DNA barcoding cannot reliably identify species of the blowfly genus Protocalliphora (Diptera: Calliphoridae)

T. L. Whitworth¹, R. D. Dawson², H. Magalon³ and E. Baudry⁴,*

¹Washington State University, 2533 Inter Avenue, Puyallup, WA 98372, USA
²University of Northern British Columbia, Prince George, British Columbia V2N 4Z9, Canada
³Laboratoire d’Ecologie, Université Paris VI, Paris 75252, France
⁴Laboratoire Ecologie, Systematique et Evolution, Université Paris-Sud, Bâtiment 362, 91405 Orsay Cedex, France

In DNA barcoding, a short standardized DNA sequence is used to assign unknown individuals to species and aid in the discovery of new species. A fragment of the mitochondrial gene cytochrome c oxidase subunit 1 is emerging as the standard barcode region for animals. However, patterns of mitochondrial variability can be confounded by the spread of maternally transmitted bacteria that cosegregate with mitochondria. Here, we investigated the performance of barcoding in a sample comprising 12 species of the blow fly genus Protocalliphora, known to be infected with the endosymbiotic bacteria Wolbachia. We found that the barcoding approach showed limited success: assignment of unknown individuals to species is possible for 60% of the species, while using the technique to identify new species would underestimate the species number in the genus by 75%. This low success of the barcoding approach is due to the non-monophyly of many of the species at the mitochondrial level. We even observed individuals from four different species with identical barcodes, which is, to our knowledge, the most extensive case of mtDNA haplotype sharing yet described. The pattern of Wolbachia infection strongly suggests that the lack of within-species monophyly results from introgressive hybridization associated with Wolbachia infection. Given that Wolbachia is known to infect between 15 and 75% of insect species, we conclude that identification at the species level based on mitochondrial sequence might not be possible for many insects. However, given that Wolbachia-associated mtDNA introgression is probably limited to very closely related species, identification at the genus level should remain possible.

Keywords: barcoding; mitochondrial introgression; insects; Wolbachia; Protocalliphora

1. INTRODUCTION

Even conservative estimates suggest that the majority of the species living on the planet are currently undescribed (e.g. Novotny et al. 2002). To achieve rapid species descriptions in the context of the current biodiversity crisis, and given the decline in the number of taxonomists, several authors have suggested the use of barcoding in taxonomy (Hebert et al. 2003a,b; Blaxter 2004; Schindel & Miller 2005). DNA barcoding is the use of a short standardized DNA sequence (in animals, a 600 bp fragment of the mitochondrial cytochrome c oxidase (COI) gene) to identify species. DNA barcoding regroups two different and relatively independent aspects: it can be used to (i) identify and assign unknown specimens to species that have previously been described and (ii) facilitate the discovery of new species.

Using a mitochondrial fragment as opposed to a nuclear one for DNA barcoding has two major advantages (Hurst & Jiggins 2005). First, because it is haploid and has highly conserved regions, the COI fragment is technically easy to amplify without cloning in a variety of species. Second, the mitochondrion has an effective population size approximately one-quarter of that of nuclear markers, and, in animals, a high evolutionary rate which therefore provides a high level of resolution. Even closely related species can usually be differentiated by using a relatively short sequence. However, these advantages are associated with a major drawback. While mitochondrial DNA was considered to be a neutral marker that reflects the history of the species, Ballard & Whitlock (2004) and Bazzin et al. (2006) have recently argued that mitochondria are in fact often under strong selection and evolve under unusual evolutionary rules when compared with other genomes. Selection can act directly on the mtDNA itself, but it can also arise indirectly from disequilibrium with other maternally transmitted DNA (Hurst & Jiggins 2005).

In insects, the endosymbiotic bacteria Wolbachia are an example of such maternally transmitted DNA. These bacteria cause a number of reproductive alterations in their hosts, including induction of thelytokous parthenogenesis, feminization of genetic males, male killing and, most commonly, the induction of sperm–egg incompatibilities termed cytoplasmic incompatibility (reviewed in Werren 1997; Stouthamer et al. 1999). These reproductive phenotypes effectively increase the frequency of infected females in the host populations, often at the expense of host fitness. Thus, when a population becomes infected with Wolbachia, the bacteria will rapidly spread and the mtDNA type associated with the initial infection will hitch-hike through the population by indirect selection. Given that between 15 and 75% of insect species harbour Wolbachia (Werren et al. 1995a; West et al. 1998;...
Japanese butterflies (Narita of *Eurema hecabe* Whitworth 1991), and between the yellow and the brown type *Protocalliphora* the rule in closely related species have been investigated, it is currently marshes (Sabrosky forests to the river valleys and from the deserts to the virtually any habitat where nidicolous birds nest from the *Protocalliphora* are known to be infected by *Wolbachia* although it is not yet known whether the bacteria induce *Protocalliphora* are widespread, occupying at the intraspecific level. At the intraspecific level, the selection on mtDNA in insects. 

**Wolbachia** can potentially influence mtDNA variation at the intra- or interspecific level. At the intraspecific level, the influence of *Wolbachia* is now well documented: numerous studies have demonstrated that selection acting on *Wolbachia* has indirectly reduced mtDNA polymorphism in the infected population or species (e.g. Shoemaker et al. 1999, 2004; Ballard 2000a,b; Jiggins 2003; see review in Hurst & Jiggins 2005). While this means that mtDNA should not be used to make inferences about population histories in *Wolbachia*-infected species, this should not cause a problem for barcoding. However, *Wolbachia* can also affect mtDNA variation between species boundaries. In insects, at least three cases are currently described where *Wolbachia* infection has caused mitochondrial introgression between closely related species: between several members of the *melanogaster* subgroup of *Drosophila* (Rousset & Solignac 1995; Ballard 2000a,b), between two sister species of sub-Saharan butterflies *Acraea encedon* and *Acraea encedana* (Jiggins 2003), and between the yellow and the brown type of *Eurema hecabe*, two yet unnamed sibling species of Japanese butterflies (Narita et al. 2006). In such situations, barcoding is impossible because different species share an identical barcode. However, due to the relative paucity of studies where mtDNA variation and *Wolbachia* infection of closely related species have been investigated, it is currently not known whether such introgressions are the exception or the rule in *Wolbachia*-infected species.

In this study, we focus on blow flies belonging to the genus *Protocalliphora*. *Protocalliphora* are the Holarctic group of species found commonly in the boreal forest and other areas. *Protocalliphora* are widespread, occupying virtually any habitat where nidicolous birds nest from the forests to the river valleys and from the deserts to the marshes (Sabrosky et al. 1989). Larval stages are obligate haematophagous parasites of nidicolous birds (Bennett & Whitworth 1991). *Protocalliphora* is the largest genus of blow flies in the Holarctic region with over 40 species described (Sabrosky et al. 1989). At least two species of *Protocalliphora* are known to be infected by *Wolbachia*, although it is not yet known whether the bacteria induce cytoplasmic incompatibility or other phenotypes in *Protocalliphora* (Baudry et al. 2003). To determine whether *Wolbachia* is involved in mtDNA introgressions, and hence the possibility of barcoding in this genus, we studied *Wolbachia* infection status and the nuclear and mtDNA polymorphism of 12 species of *Protocalliphora*.

**2. MATERIAL AND METHODS**

(a) Sampling and DNA extraction

Thirty-one *Protocalliphora* individuals, belonging to 12 species, were included in this study (table 1). There are at least 16 other *Protocalliphora* species in the Nearctic, but they were not included in this study because they are rare and very difficult to sample. When available, three individuals per species were analysed. The most closely related genus to *Protocalliphora* is the monospecific genus *Trypocalliphora* (Sabrosky et al. 1989); we therefore used three *Trypocalliphora braueri* individuals as outgroups. However, the taxonomic status of *Trypocalliphora* has been debated. Rognes (1984) considered it a valid genus, while Sabrosky et al. (1989) considered it a subspecies of *Protocalliphora*. Whitworth (2003b) evaluated each argument and concluded that Sabrosky’s conclusions were based on a misinterpretation of larval morphology and behaviour. Thus, he supported Rognes view that *Trypocalliphora* is a valid genus. Given the uncertainty about the status of *Trypocalliphora*, we also used one *Lucilia sericata* and one *Phormia regina* as outgroups.

Blow fly larvae or pupae were collected from bird nests several days after fledging of the young birds. Collections were made either directly by the authors or by naturalists, in the continental USA and Canada, except for the *Protocalliphora falcozi* individuals, which were collected from France. Emergent flies, when possible with their puparia, were placed into 95% ethanol. Species were then identified based on fly and pupal case morphology (Sabrosky et al. 1989; Whitworth 2002, 2003a,b). To minimize screening of siblings, only one individual per bird nest was subjected to molecular analysis. DNA from adult flies was extracted with QIAgen DNeasy kit, following the manufacturer’s protocol. The lower half of the abdomen of each fly was used for DNA extraction, as it contains the reproductive tissues in which *Wolbachia* is predominantly found. Extracted DNA was resuspended in 100 μl elution buffer.

(b) Nuclear analysis

We first attempted to reconstruct the phylogeny of the *Protocalliphora* genus by using nuclear sequence data (Internal Transcribed Spacers 1 and 2), but this was unsuccessful due to a very low level of substitutions between the species, major alignment problems caused by numerous indels, and the fact that the few observed substitutions between species were almost only autapomorphies.

We therefore used the amplified fragment length polymorphism technique (AFLP; Vos et al. 1995) to analyse the nuclear structure of the *Protocalliphora* genus because this technique has the ability to generate a large number of informative markers with relative ease. For each individual, genomic DNA was double-digested with EcoR1 and MseI. DNA fragments were ligated with EcoR1 and MseI adapters, generating template DNA for polymerase chain reaction (PCR) amplification (see Baudry et al. 2003 for details). A pre-selective amplification was performed using two primers complementary to the adapters and the restriction site sequences, in the following conditions: 94°C for 1 min, 56°C for 1 min 30 s and 72°C for 2 min, for a total of 35 cycles. Next, a selective PCR was performed with primers similar to the pre-selective amplification primers but with three additional bases at the 3′-end. A total of six primer combinations was used, with the following selective bases: E-TAC and M-TAC and M-GAT, E-TAC and M-CTG, E-GAT and M-ATC, E-GAT and M-CTG, E-GAT and M-CAG, with the Eco primer being fluorescently labelled with 6-FAM. The PCR products were run on an ABI 3700 Capillary DNA Sequencer, thus allowing us to estimate the size of the fragments with an error less than 0.2 bp. Fragments within the size range of 50–500 bp were kept for analyses.

The character matrix of presence or absence of bands produced by the AFLP procedure was analysed with PAUP v. 4.0 (Swofford 2002). A nuclear phylogenetic tree, rooted with *L. sericata* and *P. regina*, was constructed by parsimony analysis using a heuristic search with tree bisection–reconnection. Phylogenetic reconstruction was also performed by the neighbour-joining method (Saitou & Nei 1987) using
implemented in HY-PHY (www.hyphy.org) to test the Model test (Posada & Crandall 1998) procedure is the most appropriate for the COI and COII data. We used 1997) to determine which model of DNA sequence evolution replicates. The last column shows the Wolbachia infection status (§§2 and 3). NI designates non-infected individuals.

<table>
<thead>
<tr>
<th>species</th>
<th>nest label</th>
<th>location</th>
<th>bird host*</th>
<th>Wolbachia infectionb</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. asiovara</td>
<td>6852</td>
<td>USA, WA</td>
<td>black-billed magpie</td>
<td>NI</td>
</tr>
<tr>
<td>P. bennetti</td>
<td>7887</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>a2 A2; a2 B</td>
</tr>
<tr>
<td>P. bennetti</td>
<td>7893</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>a2 A2; a2 B</td>
</tr>
<tr>
<td>P. bennetti</td>
<td>7903</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>a2 A2</td>
</tr>
<tr>
<td>P. deceptor</td>
<td>6765</td>
<td>USA, TX</td>
<td>Carolina chickadee</td>
<td>NI</td>
</tr>
<tr>
<td>P. deceptor</td>
<td>6767</td>
<td>USA, TX</td>
<td>Bewick’s wren</td>
<td>NI</td>
</tr>
<tr>
<td>P. deceptor</td>
<td>6884</td>
<td>USA, OK</td>
<td>Bewick’s wren</td>
<td>NI</td>
</tr>
<tr>
<td>P. falcozi</td>
<td>ar1</td>
<td>France, Corsica</td>
<td>blue tit</td>
<td>a1 A1; a2 A2</td>
</tr>
<tr>
<td>P. falcozi</td>
<td>pac13</td>
<td>France, Corsica</td>
<td>blue tit</td>
<td>a1 A1; a2 A2</td>
</tr>
<tr>
<td>P. falcozi</td>
<td>fel18</td>
<td>France, Corsica</td>
<td>blue tit</td>
<td>a1 A1; a2 A2</td>
</tr>
<tr>
<td>P. halli</td>
<td>7884</td>
<td>Canada, BC</td>
<td>barn swallow</td>
<td>a2 A2; bB</td>
</tr>
<tr>
<td>P. halli</td>
<td>6998</td>
<td>USA, WA</td>
<td>barn swallow</td>
<td>a2 A2; a2 B</td>
</tr>
<tr>
<td>P. hirundo</td>
<td>6904</td>
<td>USA, WA</td>
<td>cliff swallow</td>
<td>bB</td>
</tr>
<tr>
<td>P. hirundo</td>
<td>7054-1</td>
<td>USA, WA</td>
<td>bank swallow</td>
<td>bA</td>
</tr>
<tr>
<td>P. metallica</td>
<td>6972-1</td>
<td>USA, OH</td>
<td>Carolina wren</td>
<td>aA1</td>
</tr>
<tr>
<td>P. occidentalis</td>
<td>7887</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>bA B</td>
</tr>
<tr>
<td>P. occidentalis</td>
<td>7903</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>bB</td>
</tr>
<tr>
<td>P. occidentalis</td>
<td>7025</td>
<td>USA, AZ</td>
<td>western bluebird</td>
<td>bB</td>
</tr>
<tr>
<td>P. rogesi</td>
<td>7054-2</td>
<td>USA, WA</td>
<td>bank swallow</td>
<td>a2 A2; a2 B</td>
</tr>
<tr>
<td>P. rogesi</td>
<td>7055</td>
<td>USA, WA</td>
<td>bank swallow</td>
<td>a2 B</td>
</tr>
<tr>
<td>P. rugosa</td>
<td>7887</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>bB</td>
</tr>
<tr>
<td>P. rugosa</td>
<td>7890</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>bB</td>
</tr>
<tr>
<td>P. rugosa</td>
<td>7893</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>bB</td>
</tr>
<tr>
<td>P. shannoni</td>
<td>7634</td>
<td>USA, OH</td>
<td>American Robin</td>
<td>NI</td>
</tr>
<tr>
<td>P. shannoni</td>
<td>7803</td>
<td>USA, OH</td>
<td>American Robin</td>
<td>a1 A</td>
</tr>
<tr>
<td>P. shannoni</td>
<td>6972-4</td>
<td>USA, OH</td>
<td>Carolina wren</td>
<td>a1 A</td>
</tr>
<tr>
<td>P. sialis</td>
<td>7811</td>
<td>Canada, SK</td>
<td>tree swallow</td>
<td>a2 A2; a2 B</td>
</tr>
<tr>
<td>P. sialis</td>
<td>7220</td>
<td>USA, OH</td>
<td>eastern bluebird</td>
<td>bB</td>
</tr>
<tr>
<td>outgroups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T. braueri</td>
<td>7851</td>
<td>USA, VA</td>
<td>mockingbird</td>
<td>NI</td>
</tr>
<tr>
<td>T. braueri</td>
<td>7903</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>NI</td>
</tr>
<tr>
<td>T. braueri</td>
<td>7909</td>
<td>Canada, BC</td>
<td>tree swallow</td>
<td>NI</td>
</tr>
</tbody>
</table>


b Wolbachia infection status: aA1, aA2, aB refers to the three different Wolbachia strains observed in *Protocalliphora* (§3). aA1, aA2 belong to the A super group, and aB to the B super group.

(c) Mitochondrial analysis
Two conserved primer pairs C1J-2183, C1-N-2659 and C2J-3138, TKN-3772 were used to respectively amplify a 374 bp fragment of the cytochrome oxidase I gene (COI) and a 579 bp fragment of the cytochrome oxidase II gene (COII). Thermocycle conditions were as described above. The PCR products were purified and then sequenced with an ABI 377 automatic sequencer (Perkin–Elmer). All COI and COII sequences were proof read and aligned manually.

Tree reconstruction and divergence calculation performed with the COI or COII data produced almost identical results (not shown); we therefore pooled the two datasets before analysis. Using COI and COII sequences to reconstruct the mitochondrial phylogeny of the genus, we started by performing likelihood ratio tests (Hueslenbeck & Rannala 1997) to determine which model of DNA sequence evolution is the most appropriate for the COI and COII data. We used the Model test (Posada & Crandall 1998) procedure implemented in HY-PHY (www.hyphy.org) to test hierarchically the effect of unequal base frequencies, different rates between transitions and transversions, different rates between all substitutions and rate variation over nucleotide sites. The model that best fit the dataset is a general time reversible (GTR) model with rate heterogeneity among sites (gamma distribution shape parameter of 0.167). We then used this model of sequence evolution to reconstruct a phylogenetic tree, rooted with *L. sericata* and *P. regina*, by maximum likelihood analysis (heuristic tree search with tree bisection–reconnection performed with PAUP v. 4.0 (Swofford 2002). The reliability of the tree obtained was examined using 1000 bootstrap replicates. Nucleotide sequence divergences between species were calculated with MEGA v. 3.1 (Kumar et al. 2004) using the Kimura two-parameter (K2P) model, the best metric when distances are low (Nei & Kumar 2000).

(d) Wolbachia analysis
A 454 bp fragment of the *wsp* gene was amplified by PCR, using the general *wsp* primers designed by Braig et al. (1998) for Wolbachia: *wsp* 81F and *wsp* 691R. Thermocycle conditions were 95°C for 1 min, 55°C for 1 min and 72°C for 1 min 30 s, for a total of 35 cycles. In the absence of
amplification, the PCR was repeated twice to confirm that the negative result was due to the absence of Wolbachia and not to a failure of the PCR procedure. The positive PCR products were purified and then sequenced with an ABI 377 automatic sequencer (Perkin–Elmer). In several cases, the sequencing results demonstrated the presence of two strains of Wolbachia in one individual. PCR products were then sequenced with primers specific for the two A groups (wA1 and wA2; Baudry et al. 2003) or the B group Wolbachia (Zhou et al. 1998).

3. RESULTS AND DISCUSSION

(a) Nuclear structure of the Protocalliphora genus

We have reconstructed the nuclear phylogeny of the genus using the AFLP technique. The six AFLP primer pairs used in this study generated a total of 1410 markers. Of these, 1391 (98.7%) were polymorphic and 897 (63.6%) were parsimony informative. The phylogenetic tree reconstructed by parsimony analysis from these data (figure 1) was almost identical to a neighbour-joining tree built with the same data (not shown).

Figure 1. Phylogram of the Protocalliphora genus based upon AFLP data. The tree was generated by parsimony analysis using a heuristic search with tree bisection–reconnection. Bootstrap values are shown as percentage of 1000 replicates at each node only if they are 50% or greater. The Wolbachia infection status of each individual is shown on the tree. Individuals infected with wA1, wA2 or wB Wolbachia strains are respectively represented by an open triangle, a solid triangle and a circle. Non-infected individuals are symbolized by NI.
Protocalliphora is a sister genus to COI and COII data is shown in figure 2. Ont this approximately at the same time. suggesting that most species of the genus have diverged shown, available from the authors upon request), from Internal Transcribed Spacers 1 and 2 (data not trying to construct a phylogeny with nuclear sequences values under 50%. We obtained similar results (partly unresolved, with several nodes having bootstrap cladogram, the relationships between species remain we used a very high number of characters to build the shared polymorphisms between species. Finally, although we used a very high number of characters to build the cladogram, the relationships between species remain partly unresolved, with several nodes having bootstrap values under 50%. We obtained similar results ($\S$2) when trying to construct a phylogeny with nuclear sequences from Internal Transcribed Spacers 1 and 2 (data not shown, available from the authors upon request), suggesting that most species of the genus have diverged approximately at the same time.

(b) Mitochondrial structure of the Protocalliphora genus and barcoding The phylogenetic tree representing the mitochondrial genetic structure of the Protocalliphora genus based upon COI and COII data is shown in figure 2. On this mitochondrial phylogeny, T. braueri also occupies a well-supported position outside Protocalliphora, confirming the status of Trypocalliphora as a sister genus to Protocalliphora. Second, Protocalliphora individuals always cluster by species, with bootstrap support values of 100% except in one case where the value was 96%. This confirms that morphology correctly identifies species in the Protocalliphora genus. However, note that the multilocus approach to reconstructing the nuclear DNA phylogeny may result in the species being monophyletic even if at individual loci there are sometimes shared polymorphisms between species. Finally, although we used a very high number of characters to build the cladogram, the relationships between species remain partly unresolved, with several nodes having bootstrap values under 50%. We obtained similar results ($\S$2) when trying to construct a phylogeny with nuclear sequences from Internal Transcribed Spacers 1 and 2 (data not shown, available from the authors upon request), suggesting that most species of the genus have diverged approximately at the same time.

The first noticeable characteristic of the cladogram is that T. braueri occupies a well-supported position outside Protocalliphora, confirming the status of Trypocalliphora as a sister genus to Protocalliphora. Second, Protocalliphora individuals always cluster by species, with bootstrap support values of 100% except in one case where the value was 96%. This confirms that morphology correctly identifies species in the Protocalliphora genus. However, note that the multilocus approach to reconstructing the nuclear DNA phylogeny may result in the species being monophyletic even if at individual loci there are sometimes shared polymorphisms between species. Finally, although we used a very high number of characters to build the cladogram, the relationships between species remain partly unresolved, with several nodes having bootstrap values under 50%. We obtained similar results ($\S$2) when trying to construct a phylogeny with nuclear sequences from Internal Transcribed Spacers 1 and 2 (data not shown, available from the authors upon request), suggesting that most species of the genus have diverged approximately at the same time.

The first objective of DNA barcoding is species identification. In practice, the sample to be identified is positioned in a previously characterized phylogeny, usually using neighbour-joining and/or parsimony analysis (Meyer & Paulay 2005). The test sample then receives the identity of its sister clade. Obviously, to be successful, this approach requires species to be monophyletic. In our sampling of Protocalliphora, reliable identification can therefore be obtained only for the four species mentioned above. Considering that our sampling comprises 10 species that are represented by more than one individual, this represents a success rate of only 40%.

Intraspecific divergence within species of Protocalliphora at the COI and COII loci ranged from 0.00 to 0.71%
Table 2. Divergence values between species at the COI and COII loci. (Within species divergence values are shown on the diagonal of the table. The three clusters defined by using the 1.8 or 3% divergence criterion (see Results and figure 2) are shown framed. Divergence values within each cluster are shown in bold. ND, not determined.)

<table>
<thead>
<tr>
<th>L. sericata</th>
<th>P. regina</th>
<th>T. brevior</th>
<th>Protocalliphora falcata</th>
<th>P. sialia</th>
<th>P. occidentalis</th>
<th>P. rugosa</th>
<th>P. rognesi</th>
<th>P. halli</th>
<th>P. falcozi</th>
<th>P. bennetti</th>
<th>P. sialia</th>
<th>P. hirundo</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00133</td>
<td>0.0124</td>
<td>0.0050</td>
<td>0.0043</td>
<td>0.0036</td>
<td>0.0101</td>
<td>0.0091</td>
<td>0.0126</td>
<td>0.0074</td>
<td>0.0074</td>
<td>0.0131</td>
<td>0.0032</td>
<td>0.0032</td>
</tr>
<tr>
<td>0.00133</td>
<td>0.0036</td>
<td>0.0074</td>
<td>0.0043</td>
<td>0.0036</td>
<td>0.0101</td>
<td>0.0091</td>
<td>0.0126</td>
<td>0.0074</td>
<td>0.0074</td>
<td>0.0131</td>
<td>0.0032</td>
<td>0.0032</td>
</tr>
<tr>
<td>0.0050</td>
<td>0.0074</td>
<td>0.0043</td>
<td>0.0036</td>
<td>0.0036</td>
<td>0.0101</td>
<td>0.0091</td>
<td>0.0126</td>
<td>0.0074</td>
<td>0.0074</td>
<td>0.0131</td>
<td>0.0032</td>
<td>0.0032</td>
</tr>
</tbody>
</table>

Table 2, with an average value of 0.18%. This value is close to the 0.25% value observed by Hebert et al. (2003a,b) for average intraspecific divergence in a large sample of Lepidoptera. Interspecific divergence shows a greater range, with values varying between 0.10 and 8.63%, and an average of 3.86% (table 2).

The second objective of DNA barcoding is to help in species discovery. The proposed method is to use a threshold value chosen to separate intraspecific from interspecific variation. An unidentified sequence differing from a known sequence by less than the threshold value will be considered to belong to this species, whereas if it differs by more than the threshold value, it will be considered to represent a new taxon. At least two methods have been proposed to choose a threshold. The first uses a fixed value considered suitable for the taxonomic group of interest. For example, it has been proposed that in insects, interspecific divergence almost always exceeds 3% (Hebert et al. 2003a,b) and this value can therefore be used as a threshold. Alternatively, Hebert et al. (2004) have proposed that a threshold of ten times the average intraspecific difference would be appropriate to screen for new animal species. In Protocalliphora, this would translate into using either 3 or 1.8% as a threshold for screening new species. In both cases, we obtained only three clusters of individuals (figure 2). Cluster II was the only one to correspond to one species (P. falcozi). Cluster III included two species, P. halli and P. rognesi, and cluster I comprised the remaining nine species. In this last cluster, the maximum divergence observed between species was only 1.64% and there was extensive haplotype sharing. Four species (P. sialia, P. occidentalis, P. rugosa and P. benetti) showed the same haplotype. Similarly, P. sialia and P. hirundo shared one haplotype. Using barcoding to identify new species in Protocalliphora would therefore very much underestimate the species richness of the genus, as only 3 species would be recognized instead of 12 (an underestimation of 75%). Note that this major underestimation of species number could not be significantly ameliorated by using a lower threshold value because the problem is caused mostly by the non-monophyly of the species and by the overlap between intra- and interspecific divergence values. At best, using an ad hoc threshold value of 1%, three more clusters would be recognized, but still only 50% of the species would be recognized.

(c) Wolbachia infection and mitochondrial DNA variation within Protocalliphora

It would of course be interesting to determine which factors have caused the major discrepancies observed between the mitochondrial and the nuclear structure of Protocalliphora. Since the bacterium Wolbachia was known to be present in at least one species of the genus (P. sialia, Baudry et al. 2003) and because interspecific mitochondrial introgression linked to Wolbachia infection has been described in insects (Rousset & Solignac 1995; Ballard 2000a,b; Jiggins 2003; Narita et al. 2006), Wolbachia seemed a possible candidate. We therefore determined the Wolbachia infection status of each Protocalliphora individual using a fragment of the wsp gene (table 1; figures 1 and 2).

Among the 12 species of our sample, only two, P. asiovora and P. deceptor, showed no Wolbachia-infected individuals. The Protocalliphora genus therefore seems to present a very high level of Wolbachia infection, with more
than 80% of the species harbouring the bacteria. Moreover, it is important to emphasize that the two species where we found no evidence of Wolbachia may not be infection free. The small sample size per species (one to three individuals) means that we are unlikely to detect infections within a species unless they are present at a very high frequency.

The results of the sequencing showed that only three Wolbachia strains were present in the Protocalliphora genus. Two bacteria belong to Wolbachia-A group (Werren et al. 1995b), hereafter called wA1 and wA2, and are observed in three and six of the species, respectively. The third one is a B group Wolbachia (Werren et al. 1995b), hereafter called wB, which is present in seven species. Note that these identifications are based on the wsp gene only and that the strains identity could be checked using the more powerful MLST method developed by Baldo et al. (2006). The three same strains were previously observed in P. sialia (Werren & Bartos 2001; Baudry et al. 2003). Several individuals are infected by two of the three strains. Finally, even with the small number of individuals analysed per species, we observed an intraspecific polymorphism of infection in four of the species.

There are two general explanations for the extensive sharing of Wolbachia strains and mitochondrial haplotypes among Protocalliphora species: (i) maintenance of an ancestral mitochondrial and infection polymorphism that existed prior to divergence of the infected species or (ii) movement of Wolbachia and their associated mitochondrial haplotype between species by interspecific hybridization after their speciation. In the first case, there should be a correlation between the nuclear phylogeny and the mitochondrial one, as well as with the Wolbachia infection status, i.e. closely related species at the nuclear level should have a higher tendency to share mitochondrial haplotype and Wolbachia strains. On the contrary, if mitochondrial introgression associated with Wolbachia infection did occur, we should observe cases where species not closely related at the nuclear level share very similar mitochondrial haplotype and Wolbachia strains.

In the Protocalliphora mitochondrial phylogeny, we observe three such cases (indicated by ellipses in figure 2). First, the four species that share an identical haplotype (P. sialia, P. occidentalis, P. rugosa and P. bennetti, see above) all harbour wB Wolbachia (and three of them also wA2). However, while P. sialia and P. occidentalis on one hand, and P. rugosa and P. bennetti on the other, are closely related at the nuclear level, the two pairs are not (figure 1). Second, P. metallica and P. shannoni show almost identical mtDNA haplotypes (1 bp difference), are both infected by the wA1 Wolbachia strain, but are only distantly related at the nuclear level. Similarly, P. halli and P. rogest show very similar mtDNA haplotypes (3 bp difference) and are both infected by the wA2 and wB Wolbachia strain, but are not closely related at the nuclear level. In these three cases, a parsimonious explanation for the observed pattern is that interspecific transfer of mtDNA and Wolbachia strain did occur. Of course, these two explanations are not mutually exclusive; while we believe that the observed patterns strongly suggest that interspecific mitochondrial and Wolbachia transfers did take place in at least three cases, maintenance of ancestral polymorphism in other cases is also possible.

Finally, it is interesting to note that, of the two species that show mitochondrial monophyly with a strong bootstrap support (P. deceptor and P. falcozi), the first is not infected by Wolbachia, while P. falcozi is the only species of our sample not from North America. This suggests that mitochondrial monophyly is observed in this genus only when mitochondrial introgression associated with Wolbachia transfer cannot occur, either because Wolbachia is absent or owing to geographical isolation.

(d) Barcoding in insects

In insects, three cases were already described where Wolbachia infection indirectly caused an interspecific mtDNA introgression, and our study adds a fourth one. In all four cases, the mtDNA introgression occurred between very closely related species, and was not accompanied by detectable nuclear introgression. The most probable explanation for this is that Wolbachia strains and associated mitochondrial haplotypes have been occasionally transferred from species to species by rare hybridization events. The rarity of these events, and the fact that the interspecific hybrids probably have a low fitness, would make the nuclear gene flow associated with these hybridizations negligible. In contrast, the selective advantage of Wolbachia results in its increase in frequency, and the infection and associated mtDNA haplotype spreads into the new species (Hurst & Jiggins 2005). It should be noted that interspecific mtDNA introgression associated with Wolbachia infections precludes identification at the species level based on COI barcoding. However, these introgressions are restricted to species that can hybridize, even if rarely, and therefore to very closely related species. This means that barcoding at a higher taxonomic rank, for example the genus, should remain possible in Wolbachia-infected species.

Although still controversial (e.g. Ebach & Holdrege 2005; Will et al. 2005; Rubinoff 2006), the scientific benefits expected of DNA barcoding include accelerating assignment of specimens to species that have been previously described and facilitating discovery of new species (Meyer & Paulay 2005; Savolainen et al. 2005; Lefebure et al. 2006). To produce accurate results, these two aspects of barcoding have different requirements. In comprehensively studied groups, assignment of a specimen to the correct species only requires species to be monophyletic at the mitochondrial level. Accurate species discovery also necessitate species to be monophyletic but, additionally, there should be an absence of overlap between intra- and interspecific variations, i.e. a barcoding ‘gap’ (Meyer & Paulay 2005).

Numerous studies have investigated the performance of the two aspects of barcoding. Most studies published to date suggest that barcoding achieves high accuracy in the task of assigning specimens to known species (e.g. Hebert et al. 2004; Janzen et al. 2005; Ward et al. 2005; Hajibabaei et al. 2006; Smith et al. 2006), implying that species are usually monophyletic at the mitochondrial level (but see Funk & Omland 2003). In contrast, there is a disagreement regarding the performance of barcoding for the discovery of new species. Earlier studies suggested a low error rate but they usually undersampled intraspecific variation (because very few individuals were sampled by species) and interspecific divergence (because closely related species were not always sampled; Meyer & Paulay 2005). A recent
study by Meyer & Paulay (2005) provides the first examination of barcoding performance in a comprehensively sampled, diverse group (cypraeid marine gastropods, or cowries). They found that due to a substantial overlap between intra- and interspecific variations, discovery of new species using barcoding would lead to an unacceptable error rate. In contrast, our study on the Protocalliphora genus showed a very high error rate for both aspects of barcoding, specimen identification and species discovery. Studies on the performance of barcoding in comprehensively sampled insects groups are needed to determine whether or not the Protocalliphora case is an exception in insects.

REFERENCES


