Genome assembly of the A-group *Wolbachia* in *Nasonia oneida* and phylogenomic analysis of *Wolbachia* strains revealed genome evolution and lateral gene transfer

Xiaozhu Wang[1], Xiao Xiong[1,2], Wenqi Cao[1], Chao Zhang[2], John H. Werren[3] and Xu Wang[1,4,5†]

[1]*Department of Pathobiology, Auburn University, Auburn, AL 36849*

[2]*Translational Medical Center for Stem Cell Therapy and Institute for Regenerative Medicine, Shanghai East Hospital, Shanghai Key Laboratory of Signaling and Disease Research, School of Life Sciences and Technology, Tongji University, Shanghai, China*

[3]*Department of Biology, University of Rochester, Rochester, NY 14627*

[4]*Alabama Agricultural Experiment Station, Auburn University, Auburn, AL 36849*

[5]*HudsonHudsonAlpha Institute for Biotechnology, Huntsville, AL 35806*

[†]corresponding author:

Xu Wang

Phone: (334) 844-7511

Fax: (334) 844-2618

E-mail: xzw0070@auburn.edu

ORCID: 0000-0002-7594-5004

Key words: *Wolbachia*, *Nasonia*, parasitoid wasp, lateral gene transfer, phylogenomics, 10X Genomics Chromium linked reads

Abstract

Wolbachia are obligate intracellular bacteria which commonly infect various nematode and arthropod species. Based on average depth and GC-content differences, we assembled the genome of *Wolbachia* in the parasitoid jewel wasp species *Nasonia oneida* (*w*One), using 10X Genomics Chromium linked read technology. The final draft assembly of the *w*One genome consists of 938,765 bp in 254 scaffolds and it contains 948 coding genes. Comparative analysis of sequenced *Wolbachia* genomes and phylogenetic analysis of five genes (*coxA*, *gatB*, *hcpA*, *ftsZ* and *fbpA*) revealed that *w*One belongs to the A1 supergroup. Pyrosequencing experiments validated the lack of A2 or B strains, which were likely lost during laboratory culturing. 68 polymorphisms were identified in the *w*One genome, suggesting heteroplasmy and/or gene duplications. These polymorphisms are enriched in nonsynonymous changes in 37 coding genes, with biological functions related to bacterial infection, such as transporter activity related genes. Phylogenomic analysis of 34 sequenced *Wolbachia* genomes revealed that *w*One is more closely grouped with A-*Wolbachia* in the *Drosophila simulans* fruitfly (*w*Ha) instead of A-*Wolbachia* in wasps (*w*VitA and *w*Uni). Phylogenetic analysis of 211 single ortholog genes in these 34 *Wolbachia* genomes discovered six discordant trees, indicating lateral gene transfer or recombination events not only between A and B supergroups, but also between A and E strains. Assembling bacterial genomes from host genome projects can provide an effective method for characterizing *Wolbachia* genome diversity, as well as the assembling microbial genomes in the insect gut microbiota.

Introduction

Wolbachia, alphaproteobacterial endosymbionts, are widespread and common in arthropods and filarial nematodes, either as reproductive parasites or mutualists [1-3]. It is estimated that half or more of arthropods are infected with *Wolbachia* [4, 5], possibly representing a dynamic equilibrium between gain and loss on a global scale [6-8]. The widespread distribution of *Wolbachia* is due to a combination of vertical and horizontal transmission. Not limited to vertical transmission, *Wolbachia* can also move across the species boundaries through horizontal transfer and hybrid introgression [9]. The *Wolbachia*-host interaction generally spans a range from reproductive parasitism to mutualism. *Wolbachia* can alter the host reproduction to enhance their own transmission in different ways, such as feminization of genetic males, male-killing, parthenogenetic induction, and cytoplasmic incompatibility [1, 10]. *Wolbachia pipientis* have been divided into eight supergroups (A-H) based on 16S ribosomal RNA sequences and other sequence information, including six supergroups (A,B and E-H) primarily identified in arthropods and two supergroups (C and D) commonly existed in filarial nematodes [1]. Supergroup G is now considered as a recombinant of supergroups A and B [11]. A multi-locus strain typing (MLST) system based on five house-keeping genes, (*coxA*, *gatB*, *hcpA*, *ftsZ* and *fbpA*) has been developed for *Wolbachia* [12], and is widely used for strain typing and to characterize strain variation within *Wolbachia*. However, the increasing number of genome sequences for *Wolbachia* allows for more detailed characterization of their diversity, including lateral gene transfer events.

The jewel wasp genus of *Nasonia* has been an excellent model for *Wolbachia* research [9, 13-16]. Eleven *Wolbachia* have so far been identified in the four species of *Nasonia*, including two (*w*VitA and *w*VitB) in *N. vitripennis* (NV), three (*w*NgirA1, *w*NgirA2 and *w*NgirB) in *N.*

44 *giraulti* (NG), three infections (*w*OneA1, *w*OneA2 and *w*OneB) in *N. oneida* (NO), and three
45 infections (*w*NlonA, *w*NlonBl and *w*NlonB2) in *N. longicornis* (NL) [9]. Although these strains
46 are closely related, many of the *Wolbachia* are not monophyletic (which would suggest
47 cospeciation with their hosts), but rather have diverse evolutionary origins, indicating horizontal
48 transfers from divergent host species [9].

49 Genomic studies of *Wolbachia* blossomed in the recent years since the first complete genome
50 of the A-*Wolbachia* parasite of *Drosophila melanogaster* (*w*Mel) published in 2004 [17]. Later
51 the complete genome of D-*Wolbachia* (*w*Bm) in nematode *Brugia malayi* was reported in 2005
52 [18]. A list of sequenced whole genomes of *Wolbachia* is summarized in Table S1. *Wolbachia*
53 genomes are small with a range between 0.9-1.7 Mb. In general, most of the nematode-
54 associated *Wolbachia* have smaller genomes but retain intact metabolic pathways and
55 immunology pathways, which contribute to the mutualistic relationship with the hosts [17-19].
56 However, arthropod-associated *Wolbachia* contain more prophage and ankyrin repeat encoding
57 (ANK) gene components, which may reflect their more frequent parasitic lifestyle [20, 21].
58 Furthermore, many studies have claimed lateral gene transfer (LGT) across strains and
59 supergroups, which may be mediated by bacteriophage and lead to the mosaic genomes of
60 *Wolbachia* [22-24]. Although co-infection of different strains and LGT exist in the same
61 arthropod host, the supergroups may remain genetically distinct clades [25]. Another key feature
62 of most *Wolbachia* genomes is the abundance of mobile and repetitive elements, which are
63 different from most *Rickettsiales* [1]. Noticeably, the proportion of repetitive elements in genome
64 varied widely among *Wolbachia* strains. For example, 22% of the *w*Ri genome is comprised of
65 repetitive sequences [22], compared to only 5% in the *w*Bm genome [18]. In jewel wasp
66 (*Nasonia*) species, only two *Wolbachia* strains have been sequenced (*w*VitA [26] and *w*VitB

67 [23]), both from NV. Genome sequence of additional *Wolbachia* strains in additional *Nasonia*
68 species will facilitate the comparative genomic and evolutionary analyses of this model system.

69 Because of its endosymbiotic nature, multiple different *Wolbachia* strains can be present in
70 the same host cells, allowing the potential for homologous recombination between strains [27,
71 28]. Recombination events in *Wolbachia* have been discovered in *Wsp* [29] and other genes in
72 Crustaceans [30], mites [31] and insect species including wasps [29, 32, 33], ants [34] and
73 butterflies [35]. There is no evidence of interstrain recombination in filarial nematode
74 *Wolbachia* strains [36]. Most of the previous research focused on MLST genes, or a few
75 genomes from the A and B supergroups. Therefore, whole-genome analyses in a large number of
76 *Wolbachia* strains of all supergroups are needed to identify additional homologous
77 recombination and LGT events among *Wolbachia* strains. In this study, we assembled the
78 *Wolbachia* strain in *N. oneida*, performed phylogenomic analyses on 34 genome sequenced
79 *Wolbachia* strains, and analyzed the individual gene tree to identify potential recombination
80 events at the genome level.

81

82 # Results

83 **Assembly of *Wolbachia* genome in the 10X Genomics Chromium sequencing of *N. oneida*.**

84 The *de novo* assembly of NO genome was performed using the linked reads generated by 10X
85 Genomics Chromium technology with `Supernova 2.0` assembler [37]. *Wolbachia* scaffolds
86 were identified and separated from the NO genome assembly using a custom bioinformatics
87 pipeline (Figure 1 and 2A). NO scaffolds were BLATed against bacterial genome database [38],
88 and we identified *w*One scaffolds based on the average coverage, GC-content and sequence
89 identity to known *Wolbachia* sequences (Figure 2; see Materials and Methods). *w*One scaffolds
90 have an average genome coverage of 32.85X, which is significantly lower compared to 348.94X
91 for the NO genomic scaffolds (P-value < 2.2×10^{-16} , Figure 2B). The mitochondrial scaffold
92 coverage is over 10,000 (Figure 2B). In addition, there is a significant shift in GC-content for the
93 *Wolbachia* genome (35.40%) compared to the host genome average (37.90%, P-value = $6.3 \times 10^{-}$
94 ^{12}, Figure 2C). The sufficient differences in coverage and GC-content among host genome, host
95 mitochondrial genome and the *Wolbachia* genome allow clear separation of the *w*One scaffolds.

96 The *w*One draft genome consists of 938,765 nucleotides with 35.40% GC-content (Table 1).
97 This assembly has a total of 254 scaffolds ranging in length from 1,013 to 27,795 bps with a
98 scaffold N50 of 4.74 Kb. A total of 1,096 genes were annotated in the *w*One genome with an
99 average gene length of 757 bp, including 948 coding genes and 115 pseudogenes. One of each
100 the rRNAs 5S, 16S and 23S were identified in *w*One draft genome, but only 25 tRNAs were
101 successfully identified with 9 tRNAs missing, including Arg (ACG), Arg (TCT), Asn (GTT),
102 Asp (GTC), Glu (TTC), Gly (GCC), Ile (GAT), Leu (CAA), and Met (CAT). 74.4% of genes

103 from *w*VitA, the closest sequenced *Wolbachia* strain in *Nasonia*, were present in *w*One assembly
104 (Figure 3).

105 **Comparative genomic analysis of *Wolbachia* strains in *Nasonia* species.**

106 Pairwise alignments were performed to two *Wolbachia* genomes isolated from *N. vitripennis*
107 (*w*VitA and *w*VitB; see Materials and Methods). The assembled genome size for *w*One is 15%
108 smaller than *w*VitB and 22% smaller than *w*VitA. The total number of proteins is similar to the
109 *w*VitA and *w*VitB genomes (Table 1). A dot plot revealed a better colinear relationship between
110 *w*One and *w*VitA, with only a small number of rearrangements around the origin (Figure 3A). A
111 total of 812,618 bps of *w*One genome were aligned *w*VitA genome, including 418,704 bps
112 matched in the same orientation and 393,914 bps matched in the reverse orientation. The longest
113 scaffold in the *w*One genome (SCAFFOLD1) was aligned to three *w*VitA scaffolds
114 (NZ_MUJM01000091.1, NZ_MUJM01000092.1, and NZ_MUJM01000093.1) with 96.96%
115 identify. However, when comparing *w*One and *w*VitA genomes with *w*VitB genome
116 respectively, significantly more genome rearrangements and inversions were observed (Figure
117 3B and C). This is not surprising as *w*VitB belongs to a different supergroup. A total of 566,507
118 bps of *w*One genome were aligned *w*VitB genome, including 256,621 bps matched in the same
119 orientation and 309,886 bps matched in the reverse orientation. These findings indicate that the
120 *w*One strain is from the A supergroup.

121 When comparing the gene contents of these *Wolbachia* strains, a total of 597 genes were
122 shared among genomes of *w*One, *w*VitA and *w*VitB; 230 more genes were shared between the
123 *w*One and *w*VitA genomes but not with *w*VitB genome. Among these *w*One-specific genes, a
124 large fraction belongs to hypothetical protein (N=93), transposon-related (N=25), insertion
125 sequence (IS) elements (N=19), and ankyrin repeat (N=5) genes. In particular, the *w*One genome

126 contains a greater number and some different families of IS elements compared to the genomes
127 of *w*VitA and *w*VitB, indicating a lineage-specific expansion of IS elements in *w*One (Table 2).
128 Although *w*VitA and *w*VitB infect the same host NV and *w*One infect a different host NO, the
129 gene content of *w*VitA is closer to that of *w*One than *w*VitB, as expected by their supergroup
130 affiliations and indicating that there is no rampant recombination between the wVitA and wVitB
131 genome-wide.

132 **Phylogenomic analysis of 34 *Wolbachia* genomes revealed the evolutionary relationship and**
133 **potential recombination events among strains**

134 We next compared *w*OneA to 33 other *Wolbachia* genomes (Table S1), which include 15 A-
135 group and 12 B-group strains from diverse host species. Single-gene ortholog clusters were
136 generated using the procedure described in the Methods, and 211 single-gene ortholog clusters
137 were identified that are shared between *w*OneA and the 33 other *Wolbachia* genomes (Table S1).
138 Based on this core gene set, a Neighbor-Joining phylogenetic tree of 34 *Wolbachia* genomes
139 confirmed the separation of different supergroups A (*w*Suzi, *w*Spc, *w*Ri, *w*Ha, *w*Au, *w*Mel,
140 *w*MelPop, *w*Gmm, *w*Uni, *w*DacA, *w*Nfe, *w*Npa, *w*Nfla, *w*Nleu, *w*VitA, *w*OneA), B (*w*Con,
141 *w*AlbB, *w*Stri, *w*Di, *w*No, *w*Tpre, *w*DacB, *w*VitB, Ob_Wba, *w*Bol1, *w*Pip_Mol, *w*Pip), C (*w*Oo,
142 *w*Ov), D (*w*Bm, *w*Wb), E (*w*Fol) and F (*w*Cle) with 100% bootstrap support (Figure 4).
143 Noticeably, *w*OneA is more closely related to a subset of A-*Wolbachia* found in *Drosophila*
144 (*w*Ha, *w*Ri, *w*Spc and *w*Suzi) than to *w*VitA and *w*Uni in parasitoid wasps. This pattern was
145 previously observed using MLST by five genes in *Wolbachia* [9], but is now supported for a
146 much larger data set. Our genomic analyses also support extensive horizontal movement of
147 *Wolbachia* between host species. A maximum likelihood phylogenetic tree of protein sequences

148 from these core genes was constructed with RAxML. The protein ML phylogenetic tree is highly
149 similar to the NJ tree with some minor differences (Figure S1).

150 We next examined genome for potential lateral gene transfers among the core gene set. Single
151 gene trees of coding sequences from the core gene set were constructed to check for supergroup
152 level consistency. For 205 of the 211 trees, the separation among the A to F supergroups was
153 consistent, with slight rearrangement for some *Wolbachia* strains within each supergroup.
154 However, six trees are mixed among different supergroups, presumably due to lateral gene
155 transfer or recombination events between strains (Figure 5-8); the gene are hypothetical protein
156 DEF52_00180, cytochrome c oxidase subunit II (coxB), hypothetical protein DEF52_04830,
157 NADH-quinone oxidoreductase subunit C (nuoC), molecular chaperone (DnaK), and arginine-
158 tRNA ligase (argS).

159 Among these six single gene trees, one tree showed evidence of a recombination event
160 between A/E-*Wolbachia* and B-*Wolbachia* in gene argS (Figure 8). Four trees revealed some mix
161 groupings between A-*Wolbachia* and B-*Wolbachia* in hypothetical protein DEF52_00180
162 (Figure 5), coxB (Figure 5A), hypothetical protein DEF52_04830 (Figure 5B), and nuoC (Figure
163 6A), which suggest a lateral transfer. Also, three trees grouped A-*Wolbachia* and the E-
164 *Wolbachia* *w*Fol together, including *w*Fol clustered with A-*Wolbachia* in coxB (Figure 6A) and
165 DEF52_04830 (Figure 5B), and A-*Wolbachia* *w*DacA clustered with *w*Fol in DnaK (Figure 7B).
166 The recombination between A and B supergroups in gene coxB was reported by a previous study
167 of 6 *Wolbachia* strains [25], and the remaining identified between supergroup recombination
168 events are novel findings in our study. Taken together, 97% of the single copy orthologs agree
169 with the supergroup classification in *Wolbachia*, with a few cases of likely recombination or
170 lateral transfer events between *Wolbachia* strains of different supergroups. The finding also

171 indicates that these recombination events involve relatively small regions, rather than events
172 involving many genes. The frequent gene order rearrangements observed in *Wolbachia* may
173 make larger recombination tracks between supergroups less successful, as they are more likely to
174 involve vital gene losses due to lack of synteny.

175 **Phylogenetic analysis of MLST genes of *Wolbachia* in *Nasonia*.**

176 Previous research suggested that NO contains three *Wolbachia* strains, A1, A2 and B, which
177 were acquired through a hybrid introgression from NG [9]. However, the genome sequence for
178 the current *N. oneida* strain indicates presence of only one A-group strain, even though the same
179 insect strain is present in both studies. This difference is likely due to stochastic loss of two
180 strains during laboratory culturing. This is known to happen in *Nasonia*, particularly when the
181 strains are passed through winter larval diapause [14], which involves storage under refrigeration
182 for up to 1.5 years after larval diapause induction. To determine which strain was identified and
183 assembled in our study, phylogenetic analysis was conducted with all *Wolbachia* strains in
184 *Nasonia* using the MLST gene approach [12, 39]. The phylogenetic trees were constructed for all
185 five MLST genes (Figure 9). *w*One MLST genes grouped with A1 strains of other species. The
186 closest branch is the corresponding *w*NgirA1 ortholog for all five MLST genes, indicating that
187 the assembled *w*One is the strain A1. All phylogenetic trees of five MLST genes supported the
188 conclusion that the identified *w*One in our study belongs to the A1 strain of NO. The MLST gene
189 sequence of *Wolbachia* in NO are the same as the corresponding one of *Wolbachia* in NG [9].
190 Our results are consistent with the previous findings for all five MLST genes in *w*OneA1.

191 **Loss of A2 and B *Wolbachia* in the assembled *N. oneida* strain.**

192 To determine whether there are other *Wolbachia* strains in NO, we aligned the *w*One
193 sequencing reads to NG MLST genes. No reads were mapped to *w*NgirB MLST genes,
194 suggesting the B supergroup is absent in our NO strain. For *w*NgirA1 MLST genes, the average
195 coverage is 30X which are close to the coverage of *w*One scaffolds. *w*NgirA2 MLST genes only
196 have multiple mapped reads to both *w*NgirA1 and *w*NgirA2, which is not informative to
197 determine the existence of the A2 strain. The informative SNPs in each of MLST genes between
198 *w*NgirA1 and *w*NgirA2 were further checked for read counts to ensure the inability to detect of
199 *w*OneA2. No read count was identified for A2 allele of all MLST genes, while all A1 alleles
200 were supported by at least 30 read counts (Table S2).

201 Furthermore, strain typing of *Wolbachia* was performed on NO of our study and NO genomic
202 DNA samples that are known to be infected with all three strains (A1, A2 and B), using
203 independent allele-specific pyrosequencing approach. An A/G SNP in the CoxA gene was used
204 to separate B-*Wolbachia* from A-*Wolbachia* (A allele in A1/A2-*Wolbachia* and G allele in B-
205 *Wolbachia*, Figure S2A). In gatB gene, a C/T SNP can distinguish A1-*Wolbachia* allele from
206 A2/B-*Wolbachia* (Figure S2B). The pyrosequencing results confirmed the lack of A2 and B
207 strains in the genome assembled NONY strain. All three *Wolbachia* infections (A1, A2 and B)
208 were successfully identified in the CAR262L strain DNA samples (Figure S2 and Table S3).

209 **Polymorphic genes in *w*OneA genome are enriched in transmembrane transporter genes.**

210 A total of 68 high-quality SNPs was called in the *w*OneA genome. The alternative allele
211 frequency of these identified SNPs ranges from 0.10 to 0.67 (Figure S3). Among these identified
212 SNPs, 44 were found to be located within gene regions (Table S4). 70% (31 of 44) of these SNPs
213 are non-synonymous resulting in protein sequencing changes and premature stops. BLAST2GO

214 [40] analysis identified the transmembrane transporter activity molecular function GO term is

215 significantly enriched (Figure S4) among genes containing SNPs.

216 ## Discussion

217 **Assembly of prokaryotic genome in an insect species using linked reads technology.**

218 Here we report the first *Wolbachia* genome assembled from host sequences using 10X
219 Genomics linked reads technology. Although many *Wolbachia* genomes were assembled since
220 the first *w*Mel genome paper was published, the difficulty in purifying *Wolbachia* DNA from the
221 host is still a limiting factor for the genomic studies of *Wolbachia*. Due to its intracellular life
222 style and inability of media culture, the purification of *Wolbachia* DNA from the host sample is
223 time-consuming and sometimes impossible to obtain sufficient quantity without contamination
224 from the host nuclear and mitochondrial genomes (Table S1). Three major methods have been
225 applied to purify the *Wolbachia* genomic DNA form the host DNA: 1) selection of *Wolbachia*
226 enriched materials for DNA extraction, such as host ovaries [41] or *Wolbachia* infected cell lines
227 [42]; 2) use of different filter purification methods including pulsed-field gels [17], various
228 gradient gels [22, 24], or filter columns [26]; 3) Multiple-displacement amplification (MDA) for
229 *Wolbachia* DNA enrichment and amplification [25, 42]. Despite the efforts on the purification of
230 *Wolbachia* genomic DNA, only 90% purification could be achieved. Alternative methods have
231 been applied to solve the contamination problem. For example, in *w*Bm genome project,
232 *Wolbachia* BACs were selected from the host BAC library [18]. For the *w*VitB genome project,
233 a high-density tiled oligonucleotide array was developed to enrich for *Wolbachia* gDNA [23].
234 Other studies extracted *Wolbachia* reads from the host whole genome sequence dataset, and then
235 align to the reference genome of the closely related strains, or perform *de novo* assembly using
236 the filtered reads [19, 43-46].

237 We have demonstrated an alternative method to circumvent the DNA purification step in the
238 traditional *Wolbachia* genome assembly approaches, by *de novo* assembly of the host genome
239 and *Wolbachia* scaffolds using the 10X Genomics linked reads technology. The *Wolbachia* DNA
240 fragments were labeled with unique 10X barcodes, therefore they are much less likely to be
241 misassembled into the host genome scaffolds. In this study, the microbe and the host have a 10-
242 fold coverage difference, allowing accurate identification of the bacterial scaffolds. In
243 conclusion, the *Wolbachia* genome assembled with 10X linked reads was of good quality with
244 no contamination of host nuclear and mitochondrial DNA.

245 This method does not require prior knowledge about the presence of specific microbe species.
246 Therefore, it has the potential of identifying other intracellular symbionts, as well as
247 characterizing bacterial species in insect gut microbiota at the whole-genome level.

248 **Lateral gene transfer and recombination events among *Wolbachia* genomes.**

249 The phylogenomic analysis of 34 *Wolbachia* genomes in our study is the most comprehensive
250 phylogenomic and evolutionary analysis conducted in *Wolbachia* strains to date. By including
251 almost all available *Wolbachia* genomes in NCBI, we confirmed at the genome level that these
252 *Wolbachia* strains evolved as distinct clusters (A, B, C, D, E, F supergroups) and different
253 *Wolbachia* co-infected in the same host kept the strain boundary [25]. 205 of the 211 single gene
254 trees are consistent with the strain tree. Six genes trees have major rearrangements among
255 *Wolbachia* groups (Figures 5-8), indicating potential recombination or lateral gene transfer
256 events between strains. We estimated that the homologous recombination or hybrid introgression
257 occurred in at least 3% of the core genes in the *Wolbachia* genomes, and recombination may be
258 one of the evolutionary forces shaping the *Wolbachia* genomes.

259 The six genes with distinct tree structure difference from the consensus *Wolbachia* tree
260 include coxB, nuoC, DnaK, argS, hypothetical proteins DEF52_00180 and DEF52_04830. Gene
261 coxB and nuoC are both involved in the electron transport chain. , which might also interact with
262 hypothetical proteins DEF52_00180 and DEF52_04830. The functions of hypothetical proteins
263 DEF52_00180 and DEF52_04830 in *Wolbachia* are still unclear. However, the potential
264 recombination or lateral gene transfer events identified in these genes might suggest that their
265 maintenance could be under positive selection.

266 LGT events among A and B *Wolbachia* supergroups have been documented in previous
267 studies, and we identified a few addition cases through the phylogenomic analysis among 34
268 sequenced genomes. Interesting we also discovered LGT events between A and E supergroups,
269 which was not known previously. The E group *Wolbachia* was found in Collembola, or the
270 springtails [47-49]. A recent study characterized the *Wolbachia* in 11 collembolan species, and
271 found that nearly all are E group *Wolbachia* that are monophyletic, based on phylogenetic
272 reconstruction using MLST genes [50]. Our genome analysis of the single collembolan
273 *Wolbachia* genome reveals a number of candidate lateral gene transfer events, including
274 intragenic recombination in argS between A, E and B, lateral gene transfer in coxB and DnaK,
275 and in DEF52_04830. Targeted sequencing of these genes in the additional collembolan species
276 or additional genome sequences will help reveal the origins and directions of these lateral gene
277 transfers. We further speculate that selective maintenance of such transfers could suggest a
278 possible role in the *Wolbachia* function, such as parthenogenesis induction [50].

279 It has been recently argued that MLST genotyping has little utility in phylogenetic analyses,
280 and should be supplanted by genomic studies [51]. When the MLST system was developed, it
281 was pointed out by the authors that the system would be most useful for identifying relatively

282 closely related *Wolbachia*, due to potential recombination among more divergent strains [51].
283 However, our comparison on genome sequence and MLST phylogenies indicates that MLST
284 typing of supergroups is largely valid, and that closely related *Wolbachia* based on MLST results
285 are also closely related genome-wide. This suggests that, until *Wolbachia* genome sequencing
286 becomes much less expensive and can be readily performed on single arthropods, that MLST
287 will remain a powerful tool for identification of strains and their relationships. Nevertheless,
288 caution should be exercised due to some documented recombination events within MLST genes
289 and among them. Therefore, topologies should be compared among genes for evidence of
290 discordance, rather than simply relying of phylogenetic reconstructions of concatenated
291 sequences.

292 **Loss of A2 and B *Wolbachia* in *N. oneida* lab strains**

293 The whole genome alignments between *w*One and *w*VitA (Figure 3) and the phylogenetic
294 analysis of MLST genes (Figure 9) indicated the identified strain in our study is belong to A1
295 supergroup. The results indicate either A2 and B are present in an extremely low level so that
296 cannot be detected at the current sequencing coverage, or their density is much less in our
297 studied male NO adult samples. Subsequent allele-specific pyrosequencing validation
298 experiments confirmed the absence of A2 and B-type *Wolbachia* infections. In NO DNA
299 samples from a recently collected field strain, we estimate A1 is the dominate strain and accounts
300 for 55% of the total infection, 40% of the infection came from the B strain and 5% from A2
301 strain (Figure S2). The absence of A2 and B *Wolbachia* in the lab NO strain is likely due to
302 stochastic loss during laboratory maintenance and diapause.

303 **Evolution of the *Wolbachia* genome in the *Nasonia* genus.**

304 The draft genome of *w*One is comprised of 254 scaffolds with a total length of 0.94 Gb. The
305 total size of *w*One draft genome is shorter comparing to the other two *Nasonia*-associated
306 *Wolbachia* *w*VitA and *w*VitB. The reduction of the *w*One genome size can be explained by the
307 smaller percentage of repetitive elements identified in the draft genome. Repetitive regions might
308 be underrepresented in the *w*One draft genome due to elevated average depth. However, this will
309 not account for all genome size differences because the median coverage measured in our study
310 can tolerate the fluctuation in sequence depth, and the linked-reads can assemble through
311 difficult regions. Most of the genomic regions in *w*One draft genome aligned well with their
312 corresponding regions in *w*VitA with very few rearrangements, indicating the high syntenic
313 conservation between these two strains (Figure 3A). However, there are more structural
314 differences between A-*Wolbachia* and B-*Wolbachia*, which were supported by the whole
315 genome alignments between A-*Wolbachia* (*w*One and *w*VitA) and B-*Wolbachia* *w*VitB (Figure
316 3B and C). The same pattern was observed when comparing gene contents among these three
317 *Wolbachia* in *Nasonia*, *w*One and *w*VitA shared more orthologous genes (Figure 3D). The
318 results indicate that A and B *Wolbachia* retain their genetic differences even when they infect the
319 same host, which suggests that recombination among them is not common, with the exception of
320 phage related genes [52].

321 We have sufficient sequence coverage in *w*OneA1 genome to identify segregating SNPs in
322 coding genes. These genes are enriched for transmembrane transporter function. Due to the
323 intercellular life style, the membrane proteins are critical for *Wolbachia* infection, nutrient
324 uptake and other interactions with the host cells. For SNPs with higher allele frequencies, they
325 might be fixed differences in recently duplicated genes with multiple copies. These
326 polymorphisms may be maintained by selection in the A1 strain to provide an advantage

327 compared to A2 and B. Further study is in needed to determine how these polymorphisms affect
328 gene functions, and how they are maintained in populations that effectively pass through genetic
329 bottlenecks each generation through the egg. Balancing polymorphisms are suggested by this
330 observation, which would be an intriguing topic for future study.

331

332

333 ## Materials and Methods

334 **Sample collection, genomic DNA extraction, 10X Genomic library preparation and genome**
335 **sequencing.**

336 Genomic DNA sample was extracted from 24-hour male adults of the *N. oneida* NONY
337 strain. MagAttract HMW DNA Mini Kit (Qiagen, MD) was used to isolate high molecular
338 weight genomic DNA. The quality of extracted gDNA was checked on a Qubit 3.0 Fluorometer
339 (Thermo Fisher Scientific, USA). The size distribution of the extracted gDNA was accessed
340 using the genomic DNA kit on Agilent TapeStation 4200 (Agilent technologies, CA). A 10X
341 Genomic library was constructed by using the Chromium Genome Reagent Kits v2 on 10X
342 Chromium Controller (10X Genomics Inc., CA). Chromium i7 Sample Index was used as library
343 barcode. Post library construction quality control was accessed with Qubit 3.0 Fluorometer and
344 Agilent TapeStation 4200. The constructed 10X genomic library was sequenced on a HiSeq X
345 sequencer at the Genomic Services Lab at the HudsonAlpha Institute for Biotechnology.

346 **Genome assembly and annotation of *w*One genome using linked reads.**

347 The *N. oneida* genome was assembled using the `Supernova 2.0` assembler [37] with a
348 barcode subsampling parameter of 25%. The following steps were conducted to identify *w*One
349 scaffolds in the *N. oneida* assembly (Figure 1): 1) all sequencing reads were aligned to the *N.*
350 *oneida* assembly to calculate the average and median coverage for each scaffold in the assembly;
351 2) *N. oneida* scaffolds are aligned to the bacterial sequence database using BLAT version 3.5
352 [38] to determine the percent of sequence identity to known *Wolbachia* sequences; 3) assign the
353 scaffolds to *w*One genome version 1. A scaffold was assigned to *w*One if it has at least 20%
354 sequence identity with known *Wolbachia* sequences and an average coverage of 30X or more.

355 To make use of all 10X reads, an independent assembly was performed using the MEGAHIT
356 assembler [53] with all reads. The version 1 *w*One assembly was aligned to the *N. oneida*
357 MEGAHIT assembly, and the average coverage was calculated for each scaffold in the
358 *Wolbachia* sequences to improve the *w*One assembly. The final *w*One genome assembly (version
359 2) was used in the subsequent analyses. Gene annotation was conducted through the NCBI
360 Prokaryotic Genome Annotation Pipeline [54] and tRNA genes were predicted by tRNAscan_SE
361 [55].

362 **Comparative analysis of *Wolbachia* genomes in the *Nasonia* genus**

363 To compare the genome structure among three sequenced *Wolbachia* genomes in *Nasonia*, we
364 first conducted whole genome alignment of *w*One, *w*VitA (GenBank accession
365 GCA_001983615.1) and *w*VitB (GenBank accession GCA_000204545.1) [23] genomes using
366 NUCmer in the MUMmer program suite with default parameter settings [56]. The pairwise
367 alignments (match length longer than 500bp) were visualized using Mummerplot [56].
368 Orthologous gene sets between *w*One and two other *Wolbachia* in *Nasonia* were generated based
369 on reciprocal best hits using BLAST with an E-value cutoff 10^{-5}.

370 **Phylogenomics analysis of 34 genome sequenced *Wolbachia* strains**

371 To examine the phylogeny of *Wolbachia* at the genome level, we conducted phylogenomic
372 analysis using *w*One and 33 other sequenced *Wolbachia* genomes, including *w*Mel (GenBank
373 accession GCA_000008025.1) [17], *w*Bm (GenBank accession GCA_000008385.1) [18], *w*Oo
374 (GenBank accession GCA_000306885.1) [19], *w*Pip (GenBank accession GCA_000073005.1)
375 [20], *w*Ri (GenBank accession GCA_000022285.1) [22], *w*VitB (GenBank accession
376 GCA_000204545.1) [23], *w*Bol1 (GenBank accession GCA_000333775.1) [24], *w*Ha (GenBank

377 accession GCA_000376605.1) [25], *w*No (GenBank accession GCA_000376585.1) [25],
378 *w*Gmm (GenBank accession GCA_000689175.1) [41], *w*AlbB (GenBank accession
379 GCA_000242415.3) [42], *w*VitA (GenBank accession GCA_001983615.1) [26], *w*Di (GenBank
380 accession GCA_000331595.1) [43], *w*Suzi (GenBank accession GCA_000333795.2) [44], *w*Tpre
381 (GenBank accession GCA_001439985.1) [45], *w*Wb (GenBank accession GCA_002204235.2)
382 [46, 57], *w*Ov (GenBank accession GCA_000530755.1) [57], *w*Pip_Mol (GenBank accession
383 GCA_000723225.2) [58], *w*Au (GenBank accession GCA_000953315.1) [59], Ob_Wba
384 (GenBank accession GCA_001266585.1) [60], *w*Cle (GenBank accession GCA_000829315.1)
385 [61], *w*Fol (GenBank accession GCA_001931755.2) [62], *w*Con (GenBank accession
386 GCA_003344345.1) [63], *w*DacA (GenBank accession GCA_001648025.1), wDacB (GenBank
387 accession GCA_001648015.1), *w*MelPop (GenBank accession GCA_000475015.1), *w*Nfe
388 (GenBank accession GCA_001675785.1), *w*Npa (GenBank accession GCA_001675775.1),
389 *w*Nleu (GenBank accession GCA_001675715.1), *w*Nfla (GenBank accession
390 GCA_001675695.1), *w*Spac (GenBank accession GCA_002300525.1) and *w*Stri (GenBank
391 accession GCA_001637495.1). Homologous genes and ortholog clusters among all 34
392 *Wolbachia* genomes were determined by using OrthoFinder [64] with default settings. The 211
393 core single-copy genes shared in all 34 *Wolbachia* genomes were aligned with MAFFT [65] at
394 both nucleotide and protein sequence level. These single-gene alignments were concatenated into
395 one alignment to use in the subsequent phylogenetic analysis. A Neighbor-Joining tree was
396 constructed with alignment of coding sequences of these core genes using MEGA 7.0 [66]. The
397 evolutionary distances were computed using the Nei-Gojobori method in the units of the number
398 of synonymous substitutions per synonymous site [67]. Single gene trees for these 211 core
399 genes were constructed with the GTRGAMMA model and 1000 bootstrap replicates using

400 RAxML v8.2 [68] to check the consistency of supergroup classification . In addition, for
401 phylogenetic analysis of protein sequences from this core gene set, the best-fit model of protein
402 evolution was searched by ProtTest 3 [69]. The final maximum-likelihood phylogenetic tree was
403 inferred by using RAxML v8.2 [68] with the FLU protein model (best fit model identified by
404 ProtTest 3 and 1000 rapid bootstrap replicates.

405 **Phylogenetic analysis of *Wolbachia* in *Nasonia* using MLST genes.**

406 The five MLST (Multi Locus Sequence Typing) genes [12, 70] were examined to further
407 characterize the phylogenetic relationships of *Wolbachia* strains in *Nasonia*. These genes include
408 *gatB* (aspartyl/glutamyl-tRNA (Gln) amidotransferase, subunit B), *coxA* (cytochrome c oxidase,
409 subunit I), *hcpA* (conserved hypothetical protein), *ftsZ* (cell division protein) and *fbpA* (fructose-
410 bisphosphate aldolase). The *w*One MLST genes were identified on five different genome
411 scaffolds, including *coxA* on SCAFFOLD17, *gatB* on SCAFFOLD28, *hcpA* on SCAFFOLD47,
412 *ftsZ* on SCAFFOLD73 and *fbpA* in SCAFFOLD76. MLST gene sequences of the following
413 strains in different hosts were included in the analysis: *w*NvitA, *w*NvitB in N. *vitripennis*;
414 *w*NgirA1, *w*NgirA2, *w*NgirB in N. *giraulti*; *w*NlonA, *w*NlonBl, *w*NlonB2 in N. *longicornis* [9].
415 Sequences of MLST genes from *Nasonia Wolbachia* strains were downloaded from the MLST
416 database [12]. Multiple sequence alignments were generated using MUSCLE (MUltiple
417 Sequence Comparison by Log-Expectation) with default parameters [71]. Phylogenetic analysis
418 was performed using the Maximum likelihood method in MEGA 7.0 software [66]. Bootstrap
419 tests with 1,000 replicates were used to evaluate the phylogenetic trees.

420 Previous study indicated that sequences of all MLST genes in *w*One are the same as that in
421 *w*Ngir [9]. Therefore, we can use sequences of MLST genes in *w*NgirA2 and *w*NgirB to check if
422 the *w*OneA2 and *w*OneB can be detected in the studied NO. We aligned the sequencing reads to

423 the NG MLST reference sequences from three *Wolbachia* strains in with BWA [9], calculated
424 the coverage, and further examined the alignment of each gene in Integrative Genomics Viewer
425 (IGV) [72].

426 **Strain-typing of *Wolbachia* in *Nasonia* with MLST genes using pyrosequencing.**

427 *Wolbachia* infection types were further checked in the genome sequenced NO in this study,
428 and genomic DNA samples from a recently (Summer 2018) collected wild-type CAR262L strain
429 using allele-specific pyrosequencing. Pyro PCR and sequencing primers were designed to target
430 SNP position in coxA and gatB genes in A1, A2 and B *Wolbachia* using PyroMark Assay
431 Design 2.0 (Qiagen, USA). A complete list of primers sequences could be found in Table S3.
432 The A/G SNP targeted in coxA can separate B-*Wolbachia* from A1/A2-*Wolbachia*, and the C/T
433 SNP in gatB allowed us to distinguish A1-*Wolbachia* from A2/B-*Wolbachia*. Pyrosequencing
434 was performed on a Pyromark Q48 Autoprep instrument (Qiagen, USA) using the PyroMark
435 Q48 Advanced CpG Reagents (Qiagen, USA). Briefly, the target regions in coxA and gatB genes
436 were PCR-amplified using the biotin-labeled forward primers and the reverse primers using
437 template genomic DNA samples. Then, pyrosequencing was performed on a PyroMark Q48
438 Autoprep instrument (Qiagen, USA) using the corresponding sequencing primers by following
439 manufacturer's protocol. The results were analyzed with Pyromark Q48 Autoprep software
440 (Qiagen, USA). Three technical replicates are performed for each sample.

441 ***de novo* SNP calling in *w*One genome**

442 To identify segregating polymorphisms in the *w*One genome, all sequencing reads were
443 aligned to *w*OneA1 genome using BWA software [9]. Initial SNP calling were performed using
444 SAMtools [73]. SNPs were further checked manually in IGV [72] to filtered out low-quality and

problematic SNPs. A total of 68 high-quality SNPs was kept for the subsequent analysis. Coding gene regions were extracted using BEDTools [74] to annotate genic SNPs. Gene Ontology (GO) annotation analysis was done on SNP containing genes using BLAST2GO with an E-value cutoff of 10^{-5} [40].

Acknowledgements

This project is supported by an Auburn University Intramural Grant Program Award to X.W. (AUIGP 180271) and a generous laboratory start-up fund to X.W. from Auburn University College of Veterinary Medicine. This work is supported by the USDA National Institute of Food and Agriculture, Hatch project 1018100. X.X. and W.C. are supported by Auburn University Presidential Graduate Research Fellowship. Contributions of J.H.W. were supported by US NSF IOS 1456233 and the Nathaniel and Helen Wisch Professorship. We thank HudsonAlpha Genomic Services Lab for Illumina sequencing and Ting Li for help with DNA extractions.

References

1. Werren JH, Baldo L, Clark ME. Wolbachia: master manipulators of invertebrate biology. Nat Rev Microbiol. 2008;6(10):741-51.
2. Fenn K, Blaxter M. Wolbachia genomes: revealing the biology of parasitism and mutualism. Trends Parasitol. 2006;22(2):60-5.
3. Werren JH. Biology of Wolbachia. Annu Rev Entomol. 1997;42:587-609.
4. 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(2):215-20.
5. 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(6):e38544.
6. Bailly-Bechet M, Martins-Simoes P, Szollosi GJ, Mialdea G, Sagot MF, Charlat S. How Long Does Wolbachia Remain on Board? Mol Biol Evol. 2017;34(5):1183-93.
7. Werren JH, Windsor DM. Wolbachia infection frequencies in insects: evidence of a global equilibrium? Proc Biol Sci. 2000;267(1450):1277-85.
8. Klopfstein S, van Der Schyff G, Tierney S, Austin AD. Wolbachia infections in Australian ichneumonid parasitoid wasps (Hymenoptera: Ichneumonidae): evidence for adherence to the global equilibrium hypothesis. Biological Journal of the Linnean Society. 2018;123(3):518-34.
9. Raychoudhury R, Baldo L, Oliveira DCSG, Werren JH. Modes of Acquisition of Wolbachia: Horizontal Transfer, Hybrid Introgression, and Codivergence in the Nasonia Species Complex. Evolution. 2009;63(1):165-83.
10. Stouthamer R, Breeuwer JAJ, Hurst GDD. Wolbachia pipientis: Microbial manipulator of arthropod reproduction. Annu Rev Microbiol. 1999;53:71-102.
11. Baldo L, Werren JH. Revisiting Wolbachia supergroup typing based on WSP: spurious lineages and discordance with MLST. Curr Microbiol. 2007;55(1):81-7.
12. Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, et al. Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. Appl Environ Microbiol. 2006;72(11):7098-110.
13. Bordenstein SR, O'Hara FP, Werren JH. Wolbachia-induced incompatibility precedes other hybrid incompatibilities in Nasonia. Nature. 2001;409(6821):707-10.
14. Perrot-Minnot MJ, Guo LR, Werren JH. Single and double infections with Wolbachia in the parasitic wasp Nasonia vitripennis: effects on compatibility. Genetics. 1996;143(2):961-72.
15. Breeuwer JA, Werren JH. Cytoplasmic incompatibility and bacterial density in Nasonia vitripennis. Genetics. 1993;135(2):565-74.
16. Bordenstein SR, Uy JJ, Werren JH. Host genotype determines cytoplasmic incompatibility type in the haplodiploid genus Nasonia. Genetics. 2003;164(1):223-33.
17. Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, et al. Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: A streamlined genome overrun by mobile genetic elements. Plos Biol. 2004;2(3):327-41.
18. Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, et al. The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. Plos Biol. 2005;3(4):e121.

19. Darby AC, Armstrong SD, Bah GS, Kaur G, Hughes MA, Kay SM, et al. Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res. 2012;22(12):2467-77.
20. Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, et al. Genome evolution of Wolbachia strain wPip from the Culex pipiens group. Mol Biol Evol. 2008;25(9):1877-87.
21. Pan X, Luhrmann A, Satoh A, Laskowski-Arce MA, Roy CR. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. Science. 2008;320(5883):1651-4.
22. Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, et al. The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans. Proc Natl Acad Sci U S A. 2009;106(14):5725-30.
23. Kent BN, Salichos L, Gibbons JG, Rokas A, Newton IL, Clark ME, et al. Complete bacteriophage transfer in a bacterial endosymbiont (Wolbachia) determined by targeted genome capture. Genome Biol Evol. 2011;3:209-18.
24. Duplouy A, Iturbe-Ormaetxe I, Beatson SA, Szubert JM, Brownlie JC, McMeniman CJ, et al. Draft genome sequence of the male-killing Wolbachia strain wBol1 reveals recent horizontal gene transfers from diverse sources. BMC Genomics. 2013;14:20.
25. Ellegaard KM, Klasson L, Naslund K, Bourtzis K, Andersson SG. Comparative genomics of Wolbachia and the bacterial species concept. PLoS Genet. 2013;9(4):e1003381.
26. Newton IL, Clark ME, Kent BN, Bordenstein SR, Qu J, Richards S, et al. Comparative Genomics of Two Closely Related Wolbachia with Different Reproductive Effects on Hosts. Genome Biol Evol. 2016;8(5):1526-42.
27. Jiggins FM, von Der Schulenburg JH, Hurst GD, Majerus ME. Recombination confounds interpretations of Wolbachia evolution. Proc Biol Sci. 2001;268(1474):1423-7.
28. Jiggins FM. The rate of recombination in Wolbachia bacteria. Mol Biol Evol. 2002;19(9):1640-3.
29. Werren JH, Bartos JD. Recombination in Wolbachia. Curr Biol. 2001;11(6):431-5.
30. Verne S, Johnson M, Bouchon D, Grandjean F. Evidence for recombination between feminizing Wolbachia in the isopod genus Armadillidium. Gene. 2007;397(1-2):58-66.
31. Ros VI, Fleming VM, Feil EJ, Breeuwer JA. Diversity and recombination in Wolbachia and Cardinium from Bryobia spider mites. BMC Microbiol. 2012;12 Suppl 1:S13.
32. Baldo L, Bordenstein S, Wernegreen JJ, Werren JH. Widespread recombination throughout Wolbachia genomes. Mol Biol Evol. 2006;23(2):437-49.
33. Baldo L, Lo N, Werren JH. Mosaic nature of the wolbachia surface protein. J Bacteriol. 2005;187(15):5406-18.
34. Reuter M, Keller L. High levels of multiple Wolbachia infection and recombination in the ant Formica exsecta. Mol Biol Evol. 2003;20(5):748-53.
35. Ilinsky Y, Kosterin OE. Molecular diversity of Wolbachia in Lepidoptera: Prevalent allelic content and high recombination of MLST genes. Mol Phylogenet Evol. 2017;109:164-79.
36. Foster J, Slatko B, Bandi C, Kumar S. Recombination in wolbachia endosymbionts of filarial nematodes? Appl Environ Microbiol. 2011;77(5):1921-2.
37. Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. Direct determination of diploid genome sequences. Genome Res. 2017;27(5):757-67.
38. Kent WJ. BLAT--the BLAST-like alignment tool. Genome Res. 2002;12(4):656-64.

39. Paraskevopoulos C, Bordenstein SR, Wernegreen JJ, Werren JH, Bourtzis K. Toward a Wolbachia multilocus sequence typing system: discrimination of Wolbachia strains present in Drosophila species. Curr Microbiol. 2006;53(5):388-95.
40. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674-6.
41. Brelsfoard C, Tsiamis G, Falchetto M, Gomulski LM, Telleria E, Alam U, et al. Presence of extensive Wolbachia symbiont insertions discovered in the genome of its host Glossina morsitans morsitans. PLoS Negl Trop Dis. 2014;8(4):e2728.
42. Mavingui P, Valiente Moro C, Tran-Van V, Wisniewski-Dye F, Raquin V, Minard G, et al. Whole-genome sequence of Wolbachia strain wAlbB, an endosymbiont of tiger mosquito vector Aedes albopictus. J Bacteriol. 2012;194(7):1840.
43. Saha S, Hunter WB, Reese J, Morgan JK, Marutani-Hert M, Huang H, et al. Survey of endosymbionts in the Diaphorina citri metagenome and assembly of a Wolbachia wDi draft genome. PLoS One. 2012;7(11):e50067.
44. Siozios S, Cestaro A, Kaur R, Pertot I, Rota-Stabelli O, Anfora G. Draft Genome Sequence of the Wolbachia Endosymbiont of Drosophila suzukii. Genome Announc. 2013;1(1).
45. Lindsey AR, Werren JH, Richards S, Stouthamer R. Comparative Genomics of a Parthenogenesis-Inducing Wolbachia Symbiont. G3 (Bethesda). 2016;6(7):2113-23.
46. Chung M, Small ST, Serre D, Zimmerman PA, Dunning Hotopp JC. Draft genome sequence of the Wolbachia endosymbiont of Wuchereria bancrofti wWb. Pathog Dis. 2017;75(9).
47. Fountain MT, Hopkin SP. Folsomia candida (Collembola): a "standard" soil arthropod. Annu Rev Entomol. 2005;50:201-22.
48. Czarnetzki AB, Tebbe CC. Detection and phylogenetic analysis of Wolbachia in Collembola. Environ Microbiol. 2004;6(1):35-44.
49. Vandekerckhove TTM, Watteyne S, Willems A, Swing JG, Mertens J, Gillis M. Phylogenetic analysis of the 16S rDNA of the cytoplasmic bacterium Wolbachia from the novel host Folsomia candida (Hexapoda, Collembola) and its implications for wolbachial taxonomy. Fems Microbiology Letters. 1999;180(2):279-86.
50. Ma Y, Chen WJ, Li ZH, Zhang F, Gao Y, Luan YX. Revisiting the phylogeny of Wolbachia in Collembola. Ecol Evol. 2017;7(7):2009-17.
51. Bleidorn C, Gerth M. A critical re-evaluation of multilocus sequence typing (MLST) efforts in Wolbachia. FEMS Microbiol Ecol. 2018;94(1).
52. Bordenstein SR, Wernegreen JJ. Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol. 2004;21(10):1981-91.
53. Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods. 2016;102:3-11.
54. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016;44(14):6614-24.
55. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955-64.
56. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004;5(2):R12.

57. Desjardins CA, Cerqueira GC, Goldberg JM, Dunning Hotopp JC, Haas BJ, Zucker J, et al. Genomics of Loa loa, a Wolbachia-free filarial parasite of humans. Nat Genet. 2013;45(5):495-500.
58. Pinto SB, Stainton K, Harris S, Kambris Z, Sutton ER, Bonsall MB, et al. Transcriptional Regulation of Culex pipiens Mosquitoes by Wolbachia Influences Cytoplasmic Incompatibility. Plos Pathog. 2013;9(10).
59. Sutton ER, Harris SR, Parkhill J, Sinkins SP. Comparative genome analysis of Wolbachia strain wAu. BMC Genomics. 2014;15:928.
60. Derks MF, Smit S, Salis L, Schijlen E, Bossers A, Mateman C, et al. The Genome of Winter Moth (Operophtera brumata) Provides a Genomic Perspective on Sexual Dimorphism and Phenology. Genome Biol Evol. 2015;7(8):2321-32.
61. Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. Evolutionary origin of insect-Wolbachia nutritional mutualism. Proc Natl Acad Sci U S A. 2014;111(28):10257-62.
62. Faddeeva-Vakhrusheva A, Kraaijeveld K, Derks MFL, Anvar SY, Agamennone V, Suring W, et al. Coping with living in the soil: the genome of the parthenogenetic springtail Folsomia candida. Bmc Genomics. 2017;18.
63. Badawi M, Moumen B, Giraud I, Greve P, Cordaux R. Investigating the Molecular Genetic Basis of Cytoplasmic Sex Determination Caused by Wolbachia Endosymbionts in Terrestrial Isopods. Genes (Basel). 2018;9(6).
64. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.
65. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. Methods Mol Biol. 2014;1079:131-46.
66. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016;33(7):1870-4.
67. Nei M, Gojobori T. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. Molecular Biology and Evolution. 1986;3(5):418-26.
68. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312-3.
69. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics. 2011;27(8):1164-5.
70. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595.
71. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792-7.
72. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92.
73. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.
74. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics. 2014;47:11 2 1-34.

Tables

Table 1. *w*One assembly summary statistics and comparison with *w*VitA and *w*VitB genomes.

	Number of contigs	Number of scaffolds	Contig N50 (kb)	Scaffold N50 (kb)	Number of proteins	Assembled genome size (bp)
***w*One**	347	254	3.81	4.74	948	938,765
***w*VitA**	142	N/A	13.38	N/A	1042	1,211,929
***w*VitB**	509	426	5.79	6.21	845	1,107,643

Table 2. Number of IS (Insertion element) in *w*One, *w*VitA and *w*VitB.

Family	Number in *w*One	Number in *w*VitA	Number in *w*VitB
IS3	1	0	0
IS4	6	1	0
IS5	3	1	1
IS5/IS1182	2	1	0
IS6	0	0	2
IS66	0	2	0
IS110	5	2	0
IS200/IS605	2	0	1
IS256	1	1	2
IS481	6	0	2
IS630	0	0	2
IS982	1	1	0

Figures

Figure 1. The workflow of *w*One genome assembly using 10X Genomics linked reads.

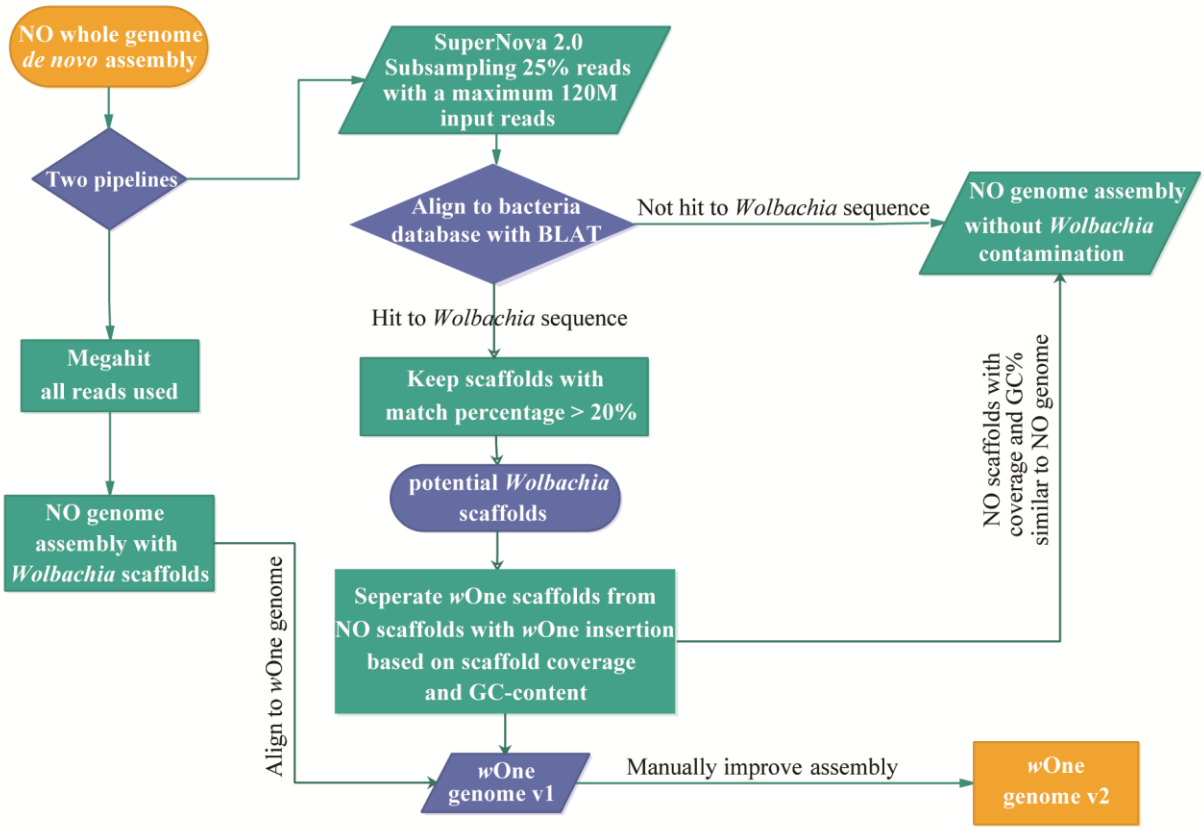


Figure 2. Comparison of average coverage and GC content between *Nasonia oneida* (NO) genome and *w*One genome.

(A) scatter plot of average coverage and GC% indicating clear separation of NO and *w*One genomes, blue square represents *w*One genome and red dot represents NO genome, the NO mtDNA was labeled as green triangle; (B) box plot of average coverage between NO and *w*One genomes, green line indicates the coverage of NO mtDNA; (B) box plot of GC% between NO and *w*One genomes, green line indicates the GC% of NO mtDNA.

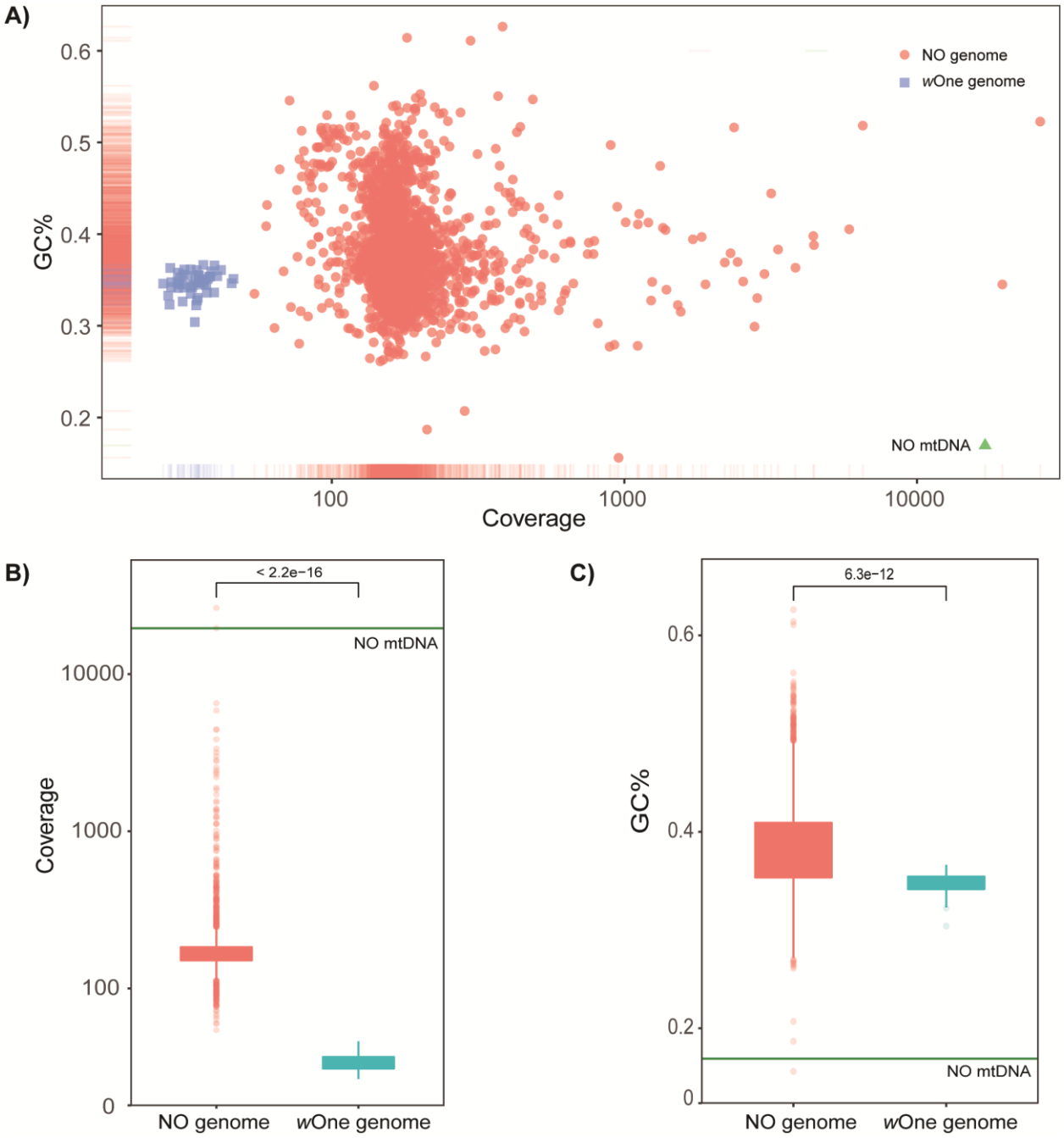


Figure 3. Comparative genomic analysis of *w*One, *w*VitA and *w*VitB genomes.
(A) Dot plot showing comparison between *w*One and *w*VitA genomes, red for a forward match and blue for a reverse match; (B) Dot plot showing comparison between *w*One and *w*VitB genomes; (C) Dot plot showing comparison between *w*VitA and *w*VitB genomes; (D) Venn Diagram showing comparison of genes and pseudogenes in *w*One, *w*VitA and *w*VitB.

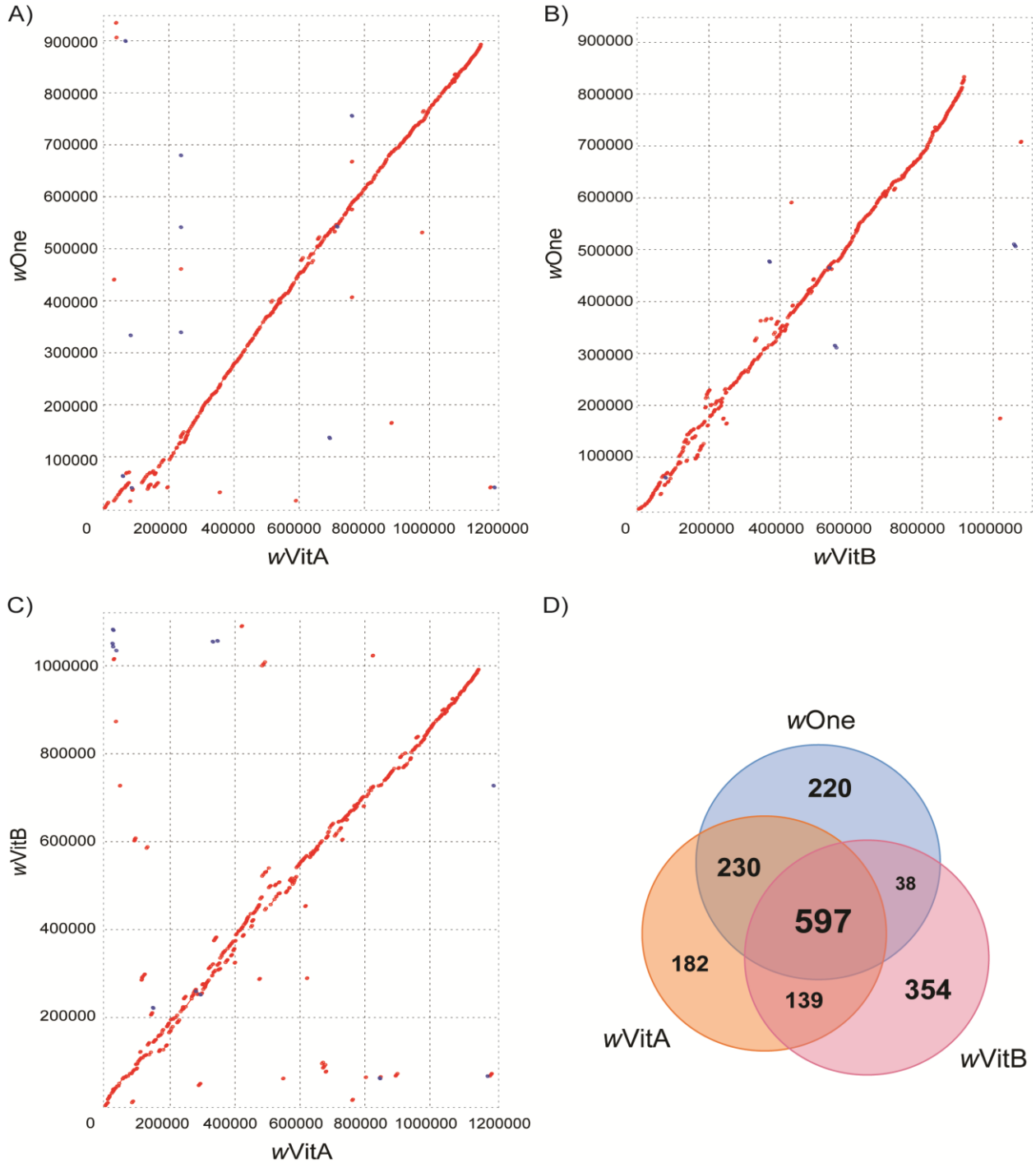


Figure 4. Phylogenomic relationships of 34 *Wolbachia* strains.
The phylogenetic tree was constructed using maximum likelihood method from a concatenated coding sequence alignment of 211 single-copy orthologous genes with RAxML. Numbers on the branches represent the support from 1000 bootstrap replicates. The supergroup classifications (A-F) represent following the color code. The host taxonomic classifications for most of *Wolbachia* strains are shown in different color code.

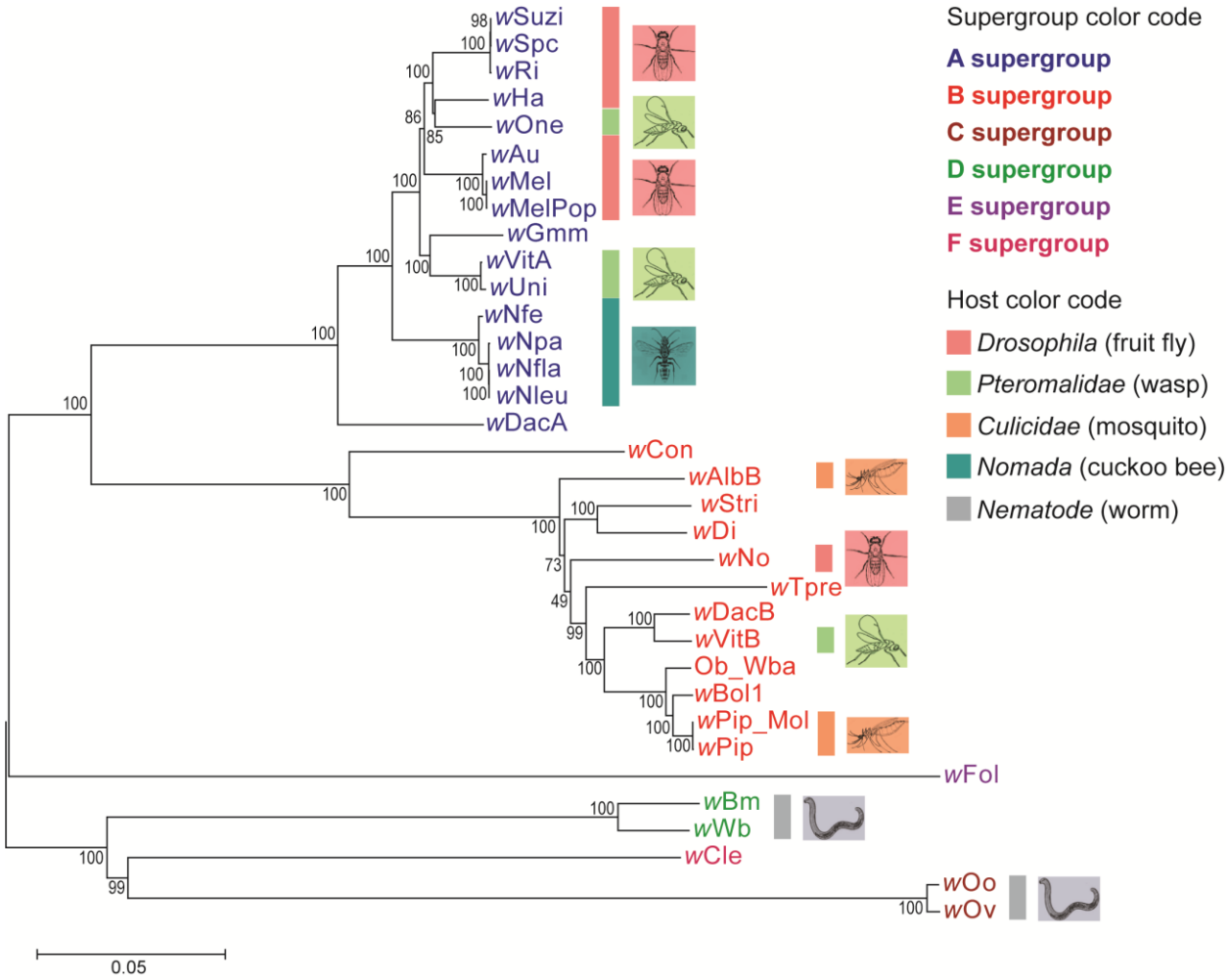


Figure 5. RAxML tree revealed lateral gene transfer between *w*One and B-*Wolbachia*.

The supergroup classifications follow the color code in Figure 4. Nucleotides at selected positions are shown in the right panels (green:A; blue: C; yellow: G; pink: T). Hypothetical protein DEF52_00180 from *w*One (A supergroup) clusters with B-*Wolbachia*.

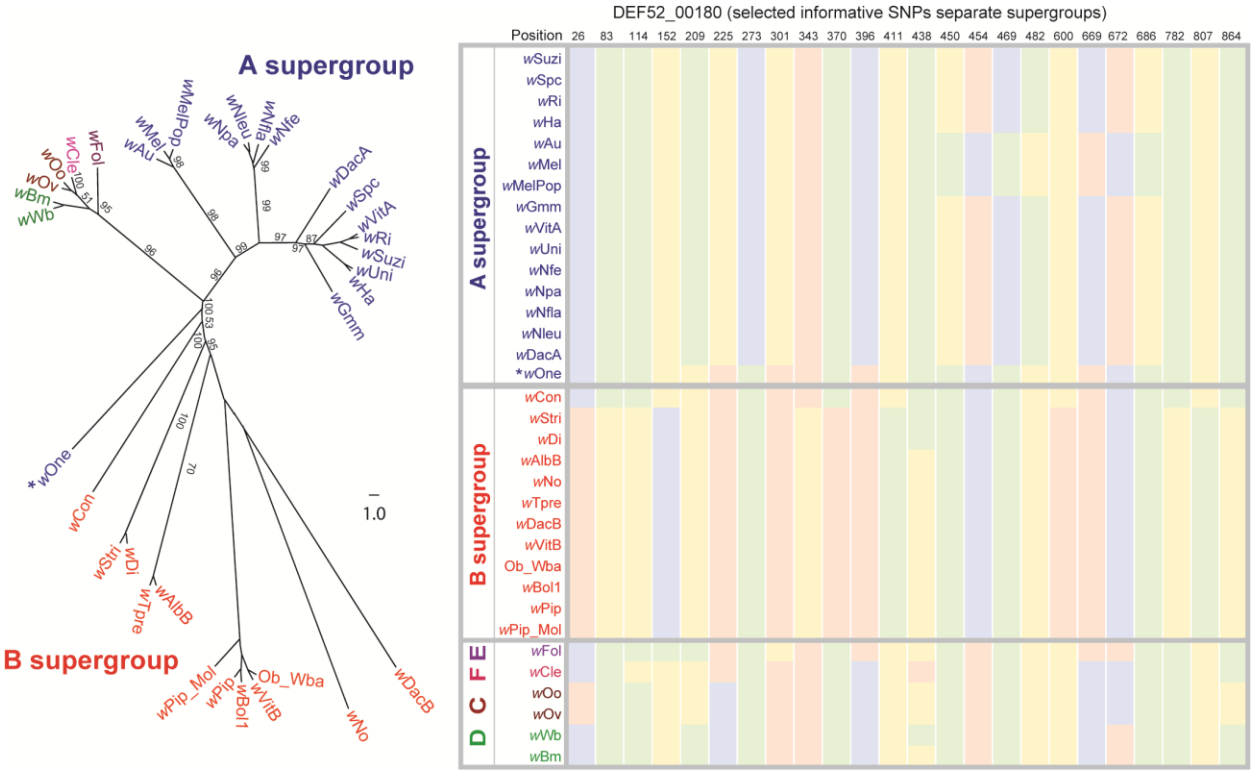


Figure 6. RAxML trees revealed lateral gene transfer between A-*Wolbachia* and B/E-*Wolbachia* in coxB and DEF52_04830 genes.

The supergroup classifications follow the color code in earlier figures. Nucleotides at selected positions are shown in the right panels. (A) In ML tree of coxB, *w*AlbB (B supergroup) and *w*Fol (E supergroup) cluster with A-*Wolbachia*, respectively; (B) In ML tree of hypothetical protein DEF52_04830, *w*Con, *w*Di and *w*Tpre and *w*AlbB (B supergroup) cluster with A-*Wolbachia*, and *w*Fol (E supergroup) clusters with A-*Wolbachia*.

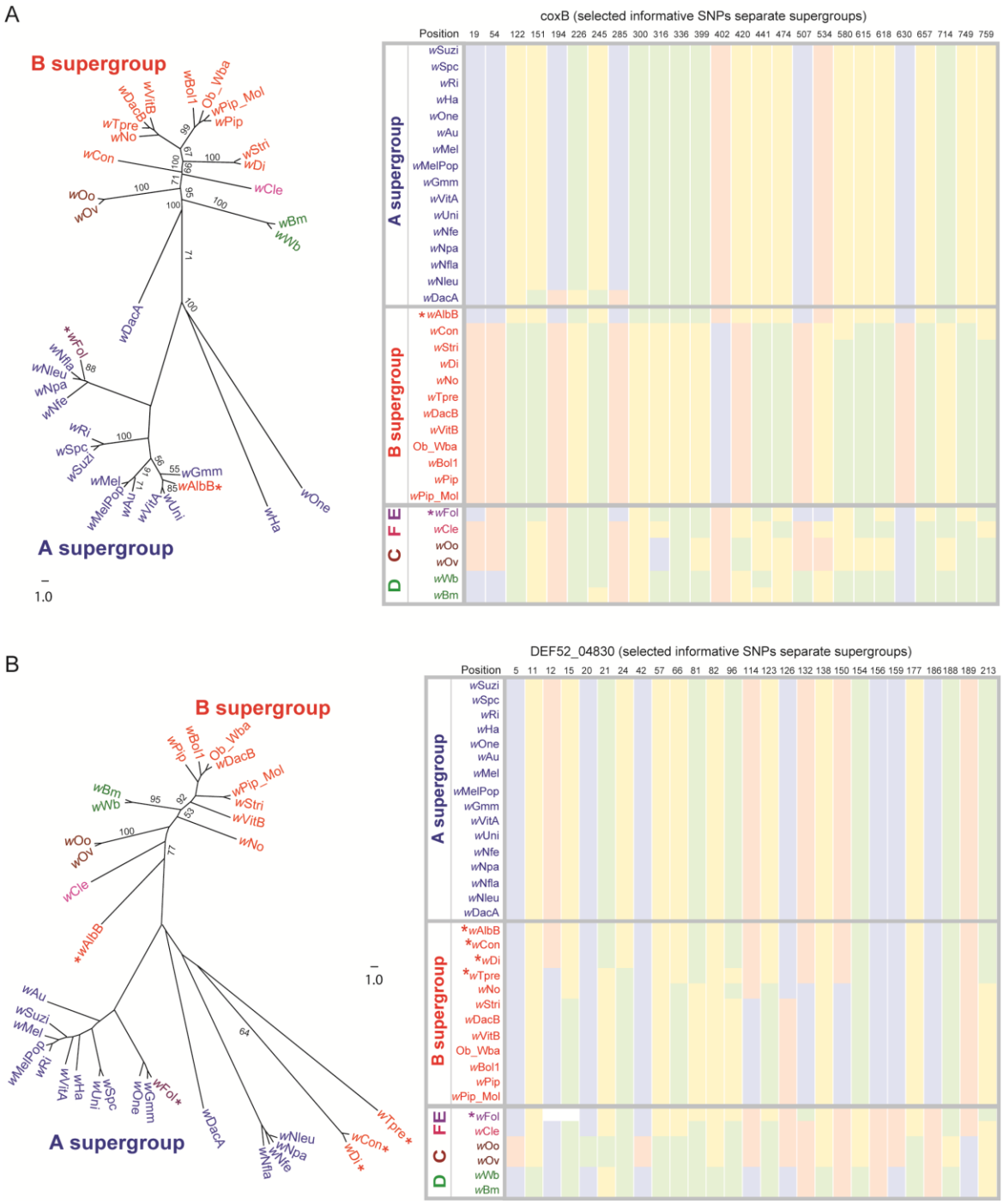


Figure 7. RAxML trees revealed lateral gene transfer between A-*Wolbachia* and B/E-*Wolbachia* in nuoC and DnaK genes.

The supergroup classifications follow the color code in earlier figures. Nucleotides at selected positions are shown in the right panels. (A) in ML tree of nuoC, *w*Con (B supergroup) clusters with A-*Wolbachia*. (B) in ML tree of DnaK, *w*DacA (A supergroup) clusters with *w*Fol (E supergroup).

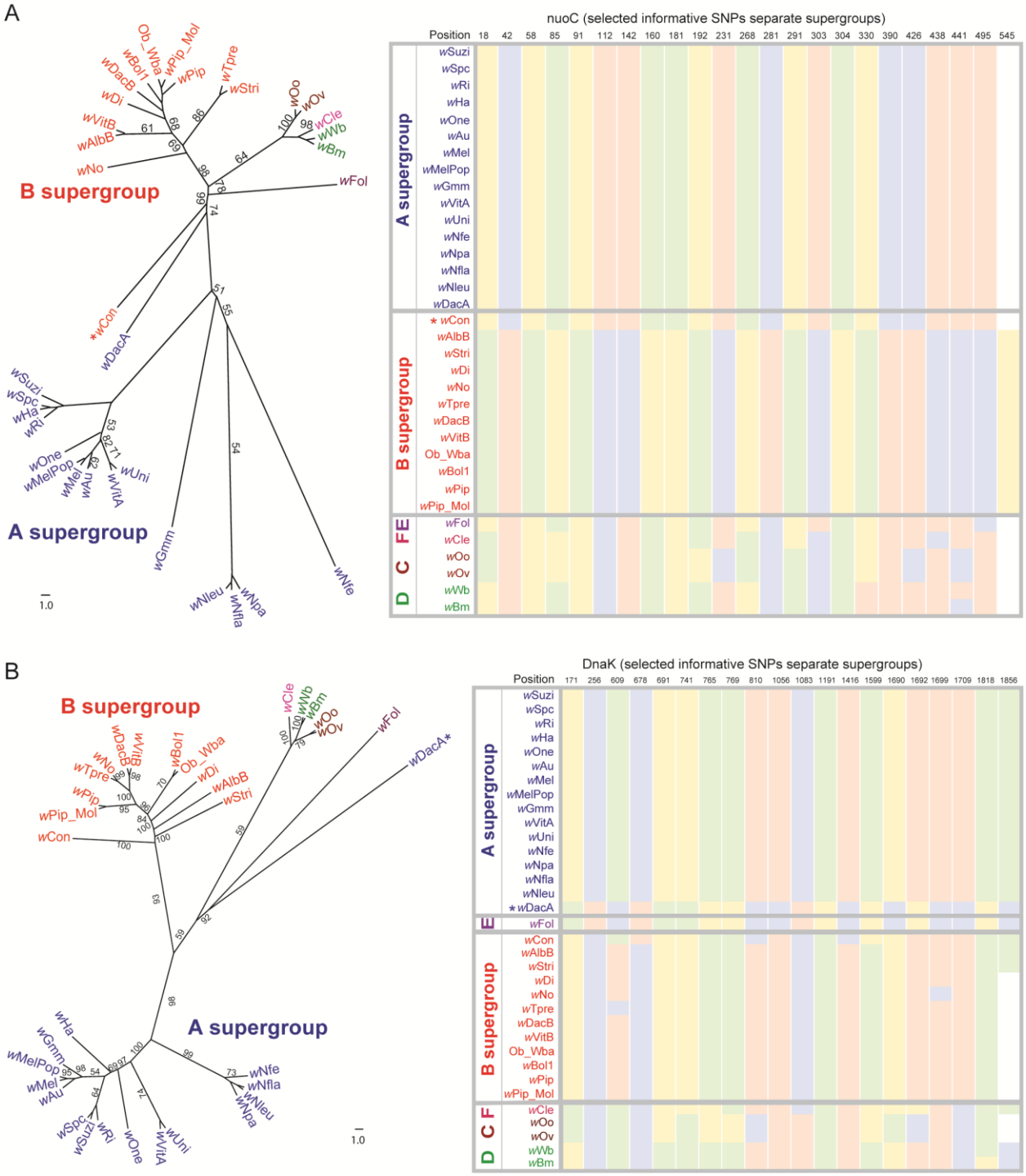


Figure 8. Distinct tree structure difference revealed in argS gene.

The supergroup classifications follow the color code in earlier figures. (A) Tree structure of argS (from starting site to 570 bp) revealed that *w*Di (B supergroup) clusters with A-*Wolbachia*; (B) Tree structure of argS (from 1141 bp to stop site) revealed that *w*DacA (A supergroup) and *w*Fol (E supergroup) clusters with B-*Wolbachia*, indicating intragenic recombination events; (C) Nucleotides at selected positions (1-570 bp in argS) supported the tree topology in (A); (D) Nucleotides at selected positions (1141-1708 bp in argS) supported the tree topology in (B).

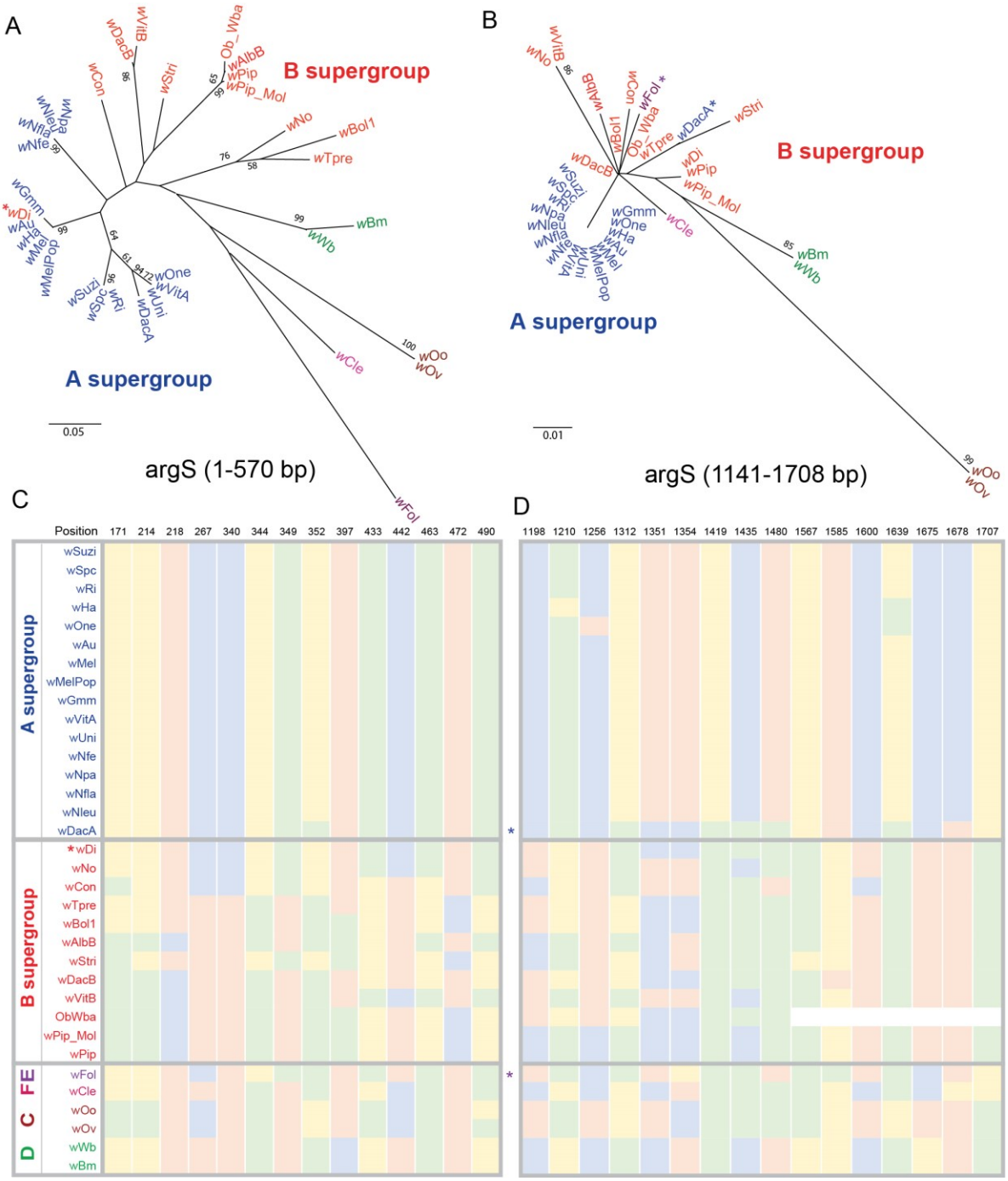
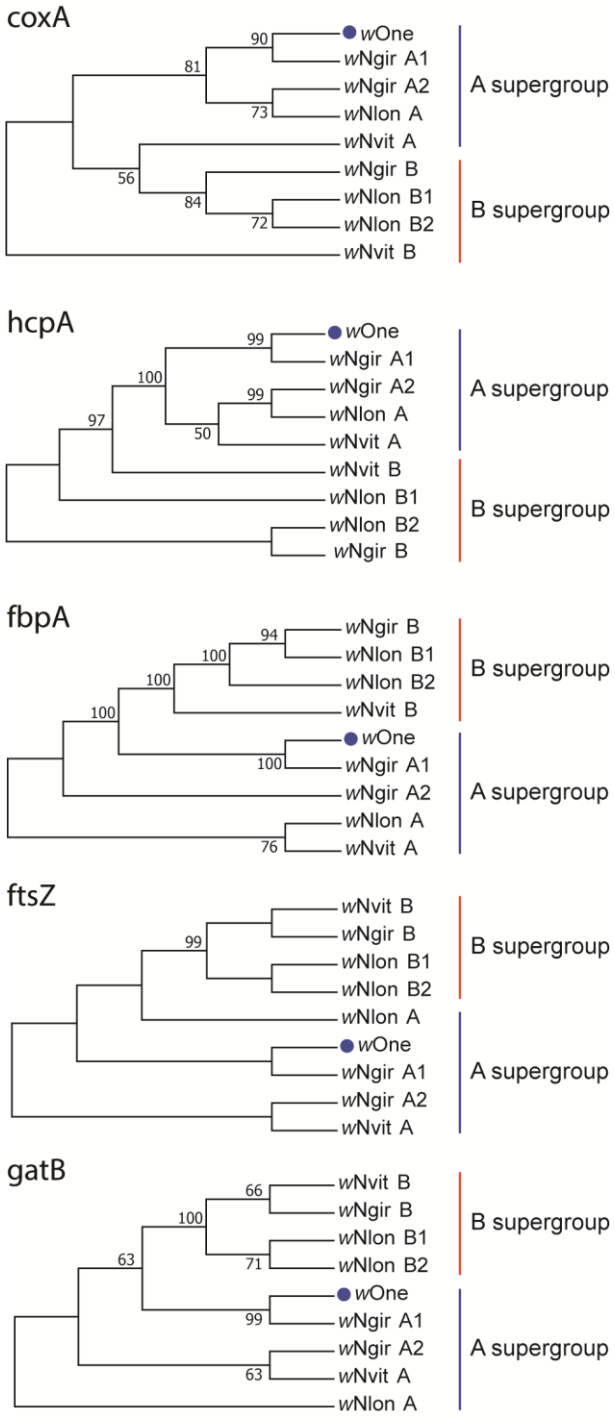


Figure 9. Phylogenetic tree of *Wolbachia* in *Nasonia* with Maximum Likelihood method based on five MLST genes.

The bootstrap consensus tree inferred from 1000 replicates. A and B supergroups were clustered into two groups in trees of all five MLST genes.



Supplemental Tables

Table S1. Summary of current sequenced *Wolbachia* whole genomes.

Strain	Super group	Host species	Genome size (Mb)	Method of bacterial DNA purification	Assembly method	Reference
*w*Mel	A	*Drosophila melanogaster*	1.268	Purified from adult flies on pulsed-field gels	TIGR Assembler	[17]
*w*Bm	D	*Brugia malayi*	1.080	Selected *w*Bm BACs from host BAC library	PHRED-PHRAP-CONSED	[18]
*w*Pip	B	*Culex pipiens*	1.482	Filter purification from preblastoderm embryos	PHRAP	[20]
*w*Ri	A	*Drosophila simulans*	1.446	Concentrated *w*Ri cells using renografin gradient and plug agarose to purify DNA	PHRED-PHRAP-CONSED	[22]
*w*VitB	B	*Nasonia vitripennis*	1.108	Enrichment of *w*VitB DNA with a high-density tiled oligonucleotide array	Velvet assembler	[23]
*w*AlbB	B	*Aedes albopictus*	1.240	Extracted from infected mosquito cell line and Multiple-displacement amplification (MDA) to purify DNA	Newbler	[42]
*w*Di	B	*Diaphorina citri*	1.241	N/A	*w*Di reads filtered from its host genome sequences	[43]
*w*Oo	C	*Onchocerca ochengi*	0.960	N/A	Generated *w*Oo genome from host assembly	[19]
*w*Ha	A	*Drosophila simulans*	1.296	Filter purification of *w*Ha cells from host preblastoderm embryos and performed MDA to purify DNA	Newbler and MIRA	[25]
*w*No	B	*Drosophila simulans*	1.302	Same method as used for *w*Ha	Newbler and MIRA	[25]
*w*Suzi	A	*Drosophila suzukii*	1.415	N/A	*w*Suzi reads filtered from its host genome sequences and assembled with Velvet and MIRA	[44]
*w*Bol1	B	*Hypolimnas bolina*	1.378	Filter purification from *w*Bol1-b-infected cells	AGRF	[24]

				and extra separation step on percoll gradient		
*w*Ov	C	*Onchocerca volvulus*	0.961	N/A	*w*Ov reads filtered from host genome sequences and assembled with Newbler	[57]
*w*Pip_Mol	B	*Culex molestus*	1.436	Unknown	Newbler	[58]
*w*Gmm	A	*Glossina morsitans morsitans*	1.020	*w*Gmm DNA extracted from ovaries of host adult female	MIRA	[41]
*w*Cle	F	*Cimex lectularius*	1.250	Extracted DNA from bacteriomes of bedbugs	PHRED-PHRAP-CONSED	[61]
*w*Au	A	*Drosophila simulans*	1.268	Filtration from host adults / cell lines and DNAse treatment	Hierarchical Genome Assembly Process (HGAP)	[59]
Ob_Wba	B	*Operophtera brumata*	1.121	N/A	Extracted Ob_Wba reads from host sequence reads and assembled with Celera	[60]
*w*VitA	A	*Nasonia vitripennis*	1.212	*w*VitA DNA purified with filter column from host pupa	Velvet and Newbler	[26]
*w*Uni	A	*Muscidifurax uniraptor*	1.049	Same method as used for *w*VitA	Velvet and Newbler	[26]
*w*Tpre	B	*Trichogramma pretiosum*	1.134	N/A	*w*Tpre sequences queried against *T. pretiosum* genome assembly	[45]
*w*Wb	D	*Wuchereria bancrofti*	1.061	N/A	Extracted wWb sequences from WGS of W. bancrofti, assembled with SPAdes	[46]
*w*Fol	E	*Folsomia candida*	1.802	N/A	Assembled from the host scaffolds	[62]
*w*Con	B	*Cylisticus convexus*	2.110	*w*Con DNA extracted from ovaries of host adult female	Newbler	[63]
*w*Spac	A	*Drosophila subpulchrella*	1.420	Unknown	ABySS	In press
*w*Stri	B	*Laodelphax striatella*	1.232	Unknown	CLC genomics workbench	Unpublished

*w*DacA	A	*Dactylopius coccus*	1.171	Unknown	MetaVelvet, Newbler, SPAdes, HGAP, and SeqManPro	Unpublished
*w*DacB	B	*Dactylopius coccus*	1.498	Unknown	Same as used for *w*DacA	Unpublished
*w*MelPop	A	*Drosophila melanogaster*	1.239	Unknown	Newbler	Unpublished
*w*Npa	A	*Nomada panzeri*	1.344	Unknown	SPAdes	Unpublished
*w*Nfe	A	*Nomada ferruginata*	1.338	Unknown	SPAdes	Unpublished
*w*Nleu	A	*Nomada leucophthalma*	1.367	Unknown	SPAdes	Unpublished
*w*Nfla	A	*Nomada flava*	1.333	Unknown	SPAdes	Unpublished

Table S2. Strain type of *w*One using informative SNPs in MLST genes.

MLST gene	SNP position	A1 allele	A1 allele count	A2 allele	A2 allele count
gatB	23	G	46	A	0
	27	A	49	G	0
	126	T	52	C	0
	147	G	55	A	0
	231	A	36	G	0
	264	T	40	C	0
	272	T	39	C	0
	292	A	45	C	0
	296	G	45	A	0
	315	C	43	T	0
	366	G	39	T	0
coxA	15	C	41	A	0
	58	T	54	C	0
	104	T	53	C	0
hcpA	126	A	48	G	0
	132	A	43	C	0
	148	A	42	G	0
	150	T	44	G	0
	179	C	51	T	0
	180	C	52	T	0
	192	C	51	T	0
	201	T	54	A	0
	225	C	55	A	0
	234	T	57	C	0
	237	A	57	G	0
	246	T	57	C	0
	249	A	61	G	0
fbpA	45	C	68	T	0
	54	G	65	A	0
	58	A	71	G	0
	63	G	68	A	0
	66	T	69	C	0
	75	A	72	G	0
	102	A	57	G	0
	114	C	57	A	0
	120	G	56	A	0
	126	T	56	C	0
	174	A	54	G	0
	185	A	53	G	0
	190	G	56	A	0
	213	G	52	A	0
	226	T	52	C	0
	252	A	62	G	0
	276	G	68	A	0
	300	A	75	G	0
	354	A	88	G	0
	400	C	77	T	0
	426	A	53	G	0

Table S3. PCR and sequencing primers of coxA and gatB genes used in pyrosequencing for strain-typing of *Wolbachia* in *nasonia*.

Gene	coxA	gatB
A1 allele	A	T
A2 allele	A	C
B allele	G	C
Forward primer	[Biotin]TTATGTTGATTGTTGCCTTACCA	[Biotin]AGGCAGATTTTGCGTTACAT
Reverse primer	CAGGATCAAAAAAGGAAGTACCA	TCACAACGAGTGCCAAATG
Sequencing primer	TCAAAAAAGGAAGTACCA	AACGAGTGCCAAATG

Table S4. List of 44 genic segregating SNPs in *w*One genome.

SNP ID	Scaffold	Position	Alleles	Coverage	Reference allele count	Alternative allele count	Gene ID	SNP effect
*w*OneSNP03	SCAFFOLD2	11049	C/A	97	83	14	yhjJ	nonsynonymous / moderate
*w*OneSNP06	SCAFFOLD2	17857	C/A	64	57	7	trxB	synonymous
*w*OneSNP08	SCAFFOLD8	11842	A/C	88	64	22	secY	nonsynonymous / moderate
*w*OneSNP09	SCAFFOLD12	5242	T/A	99	60	26	rlp	nonsynonymous / moderate
*w*OneSNP10	SCAFFOLD15	1731	A/C	88	61	27	peptidylprolyl isomerase	nonsynonymous / moderate
*w*OneSNP11	SCAFFOLD18	1313	T/C	67	59	7	hypothetical protein DEF52_01265	nonsynonymous / moderate
*w*OneSNP12	SCAFFOLD21	2687	C/G	83	60	18	cysS	nonsynonymous / moderate
*w*OneSNP13	SCAFFOLD21	8235	T/A	76	66	9	glyS	nonsynonymous / moderate
*w*OneSNP15	SCAFFOLD26	4657	A/G	76	66	9	rodA	nonsynonymous / moderate
*w*OneSNP16	SCAFFOLD26	6119	T/C	76	63	11	purF	nonsynonymous / moderate
*w*OneSNP17	SCAFFOLD28	2825	C/T	63	55	8	cox3	synonymous
*w*OneSNP18	SCAFFOLD28	6143	G/A	42	36	6	gatB	nonsynonymous / moderate
*w*OneSNP21	SCAFFOLD31	1640	G/A	47	33	9	clpA	synonymous
*w*OneSNP22	SCAFFOLD31	4365	T/G	48	41	7	TldD/PmbA family protein	nonsynonymous / moderate
*w*OneSNP23	SCAFFOLD34	1724	T/A	56	46	6	holB	nonsynonymous / moderate
*w*OneSNP24	SCAFFOLD41	2368	C/T	71	59	11	ribC	stop_gained / high
*w*OneSNP25	SCAFFOLD44	3152	A/G	54	48	6	hypothetical protein DEF52_02300	synonymous
*w*OneSNP26	SCAFFOLD47	1379	A/T	69	59	8	SPFH domain-containing protein	nonsynonymous / moderate
*w*OneSNP27	SCAFFOLD50	1912	G/A	94	82	12	sensor histidine kinase	synonymous
*w*OneSNP28	SCAFFOLD53	1239	C/T	86	56	28	pth2	nonsynonymous / moderate
*w*OneSNP30	SCAFFOLD59	3903	G/A	96	52	44	argS	nonsynonymous / moderate
*w*OneSNP31	SCAFFOLD60	1139	A/T	57	51	6	hypothetical protein DEF52_02790	nonsynonymous / moderate
*w*OneSNP32	SCAFFOLD64	736	G/A	81	73	8	secA	nonsynonymous / moderate
*w*OneSNP33	SCAFFOLD64	739	T/C	80	71	8	secA	synonymous
*w*OneSNP34	SCAFFOLD69	507	C/T	57	49	6	SAM-dependent methyltransferase	nonsynonymous / moderate
*w*OneSNP35	SCAFFOLD73	211	A/T	52	46	6	ftsZ	nonsynonymous / moderate

*w*OneSNP36	SCAFFOLD73	494	G/A	65	57	7	ftsZ	nonsynonymous / moderate
*w*OneSNP40	SCAFFOLD76	1774	T/G	62	47	11	pyrD	nonsynonymous / moderate
*w*OneSNP42	SCAFFOLD86	1204	C/G	62	53	9	yidC	nonsynonymous / moderate
*w*OneSNP43	SCAFFOLD86	1205	C/G	64	53	9	yidC	nonsynonymous / moderate
*w*OneSNP44	SCAFFOLD103	563	A/T	100	67	30	mutL	nonsynonymous / moderate
*w*OneSNP47	SCAFFOLD104	1038	G/A	73	63	7	virB4	synonymous
*w*OneSNP48	SCAFFOLD121	66	C/T	98	50	47	tuf	synonymous
*w*OneSNP49	SCAFFOLD121	75	A/T	100	45	52	tuf	synonymous
*w*OneSNP50	SCAFFOLD121	78	A/G	100	47	52	tuf	synonymous
*w*OneSNP52	SCAFFOLD136	1974	T/G	96	40	55	transposase	nonsynonymous / moderate
*w*OneSNP53	SCAFFOLD142	2081	A/G	93	48	41	hypothetical protein DEF52_04290	nonsynonymous / moderate
*w*OneSNP54	SCAFFOLD142	2101	G/A	86	44	42	hypothetical protein DEF52_04290	synonymous
*w*OneSNP57	SCAFFOLD155	1642	G/T	83	63	19	hypothetical protein DEF52_04490	nonsynonymous / moderate
*w*OneSNP58	SCAFFOLD164	552	C/T	81	51	27	dnaJ	synonymous
*w*OneSNP59	SCAFFOLD164	573	C/T	78	51	27	dnaJ	synonymous
*w*OneSNP63	SCAFFOLD189	213	T/C	54	48	6	hypothetical protein DEF52_04885	nonsynonymous / moderate
*w*OneSNP65	SCAFFOLD211	392	C/T	76	60	15	ankyrin repeat domain-containing protein	nonsynonymous / moderate
*w*OneSNP66	SCAFFOLD212	762	A/C	79	43	35	ubiH	nonsynonymous / moderate

Supplemental Figures

Figure S1. Phylogenetic analysis of 34 *Wolbachia* genomes using maximum likelihood method from a concatenated protein alignment of 211 single copy orthologous genes.

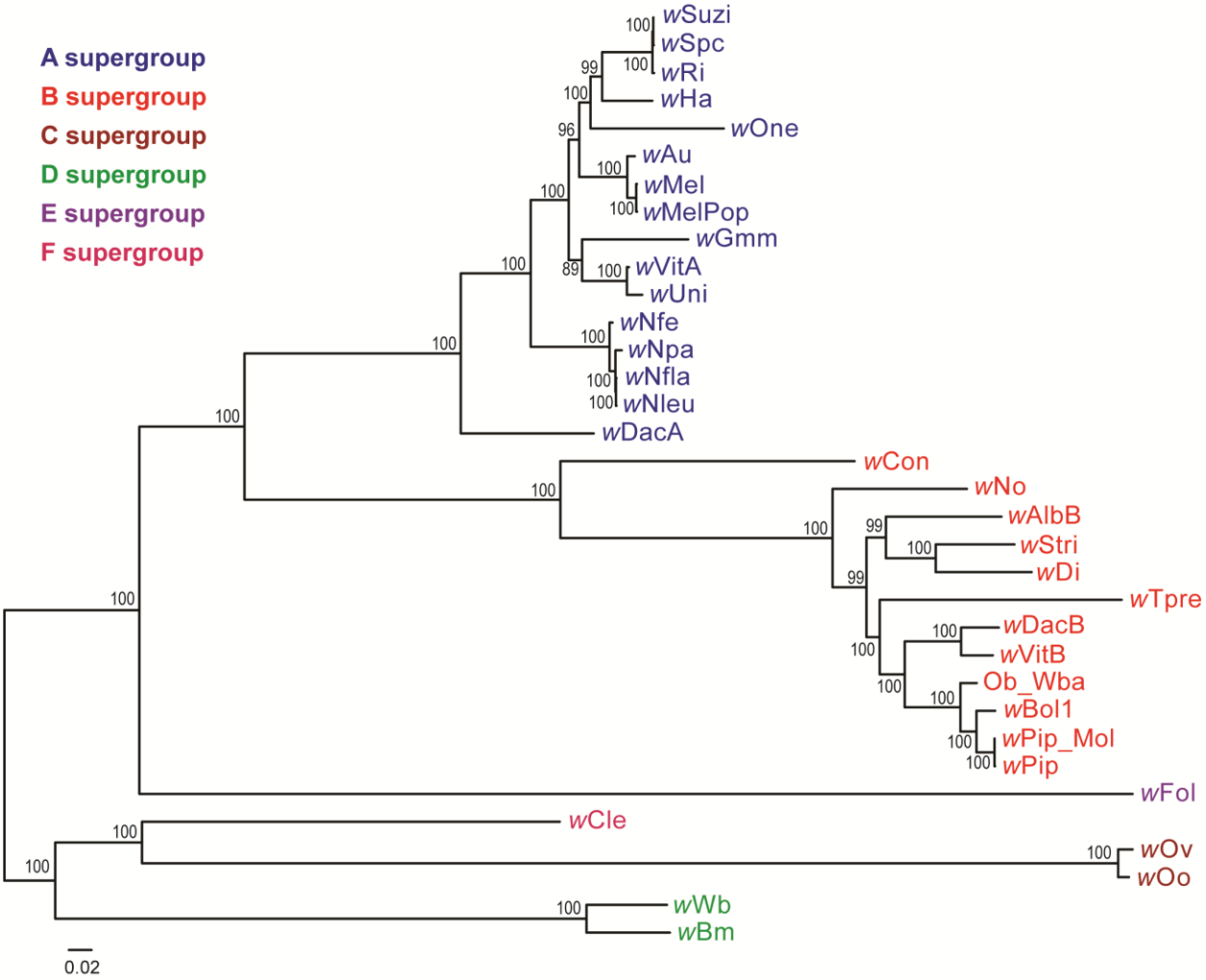


Figure S2. Allele-specific pyrosequencing for *Wolbachia* strain typing in genome sequenced *N. oneida* and the wild strain CAR262L.

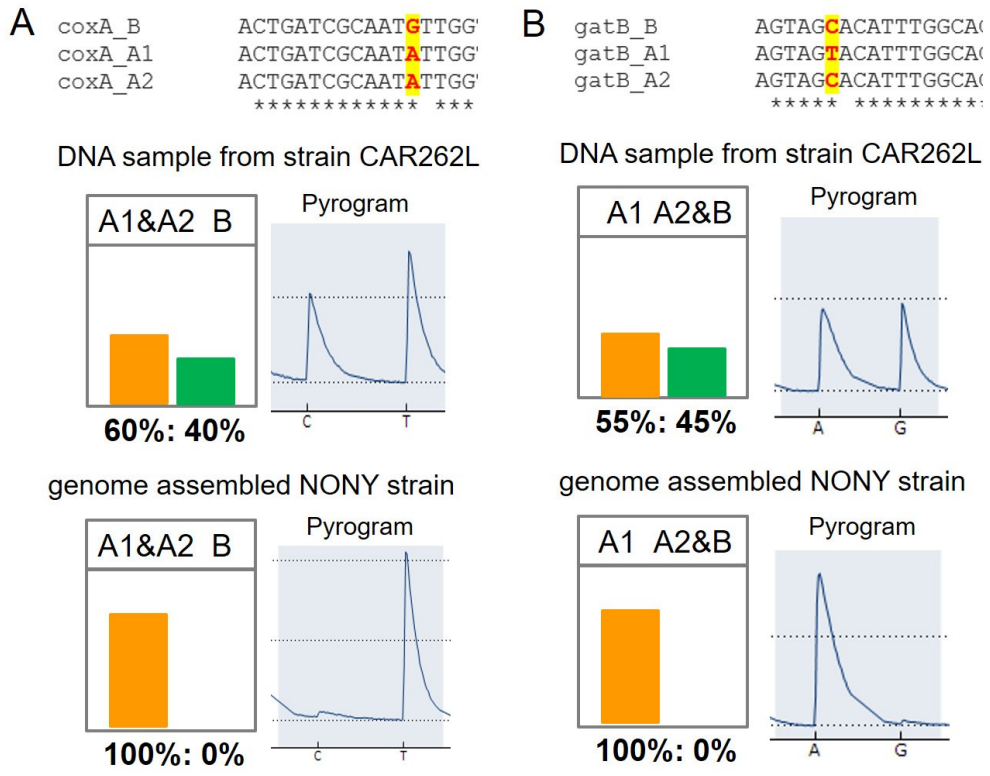


Figure S3. Alternative allele frequency of SNPs identified in *w*One genome.

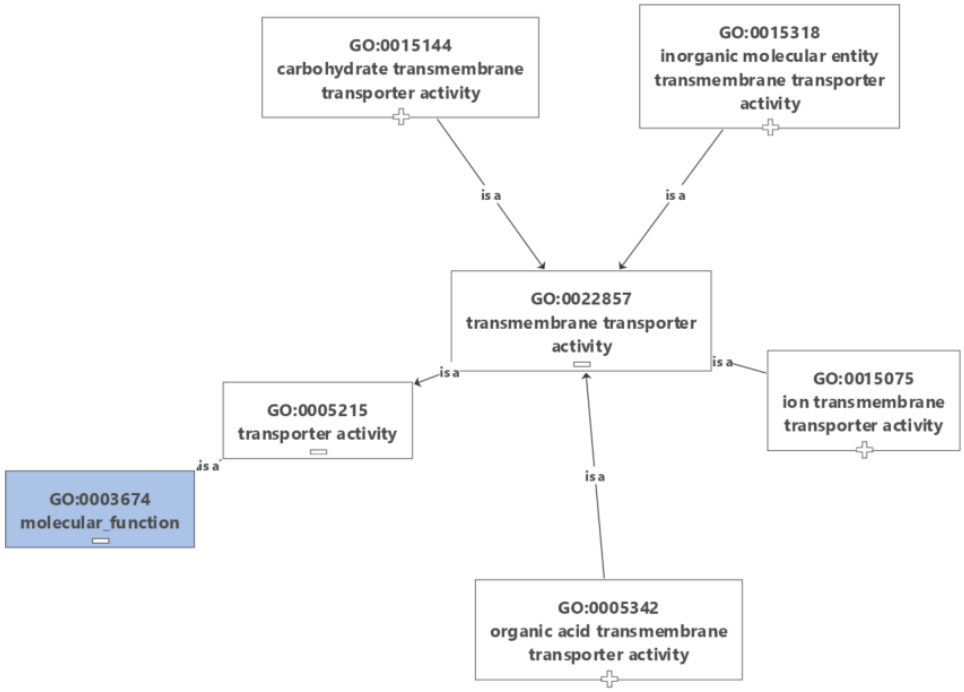



Figure S4. Enriched GO terms of genes with identified SNPs in the *w*One genome.