

## DNA barcoding mosquitoes: advice for potential prospectors

NIGEL W. BEEBE<sup>1,2</sup><sup>1</sup> *University of Queensland, St Lucia, Brisbane, Australia*<sup>2</sup> *CSIRO, Brisbane, Australia*

(Received 14 September 2017; revised 29 January 2018; accepted 10 February 2018; first published online 22 March 2018)

## SUMMARY

Mosquitoes' importance as vectors of pathogens that drive disease underscores the importance of precise and comparable methods of taxa identification among their species. While several molecular targets have been used to study mosquitoes since the initiation of PCR in the 1980s, its application to mosquito identification took off in the early 1990s. This review follows the research's recent journey into the use of mitochondrial DNA (mtDNA) cytochrome oxidase 1 (COI or COX1) as a DNA barcode target for mosquito species identification – a target whose utility for discriminating mosquitoes is now escalating. The pros and cons of using a mitochondrial genome target are discussed with a broad sweep of the mosquito literature suggesting that nuclear introgressions of mtDNA sequences appear to be uncommon and that the COI works well for distantly related taxa and shows encouraging utility in discriminating more closely related species such as cryptic/sibling species groups. However, the utility of COI in discriminating some closely related groups can be problematic and investigators are advised to proceed with caution as problems with incomplete lineage sorting and introgression events can result in indistinguishable COI sequences appearing in reproductively independent populations. In these – if not all – cases, it is advisable to run a nuclear marker alongside the mtDNA and thus the utility of the ribosomal DNA – and in particular the internal transcribed spacer 2 – is also briefly discussed as a useful counterpoint to the COI.

Key words: Mosquito, cryptic species, mitochondria, COI, COX1, rDNA, ITS2.

## INTRODUCTION

Mosquitoes are one of the world's most recognizable insects, commonly found throughout tropical and temperate zones. Highly adaptive, these insects can also exist at high elevation and have been discovered well into the Arctic circle, although the Antarctic is currently free of their presence. The mosquitoes' family of Culicidae sits within the order Diptera (two-winged flies) which is itself divided into two sub-families – Anophelinae and Culicinae – comprising a total of 112 genera. Their species' biodiversity is extensive, mostly geographically structured and constantly adapting to environments to maximize reproductive output. As of February 2018, the diversity count numbered 3554 recognized species (Harbach, 2018).

## MOSQUITO SPECIES AS A TAXONOMIC UNIT

Because mosquitoes transmit pathogens that cause disease, our studies of them are motivated not only by the desire to complete the taxonomic inventory of their biodiversity but also by their importance in the human public health arena. This emphasis has seen mosquitoes, like many species, often distinguished through the prism of the modern synthesis of evolutionary biology in which allopatric reproductive isolation leads to the emergence of distinct

species (Mayr, 1942). These potentially interbreeding units are regarded as a fundamental category of biological organization: by definition, their mating leads to the production of fertile offspring (de Queiroz, 2005). And while reproductive isolation – whether through pre- or post-mating barriers, variance or ecological selection – can draw populations apart, if this isolation is maintained for sufficient time, post-mating barriers may develop that may include changes in chromosome organization and synteny (the localization of genes on a chromosome). This reinforces genetic isolation and completes the journey of an organism to becoming a distinct species (Ayala and Coluzzi, 2005; Hoffmann and Rieseberg, 2008).

But this journey towards to what we term 'species', and their genetic discontinuity, is rarely clean-cut: there can be fascinating genetic exchanges between closely related so-called 'species' (Mallet *et al.* 2016). Additionally mosquitoes often exist rather as closely related species complexes that, when studied intimately – as with the *Anopheles gambiae* complex in Africa – can be observed to be diverging at some parts of the genome under low recombination while other parts of the genome undergo gene flow (Besansky *et al.* 2003; Reidenbach *et al.* 2012). Do we split these into separate species or group them into one? The answer, in terms of medical entomology and vector biology, is more likely to revolve around the taxa's phenotype and its ability to transmit pathogens, adapt to

Corresponding author: University of Queensland, St Lucia, Brisbane, Australia. E-mail: [n.beebe@uq.edu.au](mailto:n.beebe@uq.edu.au)

*Parasitology* (2018), **145**, 622–633. © Cambridge University Press 2018. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. This article is published online by Cambridge University Press on 17 May 2018, subject to the Cambridge Core terms of use, available at <http://www.cambridge.org/core/terms>. <https://doi.org/10.1017/S0031182018000343>

control methods and exchange advantageous genes. In this way, population genetics and population genomics are revealing mosquito species' boundaries as semipermeable, with gene exchange reflecting the genome region, blurring what are and are not fully reproductively isolated species (Crawford *et al.* 2015). In particular, strong selection for advantageous alleles such as insecticide resistance can result in adaptive introgressions that can breach what we understand as species' boundaries (Norris *et al.* 2015).

#### MOSQUITO SPECIES IDENTIFICATION

Most described species of mosquitoes have been recognized through traditional morpho-taxonomy using differential morphology as Linnaeus originally intended. Detailed morphological keys and skilled entomological technicians are often required to key out species collected in adult mosquito traps or as larvae in their aquatic environment, and regional knowledge is incredibly important to this work.

The simplest way to identify an adult mosquito is by morphology, which makes it unfortunate that mosquito scales can be easily rubbed off or are often damaged when they are collected in adult mosquito traps or not stored carefully. Perhaps even less convenient for morphological identification is the fact that mosquitoes often exist in groups of closely related sibling species in what is called a species 'complex' (all of which are isomorphic) or within a larger related species 'group' (which contains individuals with overlapping morphology that can often include a complex). The cross-mating studies that would confirm species status by observing post-mating barriers will void any natural pre-mating barriers and are labour intensive as they usually require establishment of a colony, while the use of chromosome banding patterns to distinguish closely related cryptic *Anopheles* species, as pioneered in the 1980s by Mario Coluzzi (Coluzzi *et al.* 2002), is still used today (Coetzee *et al.* 2013). Again, this method works best for mosquitoes that have giant polytene chromosomes – and this renders its utility outside of *Anopheles* limited.

Molecular genetic studies are often undertaken on mosquitoes that transmit human pathogens and these frequently discover that sibling species groups can be common within morphological taxa that can also show differences in biology (i.e. human feeding, time of feeding) and ecology (i.e. oviposition site selection, geographic distribution). These variations can result in different pathogen transmission potentials. For example, our work on the *Anopheles punctulatus* group of Southwestern Pacific malaria vectors over the last 20 years has used molecular genetics and the development of DNA-based species diagnostic tools to identify 13 sibling species from its three described

morphological species – only five of which seem to be primary malaria vectors. [See our earlier review (Beebe *et al.* 2015) for a sense of this journey from the first identifications of the species as reproductively isolated, through the development of the molecular diagnostics, and the subsequent surveys that provided new insights into these malaria vectors]. The various ramifications of this in terms of vector control and the allocation of any available public health resources both underscore the crucial urgency of this type of work.

#### DNA BARCODES FOR MOSQUITOES

Over the past few decades, mosquito taxonomists have themselves become something of an endangered species, although whether this has led to or necessitated the shift to identifying mosquitoes using DNA sequences – more recently labelled barcodes – is a moot point. The utility of a single DNA sequence and the recent growth and utility of a mitochondrial DNA (mtDNA) cytochrome oxidase I gene (COI or COX1) sequence that can be compared universally is a reasonable starting point for categorizing mosquito biodiversity. While this review focuses mostly on the mtDNA COI, other DNA sequence targets for mosquitoes have been exploited including other mtDNA targets (ND4), nuclear gene targets (the white gene) and ribosomal DNA targets [internal transcribed spacers and ribosomal DNA (rDNA) genes]. Supplementary Table 1 provides a list of these targets and associated studies, and rDNA markers are also discussed briefly below in 'Other Mosquito Barcodes'.

The application of the mtDNA COI barcode approach has consistently grown since its original suggestion in the early 2000s (Hebert *et al.* 2003), and the utility of using a single sequence such as the COI continues to fulfil the prerequisite for 'Barcode of Life Data' systems (BOLD). These are sequences that can be easily amplified with a simple protocol; their sequence region is flanked by a conserved region in which reliable primers anneal: and in this way, the organism can be capably identified at a species level. For metazoans, the COI is used; for fungi, it is the ribosomal DNA (rDNA) internal transcribed spacers (ITS); and plants utilize a multi-locus barcode (see <http://www.boldsystems.org/> for more detail).

The fact that the maternally inherited COI had already enjoyed decades of use by population geneticists and molecular systematists as an evolutionary barometer contributes to its selection for this work as it appears to be an optimal tool for inferring evolutionary and demographic history as well as molecular taxonomy (Avise *et al.* 1987). Some of the advantages of the mtDNA COI include its universality (it is carried by all eukaryotic organisms); its relatively high copy number in the cell (which

is good for PCR); and its comparatively higher substitution rate over nuclear genes (which allows for good levels of discrimination between species). It also possesses a maternal inheritance with no (or very rare) recombination in mosquitoes, and this provides a single evolutionary history. Thus, mosquitoes have become good candidates for the Barcode for Life initiative which uses a single DNA sequence to describe biodiversity by identifying species (Ratnasingham and Hebert, 2007).

#### THE MITOCHONDRIA

The mitochondria are organelles found in all eukaryotic organisms. They are likely to be a remnant ancestral bacterial endosymbiont that encodes its own independent genome of 13 coding genes – no introns are present because of its prokaryote origin, and it represents a very small fraction of the organism's actual genome size. Multiple identical copies are often present in each cell and these function as a chemical power plant for the cell generating adenosine triphosphate. For more detail on the natural history of the mitochondria – particularly with regard to its use as an evolutionary marker – see the elegant review by Ballard and Whitlock (2004).

Multiple clonal copies within each cell make the PCR amplification of the mtDNA easier than parts of the nuclear DNA where the two paternal copies also need to be separated before sequencing. This can be achieved by either cloning or through an algorithm-driven reconstruction of paternal nuclear haplotypes post-sequencing using software like PHASE (Rozas *et al.* 2003). Separating paternal sequences with an algorithm carries with it its own issues, and these can require subsequent validation through cloning and sequencing (Garrick *et al.* 2010).

Being mostly the same clonal copy, the mtDNA is noted for having a much higher mutation rate than nuclear coding genes and this facilitates their utility as a relatively quickly evolving DNA marker (Brown *et al.* 1979). However, it is becoming more evident that some mosquitoes contain copies of non-functional nuclear pseudogene sequences of mitochondrial origin – or 'numts' (Richly and Leister, 2004). These numts may initially appear as heteroplasmy (presenting more than one type of mitochondrial genome or sequence in an individual) and the overlapping chromatogram peaks they generate through Sanger sequencing can make subsequent analyses problematic as multiple sequences are being read as one. Nonetheless, despite increasing reports of multiple variant copies in mitochondria, which often come from the massively parallel sequencing of other organisms such as humans (Just *et al.* 2015), the proportion of mosquito species with numts appears to be small overall. At present, it includes *Aedes aegypti* (Black and Bernhardt, 2009; Hlaing *et al.* 2009) and the *Culex*

*pipiens* group members (Behura *et al.* 2011). Traditionally, if heteroplasmy is suspected within a species, the PCR product can be cloned and a number of clones sequenced to identify the presence of mtDNA copies. Next-generation sequencing (NGS) will also reveal heteroplasmy as it can manifest in the presence of pileup files as rare mtDNA sequences.

#### THE MTDNA COI BARCODE

DNA barcoding has enjoyed rapid growth as a large-scale initiative for investigating biodiversity and an audience of followers is keen to exploit its simplicity. This is despite persistent warnings from systematists who see its conceptual foundation as problematic because it stems from exclusive reliance on mitochondria (Goldstein and DeSalle, 2011) and because of issues with inherited symbionts manipulating the maternal line (Hurst and Jiggins, 2005). The phylogenetic method normally used for these analyses is the relatively simple Neighbour-Joining (NJ) method with the Kimura 2-parameter (K2P) model initially suggested by Herbert in 2003 (Hebert *et al.* 2003). This tends to be preferred because this simpler model permits faster analyses with large datasets. Thus, the K2P model is prevalent throughout the literature, and while it assumes that transitions and transversions occur at different rates, frequencies of nucleotides are regarded as the same and an equal substitution process applies to all three codon positions (Kimura, 1980). The mosquito's mitochondrial genome shows the strong AT-bias and it is only really the third nucleotide within the codon (and sometimes second) that is most free to change without affecting the phenotype. With improved and more subtle evolutionary models available (Zinger and Philippe, 2016), one would hope that the contemporary increased computational ability now provided by most personal computers and online servers may lead to more sophisticated evolutionary models of nucleotide evolution, given that the limitations of the K2P model may well lead to underestimations of species' richness (Barley and Thomson, 2016; Zinger and Philippe, 2016).

In mtDNA barcoding, there is a phenomenon known as the 'barcoding gap': this is the separation or distance between the mean intraspecific sequence variability and the interspecific variability for congeneric COI sequences (Meyer and Paulay, 2005). If a gap exists, you can determine a cut-off value for the data identification as there would be no overlap between the interspecific and intraspecific distances. Thus the process of identifying a field-collected specimen to its species' level is relatively straightforward if it displays minimal intraspecific variability and large interspecific variability although determining this gap requires a substantial sampling design. Some systematists favour barcode-

species identifications based on only the smallest interspecific distance as the mean interspecific distances are artificially inflated – see work by Meier *et al.* for more detail on the differentiation methodologies of species (Meier *et al.* 2008).

#### BROAD-SCALE COI BARCODING

Before the mtDNA COI was employed as a standalone entity, it was often used alongside nuclear DNA sequence regions in studies of mosquito biodiversity, and this approach allowed sibling species to be teased apart in order to study their biology, behaviour and pathogen transmission potential (see the supplementary Table 1 for information on how the COI has been co-assessed alongside other markers in mosquito studies). Although the COI was used earlier as a population genetics tool in this way, the barcode concept itself was substantially forged in the 21st century (Hebert *et al.* 2003).

The first mosquito study to employ a dedicated COI barcoding approach came from researchers involved in DNA barcoding (Cywinska *et al.* 2006). Using mosquito collections from Canada that had initially been identified to species by morphology, they combined this information with additional mosquito COI sequences pulled from Genbank (Cywinska *et al.* 2006). Again, a relatively simple model of evolution promoting computational speed was used – an NJ analysis with K2P (Kimura, 1980). Outcomes from this study were reasonably compelling for species that were distantly related within a genus and barcode congruence (evidence of the same sequences) was found between some closely related species that were still morphologically distinct. Interesting data from that first study suggests that 98% of mosquito species were <2% divergent. The small divergence recorded may reflect the limited sampling of individuals within species, although one *Aedes* species encountered in this work was found to be 3.6–3.9% divergent. Surprisingly only one pseudogene generated detectable numts, and this was encouraging because it again suggested that numts may not be common in mosquitoes.

Indeed, morphological and molecular comparisons of the COI sequence do appear congruent in studies across genus levels, as attested by morphological species' studies on mosquito diversity from Argentina (Laurito *et al.* 2013), Australia (Batovska *et al.* 2016), China (Wang *et al.* 2012), India (Kumar *et al.* 2007), Singapore (Chan *et al.* 2014), Italy (Talbalaghi and Shaikevich, 2011), Iran (Azari-Hamidian *et al.* 2009) and Pakistan (Ashfaq *et al.* 2014). This latter study from Pakistan reveals intraspecific divergences at a maximum of 2.4% from over 1600 individuals from 24 taxa – a result in line with the original Canadian work. In addition to this, the COI barcode appears to complement taxa described by morphology within the generic and

subgeneric levels where morphology still has an important utility (Torres-Gutierrez *et al.* 2016). Most of these studies were performed on endemic species (with the necessary inclusion of some ubiquitous exotics), and the utility of the COI barcode as a correlate with morphology is perhaps most valuable at this relatively broadscale level: it complements traditional morphological taxonomy in cataloguing mosquitoes from regional landscapes. Unfortunately, many of these studies employ only a small intraspecific sample size and this may deliver a biased picture of the intraspecific variation within these taxonomic units by conveying inaccurate intraspecific divergences. Because of this, sampling across the full range of the species is advised. Even museum specimens can now be regarded as biobanks as new NGS technologies can obtain adequate molecular data from old specimens and overcome issues with DNA damage in them as well (Yeates *et al.* 2016).

#### FINE-SCALE COI BARCODING

Moving on from these comparisons with taxonomy by morphology where the COI barcode and morphology appear to correlate relatively well, we travel into the darkness of closely related sibling and cryptic species groups and complexes, where species morphology is either polymorphic for diagnostic characters or isomorphic, with multiple species hidden under the same morphology. In this area, a barcode can indeed shed important light on divergent lineages or hypothetical species that may represent reproductively isolated taxa. For example, studies on *Culex* species in Australasia have identified novel divergent lineages, one of which correlates to the southern limit of Japanese encephalitis activity in the region (Hemmerter *et al.* 2007). When this study was followed up with a nuclear sequence from the acetylcholine esterase 2 (*ace2*), the nuclear marker supported the discovery of species' level reproductive isolation (Hemmerter *et al.* 2009).

However not all investigations deliver such clear-cut results: a case study from Argentina and Brazil that sought to resolve several *Culex* taxa using the COI only managed to resolve 69% of species with the remaining unresolved individuals registering as ambiguous (10%), misidentified (18%), or unidentified (3%) (Laurito *et al.* 2013). Recently diverged *Culex* mosquitoes, especially those in the *Cx. pipiens* group, can show insufficient variation at either the COI or the rDNA ITS2 for classification (Crabtree *et al.* 1995; Danabalan *et al.* 2012; Batovska *et al.* 2017) and they are notorious for showing hybrids where species' distributions overlap (Farajollahi *et al.* 2011; Tahir *et al.* 2016).

But at this finer scale, and with good field sampling, the COI barcode has proved useful in recent anopheline biodiversity studies in both Africa and

the Western Pacific (Lobo *et al.* 2015; Laurent *et al.* 2016). Given that it is maternally inherited through the female egg – and so is excluded from the direct influence of sex – the mtDNA cannot reveal reproductive isolation like the nuclear DNA. It can, however, reveal the presence of divergent lineages reflecting the kind of long-standing isolation that permits mtDNA lineages to fully differentiate into separate genetic clades (lineage sorting). On the other hand, problems may manifest when mtDNA markers are used to discriminate recently diverged species. We see this phenomenon in well-studied groups of cryptic mosquito species such as the African *An. gambiae* complex where lineages have not fully sorted into divergent clades or where mtDNA introgression may still be shuffling mitochondria between what we have previously called distinct species (Thelwell *et al.* 2000; Donnelly *et al.* 2004). Indeed the *An. gambiae* complex gives us one of the best illustrations of the complexity of determining reproductive isolation in recently diverged cryptic species as it permits us to peer into the fascinating evolutionary dynamics occurring in diverging populations (Weetman *et al.* 2014; Mallet *et al.* 2016). The *An. gambiae* complex is likely the rule rather than the exception of what can be occurring in recently diverged populations.

Despite the mitochondrial genome displaying in general a higher mutation rate and smaller effective population size (one copy of the genome) than the nuclear genome (which delivers four copies of autosomal nuclear DNA to the next generation), the mtDNA should in theory fix alleles much faster than nuclear DNA (Ballard and Whitlock, 2004). Nonetheless, incomplete lineage sorting of COI sequences between recently diverged species can result in species being overlooked because shared COI sequences still exist in both populations, as either ancestral haplotypes or through interspecies introgression events between related species (Donnelly *et al.* 2004; Bennett *et al.* 2015; Surendran *et al.* 2015). Figure 1 provides a simple graphical example of where barcode sharing occurs in the study of closely related or recently diverged species.

The mitochondria organelle can show metabolic differences that affect the phenotype, and experiments in *Drosophila* suggest its electron transport system is not impaired by introgression events (Pichaud *et al.* 2012), and so may facilitate the uptake of more fit mitochondria through such introgression events. If the mitochondria provide a selective advantage to the organism, the introgressed mitochondria can sweep through a population or species obscuring the evolutionary signal by

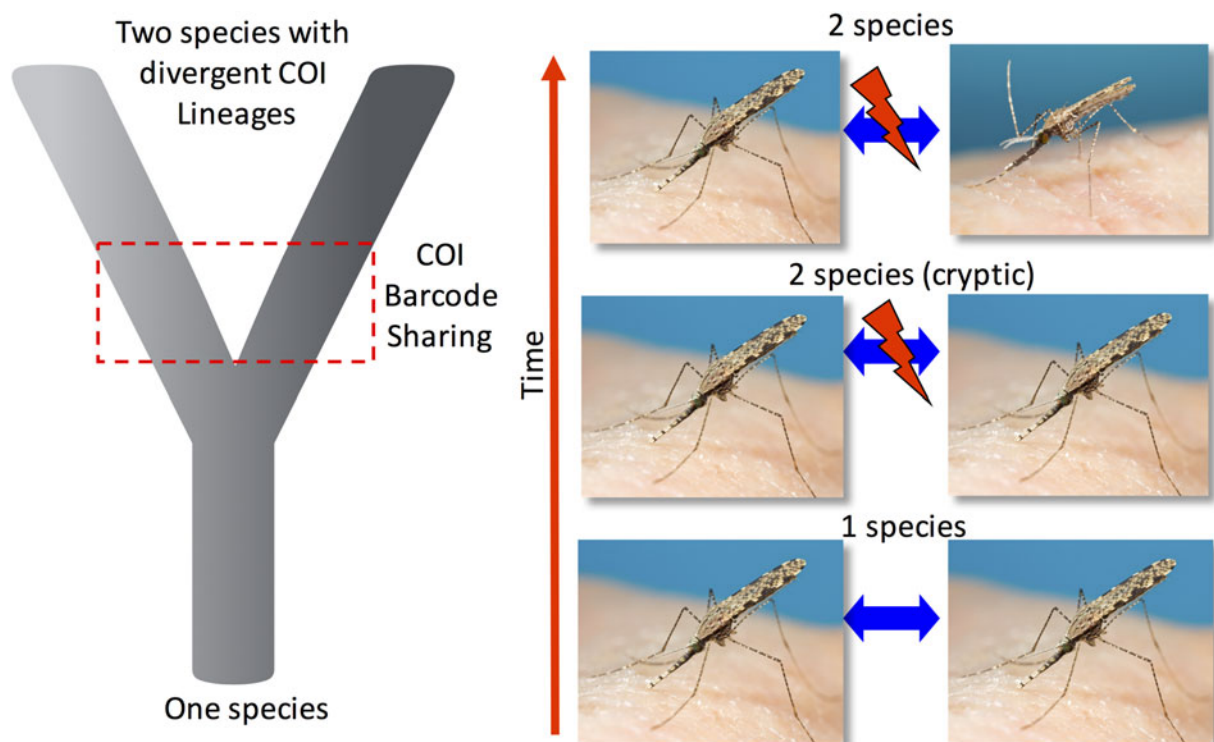


Fig. 1. A simplified representation of the progression of one species diverging into two reproductively isolated species. Often there is a period of time where the two recently discrete species are regarded as morphologically cryptic and cannot be distinguished by external morphology. During early stage of isolation shared mtDNA COI sequences can still be present and have not been fully sorted into discrete mtDNA COI lineages despite being reproductively incompatible. Additionally, mitochondrial genomes can be introgressed between closely related species through hybridization again resulting in mtDNA barcode sharing.

placing individuals in an incorrect taxonomic clade. For example, an evolutionary study of ours on two sympatric *Anopheles* species from the Southwest Pacific identified what appears to be a mitochondrial sweep from an introgression event between two related species (Ambrose *et al.* 2012). Figure 2 depicts this scenario where a genetically and geographically restricted population of the coastal restricted species *Anopheles farauti* appears paraphyletic for the COI in the phylogeny, and this introgressed population sits well within its sister species

*Anopheles hinesorum*, which also occurs inland and at elevation. Both species are reciprocally monophyletic for nuclear markers, but the *An. hinesorum*-like mtDNA sequences found in the *An. farauti* appear to have swept through the large population that spans from northeast Australia and southern New Guinea. If we had only used the COI marker on these populations, we would have mistakenly thought that *An. hinesorum* is saline water tolerant through some of its range, but this is not the case: only *An. farauti* is saline water tolerant (Sweeney *et al.* 1990).

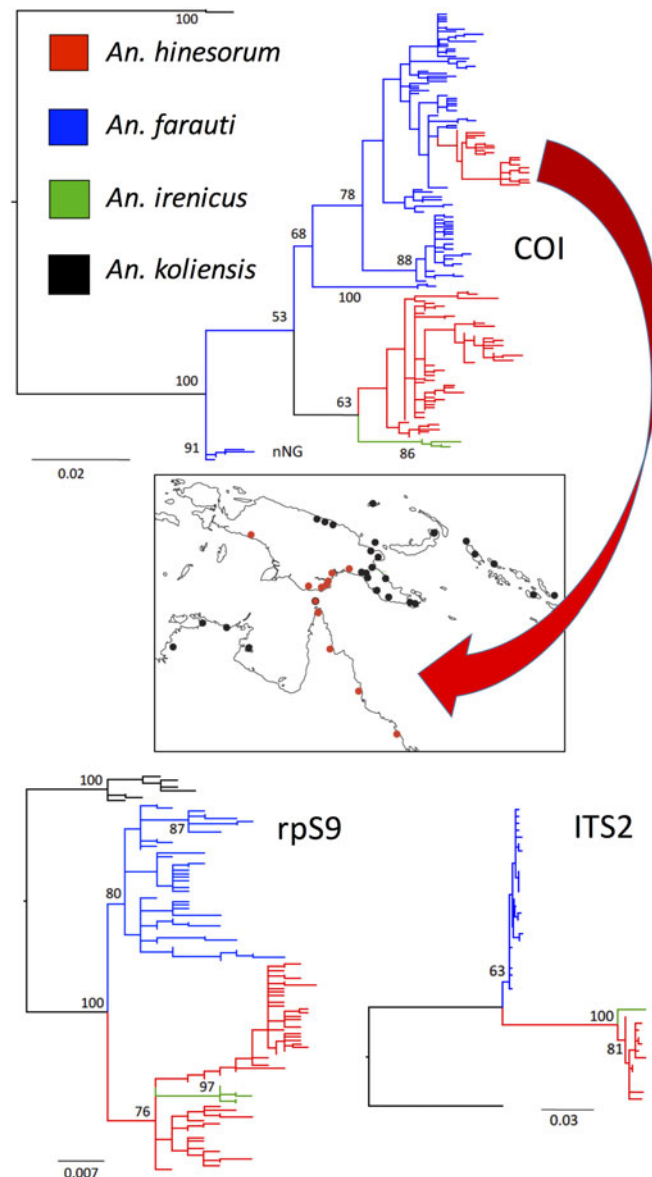


Fig. 2. Potential conflicting scenarios between the mtDNA and nuclear DNA. Cryptic species *An. farauti* (coastal), *An. hinesorum* (coastal and inland), *An. irenicus* (Solomon Islands restricted) and outgroup *An. koliensis* (New Guinea) were sequenced for the mtDNA COI and two nuclear markers [ITS2 and ribosomal protein S9 (rpS9)]. A PhyML analysis reveals that *An. farauti* and *An. hinesorum* are reciprocally monophyletic at both nuclear markers; however, a genetic and geographic population of *An. farauti* also appears paraphyletic for the mtDNA COI. This *An. farauti* COI population emerges within *An. hinesorum* (red branch in blue *An. hinesorum* tree). The distribution of this mtDNA population is also shown on the map in red and is most likely the result of past introgression with *An. hinesorum* followed by a mitochondrial sweep through a genetically and geographically restricted *An. farauti* population in northwest Australia and southern New Guinea (figure was modified from our study of Ambrose *et al.* 2012).

## INDIRECT SELECTION ON MTDNA

Arthropods often carry passenger microorganisms that exist in their cells and are passed from a female to her progeny through an egg. This phenomenon – whether its results are positive or negative – can place an indirect selection pressure on the mtDNA arising from linkage disequilibrium with the maternally inherited arthropod symbionts (Hurst and Jiggins, 2005). In mosquitoes, the effects of the maternally inherited symbiont *Wolbachia* on the evolution of mtDNA can be seen through the indirect selection they place on the mtDNA sequences/haplotype that will co-migrate through the maternal line. A neat example of this phenomenon was observed in an early study on *Drosophila* from Californian populations of *D. simulans* where a strain of *Wolbachia* swept through an uninfected population during the 1980s (at 100 km year<sup>-1</sup>) driven by the *Wolbachia*'s cytoplasmic incompatibility with uninfected wild types (Turelli *et al.* 1992). This led to the transformation of the population's mtDNA as it hitchhiked along with the symbiont. While a mechanism for *Wolbachia* cytoplasmic incompatibility has only recently been described (Beckmann *et al.* 2017; LePage *et al.* 2017), the symbiont's ability to obscure the mtDNA evolutionary signal through selected sweeps has been described in other Dipteran species such as fruit flies, where it reduces mtDNA diversity and drives new or rare haplotypes through populations (Whitworth *et al.* 2007; Nunes *et al.* 2008; Schuler *et al.* 2016). Outside of the *Cx. pipiens* group of mosquitoes (Rasgon *et al.* 2006), this phenomenon has not been well described, however artificially transformed *Wolbachia* infected *Ae. aegypti* are successfully being used to transform wild *Ae. aegypti* populations in a bid to reduce virus transmission (O'Neill, 2016), and it will be interesting to watch both the outcomes of these artificially induced selective sweeps and the ultimate genetic changes within the transformed and untransformed *Ae. aegypti* populations (Yeap *et al.* 2016).

## OTHER MOSQUITO BARCODES: RIBOSOMAL DNA

A PubMed search of literature about the utility of a DNA barcode using the terms 'mosquito, COI and species' calls up more than 150 papers. But recasting this search reveals that the most common DNA marker for mosquito identification has been the rDNA ITS2 (replacing COI with ITS2 as a search term). PubMed cites 223 papers that have employed this marker either alone or alongside other, of which 193 (~90%) have been on anopheline mosquitoes. These papers include studies on the spacers' utility as PCR-based tools for species diagnostics methods where the goal is not DNA sequencing but rather DNA genotyping for species based on specific

polymorphisms that can provide allele-specific primers (Porter and Collins, 1991), or on restriction analyses of PCR products (Beebe and Saul, 1995). It is important to note, too, that while the rDNA large subunit D2 and D3 regions are also used, the ITS2 appears to be the most published molecular marker for mosquitoes. The rDNA D2 and D3 subunit are part of the structural RNA gene, while the ITS2 is an intriguingly expedient spacer that separates two structural RNA genes (5.8S and 28S). Spliced out of the mature RNA, it appears to accommodate mutations more quickly than gene regions. Figure 3A gives a simplistic illustration of the rDNA gene family organization and the relative positioning of these markers.

Yet in many ways, the rDNA provides a peculiar DNA barcode target to use for species identification. It exists in the genome as a multicopied tandem gene family array, up to hundreds of copies in the metazoan genome, and it evolves through a non-Mendelian process (Dover, 2002). This evolutionary process is still not well described in and of itself: various theories to explain this pattern of concerted evolution where gene family units evolve together have been proposed, but we still do not have a unifying theory for this gene family evolution (Dover, 2002; Nei and Rooney, 2005; Eickbush and Eickbush, 2007). But despite the current deficiencies in detailing the rDNA evolutionary process, its intimate involvement in sex and its rapid evolutionary turnover give it crucial utility as a species-level marker because of its ability to manifest early genetic discontinuities (Bower *et al.* 2008), and to reveal cryptic species-level diversity (Paskewitz *et al.* 1993). Thus in many ways, the rDNA maintains a utility over the mtDNA COI in terms of the speed of lineage sorting of its multicopy rDNA array. As a comparable evolutionary marker for closely related species, rDNA's utility is often complementary to the COI (Alquezar *et al.* 2010; Ruiz-Lopez *et al.* 2013; Lobo *et al.* 2015). However, as it is rare to find heteroplasmy in the mitochondrial marker, it is not uncommon to observe intragenomic copy variants appearing in mosquito rDNA sequencing, making direct PCR-Sanger sequencing tricky to read and often causing chromatograms to collapse (Bower *et al.* 2008; Alquezar *et al.* 2010; Batovska *et al.* 2017). Because intra-individual sequence variants in rDNA copies that contain insertion/deletion indel variants can cause chromatograms to collapse, cloning prior to sequencing is often required. We have found that the decision to clone or not to clone can be assessed by running the rDNA PCR product through a native acrylamide gel: paired strands between different sequence variants can be visualized as they migrate more slowly in the gel. [See Fig. 3B for a graphic portrayal of visualizing intragenomic ITS2 variants as well as the

following citations (Beebe *et al.* 2001; Alquezar *et al.* 2010).]

The answer as to why the rDNA is so prevalent as a species diagnostic target for mosquitoes is probably due to its high rate of mutation and rapid DNA turnover within and between rDNA repeats (Eickbush and Eickbush, 2007; Bower *et al.* 2009; Alquezar *et al.* 2010). The rDNA gene family in *Anopheles* mosquitoes is positioned near the centromere on the sex chromosomes (Kumar and Rai, 1990). The reduced rate of recombination in this genomic landscape may influence the rate of genetic divergence (Nachman and Churchill, 1996; Stump *et al.* 2005). In this, genetic divergence can be observed to manifest, particularly in the rDNA in the face of apparent gene flow at other parts of the genome (Slotman *et al.* 2006; Weetman *et al.* 2012). Despite much of the published literature being based on *Anopheles* mosquitoes, the rDNA ITS2 spacer appears to also perform well as a species level marker with *Culex* mosquitoes (Vesgueiro *et al.* 2011). The rDNA ITS1 and ITS2 display utility as molecular diagnostics targets for other mosquito genera (Beebe *et al.* 2002; Beebe *et al.* 2007; Higa *et al.* 2010; Montgomery *et al.* 2017). The fast-

evolving rDNA transcribed spacers (such as the ITS2) have thus become useful markers for revealing early genetic discontinuities in populations and for providing species-level discrimination.

One must advocate caution as DNA sequence-based identification and the high rate of rDNA spacer evolution (turnover) can lead to geographically structured populations developing distinct ITS2 sequences as found in *An. farauti* in the Western Pacific (Beebe *et al.* 2000; Bower *et al.* 2008). The inability or perhaps complexity of detecting the presence of shared heterozygotes in these multigene family situations makes assessing gene flow signatures between geographically separated populations tricky. Indeed with sequence variation within seemingly identical species found in a Latin American anophelines study that used submitted GenBank ITS2 sequences revealed intraspecific variation ranging from 0.2 to 19% (Marrelli *et al.* 2006). The authors' wisely caution that a minimum requirement be considered for additional studies include voucher specimens, sampling for intraspecific variation and the use of other molecular markers.

For more details on the utility of these and other molecular markers used to study mosquitoes, a

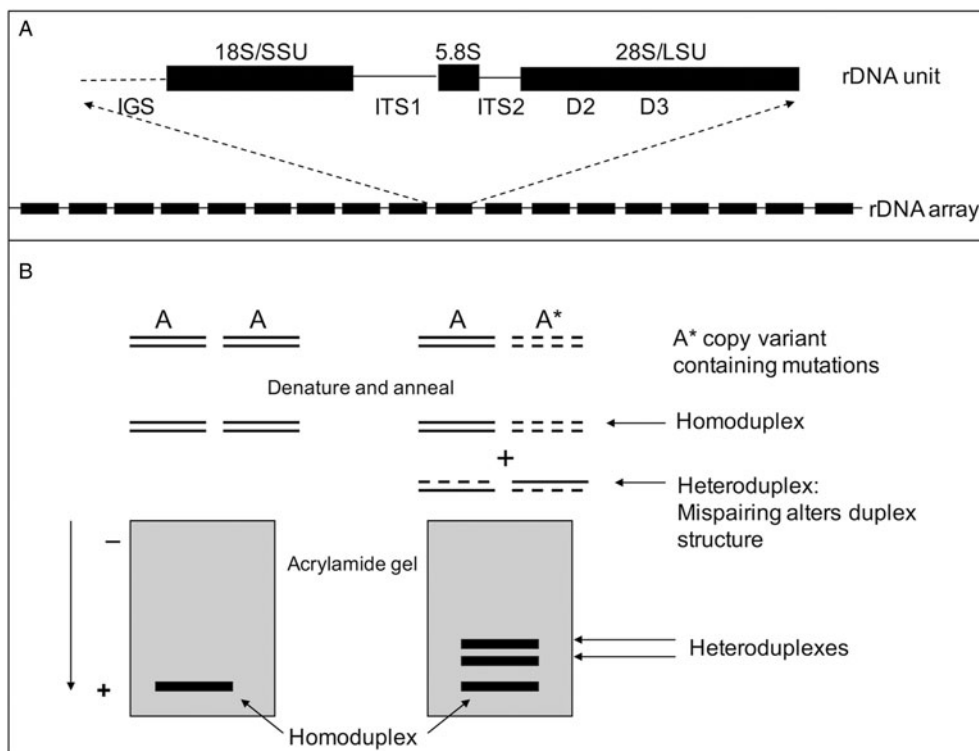


Fig. 3. The ribosomal DNA (rDNA) spacers have been popular as a genetic marker for species-level studies. The ribosomal gene family array [panel (A), rDNA unit] sits tandemly organized head to tail in the array, often 100 s of times usually positioned at a single locus in mosquitoes. The units in the array are observed to evolve together (in concert) although sequence variant copies can manifest in the array causing problems with traditional Sanger sequencing. These variant copies in the array can be seen by a simple acrylamide electrophoresis of the PCR product as variant copies that form mispaired double-stranded DNA duplexes (heteroduplexes) retard migration through the gel-forming laddering effect – more paralogues (variants) the more laddering occurs. Direct sequencing can fail and cloning was usually required prior to sequencing.



supplementary table (supplementary Table 1) is available as a summary of PubMed searches on genetic markers for mosquito identification. This table – of over 200 mosquito studies – is not exhaustive but provides an insight into the diversity of genomic regions used for taxonomic and genetic analyses of mosquito species and the ITS2 is by far the most commonly used. Indeed, the size of the ITS regions may have some bearing on the ability of the sequence to acquire non-deleterious mutations with longer sequences generally better able to accommodate changes than shorter sequences (Alquezar *et al.* 2010). This effect whereby ITS length can accommodate larger amounts of variation can be observed in studies of the neighbouring ITS1 where lengths can exceed 2000 bp and show large amounts of intraindividual and intraspecific variation that may be difficult to manage (Bower *et al.* 2008, 2009).

In regard to aligning rDNA spacer sequences, rDNA spacer sequences can contain large amounts of indels and repeat sequences and these can be tricky to align. Thus, computer-based alignments may require editing by eye (Beebe *et al.* 2000; Bower *et al.* 2008, 2009). Fortunately, there is now a ‘How to’ manual for molecular systematics to assist with ITS2 sequence alignments and to guide sequence alignments based on secondary structure (Schultz and Wolf, 2009).

#### SEQUENCING DNA BARCODES

While Sanger sequencing now provides a relatively cheap and simple means to acquire individual sequences (~\$US5.00/sequence at the time of writing), the advent of next-generation sequencing (NGS) platforms permits the parallel acquisition of DNA barcode sequences from numerous specimens simultaneously – after initial morphological classification down to species or species complex. These methods have been assessed and compared with traditional Sanger sequencing and found to be both superior and more efficient in terms of labour and cost (Shokralla *et al.* 2014; Shokralla *et al.* 2015; Batovska *et al.* 2017).

The first method employed 454 pyrosequencing on 190 Lepidoptera specimens to recover, after bioinformatics analysis, full-length DNA barcodes. Only 12.5% of a 454 sequencing run’s capacity had to be utilized to provide 143 sequence reads for each specimen. When compared with Sanger sequencing of the same 190 individuals – which delivered longer individual reads for each specimen than the 454 sequencing – the 454 showed a superior ability to discriminate species number, heteroplasmic sequences and nuclear mtDNA introgressions (Shokralla *et al.* 2014). The second method used a double dual-indexing approach on an Illumina MiSeq to identify 1010 specimens from 11 orders

of arthropods collected from a single Malaise trap sample from Area de Conservación Guanacaste in northwestern Costa Rica (Shokralla *et al.* 2015). Again, this alternative method proved better than Sanger sequencing of COI barcodes with the authors able to cite a 27% reduction in total time required, a 78% reduction in hands-on time, and a 79% reduction in laboratory costs. A similar Illumina MiSeq-based method has also been used for sequencing the ITS2 barcode of 26 species of mosquitoes collected from Australia which were compared to Sanger sequencing of the same samples (Batovska *et al.* 2017). The authors of this study also found superior resolution compared to the Sanger sequencing on the same individuals and could avoid the common difficulty of Sanger sequence chromatograms collapsing when individuals contain multiple ITS2 sequence variants with insertion/deletion indels.

The three methods described above all utilized PCR amplification of the barcodes regions prior to sequencing. Perhaps as NGS costs reduce, low-level genome skims using whole genomic DNA from individual specimens will prove to be the future for barcoding (Crampton-Platt *et al.* 2016), permitting the full reconstruction of high-copy DNA such as the mtDNA and rDNA.

#### CONCLUDING REMARKS

This review has focused mostly on the use of mtDNA COI DNA barcode for species identification of mosquitoes given that its utility for the study of mosquitoes is growing rapidly. For closely related species such as cryptic species groups and complexes, investigators should proceed with caution given problems with incomplete lineage sorting and introgression events where the same COI sequence may still appear in different species or may introgress across what we take to be species’ boundaries. In all cases, it is advisable to run a nuclear marker alongside the mtDNA, and the rDNA – in particular the ITS2 – can provide a useful counterpoint to the COI (Alquezar *et al.* 2010; Ajamma *et al.* 2016).

We are only at the beginning of this journey in linking mosquito species using initial morphological taxonomy with molecular characters. With there currently being ~3500 described species (Harbach, 2017), there is a paucity of molecular data available on these species and one would envisage this species number escalating as the many cryptic and undescribed species are assembled. It is useful to distinguish DNA barcoding from DNA-based taxonomy – both of which were proposed to support inefficiencies and difficulties using traditional morphology-based taxonomy and to permit non-taxonomists to develop species’ identification methods and tools. We often now see a process of

taxonomic understanding that combines morphological and molecular data that come together as an ‘integrative taxonomy’ (Teletchea, 2010). It seems important to reiterate that the Linnaean system of nomenclature should always be fundamental to describing the hierarchy of biodiversity down to species, and following the journey from primary literature through taxa classification by description and revision is paramount to any working understanding of species. This should also continue as advances in DNA sequencing generate hypotheses for the discovery and delineation of new species. More useful information on describing biodiversity can be found in reviews by Goldstein and DeSalle, and Kress, Garcia-Robledo et al. (Goldstein and DeSalle, 2011; Yeates et al. 2011; Kress et al. 2015). A general warning is probably useful here as the available databases are likely littered with poorly identified species and incorrect sequences so investigators should beware of spurious hits from incorrectly identified species sequences. Finally, the move to next generation sequencing for mosquito barcoding is exciting as it would allow researchers to run multiple target barcodes for less cost and effort (Batovska et al. 2017). The sharing of bioinformatics pipelines would be strongly encouraged.

#### SUPPLEMENTARY MATERIAL

The supplementary material for this article can be found at <https://doi.org/10.1017/S0031182018000343>.

#### ACKNOWLEDGEMENTS

The author would like to thank David Yeates for comments on the manuscript and Steve Doggett for the mosquito photos used in Fig. 1.

#### FINANCIAL SUPPORT

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

#### REFERENCES

Ajamma, Y. U., Villinger, J., Omondi, D., Salifu, D., Onchuru, T. O., Njoroge, L., Muigai, A. W. T. and Masiga, D. K. (2016). Composition and genetic diversity of mosquitoes (Diptera: Culicidae) on Islands and Mainland shores of Kenya's lakes Victoria and Baringo. *Journal of Medical Entomology* **53**, 1348–1363.

Alquezar, D. E., Hemmerter, S., Cooper, R. D. and Beebe, N. W. (2010). Incomplete concerted evolution and reproductive isolation at the rDNA locus uncovers nine cryptic species within *Anopheles longirostris* from Papua New Guinea. *Bmc Evolutionary Biology* **10**, 392.

Ambrose, L., Riginos, C., Cooper, R. D., Leow, K. S., Ong, W. and Beebe, N. W. (2012). Population structure, mitochondrial polyphyly and the repeated loss of human biting ability in anopheline mosquitoes from the southwest Pacific. *Molecular Ecology* **21**, 4327–4343.

Ashfaq, M., Hebert, P. D., Mirza, J. H., Khan, A. M., Zafar, Y. and Mirza, M. S. (2014). Analyzing mosquito (Diptera: culicidae) diversity in Pakistan by DNA barcoding. *PLoS ONE* **9**, e97268.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. and Saunders, N. C. (1987). Intraspecific

Phylogeography – the mitochondrial-DNA bridge between population-genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489–522.

Ayala, F. J. and Coluzzi, M. (2005). Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America* **102**(Suppl. 1), 6535–6542.

Azari-Hamidian, S., Yaghoobi-Ershadi, M. R., Javadian, E., Abai, M. R., Mobedi, I., Linton, Y. M. and Harbach, R. E. (2009). Distribution and ecology of mosquitoes in a focus of dirofilariasis in north-western Iran, with the first finding of filarial larvae in naturally infected local mosquitoes. *Medical and Veterinary Entomology* **23**, 111–121.

Ballard, J. W. O. and Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729–744.

Barley, A. J. and Thomson, R. C. (2016). Assessing the performance of DNA barcoding using posterior predictive simulations. *Molecular Ecology* **25**, 1944–1957.

Batovska, J., Blacket, M. J., Brown, K. and Lynch, S. E. (2016). Molecular identification of mosquitoes (Diptera: Culicidae) in southeastern Australia. *Ecology and Evolution* **6**, 3001–3011.

Batovska, J., Cogan, N. O., Lynch, S. E. and Blacket, M. J. (2017). Using next-generation sequencing for DNA barcoding: capturing allelic variation in ITS2. *G3 (Bethesda)* **7**, 19–29.

Beckmann, J. F., Ronau, J. A. and Hochstrasser, M. (2017). A *Wolbachia* deubiquitylating enzyme induces cytoplasmic incompatibility. *Nature Microbiology* **2**, 17007.

Beebe, N. W. and Saul, A. (1995). Discrimination of all members of the *Anopheles punctulatus* complex by polymerase chain reaction–restriction fragment length polymorphism analysis. *American Journal of Tropical Medicine and Hygiene* **53**, 478–481.

Beebe, N. W., Cooper, R. D., Morrison, D. A. and Ellis, J. T. (2000). Subset partitioning of the ribosomal DNA small subunit and its effects on the phylogeny of the *Anopheles punctulatus* group. *Insect Molecular Biology* **9**, 515–520.

Beebe, N. W., Maung, J., van den Hurk, A. F., Ellis, J. T. and Cooper, R. D. (2001). Ribosomal DNA spacer genotypes of the *Anopheles bancroftii* group (Diptera: Culicidae) from Australia and Papua New Guinea. *Insect Molecular Biology* **10**, 407–413.

Beebe, N. W., Russell, T., Burkot, T. R. and Cooper, R. D. (2015). *Anopheles punctulatus* group: evolution, distribution, and control. *Annual Review of Entomology* **60**, 335–350.

Beebe, N. W., van den Hurk, A. F., Chapman, H. F., Frances, S. P., Williams, C. R. and Cooper, R. D. (2002). Development and evaluation of a species diagnostic polymerase chain reaction–restriction fragment-length polymorphism procedure for cryptic members of the *Culex sitiens* (Diptera: Culicidae) subgroup in Australia and the southwest Pacific. *Journal of Medical Entomology* **39**, 362–369.

Beebe, N. W., Whelan, P. I., Van den Hurk, A. F., Ritchie, S. A., Corcoran, S. and Cooper, R. D. (2007). A polymerase chain reaction-based diagnostic to identify larvae and eggs of container mosquito species from the Australian region. *Journal of Medical Entomology* **44**, 376–380.

Behura, S. K., Lobo, N. F., Haas, B., deBruyn, B., Lovin, D. D., Shumway, M. F., Puiu, D., Romero-Severson, J., Nene, V. and Severson, D. W. (2011). Complete sequences of mitochondria genomes of *Aedes aegypti* and *Culex quinquefasciatus* and comparative analysis of mitochondrial DNA fragments inserted in the nuclear genomes. *Insect Biochemistry and Molecular Biology* **41**, 770–777.

Bennett, K. L., Linton, Y. M., Shija, F., Kaddumukasa, M., Djouaka, R., Misinzo, G., Lutwama, J., Huang, Y. M., Mitchell, L. B., Richards, M., Tossou, E. and Walton, C. (2015). Molecular differentiation of the African Yellow Fever Vector *Aedes bromeliae* (Diptera: Culicidae) from Its Sympatric Nonvector Sister Species, *Aedes lili*. *PLoS Neglected Tropical Diseases* **9**, 1–19. ARTN e0004250.

Besansky, N. J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., Fontenille, D., Toure, Y. and Sagnon, N. (2003). Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 10818–10823.

Black, W. C. and Bernhardt, S. A. (2009). Abundant nuclear copies of mitochondrial origin (NUMTs) in the *Aedes aegypti* genome. *Insect Molecular Biology* **18**, 705–713.

Bower, J. E., Downton, M., Cooper, R. D. and Beebe, N. W. (2008). Intraspecific concerted evolution of the rDNA ITS1 in *Anopheles farauti sensu stricto* (Diptera: Culicidae) reveals recent patterns of population structure. *Journal of Molecular Evolution* **67**, 397–411.

Bower, J. E., Cooper, R. D. and Beebe, N. W. (2009). Internal repetition and intraindividual variation in the rDNA ITS1 of the *Anopheles punctulatus* group (Diptera: Culicidae): multiple units and rates of turnover. *Journal of Molecular Evolution* **68**, 66–79.

- Brown, W. M., George, M., Jr and Wilson, A. C.** (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 1967–1971.
- Chan, A., Chiang, L. P., Hapuarachchi, H. C., Tan, C. H., Pang, S. C., Lee, R., Lee, K. S., Ng, L. C. and Lam-Phua, S. G.** (2014). DNA barcoding: complementing morphological identification of mosquito species in Singapore. *Parasites & Vectors* **7**, 1–12.
- Coetzee, M., Hunt, R. H., Wilkerson, R., Della Torre, A., Coulibaly, M. B. and Besansky, N. J.** (2013). *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa* **3619**, 246–274.
- Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M. A. and Petrarca, V.** (2002). A polytene chromosome analysis of the *Anopheles gambiae* species complex. *Science* **298**, 1415–1418.
- Crabtree, M. B., Savage, H. M. and Miller, B. R.** (1995). Development of a species-diagnostic polymerase chain reaction assay for the identification of *Culex* vectors of St. Louis encephalitis virus based on interspecies sequence variation in ribosomal DNA spacers. *American Journal of Tropical Medicine and Hygiene* **53**, 105–109.
- Crampton-Platt, A., Yu, D. W., Zhou, X. and Vogler, A. P.** (2016). Mitochondrial metagenomics: letting the genes out of the bottle. *Gigascience* **5**, 1–11.
- Crawford, J. E., Riehle, M. M., Guelbeogo, W. M., Gnome, A., Sagnon, N., Vernick, K. D., Nielsen, R. and Lazzaro, B. P.** (2015). Reticulate speciation and barriers to introgression in the *Anopheles gambiae* species complex. *Genome Biology and Evolution* **7**, 3116–3131.
- Cywinska, A., Hunter, F. F. and Hebert, P. D. N.** (2006). Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology* **20**, 413–424.
- Danabalan, R., Ponsonby, D. J. and Linton, Y. M.** (2012). A critical assessment of available molecular identification tools for determining the status of *Culex pipiens* Sl in the United Kingdom. *Journal of the American Mosquito Control Association* **28**, 68–74.
- de Queiroz, K.** (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6600–6607.
- Donnelly, M. J., Pinto, J., Girod, R., Besansky, N. J. and Lehmann, T.** (2004). Revisiting the role of introgression vs shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the *Anopheles gambiae* complex. *Heredity (Edinb)* **92**, 61–68.
- Dover, G.** (2002). Molecular drive. *Trends in Genetics* **18**, 587–589.
- Eickbush, T. H. and Eickbush, D. G.** (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics* **175**, 477–485.
- Farajollahi, A., Fonseca, D. M., Kramer, L. D. and Marm Kilpatrick, A.** (2011). Bird biting? mosquitoes and human disease: a review of the role of *Culex pipiens* complex mosquitoes in epidemiology. *Infection Genetics and Evolution* **11**, 1577–1585.
- Garrick, R. C., Sunnucks, P. and Dyer, R. J.** (2010). Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology* **10**, 1–17.
- Goldstein, P. Z. and DeSalle, R.** (2011). Integrating DNA barcode data and taxonomic practice: determination, discovery, and description. *Bioessays* **33**, 135–147.
- Harbach, R. E.** (2018) Mosquito Taxonomic Inventory In. <http://mosquito-taxonomic-inventory.info/>.
- Hebert, P. D. N., Cywinska, A., Ball, S. L. and DeWaard, J. R.** (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B – Biological Sciences* **270**, 313–321.
- Hemmerter, S., Slapeta, J., van den Hurk, A. F., Cooper, R. D., Whelan, P. I., Russell, R. C., Johansen, C. A. and Beebe, N. W.** (2007). A curious coincidence: mosquito biodiversity and the limits of the Japanese encephalitis virus in Australasia. *BMC Evolutionary Biology* **7**, 1–11.
- Hemmerter, S., Slapeta, J. and Beebe, N. W.** (2009). Resolving genetic diversity in Australasian *Culex* mosquitoes: incongruence between the mitochondrial cytochrome c oxidase I and nuclear acetylcholine esterase 2. *Molecular Phylogenetics and Evolution* **50**, 317–325.
- Higa, Y., Toma, T., Tsuda, Y. and Miyagi, I.** (2010). A multiplex PCR-based molecular identification of five morphologically related, medically important subgenus *Stegomyia* mosquitoes from the genus *Aedes* (Diptera: Culicidae) found in the Ryukyu Archipelago, Japan. *Japanese Journal of Infectious Diseases* **63**, 312–316.
- Hlaing, T., Tun-Lin, W., Sombon, P., Socheat, D., Setha, T., Min, S., Chang, M. S. and Walton, C.** (2009). Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. *BMC Genetics* **10**, 1–12.
- Hoffmann, A. A. and Rieseberg, L. H.** (2008). Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology Evolution and Systematics* **39**, 21–42.
- Hurst, G. D. D. and Jiggins, F. M.** (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B – Biological Sciences* **272**, 1525–1534.
- Just, R. S., Irwin, J. A. and Parson, W.** (2015). Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Science International: Genetics* **18**, 131–139.
- Kimura, M.** (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution* **16**, 111–120.
- Kress, W. J., Garcia-Robledo, C., Uriarte, M. and Erickson, D. L.** (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution* **30**, 25–35.
- Kumar, A. and Rai, K. S.** (1990). Chromosomal localization and copy number of 18S + 28S ribosomal RNA genes in evolutionarily diverse mosquitoes (Diptera, Culicidae). *Hereditas* **113**, 277–289.
- Kumar, N. P., Rajavel, A. R., Natarajan, R. and Jambulingam, P.** (2007). DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology* **44**, 1–7.
- Laurent, B. S., Cooke, M., Krishnankutty, S. M., Asih, P., Mueller, J. D., Kahindi, S., Ayoma, E., Oriango, R. M., Thumloup, J., Drakeley, C., Cox, J., Collins, F. H., Lobo, N. F. and Stevenson, J. C.** (2016). Molecular characterization reveals diverse and unknown malaria vectors in the Western Kenyan Highlands. *American Journal of Tropical Medicine and Hygiene* **94**, 327–335.
- Laurito, M., de Oliveira, T. M. P., Almiron, W. R. and Sallum, M. A. M.** (2013). COI barcode versus morphological identification of *Culex* (*Culex*) (Diptera: Culicidae) species: a case study using samples from Argentina and Brazil. *Memorias Do Instituto Oswaldo Cruz* **108**, 110–U152.
- LePage, D. P., Metcalf, J. A., Bordenstein, S. R., On, J. M., Perlmutter, J. I., Shropshire, J. D., Layton, E. M., Funkhouser-Jones, L. J., Beckmann, J. F. and Bordenstein, S. R.** (2017). Prophage WO genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. *Nature* **543**, 243.
- Lobo, N. F., St Laurent, B., Sikaala, C. H., Hamainza, B., Chanda, J., Chinula, D., Krishnankutty, S. M., Mueller, J. D., Deason, N. A., Hoang, Q. T., Boldt, H. L., Thumloup, J., Stevenson, J., Seyoum, A. and Collins, F. H.** (2015). Unexpected diversity of *Anopheles* species in Eastern Zambia: implications for evaluating vector behavior and interventions using molecular tools. *Scientific Reports* **5**, 17952.
- Mallet, J., Besansky, N. and Hahn, M. W.** (2016). How reticulated are species? *BioEssays* **38**, 140–149.
- Marrelli, M. T., Sallum, M. A. M. and Marinotti, O.** (2006). The second internal transcribed spacer of nuclear ribosomal DNA as a tool for Latin American anopheline taxonomy - A critical review. *Memorias Do Instituto Oswaldo Cruz* **101**, 817–832.
- Mayr, E.** (1942). *Systematics and the Origin of Species*. Columbia University Press, New York.
- Meier, R., Zhang, G. Y. and Ali, F.** (2008). The Use of mean instead of smallest interspecific distances exaggerates the size of the 'barcoding gap' and leads to misidentification. *Systematic Biology* **57**, 809–813.
- Meyer, CP and Paulay, G** (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* **3**, 2229–2238, e422.
- Montgomery, B. L., Shivas, M. A., Hall-Mendelin, S., Edwards, J., Hamilton, N. A., Jansen, C. C., McMahon, J. L., Warrilow, D. and van den Hurk, A. F.** (2017). Rapid Surveillance for Vector Presence (RSVP): Development of a novel system for detecting *Aedes aegypti* and *Aedes albopictus*. *PLOS Neglected Tropical Diseases* **11**, e0005505.
- Nachman, M. W. and Churchill, G. A.** (1996). Heterogeneity in rates of recombination across the mouse genome. *Genetics* **142**, 537–548.
- Nei, M. and Rooney, A. P.** (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics* **39**, 121–152.
- Norris, L. C., Main, B. J., Lee, Y., Collier, T. C., Fofana, A., Cornel, A. J. and Lanzaro, G. C.** (2015). Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 815–820.
- Nunes, M. D., Nolte, V. and Schlotterer, C.** (2008). Nonrandom *Wolbachia* infection status of *Drosophila melanogaster* strains with different mtDNA haplotypes. *Molecular Biology and Evolution* **25**, 2493–2498.
- O'Neill, S. L.** (2016). *Wolbachia* mosquito control: tested. *Science* **352**, 526–526.
- Paskewitz, S. M., Wesson, D. M. and Collins, F. H.** (1993). The internal transcribed spacers of ribosomal DNA in five members

- of the *Anopheles gambiae* species complex. *Insect Molecular Biology* 2, 247–257.
- Pichaud, N., Ballard, J. W. O., Tanguay, R. M. and Blier, P. U.** (2012). Naturally occurring mitochondrial DNA haplotypes exhibit metabolic differences: insight into functional properties of mitochondria. *Evolution* 66, 3189–3197.
- Porter, C. H. and Collins, F. H.** (1991). Species-diagnostic differences in a ribosomal DNA internal transcribed spacer from the sibling species *Anopheles freeborni* and *Anopheles hermsi* (Diptera: Culicidae). *American Journal of Tropical Medicine and Hygiene* 45, 271–279.
- Rasgon, J. L., Cornel, A. J. and Scott, T. W.** (2006). Evolutionary history of a mosquito endosymbiont revealed through mitochondrial hitchhiking. *Proceedings of the Royal Society B – Biological Sciences* 273, 1603–1611.
- Ratnasingham, S. and Hebert, P. D. N.** (2007). BOLD: the barcode of life data system ([www.barcodinglife.org](http://www.barcodinglife.org)). *Molecular Ecology Notes* 7, 355–364.
- Reidenbach, K. R., Neafsey, D. E., Costantini, C., Sagnon, N., Simard, F., Ragland, G. J., Egan, S. P., Feder, J. L., Muskavitch, M. A. T. and Besansky, N. J.** (2012). Patterns of genomic differentiation between ecologically differentiated M and S forms of *Anopheles gambiae* in West and Central Africa. *Genome Biology and Evolution* 4, 1202–1212.
- Richly, E. and Leister, D.** (2004). NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution* 21, 1081–1084.
- Roza, J., Sanchez-DelBarrio, J. C., Messeguer, X. and Roza, R.** (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19, 2496–2497.
- Ruiz-Lopez, F., Wilkerson, R. C., Ponsonby, D. J., Herrera, M., Sallum, M. A. M., Velez, I. D., Quinones, M. L., Flores-Mendoza, C., Chadee, D. D., Alarcon, J., Alarcon-Ormasa, J. and Linton, Y. M.** (2013). Systematics of the Oswaldo Complex (*Anopheles*, *Nyssorhynchus*) in South America. *Parasites & Vectors* 6, 1–13.
- Schuler, H., Koppler, K., Daxbock-Horvath, S., Rasool, B., Krumbach, S., Schwarz, D., Hoffmeister, T. S., Schlick-Steiner, B. C., Steiner, F. M., Telschow, A., Stauffer, C., Arthofer, W. and Riegler, M.** (2016). The hitchhiker's guide to Europe: the infection dynamics of an ongoing *Wolbachia* invasion and mitochondrial selective sweep in *Rhagoletis cerasi*. *Molecular Ecology* 25, 1595–1609.
- Schultz, J. and Wolf, M.** (2009). ITS2 sequence-structure analysis in phylogenetics: a how-to manual for molecular systematics. *Molecular Phylogenetics and Evolution* 52, 520–523.
- Shokralla, S., Gibson, J. F., Nikbakht, H., Janzen, D. H., Hallwachs, W. and Hajibabaei, M.** (2014). Next-generation DNA barcoding: using next-generation sequencing to enhance and accelerate DNA barcode capture from single specimens. *Molecular Ecology Resources* 14, 892–901.
- Shokralla, S., Porter, T. M., Gibson, J. F., Dobosz, R., Janzen, D. H., Hallwachs, W., Golding, G. B. and Hajibabaei, M.** (2015). Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Scientific Reports* 5, 1–7.
- Slotman, M. A., Reimer, L. J., Theimann, T., Dolo, G., Fondjo, E. and Lanzaro, G. C.** (2006). Reduced recombination rate and genetic differentiation between the M and S forms of *Anopheles gambiae* s.s. *Genetics* 174, 2081–2093.
- Stump, A. D., Fitzpatrick, M. C., Lobo, N. F., Traore, S., Sagnon, N., Costantini, C., Collins, F. H. and Besansky, N. J.** (2005). Centromere-proximal differentiation and speciation in *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15930–15935.
- Surendran, S. N., Truelove, N., Sarma, D. K., Jude, P. J., Ramasamy, R., Gajapathy, K., Peiris, L. B. S., Karunaratne, S. H. P. P. and Walton, C.** (2015). Karyotypic assignment of Sri Lankan *Anopheles culicifacies* species B and E does not correlate with cytochrome oxidase subunit I and microsatellite genotypes. *Parasites & Vectors* 8, 1–7.
- Sweeney, A. W., Cooper, R. D. and Frances, S. P.** (1990). Distribution of the sibling species of *Anopheles farauti* in the Cape York Peninsula, northern Queensland, Australia. *Journal of the American Mosquito Control Association* 6, 425–429.
- Tahir, H. M., Kanwal, N. and Mehwish** (2016). The sequence divergence in cytochrome C oxidase I gene of *Culex quinquefasciatus* mosquito and its comparison with four other *Culex* species. *Mitochondrial DNA A* 27, 3054–3057.
- Talbalaghi, A. and Shaikevich, E.** (2011). Molecular approach for identification of mosquito species (Diptera: Culicidae) in Province of Alessandria, Piedmont, Italy. *European Journal of Entomology* 108, 35–40.
- Teletchea, F.** (2010). After 7 years and 1000 citations: comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA* 21, 206–226.
- Thelwell, N. J., Huisman, R. A., Harbach, R. E. and Butlin, R. K.** (2000). Evidence for mitochondrial introgression between *Anopheles bwambae* and *Anopheles gambiae*. *Insect Molecular Biology* 9, 203–210.
- Torres-Gutierrez, C., Berge, E. S., Emerson, K. J., de Oliveira, T. M. P., Greni, S. and Sallum, M. A. M.** (2016). Mitochondrial COI gene as a tool in the taxonomy of mosquitoes *Culex* subgenus *Melanoconion*. *Acta Tropica* 164, 137–149.
- Turelli, M., Hoffmann, A. A. and Mckechnie, S. W.** (1992). Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila-simulans* populations. *Genetics* 132, 713–723.
- Vesgueiro, F. T., Demari-Silva, B., Malafrente, R. D., Sallum, M. A. M. and Marrelli, M. T.** (2011). Intra-genomic variation in the second internal transcribed spacer of the ribosomal DNA of species of the genera *Culex* and *Lutzia* (Diptera: Culicidae). *Memorias Do Instituto Oswaldo Cruz* 106, 1–8.
- Wang, G., Li, C. X., Guo, X. X., Xing, D., Dong, Y. D., Wang, Z. M., Zhang, Y. M., Liu, M. D., Zheng, Z., Zhang, H. D., Zhu, X. J., Wu, Z. M. and Zhao, T. Y.** (2012). Identifying the main mosquito species in China based on DNA barcoding. *PLoS ONE* 7, 1–11.
- Weetman, D., Wilding, C. S., Steen, K., Pinto, J. and Donnelly, M. J.** (2012). Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S forms. *Molecular Biology and Evolution* 29, 279–291.
- Weetman, D., Steen, K., Rippon, E. J., Mawejje, H. D., Donnelly, M. J. and Wilding, C. S.** (2014). Contemporary gene flow between wild *An. gambiae* s.s. and *An. arabiensis*. *Parasites & Vectors* 7, 279–291.
- Whitworth, T. L., Dawson, R. D., Magalon, H. and Baudry, E.** (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society B – Biological Sciences* 274, 1731–1739.
- Yeap, H. L., Rasic, G., Endersby-Harshman, N. M., Lee, S. F., Arguni, E., Le Nguyen, H. and Hoffmann, A. A.** (2016). Mitochondrial DNA variants help monitor the dynamics of *Wolbachia* invasion into host populations. *Heredity* 116, 265–276.
- Yeates, D. K., Seago, A., Nelson, L., Cameron, S. L., Joseph, L. and Trueman, J. W. H.** (2011). Integrative taxonomy, or iterative taxonomy? *Systematic Entomology* 36, 209–217.
- Yeates, D. K., Zwick, A. and Mikheyev, A. S.** (2016). Museums are bio-banks: unlocking the genetic potential of the three billion specimens in the world's biological collections. *Current Opinion in Insect Science* 18, 83–88.
- Zinger, L. and Philippe, H.** (2016). Coalescing molecular evolution and DNA barcoding. *Molecular Ecology* 25, 1908–1910.