz Foundation

See next page for additional authors

Follow this and additional works at: <http://digitalcommons.unl.edu/usarmyresearch>

Krzywinski, Jaroslav; Li, Cong; Morris, Marion; Conn, Jan E.; Lima, José B.; and Povoia, Marinete M., "Analysis of the evolutionary forces shaping mitochondrial genomes of a Neotropical malaria vector complex" (2011). *US Army Research*. 335.

<http://digitalcommons.unl.edu/usarmyresearch/335>

This Article is brought to you for free and open access by the U.S. Department of Defense at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in US Army Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Authors

Jaroslav Krzywinski, Cong Li, Marion Morris, Jan E. Conn, José B. Lima, and Marinete M. Pova



Analysis of the evolutionary forces shaping mitochondrial genomes of a Neotropical malaria vector complex

Jaroslaw Krzywinski^{a,*}, Cong Li^{b,c}, Marion Morris^a, Jan E. Conn^{d,e}, José B. Lima^f, Marinete M. Pova^g, Richard C. Wilkerson^{b,c}

^a Vector Group, Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK

^b Division of Entomology, Walter Reed Army Institute of Research, Silver Spring, MD 20910-7500, USA

^c Walter Reed Biosystematics Unit, WRAIR, Museum Support Center, MRC 534, Smithsonian Institution, 4210 Silver Hill Road, Suitland, MD 20746-2863, USA

^d Wadsworth Center, Griffin Laboratory, New York State Department of Health, 5668 State Farm Road, Slingerlands, NY 12159, USA

^e Department of Biomedical Sciences, School of Public Health, State University of New York-Albany, Empire Plaza, Albany, NY 12201, USA

^f Department of Entomology, Oswaldo Cruz Foundation, 4365, Manguinhos, Rio de Janeiro, Brazil

^g Laboratório de Malária, Seção de Parasitologia, Instituto Evandro Chagas, BR-316, km 7 s/n, B. Levilândia, 67030-000 Ananindeua, PA, Brazil

ARTICLE INFO

Article history:

Received 27 July 2010

Revised 22 December 2010

Accepted 11 January 2011

Available online 15 January 2011

Keywords:

Comparative genomics

Mitochondrial genome evolution

Replication origin

Introgressive hybridization

Malaria vectors

ABSTRACT

Many vectors of human malaria belong to complexes of morphologically indistinguishable cryptic species. Here we report the analysis of the newly sequenced complete mitochondrial DNA molecules from six recognized or putative species of one such group, the Neotropical *Anopheles albitarsis* complex. The molecular evolution of these genomes had been driven by purifying selection, particularly strongly acting on the RNA genes. Directional mutation pressure associated with the strand-asynchronous asymmetric mtDNA replication mechanism may have shaped a pronounced DNA strand asymmetry in the nucleotide composition in these and other *Anopheles* species. The distribution of sequence polymorphism, coupled with the conflicting phylogenetic trees inferred from the mitochondrial DNA and from the published *white* gene fragment sequences, indicates that the evolution of the complex may have involved ancient mtDNA introgression. Six protein coding genes (*nad5*, *nad4*, *cox3*, *atp6*, *cox1* and *nad2*) have high levels of sequence divergence and are likely informative for population genetics studies. Finally, the extent of the mitochondrial DNA variation within the complex supports the notion that the complex consists of a larger number of species than until recently believed.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Anophelines are the only mosquitoes capable of transmitting human malaria, a parasitic disease that kills up to 3 million people and debilitates hundreds of millions every year. Approximately half of the medically important vector species belong to cryptic species complexes whose morphologically indistinguishable members often vary in their feeding habits, which, ranging from anthropophily to zoophily, define their vector or non-vector status (Service and Townson, 2002). Existence of such cryptic species complexes poses serious challenges to the routine identification of vectors and, in the case of sympatric occurrence, to determining the role of each species in malaria transmission, which in turn affects rational approaches to controlling the disease. In addition, the apparent recent species divergences complicate understanding of the evolutionary relationships, the reproductive isolation, and potential gene flow among different evolutionary units within the

complexes (Besansky et al., 2003). This is not merely an academic problem, because “leaky” species boundaries may facilitate inter-specific transfer of genes or alleles important in malaria epidemiology.

The Neotropical *Anopheles* (*Nyssorhynchus*) *albitarsis* complex consists of six recognized species (Motoki et al., 2009); however, it continues to represent a taxonomic puzzle with a hitherto unclear number of constituent taxa. Moreover, the relationships among them are poorly understood, because the phylogenetic trees inferred with different molecular markers have confusingly conflicting topologies (Bourke et al., 2010; Brochero et al., 2007; Lehr et al., 2005; Merritt et al., 2005; Wilkerson et al., 2005). Members of this widely distributed complex are behaviorally and morphologically variable, and their reliable identification requires application of molecular markers. Originally, RAPD-PCR was used to recognize four species closely related to *An. albitarsis* (Wilkerson et al., 1995). An additional species was detected during a study attempting to correlate the RAPD-PCR banding patterns with the mitochondrial *cox1* gene sequence variation (Lehr et al., 2005), and a sixth species was recognized based on the analysis of the

* Corresponding author. Fax: +44 151 705 3369.

E-mail address: jarek@liv.ac.uk (J. Krzywinski).

nuclear ribosomal DNA fragment diversity (Brochero et al., 2007). Unexpectedly, but in part consistent with earlier indications (Lehr et al., 2005; Wilkerson et al., 2005; McKeon et al., 2010), the most recent large scale study of the mitochondrial *cox1* barcode region suggested existence of three new putative members within the complex (J. Ruiz, Y.M. Linton, J.E. Conn, N.S. McKeon, and R.C. Wilkerson; unpublished). Further research is required to understand their taxonomic status.

Three species of the complex are important regional malaria vectors and two are suspected vectors (Branquinho et al., 1993; references in Conn and Mirabello, 2007); the role of the other members in malaria transmission awaits elucidation. This is not an easy task, since no simple one-tube PCR assay is available for the identification of even the six species reported in Motoki et al. (2009). The ribosomal ITS2 region has been successfully targeted in multiplex PCR (with mixtures of a universal and multiple species-specific primers that amplify products sufficiently different in length for each species to be unequivocally discriminated on an agarose gel) for species recognition in various anopheline cryptic species complexes (Krzywinski and Besansky, 2003). Unfortunately, a small number of species-specific differences and intragenomic variation within the ITS2 sequences preclude use of such a strategy for the *An. albitarsis* complex (Li and Wilkerson, 2005, 2007). An alternative high copy DNA marker with fixed interspecific variation sufficient to discriminate between species within the complex would greatly facilitate further studies on the distribution of species and their contribution to malaria transmission.

In most animal taxa the mitochondrial genome is a small circular molecule with a densely packed set of 13 protein coding genes (PCGs), 22 tRNA genes, and two rRNA genes (Boore, 1999). Coding content conservation, maternal inheritance, high mutation rate, presence of variable regions adjacent to highly conserved sites suitable for primer design, and ease of amplification associated with a high copy number within the cell make the mitochondrial DNA (mtDNA) the molecule of choice for evolutionary studies. It has been widely utilized in the analyses of patterns of molecular evolution and as a marker for phylogeographic and phylogenetic inference (Gissi et al., 2008). In mosquitoes, mtDNA sequences have been used to resolve relationships at various levels of divergence (Krzywinski and Besansky, 2003; Krzywinski et al., 2001; Loaiza et al., 2010) and to estimate dates of branching of major lineages (Krzywinski et al., 2006; Moreno et al., 2010), but also used for PCR-based identification of species belonging to cryptic species complexes (Dusfour et al., 2007; Goswami et al., 2006).

To date, mitochondrial genomes of four distantly related species of *Anopheles* have been published (Beard et al., 1993; Krzywinski et al., 2006; Mitchell et al., 1993; Moreno et al., 2010). Here we report complete sequences of the mitochondrial DNA molecules from five recognized species of the *An. albitarsis* complex and from one of the new putative species, *An. albitarsis* G, whose existence was suggested earlier (Lehr et al., 2005; Wilkerson et al., 2005; McKeon

et al., 2010). The objective of this study was to obtain insight into the evolutionary history of the complex, to evaluate the extent of interspecific divergence within mtDNA of its members, and to explore the mechanisms of molecular evolution of mtDNA operating in mosquitoes at low levels of divergence. Our analyses indicated that the molecular evolution of the mtDNA had been driven by strong purifying selection and that the evolution of the complex may have involved ancient mtDNA introgression. Comparisons of the genomes supported the status of *An. albitarsis* G as a distinct species and revealed highly variable regions that are likely informative for phylogenetic and phylogeographic studies of this mosquito group. Information about the levels of interspecific mtDNA divergence may facilitate the design of PCR-based assay targeting mtDNA that would allow identification of members of the *An. albitarsis* complex. Finally, the results of sequence analyses led us to hypothesize about a likely mechanism of the mtDNA replication in mosquitoes.

2. Materials and methods

2.1. Samples

Morphological characters from Linthicum (1988) were used for identification of *An. (Nys.) albitarsis* s.l. Species within the complex were determined as described in (Wilkerson et al., 1995) and on the basis of a sequence variation within the *cox1* barcoding region. Individuals from progeny broods preserved in 100% ethyl alcohol maintained at -70°C were used for the study. Voucher specimens from each brood, including DNA, have been deposited in the Smithsonian Institution, National Museum of Natural History (NMNH). Species whose complete mtDNA sequence is reported for the first time in this study are listed in Table 1. Sequences of the *white* gene fragment were obtained from GenBank (accession numbers AY956297–AY956302, DQ228314, DQ906919) or obtained in this study (*An. albitarsis* G from Manaus, Amazonas state, Brazil, BR026–36; accession number HQ616606).

2.2. Experimental procedures

DNA was extracted from individual adult mosquitoes by phenol-chloroform extraction as described (Wilkerson et al., 1993). PCR primers were designed based on the sequence comparisons between complete mitochondrial genomes of *Anopheles gambiae* (NC_002084), *Anopheles quadrimaculatus* (NC_000875), *Anopheles funestus* (NC_008070), and *Aedes albopictus* (NC_006817), and between mtDNA sequences from members of the *An. albitarsis* complex. A list of all primers used in the study is presented in Supplementary file 1. Mitochondrial DNA fragments were amplified in 50 μl containing 50 mM KCl, 10 mM Tris–HCl, pH 8.3, 1.5 mM MgCl₂, 0.2 mM each dNTP, 2.5 U *Taq* polymerase (Invitrogen),

Table 1
Species from the *An. albitarsis* complex used in this study and source of specimens whose mtDNA has been sequenced.

Species	Collection No.	Locality	Coordinates	GenBank accession No.
<i>An. albitarsis</i> s.s.	BR501(37)	São Paulo, 6 km SW Registro, Brazil	24°36.8'S, 47°53.1'W	HQ335344
	BR511(3) ^a	Paraná, near Guaira, Brazil	24°4'S, 54°15'W	
<i>An. oryzalimnetes</i>	BR019(6)	Ceara, Paraipaba, Brazil	3°25'S, 39°13'W	HQ335345
	BR/R015(1) ^a	Bahia Itaquara, Brazil	13°26'S, 39°56'W	
<i>An. deaneorum</i>	BR700(1)	Rondonia, Ariquemes, Brazil	9°56'S, 63°04'W	HQ335347
	BR700(6) ^a			
<i>An. janconnae</i>	BR36-2	Boa Vista, Roraima, Brazil	1°45.4'N, 61°9.1'W	HQ335348
	BV29170 ^a		2°49'N, 60°40'W	
<i>An. albitarsis</i> F	CV9-7	Vichada, Colombia	6°11.9'N, 67°28.8'W	HQ335349
<i>An. albitarsis</i> G	BR026(12)	Amazonas, Manaus, Brazil	2°53'S, 60°15'W	HQ335346

^a Used to close several short sequence gaps.

10–25 pmol each primer, and 1 µl template DNA (1/100th of the DNA extracted from a single mosquito). Occasionally, for the amplification of more difficult templates, containing long runs of adenines or thymines, Takara Ex *Taq* (Takara) was used according to the manufacturer's recommendation. PCR thermal cycling included 1 min initial denaturation at 94 °C, followed by 35–40 cycles of 40 s at 94 °C, 45 s at 48–56 °C, and 1–3 min at 68 °C, and a final elongation for 10 min at 72 °C. The nuclear *white* gene fragment was amplified as described (Besansky and Fahey, 1997). PCR products were purified using either ExoSAP-IT (GE Healthcare) or QIAquick Gel Extraction Kit (Qiagen), according to manufacturers' protocols, and sequenced directly with primers used for the PCR. Sequencing was performed using ABI BigDye terminator chemistry (Perkin–Elmer Applied Biosystems) on an ABI 3700 and 3130xl Genetic Analyzer. Sequences were assembled and verified by visual inspection of electropherograms using Sequencher v. 4.1 (Gene Codes Corp.).

2.3. Sequence analysis

Annotation of the mitochondrial genomes of the *An. albitarsis* complex species was completed using sequence comparisons to the mtDNA genes of *A. gambiae* (Genbank accession number NC_002084) and *An. funestus* (NC_008070). The program tRNAscan-SE (Lowe and Eddy, 1997) was used to additionally verify the identity of tRNA genes and to determine the secondary structures of their products. All the tRNA genes were identified using tRNAscan-SE, with one exception: tRNA^{Ser(AGN)}. This particular tRNA does not fold into a typical cloverleaf structure due to a missing D-stem, a common feature of the animal tRNA^{Ser(AGN)}; its secondary structure was determined by a comparison to the tRNA^{Ser(AGN)} consensus structure proposed for *Drosophila* (Montooth et al., 2009). Sequences of the entire newly obtained genomes were aligned with the mtDNA genomes of *Anopheles darlingi* (GQ91 8273), *An. gambiae*, *An. quadrimaculatus* (NC_000875) and *Ae. aegypti* (NC_010241) using ClustalX (Thompson et al., 1997). In addition, sequences of individual genes were separately aligned and the alignments of genes belonging to the same class (rRNAs, tRNAs, protein coding) were concatenated for further analyses. Sequence statistics, pairwise divergence, and numbers of synonymous and non-synonymous substitutions were calculated using MEGA v. 4.0 (Tamura et al., 2007). The secondary structure models of the rRNA molecules were derived from the European Ribosomal RNA database (RDB) (<http://bioinformatics.psb.ugent.be/webtools/rRNA/>) and the Comparative RNA Web Site (<http://www.rna.cccb.utexas.edu/>). The models obtained from the RDB were visualized using RnaViz 2.0 (De Rijk et al., 2003). The models served as a guide to identify the positions in the rRNA genes alignment that correspond to the stem (paired bases) and unpaired (referred to as loop) regions of the rRNAs. Strand asymmetry was measured using the formulas AT-skew = $(A\% - T\%)/(A\% + T\%)$ and GC-skew = $(G\% - C\%)/(G\% + C\%)$ (Perna and Kocher, 1995). Sliding window analysis was performed using DnaSP v. 5.10.01 (Librado and Rozas, 2009), with a window size of 200 bp and a step size of 25 bp.

2.4. Phylogenetic analyses

Phylogenetic analyses were done using aligned concatenated nucleotide sequences of all mtDNA-encoded protein coding genes. Because various optimality criteria are based on different assumptions that may influence phylogenetic inference, maximum parsimony (MP), neighbor joining (NJ) maximum likelihood (ML), and Bayesian analysis were used in our study. The MP, NJ and ML were carried out in PAUP v. 4.0b10 (Swofford, 2002) using heuristic searches. The analyses were done with 1000 replications by stepwise random addition of taxa and tree-bisection reconnection branch swapping; confidence in the inferred topologies was

estimated by bootstrapping (500–1000 bootstrap pseudoreplicates, each with 10 random additions of sequences). The best-fit model (GTR + I + G) of DNA substitution used in the NJ and ML analyses was determined by the Akaike information criterion using jModeltest 0.1.1 (Posada, 2008). Bayesian inference was conducted using MrBayes v. 3.1.2 (Ronquist and Huelsenbeck, 2003) with the previously specified model of sequence evolution, but allowing the estimation of parameters separately for the 3rd codon position. All other priors were set to default values. Two independent runs with one cold and three incrementally heated chains were carried out for 3 million generations, with trees sampled every 100 generations (in addition, four independent runs with a setup as above were performed, leading to identical results as the two-run analysis). Bayesian posterior probabilities were estimated after discarding the first 25% of the retained trees as “burn-in”, the size of which was determined by checking convergence of posterior clade probabilities using the program AWTY (Nylander et al., 2008). The newly generated and published sequences of the nuclear *white* gene fragment (Merritt et al., 2005; Brochero et al., 2007) were analyzed using the same approaches as the mtDNA sequences, but with the TVM + G as the optimal model of sequence evolution selected using jModeltest. The Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999) as applied in PAUP was used to compare the likelihood scores of the ML tree with suboptimal trees found under constrained searches. The test was performed using resampling estimated log-likelihood bootstrap with 10,000 replicates. Based on the inferred ML tree, the ancestral nucleotide states were reconstructed at each node using PAUP and then the derived states within the *An. albitarsis* complex were assigned. Synapomorphies (shared derived nucleotide states) were mapped on the nodes using TNT v. 1.1 (Goloboff et al., 2008). The non-conservative amino acid substitutions were mapped onto predicted transmembrane protein topologies derived with THMM (Krogh et al., 2001). The protein coding gene-based tree topology was used to infer the ancestral tRNA sequences for each node within the *An. albitarsis* complex. Analyses of all genes were performed using the coding (sense) strand.

3. Results and discussion

3.1. Genome structure and nucleotide composition

Complete mtDNA genomes of the six identified members of the *An. albitarsis* complex have been sequenced (GenBank accession numbers HQ335344–9). The molecules vary between 15,413 and 15,474 bp in length, with most of the size heterogeneity within the AT-rich control region (Table 2). The gene content, with 37 genes, is typical of metazoan taxa (Boore, 1999), and the gene order and orientation is as described earlier in *Anopheles* (Beard et al., 1993; Krzywinski et al., 2006; Mitchell et al., 1993). The PCGs have the same lengths in all six species, while lengths of the tRNA and rRNA genes vary by up to 2 and 5 nt, respectively. Apart from the A + T rich region, the intergenic sequences comprise 51–68 nucleotides.

The newly sequenced mitochondrial genomes are almost identical to each other in nucleotide composition. Overall, the A + T content ranges between 77.1% and 77.4%, but within the PCGs it is much higher in the third codon positions, where it approaches the extremes (92.3–94.1%) found in the A + T rich region (Table 2 and Supplementary file 4). The nucleotide proportions are asymmetrically distributed on the two mtDNA strands (in this study the strand nomenclature is adopted from the *Drosophila* literature: the strand encoding the majority of genes is called the major strand, or J-strand, and the other – the minor strand, or N-strand). The asymmetry can be measured by AT and GC skew (Perna and

Table 2
Lengths and A + T composition of structural features of the *An. albitarsis* complex mitochondrial genomes.

	Length						% A + T							
	Overall	PCG	tRNA	rRNA	Intergenic	AT-rich ^a	Overall	Protein coding genes			tRNA	rRNA	AT-rich ^a	
								Overall	1st	2nd				3rd
<i>An. albitarsis</i>	15,413	11,216	1478	2120	53	575	77.1	76.1	69.6	66.7	91.9	78.7	81.2	93.2
<i>An. oryzalimnetes</i>	15,421	11,216	1477	2120	54	580	77.1	76.2	69.6	66.8	92.1	78.7	81.2	94.1
<i>An. albitarsis</i> G	15,474	11,216	1476	2125	68	615	77.1	76.0	69.4	66.7	92.0	78.7	81.6	92.6
<i>An. deaneorum</i>	15,424	11,216	1476	2121	57	581	77.2	76.2	69.4	66.6	92.6	78.3	81.4	92.3
<i>An. janconnae</i>	15,425	11,216	1477	2120	62	575	77.1	76.0	69.7	66.8	91.7	78.4	81.5	92.4
<i>An. albitarsis</i> F	15,418	11,216	1478	2121	51	578	77.4	76.3	69.7	66.8	92.5	78.7	81.6	92.9

^a AT-rich control region.

Kocher, 1995), which assumes positive values when A% and G% are higher than T% and C%, respectively, and negative values when the proportions are reversed. Analysis of the asymmetry in the PCGs reveals a lower content of As than Ts on both strands, but a bigger difference (more negative values of AT skew) on the N-strand (Table 3). In contrast, the values of the GC skew are positive for the N-strand and negative for the J-strand; the disparity between these values is particularly strong in the third codon positions, in which the functional constraints are relaxed (average GC skew across all genes encoded on the N-strand = 0.674, on the J-strand = -0.488; Table 3). Such strand asymmetry is well documented in the animal mitochondrial genomes (Perna and Kocher, 1995) and it has been attributed to a peculiar mode of mtDNA replication. Evidence suggests that in *Drosophila* mtDNA replication follows a strand-asynchronous asymmetric model (Saito et al., 2005), originally proposed for the mammalian mtDNA (Clayton, 1982). According to the model, replication in *Drosophila* initiates from the replication origin of the N-strand (O_N) and proceeds unidirectionally, with a displacement of a parental N-strand, which remains temporarily single stranded; replication of the J-strand begins only after 97% of the N-strand has been synthesized (Goddard and Wolstenholme, 1980). The single stranded DNA is more prone than double stranded DNA to damage and mutations, especially to deaminations C → T and A → G (Frederico et al., 1990; Lindahl, 1993). This observation leads to a prediction that during a single stranded state the N-strand becomes exposed to directional mutation pressure, which should result in an N-strand-specific overabundance of thymines and guanines as compared to adenines and cytosines. Indeed, such patterns have been reported for *Drosophila* and several other insect groups (Ballard, 2000; Carapelli et al., 2008; Montooth et al., 2009;

Oliveira et al., 2008), and observed in this study in members of the *An. albitarsis* complex (Table 3) and four other anopheline species (data not shown).

3.2. A + T rich region

In the newly sequenced mitochondrial genomes the A + T rich region is variable in size (575–615 bp), primarily due to a centrally located indel containing microsatellite-like AT repeats (Supplementary file 2). Upstream from the indel, about 150 bp from the beginning of the A + T rich region (counting from the end of s-rRNA gene), the J-strand harbors a 15–18 bp long T-stretch. The location and length of the T-stretch is conserved in the representatives of the three *Anopheles* subgenera, for which mtDNA sequences are available in GenBank (*An. darlingi*, subgenus *Nyssorhynchus*; *An. funestus* and *A. gambiae*, subgenus *Cellia*; *An. quadrimaculatus*, subgenus *Anopheles*). In addition, a CCCCTA hexamer adjacent to the 5'-end of the T-stretch is present in all anophelines (although not perfectly conserved in culicine mosquitoes and poorly conserved or absent in other dipterans). The remarkable conservation of these sequences across over 100 million years of anopheline evolution (Krzywinski et al., 2006), within a region regarded as the most variable in the whole mitochondrial genome, strongly suggests a functional role (see below).

The A + T rich region in insects is known to contain regulatory sequences responsible for controlling replication and transcription of the mitochondrial genome. In *Drosophila*, O_N has been mapped to the center of the A + T rich region on the major strand, and the origin of replication of the J-strand (O_J) near the tRNA^{lle} gene on the minor strand. In both cases long thymine stretches are located immediately upstream from the origins of replication (Saito et al.,

Table 3
Length, number of variable positions within the alignment, AT and GC skew values, and start/stop codons for the mtDNA-encoded protein coding genes of the *An. albitarsis* complex. The skew values were derived from average nucleotide content calculated for each gene across the six species of the complex.

Gene	Length	Variable/informative positions		All codon positions		3rd Codon positions		Start codon	Stop codon
		Within the complex	Complex – outgroup ^a	AT skew	GC skew	AT skew	GC skew		
<i>Major strand</i>									
ATP6	680	53/28	83/31	-0.126	-0.127	0.034	-0.271	ATG	TA-
ATP8	162	7/2	15/2	-0.103	-0.416	-0.090	-0.656	ATT	TAA
CO1	1537	116/62	189/73	-0.118	-0.013	0.003	-0.418	TCG	T-
CO2	685	30/17	68/22	-0.064	-0.035	-0.020	-0.134	ATG	T-
CO3	787	62/32	100/36	-0.086	-0.041	0.077	-0.506	ATG	T-
cytb	1135	80/34	133/45	-0.129	-0.019	-0.016	-0.532	ATG	T-
ND2	1024	80/40	113/50	-0.096	-0.159	0.066	-0.315	ATT	T-
ND3	352	26/13	45/14	-0.087	-0.139	0.085	-0.860	ATA	T-
ND6	524	25/10	50/17	-0.077	-0.236	0.083	-0.696	ATT	TA-
<i>Minor strand</i>									
ND1	945	64/32	110/33	-0.272	0.277	-0.150	0.659	ATA	TAA
ND4	1342	100/49	182/55	-0.236	0.239	-0.090	0.667	ATG	T-
ND4L	300	20/9	29/10	-0.262	0.281	-0.090	0.723	ATG	TAA
ND5	1741	137/68	214/78	-0.173	0.260	-0.082	0.647	GTG/ATG	T-

^a Number of variable positions when the closest outgroup, *An. darlingi*, was included in comparisons.

2005). The presence of similar T-stretches directly adjacent to the O_N in a lepidopteran and a beetle (Saito et al., 2005), and to the L-strand origin in mammals (Clayton, 1982) indicate that the long thymine homopolymers serve as a structural marker for the recognition of the mtDNA replication origins in a broad range of taxa. Therefore, the T-stretch and the adjacent conserved hexamer sequence found on the J-strand in *Anopheles* most likely signal the replication origin of the minor strand. This finding, combined with the nucleotide composition strand asymmetry predicted by the strand-asynchronous asymmetric replication model, implies that the mtDNA replication mechanisms in *Anopheles* and *Drosophila* may be very similar. However, the *Anopheles* A + T rich region lacks long stretches of thymines that could indicate the position of the O_J . In this case, a sequence capable of folding into a short stem-loop structure, found close to the tRNA^{leu} gene (Supplementary file 2), may be involved in replication initiation, as is apparently the case in *Locusta migratoria* (Saito et al., 2005).

3.3. RNA genes

The RNA genes are the most conserved elements of the newly sequenced mitochondrial genomes (Fig. 1; Supplementary file 4), consistent with the evolutionary patterns observed in other organisms (Montooth et al., 2009; Pabijan et al., 2008). There are only 43 variable positions (including six indels) across a 2125 bp sequence alignment of the two rRNA genes from the six members of the complex. The tRNA genes are more conserved than the rRNA genes: the concatenated alignment of all the 22 genes is 1479 bp long, but it contains only 26 variable sites (including three indels). Seven tRNA genes are invariant, while the remaining genes have accumulated up to three mutations. We mapped all the observed substitutions onto the tRNA secondary structures (Supplementary file 3), and, guided by the phylogeny of the complex, reconstructed ancestral states at each node for the variable positions to get insights into the processes governing the tRNA evolution. In most cases mutations are mapped to the loop regions, and only five substitutions are found within the stem regions. Among these, four led from Watson–Crick pairing to non-canonical G–U pairing that is thermodynamically stable in RNA molecules (Varani and McClain, 2000), and one resulted in a change from U–G to U–A. A small proportion of changes in the stem regions and base-pairing maintenance in each case indicates strong selection to preserve the tRNAs secondary structures. Similar signatures of negative selection have been found in *Drosophila* (Montooth et al., 2009).

Apparently, the same mechanism drives the conservation of the rRNA genes. Guided by the secondary structure models of the *Anopheles* mitochondrial rRNA molecules we identified the positions in the alignments of the rRNA genes that correspond to the stem and loop regions, and counted the number of mutated positions in each structural category. Stem and loop regions in the mosquito rRNA molecules have approximately the same length, however, twice as many mutations were found in the loop regions (Supplementary file 4). Our observation is consistent with the finding that rRNA loops evolve much faster than stems (Smit et al., 2007).

3.4. Protein coding genes

In all species the initiation codons are as reported earlier in *Anopheles* (Krzywinski et al., 2006), with the exception of the ND5, which in *An. albitarsis* starts with a canonical ATG and in other species with a GTG triplet. Use of the GTG start codon has been documented for mtDNA-encoded proteins in various organisms, including *Anopheles* (Wolstenholme, 1992; Krzywinski et al., 2006). The complete stop codons (TAA) are found only in the *nad1*, *atp8* and *nad4L* genes. The latter two genes each have an overlapping sequence with a downstream protein coding gene (*atp8/atp6* and *nad4L/nad4*), with which they are transcribed as bicistronic messages (Krzywinski et al., 2006). In the remaining genes the stop codons are truncated (T or TA; Table 3), as judged from the presence of a tRNA gene encoded on the same strand and starting immediately downstream from the in-frame T or TA nucleotides. Such codons are post-transcriptionally regenerated during mRNA polyadenylation (Ojala et al., 1981).

Across the *An. albitarsis* complex there are 795 variable nucleotide positions in the PCGs; 94% of these are silent, most (674) within the 3rd codon positions. Average pairwise divergence rates are 0.003 at non-synonymous sites (d_N) and 0.116 at synonymous sites (d_S). Overall, values of the d_N/d_S ratios are very low and range between 0.0 and 0.057 (Supplementary file 4), indicating strong purifying selection in all PCGs; however, strength of selection varies across different genes. When genes are grouped according to their respective OXPHOS complexes and the average value of the d_N/d_S ratio across the *An. albitarsis* complex is used as a measure of divergence, the genes can be ordered as follows: ND $d_N/d_S = 0.035 > COX 0.019 > CYTB 0.016 > ATPase 0.010$. Although this ranking cannot be statistically supported due to the small sample size, it indicates a trend consistent with reports from other

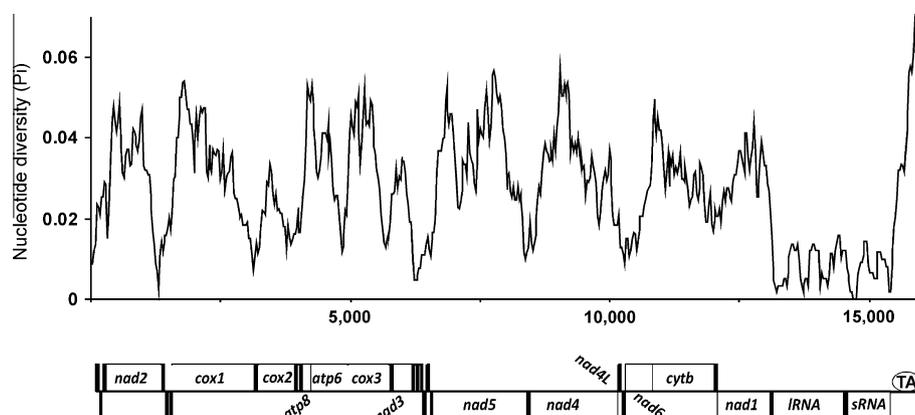


Fig. 1. Sliding window analysis (window size 200 bp, step size 25 bp) of the mtDNA nucleotide divergence between members of the *An. albitarsis* complex. Location of the genes is presented on a map of the mitochondrial genome below the graph. Major and minor strands are shown; the map starts at the tRNA^{leu} gene encoded on the major strand and ends with the non-coding A + T rich region presented as a horizontal black bar (labeled AT). Black vertical bars represent the tRNA genes, open labeled boxes represent protein coding and ribosomal RNA genes.

organisms, such as *Drosophila* (Montooth et al., 2009), fish (Doiron et al., 2002), toads (Pabijan et al., 2008) and mammals (da Fonseca et al., 2008), in which the evolutionary rate of ND is higher compared to other genes.

There are 51 non-synonymous nucleotide substitutions observed within the *An. albitarsis* complex. Fourteen of those, localized in *cox1*, *cytb*, and five *nad* genes (Table 4), lead to non-conservative amino acid changes, as indicated by negative mutability scores derived from the amino acid mutation matrix for transmembrane proteins (Jones et al., 1994). The functional significance of these changes is unknown. Most of the changes map to the subunits of the NADH:ubiquinone oxidoreductase (complex I), which is the first and the largest enzyme in the respiratory chain. Complex I is composed of 46 subunits, seven of which are encoded by the mitochondrial genes (Fearnley et al., 2007). It catalyzes the transfer of electrons from NADH to ubiquinone and the translocation of protons across the mitochondrion's inner membrane (Walker, 1992). The mtDNA-encoded ND subunits are embedded in the inner membrane and are thought to act as a proton pump (Brandt, 2006). Thus, the non-conservative amino acid substitutions in these subunits, in particular within the transmembrane domains, may interfere with the efficiency of the proton translocation process. The fact that such substitutions are tolerated likely indicates relaxed selective constraints on the structure of the subunits within the affected sites (Ingman and Gyllensten, 2007).

3.5. Phylogenetic analysis

The majority of the substitutions in the PCGs sequences are located within the 3rd codon positions, which, being prone to multiple substitutions, may be of limited utility for phylogenetic inference. However, values of the corrected pairwise distances indicate that within the *An. albitarsis* complex these sites are far from reaching saturation (Supplementary file 4), and thus are likely to provide valuable insight into the evolution of the complex. Analysis of the concatenated PCGs sequences yielded a single identical tree regardless of the inference method, with the majority of branches uniformly strongly supported. Monophyly of the *An. albitarsis* complex was strongly supported, consistent with the predicted close relationships between its members (Fig. 2A). Within the complex, species pairs *An. albitarsis* – *An. oryzalimnetes*,

An. deaneorum – *An. albitarsis* G, and *An. janconnae* – *An. albitarsis* F were inferred as strongly supported sister taxa, with the *An. deaneorum* – *An. albitarsis* G forming a basal clade. A close relationship between *An. albitarsis* and *An. oryzalimnetes* is consistent with earlier studies based on fragments of mitochondrial genes (Lehr et al., 2005; Wilkerson et al., 2005). A long terminal branch leading to *An. albitarsis* G indicated substantial divergence from other analyzed species, supporting its status of a distinct species, in agreement with the findings based on the analysis of the *cox1* barcode region (J. Ruiz, Y.M. Linton, J.E. Conn, N.S. McKeon, and R.C. Wilkerson, unpublished). The divergence between *An. albitarsis* G and its inferred sister species is comparable to the divergence between *An. albitarsis* s.s. and *An. deaneorum* (cf. table in Fig. 2A), whose distinct species status was demonstrated by laboratory crosses (Lima et al., 2004).

Unlike the mtDNA data, earlier analyses of the nuclear *white* gene supported clustering of *An. albitarsis* with *An. deaneorum* (Bourke et al., 2010; Brochero et al., 2007; Merritt et al., 2005). In this study, we reanalyzed the published and the newly generated sequences of the *white* gene fragment. The fragment covers a portion of an exon that, in the outgroup and two members of the *An. albitarsis* complex (*An. janconnae* and *An. albitarsis* G), is interrupted by a short (89–123 bp) intron. Analysis of the sequences aligned into an 883 bp long alignment yielded a strongly supported (*An. albitarsis* + *An. deaneorum*) clade, in agreement with the previous studies (Fig. 2B). We conducted the Shimodaira–Hasegawa test to compare the likelihood of the alternative tree topologies inferred using the mtDNA and the *white* gene sequences. In both cases the suboptimal trees inferred with the topological constraints enforced were significantly less likely (mtDNA tree with *An. albitarsis* – *An. deaneorum* monophyletic $P < 0.0001$; *white* gene tree with *An. albitarsis* – *An. oryzalimnetes* monophyletic $P = 0.0175$) than the respective unconstrained ML trees. In other words, the mtDNA and the *white* gene ML trees significantly better explained their respective datasets than the alternative (suboptimal) trees, indicating strong incongruence between the mtDNA and the *white* gene datasets.

3.6. mtDNA introgression as a likely cause of incongruent phylogenies

The *An. albitarsis* complex species may have radiated within a short time window during the last 2 million years (Conn and Mirabello, 2007). Although there are several potential explanations for the discordant tree topologies inferred with unlinked markers, retention of ancestral polymorphisms coupled with differential lineage sorting or introgressive hybridization are the most likely scenarios in case of such closely related species. Differentiating between these hypotheses is difficult because both processes can generate very similar phylogenetic patterns (Holder et al., 2001). However, strong support for the alternative topologies and the extent of variation within the two data sets (Fig. 2) indicate that the conflicting phylogenies are best explained by ancient mtDNA introgression (defined as stable incorporation of genetic material from one species into another, resulting from a historical interspecific hybridization).

Differences in effective population size and time to fixation dictate that differential lineage sorting is more likely to occur among nuclear rather than mitochondrial lineages. The *white* gene sequences used in our analysis were sampled for three members of the complex from two individuals, each from the extreme ends of the species range (Merritt et al., 2005), which provides a glimpse into their intraspecific polymorphism. The paucity of polymorphic sites in the analyzed individuals (0–3 variable sites per species; Fig. 2) suggests that ancestral polymorphisms may have had negligible influence on the shape of the *white* gene phylogenetic tree.

Table 4
Non-conservative amino acid substitutions in the mtDNA-encoded genes of the *An. albitarsis* complex.

Protein	Amino acid		Score	Domain ^a	Species
	Position	Change			
CO1	6	F → N	–4	Matrix	<i>An. deaneorum</i>
CYTB	2	F → V	–1	Inter	<i>An. deaneorum</i>
	8	N → M	–2	Inter	<i>An. deaneorum</i>
ND1	175	S → F	–1	Trans	<i>An. albitarsis</i> F
ND2	318	S → L	–2	Trans	<i>An. janconnae</i>
ND4	232	L → W	–2	Trans	<i>An. janconnae</i>
	349	F → I	–1	Trans	<i>An. albitarsis</i> F
ND4L	70	C → K	–3	Trans	<i>An. albitarsis</i> F
ND5	168	W → S	–3	Trans	<i>An. albitarsis</i> G
	176	M → S	–2	Inter	<i>An. albitarsis</i> <i>An. albitarsis</i> G <i>An. deaneorum</i> <i>An. janconnae</i> <i>An. albitarsis</i> F
	176	S → M	–2	Inter	<i>An. oryzalimnetes</i>
	181	F → A	–2	Inter	<i>An. deaneorum</i>
	464	I → N	–3	Trans	<i>An. albitarsis</i> G
	521	Y → N	–1	Inter	<i>An. oryzalimnetes</i>

^a Position of the substitution in either matrix, intermembrane, or transmembrane domain of the protein.

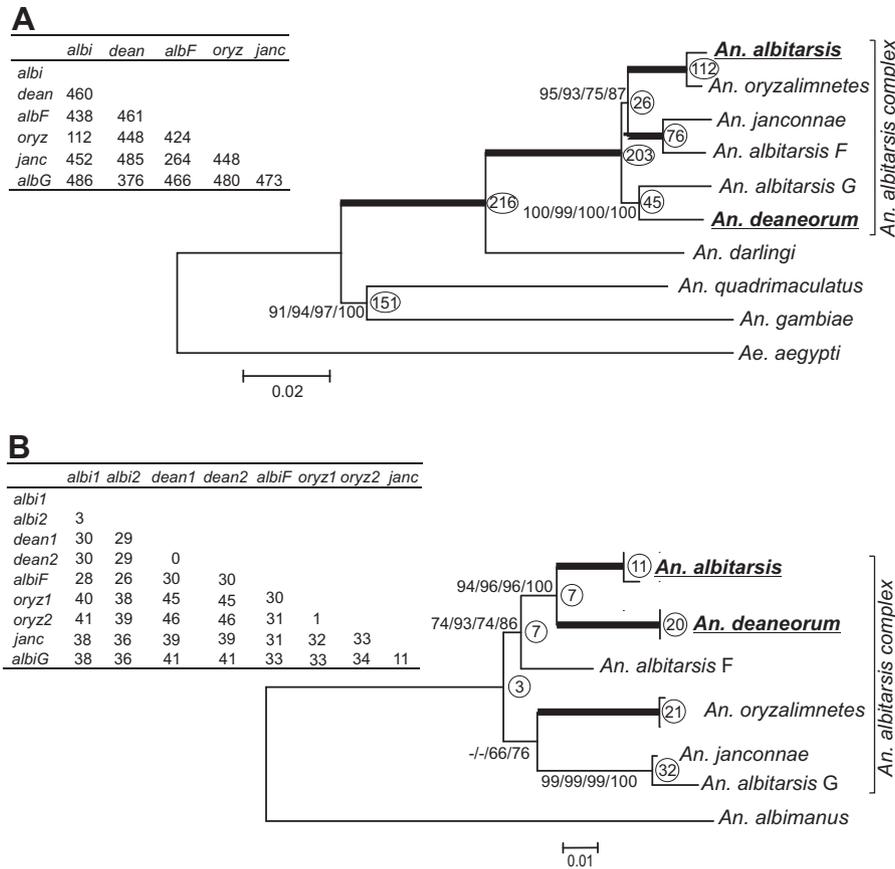


Fig. 2. Maximum likelihood phylogenetic trees of relationships within the *An. albitarsis* complex inferred from (A) all mtDNA-encoded protein coding genes and (B) a fragment of the nuclear *white* gene. Note the discordance in the inference of the closest relative to *An. albitarsis* from each data set. Thick branches denote 100% support from each inference method; otherwise support values for a branch are given in the order: maximum parsimony/neighbor joining/maximum likelihood/Bayesian posterior probability. Numbers of synapomorphies for each branch are circled. The tables accompanying the trees show the numbers of observed pairwise differences between the aligned (A) complete mitochondrial genome sequences or (B) sequences of the *white* gene fragment used for the phylogenetic reconstruction. Despite a relatively short length, the *white* gene fragment harbors considerable phylogenetic information, as demonstrated by the numbers of interspecific differences and the numbers of synapomorphies. The *white* gene MP and NJ trees (not shown) differ from the presented ML tree in the position of *An. oryzalimnetes* as a basal taxon to the (*An. albitarsis* F, (*An. albitarsis*, *An. deaneorum*) clade. The MP and NJ trees are, thus, consistent with a single loss of the *white* gene intron within the ingroup (in the common ancestor of the four species listed above). Each tree was inferred using a different member of the subgenus *Nyssorhynchus* (either *An. darlingi* or *An. albimanus*) as the closest outgroup species.

Low intraspecific polymorphisms contrast with high interspecific divergence of the *white* gene among most members of the complex. *An. albitarsis* differs from its inferred sister species *An. deaneorum* by 29 fixed nucleotide substitutions, which is comparable to an average of 35 (11–45) fixed differences for other pairwise comparisons within the complex (Fig. 2B). A strikingly different picture emerges from the mtDNA data. *An. albitarsis* and *An. oryzalimnetes*, inferred as sister taxa in the mtDNA tree, have accumulated a drastically lower number of nucleotide differences than any other species pair within the complex: 112 differences compared to an average of 440 (264–486) differences across the entire mtDNA molecule (Fig. 2A). The above numbers indicate that the interspecific coalescence events involving the *An. albitarsis* mtDNA and the *An. albitarsis* nuclear *white* gene DNA may have been widely temporally separated and that *An. albitarsis* may have shared the most recent common mitochondrial ancestor with another species much more recently than the most recent common *white* gene ancestor. The patterns observed in our data and the incongruence between mtDNA and nuclear phylogenies are consistent with an ancient mitochondrial introgression caused by hybridization between non-sister taxa (Linnen and Farrell, 2007), in this case between *An. albitarsis* and *An. oryzalimnetes*. Interspecific introgression of the mtDNA is a very common outcome of hybridization; there is ample evidence of such introgression in various animal groups (Avice, 2000; Chan and Levin, 2005; Linnen and

Farrell, 2007) and it has also been invoked to explain conflicting phylogenies within the *An. gambiae* complex (Besansky et al., 2003, 1994). Further study, including additional unlinked nuclear markers and a larger number of individuals sampled across species ranges, are needed to test our hypothesis.

3.7. Utility of mtDNA sequences as phylogenetic, population genetics and species identification markers

Although mtDNA introgression may occasionally complicate inference of the relationships, mtDNA genomes constitute an attractive source of molecular markers that can address various phylogenetic and population genetics questions. We conducted a sliding window analysis of the aligned mtDNA sequences at increasing phylogenetic depths to identify highly variable regions that may be utilized as markers in the future studies of anophelines (Fig. 1; Supplementary file 5). Comparisons within the *An. albitarsis* complex and within *Anopheles* subgenus *Nyssorhynchus* (between members of the complex and *An. darlingi*) indicated that the *nad5*, *nad4*, *cox3*, *atp6*, *cox1* and *nad2* are the most promising markers at low levels of divergence. However, variation is not equally distributed along these genes. For example, the 5' half of the *cox1*, which encompasses the barcode region, is markedly more variable than its 3' half. Because a strong correlation between the interspecific divergence and intraspecific

polymorphism has been observed in the mitochondrial genomes (Zarowiecki et al., 2007), at least some of these genes are also likely informative for population genetics studies. In contrast, higher conservation of the *cox2*, *nad3*, *nad6* and *cytb* compared to the other PCGs suggests that they may not be the optimal markers.

The *nad5*, *nad4*, and *nad2* remain the most highly variable PCGs in comparisons between deeply diverged *Anopheles* subgenera (cf. split between the subgenera *Anopheles* and *Cellia* estimated at 90–106 Mya; (Krzywinski et al., 2006)). While at such evolutionary time scales mutational saturation may have erased phylogenetic signal at many synonymous positions, these genes still contain considerable variation at the non-synonymous sites that are less likely to be saturated (cf. Supplementary file 4).

The non-coding AT-rich region is the most highly variable within the entire mtDNA genome, especially at larger evolutionary distances, but its utility is severely limited by the uncertainty regarding reliability of its alignment. In addition, the presence of low complexity sequences, such as internal homopolymer runs and dinucleotide repeats, may seriously hinder obtaining reliable sequencing data within this region.

Multiplex PCR assays targeting the nuclear ribosomal ITS2 region are fast and the most resource-efficient for routine identification of anopheline species within cryptic species complexes (Krzywinski and Besansky, 2003). However, attempts to apply multiplex PCR for the discrimination between species of the *An. albitarsis* complex failed due to inadequate variation within the ITS2 region (Li and Wilkerson, 2005). Because the mtDNA has been successfully used for the PCR-based identification of species belonging to the *An. sundaeicus* and *An. culicifacies* complexes (Dusfour et al., 2007; Goswami et al., 2006), it is likely that the mtDNA molecule could also serve as an alternative target in the *An. albitarsis* complex. The mtDNA of the six representatives of the *An. albitarsis* complex described here contain several highly conserved regions adjacent to moderately variable sequences that may be optimal targets for the design of such a multiplex PCR assay. However, analysis of the utility of these regions is beyond the scope of this study. To be meaningful, it would need an extended sample that would include mitochondrial sequences of all the species identified within the complex.

Acknowledgments

We are grateful to Martin Donnelly for a constructive discussion on the issue of potential introgression. Comments and suggestions made by two anonymous reviewers and Stephen Cameron improved the shape of the manuscript. We thank John F. Ruiz for providing the *An. albitarsis* G *white* gene sequence for the current analysis. We also thank Laboratório de Pesquisas Básicas at Instituto Evandro Chagas, Belem, Para state, Brazil for logistical support in the field. This study was funded by National Institutes of Health (USA) Grant 2R01AI54139 to JEC and by start-up funds from the Liverpool School of Tropical Medicine to JK. Material has been reviewed by the Walter Reed Army Institute of Research. There is no objection to its presentation and/or publication. The opinions or assertions contained herein are the private views of the author, and are not to be construed as official, or as reflecting true views of the Department of the Army or the Department of Defense. Parts of this research were performed at the Smithsonian Institution under a Memorandum of Understanding between the Walter Reed Army Institute of Research and the Smithsonian Institution, with institutional support provided by both organizations.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.01.003.

References

- Avise, J.C., 2000. *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Mass.
- Ballard, J.W., 2000. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J. Mol. Evol.* 51, 48–63.
- Beard, C.B., Hamm, D.M., Collins, F.H., 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.* 2, 103–124.
- Besansky, N.J., Fahey, G.T., 1997. Utility of the *white* gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae). *Mol. Biol. Evol.* 14, 442–454.
- Besansky, N.J., Powell, J.R., Caccone, A., Hamm, D.M., Scott, J.A., Collins, F.H., 1994. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. *Proc. Natl. Acad. Sci. USA* 91, 6885–6888.
- Besansky, N.J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., Fontenille, D., Toure, Y., Sagnon, N., 2003. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proc. Natl. Acad. Sci. USA* 100, 10818–10823.
- Boore, J.L., 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780.
- Bourke, B.P., Foster, P.G., Berge, E.S., Calado, D.C., Sallum, M.A., 2010. Phylogenetic relationships among species of *Anopheles* (*Nyssorhynchus*) (Diptera, Culicidae) based on nuclear and mitochondrial gene sequences. *Acta Trop.* 114, 88–96.
- Brandt, U., 2006. Energy converting NADH:quinone oxidoreductase (complex I). *Annu. Rev. Biochem.* 75, 69–92.
- Branquinho, M.S., Lagos, C.B., Rocha, R.M., Natal, D., Barata, J.M., Cochrane, A.H., Nardin, E., Nussenzweig, R.S., Kloetzel, J.K., 1993. Anophelines in the state of Acre, Brazil, infected with *Plasmodium falciparum*, *P. vivax*, the variant *P. vivax* VK247 and *P. malariae*. *Trans. Roy. Soc. Trop. Med. Hyg.* 87, 391–394.
- Brochero, H.H., Li, C., Wilkerson, R.C., 2007. A newly recognized species in the *Anopheles* (*Nyssorhynchus*) *albitarsis* complex (Diptera: Culicidae) from Puerto Carreno, Colombia. *Am. J. Trop. Med. Hyg.* 76, 1113–1117.
- Carapelli, A., Comandi, S., Convey, P., Nardi, F., Frati, F., 2008. The complete mitochondrial genome of the Antarctic springtail *Cryptopygus antarcticus* (Hexapoda: Collembola). *BMC Genomics* 9, 315.
- Chan, K.M.A., Levin, S.A., 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* 59, 720–729.
- Clayton, D.A., 1982. Replication of animal mitochondrial DNA. *Cell* 28, 693–705.
- Conn, J.E., Mirabello, L., 2007. The biogeography and population genetics of neotropical vector species. *Heredity* 99, 245–256.
- da Fonseca, R.R., Johnson, W.E., O'Brien, S.J., Ramos, M.J., Antunes, A., 2008. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics* 9, 119.
- De Rijk, P., Wuyts, J., De Wachter, R., 2003. RnaViz2: an improved representation of RNA secondary structure. *Bioinformatics* 19, 299–300.
- Doiron, S., Bernatchez, L., Blier, P.U., 2002. A comparative mitogenomic analysis of the potential adaptive value of Arctic charr mtDNA introgression in brook charr populations (*Salvelinus fontinalis* Mitchell). *Mol. Biol. Evol.* 19, 1902–1909.
- Dusfour, I., Blondeau, J., Harbach, R.E., Vythilingham, I., Baimai, V., Trung, H.D., Sochant, T., Bangs, M.J., Manguin, S., 2007. Polymerase chain reaction identification of three members of the *Anopheles sundaeicus* (Diptera: Culicidae) complex, malaria vectors in Southeast Asia. *J. Med. Entomol.* 44, 723–731.
- Fearnley, I.M., Carroll, J., Walker, J.E., 2007. Proteomic analysis of the subunit composition of complex I (NADH: ubiquinone oxidoreductase) from bovine heart mitochondria. *Methods Mol. Biol.* (Clifton, NJ) 357, 103–125.
- Frederico, L.A., Kunkel, T.A., Shaw, B.R., 1990. A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* 29, 2532–2537.
- Gissi, C., Iannelli, F., Pesole, G., 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* 101, 301–320.
- Goddard, J.M., Wolstenholme, D.R., 1980. Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. *Nucleic Acids Res.* 8, 741–757.
- Goloboff, P.A., Farris, J.S., Nixon, K.C., 2008. TNT, a free program for phylogenetic analysis. *Cladistics* 24, 774–786.
- Goswami, G., Singh, O.P., Nanda, N., Raghavendra, K., Gakhar, S.K., Subbarao, S.K., 2006. Identification of all members of the *Anopheles culicifacies* complex using allele-specific polymerase chain reaction assays. *Am. J. Trop. Med. Hyg.* 75, 454–460.
- Holder, M.T., Anderson, J.A., Holloway, A.K., 2001. Difficulties in detecting hybridization. *Syst. Biol.* 50, 978–982.
- Ingman, M., Gyllenstein, U., 2007. Rate variation between mitochondrial domains and adaptive evolution in humans. *Hum. Mol. Genet.* 16, 2281–2287.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1994. A mutation data matrix for transmembrane proteins. *FEBS Lett.* 339, 269–275.
- Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.
- Krzywinski, J., Besansky, N.J., 2003. Molecular systematics of *Anopheles*: from subgenera to subpopulations. *Annu. Rev. Entomol.* 48, 111–139.
- Krzywinski, J., Wilkerson, R.C., Besansky, N.J., 2001. Evolution of mitochondrial and ribosomal gene sequences in Anophelinae (Diptera: Culicidae): implications for phylogeny reconstruction. *Mol. Phylogenet. Evol.* 18, 479–487.

- Krzywinski, J., Grushko, O.G., Besansky, N.J., 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol. Phylogenet. Evol.* 39, 417–423.
- Lehr, M.A., Kilpatrick, C.W., Wilkerson, R.C., Conn, J.E., 2005. Cryptic species in the *Anopheles (Nyssorhynchus) albitarsis* (Diptera: Culicidae) complex: incongruence between random amplified polymorphic DNA-polymerase chain reaction identification and analysis of mitochondrial DNA COI gene sequences. *Ann. Entomol. Soc. Am.* 98, 908–917.
- Li, C., Wilkerson, R.C., 2005. Identification of *Anopheles (Nyssorhynchus) albitarsis* complex species (Diptera: Culicidae) using rDNA internal transcribed spacer 2-based polymerase chain reaction primers. *Mem. Inst. Oswaldo Cruz* 100, 495–500.
- Li, C., Wilkerson, R.C., 2007. Intragenomic rDNA ITS2 variation in the neotropical *Anopheles (Nyssorhynchus) albitarsis* complex (Diptera: Culicidae). *J. Heredity* 98, 51–59.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)* 25, 1451–1452.
- Lima, J.B.P., Valle, D., Peixoto, A.A., 2004. Analysis of reproductive isolation between sibling species *Anopheles albitarsis* sensu stricto and *Anopheles deaneorum*, two malaria vectors belonging to the Albitarsis complex (Diptera: Culicidae). *J. Med. Entomol.* 41, 888–893.
- Lindahl, T., 1993. Instability and decay of the primary structure of DNA. *Nature* 362, 709–715.
- Linnen, C.R., Farrell, B.D., 2007. Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution* 61, 1417–1438.
- Loaiza, J.R., Scott, M.E., Bermingham, E., Rovira, J., Conn, J.E., 2010. Evidence for pleistocene population divergence and expansion of *Anopheles albimanus* in Southern Central America. *Am. J. Trop. Med. Hyg.* 82, 156–164.
- Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964.
- McKeon, S.N., Lehr, M.A., Wilkerson, R.C., Ruiz, J.F., Sallum, M.A., Lima, J.B., Povoas, M.M., Conn, J.E., 2010. Lineage divergence detected in the malaria vector *Anopheles marajoara* (Diptera: Culicidae) in Amazonian Brazil. *Malaria J.* 9, 271.
- Merritt, T.J., Young, C.R., Vogt, R.G., Wilkerson, R.C., Quattro, J.M., 2005. Intron retention identifies a malaria vector within the *Anopheles (Nyssorhynchus) albitarsis* complex (Diptera: Culicidae). *Mol. Phylogenet. Evol.* 35, 719–724.
- Mitchell, S.E., Cockburn, A.F., Seawright, J.A., 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome* 36, 1058–1073.
- Montooth, K.L., Abt, D.N., Hofmann, J.W., Rand, D.M., 2009. Comparative genomics of *Drosophila* mtDNA: novel features of conservation and change across functional domains and lineages. *J. Mol. Evol.* 69, 94–114.
- Moreno, M., Marinotti, O., Krzywinski, J., Tadei, W.P., James, A.A., Achee, N.L., Conn, J.E., 2010. Complete mtDNA genomes of *Anopheles darlingi* and an approach to anopheline divergence time. *Malaria J.* 9, 127.
- Motoki, M.T., Wilkerson, R.C., Sallum, M.A., 2009. The *Anopheles albitarsis* complex with the recognition of *Anopheles oryzalimnetes* Wilkerson and Motoki, n. sp. and *Anopheles janconnae* Wilkerson and Sallum, n. sp. (Diptera: Culicidae). *Mem. Inst. Oswaldo Cruz* 104, 823–850.
- Nylander, J.A.A., Wilgenbush, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* 24, 581–583.
- Ojala, D., Montoya, J., Attardi, G., 1981. tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290, 470–474.
- Oliveira, M.T., Barau, J.G., Junqueira, A.C., Feijao, P.C., Rosa, A.C., Abreu, C.F., Azeredo-Espin, A.M., Lessinger, A.C., 2008. Structure and evolution of the mitochondrial genomes of *Haematobia irritans* and *Stomoxys calcitrans*: the Muscidae (Diptera: Calyptratae) perspective. *Mol. Phylogenet. Evol.* 48, 850–857.
- Pabijan, M., Spolsky, C., Uzzell, T., Szymura, J.M., 2008. Comparative analysis of mitochondrial genomes in *Bombina* (Anura; Bombinatoridae). *J. Mol. Evol.* 67, 246–256.
- Perna, N.T., Kocher, T.D., 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* 41, 353–358.
- Posada, D., 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256.
- Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* 19, 1572–1574.
- Saito, S., Tamura, K., Aotsuka, T., 2005. Replication origin of mitochondrial DNA in insects. *Genetics* 171, 1695–1705.
- Service, M.W., Townson, H., 2002. The *Anopheles* vector. In: Warrell, D.A., Gilles, H.M. (Eds.), *Essential Malariaology*. Arnold, London, pp. 59–84.
- Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116.
- Smit, S., Widmann, J., Knight, R., 2007. Evolutionary rates vary among rRNA structural elements. *Nucleic Acid Res.* 35, 3339–3354.
- Swofford, D.L., 2002. PAUP*4.0: Phylogenetic Analysis using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Mass., USA.
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Varani, G., McClain, W.H., 2000. The G × U wobble base pair. A fundamental building block of RNA structure crucial to RNA function in diverse biological systems. *EMBO Rep.* 1, 18–23.
- Walker, J.E., 1992. The NADH: ubiquinone oxidoreductase (complex I) of respiratory chains. *Quart. Rev. Biophys.* 25, 253–324.
- Wilkerson, R.C., Parsons, T.J., Albright, D.G., Klein, T.A., Braun, M.J., 1993. Random amplified polymorphic DNA (RAPD) markers readily distinguish cryptic mosquito species (Diptera: Culicidae: *Anopheles*). *Insect Mol. Biol.* 1, 205–211.
- Wilkerson, R.C., Parsons, T.J., Klein, T.A., Gaffigan, T.V., Bergo, E., Consolim, J., 1995. Diagnosis by random amplified polymorphic DNA polymerase chain reaction of four cryptic species related to *Anopheles (Nyssorhynchus) albitarsis* (Diptera: Culicidae) from Paraguay, Argentina, and Brazil. *J. Med. Entomol.* 32, 697–704.
- Wilkerson, R.C., Foster, P.G., Li, C., Sallum, M.A., 2005. Molecular phylogeny of Neotropical *Anopheles (Nyssorhynchus) albitarsis* species complex (Diptera: Culicidae). *Ann. Entomol. Soc. Am.* 98, 918–925.
- Wolstenholme, D.R., 1992. Genetic novelties in mitochondrial genomes of multicellular animals. *Curr. Opin. Genet. Dev.* 2, 918–925.
- Zarowiecki, M.Z., Huysse, T., Littlewood, D.T., 2007. Making the most of mitochondrial genomes – markers for phylogeny, molecular ecology and barcodes in *Schistosoma* (Platyhelminthes: Digenea). *Int. J. Parasitol.* 37, 1401–1418.