Intra-lineage microevolution of *Wolbachia* leads to the emergence of new cytoplasmic incompatibility patterns

Short title: Short-term evolution of cytoplasmic incompatibility patterns

Alice Namias[1,*], Annais Ngaku[1], Patrick Makoundou[1], Sandra Unal[1], Mathieu Sicard[1], Mylène Weill[1,*]

[1] *ISEM, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France*

*Corresponding authors

Email: alice.namias@gmail.com (AN), mylene.weill@umontpellier.fr (MW)

Abstract

Mosquitoes of the *Culex pipiens* complex are worldwide vectors of arbovirus, filarial nematode, and avian malaria agents. In these hosts, the endosymbiotic bacteria *Wolbachia* induce cytoplasmic incompatibility (CI), that is, reduced embryo viability in so-called incompatible crosses. *Wolbachia* infecting *Culex pipiens* (*w*Pip) cause CI patterns of unparalleled complexity, associated with the amplification and diversification of *cidA* and *cidB* genes, with up to six different gene copies described in a single *w*Pip genome. By repeating crosses between *Culex* isofemale lines over 17 years, we documented the emergence of a new compatibility type. Using a new sequencing method adapted to multigene families to acquire *cid* genes, we showed that some *w*Pip genomes lost specific *cidA* gene copies, thus giving rise to several sub-lineages segregating in the same cage. By linking phenotypic changes to their underlying genotypic bases, we showed that gene copies that are key for CI phenotypes originated from recombinations, not point mutations. We revealed how new CI patterns could emerge as part of a two-step process: first, local changes take place in the CI repertoires while maintaining compatibility with the surrounding mosquitoes, and then migration and secondary contact occur with the incompatible lines.

Keywords

Wolbachia, *Culex pipiens*, cytoplasmic incompatibility, microevolution, recombination

Introduction

Wolbachia are maternally transmitted endosymbiotic bacteria that infect up to 50% of arthropod species [1–3]. These bacteria are well known for their wide range of reproductive manipulations in arthropods. Their most common manipulation is cytoplasmic incompatibility (CI), that is, a reduction in embryo hatching rates in crosses between infected males and uninfected females, which is the simplest form of CI. CI results from a *Wolbachia*-induced *modification* that perturbs the first division of embryos, which infected females are able to *rescue* [4,5]. CI is thus formalized in a *modification-rescue* framework.

CI can also occur between two infected individuals if they are infected with genetically different and incompatible *Wolbachia* strains. In *Culex pipiens* mosquitoes, in which *Wolbachia* infection is fixed [6], the *w*Pip strains are responsible for a unique (to date) complexity of CI patterns based on multiple uni- and bidirectional incompatibilities [7–10]. An initial analysis of *w*Pip genetic diversity in *Culex* identified five phylogenetic groups named *w*Pip-I to *w*Pip-V [11]. Comparing *w*Pip phylogeny with crossing results revealed that these groups were good predictors of compatibility outcomes [9]: while both compatible and incompatible crosses were reported in crosses among *w*Pip groups, intra-group crosses were, with rare exceptions, compatible.

In 2013, putative CI genes were identified using a combination of proteomic and genetic approaches [12]. Then, in 2017, two functional studies confirmed that these genes were key for CI, naming them *cif* for "CI factors" [13,14]. The *cif* genes were described in *cifA-cifB* tandems, playing a central role in CI, although the precise molecular mechanisms of CI are still to be determined. Two main alternative models are still being debated: a host-modification (HM) model in which *Wolbachia* toxicity is due to the modifications of the host, with rescue occurring through a reversal of these modifications; and a toxin-antidote (TA) model in which the sperm carries a *Wolbachia* factor that is toxic unless an appropriate *Wolbachia* antidote is expressed in the egg, thus binding and neutralizing the toxin [15–18]. In *Culex,* the complexity of the crossing patterns makes the TA model far more likely: (in)compatibility would thus be based on different toxin-antidote binding abilities [16,19–21].

61 Recent publications agree with *cifA* being the rescue (resc, or antidote) gene and *cifB* being a major
62 modification (mod, or toxin) gene [13,18,22]. Sequencing of *Wolbachia* genomes from different host
63 species revealed several different pairs of *cif* genes, with distinct functional domains that categorized
64 them into five clades [23–25]. All *w*Pip genomes sequenced so far have *cif* genes from two of these
65 clades: clade I *cif*, in which the *cifB* gene has a deubiquitinase domain (the tandem is thus called *cid*,
66 with *d* denoting the deubiquitinase domain), and clade IV *cif,* in which *cifB* bears a nuclease domain
67 (thus called *cin*) [13,14]. Here, we use this functional-based nomenclature [19].

68 Before the discovery of the *cif* genes, models based on *Culex*-crossing experiments predicted that
69 several factors or pairs of *mod*/*resc* genes were required to encode the complex crossing patterns
70 induced by *w*Pip [9,26]. Studies on *cif* (*cid* and *cin*) genes present in *w*Pip genomes highlighted that
71 *cin* genes were always monomorphic, whereas *cid* genes were amplified and diversified in all
72 sequenced *w*Pip genomes [27], with up to six different copies of each gene in a single *Wolbachia*
73 *w*Pip genome. The different *cidA/cidB* copies in a given *w*Pip genome are known as "variants" and
74 the full set of all copies constitutes the "*cid* repertoire." In *w*Pip genomes, the polymorphism of both
75 *cidA* and *cidB* genes are located in two specific regions, which were named "upstream" and
76 "downstream" regions and predicted to be involved in CidA-CidB interactions [27]. This has been
77 confirmed by the recently obtained structure of a CidA-CidB cocrystal: out of the three interaction
78 regions identified, two perfectly match the previously identified upstream and downstream regions
79 for both CidA and CidB [20,21]. In addition to showing that CidA and CidB from *w*Pip bind together,
80 this recent study also showed that the different CidA-CidB variants have different binding properties,
81 thus lending further support to the toxin-antidote model for *w*Pip CI complexity [21].

82 Evolutionary changes of *Wolbachia* genomes on long time scales have been widely documented
83 in different host species. The comparison of c*if* genes in *Wolbachia* strains from different arthropod
84 hosts showed that they are quite divergent, and highly subject to lateral gene transfers [28,29] so
85 that the congruence between *Wolbachia* and *cif* phylogenies is totally disrupted [23,24]. Such lateral
86 transfers may be phage-linked, as *cif* genes are located in WO prophage regions [14,23], with lateral

87 transfers of genes in prophage WO regions being previously documented [30–33]. Transposon-
88 dependent (and thus phage-independent) transfers of *cif* genes have also been described [28]. By
89 contrast, although the ability of *Wolbachia* to rapidly adapt to new environmental conditions has
90 been suggested [34], very few studies have explored the short-term evolution of *Wolbachia*. To our
91 knowledge, the only studies linking rapid changes in phenotypes (in a few host generations) to
92 underlying genomic variations were reported for the "Octomom region" in *w*Mel and *w*MelPop,
93 showing that variations in the amplification level of this region were responsible for variations in
94 both *Wolbachia* virulence and viral protection conferred by *w*Mel [35–37].

95 CI patterns in *Culex* were previously shown to change over a few host generations [38]. Yet these
96 changes were described before the discovery of *cif* genes. Here, we observed a change in the CI
97 pattern between two isofemale lines kept in our laboratory since 2005: Slab (*w*Pip III) and Istanbul
98 (Ist, *w*Pip IV) [10]. While crosses between Slab females and Ist males were compatible from 2005 to
99 2017, we observed fully incompatible crosses for the first time in 2021. This shift in CI patterns
100 presented an opportunity to study the underlying genetic basis of CI evolution in laboratory-
101 controlled isofemale lines.

102 The emergence of a new CI phenotype may have resulted from (i) contamination, (ii) the
103 acquisition of a new toxin in some Ist *Wolbachia*, or (iii) the loss of antidotes in some Slab *Wolbachia*.
104 We took advantage of a recent methodological development that enables the rapid and extensive
105 acquisition of *cidA* and *cidB* repertoires using Nanopore Technologies sequencing [39] to address
106 these three hypotheses. We were able to rule out the contamination hypothesis and show that no
107 less than three distinct *Wolbachia* sublineages with different *cidA* repertoires now coexist in the Slab
108 isofemale line (but do not appear to coinfect the same individuals). We also found that the loss of
109 *cidA* variants (i.e., antidotes) in Slab perfectly matches variations in their rescuing ability. The *cidA*
110 variant whose presence/absence best matches the CI variations has original aminoacid combinations
111 at its interaction interface with CidB, probably resulting from a recombination.

Results

From CI phenotype shift to underlying genomic basis

In 2021, crosses between Slab females and Ist males highlighted that some of the Slab females were fully incompatible with Ist males for the first time.

At least three *cidA* repertoires coexist in the Slab isofemale line

To decipher the genomic changes that led to this recent CI shift, we sequenced the *cidA* and *cidB* repertoires present in *Wolbachia* infecting four individuals from the Ist isofemale line and eight females from the Slab isofemale line (four compatible and four incompatible) using Nanopore Technologies sequencing of *cidA* and *cidB* polymerase chain reaction (PCR) products [39]. While *Wolbachia* present in all Ist individuals had the same *cid* repertoires (S1 Table), we found polymorphism in *cid* repertoires of Slab individuals: they all had the same *cidB* (i.e., toxin) diversity, composed of *cidB-III-ae3* and *cidB-III-ag1* variants, whereas *cidA* (i.e., antidotes) repertoires varied among individuals. *Wolbachia* present in compatible and incompatible Slab females all exhibited the variants *cidA-III-alpha(5)-25* and *cidA-III-gamma(3)-12,* although additional variants were found in compatible females: *cidA-III-beta(2)-25*, either alone or accompanied by *cidA-III-beta(2)-16*. The observed differences in *cidA* repertoires can be summed up by the presence/absence of two specific *cidA* regions: the upstream region *cidA-III-beta(2)* and the downstream region *cidA-III-16* (S1 Fig).

cidA repertoire variation fully explains compatibility polymorphism

To determine if compatibility variations were linked to these specific *cidA* repertoire variations, we performed three replicates of the [Slab x Ist] cross. We isolated egg rafts from a total of 102 Slab females from the same isofemale line (28, 34, and 40 females for replicates 1, 2, and 3, respectively) and studied their hatching rates (HRs, i.e., number of hatched larvae over the number of eggs laid). Eight females produced unfertilized rafts and were thus excluded from further analyses. For the

137 remaining 94 females, HRs ranged from 0% (fully incompatible raft) to 96.4%, thus highlighting the
138 strong differences in the rescuing abilities of females. Of these 94 females, 25 (i.e., 27%) were fully
139 incompatible with Ist males.
140 Since repertoires differed in the presence/absence of two specific regions of *cidA* repertoires, we
141 designed primers and specifically screened the 94 females (S2 Table). The PCR assay confirmed that
142 three distinct *Wolbachia cidA* repertoires coexisted in the Slab isofemale line: (i) individuals hosting
143 *Wolbachia* with both the upstream *beta(2)* and downstream *16* regions, named (β+,16+); (ii)
144 individuals hosting *Wolbachia* with only the upstream region *beta(2)* but not the downstream *16*
145 region, named (β+,16-); and (iii) individuals hosting *Wolbachia* genomes with neither of these
146 regions, named (β-,16-) (S2 Table). For the sake of simplicity, we metonymically qualified mosquitoes
147 based on their *cidA* repertoires (β+,16+), (β+,16-), or (β-,16-). Of the 94 females analyzed, 61 were
148 (β+,16+), 8 (β+,16-), and 25 (β-,16-) (Fig 1).
149
150 **Fig 1. Intra-line polymorphism in *cidA* repertoires matches the compatibility patterns.** (β-,16-)
151 females have neither *cidA-III-beta(2)* nor *cidA-III-16* regions; (β+,16+) females have both regions, and
152 (β+,16-) females have the *cidA-III-beta(2)* region but not the *cidA-III-16*. The presence or absence of
153 *cidA-III-beta(2)* and *cidA-III-16* regions was determined by specific qPCRs. *N* = Number of individuals.
154
155 Egg rafts produced by (β-,16-) females had a null hatching rate, with a single exception, in which only
156 one larva hatched, resulting in a HR of 1.8%. Fully incompatible crosses were only observed with (β-
157 ,16-) mothers, as all (β+,16+) and (β+,16-) mothers were compatible with Ist males (to varying
158 extents) (Fig 1).
159

cidA repertoire variations do not explain HR variability in compatible crosses

While it was long believed that the presence of CI in *Culex* was a "yes or no" phenomenon, a recent study explored HR variations in compatible crosses with *w*Pip from different groups [40]. Examining the compatible crosses involving (β+,16+) or (β+,16-) Slab females and Ist males, we also found strong variations in HRs ranging from 3.2% to 96.4% (S2 Table). While infection levels (quantified by qPCR) in full individuals were variable, with 0.11 to 10.1 *Wolbachia* genomes per host genome and an average of 1.0 ± 0.2 (mean ± standard error), this variability did not correlate with HR variations (S2A Fig, Spearman correlation test, ρ = -0.09, p = 0.49, S2 Table). Similarly, HR variations were not correlated with the *cidA* copy numbers (total *cidA*, *cidA-III-beta*, or *cidA-III-16* copy numbers) per *w*Pip-Slab genome investigated in Slab females (Spearman correlation test, ρ = -0.13, p = 0.28; ρ = 0.01, p = 0.94; ρ = -0.17, p = 0.16 for the total *cidA*, *cidA-III-beta,* and *cidA-III-16*, respectively; S2B-S2D Figs, S2 Table).

Recombination in *cid* variants leads to changes in CidA-CidB interaction zones

Recent works revealed that there were three interaction interfaces between CidA and CidB proteins, named interfaces I to III (Fig 2A, [20,21]). The upstream region *cidA-III-beta(2)*, present in both (β+,16+) and (β+,16-), seems key to enable Slab females to rescue Ist males. This region corresponds to the first interaction interface of the CidA protein (Fig 2A). Looking closer at the protein sequences at this interface, we found that CidA-III-beta(2) has an original combination of aminoacids resulting from a recombination of the CidA-alpha and CidA-gamma regions (Fig 2B). Building a phylogenetic network of all the described *cidA* upstream regions [22,27,40,41] confirmed that *beta*-like regions were rare (only found in the *w*Pip group IV and Slab to date) and that they came from a recombination of the *cidA-gamma*-like and *cidA-alpha*-like regions, with these last two regions being

185 frequent and found in all *w*Pip groups surveyed to date (Fig 2C). We also showed that the *cidA-delta*
186 region was recombinant.
187
188 **Fig 2. Matching recombination in the CidA-CidB interaction interface I is key for rescue.** (A) CidA
189 and CidB interact head-to-tail through three distinct interfaces, shown here in green, yellow, and red,
190 respectively. The *cidA-III-beta(2)* and *cidB-IV-2* regions, which are key to the (in)compatibility
191 between Slab and Ist isofemale lines, are both exposed in the interaction interface I. (B) Alignments
192 of the protein sequences of the interaction interface I are shown for *cidA* and *cidB* (only sequence
193 between the first and last variable aminoacid belonging to interface I are represented; Wang &
194 Hochstrasser, pers. com.). (C) Network phylogenetic analysis of the *cidA* upstream and *cidB*
195 downstream regions using the Neighbor-Net method on 35 *cidA* variants and 22 *cidB* variants taken
196 from this study and [27,40,41]. For both regions, there are two major groups of variants
197 (alpha/gamma for *cidA* and 1/3 for *cidB*), with several intermediate recombinant variants (in green).
198 Variants found in the present study are shown in light green. Each edge (or set of parallel edges)
199 corresponds to a split in the dataset with a length equal to the weight of the split.
200
201 The CidB-IV-2 downstream region, present in *w*Pip-Ist, had previously been associated with
202 incompatibility when crossed with a wide range of females [22]. This region interacts with *cidA-beta*
203 at the interface I (Fig 2A) and also displays a pattern that could be explained by a recombination
204 (between CidB-1 and CidB-3 downstream regions) (Fig 2B). When looking at wider phylogenetic
205 scales based on all the previously published sequences, we found an even clearer pattern with only
206 two CidB-1 and CidB-3 downstream regions, identical among *w*Pip groups, with CidB-IV-2 being the
207 only intermediate recombinant region (Fig 2C).
208

Evolution of *cid* repertoires

Different *cidA* repertoires segregate into sublines

Since a new CI pattern emerged between 2017 and 2021, a closer look at the transmission and variation of Slab individuals' repertoires was required to better understand the precise nature of this evolution. For each of the three *cidA* repertoires found in the original Slab isofemale line, we established two isofemale sublines. We sequenced their *cidA* and *cidB* repertoires and confirmed that they were identical to those previously found for (β-,16-), (β+,16-), and (β+,16+) (S1 Fig). We confirmed that the females of these new isofemale sublines had the same CI phenotype as their respective founding females when crossed with Ist males (Table S3). We also verified that they were compatible with each other in both directions, showing that *cidA* repertoire polymorphism among sublines did not create incompatibilities (Table S4).

To test whether changes in repertoires could occur at the scale of a few host generations, we genotyped *Wolbachia* from 20 to 45 individuals of each Slab subline after about 15 generations for both *cidA* markers and found that no change had occurred since the establishment of the isofemale sublines.

We also genotyped individuals from the Slab isofemale line kept in liquid nitrogen since 2005, that is, shortly after the establishment of the isofemale line in our insectaries. Out of 30 tested mosquitoes, there were 26 (β+,16-) and 4 (β+,16+). None of them was (β-,16-), which is in accordance with the hypothesis that this genotype, leading to a new CI pattern, appeared afterwards.

No sign of stable coinfection

Using qPCR on Slab females, we found that total *cidA* copy numbers were variable, ranging from 2.2 to 6.2 copies per *w*Pip genome (quantifications normalized on the single copy *Wolbachia* gene *wsp*; S3A Fig). However, the distribution was bimodal with two peaks around three and five *cidA* copies (S3D Fig). The *cidA* copy number was homogeneous for females infected with a given *cidA* repertoire,

with around five *cidA* copies for all (β+,16+) and (β+,16-) individuals and around three *cidA* copies for all (β-,16-) individuals (S3A Fig). We also quantified the number of *cidB* copies in eight to ten females of each repertoire and found that they had equal copy numbers of *cidB* and *cidA*: around five copies in both (β+,16+) and (β+,16-) individuals and around three *cidB* copies in (β-,16-) individuals (S4 Fig). These results show that while Slab *cidB* repertoires do not differ between lines in terms of the identity of variants, they differ in terms of the *cidB* copy numbers.

The variability in these qPCR results can be explained by the intrinsic variability of qPCR. The centered data distributions in S3 and S4 Figs support the conclusion that most individuals (if not all) from all the (β+,16+), (β+,16-), and (β-,16-) sublines are not coinfected by *Wolbachia* with different *cid* repertoires.

Multiple copies of the same variant in the repertoires

qPCR results revealed that the total number of *cidA* copies was always higher than the number of distinct variants, showing that some variants must be present in multiple identical copies in a given repertoire. By combining the qPCR data and relative read coverage information from Nanopore sequencing (e.g., variant with two copies is covered twice as much as one with a single copy), we estimated the copy numbers of each variant in the different *cidA* repertoires of the Slab line. We found that (β+,16+) individuals, which have a total of five *cidA* copies, had two copies of *cidA-III-alpha(5)-25* and a single copy of each of the other three variants (*cidA-III-beta(2)-16*, *cidA-III-beta(2)-25*, and *cidA-III-gamma(3)-12*). (β+,16-) individuals also had five *cidA* copies, of which there were two copies of *cidA-III-alpha(5)-25*, a single copy of *cidA-III-beta(2)-25*, and two copies of *cidA-III-gamma(3)-12*. Lastly, (β-,16-) individuals, which had a total of three *cidA* copies, had two copies of *cidA-III-alpha(5)-25* and a single copy of *cidA-III-gamma(3)-12* (Fig 3).

Fig 3. *cidAB* repertoires of *Wolbachia* infecting the three Slab sublines. The *cidA* gene is represented in light purple and *cidB* in dark purple. Upstream and downstream variable regions are shown in

color. Regions in green are the key studied regions: *cidA-III-beta(2)* and *cidA-III-16.* The copy number of each tandem was deduced by combining the relative coverages in nanopore sequencing data with the qPCR data.

We confirmed the deduced copy numbers using qPCR specific to *cidA-III-beta(2)* and *cidA-III-16*: quantifying these two markers in 89 single Slab females, we confirmed that (β+,16+) repertoires had a single copy of *cidA-III-16*, while the other repertoires had none (S3B and S3E Figs). For *cidA-III-beta(2)*, we found that (β+,16+) individuals had two copies (corresponding to one copy of *cidA-III-beta(2)-16* and one copy of *cidA-III-beta(2)-25*), while (β+,16-) individuals had a single copy (unique copy of *cidA-III-beta(2)-25*). Finally, (β-,16-) individuals had none of these copies (S3C and S3F Figs).

cidA and *cidB* variants are always located within tandems

Previous investigations describing the *cid* repertoire of a given *Culex* line were performed by acquiring the sequences of *cidA* and *cidB* variants separately [22,27,39–41]. Here, we also amplified and Nanopore sequenced a region with both *cidA* and *cidB* variable regions in a single amplicon (S5 Fig). We could thus reconstruct the *cidA-cidB* tandems associated in the genome, resulting in the complete repertoires shown in Figure 3. To determine whether all *cid* variants were associated in tandems or whether they could stand alone in the genome, we compared their relative coverage when amplified and sequenced alone versus when amplified and sequenced in *cidA-cidB* tandem acquisitions. As these relative coverages were similar, we considered that all *cidA* and *cidB* variants were always located in tandems. Thus, (β+,16+) and (β+,16-) sublines have five *cidAB* tandems, while (β-,16-) lines have three *cidAB* tandems (Fig 3). Individuals of all sublines have *Wolbachia* with two copies of the *cidA-III-alpha(5)-25_cidB-III-ae3* tandem and one copy of the *cidA-III-gamma(3)-12_cidB-III-ag1* tandem. In addition, (β+,16+) and (β+,16-) individuals have one *cidA-III-beta(2)-25_cidB-III-ae3* tandem, although they differ in terms of their fifth tandem: *Wolbachia* in (β+,16+) individuals have a specific *cidA-III-beta(2)-16_cidB-III-ae3* tandem, while those in (β+,16-) individuals

have a *cidA-III-gamma(3)-12_cidB-III-ae3* tandem instead (Fig 3). Our results show that the *cidB* variant *"cidB-III-ae3"* is associated in tandem with four different *cidA* variants (Fig 3).

Discussion

The evolution of *Wolbachia* genomes has already been documented at large phylogenetic scales, with multiple horizontal transfers and gene losses (e.g., [31,32,37,42–45]). This is especially true for the *cif* gene family: numerous horizontal transfers have been described, resulting in a strong incongruence between *cif* gene phylogeny and *Wolbachia* phylogeny [23,24,28]. Although rapid shifts in CI patterns were previously described in *Culex* mosquitoes infected with *w*Pip [38], the *cif* genes had not been described at that time, thus preventing the study of the genetic basis of these phenotypic changes. Furthermore, we could not rule out the impact of contamination in such shifts. Here, we could link changes in CI patterns with their *Wolbachia* genomic basis for the first time.

We detected a recent shift in CI patterns between Slab (*w*Pip-III) and Ist (*w*Pip-IV) lines: while they had been compatible for 12 years from 2005 to 2017, a quarter of the crosses were fully incompatible in 2021. We looked for underlying genomic variations in their *cid* repertoires. This shift in CI could have resulted from either a variation in the *cidB* (toxin) repertoires of Ist males or a variation in the *cidA* (antidote) repertoires of Slab females. We found that *cid* repertoires did not vary in Ist but that three different *cidA* repertoires coexisted in the same Slab isofemale line: two distinct repertoires with five *cidA* copies, labelled (β+,16+) and (β+,16-), which were able to rescue the modification induced by *w*Pip-Ist, and a repertoire with three *cidA* copies, named (β-,16-), which were unable to rescue these modifications. These results indicate that some *w*Pip genomes in the Slab isofemale line lost *cidA* variants. Indeed, the screening of mosquitoes from 2005 revealed that the (β-,16-) repertoire did not exist at that time, showing that some *cidA* variants were lost in some *w*Pip-Slab genomes between 2017, when all crosses were compatible, and 2021, when the shift in CI was detected.

By studying the genomic basis of phenotypic changes, we showed that *cid* gene repertoires truly evolved in *w*Pip over a period of 4 years. While previous studies showed that *cid* genes are highly variable within and between *w*Pip groups (with up to 30 distinct variants described to date for *cidA* and 20 for *cidB* [22,27,40,41]), we found that the new (β-,16-) repertoire is composed of *cidA* and *cidB* variants present in ancestral repertoires, which are associated in an identical *cidA-cidB* tandem architecture. We thus conclude that the shift in the CI pattern results from the evolution of the repertoire and not from any type of contamination.

Rapid evolution of *cid* repertoires through gene losses

In 2017, approximately one quarter of females had a total of three *cidA-cidB* tandems, showing that two *cid* tandems had been lost. *cid* copy number distributions are bimodal, with two peaks at three and five copies and very rare with intermediate repertoires of four copies. These rare intermediates are likely due to the qPCR variability. This could mean either that (i) four-copy repertoires were a transient state that disappeared quickly due to lower fitness or drift or that (ii) four-copy repertoires did not exist, meaning that two tandems were lost simultaneously.

Here, we found that all the observed repertoires were mutually compatible. This means that mosquitoes infected by the putative four-copy repertoires would also have been compatible with all surrounding mosquitoes, making its quick disappearance unlikely. The most plausible hypothesis is that two tandems were lost simultaneously, making it likely that the two tandems were located next to each other.

Literature on CI evolution suggests that *mod* factors or toxins (*cifB* variants) could degenerate first, as their evolution is neutral for maternally transmitted *Wolbachia* [46,47]. Only once *cifB* variants have lost their function is it possible for the loss of *cifA* (*resc* factors or antidotes) [48]. In such a conceptual framework, losing *cifA* without losing *cifB* would lead to self-incompatibility, which is an immediate evolutionary dead-end. Independent from CI, the presence of *cidB* without an associated *cidA* might be toxic for the host, as shown recently in insect cells [4]. These predictions have been

337 corroborated by a large-scale phylogenetic study of *cif* genes, indicating that *cifB* genes were often
338 disrupted, while *cifA* were not [24]. However, at a microevolutionary scale, we observed that it was
339 possible to qualitatively decrease the diversity of *cidA* variants without qualitatively changing the
340 *cidB* variants: all (β+,16+), (β+,16-), and (β-,16-) repertoires described here share the same *cidB*
341 variants (*cidB-III-ag1* and *cidB-III-ae3*), while they have four, three, and two distinct *cidA* variants,
342 respectively (Fig 3). This apparent discrepancy between the predictions and our results may be
343 explained by the fact that *w*Pip genomes contain several *cid* genes. Unlike the other systems studied
344 to date where the loss of *cifA* can lead to a loss of the self-rescuing ability, *w*Pip has several *cidA*
345 variants, which makes it possible to lose one variant while still maintaining compatibility with the
346 surrounding mosquitoes. The loss of *cidA* variants required for intra-line compatibility may also
347 occur, although the mosquitoes that lost these key antidotes would be quickly wiped out by
348 selection, a hypothesis that has been corroborated by the mutual compatibility between all newly
349 established isofemale lines, regardless of their *cid* repertoire.
350 Laboratory conditions could facilitate this evolution by qualitatively reducing the *cidA* repertoire.
351 Indeed, in natural populations, multiple *cid* repertoires coexist (as shown by sequencing and revealed
352 by numerous crosses [22,27,41,49]), meaning that individuals are exposed to a wide range of toxins.
353 A greater diversity in the *cidA* repertoire enables the host to rescue a wider range of males, which is
354 thus under positive selection. While *cidA* diversity is advantageous in natural populations, in a cage
355 with an isofemale line, females are exposed to a restricted and fixed set of toxins. Loss of *cidA*
356 variants, which are not strictly required for the rescue of the *cidB* in a given cage, is not counter-
357 selected. Furthermore, as numerous gene copies can be costly (e.g., [50]), the loss of *cidA-cidB*
358 tandems could confer a higher fitness to *Wolbachia* with lower numbers of copies and thus
359 contribute to the fixation of new *Wolbachia* genomes in mosquitoes, which could explain the rapid
360 *Wolbachia* turnover. It is conceivable that the three-copy repertoire of (β-,16-) *Wolbachia*, which
361 contains two distinct variants of *cidA* and two distinct variants of *cidB*, is the minimum viable
362 repertoire with only pairs of matching toxins and antidotes.

No lasting coinfection involved in changes in CI patterns in *w*Pip

One plausible explanation for the observed shift in CI could be that Slab is coinfected by two strains of *Wolbachia* with different compatibility types and different *cid* repertoires, which are usually co-transmitted, and that some individuals recently lost one of these strains, as observed in *Aedes albopictus* [51,52]. Using common MLST markers, Baldo et al. initially found no variability in the *w*Pip strains [53], showing that if coinfection occurs, it should be between highly similar strains. The development of a *w*Pip-specific MLST led to the uncovering of *w*Pip polymorphism and the description of five monophyletic groups within *w*Pip [11]. However, while these markers highlighted *w*Pip polymorphism, no sign of coinfection was found, as every mosquito clearly hosted one of the five groups [11]. In this study focusing on *cid* genes, which are highly variable even within a *w*Pip group, we likewise found no clear signs of coinfections. We quantified several *cid* markers and analyzed the distribution of their values. In the case of coinfections with two *Wolbachia* strains with three and five *cid* copies, respectively, quantification outcomes would have varied depending on the proportion of each *Wolbachia* in each mosquito, leading to a unimodal distribution centered on an intermediate value. For the three distinct *cid* markers analyzed in 89 females, the quantifications displayed a bimodal distribution (S3D-F Figs), with a few intermediate values probably due to technical variations in the qPCR measures. While coinfection occurred at some point (all *Wolbachia* within one individual cannot be wiped out and instantly replaced by another *Wolbachia*), it had to be brief with no lasting coinfection in a single host individual. This could be explained by two alternative hypotheses: (i) an exclusion between the two distinct genotypes (three and five copies) for any non-CI-related reason, or (ii) a small effective population of *Wolbachia* (number of *Wolbachia* transmitted from mother to offspring) in *w*Pip.

Variability in hatching rates in compatible crosses not explained by *cid* repertoires of females

CI has long been described as a "yes or no" phenomenon in *Cx pipiens*, meaning that crosses are either compatible or fully incompatible [9]. However, after closely investigating the HRs, we recently showed that they could be highly variable in so-called compatible crosses [40]. We further revealed that reduced HRs were associated with cytological signs of canonical CI [40]. We suggested that these phenomena were linked to the *cid* repertoires, as the males infected with *w*Pip-IV, which had distinct *cidB* variants from other *w*Pip groups, induced more intermediate CI than other males.

Studying the cross between Slab females (*w*Pip-III) and Ist males (*w*Pip-IV), we also found intermediate HRs in "compatible" crosses. While on/off compatibility was clearly explained by the presence/absence of specific *cidA* variants, these intermediate HRs did not correlate with any of the *cidA* markers nor with the *Wolbachia* infection rate.

Once again, intermediate HRs were found in crosses involving males infected with *w*Pip-IV. By comparing *cidA* and *cidB* sequences from different wPip groups, we showed that the second interaction interface differed between CidB belonging to *w*Pip-IV and CidB from other groups, with no variation in the interface II of CidA variants (S6 Fig). This could mean that CidB variants belonging to *w*Pip group IV never perfectly bind CidA variants. The absence of a match at interface II could thus lead to a more unstable CidA-CidB interaction and explain the intermediate HR values.

Recombination as the key to the emergence of new CI patterns

A specific *cidB* region, named *cidB-IV-2*, was previously correlated with strong CI-inducing capabilities in males [22,27]. This region is present in the *w*Pip infecting the Ist males studied here. We showed that this specific region is issued from a recombination between the two main groups of *cidB* downstream regions, which are shared by all sequenced *w*Pip groups (Fig 2).

413 Furthermore, it was previously shown that the only females able to rescue the modifications induced
414 by Ist males were those infected with *w*Pip-IV as well as Slab females (*w*Pip-III, studied here) [22,27].
415 We found that the rescuing ability of Slab females depended on the presence of a specific *cidA*
416 upstream region known as *cidA-beta*: all females with this region can rescue Ist males, while females
417 that have lost it cannot. This variant was also generated by a recombination between two main
418 groups of upstream regions shared by all *w*Pip groups. Here, there were two types of recombinants,
419 the *cidA-beta* and *cidA-IV-delta* regions (Fig 2). When looking at the repertoires found in females
420 infected with *w*Pip-IV, which can all rescue Ist males, we found that they all had one recombinant
421 region in their repertoire (either *cidA-beta* or *cidA-IV-delta*). In other words, all females able to
422 rescue the toxicity of the recombinant region *cidB-IV-2* have a recombinant antidote in their
423 repertoire. Since the *cidA* upstream region (e.g., *beta* or *delta*) interacts with the *cidB* downstream
424 region (e.g., *cidB-IV-2*) in the interaction interface I [21], this finding points to recombinant-for-
425 recombinant binding (Fig 2) in which the recombinant regions can only bind similarly recombinant
426 regions. New CI phenotypes could thus emerge through the recombination of existing variants.
427
428 In conclusion, we found that CI evolved in less than 4 years in a *Culex* isofemale line. This evolution is
429 linked to the loss of some *cidA* (antidote) variants in their *cid* repertoire. Our results suggest that
430 these incompatibilities could emerge as a two-step process: first, local changes take place in the CI
431 repertoires while maintaining compatibility with the surrounding mosquitoes, and then migration
432 and secondary contact occur with the incompatible lines. Here, we showed that all repertoires
433 present in a cage were mutually compatible: changes in the repertoires modified the CI pattern only
434 when crossed with an external line. We demonstrated that new compatibility types can emerge
435 rapidly, as a result of the rapid evolution of *Wolbachia* genomes. A common point in the few studies
436 addressing *Wolbachia* microevolution is that both Octomom in *w*Mel [35–37] and *cid* genes in *w*Pip
437 are amplified genomic regions, suggesting the strong role played by gene amplification in the speed
438 of *Wolbachia* genome evolution.

This rapid evolution should therefore be considered when releasing *Wolbachia*-infected mosquitoes, inducing CI, for vector control purposes.

Materials and methods

All DNA extractions in this study were performed following the cetrimonium bromide (CTAB) protocol [54].

Isofemale lines

All the mosquito lines used were isofemale lines, i.e., lines created by rearing the offspring resulting from a single egg raft and thus from a single female. We used the following lines: Slab, infected with *Wolbachia* from the phylogenetic group *w*Pip-III, and Istanbul (Ist), infected with *Wolbachia* from the phylogenetic group *w*Pip-IV. Both lines were established in our laboratory in 2005.

We also used six newly established isofemale (sub)lines: two (β+,16+) lines bearing both the *cidA-III-beta* and *cidA-III-16* regions; two (β+,16-) lines bearing only the *cidA-III-beta* marker; and two (β-,16-) lines with neither of these markers. In short, to establish the isofemale line, a cage containing the Slab isofemale line was blood fed. After 5 days, females were isolated in glasses for egg laying. After egg laying, DNA was extracted, and females were genotyped for *cidA-III-beta* and *cidA-III-16* (S5 Table). For each genotype, the two most numerous progenies were kept in order to establish two separate isofemale lines. All isofemale lines were reared in 65 dm^3 screened cages in a single room maintained at 26°C under a 12h light/12h dark cycle. Larvae were fed with a mixture of shrimp powder and rabbit pellets, and adults were fed on a honey solution. Females were fed with turkey blood using a Hemotek membrane feeding system (Discovery Workshops, UK) to enable them to lay eggs.

Crosses

For each cross performed, around 100 virgin females were put in a cage with around 50 virgin males. After 5 days in cages, females were fed a blood meal; and 5 days later, individual females were isolated for egg laying. This allows us to precisely link a female with its egg raft. Egg rafts were deposited into 24-well plates. The eggs were photographed using a Leica camera and counted using ImageJ. Two days after egg laying, larvae issued from each egg raft were counted manually. HR was calculated as the ratio between the total number of larvae and the total number of eggs. Egg rafts with a null HR were mounted on a slide and examined with a binocular magnifier to ensure that the eggs were fertilized and thus confirm that null HR resulted from CI rather than non-fertilization.

Repertoire acquisition using Nanopore sequencing of PCR products

We acquired the *cidA* and *cidB* repertoires of eight individuals from the Slab isofemale line, four individuals from the Ist isofemale line, and two individuals from each of the newly established isofemale lines following the pipeline described in [39]. We also acquired *cidAB* combinations for all newly established isofemale lines. DNA was extracted from the individuals. In short, the target sequence was amplified using the adequate primer set (S5 Table, S6 Fig). PCR products were then sequenced by Nanopore sequencing with the MGX Platform (Montpellier) using a MinION (Oxford Nanopore Technologies). Repertoires were then acquired following the pipeline described in [39]. For *cidAB* combinations, the pipeline was adapted: *cidA* and *cidB* repertoires were first obtained, and then the combinations between all *cidA* and *cidB* were artificially created and used as a reference base to run the pipeline.

Wolbachia genotyping

Standard PCR

To genotype individuals, DNA was extracted. The presence/absence of both *cidA-III-16* and *cidA-III-beta* was determined using specific PCRs (S5 Table). For individuals negative for both markers, a PK1 PCR (S5 Table) was used to control DNA quality. This was performed on either adult females or fourth instar larvae.

Quantitative PCR

Real-time quantitative PCR was run using the LightCycler 480 system (Roche). All DNA samples were analyzed in triplicates for each quantification. A total of 89 females were used for qPCR assays.

The total number of *cidA* copies, along with the number of *cidA-III-beta* and *cidA-III-16* region copies, were quantified using relative quantitative PCR, corrected by *wsp,* a single-copy *Wolbachia* gene (following [55], primers in S5 Table). Since *wsp* is present in a single copy per haploid genome, the ratio of the target and *wsp* signal can be used to estimate the number of target copies per *Wolbachia* genome, correcting for *Wolbachia* infection level and DNA quality. For *cidA-III-beta* qPCR, quantifications were normalized on three distinct individuals with a single *cidA-III-beta* copy.

The infection level (*i.e.*, the number of *Wolbachia* bacteria per female) was assessed with qPCR using a *wsp* PCR normalized on the *Culex*-specific single-copy *ace-2* locus [56]. The ratio between *wsp* and *ace-2* signals can be used to estimate the relative number of *Wolbachia* genomes per *Culex* genome.

For *cidA-III-16/wsp* and *cidA/wsp*, the following protocol was used: 3 min at 95°C, 45 cycles of 95°C for 10 sec, 60°C for 20 sec, and 72°C for 20 sec. For *cidA-III-beta/wsp* and *ace2/wsp*, the protocol was as follows: 3 min at 95°C, 45 cycles of 95°C for 3 sec, 62°C for 10 sec, and 72°C for 15 sec. Primer concentrations were 0.6 µM for *cidA-III-16* and *cidA,* 0.4 µM for *wsp* and *ace2*, and 1 µM for *cidA-III-beta*.

Variant copy number analysis

Variant copy numbers were obtained using both the *cidA* qPCR data and the relative coverage of Nanopore sequencing. Knowing the total number of *cidA* copies, we deduced each variant copy number. For instance, in the case of three *cidA* copies, if out of the two variants present, the coverage of the first was twice the coverage of the second, it could be concluded that there were two copies of the first variant and one copy of the second.

Grouping analysis of upstream/downstream regions of variants

Phylogenetic analysis of variants was performed using all the variants already published [22,27,40,41] and present in this study. Network phylogenies of the two regions of interest (*cidA* upstream variable region and *cidB* downstream region, which respectively corresponded to nucleotides 1-800 and 1201-end) were made using Splitstree [57].

Nomenclature of *cid* variants

The nomenclature established by [27] was updated, since it could result in identical sequences being given different names in different *w*Pip groups. The fundamentals of the previous nomenclature were retained: names of variants were composed of the name of their upstream region (in Greek letters for *cidA* and in Latin letters for *cidB*) and that of their downstream region (digits). A fixed limit was established between the upstream and downstream regions for both genes (downstream regions starting at nucleotides 801 and 1201 for *cidA* and *cidB*, respectively). A phylogenetic network revealed four clusters of *cidA* upstream regions (Fig 2B), regions within a cluster differing by at most three single-nucleotide polymorphisms (SNPs). We renamed regions so that all those belonging to the same cluster were named using the same Greek letter (e.g., all named alpha). When SNPs were

present, to distinguish the sequences, we introduced a number between brackets such as alpha(1). The same was done for *cidB* downstream regions. Similarly, *cidA* downstream regions and *cidB* upstream regions were renamed so that distinct regions had different names, and identical regions had identical names, regardless of the *w*Pip group in which they were found.

For all variant names, the group in which the variant occurs is mentioned. For instance, cidA-III-alpha(5)-25) is found in a *w*Pip-III.

The correspondence between the previously used names and the names based on this new nomenclature is provided in S6 Table.

Statistical analyses and graphical representation

All statistical analyses were run using R version 4.2.0 [58]. Correlation analyses were performed using a Pearson correlation test. All boxplots and histograms were done using ggplot2 [59], and sequence representation was conducted using the function msaPrettyPrint from the msa package [60].

Data availability statement

The nucleotide sequences of the new *cidA* and *cidB* variants were deposited in GenBank. Accession numbers are provided in S7 Table. The authors declare that all other data supporting the findings of this study are available in the article and Supporting Information.

Acknowledgments

We thank Michael Turelli and Nicole Pasteur for their thorough reading and helpful comments on the manuscript and Wei Wang and Mark Hochstrasser for their input on the interface residuals. Sequencing data were generated at the Montpellier GénomiX platform.

References

1. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH. How many species are infected with *Wolbachia*? - A statistical analysis of current data. FEMS Microbiol Lett. 2008;281: 215–220. doi:10.1111/j.1574-6968.2008.01110.x
2. Zug R, Hammerstein P. Still a host of hosts for *Wolbachia*: Analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS One. 2012;7: 7–9. doi:10.1371/journal.pone.0038544
3. Weinert LA, Araujo-Jnr E V., Ahmed MZ, Welch JJ. The incidence of bacterial endosymbionts in terrestrial arthropods. Proc R Soc B Biol Sci. 2015;282: 3–8. doi:10.1098/rspb.2015.0249
4. Horard B, Terretaz K, Gosselin-Grenet AS, Sobry H, Sicard M, Landmann F, et al. Paternal transmission of the *Wolbachia* CidB toxin underlies cytoplasmic incompatibility. Curr Biol. 2022;32: 1319-1331.e5. doi:10.1016/j.cub.2022.01.052
5. Kaur R, Leigh BA, Ritchie IT, Bordenstein SR. The Cif proteins from *Wolbachia* prophage WO modify sperm genome integrity to establish cytoplasmic incompatibility. PLOS Biol. 2022;20: e3001584. doi:10.1371/journal.pbio.3001584
6. Duron O, Lagnel J, Raymond M, Bourtzis K, Fort P, Weill M. Transposable element polymorphism of *Wolbachia* in the mosquito *Culex pipiens*: evidence of genetic diversity, superinfection and recombination. Mol Ecol. 2005;14: 1561–1573. doi:10.1111/j.1365-294X.2005.02495.x
7. Laven H. Crossing experiments with *Culex* strains. Evolution (N Y). 1951;5: 370–375. doi:10.2307/2405682
8. Laven H. Speciation and evolution in *Culex pipiens*. In: Wright J, Pal R, editors. Genetics of Insect Vectors of Disease. Elsevier: Amsterdam; 1967. pp. 251–275.
9. Atyame CM, Labbé P, Dumas E, Milesi P, Charlat S, Fort P, et al. *Wolbachia* divergence and the evolution of cytoplasmic incompatibility in *Culex pipiens*. PLoS One. 2014;9: e87336. doi:10.1371/journal.pone.0087336

585 10. Duron O, Bernard C, Unal S, Berthomieu A, Berticat C, Weill M. Tracking factors modulating
586 cytoplasmic incompatibilities in the mosquito *Culex pipiens*. Mol Ecol. 2006;15: 3061–3071.
587 doi:10.1111/j.1365-294X.2006.02996.x
588 11. Atyame CM, Delsuc F, Pasteur N, Weill M, Duron O. Diversification of *Wolbachia*
589 endosymbiont in the *Culex pipiens* mosquito. Mol Biol Evol. 2011;28: 2761–2772.
590 doi:10.1093/molbev/msr083
591 12. Beckmann JF, Fallon AM. Detection of the *Wolbachia* protein WPIP0282 in mosquito
592 spermathecae: Implications for cytoplasmic incompatibility. Insect Biochem Mol Biol. 2013;43:
593 867–878. doi:10.1016/j.ibmb.2013.07.002
594 13. Beckmann JF, Ronau JA, Hochstrasser M. A *Wolbachia* deubiquitylating enzyme induces
595 cytoplasmic incompatibility. Nat Microbiol. 2017;2: 1–7. doi:10.1038/nmicrobiol.2017.7
596 14. LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI, Shropshire JD, et al. Prophage WO
597 genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. Nature.
598 2017;543: 243–247. doi:10.1038/nature21391
599 15. Beckmann JF, Bonneau M, Chen H, Hochstrasser M, Poinsot D, Merçot H, et al. The Toxin–
600 Antidote Model of Cytoplasmic Incompatibility: Genetics and Evolutionary Implications.
601 Trends Genet. 2019;35: 175–185. doi:10.1016/j.tig.2018.12.004
602 16. Namias A, Sicard M, Weill M, Charlat S. From *Wolbachia* genomics to phenotype: molecular
603 models of cytoplasmic incompatibility must account for the multiplicity of compatibility types.
604 Curr Opin Insect Sci. 2022;49: 78–84. doi:10.1016/j.cois.2021.12.005
605 17. Shropshire JD, Leigh B, Bordenstein SR, Duplouy A, Riegler M, Brownlie JC, et al. Models and
606 Nomenclature for Cytoplasmic Incompatibility: Caution over Premature Conclusions – A
607 Response to Beckmann et al. Trends Genet. 2019;35: 397–399. Available:
608 https://doi.org/10.1016/j.tig.2019.03.004
609 18. Shropshire JD, Bordenstein SR. Two-By-One model of cytoplasmic incompatibility: Synthetic
610 recapitulation by transgenic expression of *cifA* and *cifB* in *Drosophila*. PLoS Genet. 2019;15: 1–

611 20. doi:10.1371/journal.pgen.1008221

612 19. Beckmann JF, Bonneau M, Chen H, Hochstrasser M, Poinsot D, Merçot H, et al. Caution Does
613 Not Preclude Predictive and Testable Models of Cytoplasmic Incompatibility: A Reply to
614 Shropshire et al. Trends Genet. 2019;35: 399–400. doi:10.1016/j.tig.2019.03.002

615 20. Xiao Y, Chen H, Wang H, Zhang M, Chen X, Berk JM, et al. Structural and mechanistic insights
616 into the complexes formed by *Wolbachia* cytoplasmic incompatibility factors. Proc Natl Acad
617 Sci U S A. 2021;118: 1–10. doi:10.1073/pnas.2107699118

618 21. Wang H, Xiao Y, Chen X, Zhang M, Sun G, Wang F, et al. Crystal Structures of *Wolbachia* CidA
619 and CidB Reveal Determinants of Bacteria-induced Cytoplasmic Incompatibility and Rescue.
620 Nat Commun. 2022;13: 1–11. doi:10.1038/s41467-022-29273-w

621 22. Bonneau M, Caputo B, Ligier A, Caparros R, Unal S, Perriat-Sanguinet M, et al. Variation in
622 *Wolbachia cidB* gene, but not *cidA*, is associated with cytoplasmic incompatibility *mod*
623 phenotype diversity in *Culex pipiens*. Mol Ecol. 2019;28: 4725–4736. doi:10.1111/mec.15252

624 23. Lindsey ARI, Rice DW, Bordenstein SR, Brooks AW, Bordenstein SR, Newton ILG. Evolutionary
625 Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage WO of *Wolbachia*.
626 Genome Biol Evol. 2018;10: 434–451. doi:10.1093/gbe/evy012

627 24. Martinez J, Klasson L, Welch JJ, Jiggins FM. Life and Death of Selfish Genes: Comparative
628 Genomics Reveals the Dynamic Evolution of Cytoplasmic Incompatibility. Mol Biol Evol.
629 2021;38: 2–15. doi:10.1093/molbev/msaa209

630 25. Bing XL, Zhao DS, Sun JT, Zhang KJ, Hong XY, Sloan D. Genomic Analysis of *Wolbachia* from
631 *Laodelphax striatellus* (Delphacidae, Hemiptera) Reveals Insights into Its "jekyll and Hyde"
632 Mode of Infection Pattern. Genome Biol Evol. 2020;12: 3818–3831. doi:10.1093/gbe/evaa006

633 26. Nor I, Engelstädter J, Duron O, Reuter M, Sagot MF, Charlat S. On the genetic architecture of
634 cytoplasmic incompatibility: Inference from phenotypic data. Am Nat. 2013;182: E15–E24.
635 doi:10.1086/670612

636 27. Bonneau M, Atyame CM, Beji M, Justy F, Cohen-Gonsaud M, Sicard M, et al. *Culex pipiens*

637 crossing type diversity is governed by an amplified and polymorphic operon of *Wolbachia*. Nat
638 Commun. 2018;9: 319. doi:10.1038/s41467-017-02749-w
639 28. Cooper BS, Vanderpool D, Conner WR, Matute DR, Turelli M. *Wolbachia* acquisition by
640 *Drosophila yakuba*-clade hosts and transfer of incompatibility loci between distantly related
641 *Wolbachia*. Genetics. 2019;212: 1399–1419.
642 29. Madhav M, Parry R, Morgan JAT, James P. *Wolbachia* Endosymbiont of the Horn Fly
643 (*Haematobia irritans*): a Supergroup A Strain with Multiple Horizontally Acquired Cytoplasmic
644 Incompatibility Genes. Appl Environ Microbiol. 2020;86: e02589-19.
645 30. Bordenstein SR, Wernegreen JJ. Bacteriophage flux in endosymbionts (*Wolbachia*): Infection
646 frequency, lateral transfer, and recombination rates. Mol Biol Evol. 2004;21: 1981–1991.
647 doi:10.1093/molbev/msh211
648 31. Wang N, Jia S, Xu H, Liu Y, Huang D. Multiple Horizontal Transfers of Bacteriophage WO and
649 Host *Wolbachia* in Fig Wasps in a Closed Community. Front Microbiol. 2016;7: 1–10.
650 doi:10.3389/fmicb.2016.00136
651 32. Hill T, Unckless RL, Perlmutter JI. Positive Selection and Horizontal Gene Transfer in the
652 Genome of a Male-Killing *Wolbachia*. Mol Biol Evol. 2022;39: 1–17.
653 doi:10.1093/molbev/msab303
654 33. Gao S, Ren YS, Su CY, Zhu DH. High Levels of Multiple Phage WO Infections and Its
655 Evolutionary Dynamics Associated With *Wolbachia*-Infected Butterflies. Front Microbiol.
656 2022;13: 1–12. doi:10.3389/fmicb.2022.865227
657 34. Le Clec'h W, Dittmer J, Raimond M, Bouchon D, Sicard M. Phenotypic shift in *Wolbachia*
658 virulence towards its native host across serial horizontal passages. Proc R Soc B Biol Sci.
659 2017;284. doi:10.1098/rspb.2017.1076
660 35. Chrostek E, Marialva MSP, Esteves SS, Weinert LA, Martinez J, Jiggins FM, et al. *Wolbachia*
661 Variants Induce Differential Protection to Viruses in *Drosophila melanogaster*: A Phenotypic
662 and Phylogenomic Analysis. PLoS Genet. 2013;9. doi:10.1371/journal.pgen.1003896

663 36. Chrostek E, Teixeira L. Mutualism Breakdown by Amplification of *Wolbachia* Genes. PLoS Biol.
664 2015;13: 1–22. doi:10.1371/journal.pbio.1002065
665 37. Duarte EH, Carvalho A, López-Madrigal S, Costa J, Teixeira L. Forward genetics in *Wolbachia*:
666 Regulation of *Wolbachia* proliferation by the amplification and deletion of an additive
667 genomic island. PLoS Genetics. 2021. doi:10.1371/journal.pgen.1009612
668 38. Duron O, Bernard J, Atyame CM, Dumas E, Weill M. Rapid evolution of *Wolbachia*
669 incompatibility types. Proc R Soc B Biol Sci. 2012;279: 4473–4480.
670 doi:10.1098/rspb.2012.1368
671 39. Namias A, Sahlin K, Makoundou P, Bonnici I, Sicard M. Nanopore sequencing enables
672 multigenic family reconstruction despite highly frequent PCR-induced recombination. bioRxiv.
673 2022; 2022.11.14.516392. doi:10.1101/2022.11.14.516392
674 40. Sicard M, Namias A, Perriat-Sanguinet M, Carron E, Unal S, Altinli M, et al. Cytoplasmic
675 incompatibility variations in relation with *Wolbachia cid* genes divergence in *Culex pipiens*.
676 MBio. 2021;12: e02797-20. doi:10.1128/mBio.02797-20
677 41. Bonneau M, Landmann F, Labbé P, Justy F, Weill M, Sicard M. The cellular phenotype of
678 cytoplasmic incompatibility in *Culex pipiens* in the light of *cidB* diversity. PLoS Pathog.
679 2018;14: e1007364. doi:10.1371/journal.ppat.1007364
680 42. Scholz M, Albanese D, Tuohy K, Donati C, Segata N, Rota-Stabelli O. Large scale genome
681 reconstructions illuminate *Wolbachia* evolution. Nat Commun. 2020;11. doi:10.1038/s41467-
682 020-19016-0
683 43. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, et al. The *Wolbachia* genome of
684 *Brugia malayi*: Endosymbiont evolution within a human pathogenic nematode. PLoS Biol.
685 2005;3: 0599–0614. doi:10.1371/journal.pbio.0030121
686 44. Newton ILG, Clark ME, Kent BN, Bordenstein SR, Qu J, Richards S, et al. Comparative genomics
687 of two closely related *Wolbachia* with different reproductive effects on hosts. Genome Biol
688 Evol. 2016;8: 1526–1542. doi:10.1093/gbe/evw096

689 45. Wang GH, Sun BF, Xiong TL, Wang YK, Murfin KE, Xiao JH, et al. Bacteriophage WO can
690 mediate horizontal gene transfer in endosymbiotic *Wolbachia* genomes. Front Microbiol.
691 2016;7: 1–16. doi:10.3389/fmicb.2016.01867
692 46. Prout T. Some evolutionary possibilities for a microbe which causes incompatibility in its host.
693 Evolution (N Y). 1994;48: 909–911.
694 47. Turelli M. Evolution of incompatibility-inducing microbes and their hosts. Evolution (N Y).
695 1994;48: 1500–1513. doi:10.1111/j.1558-5646.1994.tb02192.x
696 48. Charlat S, Calme C, Merçot H. On the *mod resc* model and the evolution of *Wolbachia*
697 compatibility types. Genetics. 2001;159: 1415–1422. doi:10.1093/genetics/159.4.1415
698 49. Atyame CM, Duron O, Tortosa P, Pasteur N, Fort P, Weill M. Multiple *Wolbachia* determinants
699 control the evolution of cytoplasmic incompatibilities in *Culex pipiens* mosquito populations.
700 Mol Ecol. 2011;20: 286–298. doi:10.1111/j.1365-294X.2010.04937.x
701 50. Assogba BS, Milesi P, Djogbénou LS, Berthomieu A, Makoundou P, Baba-Moussa LS, et al. The
702 *ace-1* Locus Is Amplified in All Resistant *Anopheles gambiae* Mosquitoes: Fitness
703 Consequences of Homogeneous and Heterogeneous Duplications. PLoS Biol. 2016;14: 1–26.
704 doi:10.1371/journal.pbio.2000618
705 51. Kambhampati S, Rai KS, Burgun SJ. Unidirectional Cytoplasmic Incompatibility in the
706 Mosquito, *Aedes albopictus*. Evolution. 1993. p. 673. doi:10.2307/2410079
707 52. Sinkins SP, Braig HR, O'Neill SL. *Wolbachia pipientis*: Bacterial density and unidirectional
708 cytoplasmic incompatibility between infected populations of *Aedes albopictus*. Exp Parasitol.
709 1995;81: 284–291. doi:10.1006/expr.1995.1119
710 53. Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, et al. Multilocus
711 sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol.
712 2006;72: 7098–7110. doi:10.1128/AEM.00731-06
713 54. Rogers SO, Bendich AJ. Extraction of total cellular DNA from plants, algae and fungi. Plant
714 Molecular Biology Manual. Springer Netherlands; 1994. pp. 183–190. doi:10.1007/978-94-

011-0511-8_12

55. Berticat C, Rousset F, Raymond M, Berthomieu A, Weill M. High *Wolbachia* density in insecticide-resistant mosquitoes. Proc R Soc B Biol Sci. 2002;269: 1413–1416. doi:10.1098/rspb.2002.2022

56. Weill M, Berticat C, Raymond M, Chevillon C. Quantitative polymerase chain reaction to estimate the number of amplified esterase genes in insecticide-resistant mosquitoes. Anal Biochem. 2000;285: 267–270. doi:10.1006/abio.2000.4781

57. Bryant D, Moulton V. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Mol Biol Evol. 2004;21: 255–265. doi:10.1093/molbev/msh018

58. R Core Team. R: A language and environment for statistical computing. http://www.R-project.org/. In: R Foundation for Statistical Computing, Vienna, Austria. 2017.

59. Wickham H. Elegant Graphics for Data Analysis: ggplot2. Applied Spatial Data Analysis with R. 2008.

60. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. Msa: An R package for multiple sequence alignment. Bioinformatics. 2015;31: 3997–3999. doi:10.1093/bioinformatics/btv494

Supporting Information

S1 Fig. *cidA* and *cidB* variants observed in the Slab isofemale line obtained by Nanopore sequencing. For each variant, the two variable regions are shown: the upstream region, named with a Greek letter for *cidA* and a Latin letter for *cidB*, and the downstream region, named with a digit for both genes. *cidA* regions, which have a different presence in the repertoires, are represented in green. *cidB* repertoires are common to all repertoires. Females infected with (β+, 16+) or (β+, 16-) *Wolbachia* are able to rescue males infected with *w*Pip-Istanbul (compatible), whereas females infected with (β-, 16-) *Wolbachia* cannot (incompatible).

740 **S2 Fig. Variations in hatching rates in compatible crosses between Slab females and Istanbul males.**
741 Hatching rates in relation to variations in: A) *Wolbachia* infection level and numbers of copies of B)
742 total *cidA*, C) *cidA-III-beta(2)*, and D) *cidA-III-16*. None of these variations are significantly correlated
743 with hatching rates (Spearman correlation coefficient, P = 0.49, 0.28, 0.94, and 0.16 for A, B, C, and
744 D, respectively).
745
746 **S3 Fig. Variations of *cidA* copy numbers in *Wolbachia* genomes of Slab females.** A), B), and C) show
747 the numbers of copies of total *cidA*, *cidA-III-16*, and *cidA-III-beta(2)* regions, respectively, in females
748 with repertoires (β-,16-), (β+,16-), and (β+,16+). D), E), and F) show the distribution of *cidA, cidA-III-*
749 *16*, and *cidA-III-beta(2)* copy numbers. All histograms show either bi- or trimodal distributions. All
750 copy numbers were quantified using quantitative PCR relative to the single copy *Wolbachia* gene
751 *wsp*.
752
753 **S4 Fig. Variations of *cidB* copy numbers in *Wolbachia* genomes of Slab females**. *cidB* copy numbers
754 equal *cidA* copy numbers. (β-,16-) *w*Pip has around three *cidB* copies, whereas (β+,16-) and (β+,16+)
755 have around five *cidB* copies. All copy numbers were quantified using quantitative PCR relative to the
756 single copy *Wolbachia* gene *wsp*.
757
758 **S5 Fig. PCR primers used to analyze sequence variation in *cid* genes.** PCRs amplifying the full *cidA*
759 gene (burgundy arrows), variable regions of the *cidB* gene (green arrows), and the *cidAB* tandem
760 (purple arrows) were used for nanopore sequencing of repertoires. PCRs specific to *cidA-III-beta(2)*
761 (yellow) and *cidA-III-16* (orange) were used both in standard PCR and qPCR to determine the
762 presence/absence of these regions.
763
764 **S6 Fig. Sequence variation at the CidA-CidB interaction interface II.** Protein sequences are shown for
765 the span of the interaction interface II (position of the first and last aminoacids of this interface were

766 obtained thanks to Wang & Hochstrasser, pers. com.). Polymorphic positions are indicated with a
767 gray background. This interface is polymorphic in CidB, with *w*Pip-IV CidB showing a different
768 sequence from all other groups (sequences shown here are representative of all sequenced variants).
769 By contrast, CidA variants are identical in this interaction interface in all sequenced variants
770 regardless of their groups.
771
772 **S1 Table. *cid* repertoires of the Istanbul isofemale line.** Nanopore sequencing of *cid* PCR products
773 gave identical repertoires for four distinct Istanbul individuals.
774
775 **S2 Table. Crosses between Slab females and Istanbul males.** For each female, we determined the
776 hatching rate (HR) (i.e., ratio between hatched larvae and eggs). The specific *cidA-III-beta(2)*
777 upstream and *cidA-III-16* downstream regions were quantified in each female and normalized over
778 the single copy *Wolbachia* gene *wsp*. The total *cidA* and *cidB* copy numbers were also quantified and
779 normalized over *wsp*. Infection rates were assessed by quantifying the single copy gene *wsp* and
780 normalizing it on the *Cx. pipiens* single copy gene *ace2* (Berticat et al, Proc. R. Soc. Lond. B, 2002).
781
782 **S3 Table. Crosses between Slab females from each of the six newly established isofemale sublines**
783 **(two (β+,16+), two (β+,16-), and two (β-,16-)) and Istanbul males.** The hatching rates and number of
784 hatched larvae are shown for each female. Eggs were only counted in compatible crosses involving
785 (β+,16+) and (β+,16-) females. All females were screened for the presence/absence of two regions of
786 the polymorphic gene *cidA* (*cidA-III-beta(2)* and *cidA-III-16*) to confirm that they had the same
787 repertoire as the female from which the isofemale subline was founded. NF = non-fertilized.
788
789 **S4 Table. Reciprocal crosses between (β+,16+) and (β-,16-) isofemale sublines**. Number of eggs and
790 larvae and resulting hatching rates (HR). NF = non fertilized.
791

792 **S5 Table. PCR primers and programs used in this study.** qPCR programs are described in the main

793 text.

794

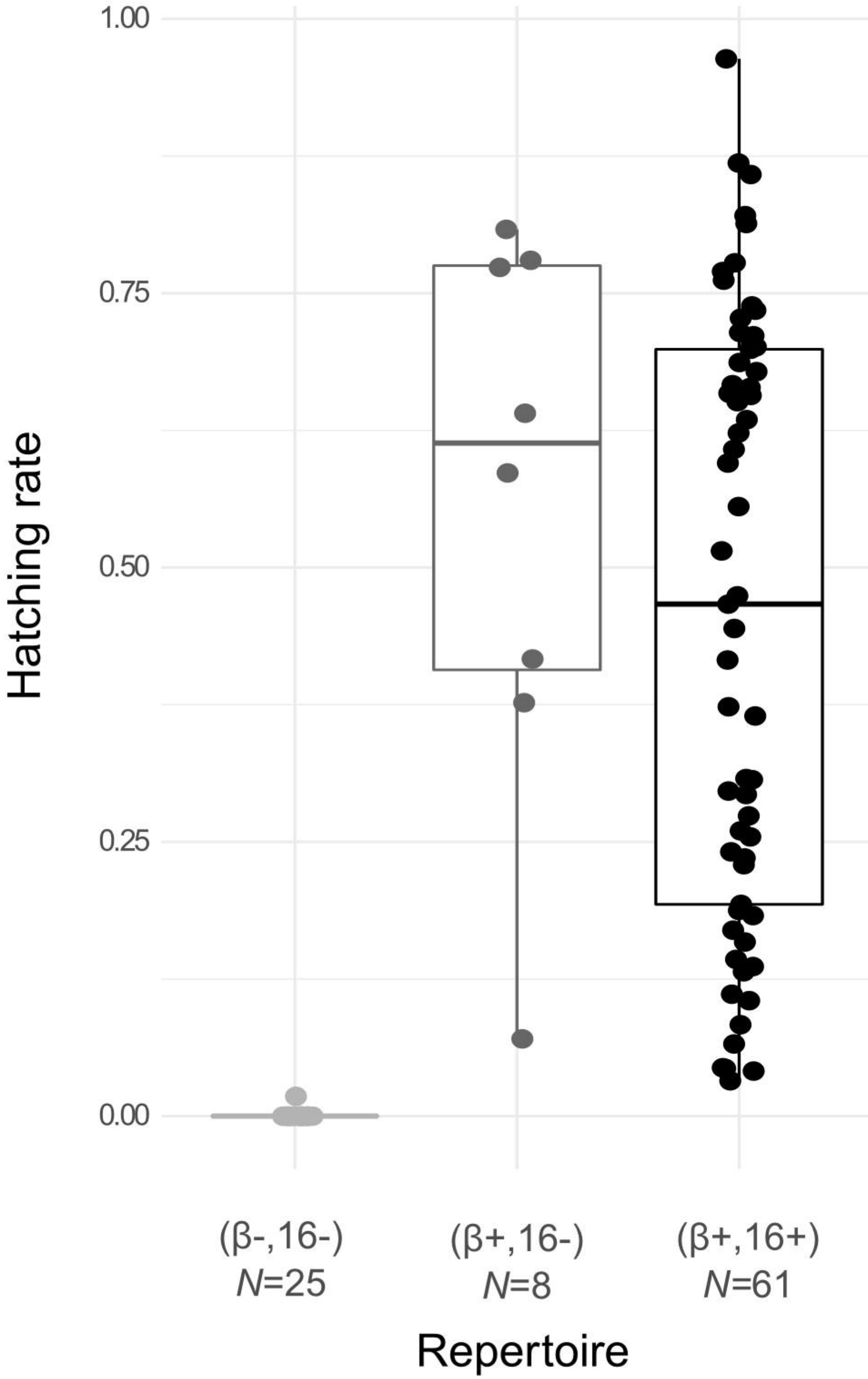
795 **S6 Table. Correspondence between variant names in the previous (Bonneau et al, Nat Com 2018)**

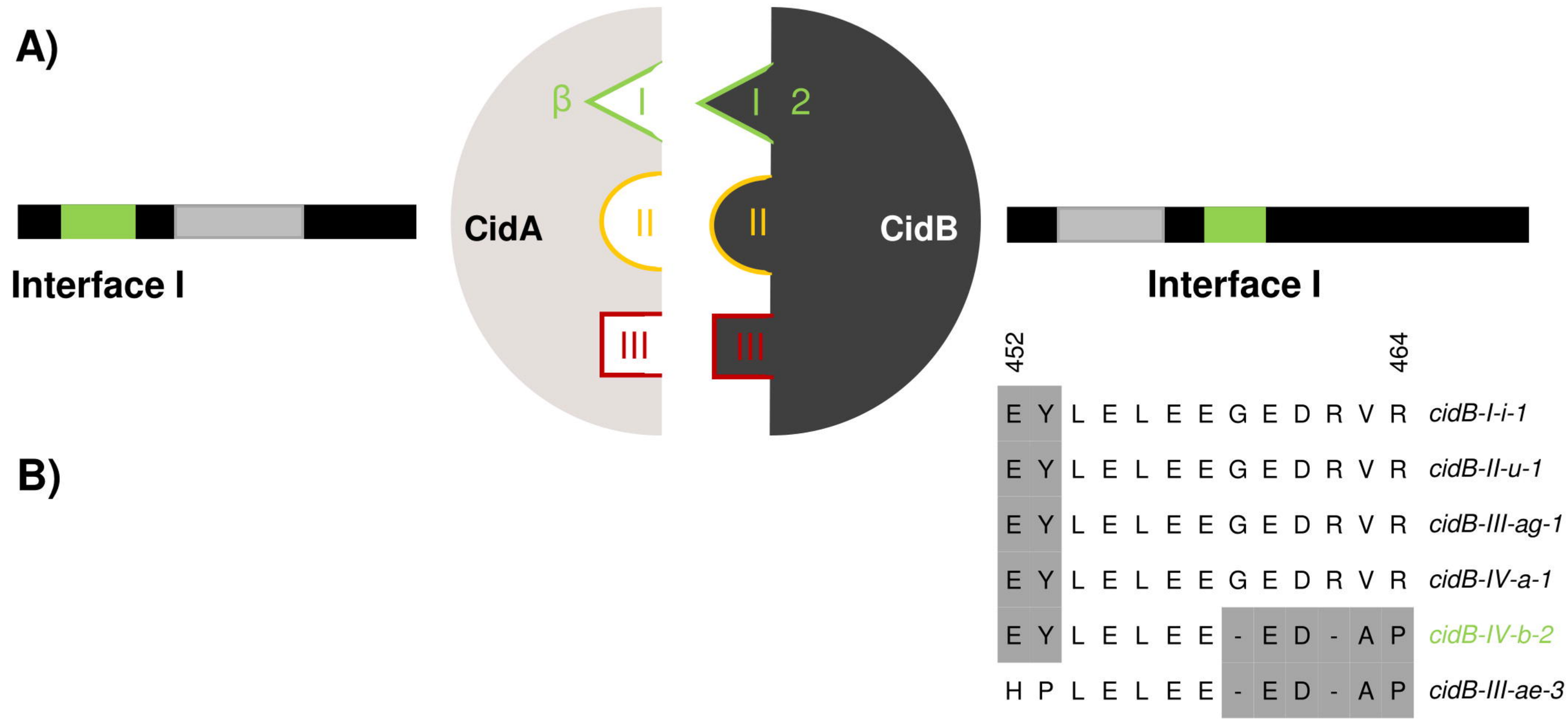
796 **and current nomenclature.**

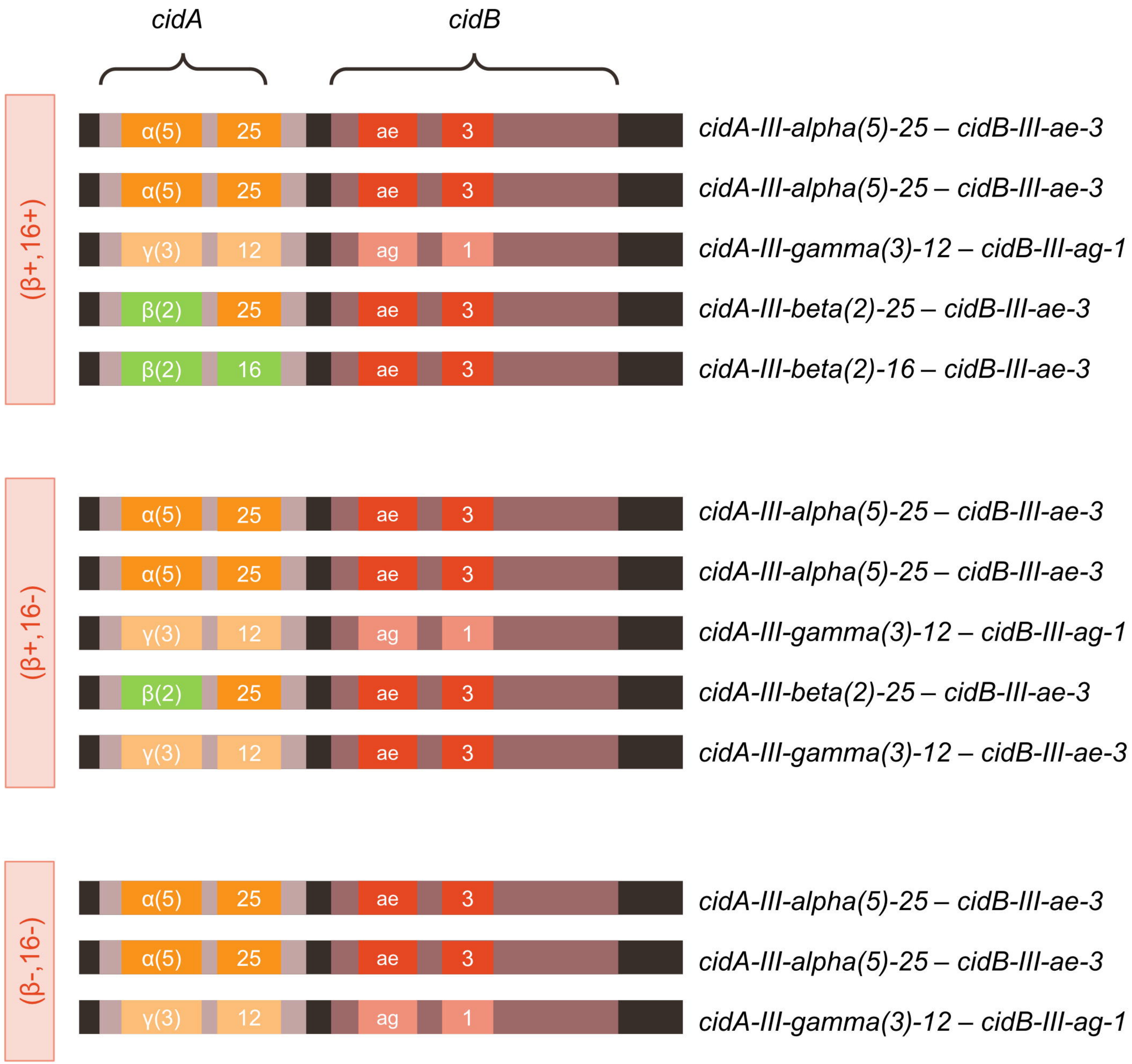
797

798 **S7 Table. GenBank accession numbers of new *cidA* and *cidB* variants used in this study.**

799


1.00
0.75
0.50
0.25
0.00
Hatching rate
(β-,16-)
N=25
(β+,16-)
N=8
(β+,16+)
N=61
Repertoire


A)
β
I
I
2
CidA
II
II
CidB
III
III
Interface I
Interface I
452
464
E Y L E L E E G E D R V R cidB-I-i-1
E Y L E L E E G E D R V R cidB-II-u-1
E Y L E L E E G E D R V R cidB-III-ag-1
E Y L E L E E G E D R V R cidB-IV-a-1
E Y L E L E E - E D - A P cidB-IV-b-2
H P L E L E E - E D - A P cidB-III-ae-3
B)
118
152
cidA-III-α(5)-25 A F E S Y K N G F S S D E L D S R G N P I T D E Y I R N S L L G G I R
cidA-III-β(2)-16 R F E S Y R N K F P P G A F D G K T E T E F G K Y V R N S L L G G I R
cidA-III-β(2)-25 R F E S Y R N K F P P G A F D G K T E T E F G K Y V R N S L L G G I R
cidA-III-γ(3)-12 R F E S Y R N K F P P G A F D G K T E T E F G K Y V R N S L L D S I K
C)
cidA-IV-gamma(5)
cidA-III-gamma(3)
cidA-II-gamma(3)
cidA-I-alpha(1)
cidA-I-alpha(2)
cidA-I-gamma(2)
cidA-I-gamma(1)
cidA-II-alpha(5)
cidA-III-alpha(5)
cidA-IV-alpha(5)
cidA-III-beta(2)
cidA-IV-beta(1)
cidA-IV-delta(1)
cidA-IV-beta(3)
0.01
cidB-I-3
cidB-II-3
cidB-III-3
cidB-IV-3
cidB-I-1
cidB-II-1
cidB-III-1
cidB-IV-1
cidB-IV-2
0,001


cidA
cidB
(β+,16+)
α(5) 25 ae 3
cidA-III-alpha(5)-25 – cidB-III-ae-3
α(5) 25 ae 3
cidA-III-alpha(5)-25 – cidB-III-ae-3
γ(3) 12 ag 1
cidA-III-gamma(3)-12 – cidB-III-ag-1
β(2) 25 ae 3
cidA-III-beta(2)-25 – cidB-III-ae-3
β(2) 16 ae 3
cidA-III-beta(2)-16 – cidB-III-ae-3
(β+,16-)
α(5) 25 ae 3
cidA-III-alpha(5)-25 – cidB-III-ae-3
α(5) 25 ae 3
cidA-III-alpha(5)-25 – cidB-III-ae-3
γ(3) 12 ag 1
cidA-III-gamma(3)-12 – cidB-III-ag-1
β(2) 25 ae 3
cidA-III-beta(2)-25 – cidB-III-ae-3
γ(3) 12 ae 3
cidA-III-gamma(3)-12 – cidB-III-ae-3
(β-,16-)
α(5) 25 ae 3
cidA-III-alpha(5)-25 – cidB-III-ae-3
α(5) 25 ae 3
cidA-III-alpha(5)-25 – cidB-III-ae-3
γ(3) 12 ag 1
cidA-III-gamma(3)-12 – cidB-III-ag-1