Review

Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts

Gregory D. D. Hurst1,* and Francis M. Jiggins2

1Department of Biology, University College London, 4 Stephenson Way, London NW1 2HE, UK
2Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, West Mains Rd, Edinburgh EH9 3JT, UK

Mitochondrial DNA (mtDNA) has been a marker of choice for reconstructing historical patterns of population demography, admixture, biogeography and speciation. However, it has recently been suggested that the pervasive nature of direct and indirect selection on this molecule renders any conclusion derived from it ambiguous. We review here the evidence for indirect selection on mtDNA in arthropods arising from linkage disequilibrium with maternally inherited symbionts. We note first that these symbionts are very common in arthropods and then review studies that reveal the extent to which they shape mtDNA evolution. mtDNA diversity patterns are compatible with neutral expectations for an uninfected population in only 2 of 19 cases. The remaining 17 studies revealed cases of symbiont-driven reduction in mtDNA diversity, symbiont-driven increases in diversity, symbiont-driven changes in mtDNA variation over space and symbiont-associated paraphyly of mtDNA. We therefore conclude that these elements often confound the inference of an organism’s evolutionary history from mtDNA data and that mtDNA on its own is an unsuitable marker for the study of recent historical events in arthropods. We also discuss the impact of these studies on the current programme of taxonomy based on DNA bar-coding.

Keywords: mtDNA; phylogeography; Wolbachia; symbiont; population genetics; selective sweep

1. INTRODUCTION

The study of evolution frequently requires understanding the history of the population, species or clade under study. In population genetics, a recent history of population bottlenecks may restrict genetic variation and thus constrain the speed of adaptation. In examining diversification over space, we need to have detailed knowledge of the different populations’ histories of colonization and the gene flow between them. In comparative analyses of processes of adaptation or molecular evolution, and in studies of historical biogeography, we require resolution of the relationships between species.

Ascertaining these patterns relies extensively on genetic markers. The most widely used of these in animals has been variation in the mitochondrial DNA (mtDNA) sequence. This choice was well reasoned. mtDNA can be easily amplified from a variety of taxa and, because it is haploid, the sequence can be obtained without cloning. Because it has a high evolutionary rate and an effective population size approximately one-quarter that of nuclear markers, it allows a chance of recovering the pattern and tempo of recent historical events without an extensive sequencing effort. Thirdly, as an area of at least low recombination, the whole molecule can be assumed to have the same genealogical history. In contrast, while stringent efforts have been made to develop and amplify nuclear markers, this often involves the refinement of primers for the target species, the sampling of several genes before one with an appropriate evolutionary rate is discovered and the separation of alleles from heterozygous individuals via cloning before sequencing. While these problems can be overcome outside ‘model’ taxa where genomic sequence data are available, they increase the effort needed to gain the required information.

Mitochondrial DNA has therefore remained the marker of choice in many population, biogeographic and phylogenetic studies. Its use has also been recommended in taxonomic studies, with the proposal that all described species are given an mtDNA sequence tag or bar-code (Hebert et al. 2003). Indeed, the use of mtDNA differentiation in defining taxonomic units has been suggested. While mtDNA is a very useful marker, its use is not without complication. It was recognized at the outset that mtDNA was strictly a marker for historical processes in females; should male and female history differ in a species, then this marker would not reflect the history of the species as a whole but that of the female portion. Further, there have been technical issues arising from the presence of nuclear integrations of mtDNA sequence. mtDNA integrated into the nuclear genome may still amplify with conserved primers aimed at mitochondrial copies, complicating or confounding analysis (Bensasson et al. 2001).

Further to these known problems, Ballard & Whitlock (2004) have recently argued that mtDNA evolution is non-neutral with sufficient regularity to question its utility as a marker for genomic history. Direct selection (selection
on mtDNA itself) and indirect selection (selection arising from disequilibrium with other maternally transmitted genes) is sufficiently common to make inferences from mtDNA data unreliable. In this review, we examine evidence for indirect selection on mtDNA in arthropods arising from disequilibrium with vertically transmitted micro-organisms, and assess the potential of this linkage to confound an interpretation of the history of a population or species.

We first note that inherited symbionts are very common in invertebrates and can have profound effects on their host. We then present theory and data that suggest that the disequilibrium of symbionts with mtDNA can confound analysis of population, biogeographic and clade history. For ease of reference, these studies are summarized in table 1. Finally, we note that widespread symbiont incidence and diversity, combined with rapid turnover, means that their presence cannot be simply tested and controlled against. We therefore conclude that mtDNA is inappropriate as a sole marker in studies of the recent history of arthropods and, potentially, other invertebrates.

2. THE INCIDENCE OF INHERITED SYMBIONTS IN ARTHROPODS

It has long been recognized that many arthropods carry passenger micro-organisms: microbes that exist inside the cells of their host and pass from a female to her progeny through the egg. These micro-organisms can broadly be classified into two kinds: beneficial to the host and parasitic. In the case of the former, treatment of the arthropod host with antibiotics produces a decrease in host fitness (and commonly infertility or death). In the case of the latter, the drive that produces the spread of the micro-organism is frequently the manipulation of host reproduction towards the survival of infected females and the daughters of infected females.

Beneficial symbionts are found widely in arthropods and other invertebrates. The majority of aphid species, for instance, depend on the presence of the bacterium Buchnera to be able to synthesize essential amino acids. Cockroaches, termites and a variety of Hemiptera (e.g. white fly), Diptera (e.g. tsetse fly), Hymenoptera (e.g. carpenter ants) and Coleoptera (e.g. weevils) likewise rely on beneficial symbionts. Feeding on depauperate diet (blood throughout life; phloem; nitrogen-poor wood) are strong predisposing factors for the occurrence of these beneficial microorganisms. The associations between beneficial symbionts and their host can be very long, as indicated by cocladogenesis of symbiont and host (e.g. Bandi et al. 1998). The observation that they have genomes that have shrunk massively over the course of symbiosis indicates that these beneficial symbionts undergo repeated selective sweeps associated with genomic rearrangements during the early parts of their association with the host (Wernegreen 2002).

Parasitic passenger micro-organisms are found even more widely. These micro-organisms show a variety of phenotypes associated with promoting the production and survival of infected daughters (which can transmit the maternally transmitted micro-organism) via negative effects on the production and survival of infected males and uninfected females (which cannot). In the simplest case, this involves creating a female bias in the host sex ratio. This is represented by cases of parthenogenesis induction in haplodiploid species and feminization in Crustacea. In a twist to this manipulation, killing of male hosts during embryogenesis is common in insects. Here, the advantage derives from removing any negative effects that male hosts may have on their sisters. Examples are found in Hemiptera, Hymenoptera, Diptera, Coleoptera and Lepidoptera. Even in Drosophila, which for ecological reasons are not the most likely to bear male-killers, there are 14 records of male-killing bacteria, indicating these parasites are widely present (Hurst & Majerus 1993). The incidence is certainly higher in other species, such as the coccinellid beetles.

The most commonly observed form of parasitism is cytoplasmic incompatibility, which differs subtly in logic from sex ratio distortion. In this manipulation, zygotes formed following fertilization of an uninfected egg with sperm from an infected male die during early embryogenesis. This behaviour produces the spread of the infection in structured populations, as the bacterial phenotype is essentially one of selectively killing uninfected individuals. Cytoplasmic incompatibility has been described in insects, mites and crustaceans, and is probably very common (although cryptic, as frequently all individuals within a population are infected; Stouthamer et al. 1999). Two micro-organisms, Wolbachia and Cardinium, are known to induce it (Breeuwer et al. 1992; Hunter et al. 2003).

In total, these parasitic interactions are common. In large surveys, repeatable over geographical regions and across arthropod groups, Wolbachia alone infects in excess of 20% of insect species at any point in time (Werren et al. 1995; Werren & Windsor 2000; Jiggins et al. 2001) and, in a limited survey, just over 50% of spiders (Rowley et al. 2004). The bacterium Cardinium infects around 7% of arthropods (Weeks et al. 2003).

The interactions between parasitic symbionts and their hosts are relatively short-lived in comparison to beneficial symbioses. While cocladogenesis of host and Wolbachia is sometimes observed (e.g. Marshall 2004), studies on focused host clades have indicated it is rather rare (Shoemaker et al. 2002). The mean lifespan of any particular interaction is therefore generally less than the mean time to speciation. This conclusion is reinforced by the observation of three actively spreading Wolbachia infections in natural populations (Turelli & Hoffmann 1991; Hoshizaki & Shimada 1995; Rieger & Stauffer 2002). Thus, we can conclude that the 20% incidence reflects a pattern where new infections must arise relatively commonly. As a final complicating feature, it should be noted that a single population may be infected with more than one strain or species of parasitic inherited micro-organism and that different populations may show different infection statuses.

Finally, recent study has revealed the widespread presence of 'secondary symbionts'. These are vertically transmitted micro-organisms that are not essential but appear to be locally beneficial (e.g. enhancing resistance to parasitoids and pathogens or adaptation to growth on a particular species of host plant; Oliver et al. 2003; Ferrari et al. 2004; Tsuchida et al. 2004). These can be common and it is known that their frequency varies geographically (Tsuchida et al. 2002). What is uncertain is the extent of
Table 1. Known effects of symbionts on mtDNA variation.

<table>
<thead>
<tr>
<th>host</th>
<th>symbiont</th>
<th>observation</th>
<th>references</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Porcellionides pruinosus</em> (Crustacea: Isopoda)</td>
<td><em>Wolbachia</em> (probably feminizing)</td>
<td>infected and uninfected woodlice from southern populations have divergent mtDNA sequences</td>
<td>Marcade et al. (1999)</td>
</tr>
<tr>
<td><em>Armadillidium vulgare</em> (Crustacea: Isopoda)</td>
<td>feminizing <em>Wolbachia</em></td>
<td><em>Wolbachia</em> associated with a subset of mitotypes in the population</td>
<td>Grandjean et al. (1993), Rigaud et al. (1999)</td>
</tr>
<tr>
<td><em>Gammarus duebeni</em> (Crustacea: Amphipoda)</td>
<td>feminizing microsporidia</td>
<td>association between infecting feminizer species and mtDNA haplotypes</td>
<td>Ironside et al. (2003)</td>
</tr>
<tr>
<td><em>Adalia bipunctata</em> (Coleoptera)</td>
<td>male-killing <em>Spiroplasma, Rickettsia</em> and <em>Wolbachia</em></td>
<td>frequency of mtDNA haplotypes associated with the bacteria in the population, not geography</td>
<td>Schulerburg et al. (2002)</td>
</tr>
<tr>
<td><em>Chelymorpha alternans</em> (Coleoptera)</td>
<td><em>Wolbachia</em></td>
<td>mtDNA diversity not different from neutrality</td>
<td>Keller et al. (2004)</td>
</tr>
<tr>
<td><em>Acraea encedana</em> (Lepidoptera)</td>
<td>male-killing <em>Wolbachia</em></td>
<td>a selective sweep has significantly reduced mtDNA diversity in infected and uninfected butterflies</td>
<td>Jiggins (2003)</td>
</tr>
<tr>
<td><em>Acraea encedon</em> (Lepidoptera)</td>
<td>male-killing <em>Wolbachia</em></td>
<td>reduced mtDNA diversity in infected but not uninfected butterflies. Two strains of <em>Wolbachia</em> associated with different mtDNA haplotypes. Geographical structure in uninfected, but not infected, mtDNA. mtDNA has introgressed with <em>Wolbachia</em> from <em>A. encedana</em></td>
<td>Jiggins (2003)</td>
</tr>
<tr>
<td><em>Ostrinia furnacalis</em> (Lepidoptera)</td>
<td>male-killing <em>Wolbachia</em></td>
<td>mtDNA sequences in infected and uninfected moths are paraphyletic</td>
<td>F. Jiggins, unpublished data</td>
</tr>
<tr>
<td><em>Orseolia oryzae</em> (Diptera)</td>
<td><em>Wolbachia</em> (effect on host unknown)</td>
<td>three highly diverged mitotypes. One is maternally inherited and associated with <em>Wolbachia</em>. The other two are paternally inherited and found in uninfected males</td>
<td>Behura et al. (2001)</td>
</tr>
<tr>
<td><em>Drosophila recens</em> (Diptera)</td>
<td>cytoplasmic incompatibility <em>Wolbachia</em></td>
<td>lower mtDNA diversity but similar nuclear DNA diversity compared with related uninfected species. Elevated rate of substitution</td>
<td>Shoemaker et al. (1999, 2004)</td>
</tr>
<tr>
<td><em>Drosophila innubila</em> (Diptera)</td>
<td>male-killing <em>Wolbachia</em></td>
<td>lower mtDNA diversity compared with nuclear DNA associated with an ancient infection. mtDNA diversity is as expected from neutral models with reduced Ne of mtDNA associated with infection</td>
<td>Dyer &amp; Jaenike (2004)</td>
</tr>
<tr>
<td><em>Drosophila santomea</em>, <em>D. yakuba</em> and <em>D. teissieri</em> (Diptera)</td>
<td><em>Wolbachia</em> (effect on host unknown)</td>
<td>these species have very similar mtDNA and <em>Wolbachia</em>, but have diverged in nuclear DNA. Probably introgression of mtDNA and <em>Wolbachia</em> between species</td>
<td>Monnerot et al. (1990), Lachaise et al. (2000)</td>
</tr>
<tr>
<td><em>Drosophila mauritiana</em> (Diptera)</td>
<td><em>Wolbachia</em> (effect on host unknown)</td>
<td>two paraphyletic mitotypes. One has introgressed together with <em>Wolbachia</em> infection from <em>D. simulans</em></td>
<td>Rousett &amp; Solignac (1995), Ballard (2000b)</td>
</tr>
<tr>
<td><em>Drosophila simulans</em> (Diptera)</td>
<td>cytoplasmic incompatibility <em>Wolbachia</em></td>
<td>selective sweep has significantly reduced mtDNA nucleotide diversity in infected populations versus uninfected. Geographical structure determined by symbiont strain present, not population history</td>
<td>Ballard &amp; Kreitman (1994), Ballard et al. (1996), Ballard (2000a), Dean et al. (2003)</td>
</tr>
</tbody>
</table>

(Continued.)
The effects of inherited symbionts

3. PASSENGER MICRO-ORGANISMS AND THE POPULATION GENETICS OF mtDNA

Inherited micro-organisms will influence the population genetics of the host’s mtDNA if they are cotransmitted and therefore in linkage disequilibrium. However, linkage disequilibrium will break down if either the symbiont or mitochondria are paternally or horizontally (infectiously) transmitted with sufficient frequency (Turelli et al. 1992). This process is equivalent to recombination breaking down the linkage between nuclear genes. Horizontal or paternal transmission of both symbionts and mtDNA has been documented in insects. Despite this, symbionts in natural populations are typically found in linkage disequilibrium with mtDNA (table 1), suggesting that such transmission is so infrequent as to be unimportant. For example, paternal transmission of both Wolbachia and mtDNA have been recorded in laboratory populations of Drosophila simulans but it is sufficiently rare that they remain in linkage disequilibrium in the field (Hoffmann & Turelli 1988; Kondo et al. 1990; Turelli et al. 1992). However, there are some species where infectious transmission of symbionts is so common that there is unlikely to be any association between the microbe and host mtDNA (Huigens et al. 2000).

Provided the assumption of linkage disequilibrium is met, if a population becomes infected with a symbiont that has sufficient drive to spread, the mtDNA type associated with the initial infection will hitchhike through the population (‘indirect selection’ on the mtDNA). This process has been recreated in laboratory populations of the mosquito Aedes albopictus infected with a Wolbachia strain that induces cytoplasmic incompatibility (Kambhampati et al. 1992). The most remarkable example, however, comes from Californian populations of D. simulans. These were originally uninfected, but during the 1980s they were invaded by a Wolbachia strain that induced cytoplasmic incompatibility (Turelli & Hoffmann 1991). This infection rapidly spread to a high prevalence and carried with it a mtDNA haplotype that was previously rare or absent in uninfected Californian populations (Turelli et al. 1992). Once the infection neared fixation, the original mtDNA haplotype was completely replaced by the haplotype linked to the symbiont. The sweeps of mtDNA seen in California D. simulans are probably regular events in insects. Despite being transient events, the spread of a new Wolbachia strain over space has also been documented in the delphacid bug Laodelphax striatellus and the fly Rhagoletis cerasi (Hoshizaki & Shimada 1995; Riegler & Stauffer 2002). These observations, together with the lack of cocladogenesis between hosts and symbionts discussed above, indicate that new interactions (which must involve symbiont spread) occur during the lifespan of many species.

It is clear from these studies that mtDNA within a population is affected by symbiont spread. Further, the ability of symbionts to spread between populations by occasional movement of hosts, and the ability of different host populations to maintain different symbiont strains, can also confound interpretation attempts to reconstruct the phylogeography of a species. The drive associated with

disequilibrium with mtDNA, which will be inversely related to the rate of these symbionts’ horizontal transmission in the field.
symbiont manipulation can even result in the spread of the symbiont into a new species following an occasional hybridization event. This will homogenize the mtDNA of different species. We now review the effects that symbionts may have on patterns within populations, between populations and between species variation in mtDNA.

4. THE EFFECT OF SYMBIONTS ON mtDNA DIVERSITY WITHIN POPULATIONS

If a population is infected by one or more symbionts, then patterns of mitochondrial polymorphism will be altered by natural selection acting on those symbionts. Depending on the recency of invasion and the number of symbionts present, they may either reduce or increase diversity. There can also be an alteration in the frequency distribution of haplotypes within the population (table 1). Unfortunately, it is these parameters that are used to infer the historical demography of populations and it is difficult to distinguish demographic from symbiont-induced effects.

The initial selective sweep that occurs as the symbiont invades, and subsequent sweeps of advantageous symbiont mutations, will reduce mtDNA diversity and skew the frequency distribution of alleles towards rare variants (Maynard-Smith & Haigh 1974; Tajima 1989b). These are similar patterns to those produced by population bottlenecks and expansions, and a selective sweep could thus easily be mistaken for these demographic processes (Tajima 1989a). Therefore, low mtDNA diversity alone should not be taken as evidence for a bottleneck or founder event, although this may sometimes be the case (see Rokas et al. (2001) for a genuine example). Instead, it is likely that many cases of low mtDNA variation are not caused by demographic events but selective sweeps of symbionts running through the population.

This is supported by the observation that parasitic symbionts are commonly associated with low mitochondrial diversity (seven cases in table 1). In at least two of these species, it has been shown that this has been caused by a selective sweep rather than a population bottleneck (Ballard & Kreitman 1994; Ballard 2000a; Jiggins 2003). These studies used an HKA test (Hudson et al. 1987) to show that the genetic diversity of mitochondrial genes was significantly less than their nuclear counterparts. This suggests that the low mtDNA diversity is caused by the symbiont, as demographic processes would have reduced the diversity of the entire genome.

These mitochondrial selective sweeps could result from selection on any maternally transmitted element, including the mitochondria themselves, symbionts or, in female heterogametic hosts, the W sex chromosome. Confirmation that selection is acting on the symbiont has come from comparisons of infected and uninfected hosts from the same population, from different populations or from different species. For example, uninfected Acraea encedona have more diverse mitochondria than Wolbachia-infected individuals from the same population (Jiggins 2003). In a study that compared different populations of D. simulans, an uninfected population in East Africa was found to have greater mtDNA diversity than infected populations from elsewhere (Dean et al. 2003). Finally, Wolbachia-infected species have been found to harbour lower levels of mtDNA diversity than closely related uninfected species (Shoemaker et al. 1999, 2004).

The effect of a selective sweep on mtDNA diversity will in part depend on the time that has elapsed since it occurred, with diversity gradually increasing after a selective sweep. Therefore, these effects may be most important for parasitic symbionts, as these tend to have short-lived associations with their hosts and frequently transmit to new host species (Werren et al. 1995).

If a symbiont is imperfectly transmitted between generations, then any uninfected offspring produced by infected females will carry the mitochondrial haplotype associated with the infection. Because this process is unidirectional, the original mtDNA lineages in the uninfected hosts will ultimately be lost and replaced by the mtDNA type associated with the symbiont (Turelli et al. 1992; Johnstone & Hurst 1996). Therefore, the selective sweep may affect uninfected, as well as infected, hosts. However, whether this occurs will depend on the time that has elapsed since the symbiont invaded and the transmission rate of the symbiont from mother to offspring (Turelli et al. 1992; Johnstone & Hurst 1996). This is illustrated by two closely related species of butterfly, Acraea encedana and A. encedon, which are infected by male-killing Wolbachia. The infection in the former species is older and has a lower transmission rate than the latter species. As expected, the selective sweep has eliminated mtDNA variation from both infected and uninfected individuals of A. encedana but has only affected infected females of A. encedon (Jiggins 2003).

So far we have only considered the simplest scenario in which a single symbiont invades an uninfected population. It is also common to find multiple strains of cytoplasmic incompatibility inducing Wolbachia co-infecting the same individuals. In this situation, the effects on mtDNA will be qualitatively similar to a single infection. However, some symbionts are polymorphic, with different infections occurring in different individuals but never co-infecting the same host. In the five cases where this has been examined, the different symbionts are associated with different mtDNA sequences (James & Ballard 2000; Schulenburg et al. 2002; Baudry et al. 2003; Jiggins 2003). Therefore, although a selective sweep may reduce the diversity of mtDNA associated with any one symbiont, high levels of diversity may be maintained within the population as a whole across different infection classes and the diversity of a population will depend on the number of symbionts it harbours. This can result in a mitochondrial genealogy with deep internal branches and short terminal branches, which could easily be mistaken as evidence for population structure and admixture (F. Jiggins, unpublished results).

Once a symbiont has invaded and reached equilibrium, the associated mtDNA will gradually accumulate mutations, and patterns of polymorphism may eventually resemble those expected under neutrality (e.g. Keller et al. 2004; Marshall 2004). However, it is notable that there are only two case studies from 19 in which mtDNA diversity in symbiont-infected species was compatible with that expected in the comparable uninfected species. There are several reasons why diversity may not return to ’normal’. The first is that mutations may occur that increase the symbiont’s fitness and are fixed by positive selection, resulting in recurrent selective sweeps through the population. Evidence for this process comes from a strong positive selection found to act on certain symbiont genes.
(Jiggins et al. 2002), which contrasts with mitochondrial genes that tend to be under a purifying selection. This process may be particularly important in parasitic interactions, where host–parasite coevolution can result in strong directional selection (Jiggins et al. 2002). However, if an advantageous mutation occurs in a beneficial symbiont, it may also cause a selective sweep of mtDNA. That this may be a regular occurrence is suggested by the observation that genome reduction of beneficial symbionts is common during the early stages of symbiosis and must be associated with selective sweeps within the population.

The second reason that diversity may fail to return to pre-infection levels is that the effective population size of mtDNA is lower after infection. There are two causes of this. First, in cases where there is inefficient transmission, only mutations in infected individuals can spread. Thus, the effective population size diminishes to that of the number of infected females (Johnstone & Hurst 1996). In cases of low prevalence infection, as found for many male-killers, significant reduction in diversity at equilibrium is therefore expected. The reduced variation in mtDNA relative to nuclear DNA in D. immigrata, compared with the same metric in D. fallax, an uninfected sibling species, can be explained by a model based on the reduced effective population size of mtDNA associated with a prevalence of infection in this species of 35% (Dyer & Jaenike 2004).

The second cause of diminished effective population size is greater efficacy of background selection in the presence of a symbiont. Background selection can be understood as a reduction in the mitochondrial effective population size to the proportion of cytoplasms free from deleterious mutations (Charlesworth et al. 1993). Its impact on mtDNA diversity will be greatest when the deleterious mutation rate is highest. Considering the relative mutation rates and genome sizes of symbionts and mtDNA (Tamura et al. 1992; Ochman et al. 1999; Sun et al. 2001; Wernegreen 2002), the total cytoplasmic mutation rate will be increased roughly tenfold in symbiont-infected hosts. However, it is unclear whether mitochondrial and symbiont mutations will have similar effects on fitness, nor whether mtDNA lies in a region of parameter space where background selection is important.

5. PHYLOGEOGRAPHIC IMPACTS: BETWEEN POPULATION EFFECTS OF PASSENGER MICRO-ORGANISMS

There are two potential effects of passenger microorganisms on mtDNA structure between populations. The first is that the migration of an infected individual from one population into an uninfected one can produce a sweep of the infection and associated mtDNA haplotype through the population. This will homogenize the haplotypes of the two populations. This, of course, happens without homogenization of nuclear genes, but if only the mtDNA is studied the populations appear to be fully mictic and unstructured. As discussed above, selective sweeps of symbionts are probably common events.

The second effect is that mtDNA differentiation between populations may be increased due to heterogeneity of the infections between populations. This could occur if natural selection maintains different symbionts in different populations despite migration between them. Alternatively, if a selective sweep occurs in one population but not another, then this could also inflate measures of population differentiation.

Drosophila simulans provides examples of both of these phenomena. As discussed previously, the spread of Wolbachia strain wRi through California was accompanied by a sweep of the mtDNA, homogenizing the haplotype of the Californian population to that of the strain from which the invading fly derived. Geographical variation in the infecting strain present is also known to maintain mtDNA differentiation between populations of D. simulans, with each infection producing a different ‘compatibility type’ (they do not rescue each other’s zygotic death phenotype). Different strains of infection have different associated mtDNA haplotypes (and populations of different infection status differ in mtDNA haplotypes) and appear isolated (James & Ballard 2000). However, examination of nuclear markers indicates they are not isolated with respect to nuclear genes; these still flow between populations (Ballard et al. 2002). This situation is also observed for the blowfly Protocalliphora in the USA, where different populations bear different strains of Wolbachia and different mtDNA types, but these patterns do not reflect the population history inferred from nuclear markers (Baudry et al. 2003).

The geographical structuring of mtDNA associated with different infections can also be seen for different infections with male-killing bacteria. The two-spot ladybird, Adalia bipunctata, bears three different male-killing infections within Europe (Hurst et al. 1999a,b). The frequency of these infections varies geographically, with one of the infections found all over Europe and the other two with more restrictive incidence. Analysis of mtDNA diversity across Europe indicates that diversity can be accounted for by variation in infection status. Geography did not explain any of the variation when infection status was controlled (Schulenburg et al. 2002). Structuring associated with different symbiont strains over space is also seen in Acraea encedon (Jiggins 2003).

6. EFFECTS ON THE mtDNA PHYLOGENY OF RECENTLY DIVERGED SPECIES

At first sight, passenger micro-organisms are expected to affect the dynamics of mtDNA within populations but not the branching pattern of mtDNA on a phylogeny. However, two case studies indicate this is not necessarily the case. First, there is the A. encedon and A. encedana species pair. These both bear a male-killing Wolbachia infection, and evidence from the Wolbachia sequence indicates they are very closely related strains (note that A. encedon also has a second, more distantly related, infection). When the phylogeny of individuals of the different species is constructed based on mtDNA, A. encedon and A. encedana individuals that bear the same Wolbachia infection bear identical mtDNA sequences, distinct from that found in uninfected A. encedon individuals. While the species clearly are distinct on morphological grounds, on the grounds of nuclear DNA sequences and in terms of genetic isolation, they appear identical on the basis of the mtDNA sequences of the infected individuals.
The effects of inherited symbionts

G. D. D. Hurst & F. M. Jiggins

The effects of inherited symbionts

G. D. D. Hurst & F. M. Jiggins

The most likely explanation for this is that rare hybridization events, although producing very little in the way of flow of nuclear genes, can produce the transfer of the male-killer and associated mitotype from species to species. This male-killer has a drive mechanism that results in its increasing in frequency, despite initially being in poorly adapted hybrid individuals, and the infection and associated mitotype spreads into the new species.

This transfer and fixation of mtDNA following hybridization would appear at first sight to be a rare event. However, it has also been observed in Drosophila. In Drosophila, mtDNA has introgressed between D. simulans and D. mauritiana, associated with Wolbachia infection (Rousset & Solignac 1995; Ballard 2000). Anecdotally, three out of four cases we have studied in our laboratories have shown evidence of symbiont-induced mtDNA introgression confounding phylogenetic estimation from an mtDNA sequence.

It is possible that symbiont-driven introgression may explain some recent case studies where mtDNA phylogenies conflict with those obtained from nuclear DNA. Shaw (2002), for instance, observes that while certain crickets of the genus Laupala in Hawaii do not differ in mtDNA sequence, they do exhibit substantial differentiation at nuclear loci. This incongruence points to selection causing introgression of the mtDNA following hybridization. A good hypothesis for this observation is that hybridization carries novel symbiont infections. The spread of the introgressed symbiont would then be associated with the spread of introgressed mtDNA, homogenizing mtDNA variation across the species boundary despite a high genetic integrity of the species as recorded on nuclear markers.

7. CONCLUSIONS

Indirect selection on mtDNA arising from disequilibrium with inherited micro-organisms is very common in arthropods and is probably widespread in many other invertebrates. While we have documented cases associated with parasitic-inherited micro-organisms, it is known that required beneficial micro-organisms are also in linkage disequilibrium with mtDNA (Funk et al. 2000; Hurtado et al. 2003). This will also produce selective sweeps as advantageous symbiont mutations spread through the population, particularly in recently evolved interactions. It is also possible that disequilibrium with secondary symbionts (beneficial symbionts that are not required: see §2) will likewise produce sweeps and additionally produce geographical structures in host mtDNA.

These associations will all confound simple interpretations of genomic history from mtDNA data. A brief description of classical explanations for patterns of mtDNA diversity, along with potential explanations for the pattern based on indirect selection associated with symbiosis, is given in table 2. It is our conclusion that the interactions occur commonly enough that the presence of these symbionts makes it both unsafe and unsatisfactory to infer patterns of genome history based on mtDNA sequence data. As a preliminary judgment, therefore, we would regard the development and use of microsatellites for intraspecific study, and nuclear coding genes for phylogenetic study, a requirement to reveal the history of nuclear DNA. This judgment is preliminary because a full test of this issue requires testing the congruence of mtDNA and nuclear DNA patterns of variability across a range of clades, where symbiont presence has not been ascertained and not (as in our reviewed studies) been used as a proband to seek the effects. We do, however, regard the above studies combined with the known high incidence of Wolbachia as an evidential ‘smoking gun’ that urges precaution in the first instant, pending formal testing.

One recent and contentious use of mtDNA sequence is in DNA bar-coding (Hebert et al. 2003; 2004a). In DNA bar-coding, the sequence of the mitochondrial cytochrome oxidase subunit 1 (COI) gene is used for the purpose of taxonomic identification and assessment of biodiversity, with the philosophy that for each species there is one bar-code (and reciprocally, one bar-code indicates a given species). DNA bar-coding relies on there being low levels of mtDNA variation within a species compared with differentiation between species and monophyly of mtDNA within species. While there may generally be low mtDNA diversity within a species, and species may frequently be delineable by mtDNA, the case studies above make it clear that symbionts can disrupt this pattern. They can homogenize biological species for mtDNA following introgression of symbionts, as in Acrea and Drosophila (cases of one bar-code, two species). They may also make one species
appear as two on the basis of high levels of intraspecific variation in a mtDNA sequence associated with possession of different symbiont strains, as in Adalia (cases of two bar-codes, one species).

While it is sure from the above and other studies that bar-coding will create some mistakes, what is unsure is the frequency of mistakes and whether this frequency exceeds tolerance limits. While a review of species polyphly on the basis of mtDNA suggested 23% of species may not be monophyletic for mtDNA sequences (Funk & Omland 2003), bar-coding tests have not revealed this pattern (Hebert et al. 2003; 2004a). However, with notable exceptions (Hebert et al., 2004b), past tests have tended not to explicitly test the ability to discriminate a range of closely related sibling species but rather a range of congenerics, many of which are relatively distantly related. Further, the sample size used when testing intraspecific variation has often been limited, with one or two extra individuals within known species obtained and found to possess very similar COI sequences to those previously found. Indeed, where they have been found to carry divergent sequences, the inference has been made that the process has revealed cryptic species, rather than that mtDNA can act as a poor marker.

Thus, while bar-coding clearly has utility for placing unknown specimens in the genera in which they belong, its utility at the species level is much less certain. The potential limitations of DNA bar-coding are well illustrated by the insect group for which we have the most detailed understanding of both taxonomy and genetic diversity, the melanogaster subgroup of Drosophila. Of the nine species in this subgroup, only three have unique barcodes. It is likely that five of the six cases where bar-coding fails are due to the presence of Wolbachia (table 1). Again, this evidence requires full evaluation across an array of clades where infection status is not a proband for the study.

Beyond the specific question of bar-coding, can we rescue the use of mtDNA as a marker? In some studies, authors have tested for and excluded the presence of Wolbachia in a particular species and thus inferred low mtDNA diversity to indicate evidence of a bottleneck (e.g. Johnson et al. 2004). However, this approach is not completely safe. Notwithstanding the possibility of direct selection on mtDNA, Wolbachia is one of many symbionts. Cardinium, for instance, infects approximately 7% of arthropods (Weeks et al. 2003). Aside from these, bacteria from many different major divisions can be found in arthropods (e.g. Hurst et al. 1999a), as can vertically transmitted microsporidia (Terry et al. 2004). Many vertically transmitted bacteria have closely related free-living forms and many remain to be discovered. Thus, polymerase chain reaction assays for the presence or absence of all symbionts are simply not possible and only a microscopy-based screen on the dissected ovaries of 100 individuals could really provide evidence that symbionts were absent. In addition to this, it is possible to find a false negative for symbiont presence if sampling is not sufficiently intensive, because symbionts are sometimes only present in a minority of individuals unless sampling is intensive (e.g. Jiggins et al. 2001). Finally, current absence (if it could be proven) is no assurance of past absence. The lack of co cladogenesis of Wolbachia and their hosts indicates that interactions are relatively short-lived in evolutionary time (Shoemaker et al. 2002). Low mtDNA diversity could be caused by a symbiont that has subsequently died out within the species.

A more promising approach is to test mtDNA datasets for the signature of natural selection that will be left by the spread of symbionts. Unfortunately, this is not simple because the demographic processes of interest will often produce similar patterns to selection acting on symbionts (Tajima 1989a). This problem can only be reliably overcome by corroborating the inferences made using mtDNA with data from nuclear genes (Rokas et al. 2001).

We therefore conclude that mtDNA alone cannot be used to reliably infer population history or the history of closely related species in arthropods as there is a very high probability of an incorrect conclusion due to indirect selection arising from the presence of a symbiont. Reciprocally, it is also problematic to infer symbiont history from mtDNA alone. We argue that mtDNA should never be used as a sole marker in studies of either genomic or symbiont history. While our arguments above derive from studies of arthropods, we would caution that these arguments may apply in a wider range of species. There are several records of inherited bacteria being present in nematodes (Adams & Eichenmuller 1963; Marti et al. 1995; Sironi et al. 1995) and also records of inherited bacteria in disequilibrium with mtDNA in molluscs (Hurtado et al. 2003). It should not be assumed that taxa lack symbionts until careful surveys have been carried out because recent history indicates that even well studied groups (e.g. filarial nematodes) can in fact be covertly infected with inherited symbionts (Sironi et al. 1995).

We wish to thank Jim Mallet for encouragement to write this piece and for comments. GH wishes to acknowledge support from the BBSRC and NERC. FJ wishes to acknowledge support from the Wellcome Trust.

REFERENCES


