bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Title page**

2 **1. Title:**

3 **Understanding *Wolbachia* acquisition and co-divergence of hosts and their**

4 **associated bacteria: *Wolbachia* infection in the *Chorthippus parallelus* hybrid zone**

5 **2. Author's names with initials:**

6 Martinez-Rodriguez, P[1, 2]; Arroyo-Yebras, F[2]; Bella, J.L[2].

7 **3. Full postal adresses**

8 **1.** INRA, Univ. Nice Sophia Antipolis, CNRS, UMR 1355-7254 Institut Sophia

9 Agrobiotech, 06900 Sophia Antipolis, France

10 **2.** Departamento de Biología, Universidad Autónoma de Madrid, C/ Darwin, 2, 28049

11 Madrid, Spain.

12 **4. Keywords:** *Wolbachia,* Codivergence, *Chorthippus parallelus,* hybrid zones.

13 **5. Name, address, fax number and electronic mail address of corresponding author:**

14 **Name** Martinez-Rodriguez, Paloma

15 **Address** BPI - Biologie des Populations Introduites.

16 INRA, Univ. Nice Sophia Antipolis, CNRS, UMR 1355-7254 Institut Sophia

17 Agrobiotech, 06900 Sophia Antipolis, France

18 **Electronic mail address:** pamartinez@paca.inra.fr

19 **5. Running title: co-divergence of *C. parallelus* and *Wolbachia***

20

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Wolbachia is one of the best known bacterial endosymbionts affecting insects and nematodes. It is estimated that it infects 40% of insect species, so epidemiologically it may be considered a pandemic species. However, the mechanisms by which it is acquired from other species (horizontal transmission) or by which it coevolves with its hosts as a result of vertical transmission across generations are not known in detail. In fact, there are few systems in which the codivergence between host and bacterium has been described.

This study goes in deep in the *Wolbachia* infection in the grasshopper *Chorthippus parallelus.* This well-known system allows us to investigate the mechanism of acquisition of various *Wolbachia* strains in a new host, and the bacterial genomic changes during bacterial-host codivergence: We describe the genetic diversity of *Wolbachia* strains infecting both subspecies of *C. parallelus* and analyse their phylogenetic relationship. We also show the emergence of new bacterial alleles resulting from recombination events in *Wolbachia* infecting hybrid hosts. Our data suggest that F strains detected in this grasshopper have co-diverged with its host, *versus* a more recent horizontal transmission of B strains. According with this, we discuss the potential role of *Wolbachia* in the dynamics of the grasshopper hybrid zone and in the divergence of the two grasshopper subspecies since the origin of their hybrid zone.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Wolbachia is one of the most widely distributed endosymbiotic bacteria, infecting about 40% of insect species. At least, 8 bacterial supergroups have been described (Zug & Hammerstein 2012, but see Gerth et al. 2014). Vertical (from females to offspring) and horizontal (across species) transmission are the two main mechanisms to explain *Wolbachia* expansion. However, the way in which the two main modes of transmission have combined during the evolutionary history of *Wolbachia* and its hosts it is not well understood (Kremer & Huigens 2011; Werren et al., 2008). On the one hand, horizontal transmission has been proposed as an essential mechanism to explain the current distribution of *Wolbachia* across species. Actually, horizontal transmission and infection loss could explain the observed phylogenetic incongruence between *Wolbachia* and its hosts or the appearance of the same *Wolbachia* strain in distantly related host species (Baudry *et al.* 2003; Keller *et al.* 2004; Martins *et al.* 2012; Raychoudhury *et al.* 2009; Shoemaker *et al.* 2003; Yun *et al.* 2011). On the other hand, vertical transmission is the predominant mode of transmission (Moran *et al.* 2008; Saridaki & Bourtzis 2010). Due to that, coevolution between *Wolbachia* and their host should be common (but it has been rarely described) (Raychoudhury *et al.* 2009; but see Bordenstein *et al.* 2009 and Gerth *et al.* 2014). Here, the *C. parallelus* hybrid zone was used to investigate this infrequently reported process due to the knowledge about the evolutionary history of this species.

The hybrid zone formed by the meadow grasshopper *Chorthippus parallelus* is considered an example of secondary contact after allopatric differentiation (Bella *et al.* 2007; Hewitt 1993; Shuker *et al.* 2005). After the last ice age, *C. p. parallelus* and *C. p. erythropus* met at the geographical barrier of the Pyrenees, giving rise to the hybrid zone that exists to this day. Currently, the hybrid zone along the valleys of Tena (Spain) and d'Ossau (France) extends over more than 40 km, where a gradient of phenotypic and genotypic characters have been

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

63 found between pure populations, located at the ends of the hybrid zone (Hewitt 2001; Hewitt
64 2011; Shucker *et al.* 2005).

65 B and F *Wolbachia* supergroups infect *C. parallelus* (Dillon et al., 2008 ; Martínez *et al.*
66 2009; Zabal-Aguirre *et al.* 2010). Previous data provide evidence of different patterns of
67 infection and coinfection by the two bacterial supergroups in pure and hybrid populations
68 throughout the Iberian Peninsula, the Pyrenees and the rest of Europe, based on infection
69 frequencies (Bella *et al.* 2010; Martinez-Rodriguez 2013; Zabal-Aguirre *et al.* 2010). It is
70 noteworthy that these bacterial biogeographical patterns clearly delineate the current
71 distribution of pure and hybrid grasshoppers.

72 Experimental crosses in the field with pure and hybrid individuals of *C. parallelus* show
73 that *Wolbachia* causes cytoplasmic incompatibility in crosses between infected and uninfected
74 individuals (unidirectional incompatibility) and in those between individuals infected with
75 different bacterial lineages (bidirectional cytoplasmic incompatibilities), as indicated by the
76 significant reduction in the number of offspring of the affected crosses. *Wolbachia* also
77 increases the fecundity of infected females (Zabal-Aguirre *et al.* 2014). In addition, the
78 bacterium induces certain cytogenetic effects in this grasshopper, this affecting the proportion
79 of abnormal spermatids and the chiasmata frequency (Sarasa *et al.* 2013). The existence of CI
80 and other mentioned effects suggest that *Wolbachia* infection could influence the dynamics of
81 the *Chorthippus* hybrid zone, reinforcing the reproductive barrier between them. Actually,
82 several theoretical studies support this fact: For example, recently Telschow *et al*. (2014)
83 report that nuclear incompatibilities (according with Dobzhansky Muller model) and
84 cytoplasmic incompatibilities could act synergistically in order to keep the existence of genetic
85 diversity after secondary contact. However, more studies are required to understand the
86 underlying processes in this particular case.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In this study and based on the multilocus system typing (MLST system) proposed by Baldo *et al.* (2006b), we (1) analyse the phylogenetic relationship between *Wolbachia* strains infecting host populations and (2) the current distribution of *Wolbachia* infection in pure and hybrid populations of this grasshopper inside and outside its hybrid zone, including populations outside the Iberian peninsula. This also serves to propose the possible influence of ancestral F *Wolbachia* in the very origin of this hybrid zone. Besides we (3) describe the greater genetic variability in *Wolbachia* strains infecting grasshopper hybrid populations *vs*. pure subspecies populations, which suggests close endosymbiont/host-genotype interactions and provides evidence of coupled evolution between both genomes. Finally, we infer (4) the modes of acquisition of *Wolbachia* in *C. parallelus* and describe (5) how the combination of vertical and horizontal modes of transmission explains current patterns of *Wolbachia* infection in *C. parallelus* and its consequences for the evolutionary history of the host.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Material and methods

Field collections

Wolbachia infection was analyzed in more than 1780 *Chorthippus parallelus* individuals collected from 21 European locations inside and outside of the hybrid zone in 2008 and 2009, with the exception of Bubion and Epping Forest populations, captured in 2002 and 2004, in the context of a *Wolbachia* infection prevalence experiment in *Chorthippus* (see Martinez-Rodriguez, 2013). The populations are grouped as indicated in Table 1. Complete data collection are indicated in supplementary table 1. Gonads were dissected and fixed in 100% ethanol.

DNA extraction, Wolbachia detection and sequencing

DNA was extracted from whole fixed ovaries and testes, as described in Martínez-Rodríguez *et al.* 2013a and 2003b. *Wolbachia* was detected by PCR amplification of a *Wolbachia 16S rRNA* gene in all sampled individual, using *Wolbachia*-specific primers (Zabal-Aguirre *et al.* 2010), followed by a second, nested PCR amplification using strain-specific primers (Martínez-Rodríguez *et al.* 2013a and 2013b) (Table 2). PCR and Nested-PCR reactions were adjusted to 25 µl: 1X buffer, 2 mM of Mg2Cl, 0.2 mM dNTP, 1.2 µM each primer, 1.25 units of BioTaq DNA polymerase (Bioline) and 100 ng genomic DNA (for the first PCR) or 0,5 µl of previous PCR product in the nested PCR). The reaction was initiated with a cycle of 95° C 30s, followed by 35 cycles of 30s at 95° C, 1 min at 54°C (first PCR) or 69° C (nested-PCR), 1min 30s at 72° C and a final cycle of 10 min at 72° C. A total of 10 µl of each amplification product were electrophoretically separated on 1% agarose gels, which were stained with 0.5 mg/ml ethidium bromide and visualized under UV light (UVIdoc, Uvitec Cambridge).

We characterized the *Wolbachia* strains using the MLST and *wsp* (*Wolbachia* surface protein) gene characterisation systems (Baldo *et al.* 2006b). The *gatB*, *coxA*, *hcpA*, *ftsZ*, *fbpA*

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and *wsp* genes were amplified in 127 selected singly infected individuals (individuals infected exclusively by F or B supergroup, according with the previous *16S rRNA* gene test, supplementary table 1) while multiple-infected individuals were discarded to avoid ambiguous chromatogram lectures and to reduce the experimental workload. Infection frequencies according with 16S gene test in the different populations are described in Martinez-Rodriguez, 2013 and supplemental table 1). These genes were amplified using previously described methods (Baldo *et al.* 2006b) with slight modifications: PCR reactions were performed in 50 µl volumes containing 2 mM of $MgCl_2$, 0.2 mM of dNTP, 30 pmoles of each primer, 1.25 U of *Taq* BIOTAQ™ DNA polymerase (Bioline) and 2 µl of DNA solution (50 ng/µl). The reaction was initiated with a cycle of 95° C 30s, followed by 35 cycles of 30s at 95° C, 1 min at 54°C (*hcpa*, *gatb*, *ftsz* and *coxa* genes) or 59° C (*wsp* and *fbpa*), 1min 30s at 72° C and a final cycle of 10 min at 72° C. A total of 10 µl of each amplification product were electrophoretically separated on 2% agarose gels, which were stained and visualized as described above. Amplified genes were purified by ExoSAP-IT (GE Healthcare) and Sanger automatically sequenced (by Stabvida, Portugal). The MLST and *wsp* sequences generated in this study have been deposited in the GenBank database under accession numbers KM078849-KM078883 (see supplementary table 2).

Sequences analyses

Further studies in this grasshopper confirm *Wolbachia* integrations in the host genome. Current genomic data confirm the absence of integrated sequences of *ftsZ*, *fbpA* and *wsp* genes. However, some incomplete reads that mapped to *coxA*, *gatB* and *hcpA* have been detected in uninfected individuals in low coverage (Funkhouser-Jones *et al.*, 2015). This forces to be cautious before confirming that the sequences obtained by PCR belong to infecting bacteria and an accurate protocol was developed to ensure this, as indicated below.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

147 Firstly, DNA was extracted from gonad tissue in order to increase the living
148 bacteria/nuclear insertions ratio. Previous studies confirm the massive presence of *Wolbachia*
149 in the grasshopper gonads of infected individuals (Martinez *et al.*, 2009). This reduces the
150 probability of Sanger sequencing *Wolbachia* insertions. Secondly, to distinguish the sequences
151 belonging to infecting *Wolbachia* and those sequences integrated into the host nucleus all
152 sequences were compared with the standard sequences to detect possible rearrangements
153 and also translated into protein, in order to detect frameshift mutations, stop codons, and
154 indels. Previous studies in a different grasshopper, *Podisma pedestris*, confirm that most
155 *Wolbachia* insertions show these types of mutations, due to the absence of evolutionary
156 constraints after integration (non-translated sequences) (Martinez-Rodriguez *et al.*
157 unpublished data). Although this reduces the probability of considering an integrated
158 sequence as belonging to "living *Wolbachia*", we are reminded that it cannot totally discard
159 this possibility. We are taking this in mind when describing and discussing our results, mainly
160 those regarding recombinant and new alleles (see below).

161 ### *Phylogenetics analysis*

162 Bayesian likelihood was inferred using a Markov Chain-Monte Carlo variant run in the
163 MrBayes 3.2.1 program (Ronquist & Huelsenbeck 2003). Phylogenies based on single and
164 concatenated MLST genes and *wsp* were reconstructed. JModeltest (Posada 2008) was used to
165 distinguish the appropriate model of evolution, the best likelihood score being chosen on the
166 basis of the AIC criteria (Akaike 1974). The selected models were GTR+I+G for concatenated
167 MLST, *ftsZ* and *gatB*; GTR+G (general time-reversible model, including gamma correction) for
168 *coxA*, *hcpA* and *wsp*; and HKY+I+G (the Hasegawa, Kishino and Yano model (Hasegawa *et al.*,
169 1985), including gamma and proportion invariant corrections) for the *16S rRNA* gene. Bayesian
170 analysis was carried out for 10^6 generations with a sample frequency of 100. The first 25% of
171 trees were considered as burn-in and thus discarded. For each locus, the level of nucleotide
172 diversity per site and the number of variable sites or Ka/Ks were estimated using DnaSP

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

software (Librado & Rozas 2009). Alignments of individual and concatenated genes with and without outgroups were screened for significant levels of recombination using RDP4 v4.16 (Martin *et al.* 2010). The analysis involved several tests including GENECONV (Padidam *et al.* 1999), MAXCHI (Maynard Smith 1992) and Chimaera (Posada & Crandall 2001). A Bonferroni correction was applied and significance was concluded for values of $p < 0.01$.

Strain characterisation

Following the MLST system (Baldo *et al.* 2006b; Maiden *et al.* 1998), we defined a *Wolbachia* strain or sequence type (ST) as being different on the basis of its unique combination of five alleles. Furthermore, strains sharing at least three alleles were considered to belong to an ST complex, a group of evolutionarily related haplotypes. This analysis was carried out using START2 (Jolley *et al.* 2001). The *wsp* system was employed as a complementary approach for strain characterisation (see Baldo *et al.* 2005, Baldo *et al.* 2006a). Alleles that were detected only once were excluded in the analysis to avoid the miss interpretation of the PCR-associated sequencing errors.

Inference of bacterial microevolution using multilocus sequence data

We inferred *Wolbachia* microevolution using ClonalFrame to identify the clonal relationships between strains, and to estimate recombination events that have disrupted the clonal inheritance (Didelot & Falush 2007). We performed five separate runs, executing 250,000 MCMC iterations for each, discarding the first 100,000 iterations as burn-in.

Biogeographical analysis

An AMOVA based on the ST frequencies detected in each population was carried out based on the estimated supergroup frequencies (some data here used from Bella *et al.* 2010 and Zabal-Aguirre *et al.* 2010) and the genetic distance between haplotypes (calculated as the Tamura–Nei distance). Locus-by-locus AMOVA and an exact test of population differentiation

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were also carried out. In addition, we tested the correlation between genetic and geographical distances with Mantel tests. Geographical distance was estimated using Geographical Distance Matrix Generator v.1.2.3 (http://biodiversityinformatics.amnh.org). All analyses were done using Arlequin 3.11 (Excoffier *et al.* 2005).

Results

1. Wolbachia diversity in C. parallelus

1.1. How many *Wolbachia strains infect C. parallelus*?

To characterize the *Wolbachia* diversity across the hybrid zone, *16S rRNA*, MLST and *wsp* genes of *Wolbachia* were sequenced from host grasshopper individuals collected in several populations, inside and outside the hybrid zone.

The reanalysed phylogenetic tree based on *Wolbachia 16S rRNA* gene sequences confirmed that *C. parallelus* are infected by at least 4 strains belonging to the F supergroup and 2 B supergroup's strains (Bella *et al.* 2010; Martínez-Rodríguez *et al.* 2013a; Zabal-Aguirre *et al.* 2010) (see supplementary Fig. 1).

In addition, we studied *Wolbachia* supergroups and strains infecting *C. parallelus*, on the basis of the five genes involved in the MLST system, and on the *wsp* gene (Baldo *et al.*, 2006b). The analysis of the sequences of the 5 MLST genes distinguish 33 different haplotypes or ST (sequence types, according with Baldo *et al.* 2006b) based on the combination of 5 loci alleles: We detected 5 different alleles of ftsZ gene, 5 alleles of gatB gene, 6 alleles of coxA gene, 5 alleles of fbpA gene, and 10 alleles of hcpA gene (see Fig. 1, 2 and supplemental Figs. S2 to S6). Nucleotide diversity and other characteristics are summarised in supplemental Table 3. The patterns of ST distribution across geographical areas will be describe after (Fig. 3 and S7-S12).

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1.2. *Bacterial recombination*

Recombination was detected in 19 STs or haplotypes. Recombinant *Wolbachia* strains were detected by two methods. Firstly, RDP4 analysis detected recombinant strains under several tests (marked "R" in Fig. 4, in contrast with parental strains, which are indicated as "F" or "B", depending on the supergroup). In addition, the appearance of alleles of the B supergroup in isolates of the F supergroup (based on most of the genetic markers), and *vice versa*, also indicates recombination events. Our analysis revealed that both supergroups have exchanged parts of their genomes in some populations of *C. parallelus*, such as those of Portalet or Tourmont, in the centre of the hybrid zone, while recombination has not been detected in the grasshopper's pure populations within or outside the hybrid zone. Some recombinants have also been detected in the north of Spain, in populations of this grasshopper characterised as hybrid on the basis of chromosomal markers (Bella *et al.* 2007).

1.3. Wolbachia phylogeny

After discarding recombinants STs to avoid artefacts, the phylogenetic tree from concatenated sequences allows distinguish 10 different strains of *Wolbachia* belonging to B supergroup and 4 different strains to F supergroup (Fig. 2).

In general, *Wolbachia* strains in *C. parallelus* are highly related. On the one hand, F strains are highly related between them. Other F strains, like *Wolbachia* infecting *Opistophthalmus granifrons (Scorpionida)* or *Cimex lectularius (Hemiptera)* are more distant. Their host is not related with *Chorthippus* (ecologically or phylogenetically). B strains are also related between them, but also with the *Wolbachia* strains infecting other Orthopteran, like *Teleogryllus taiwanemma*, and the recently detected *Wolbachia* strain infecting *Podisma pedestris*. This latest species shares habitat with *C. parallelus* (data not shown). Also, B strains (based on *16S rRNA* gene amplification) have been recently detected in other species,

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

including *Ruspolia nitidula*, *Chorthippus vagans* and *Euhorthippus chopardi*, captured in the same populations that *Chorthippus parallelus* (Martinez-Rodriguez 2013).

Phylogenetic analyses of individual genes were also carried out. Alleles were correctly characterised as belonging to the F or B supergroups (Fig. 1 and Figs. S2-S6).

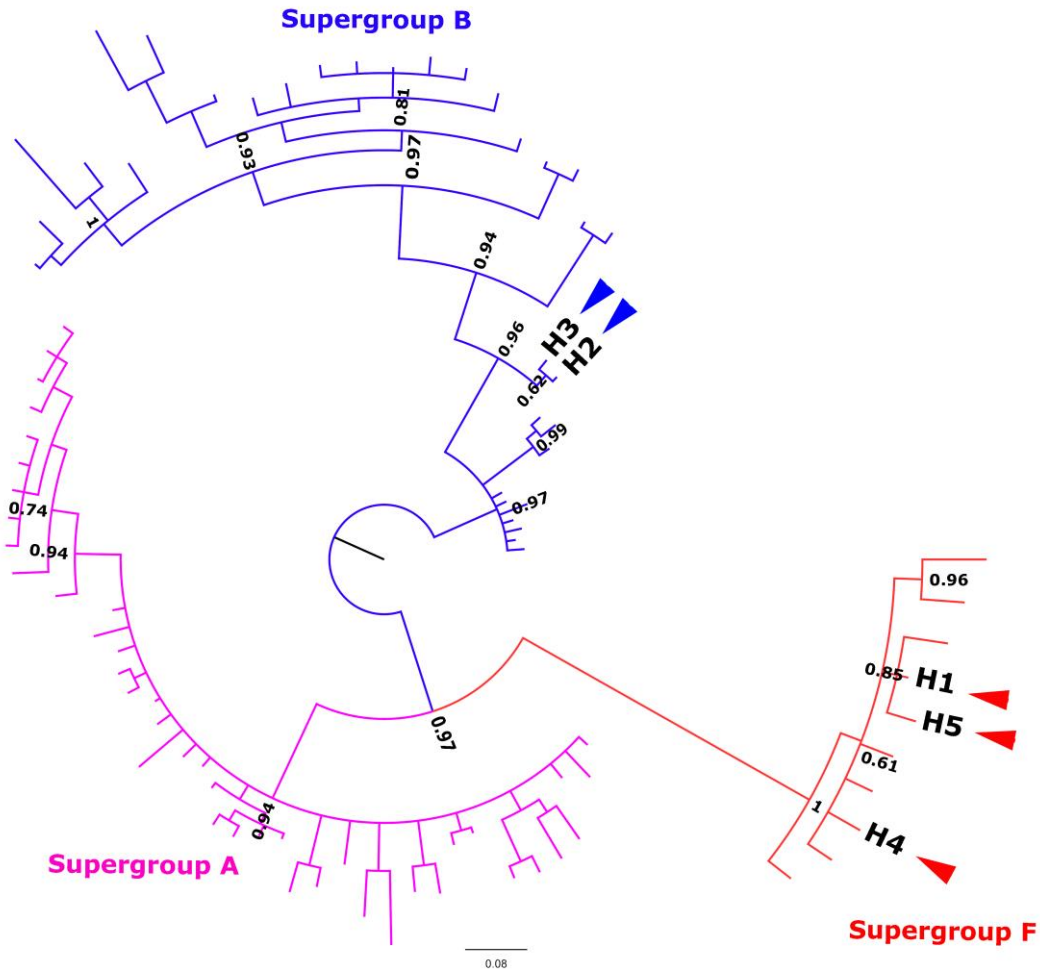


Figure 1: (Online colour figure) Summary unrooted phylogenetic tree of *fbpA alleles in several insects, including Chorthippus parallelus*, obtained by Bayesian inference. Alleles described in *C. parallelus* are named H1 to H5 *(marked as coloured arrows)*. Posterior probabilities are shown at the nodes. Other MLST genes are also analysed (Supplemental Fig. S2-S6). Sequence accession numbers are presented in Tables S13-S18.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

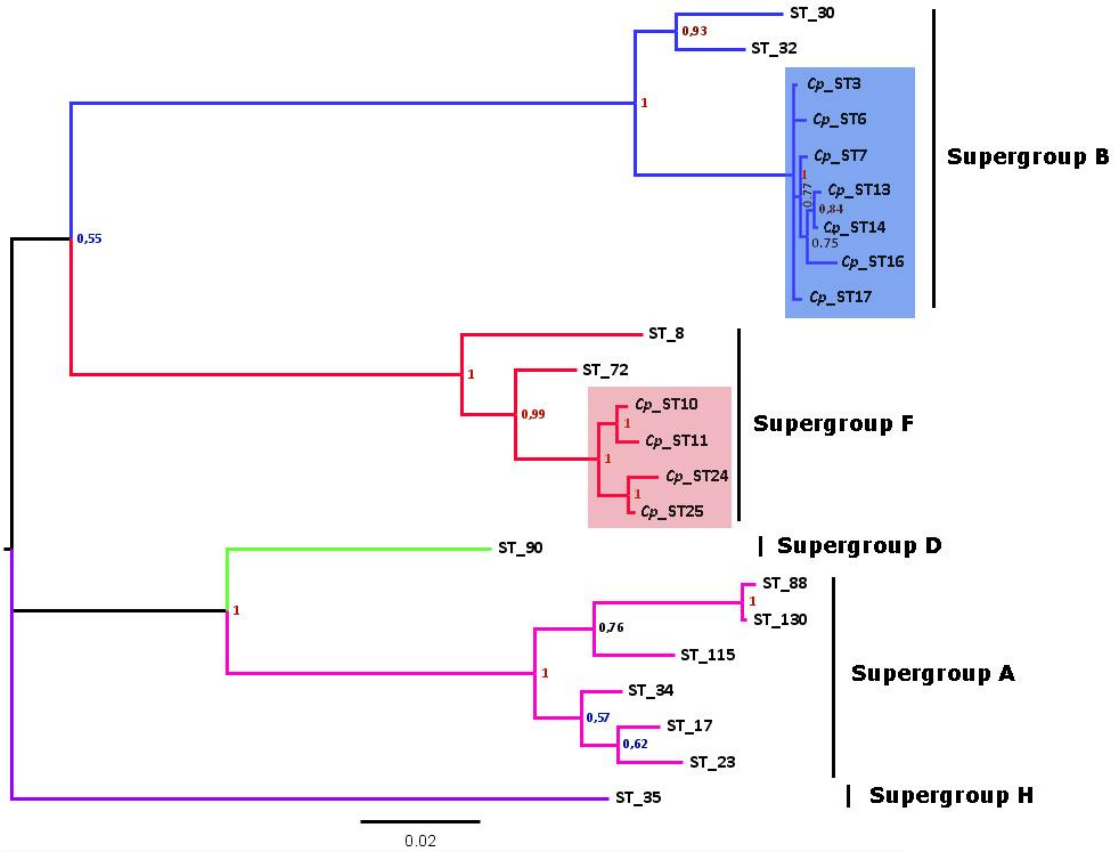


Figure 2: (Online colour figure) Phylogenetic tree of *Wolbachia* STs detected in *C. parallelus* (marked as Cp, coloured squares) excluding recombinants (see Fig. 4) obtained by Bayesian inference. The alleles described in this grasshopper bear the prefix Cp_ST. All other STs, named according to the official nomenclature, are available in the MLST database http://www.mlst.net/. Posterior probabilities are shown at the nodes.

2. *Biogeographical distribution of the Wolbachia strains:*

2.1. *Individual* loci *analyses*:

Different geographic isolates showed a high level of genetic variation within *Wolbachia* strains: Individual analysis of the 5 loci of MLST and the *wsp* gene allows us to determine a clear geographical pattern. In general, we detected alleles belonging to the B supergroup in *C. p. parallelus* and *C. p. erythropus* populations, indistinctly. By contrast, we can detect some alleles, belonging to F supergroup, specifically in some populations of *C. p. erythropus* or *C. p. parallelus* (see Fig. 3). In addition, we noted the presence of new alleles exclusive to *Wolbachia* infecting the hybrid grasshopper populations (Fig. 1, 3 and S2-S12). For instance and regarding gene *fbpA*, three alleles were identified belonging to supergroup F, and two alleles were

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

assigned to supergroup B (Fig. 1). In the first case, allele 5 has been described in European populations of *C. parallelus* and Bubion in southern Spain and in the populations of the Cantabrian region (hybrid populations, *sensu* Bella *et al.* 2007). It has also been identified in the pure population of Gabas, on the French side of the hybrid zone (ZH France, Fig. 3). Allele 1, which also belongs to supergroup F, was detected in pure populations from the centre of the Iberian Peninsula and in the South Pyrenees populations of Escarrilla, Sallent and Portalet (hybrid zone). In the case of supergroup B, allele 2 was found in most of the populations. However, in the hybrid populations of Sallent, Corral de Mulas and Portalet (hybrid zone) we also detected alleles 3 and 4. Allele 4 has also been detected in Cantabrian hybrid populations. Similar patterns have been observed for the rest of the analyzed genes.

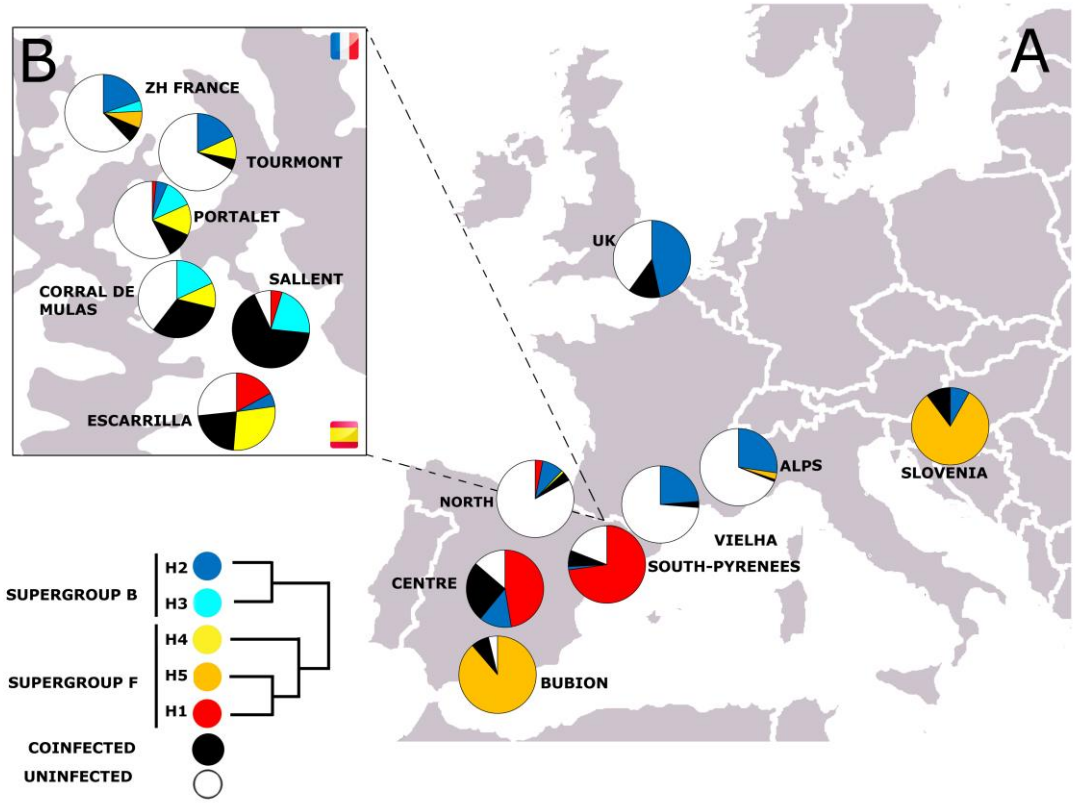


Figure 3. (Online colour figure) A) Geographical distribution of *fbpA* alleles in the *C. parallelus* populations analysed. Pyrenean hybrid zone (Tena's valley, Huesca, Spain) is zoomed in B. See Table 1 for details.

2.1.1. *Wolbachia* ST- complexes:

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

According with the MLST system implemented in START2 (Jolley *et al.* 2001), we classify the different haplotypes or ST in five ST-complexes (each one defined as a group of STs sharing a minimum of three alleles) (Fig. 4). The first ST-complex included several STs, some of them belonging to the F supergroup or recombinants highly related to this supergroup (see before, *Wolbachia* phylogeny), sampled in several non-Iberian populations from the rest of Europe and in some samples from the Basque Country in the north of Spain. The second ST complex, included isolates belonging to the B supergroup (and some recombinants, highly related to the B supergroup) widely distributed in *C. parallelus* populations, including both subespecies, but in different proportions. ST3 and ST4 complexes also include some recombinants and B strains. Finally, ST5 complex belongs to F supergroup but it shares some alleles with those of the B supergroup. They have been detected in *C. p. erythropus*, including some populations of the centre of Spain, and the Spanish region of the Pyrenees. The geographical distribution of each ST is showed in Fig. 4.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ST-complex	ST	gatB	coxA	hcpA	ftsz	fbpA	Type	n	population
ST1	2	1	2	3	2	5	R	1	ZH France
ST1	9	2	1	1	1	2	R	1	ZH France
ST1	10	2	1	1	2	1	F	2	ZH France
ST1	11	2	1	3	2	5	F	13	Bubion(3), Alps(3) and Slovenia (7)
ST1	12	2	2	3	2	5	R	11	ZH France
ST2	1	1	1	5	1	5	R	1	South Pyrenees
ST2	3	1	2	5	1	2	B	4	Escarrilla(1), Centre(3)
ST2	5	1	5	3	1	2	R	1	Vielha
ST2	6	1	5	5	1	2	B	10	Vielha (3), Slovenia(1), Alps (3), ZH France(3), Ingland (3)
ST2	7	1	5	5	1	3	B	1	ZH France
ST2	8	1	5	5	2	2	R	1	South Pyrenees
ST2	17	3	5	5	1	2	B	4	North
ST2	26	4	5	4	1	2	R	1	Escarrilla
ST3	13	3	2	2	1	3	B	2	Sallent
ST3	14	3	2	4	1	3	B	3	Sallent(2), CM(1)
ST3	16	3	2	7	1	3	B	2	CM
ST3	30	5	2	2	1	2	R	1	Portalet
ST3	32	5	6	7	1	2	B	1	Tourmont
ST4	22	4	3	9	3	3	R	1	Portalet
ST4	27	4	6	6	3	3	R	1	Portalet
ST4	29	4	6	7	3	3	R	1	Portalet
ST4	31	5	3	7	3	3	B	1	Portalet
ST4	33	5	6	7	3	3	B	1	Portalet
ST5	4	1	4	8	4	1	R	1	Escarrilla
ST5	15	3	2	6	5	3	R	1	CM
ST5	18	4	2	6	5	3	R	1	CM
ST5	19	4	2	6	5	4	R	3	CM(2), Portalet(1)
ST5	20	4	2	9	5	4	R	1	Tourmont
ST5	21	4	3	9	2	4	R	3	Portalet(2), Tourmont(1)
ST5	23	4	4	6	4	4	R	3	Escarrilla(2), CM(1)
ST5	24	4	4	6	5	4	F	4	Escarrilla(3), CM(1)
ST5	25	4	4	8	5	1	F	12	Centre(6), Pyrenees(5), Sallent(1)
ST5	28	4	6	6	5		R	5	Portalet

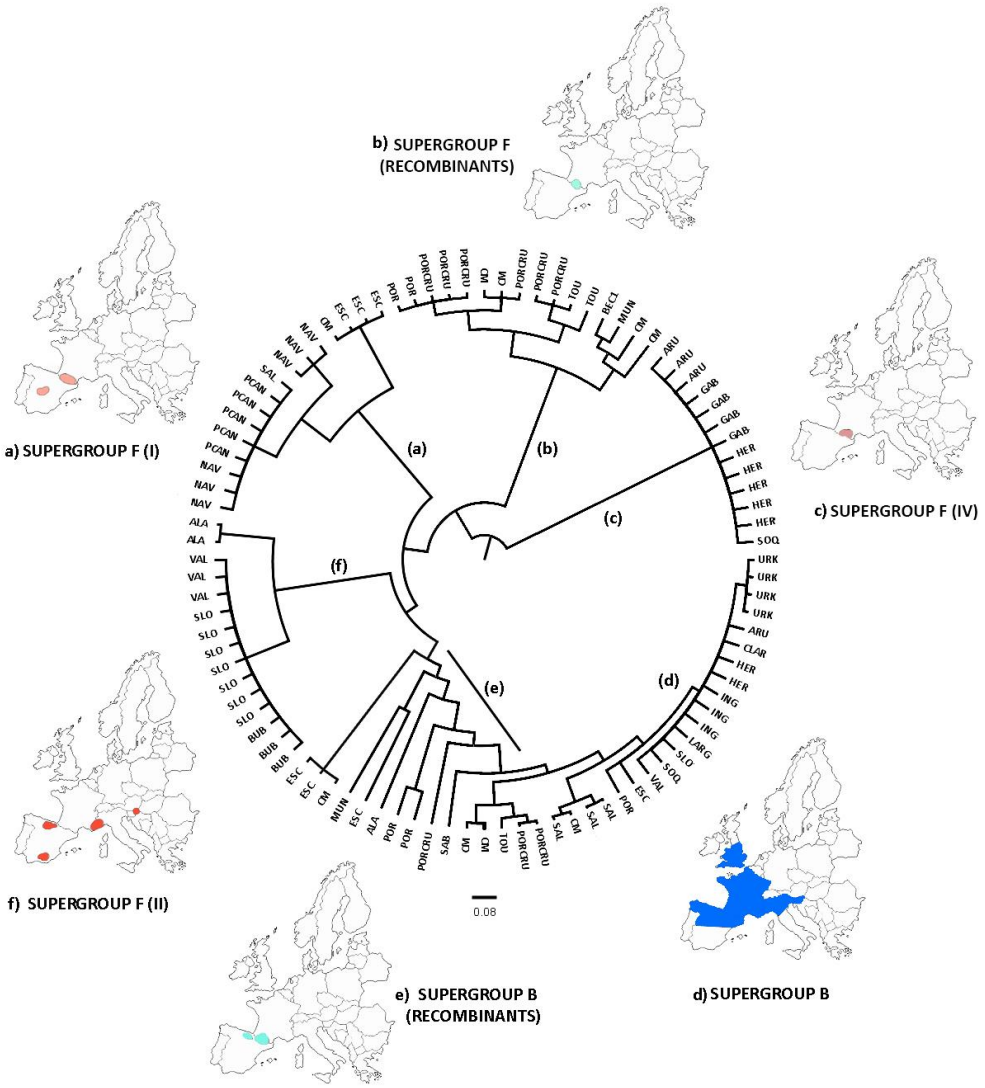
Figure 4: (Online colour figure) *Wolbachia* ST-complexes and allelic profiles described in *C. parallelus*. Note the classification in three groups: those assigned to supergroups F and B strains ("F" and "B", respectively) and those in which possible recombination events between these supergroups were observed ("R"). Alleles belonging to F supergroup (see Fig. 1 and S2-S6) are marked with different red tones, while alleles belonging to the B supergroup are marked with blue tones. STs detected in only one individual (blue) should be interpreted with caution, even if the alleles appear in more than one sample. The name of population and number of individual (parenthesis) detected in each population are also indicated.

The ClonalFrame-based analysis infers *Wolbachia* microevolution using the multilocus sequence data and considers recombination. The genealogies confirmed the genetic subdivisions in the strains of the F supergroup (Fig. 5), while B strains were grouped in the same clade. The genealogies also detected the recombinant strains that mostly appear in the

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

310 grasshopper hybrid zone. The clades also support an association between genetic and

311 geographic data.



312

313 Figure 5: ClonalFrame genealogy (Online colour figure). Maps indicate the approximate location of the
314 samples assigned to the major clades, classified with respect to their corresponding F or B supergroup. The analysis
315 distinguished three major clades of supergroup F (a, c and f), one clade belonging to supergroup B (d), and several
316 recombinant strains (b and e). This was consistent with our previous analyses. Acronyms are listed in Table 1.

317

318 The lower differentiation of isolates within geographic populations and the higher

319 differentiation of those between geographic populations suggest isolation-by-distance

320 between the bacterial F strains infecting the two grasshopper subspecies (with the exception

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of Bubion; see Discussion). The AMOVA indicated a geographic division of the F supergroup: (i) Central Iberian Peninsula and South Pyrenees populations, (ii) Pyrenean hybrid zone populations, (iii) French side of the hybrid zone, and (iv) Non-Iberian populations from the rest of Europe and Bubion (in Spain) (Table 3).

These results were also supported by the locus-by-locus AMOVA (except for *hcpA*) (Table S4) and the exact test of population differentiation (Rousset *et al.* 1992) (Table S5).

In addition, the Mantel tests confirmed that the genetic and geographic distances were correlated (rY1: 0.338, p=0.001). This correlation was stronger when the Bubion data were excluded (rY1: 0.483, p=0.003). This particular geographical distribution could be related to the biogeographical distribution of this grasshopper during the last glaciation and allows us to infer the origin of *Wolbachia* infection in *C. parallelus* and its role in establishing the hybrid zone.

3. Estimation of Wolbachia divergence dates

Previous studies suggest a synonymous divergence rate of about 0.90% per million years (MY) for bacteria. However, this bacterial molecular clock should be interpreted with caution since divergence rates may differ between bacteria species (Ochman *et al.* 1999; Ochman & Wilson 1987; Raychoudhury *et al.* 2009). Based on this estimate, the divergence between the F strains detected in the centre of Spain (Cp_ST-25) and Slovenia (Cp_ST-11) is about 3,400,000 years. On the other hand, the divergence between the F strains detected in the centre of Spain (Cp_ST-25) and the hybrid zone (Cp_ST-24) is about 1,400,000 years. B strains detected in the centre of Spain (Cp_ST-3) and the widely distributed Cp_ST-6 diverged about 250,000 years ago. The dates of divergence of strains based on the different markers are illustrated in supplementary Tables S6-S12. Substitution rate could represent a lower boundary for the mutation rate within strains (Emerson 2007). Thus, other estimation of intraspecific mutation rate (in terms of *D. melanogaster* generations (6.87E-10 per position per insect

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

346 generation in the 3rd position, see Richardson *et al.* 2012) have been used. Divergence

347 between strains could be higher (x 10) if we consider this estimation of *Wolbachia*

348 evolutionary rates.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Modes of acquisition of Wolbachia. Codivergence vs. *horizontal transmission.*

Three hypotheses about the origin of *Wolbachia* in *Chorthipus parallelus* are discussed: Firstly, an ancient co-divergence between *Wolbachia* and this orthopteroid. Secondly, the acquisition of *Wolbachia* before the subspecies divergence, and the recent co-divergence of *Wolbachia* and the host. And thirdly, the recent acquisition of *Wolbachia* by horizontal transmission.

Wolbachia codivergence with their host is extremely rare in the literature compared with horizontal acquisition between species (Raychoudhury *et al.* 2009). To distinguish between co-divergence and horizontal transmission events, a good knowledge of the recent evolutionary history of the host is required as it happens with the *C. parallelus* system which becomes a good model to study *Wolbachia* expansion. Our *Wolbachia* phylogenetic and phylogeographic data can also be interpreted in the context of its host evolution so serving to infer the *Wolbachia* transmission and evolution in this particular grasshopper and its influence in the hybrid zone.

Not discarding other mechanisms that could also be involved (some paternal transmission, infection loss, drive, etc.), our data point out two possible mechanisms to explain current *Wolbachia* infection in both *Chorthippus parallelus* subspecies: the codivergence of *Wolbachia* F strains during recent speciation of both subspecies followed by "modern" horizontal transmission of B strains from other organisms.

a) Phylogenetic relationships & common biogeography between host and bacteria.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

372 Phylogenetic data support that bacterial F strains infecting *Chorthippus* are extremely
373 similar among them, and that they are highly related each other than with any other outside
374 this grasshopper's taxa (see Fig. 1, 2 and supplemental figures S2-S6). These data support the
375 recent co-divergence between host and bacteria. Furthermore, the geographical distribution
376 of two main F bacterial lineages is largely congruent with the biogeography of *C. parallelus*
377 (Lunt et al; 1998). Cp25 and Cp24 linages infect *C. p. erythropus*, while Cp11 infects mainly *C. p.*
378 *parallelus* (except the Iberian southern population of *C. p. erythropus* of Bubion). Hybrid
379 grasshoppers are infected by variants of both lineages. Strains geographical distribution
380 supports the co-divergence between the two subspecies and the two main F strains infecting
381 *C. parallelus*.

382 However, the co-divergence between *Wolbachia* and their host should be "recent". F
383 supergroup (based on *16S rRNA* and *Ftsz* genes) has been detected in the bush cricket species
384 *Orocharis saltator* and *Hapithus agitator* (Gryllidae: Eneopterinae) but no in other Acrididae
385 (both families have diverged 300 Ma ago; see Song *et al*. 2015). In addition, both F *Wolbachia*
386 infecting *Chorthippus* are closer to strains infecting other insect orders than to this Gryllidae
387 one. This suggests that the F strain of *Wolbachia* was acquired by horizontal transmission
388 before the divergence between subspecies, followed by co-divergence between each host and
389 bacteria in their corresponding glacial refugia and during postglacial expansion.

390 By contrast, B supergroup strains infect homogeneously both subspecies, without a
391 biogeographical pattern. The variability within B supergroup is restricted to the hybrid zone, in
392 which new variants and alleles, highly related, appear. All data suggest a recent and quick
393 horizontal transmission of B strain to this host.

394 **b) Divergence time estimation:**

395 Our current data serve to estimate the divergence time of *Wolbachia* according with a
396 general bacterial molecular clock (Ochman *et al.* 1999; Ochman & Wilson 1987; Raychoudhury

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

397 *et al.* 2009). This supports that the divergent time of *Wolbachia* F strains is higher (3.4-1.4 Myr)
398 than *C. parallelus* subspecies divergence time (500.000 years according with mtDNA data,
399 Hewitt, 1996). However, this estimation could represent a lower boundary for the mutation
400 rate within species (Emerson 2007). Furthermore, we have also estimated this time of
401 *Wolbachia* divergence higher (x10) according with some specific *Wolbachia* mutation rates
402 noticed in *Drosophila* (Richardson *et al.*, 2012). However, several factors can lead to
403 inappropriate estimation of divergence dates. For instance, this estimation in *Drosophila* could
404 be inappropriate for *Chorthippus*, in which each host generation takes one year, which
405 modifies the dynamic of *Wolbachia* transmission, not discarding possible bottlenecks of
406 bacterial population, selection pressures, etc. Due to that, we think that the divergence times
407 are compatible, even when they are not coincident, and support an ancient acquisition of F
408 *Wolbachia*, followed by its co-divergence with their *Chorthippus* hosts.

409 By contrast, B *Wolbachia* divergence times are lower, and suggest a more "recent"
410 acquisition by *C. parallelus* by horizontal transmission. The existence of an extremely close B
411 strain of *Wolbachia* in a number of orthopteran species that share the same habitats also
412 support this hypothesis (Martinez-Rodriguez, 2013):

413 **c) Horizontal transmission from other taxa:**

414 We have detected extremely closely related B strains in other orthopteroids like
415 *Podisma pedestris*, *Chorthippus vagans* and *EuChorthippus chopardi* (Acrididae) but also
416 *Ruspolia nitidula* (Tettigoniidae) that share habitat with *Chorthippus* (data no shown, Martinez-
417 Rodriguez, 2013). Most genera belonging to family Acrididae diverged 50 Ma ago, and both
418 families (Acrididae and Tettigoniidae) did it 250 Ma ago (Song *et al.* 2015). The incongruence
419 between *Wolbachia* and host divergence times supports that the B supergroup could have
420 been "recently" acquired as a result of rapid expansion of the infection from other taxa

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(horizontal transmission). In this context, some species of parasitoids could be a vector for intra- or inter-specific infection transmission (unpublished data, Martinez-Rodriguez, 2013).

By contrast, a recent horizontal transmission of F strain is unlikely. There is no evidence of closely F *Wolbachia* strains currently infecting other Orthoptera. However, F strain (usual mutualist of nematodes, but also present in arthropods) could infect another insect in the past and explain the horizontal transmission of *Wolbachia* to an ancestral *C. parallelus* before subspecies divergence. Even if we cannot totally discard a recent horizontal transmission, we consider that the hypothesis of 2 independent "recent" acquisitions of 2 related F strains to this 2 geographically distant subspecies of *Chorthippus* is unlikely. In our opinion, the hypothesis of an "ancient" acquisition (>4 Myr) and consecutive co-divergence of *Wolbachia* F strains is more likely.

This hypothesis is also supported by the detection of an insertion of *Wolbachia* that coincides in homologous chromosomes of both Cpe and Cpp, while other inserts are subspecies-specific (Funkhouser-Jones *et al.* 2015, Toribio-Fernandez *et al*., *in. prep.*). This recent finding supports that an ancestral *Chorthippus* sp. was already infected by *Wolbachia*, before divergence of two subspecies.

Diversification of Wolbachia inside the hybrid zone:

The ST distribution suggests that there is a particular pattern of *Wolbachia* infection within the Pyrenean grasshopper hybrid zone and suggests that the particular interaction between host "hybrid genomes" and bacterial infection could happen. B and F supergroups are in contact in several populations of *Chorthippus parallelus*, to the point of coexisting in the same individuals (coinfection, see below). However, we only have detected these new, recombinant strains in hybrid populations (inside the Pyrenees hybrid zone but also in a hybrid population in northern Spain).

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

High recombination between *Wolbachia* strains has been reported several times (Foster *et al.* 2011; Jiggins 2002; Jiggins *et al.* 2001; Verne *et al*. 2007; Werren & Bartos 2001). In fact, recombination levels in *Wolbachia* seem to be higher than, for example, in *Neisseria meningitidis*, which is considered a bacterium with a great capacity for recombination (Jolley et al. 2005). However, different recombination rates between strains have been detected (Klasson *et al.* 2009). The recombination process serves the strains to vary and adapt rapidly, which is important for their interaction with the host. For instance, data show that mutualist strains, adapted to a particular host, have limited levels of recombination compared with other strains than potentially should adapt to a new host (Jiggins 2002; Werren & Bartos 2001). Actually, the preservation of a high number of genes of recombination guarantees genomic flexibility during recurrent host change (Darby *et al*., 2007; Hurst *et al*., 2002).

This hypothesis is pertinent to the case of *Wolbachia* strains that infect hybrid *C. parallelus*. The contact of bacteria with in a new host (the hybrid grasshoppers) could have resulted in a high bacterial recombination rate in order to adapt to this new host. It might explain why our analyses detect recombinant strains infecting grasshoppers just in hybrid populations, although the F and B supergroups are in contact in many other *C. parallelus* populations (Bella *et al*. 2010; Zabal-Aguirre *et al*. 2010; Zabal-Aguirre *et al*. 2014).

In addition, we also detected, specifically in the grasshopper hybrid zone, new bacterial alleles belonging to these recombinant strains, which have diverged from closely related B and F alleles found in other isolates. This suggests that sequences diverged rapidly after recombination. Possible explanations include that *Wolbachia* strains infecting grasshopper hybrids diverged separately of other strains (due to the isolation between hybrids and pure populations) or perhaps these strains support other evolutionary pressures (for instance, adaptation to other a new hybrid host or to the evolutionary processes involved in the hybrid zone). More studies will be needed in order to clarify this.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The origin and expansion of Wolbachia infection in C. parallelus and its effects on the dynamic of the hybrid zone:

We consider that recent codivergence is the best explanation for F *Wolbachia* strains appearance in both subespecies of *C. parallelus*, while B infection is better explained by modern horizontal transmission. According with that, we propose a possible scenario to explain the *C. parallelus Wolbachia*'s acquisition (Fig. 6).

The study of *Wolbachia* infecting *C. parallelus* divergence should be considered in the context of the last quaternary ice age in Europe and its consequences for *C. parallelus* distribution: During this glaciation the grasshopper subspecies diverged as a result of their geographic isolation in allopatry (Hewitt 1993, 1996, 1999, 2001, 2011; Serrano *et al.* 1996). After the retreat of the ice, grasshopper populations from the Iberian Peninsula colonised the Pyrenees, meeting *C. p. parallelus* coming from the Balkans, as suggested by Lunt *et al.* (1998).

The current data about *Wolbachia* infection suggest that an ancient F strain of *Wolbachia* and an ancestral host could codiverge during this period before meeting when the hybrid zone formation. In addition, a new B infection could have been acquired more recently, and expanded in the pure and hybrid populations afterwards. *Wolbachia* spread by horizontal transmission can be very effective, as previously suggested (Turelli & Hoffmann, 1991). In addition, the lack of B infection in some populations, like Bubion in Southern Spain, also points out a recent spread of infection from continental Europe (where it is massive): the isolation of these individuals and their geographical location has not permitted their infection yet. Loss of an ancestral B infection in this population seems less plausible to us, given the strain's aforementioned homogeneity and abundance.

Finally, after the hybrid zone formation, new strains would have arisen in the hybrid zone by recombination. We are reminded that the appearance of the F strains in the grasshopper populations of central and southern Spain could be explained by an alternative

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

route of colonization (from the East or the South), not discarding either other less parsimonious hypotheses.

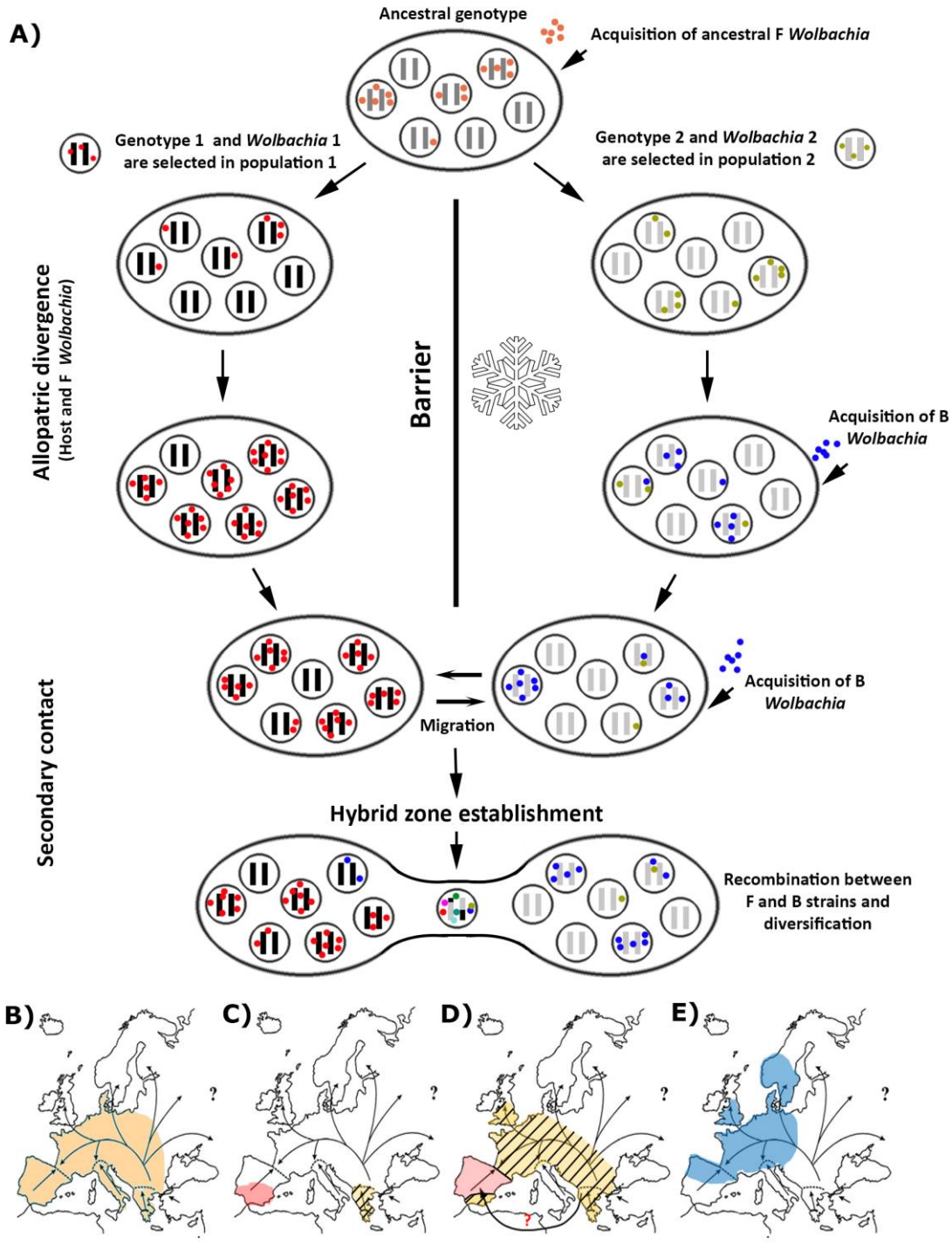


Figure 6: (Online colour figure). A) Proposed hypothesis for the origin of *Wolbachia* infection in *C. parallelus*. Each ellipse represents a population. Inner circles represent individuals. Black and grey bars indicate the host genome, while the coloured dots show the bacterial type infecting the individual. The hybrid zone would be established simultaneously with the appearance of recombinant genomes in the host, and a high bacterial diversity, induced by recombination. B) Spatial representation of the population expansion of infection: the arrows indicate the population expansion of *C. parallelus* (modified from Hewitt 2001), after the retreat of the glacial ice. Before the last

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

glaciation the infection of *Wolbachia* by the F supergroup was homogeneous. C) During the last glaciation, *C. parallelus* and F *Wolbachia* diverged in allopatry. D) After the ice disappeared, the pattern of expansion of the F infection coincided with that of the migration of its host, E) Recently, B infection has been transmitted horizontally in different European populations.

Wolbachia effects in the hybrid zone: New perspectives.

Our data suggest that *Wolbachia* already infected *C. parallelus* during the hybrid zone formation. Due to that, *Wolbachia*'s role in the hybrid zone dynamic deserves some discussion: We propose that genetic incompatibilities between the grasshopper subspecies accumulated during the divergence, together with the unidirectional and bidirectional CI that *Wolbachia* induces in the hybrid zone (Zabal-Aguirre *et al.*, 2014) thereby influencing the formation of the current grasshopper hybrid zone. More data are required to quantify the importance of CI in hybrid formation.

In the other hand at least two F strains of *Wolbachia* infect differently *C. parallelus* subspecies. New experiments should be carried out in order to verify if further CI exists, induced within those strains belonging to F supergroup. It is possible that these new bacterial lineages or STs, the result of processes of recombination between strains from supergroups F and B, and their subsequent diversification by point mutations and adaptation, would limit the incompatibility between grasshopper individuals infected by different supergroups in hybrid populations, favouring the appearance of a new, mixed bacterial and host genetic background in this area, in contrast to the pure populations on either side of the Pyrenees. This new scenario should be tested to know the current and actual role of *Wolbachia* in the *C. parallelus* hybrid zone and for a better study of this model of incipient speciation.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Acknowledgments

We are grateful to the authorities that gave us permission to collect the specimens analysed here: Parc National des Pyrenèes, Parque Nacional de Sierra Nevada, Comunidad Autónoma de Madrid, Comunidad Autónoma de Aragón, Comunidad Autónoma de Castilla y León, Parc National du Mercantour Alpes-Maritimes. Dr PL Mason (University of Glasgow, UK) helped us with the manuscript and Dr JR Dagley and Dr R Kostanjšek kindly provided the individuals from the Epping Forest (UK) and Mokronog (Slovenia), respectively. We also thank other members of our HYZO group in the Universidad Autónoma de Madrid, who helped with insect collection and handling, or contributed to discussions about our study. This work has been supported by the Spanish MINECO I+D+i grants CGL2009-08380/BOS and CGL2012-35007 and the collaboration of Chromacell.

Data accessibility

DNA sequences: GenBank database under accession numbers KM078849-KM078883

Author contribution:

Martinez-Rodriguez, P designed research, performed research, analyzed data and wrote the paper. Arroyo-Yebras, F performed research. Bella, JL designed research, analyzed data and wrote the paper. All of them, with the contribution in some cases of other members of the group and the collaborators cited in the acknowledgments section, collected the grasshoppers.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**, 716–772.

Baldo L, Bordenstein S, Wernegreen JJ, Werren JH (2006a) Widespread recombination throughout *Wolbachia* genomes. *Molecular Biology and Evolution* **23**, 437-449.

Baldo L, Hotopp JCD, Jolley KA*, et al.* (2006b) Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Applied and Environmental Microbiology* **72**, 7098-7110.

Baldo L, Lo N, Werren JH (2005) Mosaic nature of the *Wolbachia* surface protein. *Journal of Bacteriology* **187**, 5406-5418.

Baudry E, Bartos J, Emerson K, Whitworth T, Werren JH (2003) *Wolbachia* and genetic variability in the birdnest blowfly *Protocalliphora sialia*. *Molecular Ecology* **12**, 1843-1854.

Bella JL, Martínez-Rodríguez P, Arroyo-Yebras F*, et al.* (2010) *Wolbachia* Infection in the *Chorthippus parallelus* Hybrid Zone: Evidence for Its Role as a Reproductive barrier. *Journal of Orthoptera Research* **19**, 205-212.

Bella JL, Serrano L, Orellana J, Mason PL (2007) The origin of the *Chorthippus parallelus* hybrid zone: chromosomal evidence of multiple refugia for Iberian populations. *Journal of Evolutionary Biology* **20**, 568-576.

Bordenstein SR, Paraskevopoulos C, Dunning Hotopp JC, Sapountzis P, Lo N, Bandi C, Tettelin H, Werren JH, Bourtzis K. (2009) Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say. *Molecular Biology and Evolution* **26, 231-241.**

Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SG (2007) Intracellular pathogens go extreme: genome evolution in the Rickettsiales. *Trends in Genetics* **23**, 511–520.

Didelot X, Falush D (2007) Inference of Bacterial Microevolution Using Multilocus Sequence Data. *Genetics* **175**, 1251-1266.

Dillon RJ, Webster G, Weightman AJ*, et al.* (2008) Composition of Acridid gut bacterial communities as revealed by 16S rRNA gene analysis. *Journal of Invertebrate Pathology* **97**, 265-272.

Emerson BC (2007) Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Systematic Biology* **56**, 337–345

Excoffier L, Laval G, Schneider S (2005) ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.

Flanagan NS, Mason PL, Gosálvez J, Hewitt GM (2001) Chromosomal differentiation through an Alpine hybrid zone in the grasshopper *Chorthippus parallelus*. *Journal of Evolutionary Biology* **12**, 577 - 585.

Foster J, Slatko B, Bandi C, Kumar S (2011) Recombination in *Wolbachia* Endosymbionts of Filarial Nematodes? *Applied and Environmental Microbiology* **77**, 1921-1922.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

584 Funkhouser-Jones LJ, Sehnert SR, Martínez-Rodríguez P, Toribio-Fernández R, Pita M, Bella JL,
585 Bordenstein SR (2015) *Wolbachia* co-infection in a hybrid zone: discovery of horizontal
586 gene transfers from two *Wolbachia* supergroups into an animal genome. *PeerJ.* 2015;
587 3: e1479.

588 Gerth M, Gansauge M-T, Weigert A, Bleidorn C (2014) Phylogenomic analyses uncover origin
589 and spread of the *Wolbachia* pandemic. *Nature Communications* 5:5117 doi:
590 10.1038/ncomms6117

591 Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock
592 of mitochondrial DNA. *Journal of Molecular Evolution*. **22**, 160-74.

593 Hewitt GM (1993) After the ice - parallelus meets erythropus in the Pyrenees. In: *Hybrid Zones*
594 *and the Evolutionary Process* (ed. Harrison RG), pp. 140-164. Oxford University Press,
595 Oxford.

596 Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and
597 speciation. Biological Journal of the Linnean Society **58**, 247–276.

598 Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the*
599 *Linnean Society* **68**, 87-112.

600 Hewitt GM (2001) Speciation, hybrid zones and phylogeography - or seeing genes in space and
601 time. *Molecular Ecology* **10**,(3) 537-49.

602 Hewitt GM (2011) Quaternary phylogeography: the roots of hybrid zones. *Genetica* **139**, 617-
603 638.

604 Hurst GDD, Jiggins FM, Pomiankowski A (2002) Which Way to Manipulate Host Reproduction?
605 *Wolbachia* That Cause Cytoplasmic Incompatibility Are Easily Invaded by Sex Ratio
606 Distorting Mutants. *The American Naturalist* **160**, 360-373.

607 Jiggins FM (2002) The rate of recombination in *Wolbachia* bacteria. *Molecular Biology and*
608 *Evolution* **19**, 1640-1643.

609 Jiggins FM, von der Schulenburg JHG, Hurst GDD, Majerus MEN (2001) Recombination
610 confounds interpretations of *Wolbachia* evolution. *Proceedings of the Royal Society B-*
611 *biological Sciences* **268**, 1423-1427.

612 Jolley KA, Feil EJ, Chan M-S, Maiden MCJ (2001) Sequence type analysis and recombinational
613 tests (START). *Bioinformatics* **17**, 1230-1231.

614 Jolley KA, Wilson DJ, Kriz P, Mcvean G, Maiden MCJ (2005) The Influence of Mutation,
615 Recombination, Population History, and Selection on Patterns of Genetic Diversity in
616 *Neisseria meningitidis*. *Molecular Biology and Evolution* **22**, 562-569.

617 Keller GP, Windsor DM, Saucedo JM, Werren JH (2004) Reproductive effects and geographical
618 distributions of two *Wolbachia* strains infecting the Neotropical beetle, *Chelymorpha*
619 *alternans* Boh. (Chrysomelidae, Cassidinae). *Molecular Ecology* **13**, 2405-2420.

620 Klasson L, Westberg J, Sapountzis P*, et al.* (2009) The mosaic genome structure of the
621 *Wolbachia* wRi strain infecting *Drosophila simulans*. *Proceedings of the National*
622 *Academy of Sciences of the United States of America* **106**, 5725-5730.

623 Kremer N, Huigens ME (2011) Vertical and horizontal transmission drive bacterial invasion.
624 *Molecular Ecology* **20**, 3496-3498.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Lunt D, Ibrahim K, Hewitt G (1998) mtDNA phylogeography and postglacial patterns of subdivision in the meadow grasshopper *Chorthippus parallelus*. *Heredity* **80**, 633-641.

Maiden MCJ, Bygraves JA, Feil E*, et al.* (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 3140-3145.

Martin D, Lemey P, Lott M*, et al.* (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462-2463.

Martins C, Souza RF, Bueno OC (2012) Presence and distribution of the endosymbiont *Wolbachia* among Solenopsis spp. (Hymenoptera: Formicidae) from Brazil and its evolutionary history. *Journal of Invertebrate Pathology.* **109**, 287-96

Martínez P, Del Castillo P, Bella JL (2009) Cytological detection of *Wolbachia* in squashed and paraffin embedded insect tissues. *Biotechnic & Histochemistry* **84**, 347-353.

Martinez-Rodriguez P (2013) *Divergencia inducida por Wolbachia en la zona híbrida de Chorthippus parallelus (Orthoptera).* PhD Thesis. Universidad Autónoma de Madrid.

Martínez-Rodríguez P, Hernández-Pérez M, Bella J (2013a) Detection of *Spiroplasma* and *Wolbachia* in the bacterial gonad community of *Chorthippus parallelus*. *Microbial Ecolology* **66**, 211-223.

Martínez-Rodríguez P, Sarasa J, Peco B*, et al.* (2013b) Endosymbiont-free ants: molecular biological evidence that neither *Wolbachia*, *Cardinium* or any other bacterial endosymbionts play a role in thelytokous parthenogenesis in the Harvester ant species, *Messor barbarus* and *M. capitatus*. *European Journal of Entomology* **110**, 197-204.

Maynard SJ (1992) Analyzing the mosaic structure of genes. *J Mol Evol* **34**, 126-129.

Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and Evolution of Heritable Bacterial Symbionts. *Annual Review of Genetics* **42**, 165-190.

Ochman H, Elwyn S, Moran NA (1999) Calibrating bacterial evolution. *Proceedings of the National Academy of Sciences* **96**, 12638-12643.

Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. . *J. Molec. Evol.* **26**, 74-86.

Padidam M, Sawyer S, Fauquet C (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218-225. .

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Posada D, Crandall K (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc Natl Acad Sci* **98**, 13757-13762. .

Raychoudhury R, Baldo L, Oliveira DCSG, Werren JH (2009) Modes of acquisition of *Wolbachia*: Horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. *Evolution* **63**, 165-183.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

666 Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM
667 (2012) Population Genomics of the *Wolbachia* Endosymbiont in *Drosophila*
668 *melanogaster*. PLoS Genet 8(12): e1003129.

669 Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed
670 models. *Bioinformatics* **19**, 1572-1574.

671 Sarasa J, Bernal A, Fernández-Calvín B, Bella JL (2013) *Wolbachia* Induced Cytogenetical Effects
672 as Evidenced in *Chorthippus parallelus* (Orthoptera). *Cytogenet Genome Res* **139**, 36-
673 43.

674 Saridaki A, Bourtzis K (2010) *Wolbachia*: more than just a bug in insects genitals. *Current*
675 *Opinion in Microbiology* **13**, 67-72.

676 Serrano L, delaVega CG, Bella JL*, et al.* (1996) A hybrid zone between two subspecies of
677 *Chorthippus parallelus*. X-chromosome variation through a contact zone. *Journal of*
678 *Evolutionary Biology* **9**, 173-184.

679 Shoemaker DD, Ahrens M, Sheill L*,* Mescherc M, Kellerd L, Rosse, KG (2003) Distribution and
680 prevalence of *Wolbachia* infections in native populations of the fire ant *Solenopsis*
681 *invicta* (Hymenoptera : Formicidae). *Environmental Entomology* **32**, 1329-1336.

682 Shuker D, King T, Bella J, Butlin R (2005) The genetic basis of speciation in a grasshopper hybrid
683 zone. In: *Insect Evolutionary Biology* (ed. Fellowes MHG, Roff J), pp. 427–454. CABI
684 Publishing, Oxford University Press, Wallingford, Oxon, UK.

685 Song H, Amédégnato C, Cigliano MM, Desutter-Grandcolas L, Heads SW, Huang Y, Otte D ,
686 Whiting MF (2015) 300 million years of diversification: elucidating the patterns of
687 orthopteran evolution based on comprehensive taxon and gene sampling. *Cladistics*
688 **31**, 621–651.

689 Telschow A, Hilgenboecker K, Hammerstein P, Werren JH (2014) Dobzhansky-Muller and
690 *Wolbachia*-Induced Incompatibilities in a Diploid Genetic System. PLoS ONE **9**, e95488.

691 Turelli M, Hoffmann AA (1991) Rapid spread of an inherited incompatibility factor in California
692 *Drosophila*. *Nature* **353**, 440-442.

693 Verne S, Johnson M, Bouchon D, Grandjean F (2007) Evidence for recombination between
694 feminizing *Wolbachia* in the isopod genus *Armadillidium*. *Gene* **397**, 58-66.

695 Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology.
696 *Nat Rev Micro* **6**, 741-751.

697 Werren JH, Bartos JD (2001) Recombination in *Wolbachia*. *Current Biology* **11**, 431-435.

698 Yun Y, Peng Y, Liu FX, Lei C (2011) *Wolbachia* screening in spiders and assessment of horizontal
699 transmission between predator and prey. *Neotropical Entomology* **40**, 164-169.

700 Zabal-Aguirre M, Arroyo F, Bella JL (2010) Distribution of *Wolbachia* infection in *Chorthippus*
701 *parallelus* populations within and beyond a Pyrenean hybrid zone. *Heredity* **104**, 174-
702 184.

703 Zabal-Aguirre M, Arroyo F, García-Hurtado J*, et al.* (2014) *Wolbachia* effects in natural
704 populations of *Chorthippus parallelus* from the Pyrenean hybrid zone. *Journal of*
705 *Evolutionary Biology* **27**, 1136-1148.

706 Zug R, Hammerstein P (2012) Still a host of hosts for *Wolbachia*: Analysis of recent data
707 suggests that 40% of terrestrial arthropod species are infected. *PLoS ONE* **7**, e38544.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tables:

Table 1: Coordinates and designation of sampled populations of *C. parallelus*. (1) Hybrid population according to (Serrano *et al.*, 1996). (2) Pure *Chorthippus parallelus* population, with some particular cytogenetics markers (see Bella *et al*. 2007). (3) Hybrid population in northern Spain (as characterized by Bella *et al*. 2007). (4) Hybrid population, according to (Flanagan *et al.* 1999) between *C. parallelus parallelus* and additional (5) Italian subspecies.

Population nomenclature (in figures)	Population	Hybrid-pure status	Latitude	Longitude	Altitude (m)
Pyrenees-Hybrid Zone					
HZ France	Arudy (France): ARU	Pure Cpp population	43° 06' 01" N	0° 26' 38" W	411
	Gabas (France): GAB	Pure Cpp population	42° 53' 60" N	0° 25' 60" W	1020
	L'Hermine (France): HER	Hybrid population	42° 51' 46.8" N	0° 23' 30.4" W	1209
	Soques (France): SOQ	Hybrid population	42° 20' 08" N	0° 23' 52" W	1396
Tourmont	Cabaña Tourmont (France): TOU	Hybrid population	42° 49' 11" N	0° 24' 21" W	1625
Portalet	Portalet (Spain): POR/PORCRU	Hybrid population	42° 48' 03" N	0° 24' 54" W	1780
CM	Corral de Mulas (Spain): CM	Hybrid population	42° 47' 09.4" N	0° 23' 34.4" W	1569
Sallent	Sallent de Gállego (Spain): SAL	Hybrid population[1]	42° 45' 57.5" N	0° 20' 33.9" W	1343
Escarrilla	Escarrilla (Spain): ESC	Pure Cpe population	42° 43' 54.1" N	0° 18' 39.3" W	1130
Pyrenees (other)					
South-Pyrenees/Vielha	Puerto del Cantó (Spain): PCAN	Pure Cpe population	42° 22' 12.9" N	1°14'11.7" E	1725
	Muna (France): MUN	Pure Cpp population	42° 53' 53" N	0° 37' 48.8" E	544
	Vielha (Spain): VIEL	Pure Cpe population	42° 40' 25.3" N	0° 46' 26.5" E	1393
Iberian peninsula					
Centre	Navafría (Spain): NAV	Pure Cpe population	40° 59' 01.95" N	3° 49' 00.9" W	1780
	Becedas (Spain): BEC	Pure Cpe population	40° 24' 18" N	5° 38' 17.2" W	1091
Bubion	Bubion (Spain): BUB	Pure Cpe population	36° 57' 1.8" N	3° 21' 22.8" W	1332

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

North	Basque Country I (Spain): ALA	Pure Cpe population[3]	42° 58' 41.4" N	2° 44' 19.7" W	625
	Basque Country II (Spain): URK	Pure Cpe population[3]	43° 13' 59.1" N	2° 29' 22.3" W	211
Europe					
Alps	Valdieri (Italy): VAL	Pure Cp population[5]	44° 12' 19.74" N	7° 22' 47.76" E	983
	Col de L'Arche (France): CLAR	Hybrid population[4]	44° 25' 34.3" N	6° 53' 21.6" E	1942
UK	Epping Forest (England): ING	Pure Cpp population	51° 39' 36" N	0° 3' 0" E	102
Slovenia	Mokronog (Slovenia): SLO	Pure Cpp population	45° 56' 37.17" N	15° 8' 55.428" E	242 m

715

716

717 Table 2: Primer sequences used in the study.

Gene	Sequences	Amplicon size (bp)
16S rRNA (1st PCR)	16SF: 5' TTG TAG CTT GCT ATG GTA TAA CT 3'	1490
	16SR: 5' ACT GCT ACC TTG TTA CGA CTT 3'	
16S rRNA (2nd nested-PCR)	Rev: 5' TAT CCC TTC GAA TAG GTA TGA TTT 3'	750
	FF: 5' TGA GCC TAT ATT AGA TTA GCT AGT TGG TAA G 3'	
	FB: 5' GCC TAT ATT AGA TTA GCT AGT TGG TGG A 3'	
gatB	gatB_F1: 5' GAK TTA AAY CGY GCA GGB GTT 3'	471
	gatB_R1: 5' TGG YAA YTC RGG YAA AGA TGA 3'	
coxA	coxA_F1: 5' TTG GRG CRA TYA ACT TTA TAG 3'	487
	coxA_R1: 5' CT AAA GAC TTT KAC RCC AGT 3'	
hcpA	hcpA_F1: 5' GAA ATA RCA GTT GCT GCA AA 3'	515
	hcpA_R1: 5' GAA AGT YRA GCA AGY TCT G 3'	
ftsZ	ftsZ_F1: 5' ATY ATG GAR CAT ATA AAR GAT AG 3'	524
	ftsZ_R1: 5' TCR AGY AAT GGA TTR GAT AT 3'	
fbpA	fbpA_F1: 5' GCT GCT CCR CTT GGY WTG AT 3'	509
	fbpA_R1: 5' CCR CCA GAR AAA AYY ACT ATT C 3'	
wsp	wsp_F1: GTCCAATARSTGATGARGAAAC	603
	wsp_R1: CYGCACCAAYAGYRCTRTAAA)	

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

718 Table 3: Analysis of molecular variance (AMOVA) from five MLST genes for the F supergroup of *Wolbachia*
719 infecting different populations of *C. parallelus.*

Source of variation	df	Sum of squares	Variance component	Percentage of variation
Between groups	3.00	820.69	13.60	39.80
Between populations within groups	9.00	305.51	4.53	13.25
Between individuals within populations	53.00	850.90	16.05	46.96
Total	65.00	1977.11	34.19	
Indels:	**Value**	**p**		
F_{sc}	0.220	<0.0001		
F_{st}	0.530	<0.0001		
F_{ct}	0.397	<0.0001		

720

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental tables:

Table S1: Accession numbers.

gen	allele	Genebank code
CoxA	H1	KM078848
	H2	KM078849
	H3	KM078850
	H4	KM078851
	H5	KM078852
	H6	KM078853
fbpA	H1	KM078854
	H2	KM078855
	H3	KM078856
	H4	KM078857
	H5	KM078858
ftsZ	H1	KM078859
	H2	KM078860
	H3	KM078861
	H4	KM078862
	H5	KM078863
gatB	H1	KM078864
	H2	KM078865
	H3	KM078866
	H4	KM078867
	H5	KM078868
hcpA	H1	KM078869
	H2	KM078870
	H3	KM078871
	H4	KM078872
	H5	KM078873
	H6	KM078874
	H7	KM078875
	H8	KM078876
	H9	KM078877
	H10	KM078878
wsp	H1	KM078879
	H2	KM078880
	H3	KM078881
	H4	KM078882
	H5	KM078883

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

724 *Table S2*: *Wolbachia* infection frequencies in the analysed populations:

	n					%				
NAME	Ø	F	B	FB	total	Ø	F	B	FB	TOTAL
ESC2008	19	7	3	9	38	0,50	0,18	0,08	0,24	1,00
ESC2009	42	34	5	7	88	0,48	0,39	0,06	0,08	1,00
SAL2008	3	3	12	30	48	0,06	0,06	0,25	0,63	1,00
SAL2009	9	5	16	21	51	0,18	0,10	0,31	0,41	1,00
CM2008	41	9	0	4	54	0,76	0,17	0,00	0,07	1,00
CM2009	25	7	10	17	59	0,42	0,12	0,17	0,29	1,00
POR2008	36	14	2	9	61	0,59	0,23	0,03	0,15	1,00
POR2009	69	12	10	0	91	0,76	0,13	0,11	0,00	1,00
C_TOU2008	45	5	9	1	60	0,75	0,08	0,15	0,02	1,00
C_TOU2009	46	4	10	3	63	0,73	0,06	0,16	0,05	1,00
SOQ2008	20	2	7	3	32	0,63	0,06	0,22	0,09	1,00
SOQ2009	27	3	18	11	59	0,46	0,05	0,31	0,19	1,00
LHER2008	24	4	15	5	48	0,50	0,08	0,31	0,10	1,00
LHER2009	23	6	16	10	55	0,42	0,11	0,29	0,18	1,00
GAB2008	30	6	6	3	45	0,67	0,13	0,13	0,07	1,00
GAB2009	36	9	10	8	63	0,57	0,14	0,16	0,13	1,00
ARU2008	20	1	13	0	34	0,59	0,03	0,38	0,00	1,00
ARU2009	71	2	9	8	90	0,79	0,02	0,10	0,09	1,00
NAV2007	22	36	17	33	108	0,20	0,33	0,16	0,31	1,00
NAV2008	0	8	4	10	22	0,00	0,36	0,18	0,46	1,00
PCANTO2008	21	22	0	6	49	0,43	0,45	0,00	0,12	1,00
MUNA2008	21	5	19	7	52	0,40	0,10	0,37	0,14	1,00
COLLARCHE2009	16	0	5	0	21	0,76	0,00	0,24	0,00	1,00
VALDIERI I and II 2009	55	3	17	1	76	0,72	0,04	0,22	0,01	1,00
LARGENTERA2009	2	0	6	0	8	0,25	0,00	0,75	0,00	1,00
BUBION2004	2	47	0	4	53	0,04	0,89	0,00	0,08	1,00
EPPING2002	6	0	7	2	15	0,40	0,00	0,47	0,13	1,00
VIELHA2008*	28	8	1	1	38	0,74	0,21	0,03	0,03	1,00
URKAREGUI2009	35	7	6	1	49	0,71	0,14	0,12	0,02	1,00
TXABARRI2009	9	0	0	1	10	0,90	0,00	0,00	0,10	1,00
TARA2008	57	2	6	4	69	0,83	0,03	0,09	0,06	1,00
TARA2009	73	1	0	3	77	0,95	0,01	0,00	0,04	1,00
ALAV_1_2009	33	12	1	1	47	0,70	0,26	0,02	0,02	1,00
SLOVENIA2009	0	40	4	5	49	0,00	0,82	0,08	0,10	1,00
TOTAL	966	324	264	228	1782	0,54	0,18	0,15	0,13	1,00

*Table S*3: Genetic diversity:

1. P=analysed positions
2. S= total polymorphic positions
3. Eta= total frequency of mutations
4. K= average number of nucleotide differences
5. Hap= frequency of haplotypes
6. Hd= Haplotype diversity
7. VarHd, Haplotype diversity Variance
8. Pi= nucleotide diversity
9. Theta Waterson = 4Nu, where N is the effective population size, and u is the mutation rate per nucleotide (or per sequence) and per generation (following Watterson 1975, Nei 1987)."
10. Tajima D, FuLiD*and FuliF* statistics to test various predictions of the neutral theory of molecular evolution (according with Tajima 1989) (Fu & Li 1993) and their significance: ** 0.1, ***0.01, G+C= G+C content. R=Recombination (MAXCHI, (Maynard Smith 1992)).

Gene	**n**	**P**	**S**	**Eta**	**K**	**Hap**	**Hd/VarHd**	**Pi**	**ThetaNuc**	**AvNumDif**	**ThetaG**	**TajimaD**	**FuLiD***	**FuLiF***	**G+C**	**R (MAXCHI, p<0.01)**
coxa	111	402	41	42	16.25	6	0.8/0.0002	0.040	0.020	16.251	7.951	3.238**	2.122**	3.090**	0.385	No
fbpa	117	429	61	62	11.62	5	0.8/0.0001	0.068	0.028	28.579	11.621	4.625***	2.307**	3.943**	0.394	Yes: 3 (83)
ftsz	112	435	59	59	27.80	5	0.7/0.0002	0.064	0.026	27.801	11.151	4.740***	2.274**	3.965**	0.407	Yes: 1 (34)
gatB	114	370	39	39	17.30	5	0.7/0.0004	0.047	0.020	17.298	7.346	4.165***	2.092**	3.517**	0.370	No
hcpA	115	419	57	57	23.66	10	0.9/0.0002	0.056	0.026	23.662	10.719	3.815***	2.057**	3.351**	0.366	No

Tabla S4: locus by locus AMOVA implemented in ARLEQUIN.

	Among groups				Among population, between groups				Within population				Fixation index					
Locus	SSD	g.l.	Va(%)	variación	SSD	g.l.	Vb(%)	variación	SSD	g.l.	Vc(%)	variación	F_{SC}	P-valor	F_{ST}	P-valor	F_{CT}	P-valor
gatB	1437156	3	0.28	80.20	116632	9	0.02	5.86	258333	53	0.05	13.95	0.2957	0.0557	0.8605	0.0000	0.8020	0.0029
coxA	1809649	3	0.35	73.60	208533	9	0.04	7.74	465152	53	0.09	18.65	0.2934	0.0059	0.8135	0.0000	0.7360	0.0000
HcpA	1304275	3	0.17	39.07	756331	9	0.20	45.04	365152	53	0.07	15.89	0.7392	0.0000	0.8411	0.0000	0.3907	0.0166
ftsZ	1096416	3	0.20	53.22	270250	9	0.04	11.91	677273	53	0.13	34.87	0.2547	0.0274	0.6513	0.0000	0.5322	0.0039
fbpA	1280584	3	0.20	47.92	486841	9	0.11	25.65	593182	53	0.11	26.43	0.4925	0.0000	0.7357	0.0000	0.4792	0.0068

Table S5: Exact test of population differentiation following the methodology of Rousset *et al.* (1992) implemented in ARLEQUIN. Gray = populations among which differentiation is observed.

	Pto. Cantó	Centre	Escarrilla	Sallent	Alps	North	Vielha	Bubión	Slovenia	Tourmont	Portalet	C. Mulas	ZH France
Pto. Cantó													
Centre	0.00458 ± 0.0008												
Escarrilla	0.10008 ± 0.0021	0.00216 ± 0.0005											
Sallent	0.24378 ± 0.0048	0.00015 ± 0.0001	0.10440 ± 0.0042										
Alps	0.06187 ± 0.0015	0.00021 ± 0.0001	0.03494 ± 0.0017	0.39706 ± 0.0041									
North	0.10136 ± 0.0015	0.00207 ± 0.0004	≈ 0.00000	0.10483 ± 0.0028	0.03525 ± 0.0011								
Vielha	0.01245 ± 0.0010	0.00002 ± 0.0000	0.01049 ± 0.0008	0.00210 ± 0.0006	0.00247 ± 0.0005	0.01230 ± 0.0006							
Bubión	0.05564 ± 0.0030	≈ 0.00000	0.02453 ± 0.0024	0.03982 ± 0.0011	0.00212 ± 0.0004	0.02717 ± 0.0008	0.00013 ± 0.0001						

Slovenia	0.50524 ± 0.0020	0.15809 ± 0.0027	0.25167 ± 0.0021	0.70584 ± 0.0059	0.27881 ± 0.0042	0.24826 ± 0.0023	≈ 0.00000	0.42498 ± 0.0123					
Tourmont	0.01702 ± 0.0008	0.00038 ± 0.0002	0.01731 ± 0.0006	0.01048 ± 0.0013	0.00451 ± 0.0004	0.01802 ± 0.0009	≈ 0.00000	0.00146 ± 0.0004	≈ 0.00000				
Portalet	0.00802 ± 0.0009	≈ 0.00000	≈ 0.00000	0.00049 ± 0.0003	0.00033 ± 0.0002	≈ 0.00000	0.00052 ± 0.0001	0.00046 ± 0.0003	0.12316 ± 0.0015	0.00135 ± 0.0003			
C. Mulas	0.39703 ± 0.0045	0.03246 ± 0.0018	0.09899 ± 0.0013	0.57693 ± 0.0065	0.14184 ± 0.0023	0.10019 ± 0.0014	0.03558 ± 0.0014	0.37234 ± 0.0104	1.00000 ± 0.0000	0.04496 ± 0.0012	0.02705 ± 0.0015		
ZH France	0.49858 ± 0.0032	0.15600 ± 0.0035	0.24915 ± 0.0022	0.70848 ± 0.0057	0.28748 ± 0.0067	0.25371 ± 0.0019	0.14421 ± 0.0017	0.41815 ± 0.0090	1.00000 ± 0.0000	0.16910 ± 0.0021	0.12423 ± 0.0018	1.00000 ± 0.0000	

746

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S6: Rate of synonymous divergence Ks-JC, detected between different alleles for gene *coxa* of *Wolbachia* infecting *C. parallelus*. SD: Standard deviation, calculated using the formula of the standard deviation of the ratio given by Neter *et al*. (1978) as proposed by Raychoudhury *et al*. (2009).

Allele x	Allele y	Ks	±SD	Divergence (years)	±years
H1	H2	0.365	0.053	40.588.889	5.872.561
H1	H3	0.363	0.053	40.377.778	5.854.403
H1	H4	0.037	0.021	4.111.111	2.302.144
H1	H5	0.365	0.053	40.588.889	5.872.561
H1	H6	0.346	0.052	38.433.333	5.801.173
H2	H3	0.190	0.043	21.133.333	4.776.953
H2	H4	0.385	0.053	42.788.889	5.934.819
H2	H5	0.012	0.012	1.344.444	1.333.422
H2	H6	0.012	0.012	1.344.444	1.333.422
H3	H4	0.383	0.053	42.566.667	5.917.255
H3	H5	0.175	0.042	19.433.333	4.623.863
H3	H6	0.175	0.042	19.433.333	4.623.863
H4	H5	0.385	0.053	42.788.889	5.934.819
H4	H6	0.365	0.053	40.588.889	5.872.561
H5	H6	0.025	0.017	2.722.222	1.885.450

Table S7: Rate of synonymous divergence Ks-JC, detected between different alleles for gene *fbpA* of *Wolbachia* infecting *C. parallelus*. SD: Standard deviation, calculated using the formula of the standard deviation of the ratio given by Neter *et al.* (1978) as proposed by Raychoudhury *et al.* (2009).

Allele x	Allele y	Ks	±SD	Divergence (years)	±years
H1	H2	0.659	0.047	73.177.778	5.199.701
H1	H2	0.040	0.019	4.444.444	2.147.990
H1	H3	0.681	0.046	75.622.222	5.108.707
H1	H5	0.030	0.017	3.311.111	1.863.101
H2	H2	0.661	0.047	73.455.556	5.196.782
H2	H3	0.010	0.010	1.088.889	1.081.110
H2	H5	0.660	0.047	73.366.667	5.197.733
H2	H3	0.683	0.046	75.922.222	5.103.356
H2	H5	0.072	0.025	7.944.444	2.826.638
H3	H5	0.682	0.046	75.822.222	5.104.997

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S8: Rate of synonymous divergence Ks-JC, detected between different alleles for gene *gatB* of *Wolbachia* infecting *C. parallelus*. SD: Standard deviation, calculated using the formula of the standard deviation of the ratio given by Neter *et al.* (1978) as proposed by Raychoudhury *et al.* (2009).

Allele x	Allele y	Ks	±SD	Divergence (years)	±years
H1	H2	0.455	0.056	50.555.556	6.238.165
H1	H3	0.013	0.013	1.422.222	1.409.711
H1	H4	0.506	0.056	56.244.444	6.276.681
H1	H5	0.000	0.000	0	0
H2	H3	0.432	0.056	48.000.000	6.205.388
H2	H4	0.026	0.018	2.877.778	1.991.931
H2	H5	0.455	0.056	50.555.556	6.238.165
H3	H4	0.482	0.056	53.500.000	6.272.865
H3	H5	0.013	0.013	1.422.222	1.409.711
H4	H5	0.506	0.056	56.244.444	6.276.681

Table S9: Rate of synonymous divergence Ks-JC, detected between different alleles for gene *ftsZ* of *Wolbachia* infecting *C. parallelus*. SD: Standard deviation, calculated using the formula of the standard deviation of the ratio given by Neter *et al.* (1978) as proposed by Raychoudhury *et al.* (2009).

Allele x	Allele y	Ks	±SD	Divergence (years)	±years
H1	H2	0.765	0.042	85.022.222	4.659.425
H1	H3	0.020	0.014	2.188.889	1.523.871
H1	H4	0.010	0.010	1.088.889	1.079.374
H1	H5	0.739	0.043	82.055.556	4.830.665
H2	H3	0.767	0.042	85.255.556	4.648.771
H2	H4	0.739	0.043	82.055.556	4.830.665
H2	H5	0.030	0.017	3.355.556	1.887.422
H3	H4	0.030	0.017	3.311.111	1.864.552
H3	H5	0.741	0.043	82.277.778	4.822.682
H4	H5	0.713	0.045	79.188.889	4.974.128

Table S10: Rate of synonymous divergence Ks-JC, detected between different alleles for gene *hcpA* of *Wolbachia* infecting *C. parallelus*. SD: Standard deviation, calculated using the formula of the standard deviation of the ratio given by Neter *et al.* (1978) as proposed by Raychoudhury *et al.* (2009).

Allele x	Allele y	Ks	±SD	Divergence (years)	±years
H1	H2	0.454	0.051	50.488.889	5.622.233
H1	H3	0.021	0.015	2.322.222	1.612.420
H1	H4	0.454	0.051	50.488.889	5.622.233
H1	H5	0.443	0.050	49.211.111	5.599.004
H1	H6	0.065	0.025	7.177.778	2.773.234

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

H1	H7	0.387	0.049	43.044.444	5.481.804
H1	H8	0.042	0.020	4.700.000	2.268.696
H1	H9	0.111	0.032	12.277.778	3.530.923
H1	H10	0.388	0.049	43.144.444	5.488.922
H2	H3	0.494	0.051	54.933.333	5.650.077
H2	H4	0.000	0.000	0	0
H2	H5	0.000	0.000	0	0
H2	H6	0.506	0.051	56.188.889	5.655.039
H2	H7	0.032	0.018	3.511.111	1.973.525
H2	H8	0.494	0.051	54.933.333	5.650.077
H2	H9	0.464	0.051	51.544.444	5.631.027
H2	H10	0.032	0.018	3.511.111	1.975.256
H3	H4	0.494	0.051	54.933.333	5.650.077
H3	H5	0.482	0.051	53.577.778	5.637.236
H3	H6	0.065	0.025	7.188.889	2.777.667
H3	H7	0.424	0.050	47.066.667	5.565.114
H3	H8	0.042	0.020	4.711.111	2.273.247
H3	H9	0.087	0.029	9.688.889	3.180.076
H3	H10	0.425	0.050	47.177.778	5.571.425
H4	H5	0.000	0.000	0	0
H4	H6	0.506	0.051	56.188.889	5.655.039
H4	H7	0.032	0.018	3.511.111	1.973.525
H4	H8	0.494	0.051	54.933.333	5.650.077
H4	H9	0.464	0.051	51.544.444	5.631.027
H4	H10	0.032	0.018	3.511.111	1.975.256
H5	H6	0.493	0.051	54.800.000	5.645.239
H5	H7	0.0315	0.018	3.500.000	1.967.158
H5	H8	0.482	0.051	53.577.778	5.637.236
H5	H9	0.452	0.050	50.244.444	5.610.062
H5	H10	0.032	0.018	3.500.000	1.968.777
H6	H7	0.434	0.050	48.188.889	5.586.109
H6	H8	0.021	0.015	2.322.222	1.615.248
H6	H9	0.042	0.020	4.711.111	2.273.247
H6	H10	0.435	0.050	48.300.000	5.592.499
H7	H8	0.424	0.050	47.066.667	5.565.114
H7	H9	0.396	0.050	44.011.111	5.503.515
H7	H10	0.000	0.000	0	0
H8	H9	0.064	0.025	7.155.556	2.766.811
H8	H10	0.425	0.050	47.177.778	5.571.425
H9	H10	0.397	0.050	44.100.000	5.510.225

778

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table S11: Synonymous divergence Ks-JC between the main ST belonging to supergroup F detected in both subspecies of *C. parallelus*. SD: Standard deviation, calculated using the formula of the standard deviation of the ratio given by Neter *et al.* (1978) as proposed by Raychoudhury *et al.* (2009).

Allele x	Allele y	Ks	±SD	Divergence (years)	±years
ST10	ST11	0.0109	0.005	1.211.111	535.687
ST10	ST24	0.0376	0.009	4.177.778	982.023
ST10	ST25	0.0241	0.007	2.677.778	791.342
ST11	ST24	0.0445	0.010	4.944.444	1 064.878
ST11	ST25	0.0308	0.008	3.422.222	891.855
ST24	ST25	0.0131	0.005	1.455.556	587.295

Table S12: synonymous divergence Ks-JC between the main ST belonging to supergroup B detected in both subspecies of *C. parallelus*. SD: Standard deviation, calculated using the formula of the standard deviation of the ratio given by Neter *et al.* (1978) as proposed by Raychoudhury *et al.* (2009).

Allele x	Allele y	Ks	±SD	Divergence (years)	±years
ST13	ST31	0.014	0.006	1.566.667	633.171
ST6	ST31	0.012	0.005	1.311.111	579.907
ST13	ST33	0.012	0.005	1.311.111	580.524
ST14	ST31	0.012	0.005	1.311.111	579.907
ST17	ST31	0.012	0.005	1.311.111	579.677
ST3	ST31	0.009	0.005	1.044.444	518.213
ST6	ST33	0.009	0.005	1.044.444	518.764
ST7	ST31	0.009	0.005	1.044.444	518.007
ST14	ST33	0.009	0.005	1.044.444	518.764
ST16	ST31	0.009	0.005	1.044.444	518.007
ST17	ST33	0.009	0.005	1.044.444	518.564
ST3	ST33	0.007	0.004	788.889	451.377
ST7	ST33	0.007	0.004	788.889	451.202
ST13	ST32	0.007	0.004	788.889	451.292
ST16	ST33	0.007	0.004	788.889	451.202
ST6	ST16	0.007	0.004	777.778	447.947
ST32	ST31	0.007	0.004	777.778	447.733
ST3	ST13	0.005	0.003	522.222	367.622
ST3	ST16	0.005	0.003	522.222	367.476
ST6	ST14	0.005	0.003	522.222	367.622
ST6	ST17	0.005	0.003	522.222	367.476
ST6	ST32	0.005	0.003	522.222	367.622
ST7	ST13	0.005	0.003	522.222	367.476
ST7	ST16	0.005	0.003	522.222	367.334
ST13	ST16	0.005	0.003	522.222	367.476
ST14	ST32	0.005	0.003	522.222	367.622
ST17	ST32	0.005	0.003	522.222	367.476
ST32	ST33	0.005	0.003	522.222	367.691
ST3	ST6	0.002	0.002	255.556	257.478
ST3	ST14	0.002	0.002	255.556	257.478

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ST3	ST17	0.002	0.002	255.556	257.375
ST3	ST32	0.002	0.002	255.556	257.478
ST6	ST7	0.002	0.002	255.556	257.375
ST6	ST13	0.002	0.002	255.556	257.478
ST7	ST14	0.002	0.002	255.556	257.375
ST7	ST17	0.002	0.002	255.556	257.276
ST7	ST32	0.002	0.002	255.556	257.375
ST13	ST14	0.002	0.002	255.556	257.478
ST13	ST17	0.002	0.002	255.556	257.375
ST14	ST16	0.002	0.002	255.556	257.375
ST16	ST17	0.002	0.002	255.556	257.276
ST16	ST32	0.002	0.002	255.556	257.375
ST33	ST31	0.002	0.002	255.556	257.300
ST3	ST7	0.000	0.000	0	0
ST14	ST17	0.000	0.000	0	0

Table s13: Alleles includes in phylogenetic analysis of gatB

gatB_1	gatB_21	gatB_121	gatB_31	gatB_97	gatB_23	gatB_66
gatB_2	gatB_38	gatB_122	gatB_30	gatB_129	gatB_36	gatB_72
gatB_62	gatB_39	gatB_126	gatB_73	gatB_131	gatB_32	gatB_128
gatB_68	gatB_40	gatB_127	gatB_65	gatB_141	gatB_49	gatB_99
gatB_88	gatB_48	gatB_132	gatB_82	gatB_142	gatB_53	gatB_130
gatB_89	gatB_55	gatB_134	gatB_113	gatB_143	gatB_56	gatB_138
gatB_3	gatB_69	gatB_139	gatB_112	gatB_18	gatB_76	gatB_157
gatB_4	gatB_70	gatB_140	gatB_115	gatB_46	gatB_94	gatB_22
gatB_83	gatB_71	gatB_145	gatB_116	gatB_133	gatB_98	
gatB_101	gatB_125	gatB_147	gatB_117	gatB_90	gatB_123	
gatB_135	gatB_79	gatB_149	gatB_118	gatB_33	gatB_37	
gatB_137	gatB_80	gatB_150	gatB_81	gatB_34	gatB_78	
gatB_152	gatB_91	gatB_151	gatB_110	gatB_124	gatB_93	
gatB_96	gatB_100	gatB_153	gatB_111	gatB_19	gatB_42	
gatB_148	gatB_102	gatB_155	gatB_28	gatB_20	gatB_43	
gatB_5	gatB_154	gatB_158	gatB_64	gatB_24	gatB_75	
gatB_6	gatB_103	gatB_15	gatB_7	gatB_27	gatB_45	
gatB_17	gatB_104	gatB_156	gatB_47	gatB_41	gatB_54	
gatB_25	gatB_105	gatB_59	gatB_67	gatB_74	gatB_87	
gatB_136	gatB_106	gatB_95	gatB_85	gatB_35	gatB_57	
gatB_108	gatB_107	gatB_144	gatB_86	gatB_51	gatB_58	
gatB_9	gatB_146	gatB_92	gatB_8	gatB_52	gatB_60	
gatB_12	gatB_109	gatB_26	gatB_10	gatB_44	gatB_61	
gatB_13	gatB_119	gatB_114	gatB_11	gatB_50	gatB_63	
gatB_16	gatB_120	gatB_29	gatB_14	gatB_77	gatB_84	

Table s14: Alleles includes in phylogenetic analysis of coxA

coxA_1	coxA_105	coxA_131	coxA_87	coxA_129	coxA_32

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

coxA_6	coxA_106	coxA_83	coxA_91	coxA_2	coxA_61
coxA_17	coxA_122	coxA_13	coxA_112	coxA_60	coxA_10
coxA_23	coxA_123	coxA_5	coxA_27	coxA_70	coxA_37
coxA_103	coxA_126	coxA_88	coxA_69	coxA_110	coxA_58
coxA_111	coxA_133	coxA_96	coxA_94	coxA_20	coxA_130
coxA_113	coxA_139	coxA_25	coxA_82	coxA_24	coxA_84
coxA_116	coxA_9	coxA_11	coxA_95	coxA_28	coxA_50
coxA_117	coxA_80	coxA_38	coxA_97	coxA_35	coxA_53
coxA_118	coxA_22	coxA_79	coxA_92	coxA_45	coxA_77
coxA_124	coxA_78	coxA_121	coxA_93	coxA_46	coxA_44
coxA_143	coxA_68	coxA_135	coxA_30	coxA_40	coxA_48
coxA_140	coxA_16	coxA_136	coxA_31	coxA_41	coxA_49
coxA_141	coxA_119	coxA_144	coxA_55	coxA_21	coxA_52
coxA_3	coxA_137	coxA_138	coxA_56	coxA_108	coxA_59
coxA_8	coxA_64	coxA_43	coxA_63	coxA_109	coxA_72
coxA_114	coxA_73	coxA_67	coxA_89	coxA_33	coxA_104
coxA_12	coxA_132	coxA_120	coxA_90	coxA_34	coxA_57
coxA_14	coxA_134	coxA_98	coxA_29	coxA_39	coxA_62
coxA_18	coxA_26	coxA_99	coxA_54	coxA_74	
coxA_36	coxA_107	coxA_100	coxA_15	coxA_47	
coxA_51	coxA_65	coxA_101	coxA_42	coxA_75	
coxA_81	coxA_66	coxA_102	coxA_76	coxA_142	
coxA_115	coxA_125	coxA_71	coxA_127	coxA_7	
coxA_85	coxA_4	coxA_86	coxA_128	coxA_19	

Table s15: Alleles includes in phylogenetic analysis of fbpA

fbpA_1	fbpA_192	fbpA_112	fbpA_147
fbpA_2	fbpA_205	fbpA_98	fbpA_186
fbpA_92	fbpA_150	fbpA_113	fbpA_199
fbpA_10	fbpA_157	fbpA_116	fbpA_220
fbpA_21	fbpA_16	fbpA_117	fbpA_121
fbpA_142	fbpA_195	fbpA_123	
fbpA_6	fbpA_201	fbpA_65	
fbpA_91	fbpA_219	fbpA_15	
fbpA_137	fbpA_41	fbpA_58	
fbpA_181	fbpA_125	fbpA_66	
fbpA_185	fbpA_129	fbpA_79	
fbpA_141	fbpA_145	fbpA_61	
fbpA_143	fbpA_169	fbpA_154	
fbpA_155	fbpA_70	fbpA_163	
fbpA_214	fbpA_73	fbpA_179	
fbpA_4	fbpA_57	fbpA_19	
fbpA_43	fbpA_101	fbpA_36	

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

fbpA_75	fbpA_103	fbpA_47
fbpA_76	fbpA_106	fbpA_52
fbpA_176	fbpA_107	fbpA_54
fbpA_25	fbpA_50	fbpA_85
fbpA_95	fbpA_140	fbpA_96
fbpA_162	fbpA_86	fbpA_8
fbpA_207	fbpA_108	fbpA_83
fbpA_132	fbpA_110	fbpA_119

Table s16: Alleles includes in phylogenetic analysis of ftsZ

ftsZ_1	ftsZ_98	ftsZ_22	ftsZ_66	ftsZ_33
ftsZ_3	ftsZ_99	ftsZ_106	ftsZ_63	ftsZ_58
ftsZ_5	ftsZ_100	ftsZ_108	ftsZ_36	ftsZ_14
ftsZ_6	ftsZ_103	ftsZ_110	ftsZ_41	ftsZ_50
ftsZ_10	ftsZ_104	ftsZ_109	ftsZ_65	ftsZ_13
ftsZ_17	ftsZ_116	ftsZ_111	ftsZ_73	ftsZ_56
ftsZ_29	ftsZ_51	ftsZ_4	ftsZ_80	ftsZ_74
ftsZ_32	ftsZ_53	ftsZ_7	ftsZ_78	ftsZ_113
ftsZ_34	ftsZ_54	ftsZ_8	ftsZ_81	ftsZ_114
ftsZ_38	ftsZ_24	ftsZ_77	ftsZ_89	ftsZ_115
ftsZ_39	ftsZ_85	ftsZ_107	ftsZ_90	ftsZ_79
ftsZ_40	ftsZ_27	ftsZ_9	ftsZ_94	ftsZ_92
ftsZ_42	ftsZ_59	ftsZ_11	ftsZ_97	ftsZ_93
ftsZ_43	ftsZ_28	ftsZ_12	ftsZ_95	ftsZ_30
ftsZ_44	ftsZ_48	ftsZ_96	ftsZ_101	ftsZ_31
ftsZ_45	ftsZ_60	ftsZ_15	ftsZ_105	ftsZ_75
ftsZ_47	ftsZ_84	ftsZ_62	ftsZ_117	ftsZ_76
ftsZ_49	ftsZ_68	ftsZ_102	ftsZ_19	
ftsZ_52	ftsZ_61	ftsZ_112	ftsZ_91	
ftsZ_55	ftsZ_86	ftsZ_69	ftsZ_16	
ftsZ_57	ftsZ_88	ftsZ_35	ftsZ_21	
ftsZ_64	ftsZ_87	ftsZ_71	ftsZ_67	
ftsZ_70	ftsZ_26	ftsZ_18	ftsZ_37	
ftsZ_72	ftsZ_83	ftsZ_20	ftsZ_25	
ftsZ_82	ftsZ_2	ftsZ_23	ftsZ_46	

Table s17: Alleles includes in phylogenetic analysis of hcpA

hcpA_106	hcpA_49	hcpA_69	hcpA_76	hcpA_114	hcpA_119	hcpA_153
hcpA_130	hcpA_46	hcpA_96	hcpA_134	hcpA_115	hcpA_164	hcpA_158
hcpA_144	hcpA_54	hcpA_61	hcpA_18	hcpA_116	hcpA_136	hcpA_157
hcpA_13	hcpA_45	hcpA_64	hcpA_131	hcpA_112	hcpA_151	hcpA_160

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

hcpA_135	hcpA_78	hcpA_24	hcpA_95	hcpA_73	hcpA_152	hcpA_156
hcpA_12	hcpA_28	hcpA_63	hcpA_60	hcpA_117	hcpA_146	hcpA_159
hcpA_150	hcpA_133	hcpA_59	hcpA_16	hcpA_107	hcpA_102	hcpA_140
hcpA_166	hcpA_97	hcpA_92	hcpA_23	hcpA_108	hcpA_99	hcpA_155
hcpA_165	hcpA_75	hcpA_149	hcpA_83	hcpA_14	hcpA_154	hcpA_74
hcpA_137	hcpA_26	hcpA_37	hcpA_47	hcpA_124	hcpA_161	hcpA_5
hcpA_138	hcpA_86	hcpA_65	hcpA_82	hcpA_121	hcpA_9	hcpA_4
hcpA_1	hcpA_167	hcpA_11	hcpA_127	hcpA_120	hcpA_91	hcpA_15
hcpA_27	hcpA_2	hcpA_128	hcpA_21	hcpA_122	hcpA_104	hcpA_10
hcpA_68	hcpA_51	hcpA_62	hcpA_71	hcpA_101	hcpA_148	hcpA_142
hcpA_103	hcpA_98	hcpA_89	hcpA_33	hcpA_147	hcpA_145	hcpA_141
hcpA_8	hcpA_39	hcpA_90	hcpA_36	hcpA_168	hcpA_52	hcpA_30
hcpA_43	hcpA_38	hcpA_41	hcpA_34	hcpA_29	hcpA_126	hcpA_113
hcpA_32	hcpA_42	hcpA_93	hcpA_58	hcpA_20	hcpA_25	hcpA_87
hcpA_56	hcpA_94	hcpA_84	hcpA_35	hcpA_132	hcpA_100	
hcpA_57	hcpA_81	hcpA_50	hcpA_72	hcpA_88	hcpA_125	
hcpA_55	hcpA_105	hcpA_85	hcpA_77	hcpA_109	hcpA_6	
hcpA_48	hcpA_53	hcpA_67	hcpA_110	hcpA_3	hcpA_17	
hcpA_129	hcpA_44	hcpA_70	hcpA_111	hcpA_163	hcpA_143	
hcpA_79	hcpA_123	hcpA_7	hcpA_31	hcpA_19	hcpA_139	
hcpA_22	hcpA_66	hcpA_80	hcpA_118	hcpA_40	hcpA_162	

Table s18: Alleles includes in phylogenetic analysis of wsp

wsp_111	wsp_6	wsp_49
wsp_31	wsp_10	wsp_40
wsp_127	wsp_26	wsp_115
wsp_5	wsp_25	wsp_117
wsp_128	wsp_7	wsp_118
wsp_9	wsp_35	wsp_120
wsp_11	wsp_36	wsp_119
wsp_22	wsp_37	wsp_122
wsp_130	wsp_74	wsp_124
wsp_14	wsp_75	wsp_151
wsp_4	wsp_76	wsp_154
wsp_1	wsp_77	wsp_157
wsp_18	wsp_79	wsp_158
wsp_21	wsp_80	wsp_101
wsp_23	wsp_113	wsp_43
wsp_33	wsp_38	wsp_41
wsp_8	wsp_87	wsp_39
wsp_29	wsp_85	wsp_45
wsp_27	wsp_90	wsp_44
wsp_28	wsp_84	wsp_83

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

wsp_2	wsp_103
wsp_3	wsp_89
wsp_20	wsp_106
wsp_30	wsp_46
wsp_15	wsp_48

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental figures:

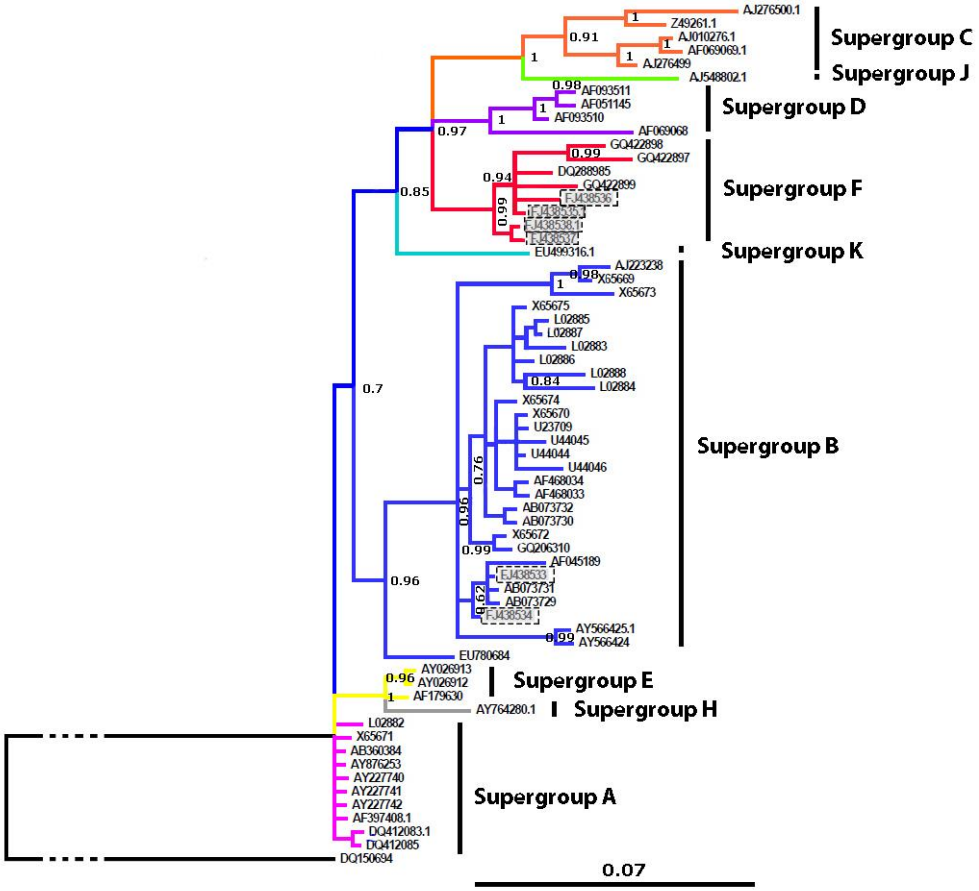
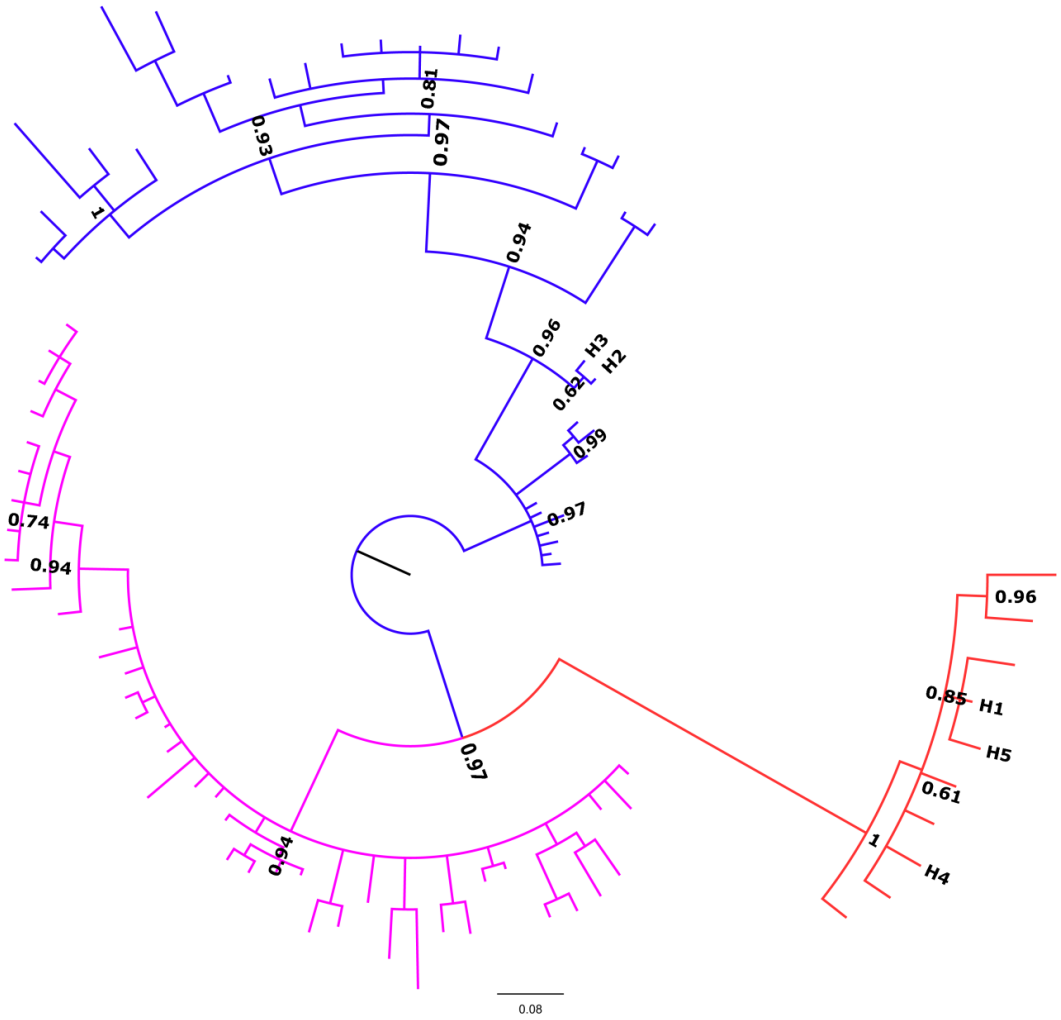


Fig. S1: (Online colour figure) Bayesian phylogenetic tree based on *16S rRNA* gene. Outgroup: *E. coli*. *Wolbachia* infection in *C. parallelus* is shaded is framed grey.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



819

820 Fig. S2: Unrooted phylogenetic tree of *fbpA*, obtained by Bayesian inference. Alleles described in *C.*
821 *parallelus* appear named H1 to H5. Posterior probabilities are shown in the nodes. The color code encodes the
822 supergroup A (pink), B (blue), D (green), F (red) and H (purple). Posterior probabilities are shown in the nodes.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

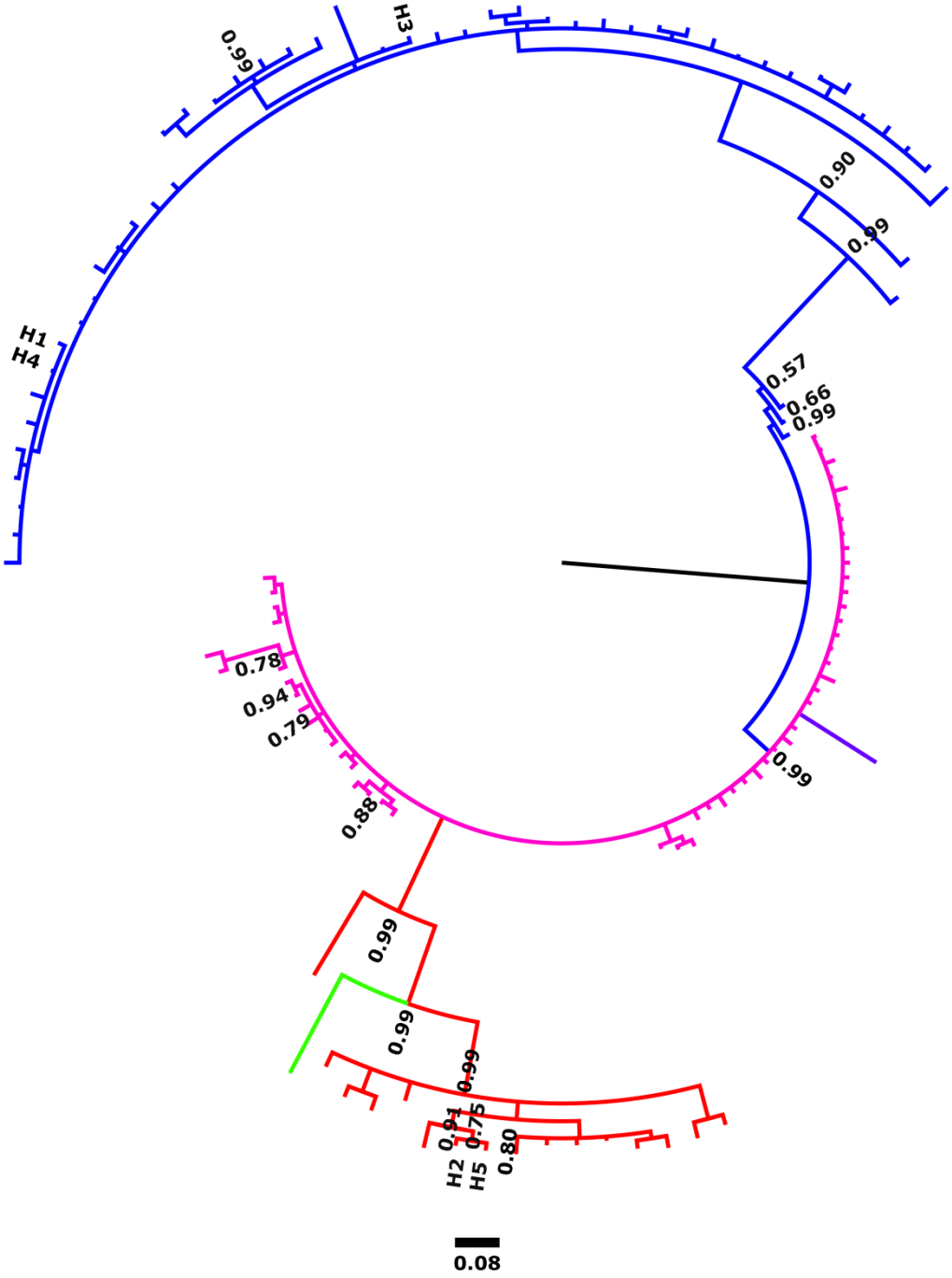


Fig. S3: Unrooted phylogenetic tree of *ftsZ*, obtained by Bayesian inference. Alleles described in *C. parallelus* appear named H1 to H5. Posterior probabilities are shown in the nodes. The color code encodes the supergroup A (pink), B (blue), D (green), F (red) and H (purple). Posterior probabilities are shown in the nodes.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

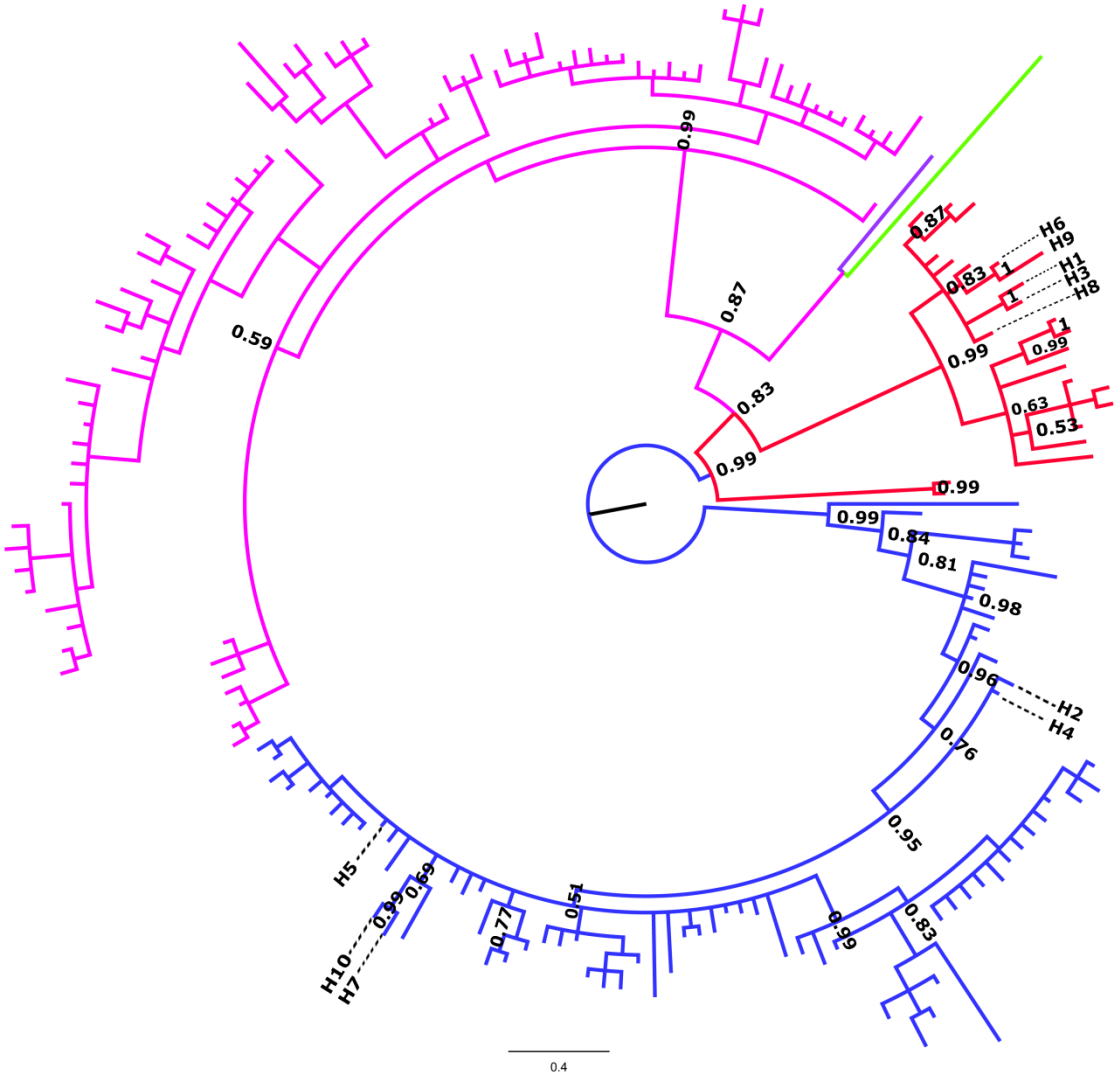


Fig. S4: Unrooted phylogenetic tree of *hcpA*, obtained by Bayesian inference. Alleles described in *C. parallelus* appear named H1 to H10. Posterior probabilities are shown in the nodes. The color code encodes the supergroup A (pink), B (blue), D (green), F (red) and H (purple). Posterior probabilities are shown in the nodes.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

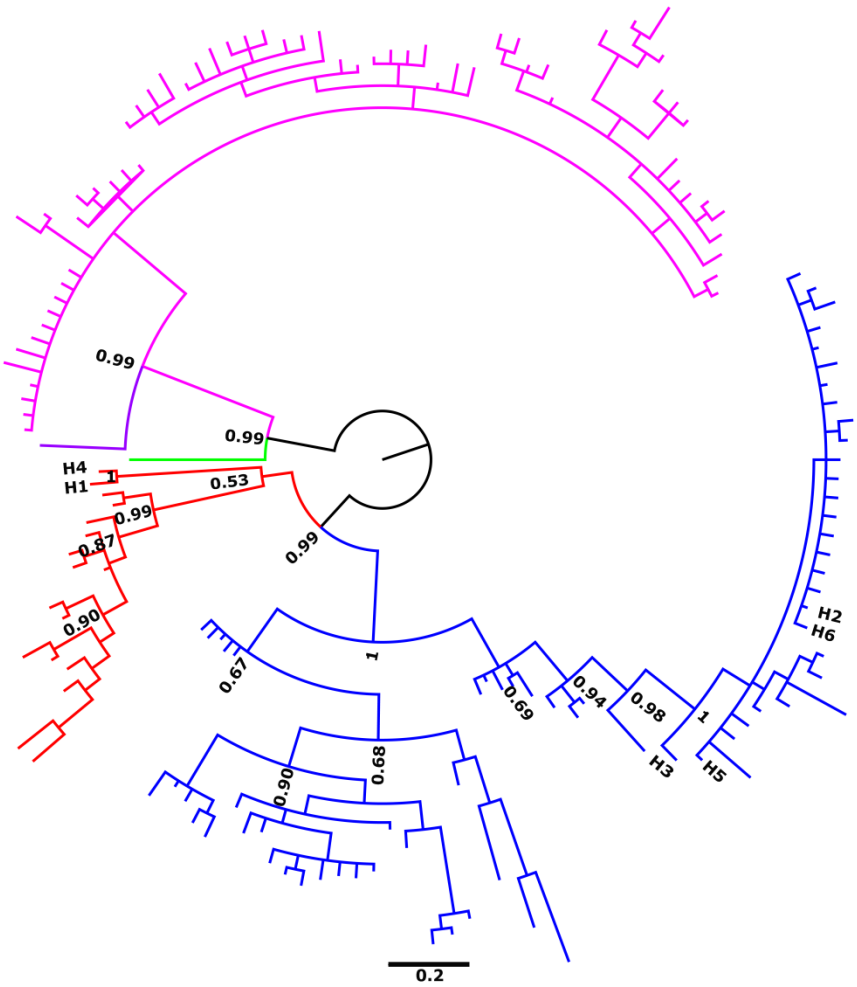
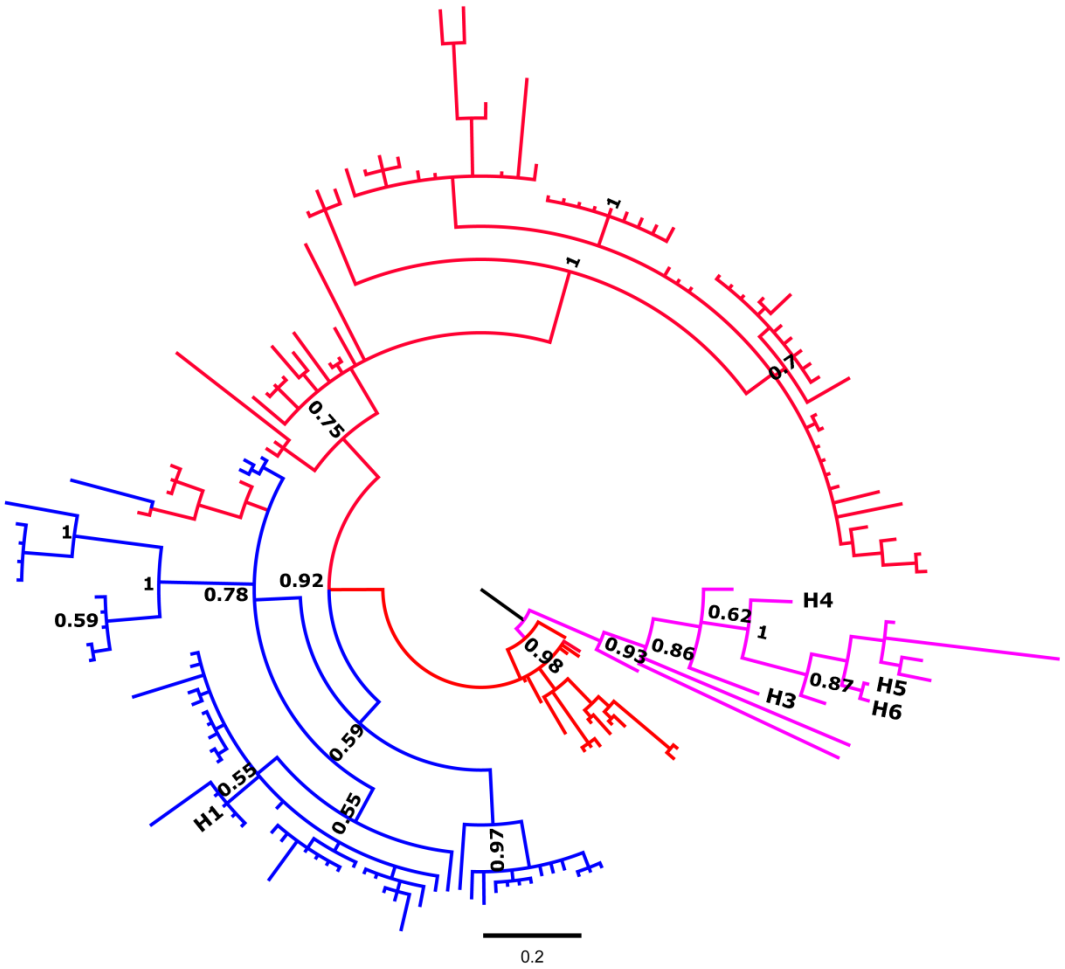


Fig. S5: Unrooted phylogenetic tree of *coxA*, obtained by Bayesian inference. Alleles described in *C. parallelus* appear named H1 to H6. Posterior probabilities are shown in the nodes. The color code encodes the supergroup A (pink), B (blue), D (green), F (red) and H (purple). Posterior probabilities are shown in the nodes.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

843

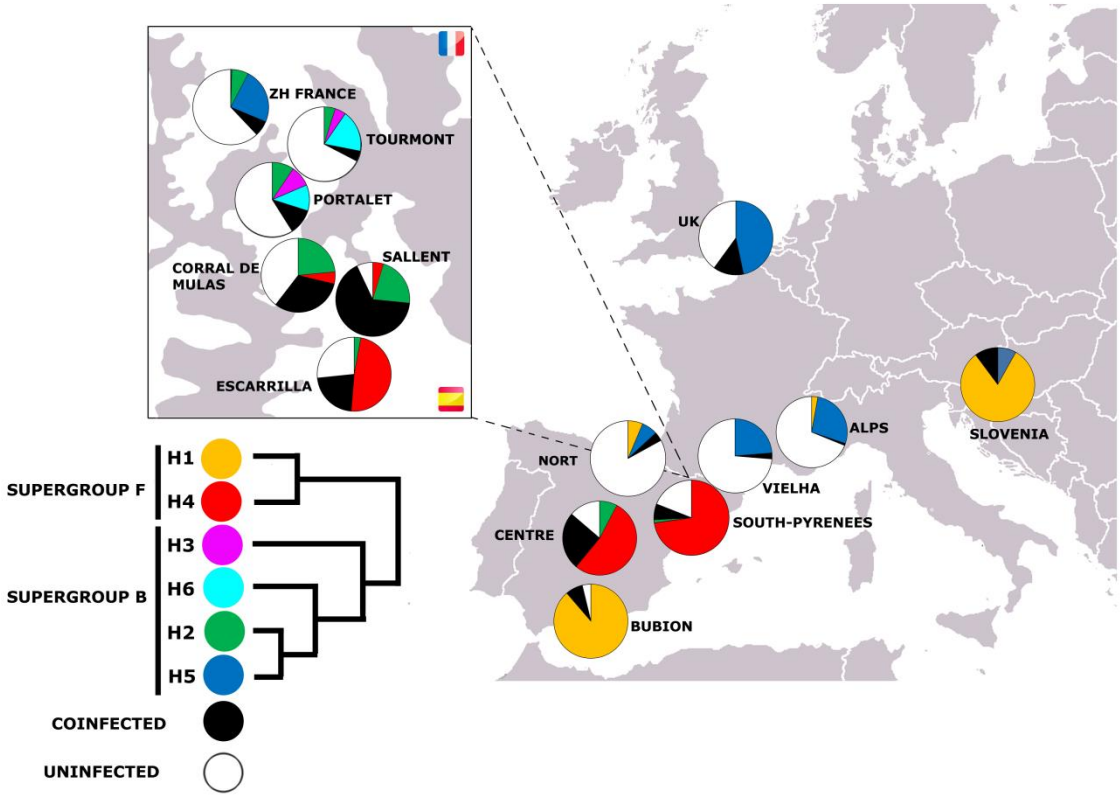
844



845

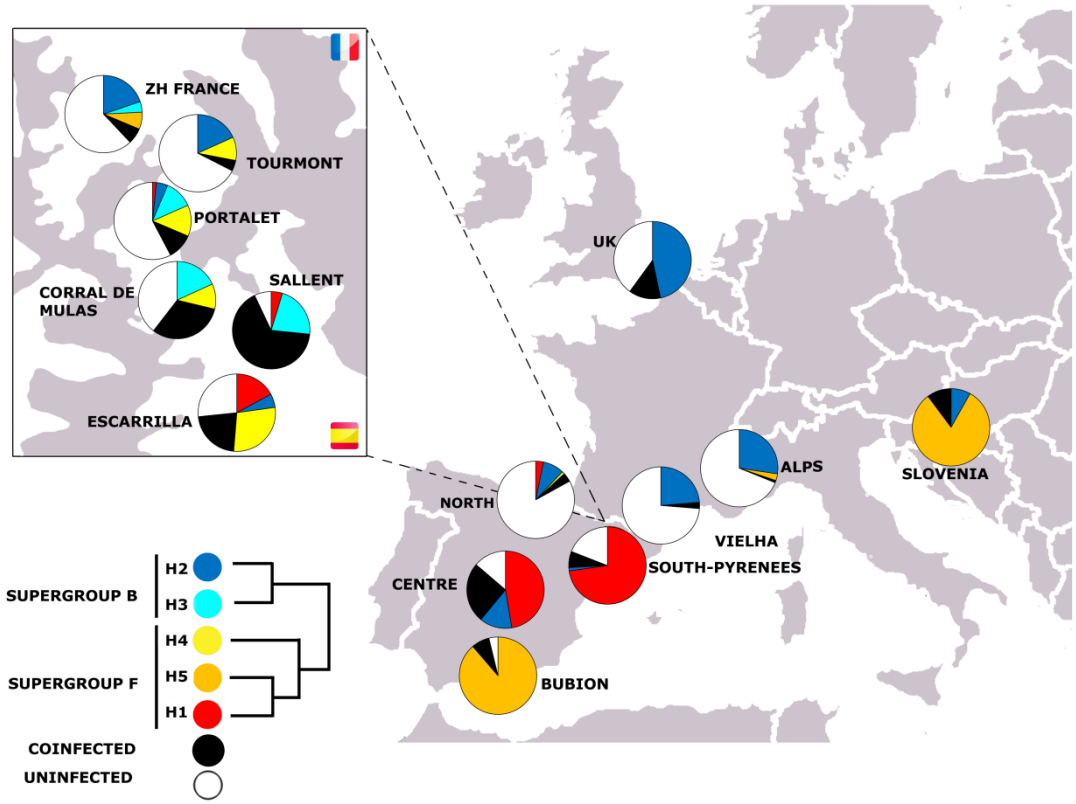
846 Fig. S6: Unrooted phylogenetic tree of *wsp*, obtained by Bayesian inference. Alleles described in *C.*
847 *parallelus* appear named H1 to H16. Posterior probabilities are shown in the nodes. The color code encodes the
848 supergroup A (pink), B (blue), D (green), F (red) and H (purple). Posterior probabilities are shown in the nodes.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



849

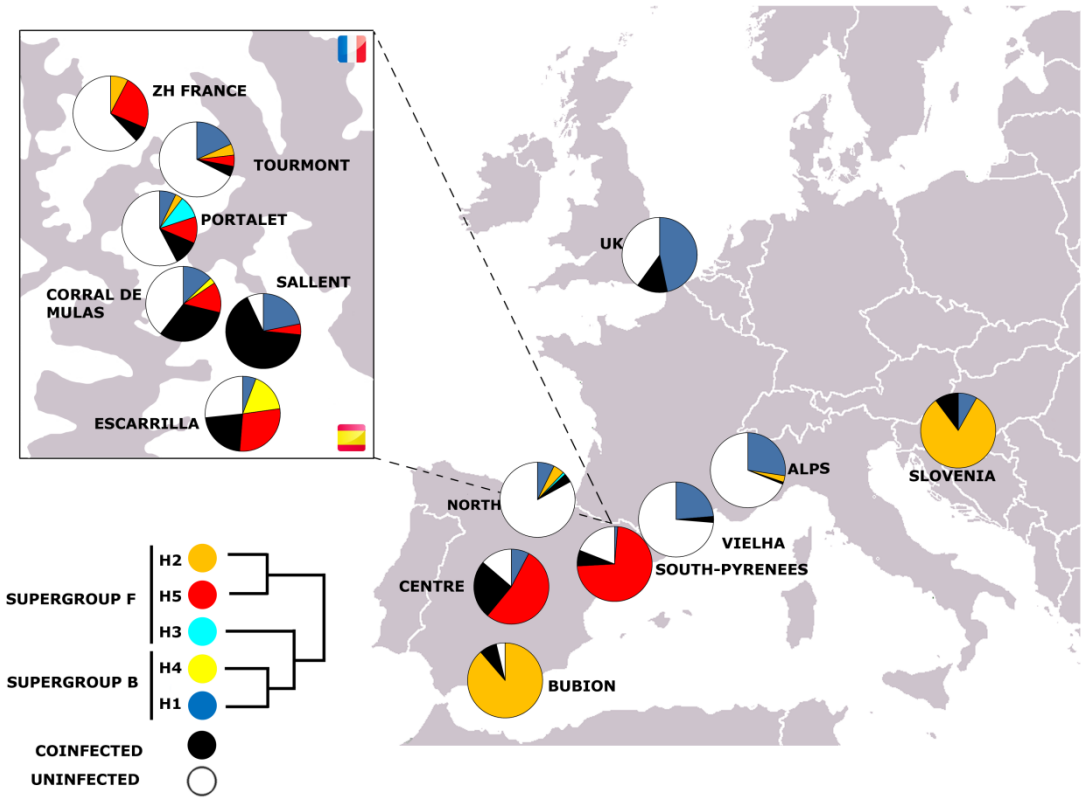
850 Fig S7: Geographical distribution of alleles detected for gene *coxA*



851

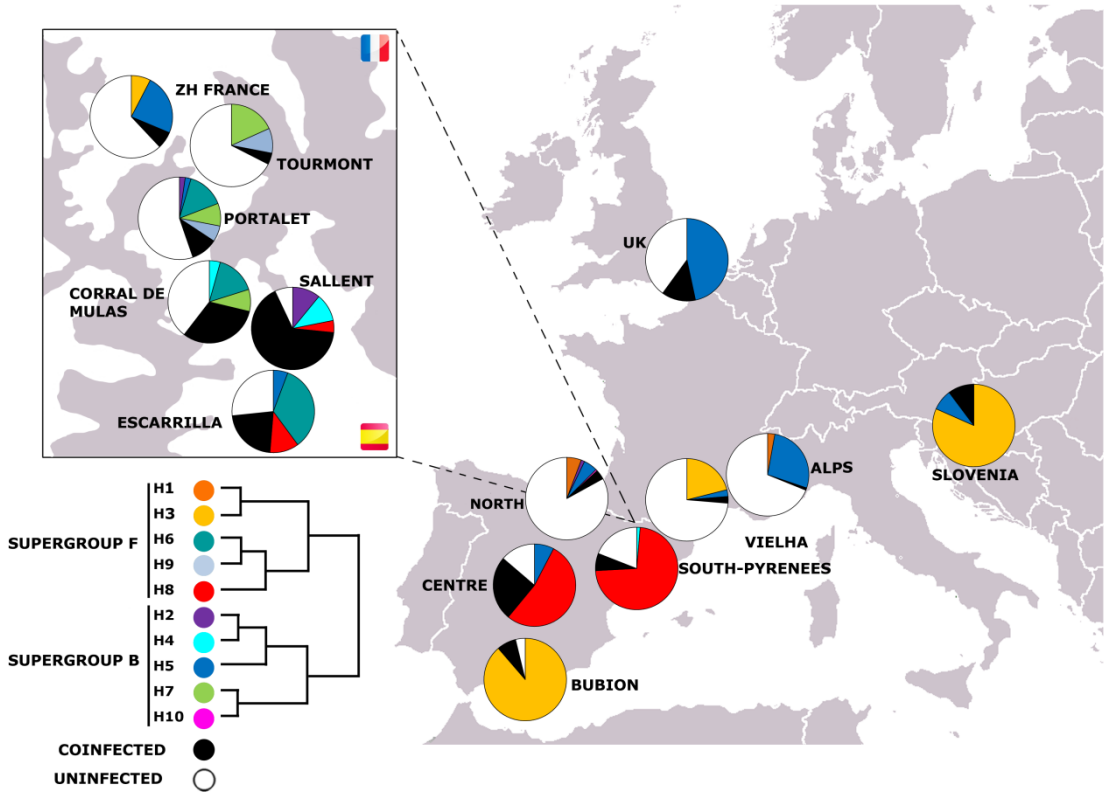
852 Fig S8: Geographical distribution of alleles detected for gene *fbpA*

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



853

854 Fig S9: Geographical distribution of alleles detected for gene *ftsZ*

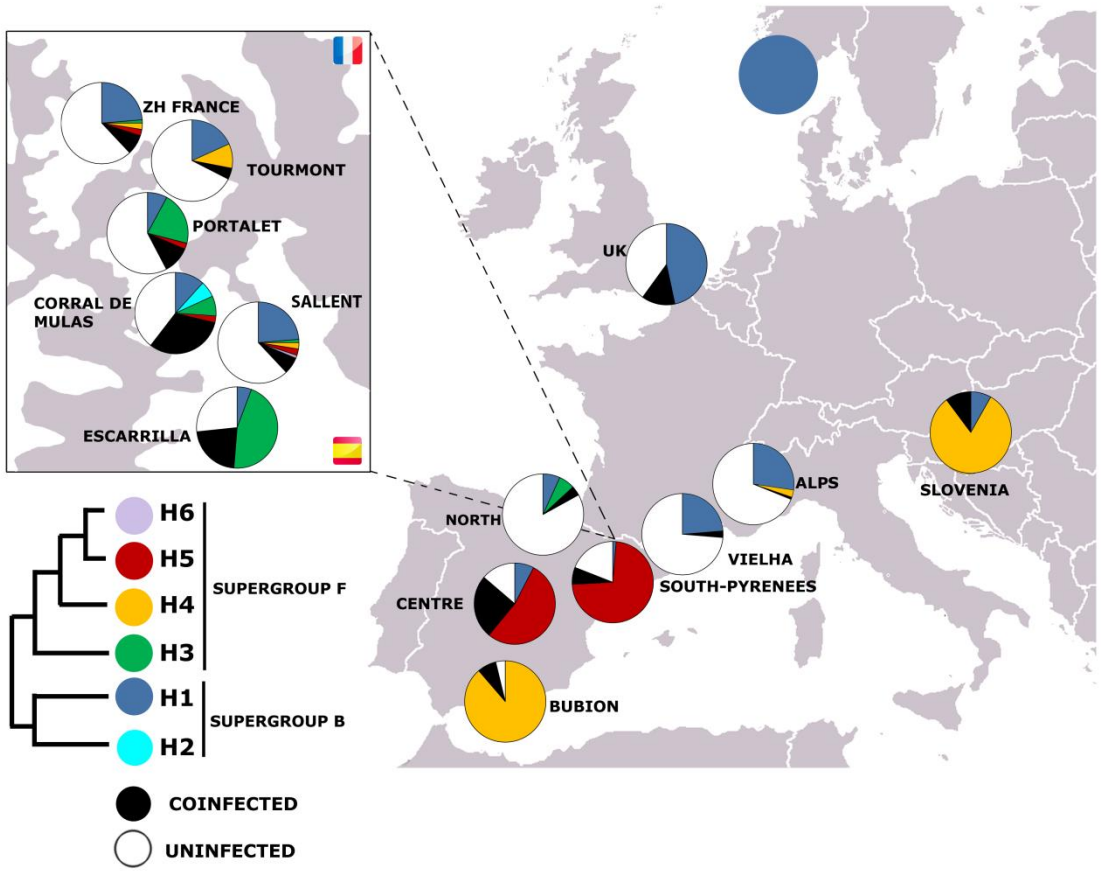


855

856 Fig S10: Geographical distribution of alleles detected for gene *hcpA*

857

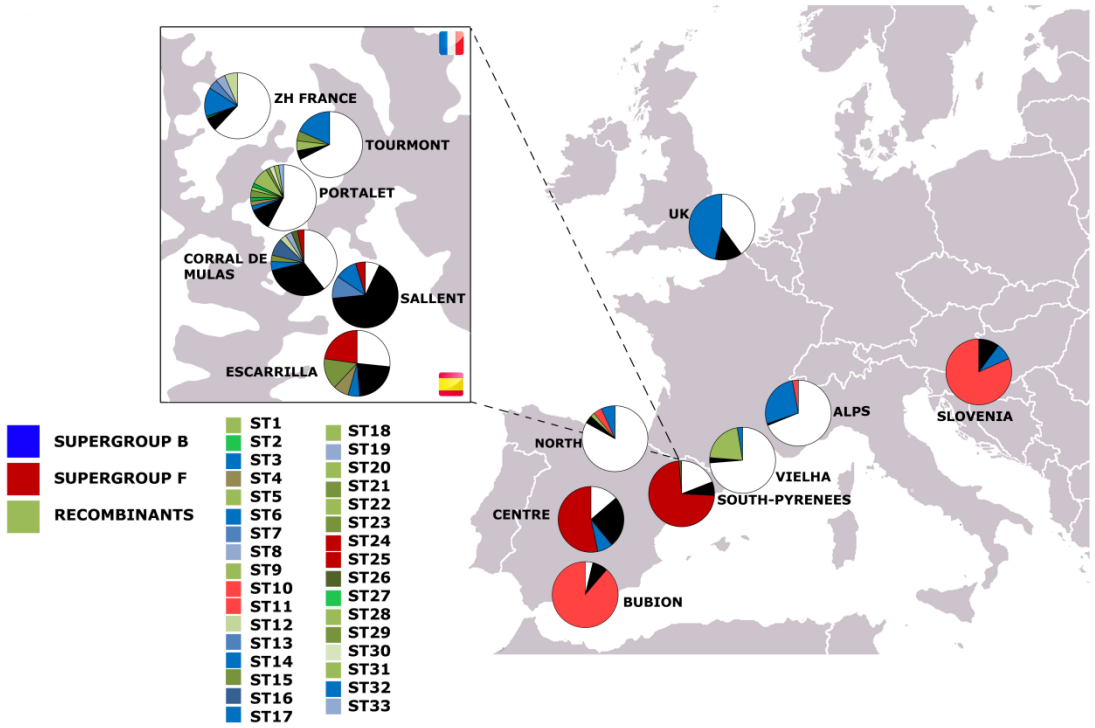
bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



858

859 Fig S11: Geographical distribution of alleles detected for gene *wsp*

860



861

862 Fig S12: Geographical distribution of ST.

bioRxiv preprint doi: https://doi.org/10.1101/044784; this version posted March 19, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplemental references.

Fu Y, Li W (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693-709.

Nei M (1987) *Molecular Evolutionary Genetics.* Columbia Univ. Press, NY.

Neter J, WassermanW, Whitmore GA (1978) *Applied statistics.* Allyn and Bacon. Boston, MA.

Rousset F, Vautrin D, Solinag M (1992) Molecular identification of *Wolbachia*, the agent of cytoplasmic incompatibility in *Drosophila simulans*, and variability in relation with host mitochondrial types. *Proceedings of the Royal Society of London Series B-biological Sciences* **247**, 163-168.

Tajima F (1989) Statistical method for testing the neutral mutationhypothesis by DNA polymorphism. *Genetics* **123**, 585-595.

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–276,