How Long Does Wolbachia Remain on Board?

Marc Bailly-Bechet, Patricia Martins-Simões, Gergely Szöllősi, Gladys Mialdea, Marie-France Sagot, Sylvain Charlat

This submission is intended as an Article,

for the "Discoveries" section of MBE.

How long does *Wolbachia* remain on board?

Marc Bailly-Bechet[1*], Patricia Martins- Simões[1,2,3*], Gergely Szöllősi[4], Gladys Mialdea[1], Marie-France Sagot[1,2] and Sylvain Charlat[1**]

1. Laboratoire de Biométrie & Biologie Evolutive - CNRS - Université Lyon 1 - Bat. Mendel, 43 boulevard du 11 novembre - 69622 Villeurbanne – France
2. Erable Team, INRIA Grenoble Rhône-Alpes, France.
3. Current address: Centre International de Recherche en Infectiologie (CIRI), INSERM U1111 – CNRS UMR5308, Equipe Pathogénie des Staphylocoques, Faculté de médecine Laennec, 7 rue Guillaume Paradin, Lyon, France
4. ELTE-MTA "Lendület" Evolutionary Genomics Research Group, Pàzmàny P. stny. 1A., 1117 Budapest, Hungary

* equally contributing authors

** Corresponding author E-mail: sylvain.charlat@univ-lyon1.fr

Abstract

Wolbachia bacteria infect about half of all arthropods, with diverse and extreme consequences ranging from sex-ratio distortion and mating incompatibilities to protection against viruses. These phenotypic effects, combined with efficient vertical transmission from mothers to offspring, satisfactorily explain the invasion dynamics of *Wolbachia within* species. However, beyond the species level, the lack of congruence between the host and symbiont phylogenetic trees indicates that *Wolbachia* horizontal transfers and extinctions do happen and underlie its global distribution. But how often do they occur? And has the *Wolbachia* pandemic reached its equilibrium? Here we address these questions by inferring recent acquisition / loss events from the distribution of *Wolbachia* lineages across the mitochondrial DNA tree of 3,600 arthropod specimens, spanning 1,100 species from Tahiti and surrounding islands. We show that most events occurred within the last million years, but are likely attributable to individual level variation (e.g. imperfect maternal transmission) rather than population level variation (e.g. *Wolbachia* extinction). At the population level, we estimate that mitochondria typically accumulate 4.7% substitutions per site during an infected episode, and 7.1% substitutions per site during the uninfected phase. Using a Bayesian time calibration of the mitochondrial tree, these numbers translate into infected and uninfected phases of approximately 7 and 9 million years. Infected species thus lose *Wolbachia* slightly more often than uninfected species acquire it, supporting the view that its present incidence, estimated here slightly below 0.5, represents an epidemiological equilibrium.

Key words: *Wolbachia*, Arthropods, Symbiosis, Horizontal transfer, Evolutionary Dynamics.

Short tittle: *Wolbachia* evolutionary dynamics

Introduction

Among the many bacterial lineages inhabiting the cytoplasm of animal cells, *Wolbachia* appears to be the most widely distributed, being present in about half of all arthropod species (Weinert et al. 2015). This patent evolutionary success relies in part on what *Wolbachia* does to its host (Werren et al. 2008). It can sterilize uninfected females (and thus benefit the infected lineage), reallocate reproductive efforts into females at the expense of males (that do not transmit the infection anyway), or protect against natural enemies and thus indiscriminately benefit individuals of both sexes (Martinez et al. 2014). These sophisticated strategies explain how *Wolbachia* can invade a population once it has made its way into at least one individual, but tell us little about the forces that govern its global distribution across the globe and the arthropod phylum. At such a large scale, the dynamics of *Wolbachia* are best seen as an epidemiological process, driven by the ability of these bacteria to jump into new host lineages before they get extinct. Although the importance of horizontal transfer and extinction rates is acknowledged by theory (Werren and Windsor 2000; Engelstädter and Hurst 2006; Zug et al. 2012), empirical information on these parameters is scarce. Many case studies have demonstrated horizontal transfers (Heath et al. 1999; Vavre et al. 1999; Huigens et al. 2000; Sintupachee et al. 2006; Raychoudhury et al. 2009; Ahmed et al. 2015; Brown and Lloyd 2015; Ahmed et al. 2016), some of which have documented possible routes of transmission, but the rate at which *Wolbachia* infections are acquired or the average duration of an infection within a lineage has not been estimated so far.

With some exceptions (Raychoudhury et al. 2009; Hamm et al. 2014), even closely related host species often have a different infection status (one species being infected but not the other) or harbour very divergent *Wolbachia* strains, suggesting a high turnover of infections. For this reason, only comparisons among closely related lineages, within species

or among sister species, will be informative to assess how divergence among hosts affects the probability of sharing an ancestral infection status, and efforts to estimate the extinction and acquisition rates must focus on this micro-evolutionary timescale. With this rationale in mind, we collected over 10,000 arthropod specimens, spanning an estimated 1,110 species, on four islands of the Society archipelago in the South Pacific. These volcanic islands emerged within the last 3 million years, as the Pacific Plate moved toward North-West over a unique hot spot (Guillou et al. 2005), each new island being in part colonised by migrants from its near and slightly older neighbours (Gillespie et al. 2008). Such stepping stone dispersal tends to produce recent splits between closely related but isolated lineages, offering the right focus to assess how variations in infection status among lineages have accumulated over the last few million years.

We compared the mitochondrial and *Wolbachia* phylogenies to infer recent events of infection loss and acquisition. Using mitochondrial branch length as a proxy for time, we show that the global rate of *Wolbachia* loss is 1.5 times higher than the rate of acquisition, so that an epidemiological equilibrium should be reached when 40% of the species are infected, neatly matching the incidence actually observed in this dataset. On average, the host mitochondria accumulate 4.7% substitutions per site during an episode of infection, and 7.1% substitutions per site during an uninfected phase. In a time-calibrated mitochondrial tree relying on a compilation of recent molecular clock studies (Pohl et al. 2009; Jansen et al. 2010; Obbard et al. 2012; Sota et al. 2013; Zhang and Maddison 2013) these numbers translate into 0.14 loss events and 0.11 acquisition events per million years.

Results

Morphological characterisation of 10,929 specimens suggested we had collected a little more than one thousand species, which was confirmed by DNA barcoding (sequencing of a standard portion of the CO1 mitochondrial gene) of 3,627 specimens that clustered into 1,110 Operational Taxonomic Units (OTUs, that is, species level molecular clusters). Details on the sampling procedures and taxonomical diversity of the specimens were presented elsewhere (Ramage et al. 2016) and are summarised in Table 1. *Wolbachia* was detected by PCR in 32% of the barcoded specimens and 40% of the OTUs (as summarised in Table 1, and presented in details in Tables S1 and S2). Sequencing of the *fbpA* gene (the most rapidly evolving of the five *Wolbachia* MLST genes; Baldo et al. 2006) provided an informative phylogenetic marker for 768 of the 1,146 infected specimens, spanning 293 of the 443 infected OTUs (Table S1, fig. S1).

How old are Wolbachia infections?

The host and symbiont molecular data provide indirect means to infer the history of their associations: while stable symbiosis should produce perfectly congruent phylogenies, infection loss and horizontal transfers produce different trees for hosts and symbionts. Cophylogenetic methods aim at using this information to trace back the history of the symbionts along the host tree. This task is however complicated by the presence of phylogenetic uncertainty and is particularly difficult to achieve for large trees, especially when loss and acquisition events are frequent. Rather than relying on a single best scenario of *Wolbachia* loss and acquisition, we thus aimed at sampling the diversity of plausible scenarios

supported by the sequence data. To this end we employed the Amalgamated Likelihood Estimation (ALE) software package (Szöllősi, Rosikiewicz, et al. 2013; Szöllősi, Tannier, et al. 2013) to produce not only the most likely loss / acquisition scenario as an output, but also a population of 1,000 scenarios, sampled according to their likelihood. Fig. S2 summarises these 1,000 scenarios, that is, the estimated probability of loss and acquisition events mapped on each branch of the host CO1 tree. The number of loss events required to reconcile the host and symbiont trees varied from 156 to 288 across the sampled scenarios (median 225), and the number of acquisitions from 206 to 242 (median 227). We used the ALE output to compute the distribution of the age of present day infections (fig. 1), taking the CO1 branch length as a proxy for time (and thus not correcting at that stage for variations in substitution rates along the arthropod tree). This analysis indicates that most infections are very recent, so that the associated mitochondrial DNA lineages have accumulated less than 1% substitution per site since the present day infection was acquired.

Quantifying the Wolbachia turnover

The number of loss and acquisition events per time unit can be modelled under Poisson point processes. We used such models and initially assumed, for the sake of simplicity, that the rates of acquisition and loss, hereafter denoted by β and γ, were homogeneous across the entire arthropod tree. Under such a model, following any loss event placed on the host tree, the probability that no acquisition has occurred after a time t should be an exponential function of t, decreasing with rate β. The same applies to the probability of no loss occurring after an acquisition event, with rate γ. We used this rationale to fit our data, and thus estimate β and γ. Specifically, for any duration t starting from a loss event, we computed the proportion of cases where no acquisition occurred (fig. 2a) and fitted $P = e^{-\beta t}$ to these data, to estimate

the acquisition rate. We proceeded similarly to estimate the loss rate (fig. 2b). This first analysis (where we assumed β and γ are homogeneous across the entire arthropod tree) resulted in a very poor fit of the model to the data, suggesting the data do not follow a single Poisson process. Indeed, on a short time scale, many more events occur than expected under this model (in fig. 2a and 2b, the left part of the full line is well below the dotted line), indicating a particularly high rate in the most recent period. In contrast, many fewer events occur than expected on a long time scale (in fig. 2a and 2b, the right part of the full line is well above the dotted line), indicating a lower "long term rate". We interpret this discrepancy as signal for a previously described phenomenon (Ho et al. 2005; Penny 2005) where non-neutral evolutionary events occur at different rates at the individual and population levels. For example, the rate of mutations is higher than the rate of substitutions (i.e. the number of mutations fixed in populations per time unit) because many deleterious mutations are lost. Similarly, in the context of *Wolbachia* infections, the rate at which new uninfected individuals are produced because of imperfect maternal transmission should be higher than the extinction rate, at which *Wolbachia* is lost from the entire population. This is because *Wolbachia* can be maintained by selection despite the constant production of uninfected individuals. In order to remove the short-term individual effects (producing polymorphism in infection status within populations) from the inference of the long-term population-level rates that are our focus, we modelled infection gain and loss as the sum of two processes. We fitted a sum of exponentials to the data, i.e., the result of two Poisson processes with different rates, one describing the signal occurring at the tips of the tree, that may be attributable to short-term individual events, while the other captures the long-term behaviour at the population level. In the following analysis, we will only report on the long-term (population) rates β_p and γ_p (for the short term rates are irrelevant to the global *Wolbachia* dynamics, and also less accurately estimated because they depend on the shortest branches of the CO1 tree, many of which carry 0

substitutions). Summing over all scenarios produced by the cophylogeny analysis, we estimate that $\beta_p = 0.14$ and $\gamma_p = 0.21$; in other words, *Wolbachia* is acquired on average 0.14 times and lost 0.21 times in the time it takes for CO1 to accumulate 1% divergence. Reciprocally, mitochondria typically accumulate 0.01/0.21 = 4.7% substitutions per site during an infected phase, and 0.01/0.14 = 7.1% during an uninfected phase.

Beyond these summary numbers that are based on the compilation of 1,000 plausible scenarios of losses and acquisitions, we estimated the range of plausible rates by analysing each scenario separately. We observed only limited variation in the estimated rates (fig. 3). Our estimate of β_p falls between 0.128 and 0.16 (per lineage per 1% CO1 distance) in 50% of the scenarios, and γ_p falls between 0.188 and 0.224.

Has Wolbachia reached its equilibrium incidence?

Under a simple epidemiological model, where all species are equally permissive to *Wolbachia*, and rates of extinction and acquisition are homogeneous across arthropod clades, we can use our estimates to predict the incidence of *Wolbachia* at equilibrium, that is, the proportion of infected species that should be reached when new *Wolbachia* acquisitions are balanced by extinctions. Having defined β_p as the rate at which uninfected species acquire *Wolbachia* per time unit (the "force of infection" in standard epidemiological terms), and γ_p as the rate at which infected species lose *Wolbachia*, a stable proportion should be reached when the total number of acquisitions and extinctions per time unit are equal, that is, when $I \times \gamma_p = U \times \beta_p$, where U and I denote the proportion of uninfected and infected species, respectively. The equilibrium should thus be reached when $\frac{I}{U} = \frac{\beta_p}{\gamma_p}$, that is (since $U + I = 1$), when $I = \frac{\beta_p}{\gamma_p + \beta_p}$. In fig. 4, we show the predicted density of this equilibrium incidence,

based on the 1,000 plausible scenarios. The maximum density strikingly matches the *Wolbachia* incidence that is actually observed in our data set. We emphasize that the equilibrium between acquisition and loss is not a hypothesis of the co-phylogeny analysis, meaning this surprising concordance is not a circular result, imposed by the analysis. In combination with the remarkable stability of the *Wolbachia* incidence across the globe (Werren et al. 1995; Werren and Windsor 2000), this result provides support for the conjecture that *Wolbachia* has reached its equilibrium incidence.

Time calibration and comparison between orders

Substantial variations in mitochondrial substitution rates occur throughout the arthropod tree (e.g. see Johnson et al. 2003; Raychoudhury et al. 2009; Obbard et al. 2012; Sota et al. 2013), but a relaxed molecular clock approach can be used to produce a time-proportional tree and thus correct at least partially for these variations. Calibration points (that is, events dated from external information) can then be used to translate branch length into absolute time. We performed such an analysis to estimate the average number of *Wolbachia* extinctions and acquisitions occurring per million years. Because of computational constraints, this required to split the analysis in 5 subtrees, each including one recent calibration point estimated from earlier molecular dating studies (Pohl et al. 2009; Jansen et al. 2010; Obbard et al. 2012; Sota et al. 2013; Zhang and Maddison 2013) (Table S3, fig. S3). As expected, the substitution rates inferred from this analysis substantially vary within and across orders, around a mean of about 1% substitutions per site per million years (fig. S4). Applying the above described double Poisson model to the time-calibrated trees, we estimate across the 1,000 ALE scenarios that uninfected lineages acquire *Wolbachia* every 9.3 million years (6 to 13.3 for 95% of the scenarios), while infected lineages lose their infection every 7 million years (5.2 to

9.6 for 95% of the scenarios). Notably, these durations are larger than the age of the islands under study, suggesting that a large part of the informative variation in infection status does not stem from recent island-related isolation events.

We can use the time-calibrated trees to assess the possibility of differences between arthropod clades in the *Wolbachia* dynamics, correcting for the potentially confounding effect of variation in CO1 substitution rates among clades. We thus estimated clade specific extinction and acquisition rates for arthropod orders represented by at least 50 species (fig. 5). Although uncertainties in time calibration call for a cautious interpretation of these numbers, we observe marked contrasts between clades. Extinction rates appear larger than the global values in Lepidoptera and Coleoptera but lower in Hymenoptera and Aranea. Acquisition rates are high in Diptera, Lepidoptera and Hemiptera, but low in Coleoptera, Hymenoptera, Aranea (suggesting parasitic and predatory lifestyles do not predispose to frequent acquisitions of *Wolbachia*).

Discussion

This study represents the first attempt to quantify *Wolbachia* dynamics at the global scale of arthropods, that is, to estimate the rate at which infections are acquired and lost, and the average duration of an infection lifetime within a host species. At the population level, we estimate that mitochondria typically accumulate 4.7% substitutions per site during an infected phase, and 7.1% during an uninfected phase. Under a relaxed molecular clock model, these numbers translate into infected and uninfected phases of approximately 7 and 9 million years. Under a simple epidemiological model, where we assume a constant force of infection, we expect that 40% of the species should be infected at equilibrium. This prediction matches the incidence observed in our dataset, suggesting the stationary state has indeed been reached, in accordance with the observed stability of *Wolbachia* incidence across wide geographic scales, documented by Werren and Windsor (2000). Notably, these authors were also the first to propose that the rates of *Wolbachia* extinction and acquisition should be related to its global incidence through some epidemiological process. However, while they relied on the equilibrium hypothesis to derive an estimate of the relative extinction / acquisition rate, here we estimated independently absolute values for the loss and acquisition rates and used these values to test (and validate) the equilibrium hypothesis.

Because of its large sample size and broad phylogenetic spectrum, this study also involved some inherent approximations and limitations that must be addressed. On the symbiont side, the fact that we estimated global values for loss and acquisition rates should not mask the possibility that some *Wolbachia* lineages might show particular dynamics. We did not detect such variations between the A and B *Wolbachia* supergroups, that are sufficiently well represented in the dataset to allow for a separate analysis (fig. S5) but this does not rule out the possibility of finer scale variations. *Wolbachia* strains that tend to occur

within multiple infections might also be more stable or more prone to horizontal transmission than the single infections on which the present study is based. Extending the analysis to a subset of the specimens through massive parallel sequencing would provide a means to assess if multiple infections have particular dynamics, and beyond the *Wolbachia* genus, to investigate potential interactions with other common maternally inherited symbionts of arthropods. The use of a single *Wolbachia* locus, a fraction of the fbpA gene, to characterise its flux across lineages, also sets some limitations to our analysis. The substitution rate in the *Wolbachia* genome is by far lower than in mitochondria (Raychoudhury et al. 2009; Richardson et al. 2012), which bounds our ability to detect horizontal transfers between closely related hosts. These might indeed go undetected if they have not been followed by mutations in the sequenced region. The use of a single marker also masks the potentially confounding effect of recombination among *Wolbachia* genomes. Although recombination is not a rare event in the *Wolbachia* history (Jiggins et al. 2001; Werren and Bartos 2001; Baldo et al. 2006), we think this has a limited confounding effect on our estimates, because most (precisely, 90%) of the *Wolbachia* acquisitions appear to occur in uninfected branches, and thus cannot be accounted for by recombination. Both of these issues could be addressed by extending the sequencing efforts to more loci, possibly the few housekeeping genes used for Multi Locus Strain Typing in *Wolbachia* (Baldo et al. 2006), but also ideally to fast evolving markers, such as mobile genetic elements, providing a phylogenetic signal on very short timescales.

On the host side, our analysis relies on DNA barcoding, which has many advantages (notably, high mutation rates and reduced effective population size, making this marker informative on short time scales), but also carries its negative aspects. Notably, the evolutionary history of mitochondria, because they do not recombine and are genetically linked to invasive elements such as *Wolbachia* itself, might more often than other loci deviate

285 from the demographic history, due to introgression or incomplete lineage sorting (Hurst and Jiggins 2005). In the context of the present analysis, however, this is not a drawback, as our aim is to estimate the rate at which *Wolbachia* jump in and out of their maternal lineage, for which mitochondria are the appropriate marker, in contrast to bi-parentally inherited nuclear genes. Potentially more problematic is the fact that the CO1-based mitochondrial tree is
290 uncertain. The CO1 gene is a rapidly evolving marker, providing good phylogenetic signal on a short timescale, but virtually uninformative for deep nodes, because of saturation. The topology of the mitochondrial tree, as well as branch length, could be better estimated by integrating the CO1 sequences of other lineages (to break long branches), and also the phylogenetic signal from nuclear housekeeping genes (to resolve the deep parts of the tree).
295 Although such improvements would certainly eliminate some of the noise in our analysis, we argue that the uncertainty in the deep nodes of the mitochondrial tree does not represent a significant concern for our estimations. Indeed, 95% of the loss and acquisition events inferred in our analysis occur at the very surface of the tree (within a distance of less than 14% substitutions per site), that is, where the CO1 phylogenetic signal is strong. Translation
300 of CO1 branch length into absolute time also represents a source of uncertainty, when it comes to estimate rates of events per million years. One possible avenue to improve the time-calibration of the CO1 tree would be to take advantage of the geological history of the archipelago to directly identify calibration points in this dataset. Finally, one should keep in mind that our estimates are based on an island microcosm, which might carry its peculiarities.
305 The co-phylogeny analysis also comprises its strengths and weaknesses. The ALE programme presents important differences compared to others usually used in the field of host / symbionts interactions (Conow et al. 2010; Merkle et al. 2010). Importantly, it takes into account the uncertainty in the symbiont tree, and thus does not infer spurious events of infection loss or acquisition in poorly resolved regions of the symbiont tree. However, the

current version of ALE runs with a single host tree, which also has an uncertain topology, as detailed above. Taking into account this side of the uncertainty could be done, at least in principle, through sampling of many plausible host trees, following a Bayesian phylogenetic inference. However, this approach is computationally inefficient, so that alternative solutions should be sought. Another important advantage of ALE is that it allows transfer from non-sampled or extinct specimens, thus relaxing a heavy and unrealistic assumption. This programme also adjusts the loss and acquisition rates by maximum likelihood, so that these values do not have to be known before the analysis. All these improvements come at a computational cost that required the analysis to be split in three sub-trees analysed independently. While this does not affect our ability to infer acquisitions of *Wolbachia* at the right place in the host tree, it hinders the detection of transfer sources: some distant branches of the host tree might be ideal source candidates, but cannot be identified as such if they are not included in the analysis. Investigating more specifically the patterns of horizontal transfer, and the contribution of phylogenetic distance or ecological connections to this phenomenon, will thus require additional methodological developments.

Our analysis revealed that the assumption of homogeneous rates of loss and acquisition along the arthropod tree is not tenable. Specifically, we inferred many more recent events and much fewer old events than would be expected under such a model. We interpret this discrepancy as evidence for high rates of individual level events (e.g. imperfect maternal transmission), and lower rates for population level events. This distinction is important and fits the view that infection loss or acquisition, at the individual level, is necessary but not sufficient for the spread of an infection or its extinction at the population level. Numerous infections appear to make it into specimens of other species, but only few of them do spread. This result emphasizes that the spread of an infection into a new host is likely associated with intense adaptive evolution on the *Wolbachia* side. Although horizontal transfer remains a rare

event in the everyday life of *Wolbachia*, it might represent a critical selective pressure, maintaining a high degree of evolvability. The striking genomic plasticity of *Wolbachia* might in part be explained by these intense episodes of selection. Similarly, the everyday loss of *Wolbachia* due to imperfect maternal transmission is not sufficient to explain extinction at the population level. *Wolbachia* extinction might rather result from evolutionary changes in the induced phenotypes, such as suppression of sex-ratio distortion by host factors (Charlat et al. 2007; Vanthournout and Hendrickx 2016) or reduction in the embryonic mortality induced by Cytoplasmic Incompatibility (CI). Notably, the latter can occur even without host suppression because CI is expected to decay by drift within populations (Turelli 1994), so that only the spread into new populations or species maintains CI at high levels in the long run.

The *Wolbachia* extinction and acquisition rates estimated here also shed light on the range of plausible evolutionary consequences of *Wolbachia* infections. In particular, it has been proposed that *Wolbachia* might contribute to increase host speciation rates, by directly reducing gene flow through CI, or more generally by driving local adaptation (Werren 1998). One condition for such effects to significantly affect speciation is their duration. We estimate that *Wolbachia* remains on average for 7 million years within a lineage, which appears by far sufficient to impact speciation rates. The possibility of an effect of *Wolbachia* on speciation rates actually raises an additional possible concern, namely, that such an effect was neglected here when estimating the *Wolbachia* loss and acquisition rates. If *Wolbachia* significantly increase the speciation rates of its hosts, this should translate into denser regions of the CO1 tree in infected clades, which would tend to increase the apparent duration of the association estimated under a Poisson model. Similarly, some possible effects of *Wolbachia* on their host extinction rates would tend to increase the estimated loss rate. Addressing these interesting but complicated issues will require more data and methodological developments.

Our study indicates that most *Wolbachia* infections seen in present day species were acquired recently. The cophylogeny analysis occasionally suggests that some infections might be ancient, but we found no clade where the two trees perfectly match. In other words, *Wolbachia* has never turned to a stable mutualistic symbiont in any of the groups under study. How comes that *Wolbachia* has stabilised in some lineages of nematodes (Comandatore et al. 2013; Lefoulon et al. 2016), but never in arthropods? Two cases are known where *Wolbachia* has become indispensable to its hosts in arthropods: the parasitoid wasp *Asobara tabida*, where uninfected females cannot produce eggs (Dedeine et al. 2001), and the bedbug *Cimex lectularius*, where *Wolbachia* produces the essential B vitamin (Nikoh et al. 2014). If *Wolbachia* can become an essential partner, why do we not see stable and long-term associations? At this stage, we are only left with speculation to answer this question. It might be that host species that have become dependent upon *Wolbachia* are threatened by the ability of these bacteria to play selfish strategies. Indeed, even an essential symbiont would benefit from the additional fitness increase associated with reproductive manipulations such as sex-ratio distortion. In the long run, this might lead to the loss of such associations, either through host extinction, replacement of *Wolbachia* by other symbionts, or simple elimination of the infection if its presence is not vital.

Under this view, the conflicting nature of the *Wolbachia* / host interaction would underlie its brevity. Interestingly, this causal relationship might also work backwards, producing a positive feedback between conflict and instability: the ability of *Wolbachia* to jump into new hosts, and its instability within a host lineage, might fuel the evolution and maintenance of selfish strategies. Beyond *Wolbachia*, the instability of associations underlies the evolution of all selfish genetic elements (that is, vertically inherited elements that can be invasive despite being harmful) (Burt and Trivers 2006). For example, transposable elements or meiotic drivers can only invade populations thanks to sex and recombination that break

associations between genes and thus open the opportunity for efficient selfish strategies. Similarly, the possibility for *Wolbachia* to reach a new and naïve host species through horizontal transfers selects for selfish invasive strategies such as sex ratio distortion or cytoplasmic incompatibility, regardless of any long term detrimental effects on host species. On a long evolutionary scale, *Wolbachia* could thus essentially be regarded as a horizontally transmitted pathogen, fitting the general notion that harmful effects can only evolve and be maintained under horizontal transmission, which uncouples the host and symbionts evolutionary trajectories.

Materials and Methods

The SymbioCode sample

The sample used in this study was obtained as part of the SymbioCode project, designed for investigating the flux of symbionts among branches of the arthropod tree, using in depth sampling in 4 islands of the Society Archipelago in French Polynesia. Details on the sampling procedure have been presented elsewhere (Ramage et al. 2016), as well as taxonomic diversity, which is also summarized in Table 1. In brief, 10,929 arthropod specimens were photographed and sorted into morpho-species following non-taxonomically focused sampling on the islands of Moorea, Tahiti, Raiatea and Huahine (Table S1). DNA was extracted from 4,837 specimens, aiming at the maximum taxonomic and geographic coverage. DNA barcoding (sequencing of a standard portion of the CO1 mitochondrial gene) was attempted on all extracts, with a 75% success rate, yielding molecular data for 3,627 specimens, where the presence of *Wolbachia* was assessed by PCR (see details below). Sequences clustered into 1,110 Operational Taxonomic Units (species-like groups) here defined on the sole basis of mtDNA data, using the Refined Single Linkage algorithm (RESL) implemented in BOLD (Ratnasingham and Hebert 2013). The SymbioCode data were deposited in the BOLD database under dataset id DS-SYMC (URL: dx.doi.org/10.5883/DS-SYMC); the mtDNA sequence data were also deposited in GenBank (BankIt1909431: KX051578 - KX055204), and the alignment is provided as Supplementary Material.

Wolbachia screening and sequencing

Wolbachia infections were screened using the 16S primers and protocols from Simões et al. (2011). The presence of *Wolbachia* DNA in extracts having produced positive 16S amplicon was further confirmed by amplifying the fructose-bisphosphate aldolase gene (*fbpA*) using primers FbpA-F1 and FbpA-R1 (Baldo et al. 2006). PCRs were performed in a total volume of 30 µl with 1.5 mM of MgCl2, 2 mM of all four dNTPS, 0.2 µM of each primer, 0.02 Units/µl EuroTaq R DNA polymerase (EUROBIO, Les Ulis, France) and 2 µl of template. The temperature profile was as follows: initial denaturation at 94°C for 120 seconds (sec); 36 cycles of 94°C for 30 sec, 56°C for 45 sec and 72°C for 90 sec; and a final extension at 72°C for 600 sec. All reactions took place in a Tetrad R Thermocycler (Bio-Rad, Hercules, CA, USA). FbpA PCR products were sanger-sequenced using both the forward and reverse PCR primers. Trace files were analysed in GENEIOUS v5.4.0 (Biomatters) (Kearse et al. 2012) as detailed elsewhere (Ramage et al. 2016) to produce 955 sequences varying in length from 152 to 467 bp. We observed no stop codons, suggesting that none of the sequences are nuclear insertions. Notably, the risk of nuclear insertions was also minimised by the systematic amplification of both 16S and fbpA to test the presence of *Wolbachia*. Sequences were deposited in the BOLD database under dataset id DS-WOLSC (URL: dx.doi.org/10.5883/DS-WOLSC) and in GenBank (BankIt1953308: KX842728-KX843321, KX843323-KX843667). The alignment and tree of the fbpA sequences used in the cophylogenetic analysis are provided as Supplementary Material (fig. S1).

Cophylogeny

We used the ALE program (Szöllősi, Rosikiewicz, et al. 2013; Szöllősi, Tannier, et al. 2013) for the cophylogeny analysis, that is, the inference of *Wolbachia* losses and acquisitions required to resolve the incongruence between host and symbiont trees. This program was

initially designed in the context of gene tree / species tree reconciliation to infer the history of gene loss, duplication and horizontal transfer, through reconciling gene trees with a known species tree. In our case, and hereafter in the text, the "gene" is the symbiont, and the "species tree" is the host tree. We will also neglect "duplication" events, which contribute to the history of genes within genomes, but were never observed in our outputs. In brief, the ALE analysis includes the following steps. The user provides a single, fully bifurcating host tree (not necessarily time-like in the "undated" version of the program that was used here; Szöllősi et al. 2015) and *k* plausible symbiont trees, sampled using a Bayesian phylogenetic inference method (in our case, $k = 5{,}000$). ALE then computes the likelihood of symbiont loss and acquisition scenarios, integrated over the *k* plausible symbiont trees, while estimating maximum likelihood rates of transfer and loss events.

The ALE program presents several features that make it the most appropriate for our analysis. First, by sampling plausible symbiont trees according to their probability, it allows us to account for this source of uncertainty when estimating the likelihood of loss / acquisition scenarios. Second, the relative costs (or rates) of loss and acquisition events are not provided *a priori* by the user but are also estimated by maximum likelihood. Finally, the program does not rely on the unrealistic assumption that all transfer events must come from the sampled part of the host tree. Instead, it allows for transfer from extinct and unrepresented species (Szöllősi, Tannier, et al. 2013).

In our analysis the maximum likelihood host tree was inferred with FastTree (Price et al. 2010) under a general time reversible model with gamma distributed rate variation among sites, constraining the relationships between arthropod orders from the topology of Regier et al. (2010). Notably, even within orders, some nodes are too deep to be inferred with confidence with CO1, which is a fast evolving marker, rapidly reaching saturation. However, 95% of the loss and acquisition events inferred occur at the very surface of the tree (within a

distance of less than 14% substitutions per site), meaning that uncertainty in the ancient nodes will have very mild consequences on our inferences. We excluded from the cophylogeny analysis 120 specimens belonging to arthropod orders represented by fewer than 10 species, because poorly populated clades carry little signal for the inference of loss and acquisition events. We further eliminated 378 specimens that were positive for *Wolbachia* from the PCR assay, but could not be sequenced, either because they were infected by multiple strains, or carried the infection at a very low density. This reduced the size of the host tree from 3,627 to 3,129. Finally, we selected only one representative sequence (the longest one) for each combination of CO1 haplotype and *Wolbachia* infection status, to remove any data that would be redundant for the cophylogeny analysis. This is equivalent to assuming that such situation derived from a single event (either loss or acquisition), thus leading to a conservative estimate in the number of infection losses and acquisitions. This reduced the size of the relevant host tree from 3,129 to 1,679 leaves. This tree was still too large to be analysed in a single ALE run and was thus split in three parts of similar size, with no consequences on our analyses as ALE does not impose that the source of transfers should be inside the tree under study. A specific version of ALE was written for the present analysis, to output not only the maximum likelihood loss / acquisition scenario, but 1,000 scenarios sampled according to their likelihood, in order to assess variation among plausible scenarios.

Time calibration of the host tree

We used BEAST to produce a time-calibrated tree under a relaxed molecular clock model that allows substitution rates to vary across branches. Because of computational constraints, the main CO1 tree was cut in 5 subtrees of similar size for this analysis (as indicated in Fig. S3). In each subtree, the FastTree topology was imposed, so that only branch length was optimised

at that stage. Because our analysis relies on recent events of loss and acquisition, and because CO1 is evolving too fast to date the deep nodes, we used recent calibration points (all younger than 10 million years). Geological records do not provide such recent calibration points; we thus used as calibration points pairs of sequences extracted from earlier studies that focused on molecular dating (table S3) (Pohl et al. 2009; Jansen et al. 2010; Obbard et al. 2012; Sota et al. 2013; Zhang and Maddison 2013). The analysis was run for 30 million generations in Beast 1.6.2 (Drummond and Rambaut 2007), with the following parameters: GTR + G substitution model: empirical base frequencies, 4 categories of Gamma, 2 partitions of codon positions (1+2 vs. 3); relaxed uncorrelated Lognormal clock model; tree prior: coalescent constant size (because recent nodes are best modelled under a coalescent process). Model convergence was checked in Tracer, and the estimated sample size exceeded 100 for all parameters. We used TreeAnnotator to export the median height tree for further analysis. Rather than estimating absolute branch length in Beast, we used this program to produce an ultrametric tree, that is to correct for mutation rate variations and estimate time-proportional branch length. We secondarily used the previously estimated ages of calibration points to translate branch length into absolute time units. To verify that our analysis captured variation in substitution rates across the arthropod tree, we computed the substitution rates in 776 clades made of closely related specimens (with a common ancestor younger than 10 million years), by dividing the sum of the branch length in the PhyML tree by the sum of the branch length in the time-calibrated tree within each clade. The results, summarised in Fig. S4, indicate a median below 3% substitutions per site per million years in all orders, with substantial variation within each order. It is known that substitution rates estimated from very recent branches (polymorphism data) tend to be larger than those inferred from between-species divergence, because slightly deleterious mutations contribute more to polymorphism than divergence (Ho et al. 2005). To assess if such an effect could bias our estimates, we

computed the substitution rate and median branch length in the above-defined 776 clades, and tested the correlation between these two variables, within each order and in the entire dataset. None of the correlation tests were significant, suggesting that variation in the distribution of branch length across clades is unlikely to introduce a bias in our substitution rate estimates.

Distribution of infection ages

We used the output of ALE to compute the age of the currently observed infections in each of the 1,000 plausible loss / acquisition scenario. Closely related infections deriving from the same acquisition event should not be regarded as independent points to estimate the age of an infection. Each point in this analysis thus corresponds to one acquisition event, rather than one infected leaf. When the CO1 distance was used as time unit, the age of an acquisition event was computed as the mean of the CO1 distances between this event (placed on a branch of the host tree) and the infected leaves deriving from this event. When the age was computed from time-calibrated trees, this calculation was more straightforward, since the time elapsed between the acquisition event and the descending infected leaves is by definition the same for all leaves.

Notably, the ALE-undated program neglects branch length in the host tree, and thus maps events on branches without specifying a particular position along the branch. We thus placed the event on the branch randomly, following a Poisson law with slow rate (0.01 event per 1% CO1 substitution). For short branches, this produces a placement similar to what would be obtained with sampling in a uniform law. On the contrary, for long branches, it favours placing the event closer to the daughter branches (where the infection status is known), avoiding a large overestimation of the infection age, which would have occurred if a uniform law had been used. The value 0.01 was chosen to be conservative, as it is lower than

our rate estimates and thus cannot inflate them, but we found that the chosen value has very little effect on our estimations (not shown).

Estimation of loss and acquisition rates

Infection losses and acquisitions were modelled as random events with a constant rate of occurrence per time, i.e. following a Poisson point process. With this rationale, we computed the distribution of the duration of the infected and uninfected states, which we fitted to the data to estimate rates. Precisely, for each acquisition event seen in the host tree, we measured the CO1 branch length (or absolute time in time-calibrated trees) elapsed between the acquisition and the first loss that occurred in the descending lineages. If more than one lineage derived from the one where the infection occurred, we summed these lineages, to compute the duration along which no loss event occurred (as illustrated in fig. S6). Similarly, following each loss event seen in the host tree, we computed the duration along which no acquisition occurred. Each acquisition event thus contributes one point to estimate the probability, as a function of time, that no loss occurred following an acquisition, while each loss event contributes one point to estimate the probability that no acquisition occurred following a loss event. Importantly, this means that the different data points are independent: two branches in the host tree contribute only one point if they share the same infection status by descent.

Technically, we fitted the *cumulated* curve, that is, for a given time t (the x axis), the probability, estimated from our data, that no event occurred in a time *t at least as long*. Using such a cumulated curve improves the fit to the model by smoothing the noise and is computationally tractable, as the exponential function is the cumulated distribution function corresponding to the Poisson point process used here. The cumulated probability distributions were fitted through Ordinary Least Squares (OLS). Another approach would have been to fit

the probability distribution itself using maximum likelihood. Our trials in doing so have shown that this approach gives an undue weight to rare events occurring deep in the tree, where most of the data uncertainty was concentrated. The OLS approach on cumulated data was thus preferred. For the single Poisson process, we fitted the cumulated probability with the function $\exp(-\lambda*t)$, with λ either the loss or acquisition rate. For the double Poisson process, where the data are explained as the sum of a fast and a slow process, the function fitted was $\alpha*\exp(-\lambda_{fast}*t) + (1-\alpha)*\exp(-\lambda_{slow}*t)$, α being the proportion of events occurring at rate λ_{fast} (i.e., imperfect maternal transmission or other individual-level events). We only present results for the population rates in the paper (slow rates), fast rates being highly dependent on the length of very short branches which are not accurately estimated because they often carry 0 substitution.

Author contributions

MBB designed and conducted the analysis and contributed to writing. PMS designed and conducted the experimental work, contributed to the analysis and to writing. GS contributed to the analysis. GM contributed to the experimental work. MFS contributed to the analysis and to writing. SC designed and conducted the project, and wrote the manuscript.

Acknowledgments

We thank Marie Cariou, Laurent Duret, Jan Engelstädter, Max Reuter and Fabrice Vavre for commenting on an earlier version of this manuscript, Hélène Henri for a gentle support on scheduling, and the following colleagues for assistance and advice in the analysis: Samuel Alison, Thomas Bigot, Damien De Vienne, Manolo Gouy, Matthew Hartfield, Benjamin Horvilleur, Nicolas Lartillot and Niklas Wahlberg. We thank the Genoscope for DNA sequencing through the “Bibliothèque du Vivant” and “Speed ID” projects. This work was supported by the Centre National de la Recherche Scientifique (CNRS) (ATIP grant SymbioCode to S.C.) and benefitted from the computing facilities of the CC LBBE/PRABI.

Software availability: https://github.com/ssolo/ALE.git

References

Ahmed MZ, Breinholt JW, Kawahara AY. 2016. Evidence for common horizontal transmission of Wolbachia among butterflies and moths. BMC Evol. Biol. 16:118.

Ahmed MZ, Li SJ, Xue X, Yin XJ, Ren SX, Jiggins FM, Greeff JM, Qiu BL. 2015. The Intracellular Bacterium Wolbachia Uses Parasitoid Wasps as Phoretic Vectors for Efficient Horizontal Transmission. PLoS Pathog. 11:1–19.

Baldo L, Dunning Hotopp JC, Jolley KA, Bordenstein SR, Biber SA, Choudhury RR, Hayashi C, Maiden MC, Tettelin H, Werren JH. 2006. Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. Appl Env. Microbiol 72:7098–7110.

Brown AN, Lloyd VK. 2015. Evidence for horizontal transfer of Wolbachia by a Drosophila mite. Exp. Appl. Acarol. 66:301–311.

Burt A, Trivers R. 2006. Genes in Conflict: The Biology of Selfish Genetic Elements. Harvard: Belknap Press.

Charlat S, Hornett EA, Fullard JH, Davies N, Roderick GK, Wedell N, Hurst GD. 2007. Extraordinary flux in sex ratio. Science. 317:214.

Comandatore F, Sassera D, Montagna M, Kumar S, Koutsovoulos G, Thomas G, Repton C, Babayan SA, Gray N, Cordaux R, et al. 2013. Phylogenomics and Analysis of Shared Genes Suggest a Single Transition to Mutualism in Wolbachia of Nematodes. Genome Biol. Evol. 5:1668–1674.

Conow C, Fielder D, Ovadia Y, Libeskind-hadas R. 2010. Jane : a new tool for the cophylogeny reconstruction problem. Algorithms Mol. Biol. 5:1–10.

Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Bouletreau M. 2001. Removing symbiotic Wolbachia bacteria specifically inhibits oogenesis in a parasitic wasp. Proc Natl Acad Sci U S A 98:6247–6252.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Engelstädter J, Hurst GDD. 2006. The dynamics of parasite incidence across host species. Evol. Ecol. 20:603–616.

Gillespie RG, Claridge EM, Goodacre SL. 2008. Biogeography of the fauna of French Polynesia : diversification within and between a series of hot spot archipelagos. 363:3335–3346.

Guillou H, Maury RC, Blais S, Cotten J, Legendre C, Guille G, Caroff M. 2005. Age progression along the Society hotspot chain (French Polynesia) based on new unspiked K-Ar ages. Bull. la Société Géologique Fr. 176:135–150.

Hamm CA, Begun DJ, Vo A, Smith CCR, Saelao P, Shaver AO, Jaenike J, Turelli M. 2014. Wolbachia do not live by reproductive manipulation alone: Infection polymorphism in Drosophila suzukii and D. Subpulchrella. Mol. Ecol. 23:4871–4885.

Heath BD, Butcher RD, Whitfield WG, Hubbard SF. 1999. Horizontal transfer of Wolbachia between phylogenetically distant insect species by a naturally occurring mechanism. Curr Biol 9:313-316.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol. Biol. Evol. 22:1561–1568.

Huigens ME, Luck RF, Klaassen RH, Maas MF, Timmermans MJ, Stouthamer R. 2000. Infectious parthenogenesis. Nature 405:178–179.

Hurst GD, Jiggins FM. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc Biol Sci 272:1525–1534.

Jansen G, Savolainen R, Vepsäläinen K. 2010. Phylogeny, divergence-time estimation,

biogeography and social parasite-host relationships of the Holarctic ant genus Myrmica (Hymenoptera: Formicidae). Mol. Phylogenet. Evol. 56:294–304.

Jiggins FM, von Der Schulenburg JH, Hurst GDD, Majerus ME. 2001. Recombination confounds interpretations of Wolbachia evolution. Proc Biol Sci 268:1423–1427.

Johnson KP, Cruickshank RH, Adams RJ, Smith VS, Page RDM, Clayton DH. 2003. Dramatically elevated rate of mitochondrial substitution in lice (Insecta : Phthiraptera). Mol. Phylogenet. Evol. 26:231–242.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Lefoulon E, Bain O, Makepeace BL, Haese C, Uni S, Martin C, Gavotte L. 2016. Breakdown of coevolution between symbiotic bacteria Wolbachia and their filarial hosts. PeerJ 4:e1840.

Martinez J, Longdon B, Bauer S, Chan YS, Miller WJ, Bourtzis K, Teixeira L, Jiggins FM. 2014. Symbionts Commonly Provide Broad Spectrum Resistance to Viruses in Insects: A Comparative Analysis of Wolbachia Strains. PLoS Pathog. 10:e1004369.

Merkle D, Middendorf M, Wieseke N. 2010. A parameter-adaptive dynamic programming approach for inferring cophylogenies. BMC Bioinformatics 11 Suppl 1:S60.

Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, Fukatsu T. 2014. Evolutionary origin of insect-Wolbachia nutritional mutualism. Proc. Natl. Acad. Sci. 111:10257–10262.

Obbard DJ, MacLennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM. 2012. Estimating divergence dates and substitution rates in the drosophila phylogeny. Mol. Biol. Evol. 29:3459–3473.

Penny D. 2005. Relativity for molecular clocks. Nature 436:183–184.

Pohl N, Sison-Mangus MP, Yee EN, Liswi SW, Briscoe AD. 2009. Impact of duplicate gene copies on phylogenetic analysis and divergence time estimates in butterflies. BMC Evol. Biol. 9:99.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490.

Ramage T, Martins-Simoes P, Mialdea G, Allemand R, Duplouy AMR, Rousse P, Davies N, Roderick GK, Charlat S. 2017. A DNA barcode-based survey of terrestrial arthropods in the Society Islands of French Polynesia: host diversity within the SymbioCode project. Eur. J. Taxon. in press.

Ratnasingham S, Hebert P. 2013. A DNA-Based Registry for All Animal Species : The Barcode Index Number (BIN) System. PLoS One 8:e66213.

Raychoudhury R, Baldo L, Oliveira DC, Werren JH. 2009. Modes of acquisition of Wolbachia: horizontal transfer, hybrid introgression, and codivergence in the Nasonia species complex. Evolution (N. Y). 63:165–183.

Regier JC, Shultz JW, Zwick A, Hussey A, Ball B, Wetzer R, Martin JW, Cunningham CW. 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463:1079–1083.

Richardson MF, Weinert LA, Welch JJ, Linheiro RS, Magwire MM, Jiggins FM, Bergman CM. 2012. Population Genomics of the Wolbachia Endosymbiont in Drosophila melanogaster.Kopp A, editor. PLoS Genet. 8:e1003129.

Simões PM, Mialdea G, Reiss D, Sagot M-F, Charlat S. 2011. Wolbachia detection: an assessment of standard PCR protocols. Mol. Ecol. Resour. 11:567–572.

Sintupachee S, Milne JR, Poonchaisri S, Baimai V, Kittayapong P. 2006. Closely related Wolbachia strains within the pumpkin arthropod community and the potential for

horizontal transmission via the plant. Microb. Ecol. 51:294–301.

Sota T, Yamamoto S, Cooley JR, Hill KBR, Simon C, Yoshimura J. 2013. Independent divergence of 13- and 17-y life cycles among three periodical cicada lineages. Proc. Natl. Acad. Sci. U. S. A. 110:6919–6924.

Szöllősi GJ, Davín AA, Tannier E, Daubin V, Boussau B. 2015. Genome-scale phylogenetic analysis finds extensive gene transfer among fungi. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 370:20140335.

Szöllősi GJ, Rosikiewicz W, Boussau B, Tannier E, Daubin V. 2013. Efficient exploration of the space of reconciled gene trees. Syst. Biol. 62:901–912.

Szöllősi GJ, Tannier E, Lartillot N, Daubin V. 2013. Lateral gene transfer from the dead. Syst. Biol. 62:386–397.

Turelli M. 1994. Evolution of incompatibility-inducing microbes and their hosts. Evolution (N. Y). 48:1500–1513.

Vanthournout B, Hendrickx F. 2016. Hidden suppression of sex ratio distortion suggests Red queen dynamics between Wolbachia and its dwarf spider host. J Evol Biol. 29:1488–1494.

Vavre F, Fleury F, Lepetit D, Fouillet P, Bouletreau M. 1999. Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations. Mol Biol Evol 16:1711–1723.

Weinert LA, Araujo-Jnr E V., Ahmed MZ, Welch JJ. 2015. The incidence of bacterial endosymbionts in terrestrial arthropods. Proc. R. Soc. B Biol. Sci. 282:20150249–20150249.

Werren JH. 1998. Wolbachia and Speciation. In: Howard D, Berlocher S, editors. Endless forms, species and speciation. Oxford: Oxford Univeristy Press. p. 245–260.

Werren JH, Baldo L, Clark ME. 2008. Wolbachia: master manipulators of invertebrate

biology. Nat. Rev. Microbiol. 6:741–751.

Werren JH, Bartos JD. 2001. Recombination in Wolbachia. Curr Biol 11:431–435.

Werren JH, Windsor DM. 2000. Wolbachia infection frequencies in insects: evidence of a global equilibrium? Proc Biol Sci 267:1277–1285.

Werren JH, Winsor D, Guo LR. 1995. Distribution of the Wolbachia among neotropical arthropods. Proc. R. Soc. London B 262:197–204.

Zhang JX, Maddison WP. 2013. Molecular phylogeny, divergence times and biogeography of spiders of the subfamily Euophryinae (Araneae: Salticidae). Mol. Phylogenet. Evol. 68:81–92.

Zug R, Koehncke A, Hammerstein P. 2012. Epidemiology in evolutionary time : the case of Wolbachia horizontal transmission between arthropod host species. J. Evol. Biol. 25:2149-60.

Tables

Taxa	Number of OTUs (infected)	Number of specimens (infected)
Diptera	305 (123)	1007 (349)
Lepidoptera	223 (81)	809 (228)
Hymenoptera	172 (64)	514 (128)
Hemiptera	133 (82)	457 (220)
Coleoptera	119 (19)	259 (27)
Araneae	50 (24)	245 (77)
Psocodea	24 (19)	73 (53)
Orthoptera	16 (12)	112 (32)
Blattodea	11 (2)	30 (7)
Other	35 (17	120 (25

Table 1. A summary of the taxonomic diversity and *Wolbachia* infection frequencies in the SymbioCode sample. The number of infected OTUs and specimens are indicated in parenthesis.

Figure legends

Fig. 1. Distribution of the ages of present day infections, using CO1 branch length as a proxy for time. Each point of this distribution is an acquisition event that led to a present day infection in one of the 1,000 loss / acquisition scenarios produced by ALE.

Fig. 2. Dynamics of *Wolbachia* acquisition (A) and extinction (B). Data (solid lines), single Poisson model (dotted lines) and double Poisson model (dashed lines). A: proportion of paths in the host tree (each starting with a *Wolbachia* loss event) remaining uninfected after a time *t*. B: proportion of paths in the host tree (starting with an acquisition event) remaining infected after a time *t*. CO1 distance (number of substitutions per site) is taken as a proxy for time. β_p and γ_p are population rates, explaining the long-term dynamics, while β_i and γ_i are individual rates, explaining the recent dynamics. The fast rates β_i and γ_i apply to a proportion α of all events.

Fig. 3. Distribution of the estimated rates of extinction (x axis) and acquisition (y axis), taking CO1 branch length as time unit, across the 1,000 reconciliation scenarios sampled. Grey levels indicate relative density.

Fig. 4. Distribution of the predicted global *Wolbachia* incidence at equilibrium. The dashed line indicates the observed incidence in our dataset.

Fig. 5. Distributions of the estimated extinction and acquisition rates (A and B, respectively) for arthropod orders represented by at least 50 species in our data set. Because the number of events was small within each order, the estimation of the variability was done by bootstrapping repeatedly (10000 times) 100 scenario out of the 1,000 plausible loss / acquisition scenarios and computing the rates on these data. Dotted lines indicate the global rates.

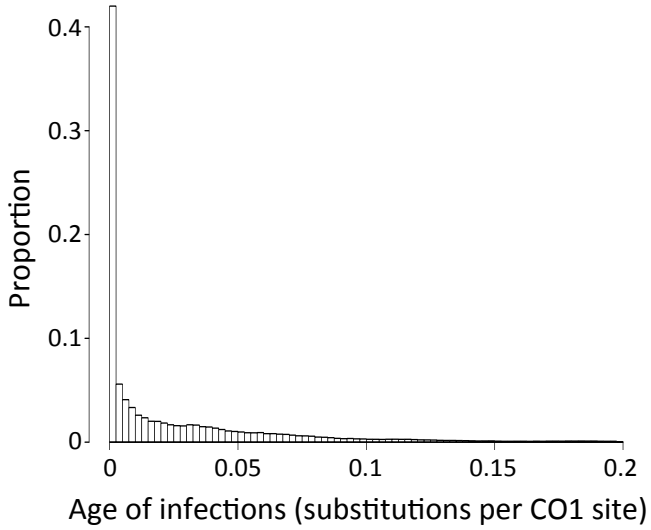
0.4
0.3
0.2
0.1
0
Proportion
0
0.05
0.1
0.15
0.2
Age of infections (substitutions per CO1 site)

Figure 1.

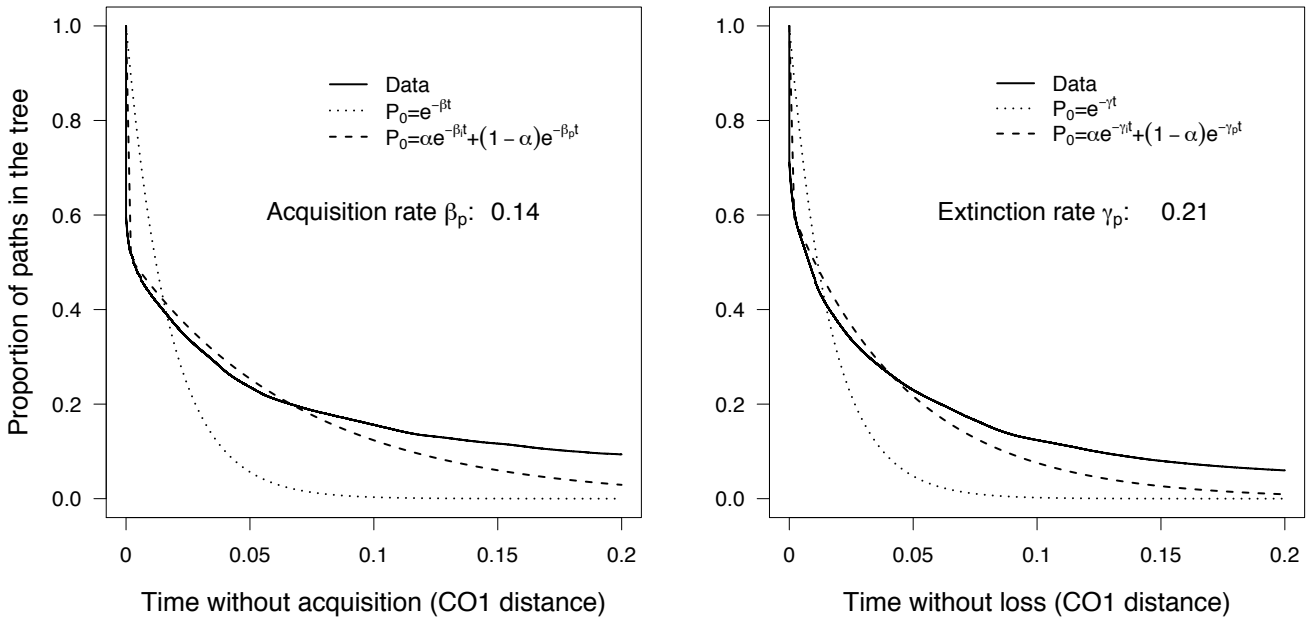
Proportion of paths in the tree
1.0
0.8
0.6
0.4
0.2
0.0
0
0.05
0.1
0.15
0.2
Time without acquisition (CO1 distance)
Data
$P_0=e^{-\beta t}$
$P_0=\alpha e^{-\beta t}+(1-\alpha)e^{-\beta_p t}$
Acquisition rate β_p: 0.14
Time without loss (CO1 distance)
Data
$P_0=e^{-\gamma t}$
$P_0=\alpha e^{-\gamma t}+(1-\alpha)e^{-\gamma_p t}$
Extinction rate γ_p: 0.21

Figure 2.

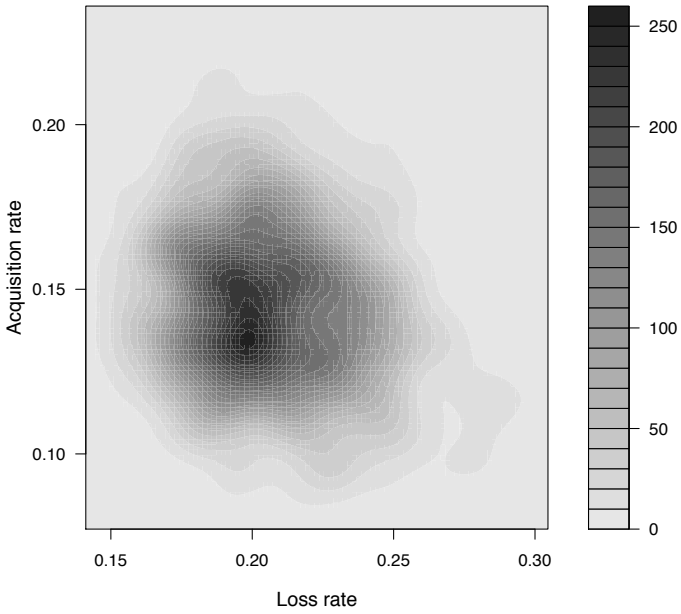
Acquisition rate
0.20
0.15
0.10
0.15
0.20
0.25
0.30
Loss rate
250
200
150
100
50
0

Figure 3.

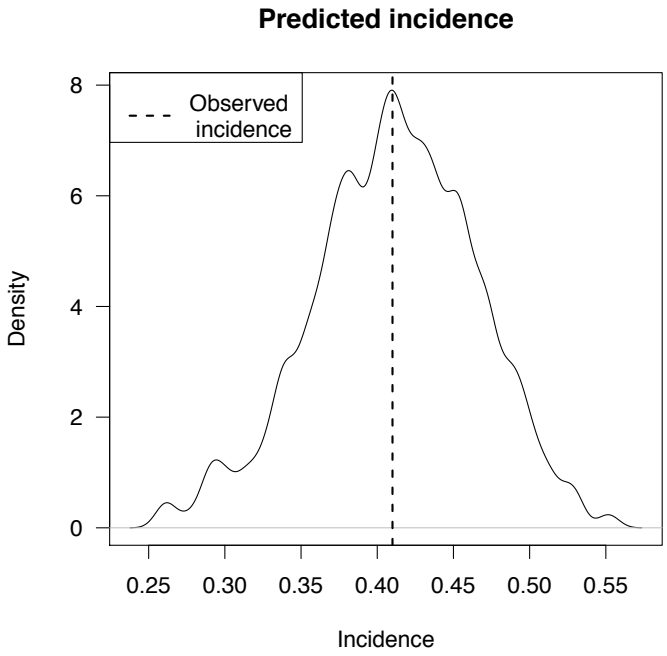
Predicted incidence
Observed incidence
Density
0
2
4
6
8
0.25
0.30
0.35
0.40
0.45
0.50
0.55
Incidence

Figure 4.

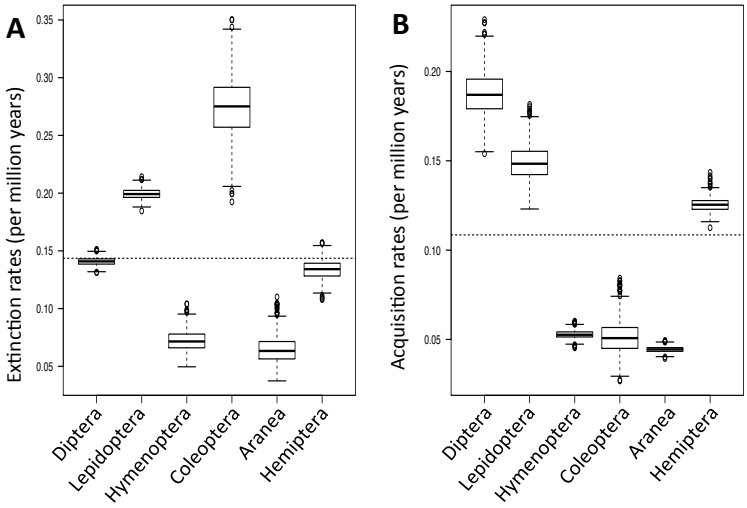
A
Extinction rates (per million years)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
Diptera
Lepidoptera
Hymenoptera
Coleoptera
Aranea
Hemiptera
B
Acquisition rates (per million years)
0.20
0.15
0.10
0.05
Diptera
Lepidoptera
Hymenoptera
Coleoptera
Aranea
Hemiptera

Figure 5.