Diversification of Wolbachia Endosymbiont in the Culex pipiens Mosquito
Célestine Atyame Nten, Frédéric Delsuc, Nicole Pasteur, Mylène Weill, Olivier Duron

To cite this version:
Célestine Atyame Nten, Frédéric Delsuc, Nicole Pasteur, Mylène Weill, Olivier Duron. Diversification of Wolbachia Endosymbiont in the Culex pipiens Mosquito. Molecular Biology and Evolution, Oxford University Press (OUP), 2011, 28 (10), pp.2761-2772. 10.1093/molbev/msr083. hal-01924446

HAL Id: hal-01924446
https://hal-sde.archives-ouvertes.fr/hal-01924446
Submitted on 16 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Diversification of *Wolbachia* endosymbiont in the *Culex pipiens* mosquito

Célestine M. Atyame*, Frédéric Delsuc*, Nicole Pasteur*, Mylène Weill* and Olivier Duron*

* Institut des Sciences de l’Evolution
Université Montpellier 2, CC 054
CNRS, UMR5554
Place Eugène Bataillon
34095 Montpellier cedex 05
France

Corresponding author: Olivier Duron; E-mail: olivier.duron@univ-montp2.fr

Type of article: Research Article

Keys words: Endosymbiosis, *Wolbachia*, mitochondrial hitchhiking

Running title: Diversification of *Wolbachia* endosymbiont
Abstract

The α-proteobacteria *Wolbachia* are among the most common intracellular bacteria and have recently emerged as important drivers of arthropod biology. *Wolbachia* commonly act as reproductive parasites in arthropods by inducing cytoplasmic incompatibility (CI), a type of conditional sterility between hosts harboring incompatible infections. In this study, we examined the evolutionary histories of *Wolbachia* infections, known as *wPip*, in the common house mosquito *Culex pipiens*, which exhibits the greatest variation in CI crossing patterns observed in any insect. We first investigated a panel of twenty *wPip* strains for their genetic diversity through a multi-locus scheme combining thirteen *Wolbachia* genes. Because *Wolbachia* depend primarily on maternal transmission for spreading within arthropod populations, we also studied the variability in the co-inherited *Cx. pipiens* mitochondria. In total, we identified fourteen *wPip* haplotypes, which all share a monophyletic origin and clearly cluster into five distinct *wPip* groups. The diversity of *Cx. pipiens* mitochondria was extremely reduced, which is likely a consequence of cytoplasmic hitchhiking driven by a unique and recent *Wolbachia* invasion. Phylogenetic evidence indicates that *wPip* infections and mitochondrial DNA have co-diverged through stable co-transmission within the cytoplasm and shows that a rapid diversification of *wPip* has occurred. The observed pattern demonstrates that a considerable degree of *Wolbachia* diversity can evolve within a single host species over short evolutionary periods. In addition, multiple signatures of recombination were found in most *wPip* genomic regions, leading us to conclude that the mosaic nature of *wPip* genomes may play a key role in their evolution.
Introduction

Arthropods are commonly infected by maternally transmitted endosymbionts. Although some endosymbionts confer direct benefits to their hosts by providing anabolic functions or resistance to pathogens (Haine 2008; Moran, McCutcheon, and Nakabachi 2008), others are associated with alterations of host reproduction (Werren, Baldo, and Clark 2008; Engelstadter and Hurst 2009). These symbionts represent reproductive parasites that include diverse unrelated bacteria, among which the alpha-proteobacteria *Wolbachia* are the most widespread (Duron et al. 2008; Hilgenboecker et al. 2008). In some host species, the successful spread of *Wolbachia* is achieved by biasing the host’s sex ratio toward the production of females (the transmitting sex) through the induction of parthenogenesis, feminization or male-killing. More commonly, *Wolbachia* are able to induce a form of conditional sterility, termed cytoplasmic incompatibility (CI), between infected males and uninfected females or females infected by incompatible strains (Werren, Baldo, and Clark 2008; Engelstadter and Hurst 2009). Such manipulations enable *Wolbachia* to spread through arthropod populations and may drive arthropod evolution through their effects on host phenotypes (Moran, McCutcheon, and Nakabachi 2008; Werren, Baldo, and Clark 2008; Engelstadter and Hurst 2009).

The dynamics of *Wolbachia* infections within the common house mosquito, *Culex pipiens* complex, remain poorly understood. The most common members of the complex are the subspecies *C. p. quinquefasciatus* (Say) and *C. p. pipiens* (L.), representing the southern and northern mosquito populations, which are ubiquitous in tropical and temperate regions, respectively (Barr 1982). Members of the *C. pipiens* complex exhibit the greatest variation of CI crossing types observed in arthropods thus far (Laven 1967; O'Neill and Paterson 1992; Guillemaud, Pasteur, and Rousset 1997; Duron et al. 2006; Atyame et al. in press). However, an early genotyping approach using the *ftsZ* gene failed to reveal any polymorphism between...
incompatible Wolbachia strains (Guillemaud, Pasteur, and Rousset 1997). Further analyses used sequences of published complete Wolbachia genomes to characterize polymorphic molecular markers. Genomes of Wolbachia strains infecting arthropod are scattered with mobile genetic elements (MGE), such as prophages and transposable elements, which can represent more than 20% of genome content (Wu et al. 2004; Klasson et al. 2008; Salzberg et al. 2009; Klasson et al. 2009). Additionally, they contain an unusual high number of genes encoding proteins with ankyrin (ANK) motif(s), which possibly mediate specific protein-protein interactions (Sinkins et al. 2005; Duron et al. 2007; Walker et al. 2007). We further developed genotyping approaches using ANK and MGE markers and used them to identify more than 100 genetically distinct Wolbachia strains (referred to as wPip strains) in natural populations of Cx. pipiens (Duron et al. 2005; Duron et al. 2006; Duron et al. 2007; Atyame et al. in press; Duron, Raymond, and Weill in press).

In this study, we characterized the evolutionary history of Wolbachia infections in the Cx. pipiens complex by examining the association between the wPip strains and Cx. pipiens mitochondrial DNA variation. The predominant mode of Wolbachia transmission within a species is vertical, via the egg cytoplasm (Werren, Baldo, and Clark 2008; Engelstadter and Hurst 2009). Because Wolbachia and mitochondrial genomes are co-transmitted and, therefore, in linkage disequilibrium, the spread of Wolbachia will strongly affects a host’s mtDNA diversity through indirect selection (review in Hurst and Jiggins 2005). However, exceptions to strict vertical transmission have been found; in some cases, Wolbachia are also transferred through horizontal transmission both within and among different host species, although the mechanisms of transfer are not well understood (Ahrens and Shoemaker 2005; Baldo et al. 2008; Raychoudhury et al. 2009). Consequently, the wide distribution of Wolbachia among arthropods is generally assumed to result from complex interactions.
between vertical and horizontal modes of transmission, modulated by their capacity to alter
host reproduction.

Here, we analyzed wPip variability and the associated mtDNA diversity in twenty Cx. pipiens
lines encompassing uni- and bidirectionally incompatible strains that originated from different
geographic areas. Our results showed that the wPip strains form a monophyletic clade of
closely related bacteria and that Cx. pipiens harbors a low level of mitochondrial variability,
which is a probable consequence of a recent Wolbachia invasion through cytoplasmic
hitchhiking. Investigation of wPip sequences revealed extensive recombination between wPip
strains, although multiple infections within a single mosquito were never detected using our
markers. However, a congruence between wPip and mtDNA phylogenies was shown,
demonstrating that Wolbachia mainly use maternal inheritance to spread through Cx. pipiens
populations. The evolutionary implications of horizontal transfers and the question of whether
the Cx. pipiens - Wolbachia association is a unique case or a representative example is
discussed.

Materials and Methods

Mosquito collection

Twenty Cx. pipiens lines from a broad geographical range were examined (Table 1). This
collection encompassed the two main Cx. pipiens subspecies, Cx. p. pipiens and Cx. p.
quinquefasciatus, which are naturally infected with compatible and incompatible wPip strains
(for more details, see Duron et al. 2006; Duron et al. 2007). The study also included the two
lines for which the wPip genome has been sequenced, wPip(Pel) (GenBank AM999887;
Klasson et al. 2008) and wPip(JHB) (ABZA01000000; Salzberg et al. 2009).
**Wolbachia markers**

The wPip strains were first genotyped for the five housekeeping genes developed for the *Wolbachia* MultiLocus Strain Typing (MLST) methodology, *gatB*, *coxA*, *hcaA*, *ftsZ* and *fbpA* (Baldo et al. 2006), and the *Wolbachia* surface protein gene *wsp* (Braig et al. 1998). The MLST system is classically used to characterize the eight supergroups (A to I) currently recognized within the *Wolbachia* genus (Lo et al. 2007; Ros et al. 2009). The polymorphism of seven additional genes was also examined: the DNA mismatch repair protein gene *MutL* (one copy in the wPip(Pel) genome), three ANK genes, *ank2* (one copy), *pk1* (three identical copies), and *pk2* (two identical copies), and three phage genes, the methylase gene *GP12* (four identical copies), the putative secreted protein gene *GP15* (also known as *VrlC*; one copy) and the regulatory protein gene *RepA* (one copy). None of these genes was amplified from *Wolbachia*-free *Cx. pipiens* lines, which confirmed their *Wolbachia* origin. A total of 13 *Wolbachia* genes, encompassing 19 distinct loci with a wide distribution along the wPip(Pel) chromosome, were examined (Figure 1, Table S1).

**Culex pipiens** mitochondrial markers

The complete mitochondrial genome of the Pel line (15,587 bp) was obtained through BLAST searches of the database of wPip(Pel) contig DNA sequences from the Wellcome Trust-Sanger Institute website (http://www.sanger.ac.uk/Projects/W_pipientis/) using the mitochondrial genome of *Aedes albopictus* (GenBank AY072044) as a probe. A similar approach to obtain the mitochondrial sequences of the JHB line from the VectorBase website (http://www.vectorbase.org/) showed several divergent mitochondrial sequences, some of which exhibited low read coverage, making the JHB sequences unreliable for further analysis.
A set of primers (Table S2) was designed from the Pel sequences and further used to obtain the complete mitochondrial genomes (with the exception of the A+T-rich region) of additional Cx. pipiens lines. Specific primers were also used to amplify a 613 bp fragment from the NADH dehydrogenase subunit 2 (ND2) gene, a 1,132 bp fragment from the NADH dehydrogenase subunit 5 (ND5) gene and an 852 bp fragment from the cytochrome b (cytb) gene from all the investigated Cx. pipiens lines (Table S2).

PCR amplification and sequencing
DNA was extracted from individual mosquitoes using a CetylTrimethylAmmonium Bromide (CTAB) protocol (Rogers and Bendich 1988). Amplification conditions were 3 min at 94°C, followed by 30 cycles of 94°C for 30 s, 52°C for 30 s (58°C for MutL), and 72°C for 1 to 1.5 min depending on the fragment size. Amplified fragments were run in agarose gel (1.5%) electrophoresis. The QIAquick gel extraction kit (QIAGEN, Valencia, CA) was then used to purify the PCR products. Sequences were obtained directly for purified products using an ABI Prism 3130 sequencer with the BigDye Terminator Kit (Applied Biosystems). The sequences have been deposited in the GenBank database (accession numbers in Tables S1 and S2).

Sequence analyses
Sequence alignments were carried out using CLUSTALW (Thompson, Gibson, and Higgins 2002) and corrected using MEGA (Tamura et al. 2007). The GBLOCKS program (Castresana 2000) with default parameters was used to remove poorly aligned positions and to obtain nonambiguous sequence alignments. The number of variable sites (VI), nucleotide diversity (\(\pi\)), G+C content and the ratios of nonsynonymous versus synonymous substitutions (\(Ka/Ks\)) were computed using DNASP (Librado and Rozas 2009). Nonrandom associations between each pair of loci were estimated through the measure of allele linkage disequilibrium (LD)
using the $D'$ statistic (Lewontin 1964; Hedrick 1987). The exact test procedure implemented in GENEPOP (Raymond and Rousset 1995) was further used to test LD significance. Statistical analyses for intragenic recombination were performed with the Sawyer’s test implemented in GENECONV (Sawyer 1999). A Bonferroni adjustment correction for multiple testing was applied (Hochberg 1988).

Annotation of the *Cx. pipiens* mitochondrial genome was based on alignments with mitochondrial sequences from *Ae. albopictus* (AY072044), *Ae. aegypti* (EU352212), *Anopheles gambiae* (L20934), *An. funestus* (DQ146364) and *An. quadrimaculatus* (L04272).

Tree reconstruction

Phylogenetic relationships were evaluated for *Wolbachia* and *Cx. pipiens* mitochondrial sequences. The best-fitting models of sequence evolution for each dataset were determined using the Akaike information criterion with MODELTEST v3.7 (Posada and Crandall 1998). The selected model was the General Time Reversible model, with Gamma distributed among-site rate variation (GTR+G) for both *Wolbachia* and mitochondrial sequence datasets.

Bayesian inferences (BI) were used to reconstruct phylogenies using Mr. Bayes v 3.1.2 (Ronquist and Huelsenbeck 2003). Two independent replicates of four Metropolis-coupled Monte Carlo Markov chains were run for 2,000,000 generations with Model parameters and trees sampled every 200 generations. Bayesian posterior probabilities were obtained from the 50% majority-rule consensus of the sampled trees after discarding the initial burn-in period. The resulting phylogenetic trees were visualized and edited in MEGA (Tamura et al. 2007).

*Wolbachia* genes were also analyzed within a phylogenetic network framework to account for potentially conflicting signals due to recombination (Fitch 1997). A phylogenetic network
was constructed based on uncorrected $P$-distances using the Neighbor-net method (Bryant and Moulton 2004) implemented in SPLITSTREE (Huson and Bryant 2006). Neighbor-net is a distance-based method to construct a network as a generalization of all possible phylogenetic trees that can be reconstructed from conflicting signals in the data.

Assessing the maximum age of mitochondrial sweep

We used the mitochondria data to infer the maximum age of the Wolbachia infection in Cx. pipiens following the method of Rich et al. (1998). This method assumes that selection only occurs at the protein level and that DNA polymorphism in degenerate sites is neutral. We used fourfold and twofold synonymous sites from protein coding mtDNA sequences to assess the age of the sweep. The number of twofold and fourfold synonymous sites in each coding region was computed with MEGA (Tamura et al. 2007), and a conservative Jukes-Cantor correction was applied for multiple hits.

Results

Monophyletic origin of the wPip strains

The MLST and wsp genes did not exhibit sequence variation between the wPip strains (eight strains were examined here, i.e. wPip(Sl), wPip(Tn), wPip(Ko), wPip(Lv), wPip(Is), wPip(Mc), wPip(Pel) and wPip(JHB)), establishing that these strains are very closely related. The wPip MLST sequence data were compared to sequences from 18 other strains belonging to five distinct Wolbachia supergroups (A, B, D, F and H). The phylogenetic tree obtained from the 2,079-bp concatenated MLST genes revealed that the wPip strains form a robust monophyletic clade within the B supergroup, which is closely related to the wBol1 strain present in the butterfly Hypolimnas bolina (identity > 99.9%; Figure S1).
High variability of wPip genomes

Seven of the examined Wolbachia genes were polymorphic among the wPip genomes: the DNA mismatch protein gene MutL, three ANK genes ank2, pk1 and pk2, and three phage genes GP12, GP15 and RepA. Analyses revealed considerable allelic variability among the 20 wPip strains, with 2-8 alleles being found per gene (Table S3). This polymorphism was mainly due to nucleotide substitutions, insertions or deletions (indels); note that an insertion of the Trl transposon (also known as ISWpi1; see Duron et al. 2005; Cordaux 2008) was observed within the RepA sequence of three wPip strains. An additional source of variability arose from a GP15 deletion in the wPip(JHB) genome. A letter was attributed to each distinct allele of the seven genes, the combination of which identified 14 wPip haplotypes among the twenty strains (Table 2).

Although the prophage related genes pk1, pk2 and GP12 were found to be present in several copies in the wPip(Pel) genome, divergent copies were never amplified from our wPip strains: direct sequences of PCR products were easily readable and showed no overlapping peaks. This indicates that the different copies (if any) present in each wPip strain examined here are identical, as observed in wPip(Pel). It further shows that only mono-wPip-infections (or multi-infections by closely related wPip strains) are present within Cx. pipiens individuals.

High recombination in wPip genomes

Recombination, both within and between Wolbachia genes, can blur molecular signals and result in misleading observations related to strain relationships. For this reason, we checked the possibility of recombination among the seven polymorphic markers obtained here (MutL, ank2, pk1, pk2, GP12, GP15 and RepA). Pairwise tests for intergenic recombination revealed
significant linkage disequilibrium (LD) for MutL, ank2, pk1, GP12, and GP15 (Table S4).

Alleles at these five loci are not randomly associated and are stably co-transmitted within the wPip chromosome. However, nonsignificant LD was found between pk2 and RepA and between these two genes and the other five genes, showing that recombination has disrupted genome clonality by shuffling the RepA and pk2 alleles among wPip strains. Intragenic recombination was also detected for at least five genes by Sawyer’s test (MutL, pk1, pk2, GP12 and GP15; Table S3). Intragenic recombination results in identical nucleotides or amino acid motifs in wPip strains divergent at other loci, which are readily apparent through the examination of sequence alignments (Figure S2).

Gene rearrangements in wPip genomes

Genome organization of wPip strains was analyzed by comparing the locations of the thirteen genes surveyed in this study in the wPip(Pel) chromosome and in the five major wPip(JHB) contigs presently available (Figure S3). There are several rearrangements distinguishing these genomes, in which diverse genes have been inverted (e.g. gatB, coxA), translocated (MutL, RepA), duplicated or deleted (three and one pk1 copies are found in wPip(Pel) and wPip(JHB), respectively). Notably, rearrangements are not limited to phage regions, which are prone to movements within and between genomes, but also affect housekeeping genes.

Inference of wPip strain relationships

Phylogenetic analyses of the 20 wPip strains using the six wPip genes, MutL, ank2, pk1, pk2, GP12 and GP15, revealed significant topological incongruence, as expected for a dataset affected by recombination (Figure S4). For instance, the wPip(Sl), wPip(Bf-B) and wPip(Mc) strains are genetically similar for four markers (ank2, pk1, GP12 and GP15), but appear distantly related for two others markers (MutL and pk2).
To assess \( w \)Pip strain relationships, we performed phylogenetic analyses based on the concatenated sequences of the seven genes. The concatenated tree deduced from Bayesian inference splits the \( w \)Pip clade into five groups (designated \( w \)Pip-I to \( w \)Pip-V; Figure 2). However, recombination can create artificial grouping of \( w \)Pip strains, and network analysis was thus conducted to visualize recombination effects, which were illustrated by multiple boxes (Figure 3). The evolutionary history of \( w \)Pip strains appears as a complex network with multiple pathways interconnecting strains, emphasizing the mosaic nature of \( w \)Pip genomes. Interestingly, despite recombination, the network analysis was congruent with the Bayesian tree in recovering the same five \( w \)Pip groups with strong bootstrap values.

A spatial structuring of \( w \)Pip diversity emerged when the geographic distribution of \( w \)Pip groups was examined, despite the limited number (19) of strains. The most common group, \( w \)Pip-I, is distributed widely from Asia to Europe (Figure 4), and all \( w \)Pip strains recently identified at La Réunion island (Indian Ocean) by Atyame et al. (in press) belong to that group. The \( w \)Pip-V group is only found in East Asia, and the \( w \)Pip-II and \( w \)Pip-III groups have an apparently discontinuous distribution, with strains being found in very distant geographic areas (for example, the \( w \)Pip-II strains are from Australia and Europe).

Low mitochondrial diversity in \( Cx. \) pipiens

\( Cx. \) pipiens Pel mitochondrial DNA exhibits classical features found in other mosquito species that have been analyzed. It contains tightly packed genes with high A+T content (78.2%). There are twenty-two genes coding tRNAs, two coding ribosomal RNAs, thirteen genes coding subunits of enzymes involved in oxidative phosphorylation and, finally, an A+T rich noncoding region (Figure S5 and Table S5). The genes are arranged along the
chromosome in a manner similar to that of other mosquito species (Beard, Mills Hamm, and Collins 1993; Mitchell, Cockburn, and Seawright 1993; Krzywinski, Grushko, and Besansky 1997).

The complete *Cx. pipiens* mitochondrial genome (14,856 bp without the A+T rich region) was sequenced from the lines Ko, Tn, Sl and Is and compared to the Pel genome. Overall, the five mtDNA sequences displayed a very low variability, with only 36 variable nucleotidic positions being found (ca. 2%), and two sequences were strictly identical (lines Ko and Tn).

Among the 13 protein-coding genes, five genes (*atp8, atp6, ND3, ND4L, and ND6*) showed no polymorphism, whereas *ND2, ND5*, and *cytb* were the most polymorphic (Figure S5).

A likely explanation of the low mtDNA diversity in *Cx. pipiens* populations is that cytoplasmic hitchhiking has occurred during *Wolbachia* invasion, as suggested earlier by Guillemaud et al. (Guillemaud, Pasteur, and Rouset 1997) and Rasgon et al. (Rasgon, Cornel, and Scott 2006). To confirm this hypothesis, we compared the nucleotide diversity per site (π) at four mitochondrial loci in the eleven Culicidae species for which the presence or absence of *Wolbachia* has been documented (Table 3 and supplementary materials). Only two species, *Cx. pipiens* and *Ae. albopictus*, are known to be infected, whereas *Wolbachia* infection was never found in the nine other species. These two *Wolbachia*-infected species harbor significantly lower mtDNA diversity than the uninfected species (Wilcoxon test, W=10, P = 0.008). For instance, the worldwide mtDNA diversity of *Cx. pipiens* is lower than the diversity observed in the North American populations of *Cx. tarsalis*, an uninfected species (Venkatesan et al. 2007). The low diversity of mtDNA observed in the *Cx. pipiens* and *Ae. albopictus* populations led us to conclude that *Wolbachia* is most likely the causative agent of mitochondrial sweeps in these taxa.
Recent mitochondrial sweep in *Cx. pipiens* complex

We then assessed the date of the mitochondrial sweep using the nucleotide divergence of 13 protein-coding mtDNA genes from the *Cx. pipiens* Is line and *Ae. albopictus* (GenBank AY072044). We estimated the substitution rate for these genes at twofold and fourfold degenerate sites with a conservative Jukes–Cantor correction. The genera *Culex* and *Aedes* diverged approximately 172 to 226 My ago (Reidenbach et al. 2009). Using the most recent estimate (172 My), the mtDNA substitution rates (substitution/site/year) were estimated at $5.1 \times 10^{-8}$ and $19 \times 10^{-8}$ for the twofold and fourfold degenerate sites, respectively, whereas, when using the oldest estimate (226 My), the substitution rates were $39 \times 10^{-9}$ and $15 \times 10^{-8}$.

Among the mitochondrial genomes of five *Cx. pipiens* lines (Is, Sl, Tn, Ko and Pel), we observed 13 and 10 nucleotide differences among the twofold (n=2,938) and fourfold (n=1,343) synonymous sites, respectively. Thus, this dates the *Cx. pipiens* mitochondrial sweep between 12,000 and 16,000 years before present (95% confidence interval, if *Culex* and *Aedes* diverged 172 My ago) or between 16,000 and 21,000 (95% confidence interval, if the two genera diverged 226 My ago). It is possible that the date of the mitochondrial sweep is even more recent as it was assumed here that the substitution rates are constant, an assumption known to overestimate divergence times (Ho et al. 2005).

Clear co-divergence of wPip and *Cx. pipiens* mitochondria

The co-divergence of mitochondria and wPip was assessed by studying the sequences of the three polymorphic *Cx. pipiens* mtDNA genes (*ND2, ND5* and *cytb*), encompassing 2,549 bp (16.4% of the whole mitochondrial genome).
Analysis of the *Cx. pipiens* mtDNA sequences among the 19 lines indicated the presence of 14 haplotypes (named *pi*1 to *pi*14), which differed overall at 22 variable nucleotide sites (Table S6). The mtDNA of the *Cx. pipiens* lines differed by only one to nine nucleotides, confirming their very high homology (99.6-99.9%). Phylogenetic analyses revealed two main mitochondrial lineages (*pi*1 to *pi*5 and *pi*6 to *pi*14) with strong branching support (Figure 5A).

The concatenated mtDNA phylogeny and the wPip phylogeny were congruent (Figure 5A and B). A significant association was found between mtDNA haplotypes and wPip haplotypes (Fisher’s exact test, *P* = 3×10⁻⁵), as well as wPip groups (*P* = 8×10⁻⁴). This demonstrates that wPip infections and mitochondrial DNA have co-diverged through stable co-transmission within the cytoplasm in *Cx. pipiens* populations. Hence, the two main mitochondrial lineages parallel the wPip divergence pattern and strongly confirm the wPip phylogeny. Additionally, *Cx. pipiens* subspecies are not significantly associated with wPip haplotypes (*P* = 0.37), wPip groups (*P* = 0.26) or mtDNA haplotypes (*P* = 0.10). Thus, *Cx. pipiens* nuclear genomes have not co-diverged with mitochondria and wPip infections and exhibit a different evolutionary history.

**Discussion**

Here, we examined 20 isolates of *Wolbachia* and their associated mitochondria within the *Cx. pipiens* complex. The combined use of *Wolbachia* and host mtDNA multi-locus sequencing revealed the processes driving the evolution of *Wolbachia* infections in this mosquito and raised the question of their likeliness to occur in other *Wolbachia*-arthropod associations.

*Wolbachia* and mitochondrial markers reveal a recent diversification of wPip strains.
The examined MLST and wsp markers showed that wPip infections form a robust monophyletic clade within the B group of *Wolbachia*, confirming the results of Baldo et al. (Baldo et al. 2006). Although these markers are widely used to characterize the genetic diversity of *Wolbachia*, even within a host species (e.g. Dedeine et al. 2004; Baldo et al. 2006; Baldo et al. 2008; Raychoudhury et al. 2009), they displayed no variation among the wPip infections in *Cx. pipiens*, which shows that wPip strains have a unique and recent evolutionary origin. As observed in other *Wolbachia* infected species (Hurst and Jiggins 2005), mitochondrial diversity was low in populations of *Cx. pipiens* (and highly significantly lower than in non-*Wolbachia*-infected mosquitoes), suggesting that *Wolbachia* have affected mitochondrial polymorphism in this species through cytoplasmic hitchhiking. The observed polymorphism of mitochondrial protein coding genes indicates that the mitochondrial sweep due to the spread of *Wolbachia* occurred within the last 21,000 years. This dating is within the range of values classically estimated for other *Wolbachia* host species of ca. < 100,000 years, (Jiggins 2003; Keller et al. 2004; Duplouy et al. 2010), and the evolutionary pathway of wPip in *Cx. pipiens* could be similar to other *Wolbachia*/arthropod associations.

Multi-locus typing using seven wPip polymorphic markers, including domains of the MGE and ANK genes, allowed the identification of 14 distinct wPip haplotypes, which cluster into five distinct wPip groups. This typing approach also established that the two published wPip genomes, wPip(Pel) and wPip(JHB), are genetically very close to each other compared to strains belonging to other wPip groups, in spite of their genomic differences (Salzberg et al. 2009). The variability of the investigated mitochondrial markers corroborates the inferences made from the wPip markers; thus, in *Cx. pipiens*, different mitochondrial haplotypes may indicate that wPip infections are different. Overall, the observed genetic diversity indicates
that, after the spread of *Wolbachia*, diversification of *wPip* and *Cx. pipiens* mitochondria occurred.

The diversity found for *wPip* exhibits geographic variations. A remarkable degree of diversity was found in the Mediterranean area, where four of the five *wPip* groups are found, whereas a reduced diversity was observed in other regions. The most common group, *wPip*-I, has a wide distribution (Asia, Africa and Europe) and was also recently reported at La Réunion Island (Indian Ocean) (Atyame et al. in press). In contrast, some *wPip* groups have a discontinuous distribution, as exemplified by the *wPip*-II strains, which were found in Europe and in Australia. Such a geographic pattern is likely to be a consequence of a recent worldwide expansion due to human activity (Raymond et al. 1991; Fonseca et al. 2004; Fonseca et al. 2006) or/and to selective advantages, possibly including CI selection. However, the 20 *wPip* infections investigated in this study represent a restricted sampling, occasionally from old mosquito colonies, and further investigations are required to improve our knowledge of the spatial structure of the *wPip* groups worldwide.

*wPip* strains are independent of *Cx. pipiens* subspecies

Strict vertical transmission must have favored the co-divergence of *wPip* and mtDNA within shared cytoplasm. However, there was no clear association between *Cx. pipiens* subspecies (nuclear diversity) and cytoplasmic diversity (i.e. *Wolbachia* and mtDNA): identical *wPip* strains and identical mitochondrial haplotypes were found in the two subspecies, *Cx. p. pipiens* and *Cx. p. quinquefasciatus*. A likely explanation for this is that the transfer of cytoplasm between *Cx. pipiens* subspecies occurred through hybridization events, as observed in *Drosophila* species (Rousset and Solignac 1995; Ballard 2000) and in butterfly species (Jiggins 2003; Narita et al. 2006; Charlat et al. 2009). In *Cx. pipiens*, this hypothesis is well
supported by the many reports of genetic introgression between the two subspecies in areas where they come into contact (Cornel et al. 2003; Fonseca et al. 2004). Hence, we can predict that DNA barcoding programs using mtDNA will fail to discriminate between *Cx. p. pipiens* and *Cx. p. quinquefasciatus*. Overall, these observations support the call of Hurst and Jiggins (Hurst and Jiggins 2005) to not use mtDNA alone as a reliable means of taxa resolution.

**Intense recombination impacts the structure of *wPip* genomes**

The existence of extensive recombination among *wPip* strains sheds light on the mechanisms shaping the evolution of *wPip* genomes since recombination can influence the adaptive dynamics of *Wolbachia* by creating new alleles and thus allow the emergence of new phenotypes. Recombination between distant *Wolbachia* genomes has been previously documented (Jiggins et al. 2001; Bordenstein and Wernegreen 2004; Baldo, Lo, and Werren 2005; Gavotte et al. 2007), although in this study, we found recombination among very closely related *Wolbachia* genomes. Evidence of recombination was found at almost all of the examined *wPip* loci, WO-phage genes, as well as nonrelated phage loci. This shows that a high level of gene flow occurs among the *Wolbachia* genomes in *Cx. pipiens*. Hence, the *wPip* strains do not form a set of clones in which evolution is independent but, rather, represent a large population of bacteria exchanging genetic information through lateral transfers. Although no instances of multiple infection were detected using our markers, we must assume that they occur, at least during a period long enough to allow recombination between strains.

Another consequence of recombination is that it can lead to misinterpretation of phylogenetic relationships between strains. However, despite the extensive recombination observed, the *wPip* and mitochondrial phylogenies are congruent: recombinations have not disrupted our grasp of the evolutionary history of *wPip* strains, probably because the contribution of recombinant regions in the phylogeny is weak compared to the diversity existing in
nonrecombinant DNA fragments. Therefore, as suggested by Baldo et al. (Baldo et al. 2006),
the use of a multi-locus approach, rather than single locus analysis, is required for a correct
understanding of the evolutionary history of Wolbachia infections.

The Cx. pipiens-Wolbachia association, a unique case?
The high number of wPip strains, which is still certainly underestimated, makes the Cx.
pipiens system remarkable because lower diversity is usually reported in Wolbachia of other
host species (e.g. Vavre et al. 1999; Mercot and Charlat 2004; Charlat et al. 2006; Arthofer et
al. 2009). However, it is possible that genetic variations of Wolbachia in other host species
could have been missed due to the methodology generally used to characterize these bacteria,
as it is generally assumed that a single Wolbachia strain is present within a host species when
the MLST or wsp markers are not variable. In D. melanogaster, a single strain, wMel, was
presumed to be present until Riegler et al. (Riegler et al. 2005) identified five distinct
genotypes by examining transposon insertion sites and chromosomal inversions. More recent
studies have reported different Wolbachia haplotypes solely on the basis of WO phage genes
in various host species, including crickets, beetles and butterflies (Charlat et al. 2009; Chafee
et al. 2010). Hence, the classical MLST system is well suited to characterize Wolbachia
belonging to distinct clades, but specific species typing systems based on markers with rapid
sequence evolution need to be developed to investigate the Wolbachia diversity that probably
exists in most associations.

Finally, the question remains of whether the Cx. pipiens-Wolbachia association is unique in
term of its extremely large CI diversity (e.g. Laven 1967; O'Neill and Paterson 1992;
Guillemaud, Pasteur, and Rousset 1997; Duron et al. 2006). We have clearly demonstrated
that the diversity of crossing types in this species is independent of nuclear backgrounds and
relies solely on wPip variability (Duron et al. 2006; Atyame et al. in press). The reason that a similar CI system has not been reported in other Wolbachia-infected species remains a matter for speculation, but it is possible that the crossing studies conducted in Cx. pipiens have been more exhaustive than in any other species because of the intensive investigations that have been carried out for clarifying its systematics and studying the inheritance of morphological characters since the 1930s (e.g. Marshall and Staley 1937; Roubaud 1941; Laven 1958; Rozeboom 1958; Laven 1967; Barr 1975; Narang and Seawright 1982; Irving-Bell 1983). As a result, the high variability of CI crossing types was investigated much earlier than the causative agent was identified by Yen and Barr in 1971. Comparatively little work on the variability of the effects of Wolbachia infection has been conducted in most arthropods, except in Drosophila species, such as D. simulans, where five distinct crossing types associated with distinct Wolbachia infections have thus far been identified (for review see Mercot and Charlat 2004). Therefore, the possibility of the existence of variable reproductive phenotypes in other host species remains to be examined.

In conclusion, the use of multi-locus typing combining Wolbachia and mitochondrial markers highlights the processes underlying the evolutionary dynamics of wPip infections. The diversification inside the wPip clade shows that a considerable amount of Wolbachia diversity can be generated within a single host species in a short period of time. Further investigations should examine the roles of recombination and MGE in the adaptive capacities of Wolbachia. In particular, this could explain rapid changes of interactions between Wolbachia and their hosts (Weeks et al. 2007; Echaubard et al. 2010) and play a key role in the evolution of phenotypes induced by Wolbachia. Finally, an important question now is to determine whether the Cx. pipiens - Wolbachia association is a unique case or, rather, a representative example.
Acknowledgments

We are grateful to Emmanuel Douzery for his advice for analyses, Jennifer Bernard, Arnaud Berthomieu, Patrick Makoundou and Sandra Unal for technical assistance, and two anonymous referees for their valuable comments to the manuscript. We acknowledge financial support from the PIR “Maladies Infectieuses et Environnement” from the CNRS. All sequence data were obtained on the Environmental Genomic Platform of the IFR Montpellier-Environnement-Biodiversité. This is contribution 2011.nx of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554 CNRS – Université Montpellier 2).

Literature Cited


<table>
<thead>
<tr>
<th>Mosquito line</th>
<th>Abbreviation</th>
<th>Wolbachia strain</th>
<th>Culex pipiens subspecies</th>
<th>Origin</th>
<th>Year of collection</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pel</td>
<td>Pel</td>
<td>wPip(Pel)</td>
<td>quinquefasciatus</td>
<td>Sri Lanka</td>
<td>1984</td>
<td>(Klasson et al. 2008)</td>
</tr>
<tr>
<td>JHB</td>
<td>JHB</td>
<td>wPip(JHB)</td>
<td>quinquefasciatus</td>
<td>South Africa</td>
<td>2001</td>
<td>(Salzberg et al. 2009)</td>
</tr>
<tr>
<td>Manille-A</td>
<td>Ma-A</td>
<td>wPip(Ma-A)</td>
<td>quinquefasciatus</td>
<td>Philippines</td>
<td>2003</td>
<td>(Duron et al. 2006)</td>
</tr>
<tr>
<td>Manille-B</td>
<td>Ma-B</td>
<td>wPip(Ma-B)</td>
<td>quinquefasciatus</td>
<td>Philippines</td>
<td>2003</td>
<td>(Duron et al. 2006)</td>
</tr>
<tr>
<td>Kara-C</td>
<td>Ka-C</td>
<td>wPip(Ka-C)</td>
<td>quinquefasciatus</td>
<td>China</td>
<td>2003</td>
<td>(Duron, Fort, and Weill 2006)</td>
</tr>
<tr>
<td>MaClo</td>
<td>Mc</td>
<td>wPip(Mc)</td>
<td>quinquefasciatus</td>
<td>California</td>
<td>1984</td>
<td>(Duron et al. 2005)</td>
</tr>
<tr>
<td>Slab</td>
<td>Sl</td>
<td>wPip(Sl)</td>
<td>quinquefasciatus</td>
<td>California</td>
<td>1950</td>
<td>(Duron et al. 2005)</td>
</tr>
<tr>
<td>Cotonou-A</td>
<td>Cot-A</td>
<td>wPip(Cot-A)</td>
<td>quinquefasciatus</td>
<td>Benin</td>
<td>2005</td>
<td>This study</td>
</tr>
<tr>
<td>Cotonou-B</td>
<td>Cot-B</td>
<td>wPip(Cot-B)</td>
<td>quinquefasciatus</td>
<td>Benin</td>
<td>2005</td>
<td>This study</td>
</tr>
<tr>
<td>Australie</td>
<td>Au</td>
<td>wPip(Au)</td>
<td>hybrid (quinquefasciatus/pipiens)</td>
<td>Australia</td>
<td>2004</td>
<td>(Duron, Fort, and Weill 2006)</td>
</tr>
<tr>
<td>El Palmar-A</td>
<td>Ep-A</td>
<td>wPip(Ep-A)</td>
<td>pipiens</td>
<td>Spain</td>
<td>2005</td>
<td>(Duron et al. 2007)</td>
</tr>
<tr>
<td>El Palmar-B</td>
<td>Ep-B</td>
<td>wPip(Ep-B)</td>
<td>pipiens</td>
<td>Spain</td>
<td>2005</td>
<td>(Duron et al. 2007)</td>
</tr>
<tr>
<td>LaVar</td>
<td>Lv</td>
<td>wPip(Lv)</td>
<td>pipiens</td>
<td>France</td>
<td>2003</td>
<td>(Duron et al. 2005)</td>
</tr>
<tr>
<td>Bifa-A</td>
<td>Bf-A</td>
<td>wPip(Bf-A)</td>
<td>pipiens</td>
<td>France</td>
<td>2002</td>
<td>(Duron et al. 2006)</td>
</tr>
<tr>
<td>Bifa-B</td>
<td>Bf-B</td>
<td>wPip(Bf-B)</td>
<td>pipiens</td>
<td>France</td>
<td>2002</td>
<td>(Duron et al. 2006)</td>
</tr>
<tr>
<td>Kol</td>
<td>Ko</td>
<td>wPip(Ko)</td>
<td>pipiens</td>
<td>Crete</td>
<td>2002</td>
<td>(Duron, Fort, and Weill 2006)</td>
</tr>
<tr>
<td>Keo-A</td>
<td>Ke-A</td>
<td>wPip(Ke-A)</td>
<td>pipiens</td>
<td>Cyprus</td>
<td>2003</td>
<td>(Duron et al. 2006)</td>
</tr>
<tr>
<td>Keo-B</td>
<td>Ke-B</td>
<td>wPip(Ke-B)</td>
<td>pipiens</td>
<td>Cyprus</td>
<td>2003</td>
<td>(Duron et al. 2006)</td>
</tr>
<tr>
<td>Tunis</td>
<td>Tn</td>
<td>wPip(Tn)</td>
<td>pipiens</td>
<td>Tunisia</td>
<td>1992</td>
<td>(Duron et al. 2005)</td>
</tr>
<tr>
<td>Istanbul</td>
<td>Is</td>
<td>wPip(Is)</td>
<td>pipiens</td>
<td>Turkey</td>
<td>2003</td>
<td>(Duron et al. 2005)</td>
</tr>
</tbody>
</table>

**Table 1.** Description of *Culex pipiens* lines and wPip strains.
<table>
<thead>
<tr>
<th>Strain</th>
<th>Gene</th>
<th>Haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>w Pip(Pel)</td>
<td>a a a a a a a</td>
<td>A</td>
</tr>
<tr>
<td>w Pip(Cot-A)</td>
<td>a a a a a a a</td>
<td>B</td>
</tr>
<tr>
<td>w Pip(Cot-B)</td>
<td>a a a a a a a</td>
<td>C</td>
</tr>
<tr>
<td>w Pip(Ko)</td>
<td>a a a a a a a</td>
<td>D</td>
</tr>
<tr>
<td>w Pip(Tn)</td>
<td>a a a a a a a</td>
<td>E</td>
</tr>
<tr>
<td>w Pip(Ma-B)</td>
<td>a a a a a a a</td>
<td>F</td>
</tr>
<tr>
<td>w Pip(JHB)</td>
<td>a a a a a - a</td>
<td>G</td>
</tr>
<tr>
<td>w Pip(Ep-A)</td>
<td>a a a a d a a</td>
<td>H</td>
</tr>
<tr>
<td>w Pip(Ep-B)</td>
<td>a a a a d a a</td>
<td>I</td>
</tr>
<tr>
<td>w Pip(Bf-A)</td>
<td>a a a a a a b</td>
<td>J</td>
</tr>
<tr>
<td>w Pip(Lv)</td>
<td>b e c a b b a</td>
<td>K</td>
</tr>
<tr>
<td>w Pip(Au)</td>
<td>d e c a b f a</td>
<td>L</td>
</tr>
<tr>
<td>w Pip(Ke-A)</td>
<td>c e c a d e a</td>
<td>M</td>
</tr>
<tr>
<td>w Pip(Ke-B)</td>
<td>c e c a e e a</td>
<td>N</td>
</tr>
<tr>
<td>w Pip(Sl)</td>
<td>e b b b b c b</td>
<td></td>
</tr>
<tr>
<td>w Pip(Bf-B)</td>
<td>e b b c b c b</td>
<td></td>
</tr>
<tr>
<td>w Pip(Mc)</td>
<td>b b b b c b</td>
<td></td>
</tr>
<tr>
<td>w Pip(Is)</td>
<td>c c d a c d a</td>
<td></td>
</tr>
<tr>
<td>w Pip(Ka-C)</td>
<td>f d e a f g a</td>
<td></td>
</tr>
<tr>
<td>w Pip(Ma-A)</td>
<td>f d e a g g a</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Allelic profiles of the seven polymorphic wPip genes in 20 wPip strains. Letters A to N represent the 14 wPip haplotypes. Dash indicates a gene deletion.
<table>
<thead>
<tr>
<th>Taxon (subfamily, species)</th>
<th><strong>cytb</strong></th>
<th><strong>ND4</strong></th>
<th><strong>COI</strong></th>
<th><strong>COII</strong></th>
<th><strong>Wolbachia infection (reference)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Culicinae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Culex pipiens</em></td>
<td>0.0021 (19)</td>
<td>0.0004 (14)</td>
<td>0.0000 (24)</td>
<td>0.0007 (8)</td>
<td>yes (Hertig and Wolbach 1924)</td>
</tr>
<tr>
<td><em>Cx. tarsalis</em></td>
<td>_</td>
<td>0.0116 (64)</td>
<td>_</td>
<td>_</td>
<td>no (Rasgon and Scott 2004)</td>
</tr>
<tr>
<td><em>Culex sp.</em></td>
<td>_</td>
<td>0.0090 (10)</td>
<td>_</td>
<td>_</td>
<td>no (Rasgon, Cornel, and Scott 2006)</td>
</tr>
<tr>
<td><em>Aedes aegypti</em></td>
<td>0.0094 (16)</td>
<td>0.0202 (46)</td>
<td>_</td>
<td>_</td>
<td>no (Kittayapong et al. 2000)</td>
</tr>
<tr>
<td><em>Ae. albopictus</em></td>
<td>0.0043 (14)</td>
<td>_</td>
<td>0.0039 (23)</td>
<td>_</td>
<td>yes (O'Neill et al. 1992)</td>
</tr>
<tr>
<td><em>Ae. caspius</em></td>
<td>_</td>
<td>_</td>
<td>0.0094 (7)</td>
<td>0.0063 (21)</td>
<td>no (Ricci et al. 2002)</td>
</tr>
<tr>
<td><em>Ae. vexans</em></td>
<td>_</td>
<td>_</td>
<td>0.0185 (7)</td>
<td>0.0084 (7)</td>
<td>no (Kittayapong et al. 2000; Ricci et al. 2002)</td>
</tr>
<tr>
<td><strong>Anophelinae</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Anopheles aconitus</em></td>
<td>_</td>
<td>_</td>
<td>0.0053 (13)</td>
<td>0.0066 (35)</td>
<td>no (Kittayapong et al. 2000)</td>
</tr>
<tr>
<td><em>An. funestus</em></td>
<td>0.0066 (11)</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>no (Ricci et al. 2002)</td>
</tr>
<tr>
<td><em>An. gambiae</em></td>
<td>_</td>
<td>_</td>
<td>0.0053 (48)</td>
<td>_</td>
<td>no (Ricci et al. 2002)</td>
</tr>
<tr>
<td><em>An. maculipennis</em></td>
<td>_</td>
<td>_</td>
<td>0.0063 (62)</td>
<td>_</td>
<td>no (Ricci et al. 2002)</td>
</tr>
</tbody>
</table>

**Table 3.** Nucleotide diversity ($\pi$) of mitochondrial genes in eleven mosquito species (Culicidae).
Figure legends

Figure 1. Map of the wPip(Pel) genome showing the position of the thirteen genes examined. Number in brackets (1 to 4) indicates identical copies of a given gene located in different positions along the chromosome. Black boxes indicate the locations of prophage regions; the MLST and wsp genes are reported in grey.

Figure 2. Phylogenetic tree of wPip strains obtained from concatenated data set (MutL, ank2, pk1, pk2, GP12, GP15 and RepA sequences) by Bayesian analysis. Posterior probabilities obtained are shown at major nodes. The scale bar is in units of substitutions/site.

Figure 3. Network analysis obtained from concatenated data set (MutL, ank2, pk1, pk2, GP12, GP15 and RepA sequences) using the Neighbor-net method. Each edge (or a set of parallel edges) corresponds to a split in the data set and has length equal to the weight of the split. Incompatible splits, produced by recombination are represented by boxes in the network. Only bootstrap values for major grouping are indicated. The five wPip groups (highlighted) are connected by multiple pathways resulting from recombination between Wolbachia genomes.

Figure 4. Distribution of wPip haplotypes and wPip groups in Culex pipiens populations. Letters and symbols represent the wPip haplotypes and wPip groups, respectively. * wPip haplotypes recently described by Atyame et al. (Atyame et al. in press).
Figure 5. Comparisons between phylogeny of *Culex pipiens* mitochondria and phylogeny of the wPip strains. **A**, Mitochondrial phylogeny constructed using Bayesian inferences based on *ND2*, *ND5* and *cytb* concatenated sequences. Names on branches indicate the mtDNA haplotypes (*piI* to *piXIV*). **B**, wPip phylogeny obtained from concatenated data set (*MutL*, *ank2*, *pk1*, *pk2*, *GP12*, *GP15* and *RepA* sequences). The five wPip groups are reported. The scale bar is in units of substitutions/site.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5