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

Short title: Molecular bases of CI patterns emergence

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

The mosquitoes of the *Culex pipiens* complex are worldwide vectors of arboviruses, filarial nematodes and avian malaria agents. In these hosts, the endosymbiotic bacteria *Wolbachia* induce cytoplasmic incompatibility (CI), a reduced embryo viability in so-called incompatible crosses. The *Wolbachia* infecting *Culex* (*w*Pip) cause CI patterns of an unparalleled complexity, associated with the amplification and diversification of their *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 multi-genes families to acquire *cid* genes, we show that some *w*Pip genomes lost specific gene copies, giving rise to several sub-lineages segregating within a cage. Linking phenotypic changes with their underlying genotypic bases, we show that gene copies which are key for CI phenotypes originated from recombination, rather than point mutations. We show how new CI patterns can emerge as a two-step process, with (i) local changes in CI genes, while maintaining compatibility with surrounding lines (all sub-lineages being mutually compatible) and (ii) reveal of those incompatibility by migration and exposure to hosts infected with incompatible lines.

Keywords

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

Introduction

Wolbachia are maternally-transmitted endosymbiotic bacteria which infect up to 50% of arthropod species [1–3]. These bacteria are well known for their wide range of reproductive manipulations in arthropods. Their major manipulation is cytoplasmic incompatibility (CI), a reduction in embryos' hatching rates in crosses between infected males and uninfected females (in its simplest form). CI results from a *Wolbachia*-induced *modification*, perturbing embryos' first division, that infected females are able to *rescue* [4,5]. CI is thus formalized in a global *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, where *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]. A first analysis of *w*Pip genetic diversity in *Culex* identified 5 phylogenetic groups (named *w*Pip-I to *w*Pip-V, [11]). Confronting *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 very few exceptions, compatible.

In 2013, putative CI genes were identified by a combination of proteomics and genetics approaches [12]. It was four years later, in 2017, that 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, with a central role in CI, although precise molecular mechanisms of CI are still to be determined. Two main alternative models are still debated: a host-modification (HM) model, stating that toxicity occurs through modifications of the host, and rescue through a reversal of these modifications; and a toxin-antidote (TA) model, in which the sperm carries a *Wolbachia* factor which is toxic, unless there is the appropriate *Wolbachia* antidote in the egg that binds and neutralizes the toxin [15–18]. In *Culex,* the complexity of the crossing patterns makes the TA model far more likely: (in)compatibility would simply be based on different toxin-antidote binding abilities [16,19–21].

61 Recent publications all agree with *cifA* being the rescue (resc, or antidote) gene, and *cifB* 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, discriminating
64 them into 5 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 bears a deubiquitinase domain (the tandem is thus also
66 called *cid*; *d* standing for deubiquitinase domain), and clade IV *cif,* in which *cifB* bears a nuclease
67 domain (thus called *cin*) [13,14]. Here, we use the functional-based nomenclature [19].

68 Before knowing the *cif* genes, models based on *Culex* crossing experiments predicted that several
69 factors or pairs of *mod*/*resc* genes were required to encode the complex crossing patterns induced
70 by *w*Pip [9,26]. Investigations on *cif* (*cid* and *cin*) present in *w*Pip genomes highlighted that *cin* genes
71 were always monomorphic while the *cid* genes were amplified and diversified in all sequenced *w*Pip
72 genomes [27], with up to 6 different copies of each gene within a single *Wolbachia w*Pip genome.
73 The different *cidA/cidB* copies in a given *w*Pip genome are called 'variants' and full set of variants
74 constitutes the '*cid* repertoire'. In *w*Pip genomes, polymorphism of both *cidA* and *cidB* genes are
75 located in two specific regions, which were named 'upstream' and 'downstream' regions that were
76 predicted to be involved in CidA-CidB specific interactions [27]. This has been confirmed by the
77 recently obtained structure of a CidA-CidB co-crystal: out of three interaction regions identified, two
78 perfectly match the previously identified 'upstream' and 'downstream' regions for both CidA and
79 CidB [20,21]. In addition to showing that CidA and CidB from *w*Pip bind together, this recent study
80 also showed that the different CidA-CidB variants have different binding abilities, further fueling the
81 toxin-antidote model for *w*Pip CI complexity [21].

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

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

95 CI patterns in *Culex* have also been shown to change over a few host generations [38]. Yet, these
96 changes were described before the discovery of the *cif* genes. Here, we observed a change in the CI
97 pattern between two isofemale lines maintained in our laboratory since 2005: Slab (*w*Pip III) and
98 Istanbul (*w*Pip IV) [10]. While crosses between Slab females and Ist males were compatible for 12
99 years (2005-2017), we observed fully incompatible crosses for the first time in 2021. Such CI shift was
100 an opportunity to study the underlying genetic basis of CI evolution in laboratory controlled
101 isofemale lines.

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

112

Results

From CI phenotype shift to underlying genomic bases

In 2021, for the first time, crosses between Slab females and Istanbul (Ist) males highlighted that a part of the Slab females were now fully incompatible with Istanbul males.

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

In order to decipher the genomic changes that led to this recent CI shift, we sequenced the *cidA* and *cidB* repertoires present in the *Wolbachia* infecting 4 individuals from the Ist isofemale line and 8 females from the Slab isofemale line (4 compatible and 4 incompatible), using Nanopore Technologies sequencing of *cidA* and *cidB* 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. toxins) diversity, composed of *cidB-III-ae3* and *cidB-III-ag1* variants, while *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,* but 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

In order to determine if compatibility variations were indeed 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 replicate 1, 2 and 3 respectively) and studied their hatching rates (HR, number of hatched larvae over the number of eggs laid). Eight females were determined as having produced unfertilized rafts and excluded from

138 further analyses. For the remaining 94 females, HRs ranged from 0% (fully incompatible raft) to
139 96.4%, highlighting strong differences in females' rescuing abilities. Out of these 94 females, 25 (i.e.
140 27%) were fully incompatible with Ist males.
141 Since repertoires differed in the presence/absence of two specific regions of *cidA* repertoires, we
142 designed primers and specifically quantified the 94 females (S2 Table). The PCR assay confirmed that
143 3 distinct *Wolbachia cidA* repertoires coexisted within the Slab isofemale line: (i) individuals hosting
144 *Wolbachia* bearing both the upstream '*beta(2)*' and the downstream '*16*' regions, named (β+,16+);
145 (ii) individuals hosting *Wolbachia* bearing only the upstream region *beta(2)*, but not the downstream
146 region *16*, named (β+,16-); and (iii) individuals hosting *Wolbachia* genomes bearing none of these
147 two regions, named (β-,16-) (S2 Table). For ease of reading, we will metonymically qualify
148 mosquitoes based on their *cidA* repertoires (β+,16+), (β+,16-) or (β-,16-). Out of the 94 females
149 analyzed, 61 were (β+,16+), 8 (β+,16-) and 25 (β-,16-) (Fig 1).
150
151 **Fig 1. Intra-line polymorphism in *cidA* repertoires matches compatibility patterns.** (β-,16-) females
152 have neither *cidA-III-beta(2)* nor *cidA-III-16* regions; (β+,16+) females have both and (β+,16-) females
153 have the *cidA-III-beta(2)* region but not the *cidA-III-16* one. Presence or absence of *cidA-III-beta(2)*
154 and *cidA-III-16* regions was determined by specific qPCRs. *N* = Number of individuals.
155
156 All but one (β-,16-) females produced egg-rafts with no hatching (the only one for which larvae
157 hatched corresponded to a HR of 1.8%). Such fully incompatible crosses were only observed with (β-
158 ,16-) mothers, (β+,16+) and (β+,16-) mothers being compatible (to various extents) with Ist males (Fig
159 1).
160

cidA repertoire variations does not explain HR variability in compatible crosses

While it was long believed that CI in *Culex* was a "yes or no" phenomenon, a recent study explored HR variations in compatible crosses involving *w*Pip from different groups [40]. Here, looking at compatible crosses involving (β+,16+) or (β+,16-) Slab females and Ist males, we also found strong variations in HRs which ranged 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, with 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; and ρ = -0.17, p = 0.16 for the total *cidA*, *cidA-III-beta* and *cidA-III-16* respectively, S2B-D Figs, S2 Table).

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

Recent works disclosed 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 for 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) displays an original combination of aminoacids, possibly issued from a recombination between the CidA-alpha and CidA-gamma regions. Building a phylogenetic network of all described *cidA* upstream regions [22,27,40,41] confirmed that *beta*-like regions were rare (only found in *w*Pip group IV and in Slab to date) and issued from a recombination between the regions *cidA-gamma*-like and *cidA-alpha*-like, these last two regions being frequent and

186 found in all *w*Pip groups surveyed so far (Fig 2B). We also showed that the *cidA-delta* region was
187 recombinant.
188
189 **Fig 2. Matching recombination in a CidA-CidB interaction interface I is key for rescue.** (A) CidA and
190 CidB interact head to tail through three distinct interfaces, shown here in green, yellow and red
191 respectively. (B) The regions *cidA-III-beta(2)* and *cidB-IV-2*, key to in the (in)compatibility between
192 Slab and Istanbul isofemale lines, are both exposed in the interaction interface I and shown in green.
193 Alignments of the protein sequences of the interaction interface I are shown for *cidA* and *cidB* (the
194 sequence between the first and the last variable aminoacid belonging to interface I is represented;
195 Wang & Hochstrasser, pers. com.). (C) Network phylogenetic analysis of the *cidA* upstream and *cidB*
196 downstream region, using the Neighbor-net method on variants from [27,40,41] and this study. For
197 both regions, there are two major groups of variants (alpha/gamma for *cidA* and 1/3 for *cidB*), with
198 few intermediate recombinant variants (in green). Variants used in the present study are shown in
199 light green. Each edge (or set of parallel edges) corresponds to a split in the dataset and has a length
200 equal to the weight of the split.
201
202 The downstream region CidB-IV-2, present in *w*Pip-Ist, had previously been associated with
203 incompatibility when crossed with a wide range of females [22]. This region interacts with *cidA-beta*
204 at the interface I (Fig 2A) and also displays a pattern that could be explained by a recombination
205 (between CidB-1 and CidB-3 downstream regions) (Fig 2B). When looking at wider phylogenetic
206 scales, relying on all the previously published sequences, we found a pattern even clearer with only
207 two downstream CidB-1 and CidB-3 regions, identical among *w*Pip groups and CidB-IV-2 being the
208 only intermediate, recombinant region (Fig 2C).
209

cid repertoires evolution

The different *cidA* repertoires segregate into sublines

Since a new CI pattern emerged between 2017 and 2021, a closer look at Slab individuals' repertoires transmission and variation was required to better understand the precise nature of this evolution. For each of the three *cidA* repertoires found in the original isofemale Slab 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 females of these new isofemale sublines had the same CI phenotype as their respective founding females when crossed with Ist males (S3Table). We also checked that they were compatible with each other in both directions showing that *cidA* repertoire polymorphism among sublines did not create incompatibilities (S4 Table).

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

We also genotyped individuals from the Slab isofemale line kept in liquid nitrogen since 2005, *i.e.* shortly after the isofemale line establishment in our insectaries. Out of 30 tested mosquitoes, there were 26 (β+,16-), 4 (β+,16+) but none of them was (β-,16-), in accordance with the hypothesis that this genotype, leading to 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-) (S3A Fig). We also quantified the number of *cidB* copies in 8-10 females of each repertoire and found that *cidB* copy numbers equal *cidA* copy numbers: 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 variants' identity, they differ in terms of *cidB* copy numbers.

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

Multiple copies of the same variant in the repertoires

qPCR results also 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 within a given repertoire. By combining the qPCR data and relative reads coverage information from Nanopore sequencing (e.g. a 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 found in Slab. We found that (β+,16+) individuals, which have a total of 5 *cidA* copies, had 2 copies of *cidA-III-alpha(5)-25*, and a single copy of each of the three other variants (*cidA-III-beta(2)-16*, *cidA-III-beta(2)-25* and *cidA-III-gamma(3)-12*). (β+,16-) individuals also had 5 *cidA* copies, out of which there were 2 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 3 *cidA* copies, had 2 copies of *cidA-III-alpha(5)-25* and a single copy of *cidA-III-gamma(3)-12* (Fig 3).

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

Regions colored in green are the key studied regions: *cidA-III-beta(2)* and *cidA-III-16.* 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 of *cidA-III-beta(2)* and *cidA-III-16*: quantifying those two markers in 89 single Slab females, we confirmed that (β+,16+) repertoires had a single copy of *cidA-III-16* while other repertoires had none (S3B and S3E Figs). For *cidA-III-beta(2)*, we found that (β+,16+) individuals had 2 copies (corresponding to 1 copy of *cidA-III-beta(2)-16* and 1 copy of *cidA-III-beta(2)-25*), while (β+,16-) individuals had a single copy (unique copy of *cidA-III-beta(2)-25*), and (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 done by acquiring the sequences of *cidA* and *cidB* variants separately [22,27,39–41]. Here, we also amplified and Nanopore sequenced a region including both *cidA* and *cidB* variable regions in a single amplicon (S5 Fig). We were thus able to reconstruct the *cidA-cidB* tandems associated in the genome, resulting in the complete repertoires presented in Fig 3. In order to know if all *cid* variants were associated in tandems or if they could stand alone in the genome, we compared their relative coverage when amplified and sequenced alone *vs* amplified and sequenced in *cidA-cidB* tandems acquisitions. These relative coverages being similar, we considered that all *cidA* and *cidB* variants were always located within tandems. Thus, (β+,16+) and (β+,16-) sublines have 5 *cidAB* tandems while (β-,16-) lines have 3 *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. (β+,16+) and (β+,16-) individuals have in addition one *cidA-III-beta(2)-25_cidB-III-ae3* tandem; but they differ by their fifth tandem: *Wolbachia* in (β+,16+) individuals have a specific *cidA-III-beta(2)-16_cidB-III-ae3* while those in (β+,16-) individuals have one *cidA-III-gamma(3)-12_cidB-III-ae3* instead

(Fig 3). Our results show that the *cidB* variant *"cidB-III-ae3"* associates 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* genes family: numerous horizontal transfers have been described, resulting in a strong incongruence between *cif* genes phylogeny and *Wolbachia* phylogeny [23,24,28]. However, until now, a single set of studies had linked *Wolbachia* genotype with rapid changes in phenotypes: the Octomom gene region in *w*Mel and *w*MelPop genomes [35–37]. While rapid shifts in CI patterns were previously described in *Culex* mosquitoes infected with *w*Pip [38], the *cif* genes were not described at that time, preventing to study the genetic bases of those phenotypic changes. Furthermore, the impact of contamination in such shifts could not be ruled out. Here, we could link changes in CI patterns with their *Wolbachia* genomic bases for the first time.

We detected a recent shift in CI pattern between Slab (*w*Pip-III) and Ist (*w*Pip-IV) lines: while they were 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 either from a variation in Ist males' *cidB* (toxins) repertoires males or from a variation in Slab females' *cidA* (antidotes) repertoires. 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 5 *cidA* copies, named (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-), unable to rescue these modifications. These results indicate that some *w*Pip genomes within the Slab isofemale line have lost *cidA* variants. Indeed, 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 where compatible, and 2021, when the shift in CI was detected.

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

Rapid evolution of *cid* repertoires through gene losses

In 2017, approximately a quarter of the 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 5 copies, and very rare intermediate 4-copy repertoire. These rare intermediates are likely due to qPCR variability. In this case, it could either mean that (i) 4-copy repertoires were a transient state that disappeared quickly, due to e.g. a lower fitness or drift; or (ii) that 4-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 putative 4-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 those two tandems are 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]. It is only once *cifB* have lost its function that *cifA* (*resc* factors, or antidotes) loss would be possible [48]. In such conceptual framework, losing a *cifA* without losing a *cifB* would lead to self-incompatibility, being an immediate

337 evolutionary dead-end. Independently from CI, the presence of *cidB* without an associated *cidA*
338 might be toxic for the host, as shown recently in insect cells [4]. Such predictions have been
339 corroborated by a large-scale phylogenetic study of *cif* genes, indicating that *cifB* genes were often
340 disrupted, while *cifA* were not [24]. However here, at a microevolutionary scale, we observed that it
341 was possible to decrease qualitatively the diversity in the *cidA* variants without changing qualitatively
342 the *cidB* variants: all (β+,16+), (β+,16-) and (β-,16-) repertoires described share the same *cidB*
343 variants (*cidB-III-ag1* and *cidB-III-ae3*), while they have respectively four, three or two distinct *cidA*
344 variants (Fig 3). Such an apparent discrepancy between predictions and our results is explained by
345 the fact that *w*Pip genomes contain several *cid* genes. Unlike the other systems studied to date,
346 where loss of *cifA* can lead to a loss of auto-rescuing ability, *w*Pip displays several *cidA* variants,
347 making possible to lose one and still maintain compatibility with surrounding mosquitoes. Loss of
348 *cidA* variants required for intra-line compatibility may also occur, but mosquitoes having lost such
349 key antidotes would be quickly wiped out by selection, an hypothesis corroborated by the mutual
350 compatibility between all newly-established isofemale lines, regardless of their *cid* repertoire.
351 Laboratory conditions could potentially facilitate such evolution by qualitative reduction of the *cidA*
352 repertoire. Indeed, in natural populations, multiple *cid* repertoires coexist (as shown by sequencing
353 and revealed by numerous crosses [22,27,41,49]), meaning that individuals are exposed to a wide
354 range of toxins. A greater diversity in the *cidA* repertoire enables to rescue a wider range of males
355 and will thus be under positive selection. While *cidA* diversity is advantageous in natural populations,
356 in a cage with an isofemale line, females are exposed to a restricted and fixed set of toxins. Loss of
357 *cidA* variants which are not strictly required to rescue these toxins will not be counter-selected.
358 Furthermore, numerous gene copies can be costly (e.g. [50]), loss of *cidAB* tandems could thus confer
359 a higher fitness to the *Wolbachia* with lower numbers of copies and participate to the fixation of new
360 *Wolbachia* genomes in mosquitoes, explaining the rapid *Wolbachia* turnover. It is conceivable that
361 the 3-copy repertoire of (β-,16-) *Wolbachia*, containing two distinct variants of *cidA* and two distinct

variants of *cidB*, is the minimum repertoire viable configuration, containing only pairs of matching toxins and antidotes.

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

A plausible explanation for the observed shift in CI could have been that Slab is co-infected 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 those strains, as observed for instance in *Aedes albopictus* [51,52]. At first, using common MLST markers, authors found no variability at all among *w*Pip strains [53], showing that if coinfection occurs, it must be between highly similar strains. The development of a *w*Pip-specific MLST enabled to uncover *w*Pip polymorphism and to describe five monophyletic groups within *w*Pip [11]. Yet, while these markers highlighted *w*Pip polymorphism, no sign of coinfection was found either, every mosquito clearly hosting a single one of the five groups [11]. Here, working on *cid* genes, highly variable even within a *w*Pip group, we found no clear signs of coinfections either. We quantified several *cid* markers, and analyzed their values distribution. Had there been coinfections of two *Wolbachia* strains with respectively three and five *cid* copies, quantifications 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, quantifications displayed a bimodal distribution (S3D-F Figs), with few intermediate values certainly due to technical variation in qPCR measures. While coinfection had to happen at some point (all the *Wolbachia* within an individual cannot be wiped out and replaced by another *Wolbachia* instantly), it had to be brief with no lasting coinfection within a single host individual. This could be explained by two alternative hypotheses: (i) an exclusion between the two distinct genotypes (3 and 5-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 is not explained by females' *cid* repertoires

CI has long been described as a 'yes or no' phenomenon in *Cx pipiens*, meaning that crosses were either compatible or fully incompatible [9]. Yet, having a closer look at hatching rates, we recently showed that they could be highly variable in so-called compatible crosses [40]. We further showed that reduced hatching rates were associated with cytological signs of canonical CI [40]. These phenomena were suggested to be linked with *cid* repertoires, as males infected with *w*Pip-IV, having distinct *cidB* variants from other *w*Pip groups, induced more intermediate CI than other males.

Here, 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 hatching rates did not correlate with any of the *cidA* markers studied, nor with *Wolbachia* infection rate.

Yet, 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 differs 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. Such lack of match at interface II could result in a more labile CidA-CidB interaction leading to intermediate HR values.

Recombination seems key to the emergence of new CI pattern

A specific *cidB* region, named *cidB-IV-2*, was previously correlated with strong CI-inducing abilities 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 2B).

Furthermore, it was previously shown that the only females able to rescue the modifications induced by Ist males were females infected with *w*Pip-IV and Slab females (*w*Pip-III, studied here) [22,27]. We found that the rescuing ability of Slab females depended on the presence of a specific *cidA* upstream region, *cidA-beta*, all females having this region being able to rescue Ist males, while females having lost this region were not. This variant was also generated by a recombination between two main groups of upstream regions shared by all *w*Pip groups. This time, there were two types of recombinants, the *cidA-beta* regions and the *cidA-IV-delta* region (Fig 2B). When looking at the repertoires found in females infected with *w*Pip-IV, which can all rescue Ist males, we found that all of them had one recombinant region in their repertoire (either *cidA-beta* or *cidA-IV-delta*). This means that all females able to rescue the toxicity of the recombinant region *cidB-IV-2* have a recombinant antidote in their repertoire. Since the upstream *cidA* region (e.g. *beta* or *delta*) interacts with the downstream *cidB* region (e.g. *cidB-IV-2*) in the interaction interface I [21], this finding suggests a recombinant-for-recombinant binding (Fig 2), in which recombinant regions can only bind similarly recombinant regions. New CI phenotypes could thus emerge through the recombination of existing variants.

In conclusion, we found that CI evolved in less than four years within a *Culex* isofemale line. Such evolution is linked to the loss of some *cidA* (antidotes) variants in their *cid* repertoire. Our results suggest that incompatibilities could emerge as a two-step process: local changes in CI repertoires,

while maintaining compatibility with surrounding mosquitoes, then migration and secondary contact with incompatible lines. Indeed, here, we showed that all repertoires present within a cage were mutually compatible: changes in repertoires led to a change in CI pattern only when crossed with an external line. We demonstrated that new compatibility types can emerge rapidly, as a result of the quick evolution of *Wolbachia* genomes. A common point in the few studies addressing *Wolbachia* microevolution is that both Octomom in *w*Mel [35–37], and *cid* genes in *w*Pip are amplified genomic regions, suggesting a strong role of gene amplification in the speed of *Wolbachia* genome evolution. Such rapid evolution has to be considered when releasing *Wolbachia*-infected mosquitoes, inducing CI, for vector control purposes.

Materials and methods

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

Isofemale lines used in this study

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 have been established at the lab in 2005.

We also used 6 newly-established isofemale (sub)lines: 2 "(β+,16+)" lines, bearing both the *cidA-III-beta* and the *cidA-III-16* regions; 2 "(β+,16-)" lines bearing only the *cidA-III-beta* marker, and 2 "(β-,16-)" lines bearing none of those markers. Briefly, for isofemale line establishment, a cage containing the Slab isofemale line was blood fed. After five 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 2 most numerous progenies were kept 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 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 five days in cages, females were fed a blood meal and five days later, individual females were isolated for egg-laying. This enabled to have a precise correspondence between a female and its egg-raft. Egg-rafts were deposited into 24-well plates, a picture of the eggs was taken using a Leica camera, and eggs were 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 checked with a binocular magnifier to ensure that eggs were fertilized, and that null HR resulted from CI rather than unfertilization.

Repertoire acquisition through Nanopore sequencing of PCR products

We acquired the *cidA* and *cidB* repertoires of 8 individuals from the Slab isofemale line, 4 individuals from the Ist isofemale line and of 2 individuals per newly-established isofemale line, following the pipeline described in [39]. We also acquired *cidAB* combinations for all newly established isofemale lines. Briefly, DNA was extracted from the individuals. The target sequence was amplified using the adequate primer set (S5 Table, S6 Fig). PCR products were then sequenced by MGX Platform (Montpellier) by Nanopore sequencing, 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, then combinations between all *cidA* and *cidB* were artificially created and used as a reference base to run the pipeline.

Wolbachia genotyping

Classic PCR

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

Quantitative PCR

Real-time quantitative PCR were run using the LightCycler 480 system (Roche). All DNA samples were analyzed in triplicates, for each quantification used. 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* regions 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 could 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 by 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

3 min at 95°C, 45 cycles of 95°C for 3 sec, 62°C for 10 sec, and 72°C for 15 sec. Primers concentration were respectively 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, if there are three *cidA* copies and if, out of the two variants present, the coverage of the first one is twice the coverage of the second, it is concluded that there are two copies of the first one and one copy of the second.

Grouping analysis of upstream/downstream regions of variants

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

Nomenclature of *cid* variants

The nomenclature established by [27] was updated, since it could result in identical sequences being called differently in different *w*Pip groups. The basics of the previous nomenclature were kept: variants names are composed of the name of their upstream region (in Greek letters for *cidA* and in Latin letters for *cidB*) and of 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 showed there were

4 clusters of *cidA* upstream regions (Fig 2B), regions within a cluster differing by at most 3 SNPs. We renamed regions so that all regions belonging to the same cluster were named by the same Greek letter (e.g. all named alpha). When SNPs were present, to distinguish sequences, we introduced a number between brackets [e.g. 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 are found.

In all variants names, the group in which the variant is found is mentioned [e.g. cidA-III-alpha(5)-25) is found in a *w*Pip-III].

The correspondence between previous names and names with this new nomenclature is given in S6 Table.

Statistical analyses and graphical representation

All statistical analyses were run using R version 4.2.0 [58]. Correlation analyses were done using a Pearson correlation test. All boxplots and histograms were done using ggplot2 [59], and sequence representation was done 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 and accession numbers are provided in S7 Table. The authors declare that all other data supporting the findings of this study are available within the article and Supporting information.

Acknowledgments

We thank Michael Turelli and Nicole Pasteur for thorough reading and helpful comments on the manuscript; and Wei Wang and Mark Hochstrasser for inputs on interface residuals. Sequencing data were generated on 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.

578 doi:10.2307/2405682

579 8. Laven H. Speciation and evolution in *Culex pipiens*. In: Wright J, Pal R, editors. Genetics of
580 Insect Vectors of Disease. Elsevier: Amsterdam; 1967. pp. 251–275.

581 9. Atyame CM, Labbé P, Dumas E, Milesi P, Charlat S, Fort P, et al. *Wolbachia* divergence and the
582 evolution of cytoplasmic incompatibility in *Culex pipiens*. PLoS One. 2014;9: e87336.
583 doi:10.1371/journal.pone.0087336

584 10. Duron O, Bernard C, Unal S, Berthomieu A, Berticat C, Weill M. Tracking factors modulating
585 cytoplasmic incompatibilities in the mosquito *Culex pipiens*. Mol Ecol. 2006;15: 3061–3071.
586 doi:10.1111/j.1365-294X.2006.02996.x

587 11. Atyame CM, Delsuc F, Pasteur N, Weill M, Duron O. Diversification of *Wolbachia*
588 endosymbiont in the *Culex pipiens* mosquito. Mol Biol Evol. 2011;28: 2761–2772.
589 doi:10.1093/molbev/msr083

590 12. Beckmann JF, Fallon AM. Detection of the *Wolbachia* protein WPIP0282 in mosquito
591 spermathecae: Implications for cytoplasmic incompatibility. Insect Biochem Mol Biol. 2013;43:
592 867–878. doi:10.1016/j.ibmb.2013.07.002

593 13. Beckmann JF, Ronau JA, Hochstrasser M. A *Wolbachia* deubiquitylating enzyme induces
594 cytoplasmic incompatibility. Nat Microbiol. 2017;2: 1–7. doi:10.1038/nmicrobiol.2017.7

595 14. LePage DP, Metcalf JA, Bordenstein SR, On J, Perlmutter JI, Shropshire JD, et al. Prophage WO
596 genes recapitulate and enhance *Wolbachia*-induced cytoplasmic incompatibility. Nature.
597 2017;543: 243–247. doi:10.1038/nature21391

598 15. Beckmann JF, Bonneau M, Chen H, Hochstrasser M, Poinsot D, Merçot H, et al. The Toxin–
599 Antidote Model of Cytoplasmic Incompatibility: Genetics and Evolutionary Implications.
600 Trends Genet. 2019;35: 175–185. doi:10.1016/j.tig.2018.12.004

601 16. Namias A, Sicard M, Weill M, Charlat S. From *Wolbachia* genomics to phenotype: molecular
602 models of cytoplasmic incompatibility must account for the multiplicity of compatibility types.
603 Curr Opin Insect Sci. 2022;49: 78–84. doi:10.1016/j.cois.2021.12.005

604 17. Shropshire JD, Leigh B, Bordenstein SR, Duplouy A, Riegler M, Brownlie JC, et al. Models and
605 Nomenclature for Cytoplasmic Incompatibility: Caution over Premature Conclusions – A
606 Response to Beckmann et al. Trends Genet. 2019;35: 397–399. Available:
607 https://doi.org/10.1016/j.tig.2019.03.004
608 18. Shropshire JD, Bordenstein SR. Two-By-One model of cytoplasmic incompatibility: Synthetic
609 recapitulation by transgenic expression of *cifA* and *cifB* in *Drosophila*. PLoS Genet. 2019;15: 1–
610 20. doi:10.1371/journal.pgen.1008221
611 19. Beckmann JF, Bonneau M, Chen H, Hochstrasser M, Poinsot D, Merçot H, et al. Caution Does
612 Not Preclude Predictive and Testable Models of Cytoplasmic Incompatibility: A Reply to
613 Shropshire et al. Trends Genet. 2019;35: 399–400. doi:10.1016/j.tig.2019.03.002
614 20. Xiao Y, Chen H, Wang H, Zhang M, Chen X, Berk JM, et al. Structural and mechanistic insights
615 into the complexes formed by *Wolbachia* cytoplasmic incompatibility factors. Proc Natl Acad
616 Sci U S A. 2021;118: 1–10. doi:10.1073/pnas.2107699118
617 21. Wang H, Xiao Y, Chen X, Zhang M, Sun G, Wang F, et al. Crystal Structures of *Wolbachia* CidA
618 and CidB Reveal Determinants of Bacteria-induced Cytoplasmic Incompatibility and Rescue.
619 Nat Commun. 2022;13: 1–11. doi:10.1038/s41467-022-29273-w
620 22. Bonneau M, Caputo B, Ligier A, Caparros R, Unal S, Perriat-Sanguinet M, et al. Variation in
621 *Wolbachia cidB* gene, but not *cidA*, is associated with cytoplasmic incompatibility *mod*
622 phenotype diversity in *Culex pipiens*. Mol Ecol. 2019;28: 4725–4736. doi:10.1111/mec.15252
623 23. Lindsey ARI, Rice DW, Bordenstein SR, Brooks AW, Bordenstein SR, Newton ILG. Evolutionary
624 Genetics of Cytoplasmic Incompatibility Genes *cifA* and *cifB* in Prophage WO of *Wolbachia*.
625 Genome Biol Evol. 2018;10: 434–451. doi:10.1093/gbe/evy012
626 24. Martinez J, Klasson L, Welch JJ, Jiggins FM. Life and Death of Selfish Genes: Comparative
627 Genomics Reveals the Dynamic Evolution of Cytoplasmic Incompatibility. Mol Biol Evol.
628 2021;38: 2–15. doi:10.1093/molbev/msaa209
629 25. Bing XL, Zhao DS, Sun JT, Zhang KJ, Hong XY, Sloan D. Genomic Analysis of *Wolbachia* from

630 *Laodelphax striatellus* (Delphacidae, Hemiptera) Reveals Insights into Its "jekyll and Hyde"
631 Mode of Infection Pattern. Genome Biol Evol. 2020;12: 3818–3831. doi:10.1093/gbe/evaa006
632 26. Nor I, Engelstädter J, Duron O, Reuter M, Sagot MF, Charlat S. On the genetic architecture of
633 cytoplasmic incompatibility: Inference from phenotypic data. Am Nat. 2013;182: E15–E24.
634 doi:10.1086/670612
635 27. Bonneau M, Atyame CM, Beji M, Justy F, Cohen-Gonsaud M, Sicard M, et al. *Culex pipiens*
636 crossing type diversity is governed by an amplified and polymorphic operon of *Wolbachia*. Nat
637 Commun. 2018;9: 319. doi:10.1038/s41467-017-02749-w
638 28. Cooper BS, Vanderpool D, Conner WR, Matute DR, Turelli M. *Wolbachia* acquisition by
639 *Drosophila yakuba*-clade hosts and transfer of incompatibility loci between distantly related
640 *Wolbachia*. Genetics. 2019;212: 1399–1419.
641 29. Madhav M, Parry R, Morgan JAT, James P. *Wolbachia* Endosymbiont of the Horn Fly
642 (*Haematobia irritans*): a Supergroup A Strain with Multiple Horizontally Acquired Cytoplasmic
643 Incompatibility Genes. Appl Environ Microbiol. 2020;86: e02589-19.
644 30. Bordenstein SR, Wernegreen JJ. Bacteriophage flux in endosymbionts (*Wolbachia*): Infection
645 frequency, lateral transfer, and recombination rates. Mol Biol Evol. 2004;21: 1981–1991.
646 doi:10.1093/molbev/msh211
647 31. Wang N, Jia S, Xu H, Liu Y, Huang D. Multiple Horizontal Transfers of Bacteriophage WO and
648 Host *Wolbachia* in Fig Wasps in a Closed Community. Front Microbiol. 2016;7: 1–10.
649 doi:10.3389/fmicb.2016.00136
650 32. Hill T, Unckless RL, Perlmutter JI. Positive Selection and Horizontal Gene Transfer in the
651 Genome of a Male-Killing *Wolbachia*. Mol Biol Evol. 2022;39: 1–17.
652 doi:10.1093/molbev/msab303
653 33. Gao S, Ren YS, Su CY, Zhu DH. High Levels of Multiple Phage WO Infections and Its
654 Evolutionary Dynamics Associated With *Wolbachia*-Infected Butterflies. Front Microbiol.
655 2022;13: 1–12. doi:10.3389/fmicb.2022.865227

656 34. Le Clec'h W, Dittmer J, Raimond M, Bouchon D, Sicard M. Phenotypic shift in *Wolbachia*
657 virulence towards its native host across serial horizontal passages. Proc R Soc B Biol Sci.
658 2017;284. doi:10.1098/rspb.2017.1076
659 35. Chrostek E, Marialva MSP, Esteves SS, Weinert LA, Martinez J, Jiggins FM, et al. *Wolbachia*
660 Variants Induce Differential Protection to Viruses in *Drosophila melanogaster*: A Phenotypic
661 and Phylogenomic Analysis. PLoS Genet. 2013;9. doi:10.1371/journal.pgen.1003896
662 36. Chrostek E, Teixeira L. Mutualism Breakdown by Amplification of *Wolbachia* Genes. PLoS Biol.
663 2015;13: 1–22. doi:10.1371/journal.pbio.1002065
664 37. Duarte EH, Carvalho A, López-Madrigal S, Costa J, Teixeira L. Forward genetics in *Wolbachia*:
665 Regulation of *Wolbachia* proliferation by the amplification and deletion of an additive
666 genomic island. PLoS Genetics. 2021. doi:10.1371/journal.pgen.1009612
667 38. Duron O, Bernard J, Atyame CM, Dumas E, Weill M. Rapid evolution of *Wolbachia*
668 incompatibility types. Proc R Soc B Biol Sci. 2012;279: 4473–4480.
669 doi:10.1098/rspb.2012.1368
670 39. Namias A, Sahlin K, Makoundou P, Bonnici I, Sicard M. Nanopore sequencing enables
671 multigenic family reconstruction despite highly frequent PCR-induced recombination. bioRxiv.
672 2022; 2022.11.14.516392. doi:10.1101/2022.11.14.516392
673 40. Sicard M, Namias A, Perriat-Sanguinet M, Carron E, Unal S, Altinli M, et al. Cytoplasmic
674 incompatibility variations in relation with *Wolbachia cid* genes divergence in *Culex pipiens*.
675 MBio. 2021;12: e02797-20. doi:10.1128/mBio.02797-20
676 41. Bonneau M, Landmann F, Labbé P, Justy F, Weill M, Sicard M. The cellular phenotype of
677 cytoplasmic incompatibility in *Culex pipiens* in the light of *cidB* diversity. PLoS Pathog.
678 2018;14: e1007364. doi:10.1371/journal.ppat.1007364
679 42. Scholz M, Albanese D, Tuohy K, Donati C, Segata N, Rota-Stabelli O. Large scale genome
680 reconstructions illuminate *Wolbachia* evolution. Nat Commun. 2020;11. doi:10.1038/s41467-
681 020-19016-0

682 43. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, et al. The *Wolbachia* genome of
683 *Brugia malayi*: Endosymbiont evolution within a human pathogenic nematode. PLoS Biol.
684 2005;3: 0599–0614. doi:10.1371/journal.pbio.0030121
685 44. Newton ILG, Clark ME, Kent BN, Bordenstein SR, Qu J, Richards S, et al. Comparative genomics
686 of two closely related *Wolbachia* with different reproductive effects on hosts. Genome Biol
687 Evol. 2016;8: 1526–1542. doi:10.1093/gbe/evw096
688 45. Wang GH, Sun BF, Xiong TL, Wang YK, Murfin KE, Xiao JH, et al. Bacteriophage WO can
689 mediate horizontal gene transfer in endosymbiotic *Wolbachia* genomes. Front Microbiol.
690 2016;7: 1–16. doi:10.3389/fmicb.2016.01867
691 46. Prout T. Some evolutionary possibilities for a microbe which causes incompatibility in its host.
692 Evolution (N Y). 1994;48: 909–911.
693 47. Turelli M. Evolution of incompatibility-inducing microbes and their hosts. Evolution (N Y).
694 1994;48: 1500–1513. doi:10.1111/j.1558-5646.1994.tb02192.x
695 48. Charlat S, Calme C, Merçot H. On the *mod resc* model and the evolution of *Wolbachia*
696 compatibility types. Genetics. 2001;159: 1415–1422. doi:10.1093/genetics/159.4.1415
697 49. Atyame CM, Duron O, Tortosa P, Pasteur N, Fort P, Weill M. Multiple *Wolbachia* determinants
698 control the evolution of cytoplasmic incompatibilities in *Culex pipiens* mosquito populations.
699 Mol Ecol. 2011;20: 286–298. doi:10.1111/j.1365-294X.2010.04937.x
700 50. Assogba BS, Milesi P, Djogbénou LS, Berthomieu A, Makoundou P, Baba-Moussa LS, et al. The
701 *ace-1* Locus Is Amplified in All Resistant *Anopheles gambiae* Mosquitoes: Fitness
702 Consequences of Homogeneous and Heterogeneous Duplications. PLoS Biol. 2016;14: 1–26.
703 doi:10.1371/journal.pbio.2000618
704 51. Kambhampati S, Rai KS, Burgun SJ. Unidirectional Cytoplasmic Incompatibility in the
705 Mosquito, *Aedes albopictus*. Evolution. 1993. p. 673. doi:10.2307/2410079
706 52. Sinkins SP, Braig HR, O'Neill SL. *Wolbachia pipientis*: Bacterial density and unidirectional
707 cytoplasmic incompatibility between infected populations of *Aedes albopictus*. Exp Parasitol.

1995;81: 284–291. doi:10.1006/expr.1995.1119

53. Baldo L, Hotopp JCD, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, et al. Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. Appl Environ Microbiol. 2006;72: 7098–7110. doi:10.1128/AEM.00731-06

54. Rogers SO, Bendich AJ. Extraction of total cellular DNA from plants, algae and fungi. Plant 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 previously described variable regions are shown: the upstream

733 region, named with a Greek letter for *cidA* and with a Latin letter for *cidB*, and the downstream
734 region, named with a digit for both genes. *cidA* regions whose presence differ among repertoires are
735 represented in green. *cidB* repertoires are common to all repertoires. Females infected with (β+, 16+)
736 or a (β+, 16-) *Wolbachia* are able to rescue males infected with *w*Pip-Ist (compatible), while females
737 infected with (β-, 16-) *Wolbachia* cannot (incompatible).
738
739 **S2 Fig. Variations of hatching rates in compatible crosses between Slab females and Ist males, in**
740 **relation to variations in: A) *Wolbachia* infection level; and numbers of copies of B) total *cidA*; C)**
741 ***cidA-III-beta(2)*; and D) *cidA-III-16*.** None of these variations are significantly correlated with hatching
742 rates (Spearman correlation coefficient, P = 0.49, 0.28, 0.94 and 0.16 for A, B, C and D respectively).
743
744 **S3 Fig. Variations of *cidA* copy numbers in *Wolbachia* genomes of Slab females.** A), B) and C) show
745 the numbers of copies of total *cidA*, *cidA-III-16* and *cidA-III-beta(2)* regions, respectively in females
746 with repertoires (β-,16-), (β+,16-), and (β+,16+). D), E) and F) show the distribution of *cidA, cidA-III-16*
747 and *cidA-III-beta(2)* copy numbers. All histograms show either bi or trimodal distributions. All copy
748 numbers were quantified using quantitative PCR relative to the single copy *Wolbachia* gene *wsp*.
749
750 **S4 Fig. Variations of *cidB* copy numbers in *Wolbachia* genomes of Slab females**. *cidB* copy numbers
751 equal *cidA* copy numbers, (β-,16-) *w*Pip having around 3 *cidB* copies, while (β+,16-) and (β+,16+) have
752 around 5 *cidB* copies. All copy numbers were quantified using quantitative PCR relative to the single
753 copy *Wolbachia* gene *wsp*.
754
755 **S5 Fig. PCR primers used to analyze sequence variation in *cid* genes.** PCRs amplifying the full *cidA*
756 gene (burgundy arrows), variable regions of the *cidB* gene (green arrows) and the *cidAB* tandem
757 (purple arrows) were used for nanopore sequencing of repertoires. PCRs specific of *cidA-III-beta(2)*

758 (yellow) and *cidA-III-16* (orange) were used both in classic PCR and qPCR to determine the
759 presence/absence of these regions.
760
761 **S6 Fig. Sequence variation at the CidA-CidB interaction interface II.** Protein sequences are shown for
762 the span of the interaction interface II (position of the first and last aminoacids of this interface were
763 obtained thanks to Wang & Hochstrasser, pers. com.). Polymorphic positions are shown with a grey
764 background.
765 This interface is polymorphic in CidB, the *w*Pip-IV CidB showing a different sequence from all other
766 groups (sequences shown here are representative of all sequenced variants). In contrast, CidA
767 variants are identical in this interaction interface in all sequenced variants whatever their groups.
768
769 **S1 Table. *cid* repertoires of the Istanbul isofemale line.** Nanopore sequencing of *cid* PCR products
770 gave identical repertoires for 4 distinct Ist individuals.
771
772 **S2 Table. Crosses between Slab females and Ist males.** For each female, we determined the hatching
773 rate (HR) (i.e. ratio between hatched larvae and eggs). The specific upstream *cidA-III-beta(2)* and
774 downstream *cidA-III-16* regions were quantified in each female and normalized over the single copy
775 *Wolbachia* gene *wsp*. The total *cidA* and *cidB* copy number was also quantified and normalized over
776 *wsp*. Infection rates were assessed by quantifying the single copy gene *wsp* and normalizing it on the
777 *Cx. pipiens* single copy gene *ace2*.
778
779 **S3 Table. Crosses between Slab females from each of the 6 newly established isofemale sublines [2**
780 **(β+,16+), 2 (β+,16-) and 2 (β-,16-)] and Istanbul males.** Hatching rates and number of hatched larvae
781 are shown for each female. Eggs were only counted in compatible crosses, involving (β+,16+) and
782 (β+,16-) females. All females were screened for the presence/absence of two regions of the

783 polymorphic gene *cidA* (*cidA-III-beta(2)* and *cidA-III-16*) to confirm that they had the same repertoire

784 as the female from which the isofemale subline was founded. NF = non fertilized

785

786 **S4 Table. Reciprocal crosses between a (β+,16+) and a (β-,16-) isofemale subline**. Number of eggs

787 and larvae and resulting hatching rates (HR). NF = non fertilized

788

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

790 text.

791

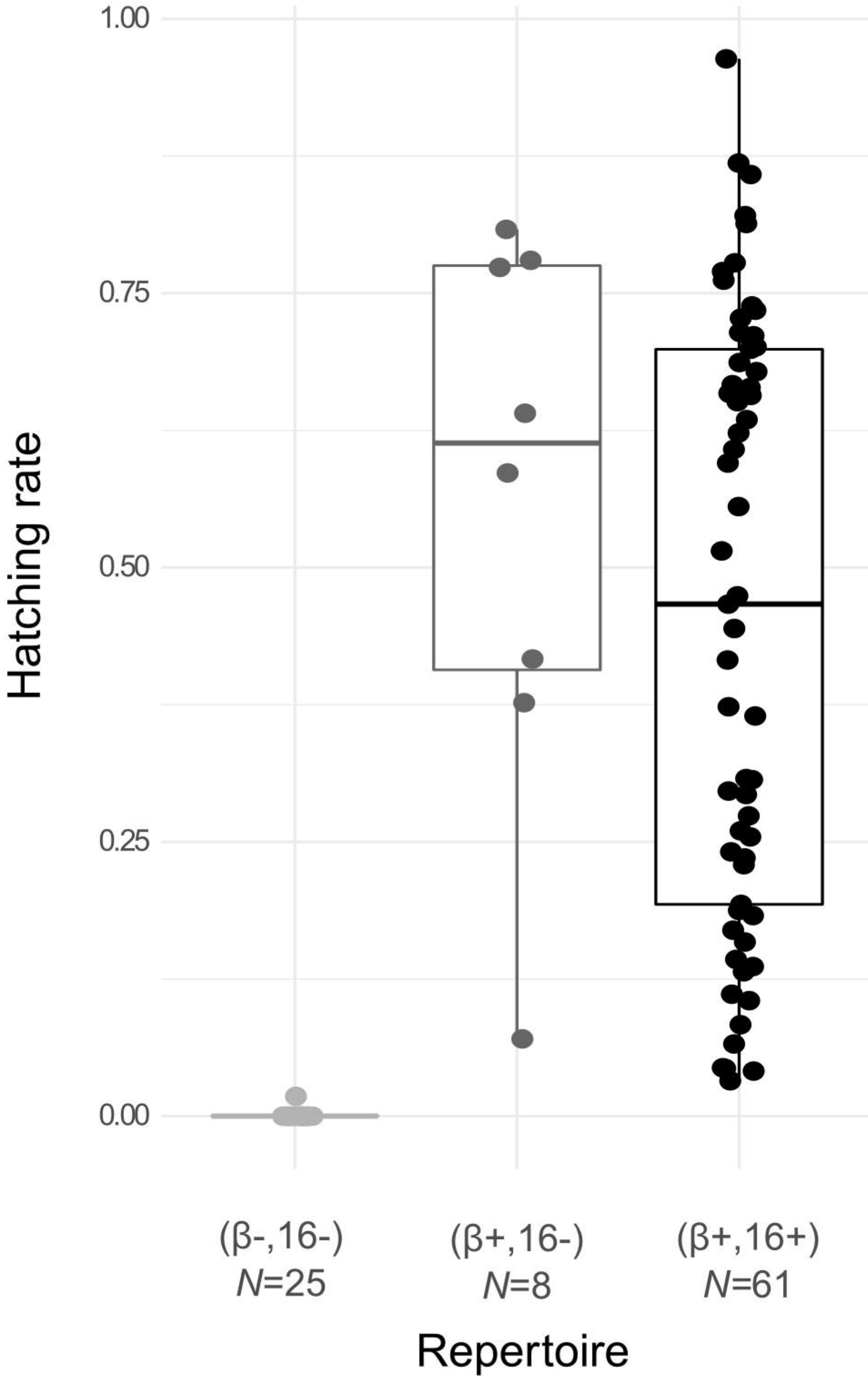
792 **S6 Table. Correspondence between variants names in the old (Bonneau et al, Nat Com 2018) and in**

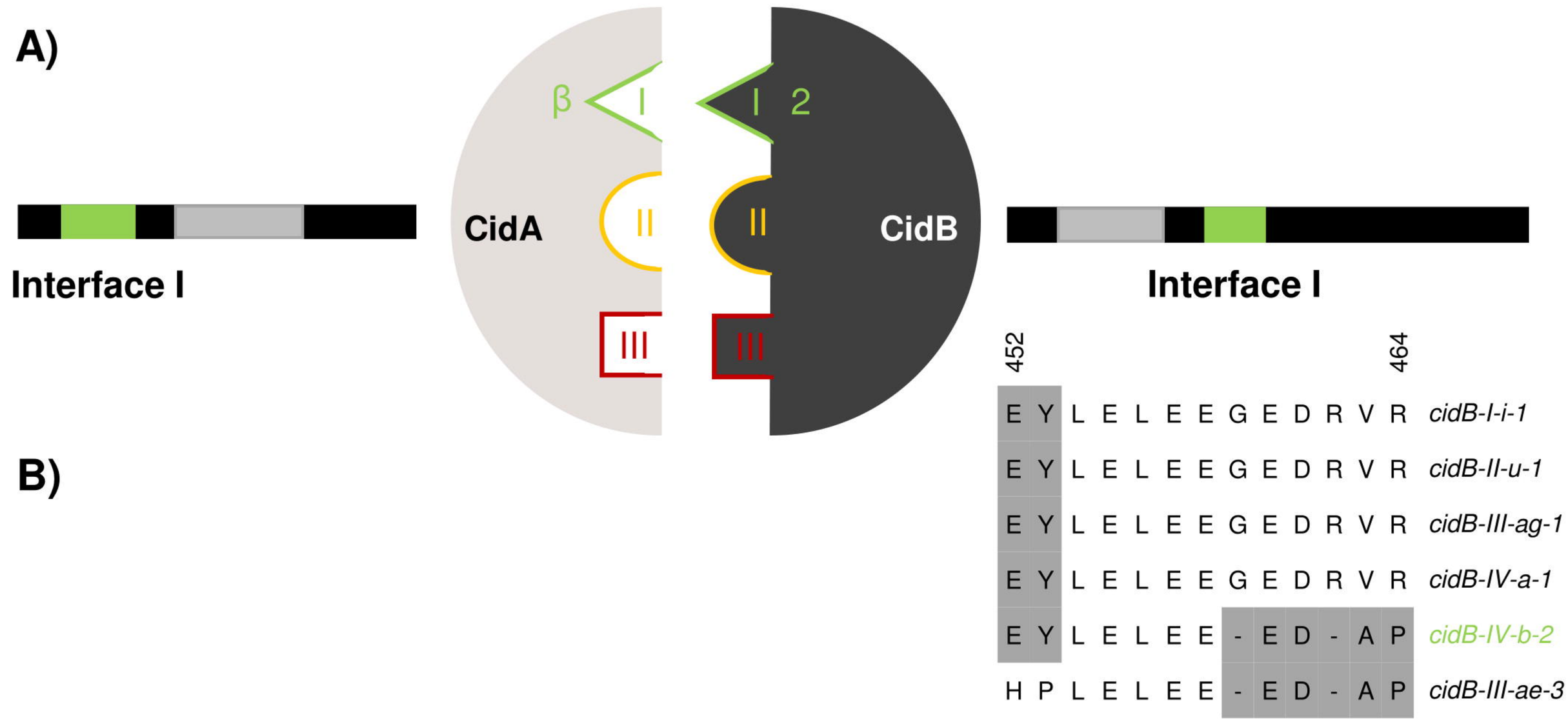
793 **current nomenclature.**

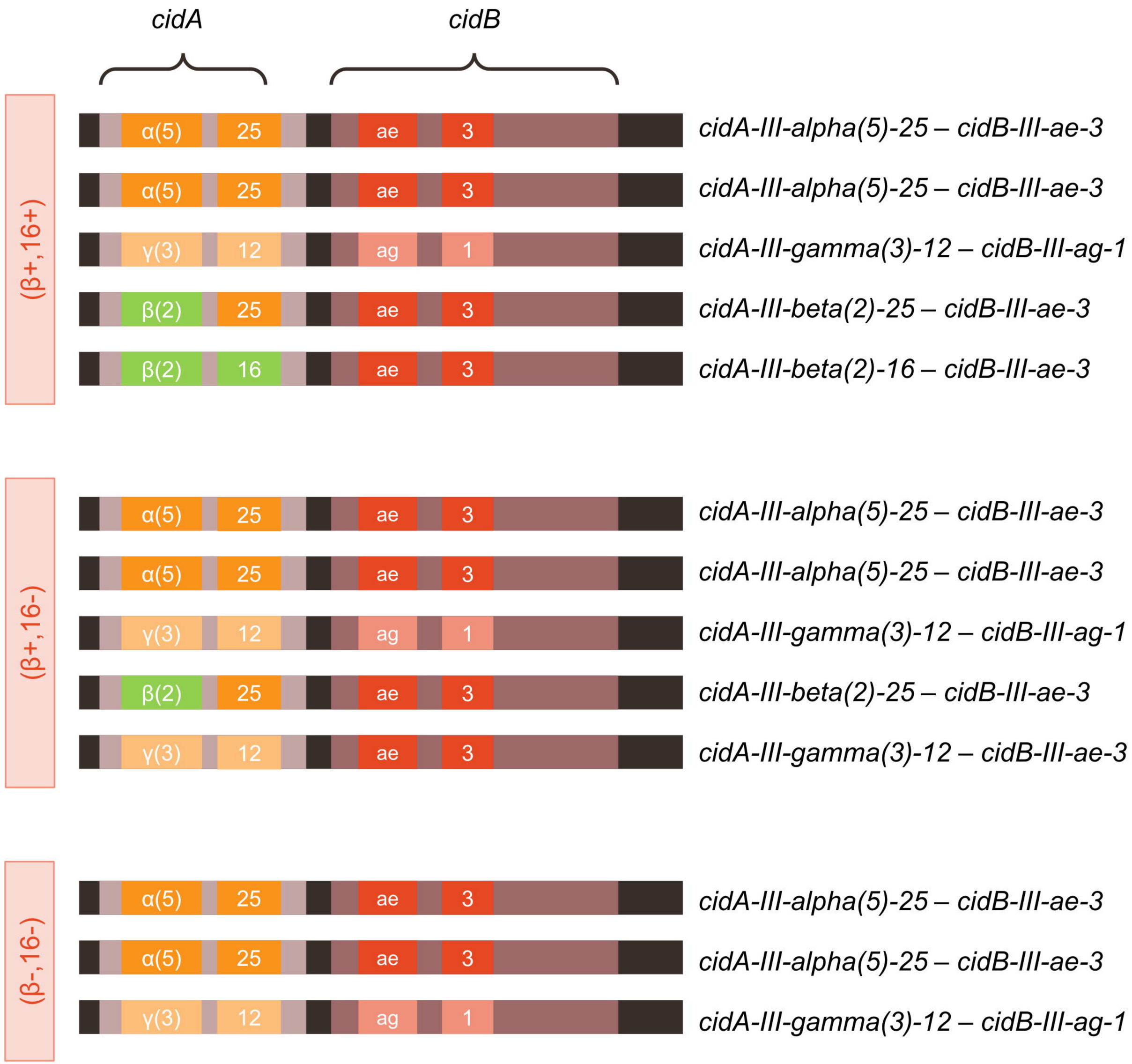
794

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

796


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