A multi-locus approach to barcoding in the *Anopheles strodei* subgroup (Diptera: Culicidae)

Brian Patrick Bourke¹*, Tatiane Porangaba Oliveira¹, Lincoln Suesdek², Eduardo Sterlino Bergo³ and Maria Anice Mureb Sallum¹

**Abstract**

**Background:** The ability to successfully identify and incriminate pathogen vectors is fundamental to effective pathogen control and management. This task is confounded by the existence of cryptic species complexes. Molecular markers can offer a highly effective means of species identification in such complexes and are routinely employed in the study of medical entomology. Here we evaluate a multi-locus system for the identification of potential malaria vectors in the *Anopheles strodei* subgroup.

**Methods:** Larvae, pupae and adult mosquitoes (n = 61) from the *An. strodei* subgroup were collected from 21 localities in nine Brazilian states and sequenced for the *COI*, ITS2 and *white* gene. A Bayesian phylogenetic approach was used to describe the relationships in the Strodei Subgroup and the utility of *COI* and ITS2 barcodes was assessed using the neighbor joining tree and “best close match” approaches.

**Results:** Bayesian phylogenetic analysis of the *COI*, ITS2 and *white* gene found support for seven clades in the *An. strodei* subgroup. The *COI* and ITS2 barcodes were individually unsuccessful at resolving and identifying some species in the Subgroup. The *COI* barcode failed to resolve *An. albertoi* and *An. strodei* but successfully identified approximately 92% of all species queries, while the ITS2 barcode failed to resolve *An. arthuri* and successfully identified approximately 60% of all species queries. A multi-locus COI-ITS2 barcode, however, resolved all species in a neighbor joining tree and successfully identified all species queries using the “best close match” approach.

**Conclusions:** Our study corroborates the existence of *An. albertoi*, *An. CP Form* and *An. strodei* in the *An. strodei* subgroup and identifies four species under *An. arthuri* informally named A-D herein. The use of a multi-locus barcode is proposed for species identification, which has potentially important utility for vector incrimination. Individuals previously found naturally infected with *Plasmodium vivax* in the southern Amazon basin and reported as *An. strodei* are likely to have been from *An. arthuri* C identified in this study.

**Keywords:** Culicidae, *Anopheles*, Strodei Subgroup, Identification, Barcode, COI, ITS2, White, Atlantic Forest, New species

**Background**

One of the most important goals of medical entomology is to develop approaches that effectively identify the roles of insect species in transmitting infectious pathogens. The incrimination of a pathogen vector requires demonstrating that the species feeds on humans, an association in time and space between the species and the occurrence of human infections, repeated isolation of the pathogen from the species, and the transmission of the pathogen by the species under controlled experimental conditions [1]. Fundamental to the process of incrimination is an ability to resolve and identify species effectively. However, many vector species are morphologically indistinguishable from close relatives yet they can exhibit a range of genetic, biological and morphological variation [2]. Such species form cryptic species complexes and their existence makes the task of vector incrimination more difficult. Molecular approaches are now routinely used to help resolve such complexes and have become essential tools in the study of medical entomology and infectious disease transmission.

---

* Correspondence: bpbourke@usp.br
¹Departamento de Epidemiologia, Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo-SP, Brazil
Full list of author information is available at the end of the article

© 2013 Bourke et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
The phylogenetic analysis of species complexes employs markers with relatively high rates of substitution that are likely to track recently diverged species. A multi-locus approach can reconstruct more robust evolutionary relationships, discover previously unknown lineages in species and inform the search for latent morphological differences. Recently, DNA barcoding initiatives have proposed approaches that employ “sequence diversity in short, standardized gene regions to aid species identification and discovery in large assemblages of life” [3]. Various molecular markers [4-6] have been employed but it is cytochrome c oxidase I (COI) that has gained acceptance as the “gold standard” barcode for animals. The internal transcribed spacer region 2 (ITS2) has also been employed as a barcode region, primarily for plants but increasingly for animals [5]. The success of the barcoding approach is related to inter-specific variation exceeding intra-specific variation (the existence of the “barcoding gap”), and the analysis to date has generally been performed using clustering (neighbor joining tree monophyly) or pairwise genetic distances [7]. Recently diverged or incipient species, however, may be frequently misidentified due to incomplete lineage sorting of ancestral polymorphisms [8-10]. While barcoding is therefore a useful approach to determine minimum estimates of species numbers in cryptic species complexes although see [11], multi-locus and multi-data (genetic/morphological/ecological) approaches are likely to be more effective at elucidating the full extent of species diversity within these systems.

The current study focuses on species diversity within the Neotropical Strodei Subgroup of Anopheles (Nyssorhynchus) mosquitoes. This Subgroup is currently comprised of five species (Anopheles albertoi Unti, Anopheles arthuri Unti, Anopheles rondoni (Neiva and Pinto) and Anopheles strodei Root), which are distributed through much of Central and South America, from Panama to Argentina [13,14], although several additional taxa have been frequently misidentified. Anopheles strodei was first described using morphological characters of the adult male, fourth-instar larvae and pupae from specimens from Juiz de Fora, Minas Gerais State, Brazil [15]. Later, An. albertoi, An. arthuri, An. artigasi Unti, and An. lloydii Unti were described based on egg characteristics and Anopheles ramosi Unti by the fourth-instar larvae [16,17]. The type localities of An. albertoii, An. arthuri, An. artigasi, An. ramosi are all from Vale do Paraíba, São Paulo state, Brazil, whereas that of An. lloydii is an unspecified location in Panama. Further examination of An. strodei based on adult female, larvae [13] and egg [18] morphology and patterns of the salivary polytene chromosome [19] showed high levels of polymorphism throughout its range and led Faran [13] to synonymize An. strodei, An. albertoi, An. arthuri, An. artigasi, An. lloydii, An. ramosi and An. strodei into a single species. A recent study of COI gene and white gene [12] sequences allowed the resurrection of An. albertoi and An. arthuri from synonymy with An. strodei, and revealed an undescribed taxon, preliminarily named An. CP Form.

Although Neotropical Anopheles species are known vectors of filariasis (Wuchereria bancrofti Cobbold [20]), arboviruses (Anopheles A Virus [21]) and malaria [22], the importance of the Strodei Subgroup in vectoring parasites is largely unknown. Anopheles strodei, however, has previously been found naturally infected with Plasmodium vivax Grassi & Feletti in Ariquemes, Rondônia, in the Amazon region, [23] although it remains unknown whether this record refers to An. strodei s.s. or another member of the Strodei Subgroup. The continental distribution of this complex confounds efforts to comprehensively describe species diversity and, ultimately, vectorial capacity. Our study seeks to provide a more complete understanding of species diversity and distribution in the Strodei Subgroup by performing a multi-locus DNA analysis of specimens collected from across Brazil. We will first resolve species relationships with a Bayesian approach using the COI, ITS2 and white gene. We will then test the utility of the COI barcode and the less frequently employed ITS2 barcode for species identification in the An. strodei subgroup.

Methods
Mosquito collection
Collection localities and identity of the specimens included in this study can be found in Table 1. These specimens were either offspring of females caught in the field using a Shannon trap or larvae and pupae collected from immature habitats, which were then raised to adulthood. Species identification of all but two specimens was based on adult male genitalia, fourth-instar larval characteristics or scanning electron micrographs of the egg. Individuals from An. arthuri displayed substantial variation in male genitalia and so were identified as An. arthuri sensu lato.

DNA Extraction
DNA was extracted from each specimen according to the animal tissue DNA extraction protocol provided by the QIAgen DNeasy Blood and Tissue Kit (QIAgen Ltd, Crawley, UK). All extractions were diluted to 200 μL with the buffer provided and extraction solutions were retained for storage at −80°C in the entomological frozen collection of the Faculdade de Saúde Pública, Universidade de São Paulo, Brazil.

COI gene
The gene was amplified using LCO- 1490 (5′-GGT CAA CAA ATC ATA AAG ATA TTG G-3′) and HCO-2198
Table 1 Sample information, including specimen numbers, species, localities, geographical coordinates, and Genbank accession numbers

| Specimen | Species       | Locality (state)   | Latitude   | Longitude  | COI GenBank accession no. | ITS2 GenBank accession no. | White GenBank accession no. |
|----------|---------------|--------------------|------------|-----------|---------------------------|------------------------------|-----------------------------|
| MG07 12 4 | *Anopheles albertoi* | Frutal (Minas Gerais) | -20.025278 | -49.075500 | GU226678                  | FJ178885                     | GU226747                    |
| MG07 3 4  | *Anopheles albertoi* | Frutal (Minas Gerais) | -20.025278 | -49.075500 | GU226676                  | FJ178889                     | GU226742                    |
| MG07 7 10 | *Anopheles albertoi* | Frutal (Minas Gerais) | -20.025278 | -49.075500 | GU226677                  | FJ178886-FJ178888            | GU226743-GU226746           |
| CE12 1 1  | *Anopheles arthuri* | Ubajara (Ceará)     | -3.8867500 | -41.001250 | KC330250                  | KC330268                     | KC330325                    |
| CE12 4 6  | *Anopheles arthuri* | Ubajara (Ceará)     | -3.8867500 | -41.001250 | KC330251                  | KC330269                     | KC330326                    |
| MG07 7 10 | *Anopheles arthuri* | Frutal (Minas Gerais) | -20.025278 | -49.075500 | GU226677                  | FJ178886-FJ178888            | GU226743-GU226746           |
| CE17 3 1  | *Anopheles arthuri* | Ubajara (Ceará)     | -3.8867500 | -41.001250 | KC330250                  | KC330268                     | KC330325                    |
| CE17 5 1  | *Anopheles arthuri* | Ubajara (Ceará)     | -3.8867500 | -41.001250 | KC330251                  | KC330269                     | KC330326                    |
| CE20 10 4 | *Anopheles arthuri* | Ubajara (Ceará)     | -3.8867500 | -41.001250 | KC330250                  | KC330268                     | KC330325                    |
| G07 1 3  | *Anopheles arthuri* | Itarumã (Goiás)    | -18.906128 | -51.024917 | KC330244                  | KC330274                     | KC330319                    |
| G07 2 1  | *Anopheles arthuri* | Itarumã (Goiás)    | -18.906128 | -51.024917 | KC330245                  | KC330271                     | KC330321                    |
| G07 3 10 5 | *Anopheles arthuri* | Itarumã (Goiás)    | -18.906128 | -51.024917 | KC330246                  | KC330276                     | KC330321                    |
| G07 6 10 2 | *Anopheles arthuri* | Itarumã (Goiás)    | -18.906128 | -51.024917 | KC330247                  | KC330279-KC330282           | KC330322                    |
| MG03 102 | *Anopheles arthuri* | Frutal (Minas Gerais) | -19.981278 | -49.065028 | GU226679                  | FJ178881-FJ178884           | GU226748                    |
| MG04 102 | *Anopheles arthuri* | Frutal (Minas Gerais) | -19.988472 | -49.033361 | GU226680                  | FJ178880                     | GU226751                    |
| MG07 106 | *Anopheles arthuri* | Frutal (Minas Gerais) | -20.025278 | -49.076500 | GU226683                  | FJ178879                     | GU226752                    |
| MG07 100 | *Anopheles arthuri* | Frutal (Minas Gerais) | -20.025278 | -49.076500 | GU226685                  | GU226706                     | GU226753-GU226755           |
| MG07 20 2 | *Anopheles arthuri* | Frutal (Minas Gerais) | -20.025278 | -49.076500 | GU226686                  | GU226707-GU226711           | GU226750                    |
| MG07 6 3  | *Anopheles arthuri* | Frutal (Minas Gerais) | -20.025278 | -49.076500 | GU226681                  | GU226700-GU226705           | GU226749                    |
| MG04 1  | *Anopheles arthuri* | Goianá (Minas Gerais) | -21.538836 | -43.200856 | GU226682                  | GU226723-GU226725           | GU226757                    |
| MG03 4  | *Anopheles arthuri* | Oliveira (Minas Gerais) | -20.746389 | -44.915278 | KC330257                  | KC330283                     | KC330332                    |
| MG03 11 2 | *Anopheles arthuri* | Oliveira (Minas Gerais) | -20.745598 | -44.915613 | KC330258                  | KC330284                     | KC330340                    |
| MG33 12 6 | *Anopheles arthuri* | Oliveira (Minas Gerais) | -20.745598 | -44.915613 | KC330259                  | KC330285                     | KC330334                    |
| MG33 13 7 | *Anopheles arthuri* | Oliveira (Minas Gerais) | -20.508785 | -44.770600 | KC330256                  | KC330286-KC330288           | KC330331                    |
| MG34 1  | *Anopheles arthuri* | Oliveira (Minas Gerais) | -20.712500 | -44.974444 | KC330260                  | KC330289                     | KC330335                    |
| MG34 9  | *Anopheles arthuri* | Oliveira (Minas Gerais) | -20.712500 | -44.974444 | KC330261                  | KC330290                     | KC330336                    |
| MG35 11 | *Anopheles arthuri* | São Franciscio de Paula (Minas Gerais) | -20.754444 | -44.917222 | KC330262                  | KC330291                     | KC330338                    |
| MG44 14 2 | *Anopheles arthuri* | Oliveira (Minas Gerais) | -20.768428 | -44.878209 | KC330263                  | KC330292                     | KC330339                    |
| RO29 18 | *Anopheles arthuri* | Campo Novo de Rondônia (Rondônia) | -10.637639 | -65.499833 | KC330248                  | KC330293                     | KC330323                    |
| RO31 103 | *Anopheles arthuri* | Campo Novo de Rondônia (Rondônia) | -10.637639 | -65.499833 | KC330249                  | KC330294                     | KC330324                    |
| RO8 1  | *Anopheles arthuri* | Monte Negro (Rondônia) | -10.266639 | -63.555389 | GU226681                  | GU226727                     | GU226759                    |
| RO8 104 | *Anopheles arthuri* | Monte Negro (Rondônia) | -10.266639 | -63.555389 | GU226690                  | GU226728                     | GU226760                    |
| RO8 109 | *Anopheles arthuri* | Monte Negro (Rondônia) | -10.266639 | -63.555389 | GU226689                  | GU226729                     | GU226761                    |
| Specimen Number | Anopheles species | Localities | Geographical Coordinates | Genbank Accession Numbers |
|-----------------|-------------------|------------|--------------------------|---------------------------|
| SP31 120        | Anopheles arthuri | Inubia Paulista (São Paulo) | -21.681417, -50.919889 | GU226687, GU226699, GU226718-GU226722, GU226758 |
| MG15 1 1        | Anopheles CP Form | Coronel Pacheco (Minas Gerais) | -21.635819, -43.19267 | JN413711, KC330265, KC330316 |
| MG15 6 12       | Anopheles CP Form | Coronel Pacheco (Minas Gerais) | -21.635819, -43.19267 | JN413712, KC330266, KC330317 |
| MG15 9 6        | Anopheles CP Form | Coronel Pacheco (Minas Gerais) | -21.635819, -43.19267 | KC330243, KC330267, KC330318 |
| PR21 110        | Anopheles CP Form | Foz do Iguaçu (Paraná) | -54.546528, -25.454583 | GU226691, FJ178890, GU226762 |
| BA23 3          | Anopheles strodei | São José da Vitória (Bahia) | -15.087060, -39.341560 | KC330296, KC330308 |
| BA25 4          | Anopheles strodei | São José da Vitória (Bahia) | -15.090910, -39.343700 | KC330297, KC330309 |
| ES09 1          | Anopheles strodei | Santa Teresa (Espírito Santo) | -19.916667, -40.600000 | GU226664, FJ178875, GU226730 |
| ES09 3          | Anopheles strodei | Santa Teresa (Espírito Santo) | -19.916667, -40.600000 | GU226665, FJ178874, GU226731 |
| MG27 108        | Anopheles strodei | Coronel Pacheco (Minas Gerais) | -21.587778, -43.265833 | GU226669, GU226693, GU226735 |
| MG30 102        | Anopheles strodei | Coronel Pacheco (Minas Gerais) | -21.587778, -43.265834 | GU226670, GU226694, GU226736 |
| MG33 9 1        | Anopheles strodei | Oliveira (Minas Gerais) | -20.745598, -44.915613 | KC330242, KC330298, KC330333 |
| PR20 4 3        | Anopheles strodei | São Miguel do Iguaçu (Paraná) | -25.265361, -54.545833 | KC330233, KC330299, KC330307 |
| PR29 23 3       | Anopheles strodei | Foz do Iguaçu (Paraná) | -25.480556, -54.586667 | GU226671, GU226695, GU226737 |
| RS37 9 8        | Anopheles strodei | Maquiné (Rio Grande do Sul) | -29.589556, -50.262639 | KC330236, KC330300, KC330310 |
| SP07 6          | Anopheles strodei | Buri (São Paulo) | -23.800000, -48.566670 | GU226674, FJ178878, GU226740 |
| SP104 18 1      | Anopheles strodei | Pindamonhangaba (São Paulo) | -22.960472, -45.452083 | KC330240, KC330299, KC330314 |
| SP105 10 12     | Anopheles strodei | Pindamonhangaba (São Paulo) | -22.999333, -45.495361 | KC330241, KC330302, KC330315 |
| SP27 1          | Anopheles strodei | Lucélia (São Paulo) | -21.618861, -50.940000 | GU226672, GU226696, GU226738 |
| SP29 121        | Anopheles strodei | Lucélia (São Paulo) | -21.618861, -50.940000 | GU226675, GU226698, GU226741 |
| SP31 101        | Anopheles strodei | Inubia Paulista (São Paulo) | -21.681417, -50.919889 | KC330232, GU226697, KC330306 |
| SP56 33         | Anopheles strodei | Mariporã (São Paulo) | -23.318889, -46.586944 | KC330238, KC330303, KC330312 |
| SP56 8          | Anopheles strodei | Mariporã (São Paulo) | -23.318889, -46.586944 | KC330237, KC330304, KC330311 |
| SP66 15 1       | Anopheles strodei | Dourado (São Paulo) | -22.134694, -48.391722 | KC330239, KC330305, KC330313 |
| VP05 11A        | Anopheles strodei | Pindamonhangaba (São Paulo) | -22.959750, -45.452389 | GU226668, FJ178877, GU226734 |
| VP06 5 2        | Anopheles strodei | Pindamonhangaba (São Paulo) | -22.959750, -45.452389 | GU226667, FJ178876, GU226733 |
| VP06 6 4        | Anopheles strodei | Pindamonhangaba (São Paulo) | -22.959750, -45.452389 | GU226666, GU226692, GU226732 |
(5′-TAA ACT TCA GGG TGA CCA AAA ATC A-3′) primers [24]. The Polymerase Chain Reaction (PCR) was carried out in a 25-μL aqueous reaction mixture containing 1 μL of DNA extraction solution, 1X PCR buffer (Invitrogen), 1.5 mM MgCl2 (Invitrogen), 1.25 μL dimethyl sulfoxide (Sigma, St. Louis, MO, USA), 0.1 μM of each primer, 0.2 mM each dNTP (Amresco, Solon, OH, USA) and 1.25 U Taq Platinum polymerase (Invitrogen). The reaction proceeded under the following temperature profile: 95°C for 2 min, 35 cycles of 94°C for 1 min, 57°C for 1 min and 72°C for 1 min and a final extension at 72°C for 7 min.

**ITS2 region**

This region was amplified using 5.8SF (5′-ATC ACT CGG CTC GTG GAT CG-3′) and 28SR (5′-ATG CTT AAA TTT AGG GGG TAG TC-3′) primers [25]. The PCR was carried out in a 25-μL aqueous reaction mixture containing 1 μL of DNA extraction solution, 1X PCR buffer (Invitrogen), 1.5 mM MgCl2 (Invitrogen), 1.25 μL dimethyl sulfoxide (Sigma), 0.1 μM of each primer, 0.2 mM each dNTP (Amresco) and 1.25 U Taq Platinum polymerase (Invitrogen). The reaction proceeded under the following temperature profile: 94°C for 2 min, 34 cycles of 94°C for 30 s, 57°C for 30 s and 72°C for 30 s and a final extension at 72°C for 10 min. ITS2 amplicons that yielded ambiguous sequence chromatograms, which is suggestive of intragenomic variation, were purified using PEG precipitation (20% polyethylene glycol 8,000/2.5 M NaCl) and then cloned into pGem-T Easy Vector (Promega, Madison, WI).

**White gene**

This gene was amplified using WZ2E and WZ11 primers [26]. This amplification product then served as a template in a sequencing reaction using internal primers W1F (5′-GAT CAA RAA GAT GTG YGA CTC GTT-3′) and W2R (5′-GCC ATC GAG ATG GAG GAG CTG-3′). Both PCRs were carried out in a 25-μL aqueous reaction mixture containing 1 μL DNA extraction solution, 1X PCR buffer (Invitrogen), 1.5 mM MgCl2 (Invitrogen), 2.5 μL of dimethyl sulfoxide (Sigma), 2.0 μM of each primer, 0.2 mM each dNTP (Amresco) and 2.5 U Taq Platinum polymerase (Invitrogen). Both PCRs proceeded under the following temperature profile: 94°C for 5 min, 35 cycles at 94°C for 30 s, an annealing temperature of 50°C for 1 min and then 72°C for 2 min followed by a final extension at 72°C for 10 min. Any *white* amplicons that yielded ambiguous sequence chromatograms were purified using PEG precipitation (20% polyethylene glycol 8,000/2.5 M NaCl) and then cloned into pGem-T Easy Vector (Promega).

**Sequencing and sequence alignment**

Sequencing reactions were carried out in both directions using a Big Dye Terminator cycle sequencing kit v3.1 (Applied Biosystems, Foster City, CA, USA) and Applied Biosystems 3130 DNA Analyzer (Applied Biosystems). The *COI* and *white* gene sequences were aligned first by nucleotides using the Muscle algorithm [27] implemented in SeaView [28] and then by amino acid using TranslatorX [29].

The ITS2 sequences were annotated for the 5.8S and 28S ends using the ITS2 annotation tool [30] in the ITS2 Database [31]. ITS2 secondary structure was then predicted for each sequence using Mfold [32] and the sequence that gave the lowest minimum free energy, ΔG, was used as a template to model the secondary structure of sequences using the Custom Modeling tool at the ITS2 Database. Sequences with secondary structures were then aligned and edited in 4Sale [33,34]. Sequence edits were performed in Bioedit [35].

**Phylogenetic analysis**

Bayesian analysis was applied to *COI*, ITS2, *white* and combined gene sequence data using partitioning schemes to allow different partitions to have their own model characteristics (composition, rate matrix and among-site variation) and to allow for among-partition rate variation. Optimal evolutionary models were determined for each partition using the Akaike Information Criterion (AIC) in jModelTest 2 ([36]; Additional file 1). Optimal partition schemes were calculated using Bayes factors [37]. All Bayesian analyses were performed using MrBayes [38] on Bioportal [39] and each analysis consisted of two simultaneous runs, which were then repeated to provide confirmation of convergence of posterior probability distribution. While all ITS2 clones were included in the isolated gene analysis, only a single randomly selected ITS2 clone from each individual was included in the combined gene analysis.

For all Bayesian analyses, each run was 12 million generations long and the first six million were discarded as burn-in. The Metropolis-coupled Markov chain Monte Carlo strategy was used with six heated chains; adequate mixing was achieved by setting the chain temperature to between 0.1 and 0.2. Convergence of topology between the two runs was monitored using the average standard deviation of split frequencies - this index consistently fell to below 0.015 in the post-burn-in samples. Convergence was also monitored by noting the potential scale reduction factor values - these values were all approximately 1.0 in the post-burn-in samples. Consensus trees were constructed containing nodes with posterior probability support greater than 70%. Trees were drawn using the R package APE [40].

**Barcoding analysis**

Individual pairwise Kimura-two-parameter (K2P) [41] distance matrices were constructed for *COI*, ITS2 and
combined COI-ITS2 using APE. All ITS2 clones were included in this analysis, and these were combined with the corresponding COI sequence for each individual in the combined COI-ITS2 dataset. K2P Neighbor Joining (NJ) trees were constructed using Mega [42], with 10,000 bootstrap replicates. Minimum inter-specific and maximum intra-specific distances for each individual was calculated using the R package SPIDER [43]. The utility of these genes for barcoding was further tested using the “Best Close Match” (BCM) algorithm in TaxonDNA v1.7.8 [44]. This algorithm involves matching the query sequence to the most similar barcode within a specified species threshold. The query is then assigned the species name if it is within the 95th percentile of all intraspecific distances. The use of such a threshold offers advantages over arbitrary species identification thresholds as it is rigorously derived and can account for differences in mutation rate among loci and divergence among taxa.

Results

Phylogenetic analysis

A total of 61 individuals from the Strodei Subgroup were included in the analysis. After alignment these yielded 53 unique COI sequences of 638 base pairs in length, 49 unique ITS2 sequences of 432 base pairs in length, and

Figure 1 Bayesian tree of combined COI, white and ITS2 sequences from the Anopheles strodei subgroup. The data were partitioned by gene and codon. Numbers at branches indicate Bayesian posterior probability (> 70%). Anopheles galvaoi was included as an outgroup.
57 unique white sequences of 716 base pairs in length (including the intron of 109 base pairs in length). This gave a combined data set of 61 unique sequences of 1786 base pairs in length. Anopheles kompi Edwards (COI and white GenBank accession no. JF923715 and JN413731, respectively), Anopheles lutzii Cruz (COI and white GenBank accession no. JF923668 and JN392485, respectively), and Anopheles galvaoi Causey (COI, ITS2 and white GenBank accession numbers were KC330264, KC330295 and KC330337, respectively) were used as outgroup taxa. Anopheles kompi and An. lutzii could not be aligned at the ITS2 locus. The ITS2 locus was left un-partitioned for the Bayesian analysis, whereas, the best partition schemes for COI and white were those that partitioned by codon position with among-partition rate variation. The best partition scheme for the combined locus dataset was one that partitioned by locus and codon position.

Results of Bayesian analyses showed support for six clades in the combined gene tree (Figure 1). Anopheles CP Form was resolved from all other individuals across all gene trees. In the white gene (Figure 2), it was found as a sister to one of the outgroup taxa (An. galvaoi) and to a clade containing the remaining An. strodei subgroup. Anopheles arthuri s.l. individuals were resolved from others across all gene trees (Figures 1, 2, 3, and 4). There was no evidence for divergence among An. arthuri s.l. individuals at ITS2 and white genes, and at

Figure 2 Bayesian tree of white sequences from the Anopheles strodei subgroup. The data were partitioned by codon position with among partition variation. Numbers at branches indicate Bayesian posterior probability (≥ 70%). Anopheles kompi, Anopheles lutzii and Anopheles galvaoi were included as outgroup taxa.
the ITS2 locus there was intra-genomic variation. Individuals that required cloning yielded between 2 and 6 clones and this intra-genomic variation (0.26% - 1.09% K2P) frequently exceeded inter-genomic variation. However, An. arthuri s.l. was resolved into four geographically meaningful clades in the COI gene tree (Figure 4). These four clades were found across Brazil (Figure 5), in the central/southern Brazilian states of Goiás, Minas Gerais and São Paulo (72% Bayesian Posterior Probability, BPP; herein denoted An. arthuri A), the northern state of Ceará (91% BPP; denoted An. arthuri B), the western Amazonian state of Rondônia (94% BPP; denoted An. arthuri C) and southern Minas Gerais state (100% BPP; denoted An. arthuri D), with the last being a sister to the Ceará clade (87% BPP). An. arthuri A, An. arthuri B, An. arthuri C, & An. arthuri D

Barcoding analysis
The Barcode NJ tree for COI (Figure 6) shows six clear groups. Individuals from An. arthuri s.l. can be found in the same four separate groups as found in the phylogenetic analysis. Figure 7 (a) shows a histogram of all intra- and inter-specific K2P COI differences among individuals and Figure 7 (b) shows a histogram of maximum intra- and minimum inter-specific K2P COI differences among individuals, when ordered into clades as defined by the phylogenetic analysis. Distances are measured in 0.001 (0.1%) intervals. There are no barcoding gaps present in either histogram, and the intra- versus inter-specific distances shows a very high degree of overlap.

The Barcode NJ tree for ITS2 (Figure 8) shows four clear groupings – An. arthuri s.l., An. CP Form, An. albertoi, and An. strodei. Figure 7 (c) and (d) show histograms of all intra- and inter- specific K2P ITS2 distance among individuals, and maximum intra- and minimum inter-specific K2P ITS2 distances among individuals, respectively, when ordered into clades as defined by the phylogenetic analysis. Again, there are no barcoding gaps present, and the intra- versus inter-specific distributions shows a very high degree of overlap.

The BCM analyses further explored the intra- and inter-specific distances in the COI (Additional file 2) and ITS2 (Additional file 3) barcodes. Threshold values for 95% of all intra-specific distances were determined for each barcode to evaluate whether a query (matching a test sequence to a reference sequence) had a close enough barcode match for identification. These were 1.92% for COI and 1.06% for ITS2. In total, 91.80% (n =...
56) of queries were correctly identified by the COI barcode according to the BCM criteria. The COI barcode was highly effective at correctly identifying queries from An. CP Form, An. arthuri A, An. arthuri B, An. arthuri C, and An. arthuri D. All queries from these five species were successfully matched to their respective species groups. However, all three queries from An. albertoi and two from An. strodei were not successfully matched. The three An. albertoi queries were incorrectly matched to An. strodei, the first An. strodei query was incorrectly matched to An. albertoi and the second An. strodei query was ambiguous as it was matched equally to both An. albertoi and An. strodei. The highest levels of intraspecific distances among all seven species were consistently from An. albertoi and An. strodei. Although intraspecific comparisons in the study ranged from 0% to 2.58%, all of the intraspecific comparisons above 1.27% (n = 232) were among An. albertoi and An. strodei COI barcodes and intraspecific comparisons above 2.00% (n = 32) were solely from An. strodei COI barcodes.

The BCM analysis for the ITS2 barcode found that only 59.55% (n = 53) of queries were correctly identified. All An. CP Form, An. albertoi and An. strodei queries were correctly matched to their respective species.
However, 39.32% \( (n = 35) \) of queries were ambiguous and 1.12% \( (n = 1) \) were incorrect and these came entirely from the \textit{An. arthuri} species.

The COI barcode, therefore, correctly identified all \textit{An. CP Form}, \textit{An. arthuri A}, \textit{An. arthuri B}, \textit{An. arthuri C}, and \textit{An. arthuri D}, while the ITS2 barcode correctly identified all \textit{An. CP Form}, \textit{An. albertoi} and \textit{An. strodei} individuals. A combined COI-ITS2 barcode was therefore tested first using a NJ tree (Figure 9) and then using the BCM analysis (with a 95% intraspecific variation threshold of 1.11%; Additional file 4). The results showed that all species could be resolved using the NJ tree and all BCM queries successfully identified \textit{An. CP Form}, \textit{An. arthuri}, \textit{An. strodei}, \textit{An. arthuri A}, \textit{An. arthuri B}, \textit{An. arthuri C}, and \textit{An. arthuri D}. This was despite maintaining a small degree of overlap between intra- and inter-specific distances due to inflated levels of genetic variation in \textit{An. strodei} (Figure 7 (e) and (f)).

**Discussion**

A recent study has added two additional species \((An. albertoi\) and \textit{An. arthuri}) to the \textit{An. strodei} subgroup [12]. It also found support for a distinct morphological form, referred to as “CP Form”, based on a single individual captured in the state of Paraná. In the current study we identified seven distinct lineages, of which three represented currently recognized species \((An. strodei, An. arthuri s.s./An. arthuri A and An. albertoi)\), and four are undescribed \((An. arthuri B, An. arthuri C, An. arthuri D and An. CP form)\).

The first important observation of the phylogeny is several incongruences among topologies generated from the DNA sequences. While ITS2 resolves \textit{An. strodei} and \textit{An. albertoi}, it fails to identify lineages within \textit{An. arthuri} s.l. The COI region, however, clearly resolves four \textit{An. arthuri} s.l. lineages, but fails to resolve \textit{An. albertoi} and \textit{An. strodei}. Differences between the gene genealogies and the species genealogy could be the result of incomplete lineage sorting or, in the case of ITS2, incomplete concerted evolution. In relation to incomplete lineage sorting, ancestral haplotypes can be retained in cases of recent speciation and/or large breeding populations, potentially resulting in the obscuring of phylogenetic signal among species. This process may explain the inability to resolve \textit{An. strodei} and \textit{An. albertoi} at the COI gene. Incomplete concerted evolution occurs when the rate of homogenization among copies in the ITS2 multi-gene family is insufficient to bring about fixation, potentially resulting in intra-genomic variation and shared haplotypes among closely related species. This process appears to be the cause of high levels of intra-genomic variation in several species of \textit{Anopheles}.
[45-49] and can potentially blur phylogenetic signal in some species, as appears to be the case among the *An. arthuri* s.l. lineages in the current study.

Our phylogenetic analysis supports distinction of *An. albertoi* and *An. arthuri* s.l. as in previous work [12], but also further splits *An. arthuri* s.l. into four distinct lineages (at the COI and combined gene tree). These lineages are geographically and ecologically distinct, and are herein referred to as *An. arthuri* A (from a central/southern Brazilian region of Goiás, Minas Gerais, and São Paulo), *An. arthuri* B (from the northern Brazilian state of Ceará), *An. arthuri* C (from the Amazonian state of Rondônia) and *An. arthuri* D (from southern Minas Gerais). The *An. arthuri* A lineage can be found in the Interior Forest Subregion of the Atlantic Forest, where seasonal semi-deciduous forest dominates [50]. Individuals from this lineage were found on both the western and eastern slopes of the Brazilian Highlands (Figure 5). Three of these individuals (MG07_1_100, MG07_10_106 and MG07_18_100) were previously included in an assessment of egg morphology using scanning electron microscopy [12] and were found to be representative of the *An. arthuri* type specimen. It is therefore likely that *An. arthuri* A identified in this study is representative of *An. arthuri* s.s. The *An. arthuri* B lineage is found in the Brejos Nordestinos Subregion of the Atlantic Forest. This subregion marks the extreme northern reach of the Atlantic Forest and consists mainly of seasonal semi-deciduous forest or dense ombrophilous forest “islands” covering isolated plateaus, which are surrounded by arid Caatinga lowlands [50]. Whereas the Atlantic Forest was until recently largely contiguous, the forests of Brejos Nordestinos were isolated much earlier, during the climatic cycles of the Pleistocene [51]. Populations from these forest islands are therefore likely to be subject to greater levels of divergence via genetic drift and barriers to gene flow. The *An. arthuri* C lineage is found in the southern reaches of the Amazonian river basin, to the north and west of the Parecis Mountains. We found no evidence for the presence of *An. strodei* in this region and that it is likely that previous reports of *An. strodei* found naturally infected with *Plasmodium vivax* in Rondônia [23] actually may refer to *An. arthuri* C. The ranges of *An. arthuri* A, *An. arthuri* B and *An. arthuri* C lineages are thus ecologically divergent, and appear to be highly allopatric (lineage sampling localities separated by more than 1600 km). Two individuals also exist which were collected from Oliveira in the state of Minas Gerais with COI haplotypes that are significantly distinct from all others in the complex (>2.92% variation). These individuals were collected from a site in the Rio Pará Valley, near the headwaters of the São Francisco and the Paraná

**Figure 6** Bootstrapped neighbor joining tree of COI sequences from the *Anopheles strodei* subgroup. Constructed with Kimura’s two parameter (K2P) distances and supported by 10,000 bootstrap replicates. All clades have greater than 70% bootstrap support.
Rivers, at an altitude of approximately 1,000 meters, in a largely un-forested landscape at the interface of Brazil’s Atlantic Forest and Cerrado eco-regions. They are found locally sympatric with *An. strodei* and *An. arthuri* in this mountain valley but their absence from all other localities indicates that this species may be confined to mountainous areas in the Brazilian Highlands. Their distinction from other species may have been shaped by the considerable topographical structure in this region, serving as a barrier to gene flow and isolating them from other populations, and the varying selective pressures that potentially exist across the enclosed humid habitat of the Atlantic Forest and the open dry habitat of the Cerrado. These distinct Rio Pará Valley haplotypes are,
therefore, tentatively identified as *An. arthuri* D, but clearly further sampling in more northerly localities in the São Francisco Valley is required to determine whether this represents a distinct species.

Previous analysis of the *An. strodei* subgroup found that *An. albertoi* can be distinguished morphologically, from its sister species by differences in the eggs (absence of a float) and male genitalia, and genetically, with the white and combined white-COI genes [12]. Using *An. albertoi* individuals from the study of Sallum et al. [12], we again differentiated this species from *An. strodei* and provide further genetic support for this lineage at the ITS2 gene. We have found the distribution of this species straddles the Brazilian Highlands, with individuals identified from the coastal forest of Serra do Mar in the state of São Paulo and the interior forest of the state of Minas Gerais, where it is found locally sympatric with *An. arthuri* A. The sampling associated with *An. strodei* is the most extensive among species in the study. Samples came from 14 different localities in six Brazilian states, some of which are separated by more than 2,000 km. Although there was genetic and morphological support for this species, the substantial range of intra-specific distance at COI (0–2.58%) can be contrasted with intra-specific distances found in other species in this study (all less than 1.59%) and the 1% species identification threshold proposed in Ratnasingham and Hebert [3]. Comparable data, i.e. intra-specific pairwise distance ranges, from other studies of *Anopheles* species are scant, but higher intra-specific COI distances have been observed across a range of well supported species from the butterfly family Lycaenidae Leach [8]. Although the distribution of *An. strodei* haplotypes does not demonstrate geographic partitioning and there is no apparent variation in morphology or habitat, the levels of intra-specific variability present may be indicative of a high degree of cryptic population genetic structure. A comprehensive population genetic study, which includes more samples (n > 20) from each of the 14 *An. strodei* localities detailed here, would help address this question and lead to a better understanding of the nature of genetic variation in this species.

The *An. CP Form* individuals have previously been resolved from other species in the *An. strodei* subgroup based on differences observed in the male genitalia of a single individual collected in Foz do Iguaçu in the state of Paraná [12]. In the current study we have included additional individuals morphologically identified as *An. CP Form* from Coronel Pacheco in the state of Minas Gerais and have found that all CP Form individuals can be resolved genetically across multiple genes. Although the *An. CP Form* collection sites (Foz do Iguaçu, Paraná and Coronel Pacheco, Minas Gerais) are confined to the Interior Forest subregion of the Atlantic Forest, they are
highly disparate, separated by more than 1,500 km. This lineage’s geographic distribution is further extended by its identification in the coastal state of Espírito Santo [52]. In addition, the lineage is found locally sympatric with other species from the An. strodei subgroup, namely An. strodei in the west, and both An. strodei and An. arthuri in the east.

Generally, the most closely related species in the complex, i.e. within the An. strodei/An. albertoi clade and within the An. arthuri clade, are not found sympatrically, which may indicate allopatric speciation is the most important mode of speciation in this complex. However, the one exception to this pattern is species that are found in Rio Pará Valley. Here we find both An. arthuri A and An. arthuri D (as well as An. strodei). It may be that the An. arthuri D clade represents a Brazilian Highland endemic as it has been unreported among more southerly and easterly localities, and that the southern limits of its range overlap with the northern limits of its sister species. However, further sampling through more northern localities of the São Francisco Valley and Brazilian Highlands is necessary to identify the breeding range of these species.

No single barcode was found to be effective at resolving all species identified from the phylogenetic analysis of the An. strodei subgroup. Neither COI nor ITS2 alone proved to be reliable as barcodes, largely because of their inability to resolve An. albertoi/An. strodei and An. arthuri species, respectively (as is evidenced by the considerable overlap between intra- and inter-specific differences). Many barcoding studies have demonstrated that the existence of substantial barcoding gaps permits effective species identification and discovery [7,53,54]. In closely related species, such as those found in species complexes, overlapping intra- and inter-specific variation are more likely and mainly due to processes such as incomplete lineage sorting [55]. However, although identification success generally declines with increasing overlap between intra- and inter-specific distances, studies have also shown that the existence of the barcoding gap does not predict the identification success of DNA barcoding [56,57]. In the current study we found that, although the COI and ITS2 barcodes do not have a barcoding gap and exhibit considerable overlap among the species identified through phylogenetic and morphological analysis, a combined COI-ITS2 barcode reduced the extent of overlap and provided a useful tool for species identification in the complex. An important advantage that the COI barcode has over the ITS2 barcode is the relative ease with which it can be aligned. The ITS2
barcode is highly variable in relation to indels, and alignment of ITS2 sequences in *Anopheles* becomes extremely difficult in any other species other than close relatives. Therefore, while the COI-ITS2 barcode may provide an effective species tool in other anopheline species complexes, ITS2 sequence alignment is a mitigating factor for its use in more distantly related species.

Several studies have demonstrated that the extent and scale of intra-specific sampling and the inclusion of closely related species can have a significant impact on the global application of barcodes [58-60]. While intra-specific variation will tend to increase with increased geographical sampling, due to isolation by distance and geographic structure, inter-specific variation will tend to decrease due to the inclusion of more closely related allopatrically distributed species [61]. The current study has attempted to sample from a diverse range of localities from across the complexes’ distribution (in nine Brazilian states) but most of the newly and tentatively identified species are clearly under-represented, numerically and geographically, particularly in the case of *An. albertoi* (n = 3) and *An. arthuri* D (n = 2). Also, although *An. arthuri* C is better represented in the study than the previous two species, the geographic distribution of these samples is quite limited versus potential *An. arthuri* C breeding habitat in the Amazon basin. Recent studies have found that sample sizes used in DNA barcoding are generally low [60,61] and that a sampling strategy of less than 20 individuals per species is unlikely to adequately represent intra-specific variation [60]. The shortcomings of the current study can therefore be addressed by future sampling in the anatomically disparate localities, particularly within the Brazilian Highlands and the Amazon basin.

**Conclusion**

We identified seven possible species in the *Anopheles strodei* subgroup, three of which are reported here for the first time. The role of these as potential vectors of malaria is largely unknown but *An. strodei* individuals previously found naturally infected with *Plasmodium vivax* in the Amazon region are likely to be *An. arthuri* C identified herein. We found poor support for the use of a single barcode for species identification in this Subgroup. Although single barcodes may be useful to estimate minimum levels of speciosity in complexes, we found significant numbers of ambiguous or incorrect query matches when using this approach and would caution against their use for effective species identification in Anopheline species complexes. Instead, we propose a combined COI-ITS2 barcode as a potentially useful tool for species identification in the *An. strodei* complex, but recommend further sampling of intra-specific variation in order to more effectively assess the utility of this multi-locus barcode.

**Additional files**

**Additional file 1:** Models used for locus and codon positions.

**Additional file 2:** Identification of species based on “best close match” and COI sequences.

**Additional file 3:** Identification of species based on “best close match” and ITS2 sequences.

**Additional file 4:** Identification of species based on “best close match” and combined COI-ITS2 sequences.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

MAMS and BPB conceived and designed the experiments. TEM carried out the molecular laboratory work. ESB and MAMS did the field collections and identified the specimens. BPB performed the data analysis. BPB wrote the paper, with contributions from MAMS. All authors read and approved the final manuscript.

**Acknowledgements**

We are grateful to the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, No. 2011/20397-7 to MAMS), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, No. 301666/2011-3 to MAMS) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, No. 29380.05274/2011-24 to LSR) for financial support. We are also indebted to Denise Sant’Ana who kindly helped us with field collections.

**Author details**

1 Departamento de Epidemiologia, Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo-SP, Brazil. Laboratório de Parasitologia, Instituto Butantan, Avenida Vital Brasil, São Paulo-SP, Brazil. 2Superintendência de Controle de Endemias, Secretaria de Estado da Saúde de São Paulo, Araraquara-SP, Brazil.

**Received:** 20 November 2012 **Accepted:** 14 April 2013 **Published:** 19 April 2013

**References**

1. Barnett HC: The transmission of western equine encephalitis virus by the mosquito *Culex tarsalis* Coq, Am J Trop Med Hyg 1956, 5:86–98.
2. Rosa-Freitas MG, Lourenço-de-Oliveira R, de Carvalho-Pinto CJ, Flores-Mendoza C, Silveira-Machado E, Do-Nascimento FF: Anopheline species complexes in Brazil. Current knowledge of those related to malaria transmission. Mem Inst Oswaldo Cruz 1998, 93:651–655.
3. Ratnasingham S, Hebert PDN: BOLD: The Barcode of Life Data System (www.barcodinglife.org). Mol Ecol Notes 2007, 7:355–364.
4. Hebert PD, Cywinska A, Ball SL, deWaard JR: Biological identifications through DNA barcodes. Proc R Soc Lond B 2003, 270:313–321.
5. Yao H, Song J, Liu C, Luo K, Han J, Li Y, Pang X, Xu H, Zhu Y, Xiao P, Chen S: Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals. PLoS One 2010, 5(10):e13102.
6. Dai QY, Gao Q, Wu CS, Chesters D, Zhu CD, Zhang AB: Phylogenetic reconstruction and DNA barcoding for closely related pine moth species (*Dendrolimus*) in China with multiple gene markers. PLoS One 2012, 7(4):e32544.
7. Hebert PD, Stoeckle MY, Zemlak TS, Francis CM: Identification of birds through DNA barcodes. PLoS Biol 2004, 2:1312.
8. Womersley MJ, Fiedler K: Does the DNA barcoding gap exist? - A case study in blue butterflies (Lepidoptera: Lycaenidae). Front Zool 2007, 4:8.
9. Hickerson MJ, Meyer CP, Moritz C: DNA barcoding will often fail to discover new animal species over broad parameter space. Syst Biol 2006, 55:729–739.
10. Hebert PD, Penton EH, Sums JM, Janzen DH, Hallwachs W: Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci USA 2004, 101:14812–14817.
11. Song H, Buhay JE, Whiting MF, Crandall KA: Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. Proc Natl Acad Sci USA 2008, 105(38):13486–13491.
12. Sallum MA, Foster PG, Dos Santos CL, Flores DC, Motoki MT, Bergo ES: Resurrection of two species from synonymy of Anophelines.
Utility of the white gene in estimating diversity metazoan invertebrates. Natural malaria infections in anophelines in Rondonia State, Brazilian CT:

MUSCLE: multiple sequence alignment with high accuracy and prediction.

BioEdit: a user-friendly biological sequence alignment editor and synchronous RNA sequence and secondary structure alignment and prediction.

Synchronous visual analysis and DNA barcoding across insect orders.

The importance of data partitioning and the utility of flavivirus and flavivirus-like sequences for pandemics. Mol Phylogenet Evol 2006, 39(2):358–367.

Malaria transmission through boundary vectors. Malar J 2008, 7:94.

Anopheles strodei (Diptera: Culicidae). spacer (ITS2) varies regionally in the Neotropical malaria vector, Anopheles aquasalis (Diptera: Culicidae). transcribed spacer rDNA in Neotropical malaria vector, Anopheles crucians complex (Diptera: Culicidae). J Med Entomol 2004, 41:392–401.

Fairley TL, Kilpatrick CW, Conn JE: Intragenomic heterogeneity of internal transcribed spacer rDNA in Neotropical malaria vector, Anopheles aquasalis (Diptera: Culicidae). J Med Entomol 2005, 42:795–800.

L. C; Willkonson RC: Intragenomic rDNA ITS2 variation in the Neotropical Anopheles (Nyssorhynchus) obitans complex (Diptera: Culicidae), J Hered 2006, 98(5):51–59.

Motoki MT, Bourke BP, Bergo ES, Silva AM, Sallum MAM: Systematic notes of Anopheles konderi and its first record in Paraná State, Brazil. J Am Mosq Control Assoc 2011, 27(1):91–100.

Gilindo-Leal C, Câmara KC: Atlantic Forest Hotspot Status: an Overview. In The Atlantic Forest of South America: Biodiversity status, threats, and outlook. Edited by Gilindo-Leal C, Câmara KC, Washington DC: Island Press; 2003:9–11.

Tabarelli M, Santos MM: Uma breve descrição sobre a história natural dos Brejos Nordestinos. In Brejos de Altitude em Pernambuco e Paraíba, História Natural, Ecologia e Conservação. Edited by Porto KC, Cabral JIP, Tabarelli M, Brasília: M. Ministério do Meio Ambiente; 2004:17–19.

Foster PG, Bergo ES, Bourke BP, Oliveira TMF, Nagaki SS, Sant’Anna DC, Sallum MAM: DNA barcoding and species confirmation in Anophelinae (Diptera: Culicidae), PLoS One 2013, 8(2):e54063.

Barrett RDH, Hebert PD: Identifying spiders through DNA barcodes. Can J Zool 2005, 83:481–491.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD: DNA barcodes distinguish species of tropical Lepidoptera. Proc Natl Acad Sci USA 2006, 103:968–971.

Elías M, Hill R, Willmott KR, Dasmahapatra KK, Brower AVZ, Mallet J, Jiggins CD: Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc Biol Sci 2007, 274:2881–2889.

Meier R, Zhang G, All F: The use of mean instead of smallest interspecific distances exaggerates the size of the barcoding gap and leads to misidentification. Syst Biol 2008, 57:803–813.

Virgilio M, Bacceljau N, Neves D, Meyer M: Comparative performances of DNA barcoding across insect orders. BMC Bioinformatics 2010, 11:206.

Funk DJ, Omland KE: Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst 2003, 34:397–423.

Miotz C, Giorio C: DNA barcoding: promise and pitfalls. PLoS Biol 2004, 2(10):e354.

Bergsten J, Bitton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, Hendrich L, Geier J, Herrmann J, Foster GN, Ribeira I, Nilsson AN, Barraclough TG, Vogler AP: The effect of geographical scale of sampling on DNA barcoding. Syst Biol 2012, 61:851–869.

Zhang A, He LJ, Crouzier RH, Muster C, Zhu CD: Estimating sample sizes for DNA barcoding. Mol Phylogenet Evol 2010, 54:1035–1039.

doi:10.1186/1756-3305-6-111

Cite this article as: Bourke et al.: A multi-locus approach to barcoding in the Anopheles strodei subgroup (Diptera: Culicidae), Parasites Vectors 2013 6:111.