**Culex pipiens** crossing type diversity is governed by an amplified and polymorphic operon of *Wolbachia*

Manon Bonneau\(^1\), Celestine Atyame\(^1,2\), Marwa Beji\(^3\), Fabienne Justy\(^1\), Martin Cohen-Gonsaud\(^4\), Mathieu Sicard\(^1\) & Mylène Weill\(^1\)

*Culex pipiens* mosquitoes are infected with *Wolbachia* (wPip) that cause an important diversity of cytoplasmic incompatibilities (CIs). Functional transgenic studies have implicated the cidA-cidB operon from wPip and its homolog in wMel in CI between infected *Drosophila* males and uninfected females. However, the genetic basis of the CI diversity induced by different *Wolbachia* strains was unknown. We show here that the remarkable diversity of CI in the *C. pipiens* complex is due to the presence, in all tested wPip genomes, of several copies of the cidA-cidB operon, which undergoes diversification through recombination events. In 183 isofemale lines of *C. pipiens* collected worldwide, specific variations of the cidA-cidB gene repertoires are found to match crossing types. The diversification of cidA-cidB is consistent with the hypothesis of a toxin–antitoxin system in which the gene cidB co-diversifies with the gene cidA, particularly in putative domains of reciprocal interactions.

---

\(^1\)Institut des Sciences de l’Evolution de Montpellier (ISEM), UMR CNRS-IRD-EPHE-Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier, France.  
\(^2\)Processus Infectieux en Milieu Insulaire Tropical (PIMIT), UMR CNRS-INSERM-IRD-Université de La Réunion, Sainte-Cloïlde, Ile de La Réunion, 97490, France.  
\(^3\)Institut Pasteur Tunis, Laboratory of Epidemiology and Veterinary Microbiology, University of Tunis El Manar, 1068 Tunis, Tunisia.  
\(^4\)Centre de Biochimie Structurale (CBS), UMR CNRS-INSERM-Université de Montpellier, 29 rue de Navacelles, 34090 Montpellier, France. Mathieu Sicard and Mylène Weill contributed equally to this work. Correspondence and requests for materials should be addressed to M.S. (email: mathieu.sicard@umontpellier.fr) or to M.W. (email: mylene.weill@umontpellier.fr)
The most common way by which Wolbachia bacteria spread within insect populations is cytoplasmic incompatibility (CI), which results in the early death of embryos when males infected with a given Wolbachia strain mate with uninfected females or females infected with an incompatible strain. The most popular model to date for conceptualizing CI involves the secretion, by Wolbachia, of a “modification” factor (mod, or toxin) in the sperm that impairs early embryogenesis, and a Wolbachia “rescue” factor (resc, or antitoxin) produced in the oocyte, which allow for a viable, diploid zygote to develop if the cross is compatible. Previous studies searching for Wolbachia effectors involved in CI by genomic comparison between mutualistic and manipulative Wolbachia revealed a higher number of genes with ankyrin repeats in the latter. Ankyrin repeat genes were thus considered as potential mod and resc candidates due to their role in protein–protein interactions and cellular cycle regulation. However, to date no correlation between the distributions of any ankyrin repeat genes and CI was demonstrated. The first convincing candidate gene for involvement in CI encodes a protein called CidA (WP0282), secreted by Wolbachia into Culex pipiens sperm. The cidA gene is part of an operon also containing a second gene, cidB (WP0283). CidA and CidB proteins directly interact with each other as demonstrated by pull-down experiments. The CidA and cidB genes (named cifA and cifB in wMel) are associated with the prophage WO modules, and paralogs of these genes have been found in all published genomes from Wolbachia strains known to induce CI in insects. The co-expression of cidA/cidB (or cifA/cifB) transgenes in Drosophila melanogaster reproduces the disturbance of the first embryonic division hallmark of CI, leading to embryo death. In this context, disruption of the deubiquitylating (DUB) domain of CidB restores normal embryogenesis, demonstrating the contribution of this domain to the “mod” factor function. The role of CidA in the mod-resc CI system is much more debated. Data obtained on yeast suggest that CidA may rescue CidB, but CifA has also been described as a CifB elicitor in D. melanogaster.

cidA-cidB operon is involved in the CI induced by infected males in the progenies of uninfected females, but the genetic basis of CI diversity (i.e., compatibility, uni or bidirectional incompatibility) between hosts infected with different Wolbachia strains has yet to be investigated. We addressed this question in C. pipiens, a powerful model in which hundreds of crosses between lines sampled worldwide have revealed unprecedented CI diversity. This unique CI diversity is solely governed by wPip diversification in this species complex since no other manipulative endosymbionts are present and no host genetic background influence has been yet demonstrated. All C. pipiens individuals are infected with Wolbachia, which recently diverged into five distinct phylogenetic groups (wPipI to wPipV). The wPip strains of the same group generally ensure compatibility between their hosts, whereas intergroup crosses are more likely to be incompatible. An analysis of multiple crosses concluded that each Wolbachia (wPip) genome must contain several mod and resc factors to account for the high diversity of CI in C. pipiens. These multiple mod and resc factors could theoretically be encoded by different copies (i.e., variants) of same mod/resc genes or different mod and resc genes within the same wPip genome. It therefore seems likely that (i) different copies (i.e., variants) of cidA-cidB, (ii) paralogs of these genes resulting from ancient duplication events, or (iii) totally unrelated genes, could be involved in CI diversity in C. pipiens.

We thus investigated the contribution of polymorphism of the cidA-cidB operon and its cinA-cinB paralog with predicted nuclease activity to CI diversity in C. pipiens with the major hypothesis that if these genes encode for CI diversity determinants, they should be different between C. pipiens lines with different crossing types. We show that all tested wPip genomes contain several variants of the cidA-cidB operon while it is not the case for cinA-cinB operon. The cidA-cidB repertoires of variants differ between wPip strains from different groups exhibiting different crossing types. In 180 isofemale lines of C. pipiens exclusively infected with wPipIV strains, specific variations of the cidA-cidB gene repertoires are found to match crossing type
variations. These variations occurred only in specific domains for both proteins, which consist exclusively of protein–protein interaction motifs and thus might be involved in the physical interaction between CidA and CidB.

Results

Illumina sequencing of several wPip strains. We first used Illumina technology to sequence three wPip genomes from strains inducing different crossing types: two strains from group IV, originating from Algeria and Turkey; and one strain from group I, originating from Tunisia. The mapping of the reads onto the reference genome wPip_Pel19 revealed an abnormally high coverage of the WOPip1 region, which includes the cidA and cidB genes. The coverage for this region was, on average, three times greater than that for 14 single-copy housekeeping genes (Supplementary Fig. 1). Moreover, the cidA-cidB operon was highly variable between and within the three isofemale lines sequenced (Fig. 1a). Contrary to the assembly of wPip_Pel

**Table 1** Crossing types of the four mosquito lines each infected with Wolbachia wPip from different groups (I, II, III, and IV)

|                           | ♂                      |    |    |    | ♂    | Crossing type |
|---------------------------|------------------------|----|----|----|------|---------------|
|                           | Tunis                  |    |    |    | Lavar| Maclo | Istanbul      |
| ♂                         | Tunis (wPipl)          | +  | -  | +  | -    | A    |
|                           | Lavar (wPiplI)         | +  | +  | +  | -    | B    |
|                           | Maclo (wPiplII)        | +  | +  | +  | +    | C    |
|                           | Istanbul (wPiplIV)      | -  | -  | -  | +    | D    |

* + means that the cross is compatible (many larvae hatch from the egg rafts) and – indicates an incompatible cross (no larvae hatch from the egg rafts). Each isofemale line has a different crossing type, as deduced from reciprocal crosses between the four lines.
**Fig. 3** The repertoire of CidA variants in the different wPip groups. a Localization of polymorphic zones within CidA proteins. Schematic representation of the CidA protein with polymorphic zones between CidA variants of the four wPip groups highlighted in black. b The polymorphism between CidA variants is distributed as blocks of variable amino acids. Protein sequences alignment of the CidA variants found in the four Wolbachia strains Tunis, Lavar, Maclo, and Istanbul. The first sequence is used as a reference to determine the polymorphic region. For more clarity, only polymorphic positions of the alignment are represented and amino-acid positions are not contiguous. When more than two contiguous amino acids were variable the “-” symbol was used between the first and the last variable position of the zone. Colors show polymorphic blocks of residues present in variants regardless of their phylogenetic wPip group (I-IV). However, no variant (i.e., complete CidA sequence) is common to all strains Tunis, Lavar, Maclo, and Istanbul. The variable domains, displayed no polymorphism within and between wPip strains (Fig. 1b), ruling out their role in CI diversity.10 These findings suggest a previously unsuspected diversity of cidA-cidB copies within wPip genomes, resulting from genome amplifications and divergence that may account for CI diversification in *C. pipiens*.

**High cidA-cidB diversity between wPip groups.** We explored the association between *cidA-cidB* genetic variants and CI diversity in *C. pipiens* further, by investigating the variability of these genes in four isofemale lines differing by their geographical origin, each infected with a wPip strain from a different group (I-IV) and exhibiting a different crossing type (Table 1). Each of these isofemale lines was founded with one initial egg-raft from a single female. Cloning and sequencing revealed a large number of *cidA-cidB* variants in each of these wPip strains. The variable domains that differ between variants were not contiguous but restricted to some specific zones of the proteins (Figs. 3a and 4a). Within a single line there were up to six different variants of *cidA* (Fig. 3b), and up to four different variants of *cidB* (Fig. 4b). Some specificity in the *cidA-cidB* repertoire was also detected, with none of the variants shared between strains from different wPip groups (Figs. 3b and 4b). Network analyses of *cidA* and *cidB* suggested that the differences between copies stemmed mostly from block rearrangements within and/or between wPip genomes (Figs. 3c and 4c). These rearrangements were never observed in the region corresponding to the CidB catalytic domain, but in other regions of the *CidA* and CidB proteins (Figs. 3 and 4; Supplementary Fig. 2 and 3). The stability of the *cidA-cidB* repertoire over a decade (between 2006 and 2017) in the four polymorphic copies of *cidA* and *cidB* genes are present in multiple copies in wPip genome. The copy number ratio between *cidA-cidB* genes and *wsp* was always between 4 and 6 (Table 2). We also found that all copies of these genes were transcribed. Indeed, in the wPip strain from Istanbul the transcripts of the six polymorphic copies (i.e., variants) of *cidA* and the four polymorphic copies of *cidB* were detected in RNA extract after reverse transcription-PCR, cloning, and sequencing.

**cidA-cidB repertoire variations are linked to crossing type.** We then focused on a simpler situation, by analyzing CI diversity within wPip group IV. We used six isofemale lines with two different crossing types. These lines had the same rescue properties but two different mods.13 Males from "incompatible" lines sterilize females whenever they host a wPip group I, II, or III strain, whereas males from "compatible" lines always produce progenies with these females (Fig. 5). We compared the *cidA-cidB* repertoires of "compatible" and "incompatible" lines. We found that one CidA variant [CidA_IV(α)] was present in all wPip IV lines, regardless of their crossing type or geographic origin.
Strikingly, *cidA* was shared between compatible and incompatible lines (Fig. 5). Such a ubiquitous distribution is expected for the resc factor mediating compatibility between all isofemale lines hosting *w*PiP strains from the same group. By contrast, no CidB variant was shared between compatible and incompatible lines (Fig. 5). Strikingly, *cidA* and *cidB* repertoire variations were clearly associated with crossing type variations, as *cidA_IV* (a/2) and *cidB_IV* (a/2) variants were found uniquely in “incompatible” lines (Fig. 5). This may suggest that both *cidA_IV* (a/2) and *cidB_IV* (a/2) are required for the incompatibility phenotype (Fig. 5).

We further investigated the distribution of these two specific *cidA/cidB* variants [*(cidA_IV* (a/2) and *cidB_IV* (a/2))] at a larger scale, by performing diagnostic restricted fragment length
polymorphism (RFLP) tests on 180 isofemale lines from 15 populations [13 in Algeria and Tunisia; 1 each in China and Turkey], for which we determined crossing types. A very strong association was established between the presence of both cidA IV (δ) and cidB IV(a/2) in “incompatible” isofemale lines and their absence from “compatible” isofemale lines: all 17 “incompatible” lines had both these variants, whereas 147 of the 163 “compatible” lines had neither of them ($\chi^2 = 78$, df = 1, $p < 2.2e-16$). Among the 16 “compatible” isofemale that appeared discordant: (i) 8 were truly discordant as they display both cidA IV(δ) and cidB IV(a/2) while being compatible but (ii) 8 were only discordant for cidA as they displayed cidA IV(δ) but lacked cidB IV(a/2) (as confirmed by cloning and sequencing; Supplementary Data 1). To summarize the data: the absence of these variants is always associated with “compatible” crossing type while its presence is mostly associated with “incompatible” crossing type with few discrepancies.

Putative ankyrin interaction domains between CidA and CidB.

The analysis of CidA and CidB polymorphic regions revealed that the sequence variations associated with crossing type variations are restricted to about 30 amino-acid positions, for both proteins (Fig. 6a). Extensive fold-recognition analysis of the CidA and CidB proteins were performed using the @tome2 in-house server21. Except for the DUB protease domain, no tridimensional model could be firmly obtained. Nevertheless, both CidA and CidB exhibit sequence identity homology (under 20%) with protein–protein interaction domains with repeated structure. Careful examination of hydrophobic patterns suggests internal helical structure repetitions as found in ankyrin or HEAT repeats22. The size of ~33 amino acids is the universal expected size for tandem ankyrin repeat domains6. However, due to the lack of available homolog sequences in databases, it is difficult to determine the helical structure repetition boundaries in both N- and C-terminal part of the protein domains. As the protein variations matching with crossing type variations occur in putative ankyrin domains, we hypothesize that CidA and CidB co-diversify in these domains because they are involved in their reciprocal interactions (Fig. 6).

Discussion

In the C. pipiens mosquitoes, the genetic architecture of the Wolbachia cidA-cidB operon is characterized by variable multicopies associated to the unique CI diversification observed in this species. These different copies of cidA and cidB likely result from genic amplification followed by diversification since we revealed a high diversity within each wPip strain. As these genes are associated with the prophage WO modules11, the number of cidA-cidB copies assessed by q-PCR may not reflect the exact number of cidA-cidB copies in a single wPip genome since viral particle multiplication resulting from lytic phage activity can influence the accuracy of this quantification23. However, the diversity of cidA-cidB sequences in a single mosquito, the geographical and temporal stability of sequences found in wPipIV group clearly point out that several copies of the operon do co-exist within the same wPip genome. As cidA and cidB are amplified in all wPip genomes from each C. pipiens individual we sequenced so far, the presence of only one copy of the cidA/cidB operon in wPip_Pel seems to result from an error during assembly likely caused by too many repeats. The reason(s) why such amplification and diversification of cidA/cidB has evolved in wPip is still an open
but are “incompatible”, see Fig. 5 for definition) is highlighted in red. The corresponding variations in protein sequences are reported below the schematic representation of the full-length proteins, with the variable residues highlighted in red. Other sequence variations not predicted to be involved in CidA variations are shown in black. In CidA_IV and CidB_IV variants sequences detected in all sequenced wPip strains from group IV. CidA_IV variants display two regions of polymorphism resulting from recombination revealed by block colors: the upstream one from 118 to 152 aa and the downstream one from 336 to 401 aa. Four possible sequences (noted a, b, γ, or δ) were found in the upstream polymorphic region followed by one of the two sequences in the downstream region (noted 1 or 2). Only the upstream polymorphic region of CidA is associated with crossing type variations. CidB_IV variants display two regions of polymorphism resulting from recombination: the upstream one from 242 to 278 aa and the second one from 450 to 481 aa. Two possible sequences (noted a and b) were found in the upstream polymorphic region followed by one of the three possible sequences in the downstream polymorphic region (noted 1, 2, and 3). Only the downstream region is associated with crossing type variations.

question. However, this diversity likely reflects high levels of intra- or intergenome recombination, which may be favored by multiple infections with different wPip and/or promoted by lytic phage activity in C. pipiens²⁴.

In a more global context, our results provide insights into the respective roles of CidA and CidB in the mod/resc model. Indeed, the association between cidA_IV(8)-cidB_IV(a/2) presence/absence and incompatible/compatible crossing types, together with the results of functional studies¹⁰,¹¹, demonstrate that CidB is involved in the mod function responsible for the variation in crossing types. However, some discrepancies were found in few isofemale lines, which harbor cidA_IV(δ) and/or cidB_IV(a/2) but are “compatible”. These discrepancies suggest that either (i) our strategy, involving larval field sampling, the establishment of isofemale lines and crossing experiments, generated a few errors or (ii) other genetic or epigenetic mechanisms may disrupt the relationship between genomic potential and the resulting crossing type.

The role of CidA is still unclear. The association of cidA_IV(δ) with variation in crossing types does not necessarily implicate CidA in the mod function. Indeed, as CidA and CidB interact with each other⁹, their co-diversification could result in specific CidA variations whenever involved in the mod (i.e., toxin) or in resc (i.e., antitoxin) function. Several lines of evidence suggest that CidA could be the antitoxin of CidB and that the cidA-cidB operon encodes a classic toxin–antitoxin system: (i) the simple two-gene structure of the operon, with the putative antitoxin gene located upstream from the putative toxin gene⁹; (ii) the overexpression of the antitoxin gene relative to the toxin gene¹ⁱ; (iii) the ability of the two proteins to form a complex¹⁰; and (iv) the required co-expression of cidA and cidB for the production of live transgenic D. melanogaster and Saccharomyces cerevisiae¹⁰. Additional argument of CidA being an antitoxin is provided by the presence of cidA_IV(α), a ubiquitous CidA variant in wPip strains within the same group wPipIV, potentially accounting for their hosts’ compatibility. Moreover, if CidA is indeed the antitoxin, then a new toxin (i.e., CidB) produced by block recombination between variants would be positively selected only if a compatible CidA, capable of interaction with this new toxin, was already present in the same wPip genome to prevent its producer from being killed²¹–²². This is coherent with our results since we found more cidA variants than cidB variants in the same isofemale line (except for Tunis, which has as many cidA as cidB).

Most toxin–antitoxin systems require an interaction between both proteins²³. Pull-down experiments have demonstrated that CidA can interact with CidB¹⁰. Our analysis of CidA-CidB repertoire in wPipIV group showed that genetic variations linked to crossing type variations occurred only in specific domains of about ~30 amino acids in length, for both proteins (Fig. 6a). These domains consist exclusively of protein–protein interaction motifs (ankyrin or HEAT repeats) and thus might be involved in the reciprocal interaction between CidA and CidB. The variability of these domains that co-diversify may account for specific interaction between one variant of CidA and one variant of CidB.
Fig. 7 cidA_IV gene polymorphic regions and PCR-RFLP tests for specific variants. a Schematic representation of the architecture of cidA_IV polymorphism. The black line represents the cidA_IV sequence with a focus on the two polymorphic regions. Non-polymorphic regions were shortened and represented as a dashed line. Numbers in red under the line represent nucleotide positions, and the numbers in blue indicate amino-acid positions. The overlapping oligonucleotides used for PCR amplification are represented by arrows numbered as in Supplementary Table 2. Both polymorphic regions were studied, but only the upstream matched the compatibility profile; its four different sequences (α, β, γ, and δ) were followed by one of the two possible sequences (1 or 2) in the downstream polymorphic region. A different color code was used for α (purple), β (light blue), γ (light green), and δ (yellow) sequences in the upstream part of the cidA gene. b The repertoire of cidA_IV variants is different in “compatible” and “incompatible” lines. cidA_IV(δ) is present only in lines with “incompatible” crossing type. The names of the C. pipiens lines used to set-up the PCR-RFLP (c) test are highlighted in red. c PCR-RFLP tests for distinguishing between cidA_IV variants on the basis of the upstream polymorphic region. A 778 bp fragment was amplified with primers 1/13. Double digestion with Apol and Hpy188I distinguished between cidA_IV(α) (six fragments: 471; 122; 57; 53; 51; and 24 bp), cidA_IV(β) and cidA_IV(γ) (six fragments: 441; 122; 83; 57; 51; and 24 bp), and cidA_IV(δ) (five fragments: 524; 122; 57; 51; and 24 bp). Panel to the left of the electrophoresis gel: PCR-RFLP on DNA from clones; right panel: PCR-RFLP on DNA from the Istanbul, Harash, Ichkeul 13, and Ichkeul 09 lines. On the right of the gel, a schematic representation of the PCR-RFLP profiles of these lines. Digestion bands that are specific of the variants are represented with the color code established in a. Bands that are not used to discriminate between the variants are represented in black on the bottom of the schematic gel.

Such variability in interactions between CidA and CidB could drive the unrivaled diversity of crossing types in C. pipiens.

Methods

Isofemale line construction. C. pipiens larvae and pupae were collected in the field and reared to adulthood in the laboratory. Females were then fed on blood to lay eggs that served to establish isofemale lines. Each egg raft (containing 100–300 eggs) was individually isolated for hatching, and the Wolbachia strain was transferred by feeding the female a filter with Wolbachia cells. Isofemale lines were established by rearing the offspring resulting from a single egg raft (thus from a single female). We established 162 isofemale lines for this study. We also used 21 isofemale lines from laboratory stocks of various geographic origins (Supplementary Data 1 and Supplementary Table 1). Isofemale lines were reared in 65 dm³ screened cages kept in a single room at 22–25°C, under a 12 h light/12 h dark cycle. Larvae were fed with a mixture of shrimp powder and rabbit pellets, and adults were fed on honey solution.

Crossing type determination. Crossing types were characterized by crossing males (25–50 virgin males) from each of the studied isofemale lines with females (25–50 virgin females) from four reference laboratory isofemale lines hosting different Wolbachia strains [Tunis (wPipI), Lavar (wPipII), Maclo (wPipIII), and Istanbul (wPipIV)] or with females from Tunis only (Supplementary Data 1 and Supplementary Table 1).

Illumina sequencing. The genomes of three Wolbachia strains from three C. pipiens isofemale lines were fully sequenced: Tunis from group wPipII, Harash and Istanbul from group wPipIV. Wolbachia is an intracellular uncultivable bacterium. To obtain satisfactory amounts of Wolbachia DNA, we therefore developed a protocol based on that described by Ellegaard et al.27. This protocol involved Wolbachia enrichment on egg rafts immediately after oviposition, to maximize the ratio of Wolbachia to Culex genomes. The freshly laid egg rafts were washed in bleach to crack their chorion. They were then rinsed in water and homogenized in phosphate-buffered saline (1× PBS). Wolbachia cells were isolated from each egg raft separately and concentrated. Each egg raft was crushed with a sterile pestle and the resulting suspension was centrifuged at 400 × g for 5 min at 4 °C to remove the cell debris. The supernatant was then centrifuged again at 6000 × g for 5 min at 4 °C, to obtain a pellet of Wolbachia cells. These cells were resuspended in 1× PBS and passed through a filter with 5 μm pores (GV5 Filter Technology), and a filter with 2.7 μm pores (Whatman), to remove any remaining particles bigger than Wolbachia. The resulting filtrate was centrifuged at 6900 × g for 15 min at 4 °C to obtain a pellet of Wolbachia cells. Multiple-displacement amplification was carried out directly on the Wolbachia pellet, with the Repl-g- mini kit (Qiagen). For each strain, five amplifications were performed independently, to randomize potential amplification errors, and the amplicons were then pooled for sequencing. Illumina sequencing was performed by the 3 kb Long Jumping Distance Mate-Pair method, on a HiSeq2000 instrument (Eurofins MWG Operon platform in Ebersberg, Germany), generating 2 × 100 bp sequences for each fragment.

Cloning and Sanger sequencing of the cidA and cidB genes. For all experiments, total DNA was extracted with the DNeasy Blood & Tissue Spin-Column kit (Qiagen; bench protocol: animal tissues). For amplification of the full-length cidA gene, we used two pairs of overlapping primers, 1/2 and 3/4 (Fig. 7, Supplementary Table 2). For amplification of the full-length cidB gene, we used four pairs of overlapping primers: 5/6; 7/8; 9/10; and 11/12 (Fig. 8, Supplementary Table 2). We used cidA fragments amplified with primers 1/2 and cidB fragments amplified with primers 7/8 for cloning experiments, because these two regions contain all the polymorphism observed in cidA and cidB variants (Figs. 3 and 4). PCR products were extracted from the electrophoresis gel with the W-tract disposable gel extraction tool (Star Lab) and purified with the QIAquick Gel Extraction Kit (Qiagen). The TOPO TA cloning Kit pCR 2.1-TOPO Vector (Invitrogen) was used for ligation. Before electroporation, the DNA was dialyzed with VSWP membrane...
Fig. 8 cidB IV gene polymorphic regions and RFLP tests for specific variants. a) Schematic representation of the polymorphism architecture of cidB IV. The black line represents the cidB IV sequence with a focus on the polymorphic region. Non-polymorphic regions were shortened and represented as a dashed line. Numbers in red represent nucleotide positions and those in blue indicate amino-acid positions. The overlapping oligonucleotides used for amplification are represented by arrows numbered as in Supplementary Table 2. Both polymorphic regions are indicated but only the downstream (on the left, base pairs 1347–1428, amino acids 450–481) matches the compatibility profiles. The two sequences of the upstream polymorphic zone (a and b) are followed by one of the three possible sequences (1, 2, and 3) in the downstream polymorphic region. A different color code was used for 1 (orange), 2 (blue), and 3 (green) different sequences in the downstream part of the cidB gene. b) The cidB IV variant repertoire differs between “compatible” and “incompatible” lines. cidB IV(a/2) is present only in lines with “incompatible” crossing type. The names of the C. pipiens lines used to set up the PCR-RFLP (c) test are highlighted in red. c) PCR-RFLP tests for distinguishing between cidB IV variants on the basis of the second polymorphic region. A 1267–1276 bp fragment was amplified with primers 7/8. Double digestion with BanI and TaqI distinguished between cidB IV(1) (three fragments: 892; 239; and 145 bp), cidB IV(2) (two fragments: 1028 and 239 bp), and cidB IV(3) (three fragments: 861; 239; and 167 bp). Panel to the left of the electrophoresis gel: PCR-RFLP on DNA from clones; right panel: PCR-RFLP on DNA from lines. The panel on the right shows schematic representations of the PCR-RFLP profiles for these lines. Digestion bands that are specific of the variants are represented with the color code established in the a. Bands that are not used in to discriminate between the variants are represented in black on the bottom of the schematic gel.

Gene copy number quantification. The copy numbers of the cidA and cidB genes were estimated per Wolbachia genome, for four laboratory lines, on male and female adults and a pool of larvae at the earliest stage, by real-time qPCR with a Roche Light Cycler. Three PCRs were performed, in triplicate, on each sample, one for the Wolbachia wsp locus present as single copy per genome (primer pair 18/19), one for cidA (primer pair 14/15), and one for cidB (primer pair 16/17). These primer pairs specifically bind to a region conserved in all wsp groups for each gene (Supplementary Table 2). The ratio between cidA or cidB and wsp signals was used to estimate the relative number of copies of cidA or cidB per Wolbachia genome.

cidA-cidB expressed repertoire. Fresh mosquito samples from adult females and males and first (L1) and last (L4) instar larvae were used for RNA extraction with Trizol (Life Technologies). The RNA obtained was treated with DNase with the TURBO DNA-free Kit (Life Technologies), in accordance with the manufacturer’s instructions. We reverse-transcribed 2 µg of RNA with SuperScript III Reverse Transcriptase Kit in triplicate. Three PCRs were performed, in triplicate, on each sample, one for the Wolbachia wsp locus present as single copy per genome (primer pair 18/19), one for cidA (primer pair 14/15), and one for cidB (primer pair 16/17). These primer pairs specifically bind to a region conserved in all wsp groups for each gene (Supplementary Table 2). The ratio between cidA or cidB and wsp signals was used to estimate the relative number of copies of cidA or cidB per Wolbachia genome.
Statistical analyses. All statistical analyses were performed with R version 3.0.2 software. We compared the proportions of “compatible” and “incompatible” isolates harboring Wolbachia Wf6 and cidB IV(a2) in a two-sample test for equality of proportions with continuity correction based on $\chi^2$.

Mapping of illumina reads on the reference genome. Illumina Long Jumping Distance sequencing data were trimmed with Trimmmomatic version 0.36. Paired reads and single reads for each line were compiled into Fasta files and mapped onto the genome with wPip_Pel, with bwa mem and default parameters. Mappings were visualized with IGV version 2.3.68.

For each Wolbachia strain, the coverages of WP0282 (cidA), WP0283 (cidB), WP0294 (cinA), WP0295 (cinB), and MLST genes 14.3, 19 were calculated with genomecov from bedtools, with the -d option, to calculate “per-base” genome coverage. Several adjustments were made: (i) coverage was normalized by dividing the mean coverage on the whole reference genome of the strain concerned, to eliminate bias due to the higher coverage for strains for which there were larger numbers of reads in the Illumina data set; (ii) polymorphic regions of cidA and cidB were removed, due to the considerable variability of coverage in these regions; and (iii) the region of cidB from position 2461 to the end of the gene was removed because this region is identically duplicated in the wPip_Pel genome, artificially increasing its coverage.

Data availability. The Illumina raw reads data of the three wPip genomes have been deposited in the Sequence Read Archive with Bioproject PRJNA393398 and BioSample SAMN07315580 (wPip Tunis), SAMN07327402 (wPipLV Harash), and SAMN07327406 (wPipLV Istanbul). Nuchistidic and proteic sequences of cidA-cidB variants have been deposited in GenBank and accession numbers are available in the Supplementary Table 3. The authors declare that all other data supporting the findings of this study are available within the article and its Supplementary Information files, or are available from the authors upon request.

Received: 17 August 2017 Accepted: 21 December 2017
Published online: 22 January 2018

References
1. Yen, J. H. & Barr, A. R. New hypothesis of the cause of cytoplasmic incompatibility in Culex pipiens. Nature 232, 657–658 (1971).
2. Poinsoet, D., Charlat, S. & Merçot, H. On the mechanism of Wolbachia-induced cytoplasmic incompatibility: confronting the models with the facts. BioEssays 25, 259–265 (2003).
3. Hurst, L. D. The evolution of cytoplasmic incompatibility or when spite can be successful. J. Theor. Biol. 148, 269–277 (1991).
4. Wilster, H. Biology of Wolbachia. Annu. Rev. Entomol. 42, 587–609 (1997).
5. Foster, J. et al. The Wolbachia genome of Broughtula malayai endosymbiont evolution within a human pathogenic nematode. PLoS Biol. 3, e599–e614 (2005).
6. Song, X. et al. The diversity and evolution of Wolbachia ankyrin repeat domain proteins. PLoS ONE 8, e53990 (2013).
7. Sedgwick, S. G. & Smerdon, S. J. The ankyrin repeat: a diversity of interaction on a common structural framework. Trends Biochem. Sci. 24, 311–316 (1999).
8. Duron, O. et al. Variability and expression of ankyrin domain genes in Wolbachia variants infecting the mosquito Culex pipiens. J. Bacteriol. 189, 4442–4448 (2007).
9. Beckmann, J. F. & Fallon, A. M. Detection of the Wolbachia protein WP0282 in mosquito spermaticca: Implications for cytoplasmic incompatibility. Insect Biochem. Mol. Biol. 43, 867–878 (2013).
10. Beckmann, J. F., Ronau, J. A., Hochstrasser, M., Tillett, D. & Ginalska, K. A. Wolbachia debiquitylating enzyme induces cytoplasmic incompatibility. Nat. Microbiol. 2, 17007 (2017).
11. Legape, D. P. et al. Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. Nature 543, 243–247 (2017).
12. Sutton, E. R., Harris, S. R., Parkhill, J. & Sinkins, S. P. Comparative genome analysis of Wolbachia strain wAu. BMC Genomics 15, 928 (2014).
13. Atayem, C. M. et al. Wolbachia divergence and the evolution of cytoplasmic incompatibility in Culex pipiens. PLoS ONE 9, 21–26 (2014).
14. Laven, H., Wright, J. W. & Pal, R. in Genetics of Insect Vectors of Disease (eds Wright, J. W. & Pal, R.) 251–275 (Elsevier, Amsterdam, 1967).
15. Atayem, C. M. et al. Multiple Wolbachia determinants control the evolution of cytoplasmic incompatibilities in Culex pipiens mosquito populations. Mol. Ecol. Res. 20, 286–298 (2016).
16. Atayem, C. M., Delsuc, F., Pasteur, N., Weill, M. & Duron, O. Diversification of Wolbachia endosymbiont in the Culex pipiens mosquito. Mol. Biol. Evol. 28, 2761–2772 (2011).
17. Nor, I. et al. On the genetic architecture of cytoplasmic incompatibility: inference from phenotypic data. Am. Nat. 182, E15–E24 (2013).
18. Klassen, L. et al. Inducible evolution of Wolbachia strain wPip from the Culex pipiens group. Mol. Biol. Evol. 25, 1877–1887 (2008).
19. Baldo, L. et al. Multilocus sequence typing system for the endosymbiont Wolbachia pipiens. Appl. Environ. Microbiol 72, 7098–7110 (2006).
20. Lo, N. et al. Taxonomic status of the intracellular bacterium Wolbachia pipiens. Int. J. Syst. Evol. Microbiol. 57, 654–657 (2007).
21. Pons, J. L. & Labesse, G. in Symbionome II: A new pipeline for comparative modeling of protein-ligand complexes. Nucleic Acids Res. 37, 485–491 (2009).
22. Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. Protein repeats: structures, functions, and evolution. J. Struct. Biol. 134, 117–131 (2001).
23. Bordenstein, S. R., Marshall, M. L., Fry, A. J., Kim, U. & Wernegreen, J. J. The Tripartite associations between bacteriophages, Wolbachia, and arthropods. PLoS. Pathog. 2, e43 (2006).
24. Wright, J. D., Sjostrand, F. S., Portaro, J. K. & Barr, A. R. The ultrastructure of the rickettsia-like microorganism Wolbachia pipiens and associated virus-like bodies in the mosquito Culex pipiens. J. Ultra. Res. 63, 79–85 (1978).
25. Aakre, C. D. et al. Evolving new protein–protein interaction specificity through promiscuous intermediates. Cell 163, 594–606 (2015).
26. Dumas, E. et al. Population structure of Wolbachia and cytoplasmic introgression in a complex of mosquito species. BMC Evol. Biol. 13, 181 (2013).
27. Ellegaard, K. M., Klasson, L., Näslund, K., Bourtizis, K. & Andersson, S. G. E. Comparative genomics of Wolbachia and the bacterial species concept. PLoS Genet. 9, e1003813 (2013).
28. Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: a multipurpose graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. 27, 221–224 (2010).
29. Bryant, D. & Moulton, V. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21, 255–265 (2004).
30. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267 (2006).
31. Paraskevopoulos, C., Bordenstein, S. R., Wernegreen, J. J., Wilen, J. H. & Bourtizis, K. Toward a Wolbachia multicellus sequence typing system: Discrimination of Wolbachia strains present in Drosophila species. Curr. Microbiol. 53, 388–395 (2006).

Acknowledgements
We would like to thank Sylvain Charlat, Pierrick Labbé, Benjamin Lopin, Nicole Pasteur, and François Roussel for helpful comments on the manuscript. We also thank Khalid Belhik for bioinformatic analyses performed on the Montpellier Bioinformatics Biodiversity cluster computing platform. Frédéric Delsuc for phylogenetic analyses, and Arnaud Berthomieu, Ali Bouattour, Emile Dumas, Patrick Makoundou, and Sandra Unal for technical support. Sequencing data were produced through the technical facilities of the LabEX ‘Centre Méditerranéen de l’Environnement et de la Biodiversité’ in the GENISQ platform. This work was funded by the French ANR (project ‘CIAWOL’ ANR-16-CE02-0006-01 and ANR-10-BINF-0003).

Author contributions
M.B., M.S., and M.W. conceptualized and designed the study; M.B., C.A., MaBe, and F.J. performed the experiments; M.B., M.C.-G., M.S., and M.W. analyzed and interpreted the data; M.B., M.S., and M.W. wrote the manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-017-02749-w.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.