AMERICAN UNIVERSITY OF BEIRUT

INVESTIGATING EVOLUTIONARY TRANSITIONS IN WOLBACHIA PIPIENTIS' PROPENSITY TO INDUCE DIFFERENT REPRODUCTIVE HOST PHENOTYPES

by
CARINE OUSSAMA HOURY

A thesis
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Department of Biology
of the Faculty of Arts and Sciences
at the American University of Beirut
Beirut, Lebanon
September 2016

AMERICAN UNIVERSITY OF BEIRUT

INVESTIGATING EVOLUTIONARY TRANSITIONS IN WOLBACHIA PIPIENTIS' PROPENSITY TO INDUCE DIFFERENT REPRODUCTIVE HOST PHENOTYPES

by

CARINE OUSSAMA HOURY

Approved by:

Dr. Heinrich Burggraf Zu Dohna-Schlobi, Assistant Professor — Advisor
Department of Biology

Dr. Khouzama Knio, Professor — Member of Committee
Department of Biology

Dr. Zakaria Kambris, Assistant Professor — Member of Committee
Department of Biology

Date of thesis defense: September 9, 2016

AMERICAN UNIVERSITY OF BEIRUT

THESIS, DISSERTATION, PROJECT RELEASE FORM

Student Name: Houry Carine Oussama

Last First Middle

● Master's Thesis ○ Master's Project ○ Doctoral Dissertation

■ I authorize the American University of Beirut to: (a) reproduce hard or electronic copies of my thesis, dissertation, or project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

☐ I authorize the American University of Beirut, **three years after the date of submitting my thesis, dissertation, or project,** to: (a) reproduce hard or electronic copies of it; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes.

_____

Signature Date

ACKNOWLEDGMENTS

Special thanks are to my advisor Dr. Heinrich Burggraf Zu Dohna-Schlobi, for his guidance and time during his research leave. I would also thank Dr. Zakaria Kambris for my time in his laboratory.

I would like to express my sincere gratitude to my family, my fiancé (my husband-to-be) and my friends for their continuous support and motivation.

AN ABSTRACT OF THE THESIS OF

Carine Oussama Houry for Master of Science
Major: Biology

Title: Investigating evolutionary transitions in *Wolbachia pipientis'* propensity to induce different reproductive host phenotypes

The endosymbiotic bacterium *Wolbachia* infects wide range of arthropods and their relatives. It is an intracellular parasite transmitted through the egg from mother to offspring. *Wolbachia* can spread and persist through various means of host reproductive manipulation. These manipulations include cytoplasmic incompatibility, the selective killing of male offspring, induction of parthenogenesis, feminization of genetic males.

It is unclear to what degree the induced host phenotype is determined by Wolbachia genetics and how *Wolbachia* evolved such a wide array of host manipulations. We assessed the phylogenetic signal within our studied trait, i.e. the *Wolbachia*-induced host phenotype, by comparing the likelihood of a model based on a star tree to the likelihood of the model based on our actual tree. Our results suggest that the model based on the actual tree has the higher likelihood inferring that the *Wolbachia*-induced host phenotype is dependent on the *Wolbachia* strain.

To understand transition rates from one host phenotype to another, we fitted contrasting models for evolutionary transitions to a phylogenetic tree and determined which one is more consistent with data. We used two model selection procedures: Bayes factor approximation from harmonic mean and model likelihood comparisons with optimizing parameters. The former approach favors a model with no constraints imposed on the data oppositely to the latter approach which favors a model with a single transition rate from and to all host phenotypes. Our results suggest that the estimations based on maximum likelihood are more conclusive, in view of drawbacks of Bayes factor approximation from harmonic mean and maximum parsimony considerations.

CONTENTS

ILLUSTRATIONS

TABLES

CHAPTER I

INTRODUCTION

Wolbachia is one of the most common parasitic microbes that infect a wide range of arthropods and their relatives. Estimates of the percentage of insect species infected by *Wolbachia* range from 40 to 65% (Hilgenboecker, Hammerstein, Schlattmann, Telschow, & Werren, 2008; Zug & Hammerstein, 2012). In addition, *Wolbachia* is an obligate mutualist of a number of filarial nematode species (Bandi, Anderson, Genchi, & Blaxter, 1998).

Wolbachia strains infecting arthropods and nematodes belong to a single species, *Wolbachia pipientis*. However, the wide diversity in these strains has resulted in their assignment into sixteen reported *Wolbachia* supergroups, named A to F and H to Q (Augustinos et al., 2011; Bing et al., 2014; Fukui et al., 2015; Glowska, Dragun-Damian, Dabert, & Gerth, 2015; Haegeman et al., 2009; O 'neill, Giordano, Colbert, Karrf, & Robertson, 1992; Ros, Fleming, Feil, & Breeuwer, 2009), with the supergroups A-D being the most studied. Supergroup G is no longer considered a separate *Wolbachia* supergroup as it has been reported to be a recombinant between supergroups A and B (Baldo & Werren, 2007). There is a considerable interest in *Wolbachia* due to its wide host range, its complex co-evolution with its hosts and the potential to use this endosymbiont as a tool for biological control of arthropods and nematodes (Augustinos et al., 2011; Iturbe-Ormaetxe, Walker, & O' Neill, 2011; Kageyama, Narita, & Watanabe, 2012; Slatko, Taylor, & Foster, 2010).

Wolbachia is an intracellular parasite transmitted through the egg from mother to offspring. Strains manipulate the reproductive phenotype of the host to increase the transmission (Werren et al., 2016). The hosts are targeted at different developmental stages ranging from the spermatogenesis stage to the pupal stage (Ma, Vavre, & Beukeboom, 2014). Some strains have evolved strategies to distort the host sex ratio towards females thus increasing the endosymbiont spread (Cordaux, Bouchon, & Grève, 2011; Duron et al., 2008; Partridge & Hurst, 1998; J. H. Werren, Baldo, & Clark, 2008). These strategies include the selective killing of male offspring, induction of parthenogenesis and the feminization of genetic males (J. H. Werren et al., 2008). Such strategies result in female-biased sex ratios. Other strains induce cytoplasmic incompatibility which reduces the offspring viability in a cross between an uninfected female and an infected male(O'Neill, Giordano, Colbert, Karr, & Robertson, 1992). CI decreases the offspring production of uninfected females. *Wolbachia* is the most successful bacterial endosymbiont involved in reproductive parasitism not only because it is widespread, but also because it has one of the widest range of different host reproductive manipulations (Hilgenboecker et al., 2008).

Male-killing *Wolbachia* strains act by repressing the host masculinizing gene, forcing the development of a female sex, in species where the two sexes have different numbers of chromosomes. In these cases, host developing as a sex that does not match the genetic makeup can be fatal to the host, since the gene dose is disturbed (Fukui et al., 2015). Parthenogenesis typically occurs in hymenopteran species; fertilized eggs produce diploid females while unfertilized eggs produce haploid males parthenogenetically (Wenseleers & Billen, 2000). Induction of parthenogenesis involves making the

reproduction of the host independent of fertilization. *Wolbachia* alters the cell cycle, forcing diploidy in unfertilized cells, hence the development of female progeny parthenogenetically (Ma et al., 2014). Feminization induction, another host reproductive manipulation by *Wolbachia* acts by suppressing an androgenic gland in developing males ((Werren, 1997). Feminization seems happen in crustaceans more frequently than in insects due to the ease of the manipulation of sexual phenotypes in the former (Ma et al., 2014). Cytoplasmic incompatibility is a post-zygotic reproductive separation that occurs when infected males and uninfected females mate or when mates are infected with different strains of the *Wolbachia* (Hoffmann, Turelli, & Harshman, 1990; O'Neill et al., 1992; Werren, 1997). *Wolbachia* can spread readily in populations through cytoplasmic incompatibility, since infected females have an advantage over uninfected females because the former can reproduce through mating with uninfected and infected males (Werren, 1997). Through these reproductive manipulations, *Wolbachia* endosymbionts increase their own transmission at the expense of the arthropod host's fitness. It has been reported that these manipulations generate a two-arms race with the host over offspring sex, subsequently driving evolutionary changes in host reproduction and sex determination mechanisms (Beukeboom, 2012; Cordaux et al., 2011; Hatcher et al., 1999; Hurst & Werren, 2001; Stouthamer et al.,2010; Werren & Beukeboom, 1998). In filarial nematodes, the mechanisms induced by *Wolbachia* are rather mutualistic. Elimination of *Wolbachia* from host nematodes usually results in either death or sterility of the hosts (Hoerauf et al., 2003).

It is unclear how *Wolbachia* evolved such a wide array of host manipulations.

Inducing a particular host phenotype often requires a complex set of specific mechanisms to interfere with the host's reproductive physiology (Werren, 1997). One would therefore expect that multiple changes in the genetic make-up of *Wolbachia* would be necessary for *Wolbachia* to make the transition from inducing one host phenotypes to another (Jaenike, 2007). If multiple genetic changes are required for a shift in host phenotype, then closely related *Wolbachia* strains should induce the same host phenotype. However, this view is not always supported by experimental and phylogenetic studies. While a given *Wolbachia* strain usually induces the same phenotype when transfected into novel hosts (Boyle et al., 1993; Braig et al., 1994; Charlat et al., 2004; Clancy & Hoffmann, 1997; Grenier et al., 1998; Poinsot et al.,1998; Riegler et al., 2004; Ruang-Areerate & Kittayapong, 2006; Zabalou et al., 2004), there are also counter-examples where the induced host phenotype changes when a *Wolbachia* strain is transfected into different hosts (Jaenike, 2007). Similarly, there are cases where very closely related strains induce different host phenotypes (Hornett et al., 2008; Jiggins et al., 2002). Hence, even though the biology of *Wolbachia*-host interaction suggests that the induced host phenotype is mainly determined by the genetic make-up of *Wolbachia*, this is not always supported by the empirical evidence.

One way to determine to what degree the induced host phenotype is driven by *Wolbachia* genetics is through comparative phylogenetic methods. Comparative phylogenetic methods use information from the phylogenetic relationships between lineages to test evolutionary hypotheses of trait changes (Revell et al., 2008). An important task in comparative methods is to fit alternative models for the evolutionary transitions

between traits and select the model that is most consistent with the data. There are several methods for model selection in comparative phylogenetics ((Harmon et al.,2008; Pagel, et al., 2004), each relying on different approximations. Some of these approximations have been criticized (Lartillot & Philippe, 2006) and currently there is no consensus about which method provides most reliable inferences.

The goal of our analysis is to determine to what extent the induced host phenotype is an intrinsic property of *Wolbachia* and to estimate the rates at which *Wolbachia* strains switch from inducing one host phenotype to another, relative to the genetic divergence between the strains. We applied two complementary methods of comparative phylogenetics for estimating transition rates between host phenotype to determine whether our results are robust with respect to the methodology.

CHAPTER II

MATERIALS AND METHODS

A. Sequence collection and alignment

Sequences were downloaded from the MLST *Wolbachia* isolate database (http://pubmlst.org/wolbachia/ on April 28, 2016) that stores information about the strain profile including its taxonomic, biological and genetic information (Baldo et al., 2006). Sequences were downloaded for the six loci *gatB*, *coxA*, *hcpA*, *fbpA*, *ftsZ*, and *wsp*. Only sequences of strains with a determined host phenotype and with sequence fragments available for all loci were selected. Seven multiple sequence alignments were created, one per locus and one for sequences obtained from concatenating all loci. The alignments were estimated using MAFFT version 7 (http://mafft.cbrc.jp/alignment/server/) (Katoh & Standley, 2013).

B. Tree reconstruction

Phylogenetic trees were reconstructed from alignments using MrBayes v.3.2 (Ronquist et al., 2012) selecting a general time reversible substitution model with a proportion of invariable sites and a gamma-shaped distribution of substitution rates across sites (GTR + I + □). Two Monte-Carlo Markov chains were run for 10 million generations with a sample frequency of 100 to generate samples from posterior distributions of trees and substitution rate parameters, given the alignment data. Trees were estimated for the alignment of single loci and for the alignment of concatenated sequences.

C.Rate estimation overview

The phylogenetic trees based on the concatenated sequences were used to estimate transition rates of *Wolbachia* strains between their different ways of manipulating the reproductive phenotype of their hosts. The induced host phenotype was treated as a discrete trait of *Wolbachia* and six different models for the transition rates between the trait states along the phylogenetic tree were estimated. The six different models were obtained by combining two ways of classifying the host phenotypes with three parameter constraints for transition rates. Host phenotype manipulations were either classified into five states (induction of parthenogenesis, feminization of genetic males, male killing, cytoplasmic incompatibility and "others") or into two states (cytoplasmic incompatibility and non-cytoplasmic incompatibility). The three parameter constraints for transition rates between host phenotype manipulations were (i) a model with a single rate for all transitions, (ii) a model with symmetric rates, i.e. the rate from phenotype *A* to *B* equals the rate from phenotype *B* to *A* and (iii) a model with no constraints. Transition rates for each of the six models were estimated using Bayesian Monte Carlo Markov chain methods and maximum likelihood.

D. Bayesian Monte Carlo Markov Chain

Posterior distributions of the transition rates of strains between the different host phenotypes were estimated based on the posterior distribution of phylogenetic trees using BayesTraits (Pagel et al., 2004). A uniform prior was used for the transition rates with a prior interval ranging from 0 to 2000. Convergence of parameter distributions to a

stationary distribution was tested according to Geweke's convergence diagnostic (Geweke 2005; R package *coda*). The fit of the three parameter constraints was compared by calculating a Bayes factor from harmonic means of the likelihood (Kass & Raftery, 1995). Ancestral state probabilities at nodes were estimated using Bayestraits command "AddMRCA". Calculation of convergence diagnostics, variance analysis and tree plotting were performed using R (R Core Team, 2015).

E.Maximum likelihood

Transition rates were also fitted via maximum likelihood to the consensus tree obtained from the trees produced by MrBayes (Harmon et al. 2008, R package *geiger*). The Akaike's information criterion (AIC) values were compared between all three parameter constraints (Akaike, 1973; Hu, 1987). In addition, a null-model with a star tree (i.e. no phylogenetic structure) was fitted to the data using the R package *geiger* (Harmon et al. 2008). The likelihood of the null-model was compared to the likelihood of the model based on the actual tree to estimate the strength of the phylogenetic signal of host phenotypes. This analysis was performed for the alignment of each locus and the alignment of the concatenated sequences.

CHAPTER III

RESULTS

The phylogenetic tree based on the concatenated genes resolved the major *Wolbachia* supergroups (A, B and D) (Figure 1). The major clades are supported by high posterior probability values. Host phenotypes cluster on clades in a non-random pattern. However, we can detect a few exceptions. For example, some strains with very similar gene sequences induce different host phenotypes. For example, the *Wolbachia* strain infecting *Hypolimnas bolina* (Linnaeus, 1758) that induces male killing forms a monophyletic group with the strain infecting the same host but that induces another phenotype (Figure 1). These two strains' sequences are 85% identical. To investigate whether the genetic relatedness between different strains on the tree is connected to the transition between host phenotypes, we fitted a null-model with a star tree (i.e. no phylogenetic structure) to the data and compared it to the model based on the actual tree estimated from the alignment of concatenated genes. The null-model produces a lower likelihood than the model based on the actual tree, indicating that host phenotype has a detectable phylogenetic signal on a *Wolbachia* tree and therefore behaves to some degree as *Wolbachia* trait (Table 1).

Trees based on the alignment of each locus alone show a less obvious level of clustering of phenotype traits (Figures 2, 3, 4, 5, 6, 7). We performed the same analysis of likelihood comparisons of the model based on the actual trees to the model based on a star tree for the trees based on the alignment of each locus (*gatB*, *coxA*, *hcpA*, *fbpA*, *ftsZ*, *wsp).*

Our aim was to check whether these genes are implicated in host phenotype. Interestingly, the trees based on the genes *gatB*, *coxA*, *hcpA*, *fbpA* and *wsp* all had a higher likelihood than the star tree (Table 4). However, the model based on the tree estimated from the alignment of concatenated genes produced the highest likelihood compared to the models based on the alignment of single loci. This implies that the phylogenetic tree estimated from concatenated genes alignment is more reliable than the trees estimated from the alignment of single loci.

The best model for the transition rates between different host phenotypes differed between Bayesian Monte Carlo Markov Chain and the likelihood method. According to the Bayes factors, the unconstrained model fits the dataset substantially better than the symmetric model (log BF=19.34, Table 2) and better than the single rate model (log BF = 5.72, Table 2). The Bayes factor implies that the unconstrained model is almost 20 times more likely than the symmetric model and almost five times more likely than the single-rate model, given the data and equal prior odds for both models. Convergence tests show that all rates posterior distributions converge to a stationary distribution. The model with two states (CI and non-CI) show that the unconstrained model fits the dataset slightly better than the symmetric model (log BF=2.98, Table 2). The frequency histograms of posterior distribution of rates when an unconstrained model is fitted to the data for multistate and binary analyses do not show any clear evidence of asymmetry among rates of transition from inducing one phenotype to another (Figures 8 and 9).

Unlike the Bayes factor harmonic mean approximations, maximum likelihood rate fitting favors a single rate model. A single rate model has the lowest AIC value and the

highest likelihood among the models fitted to the data, regardless of whether the host phenotype was treated as a multi-state or binary character (Table 3).

Parameter restrictions	Number of states	Number of free parameters	Log-Likelihood (harmonic mean)
Unrestricted rates	5	20	-80.43
Symmetric rates	5	10	-90.10
Single rate	5	1	-83.29
Unrestricted rates	2	2	-41.19
Symmetric rates	2	1	-42.68
Single rate	2	1	-42.68

Table 1: Comparison of the three different models (Unrestricted rates, symmetric rates and single rate)

Number of states	Complex model	Simple model	Bayes Factor
5	Unrestricted rates	Symmetric rates	19.34
5	Unrestricted rates	Single rate	5.72
5	Symmetric rates	Single rate	-13.62
2	Unrestricted rates	Symmetric rates	2.98
2	Unrestricted rates	Single rate	2.98
2	Symmetric rates	Single rate	

Table 2: Comparison of the Bayes factors of the three different models (Unrestricted rates, symmetric rates and single rate)

Parameter restrictions	Number of states	Number of free parameters	Log-Likelihood	Akaike's Information Criterion
Unrestricted rates	5	20	-70.11	180.23
Symmetric rates	5	10	-80.43	180.86
Single rate	5	1	-84.80	171.60
Unrestricted rates	2	2	-39.11	82.23
Symmetric rates	2	1	-39.14	80.28
Single rate	2	1	-39.14	80.28

Table 3: Comparison of likelihood and Akaike's Information Criterion (AIC) values of models with different rate constraints for multi-state and binary analyses

Gene	Log-likelihood
wsp	-81.64
hcpA	-83.48
ftsZ	-92.82
fbpA	-81.12
coxA	-87.03
gatB	-84.28
Concatenated genes	-79.03
Null model	-89.28

Table 4: Comparison of likelihood values of the null-model to the likelihood of the model based on the actual tree for each locus and for concatenated genes

Ancestral phenotype	Probability for node A	Probability for node B
Cytoplasmic incompatibility (CI)	0.44	0.20
Male killing (MK)	0.14	0.20
Parthenogenesis induction (PI)	0.14	0.20
Feminization induction (FI)	0.14	0.20
Other	0.14	0.20

Table 5: Comparison of the probabilities of each host phenotype at nodes

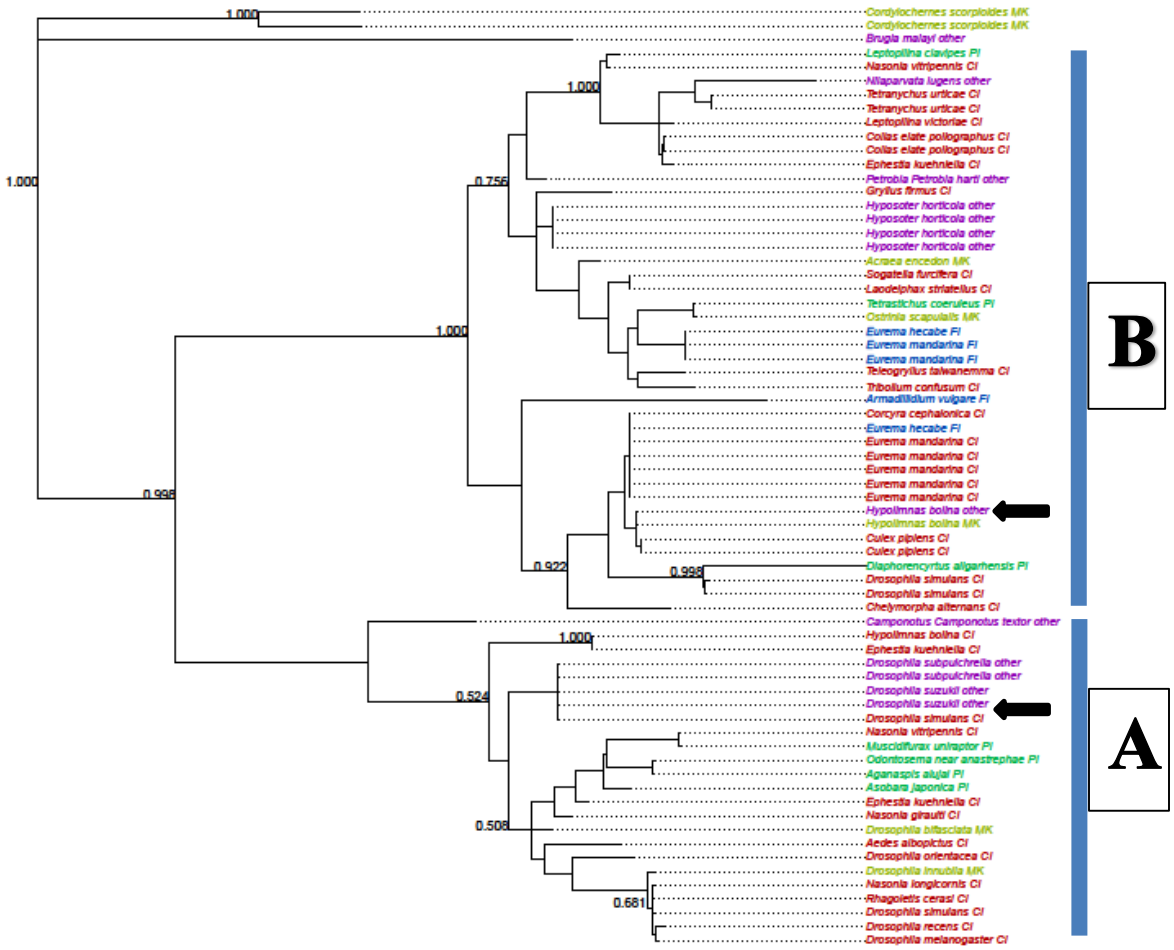

Figure 1: Bayesian MCMC inference phylogeny based on the six concatenated genes (*gatB, coxA, hcpA, fbpA, ftsZ and wsp*). Tree branches are colored by host phenotype indicated at tip labels (CI = cytoplasmic incompatibility, MK = male killing, FI = feminization induction, PI = parthenogenesis induction).Major clades are shown on the tree.Posterior probability values are shown at major nodes. The arrows point to the close *Wolbachia* strain sequences inducing a different phenotype in their host.

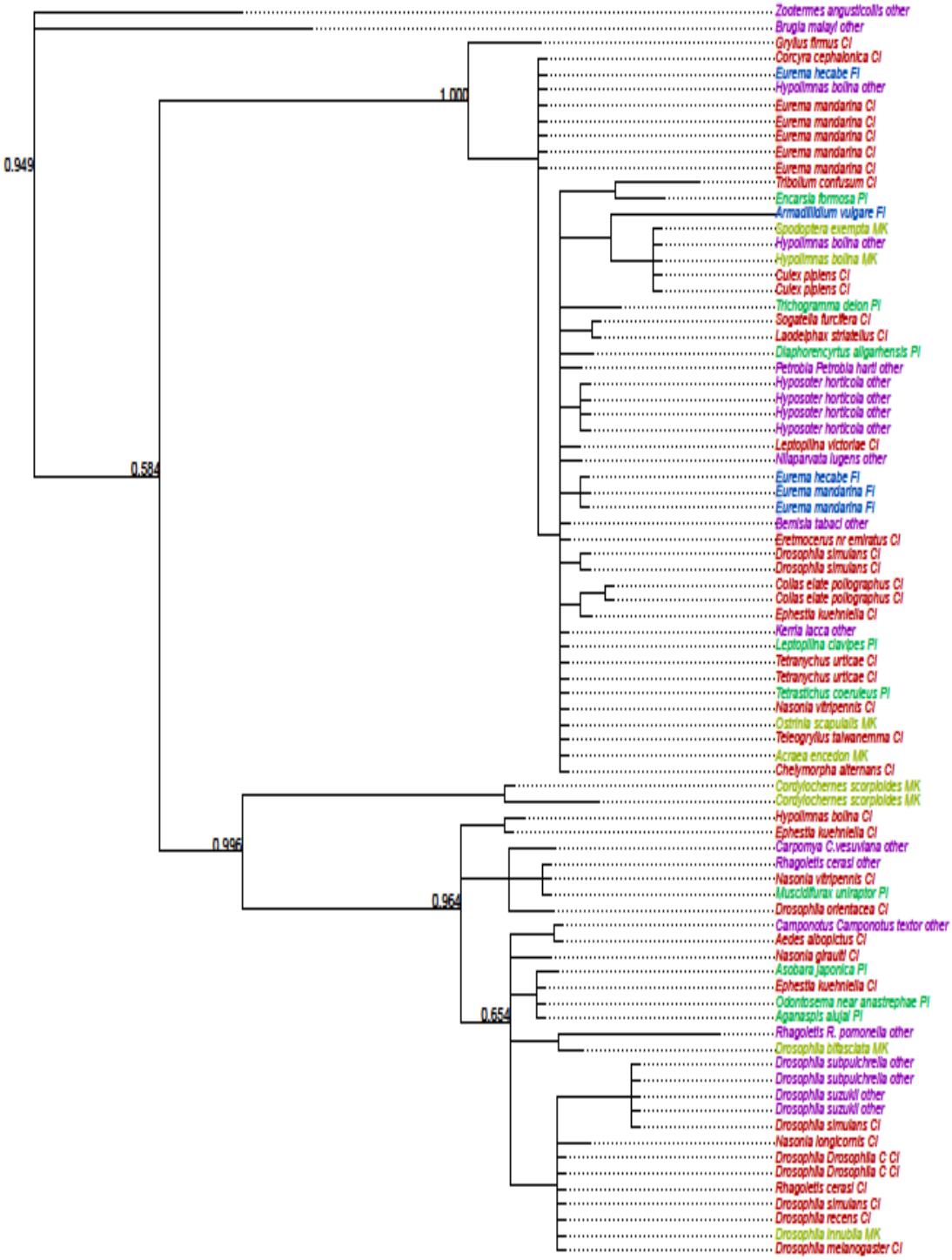

Figure 2: Bayesian MCMC inference phylogeny based on the *gatB* gen

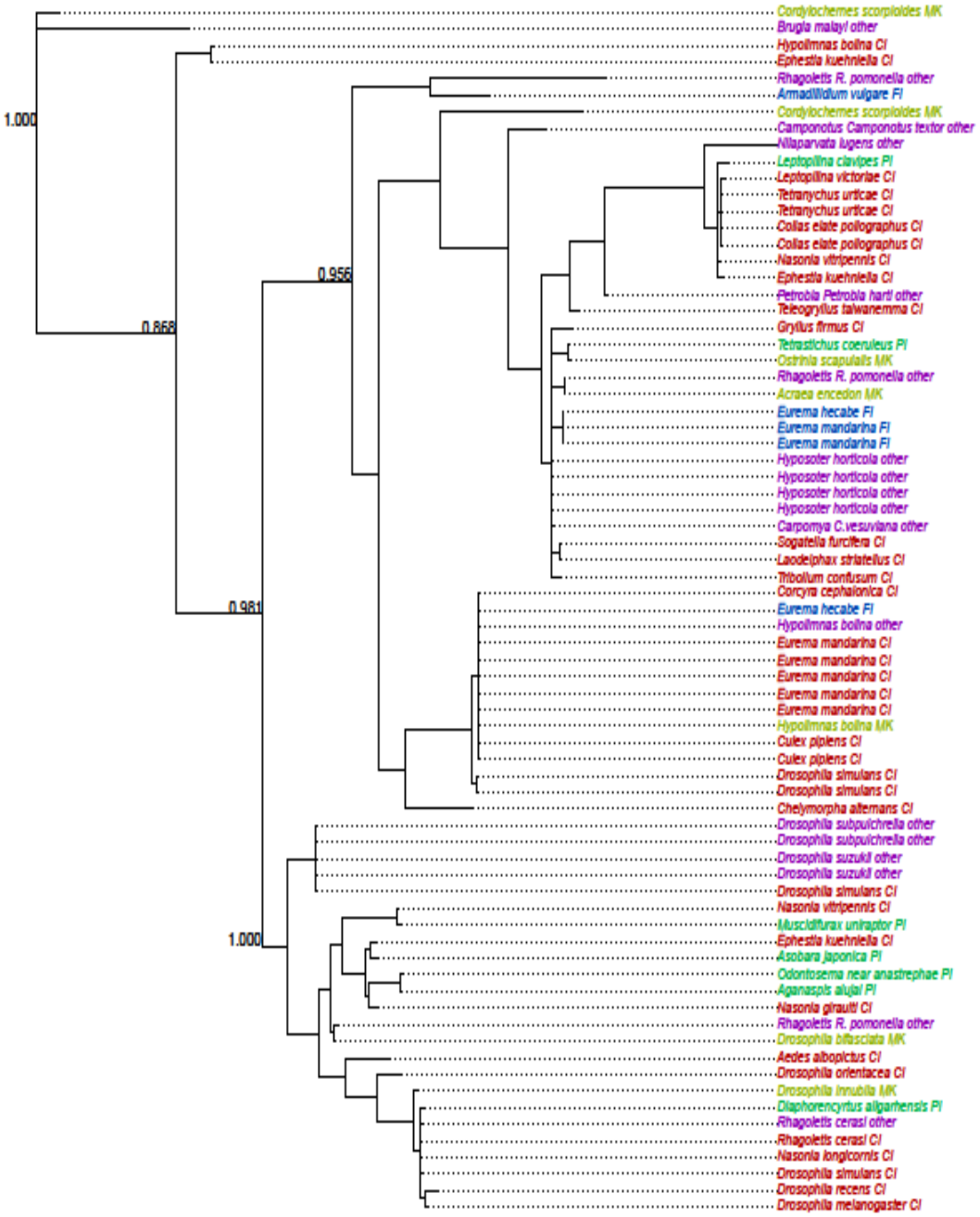

Figure 3: Bayesian MCMC inference phylogeny based on the *wsp* gene

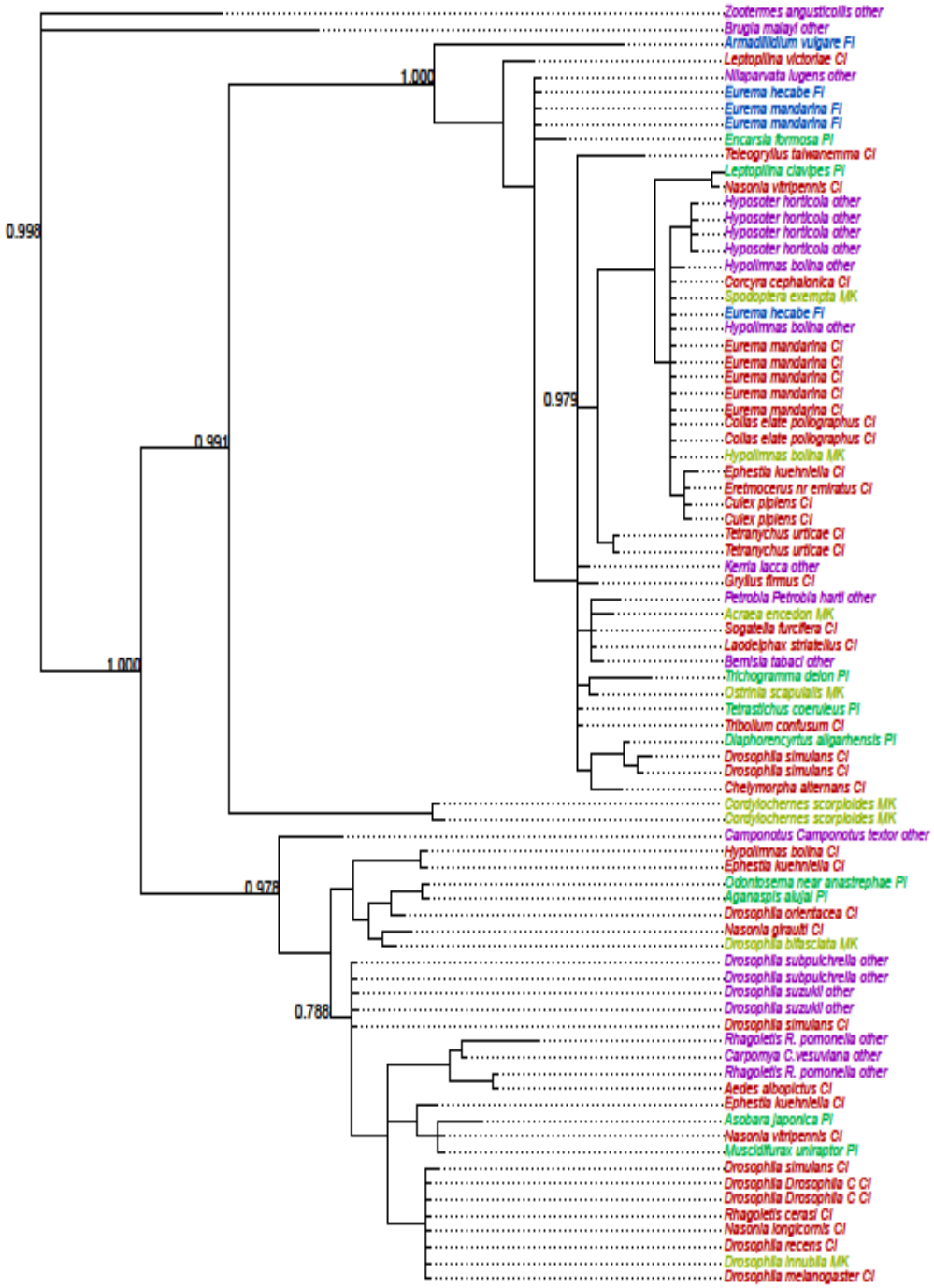

Figure 4: Bayesian MCMC inference phylogeny based on the *hcpA* gene.

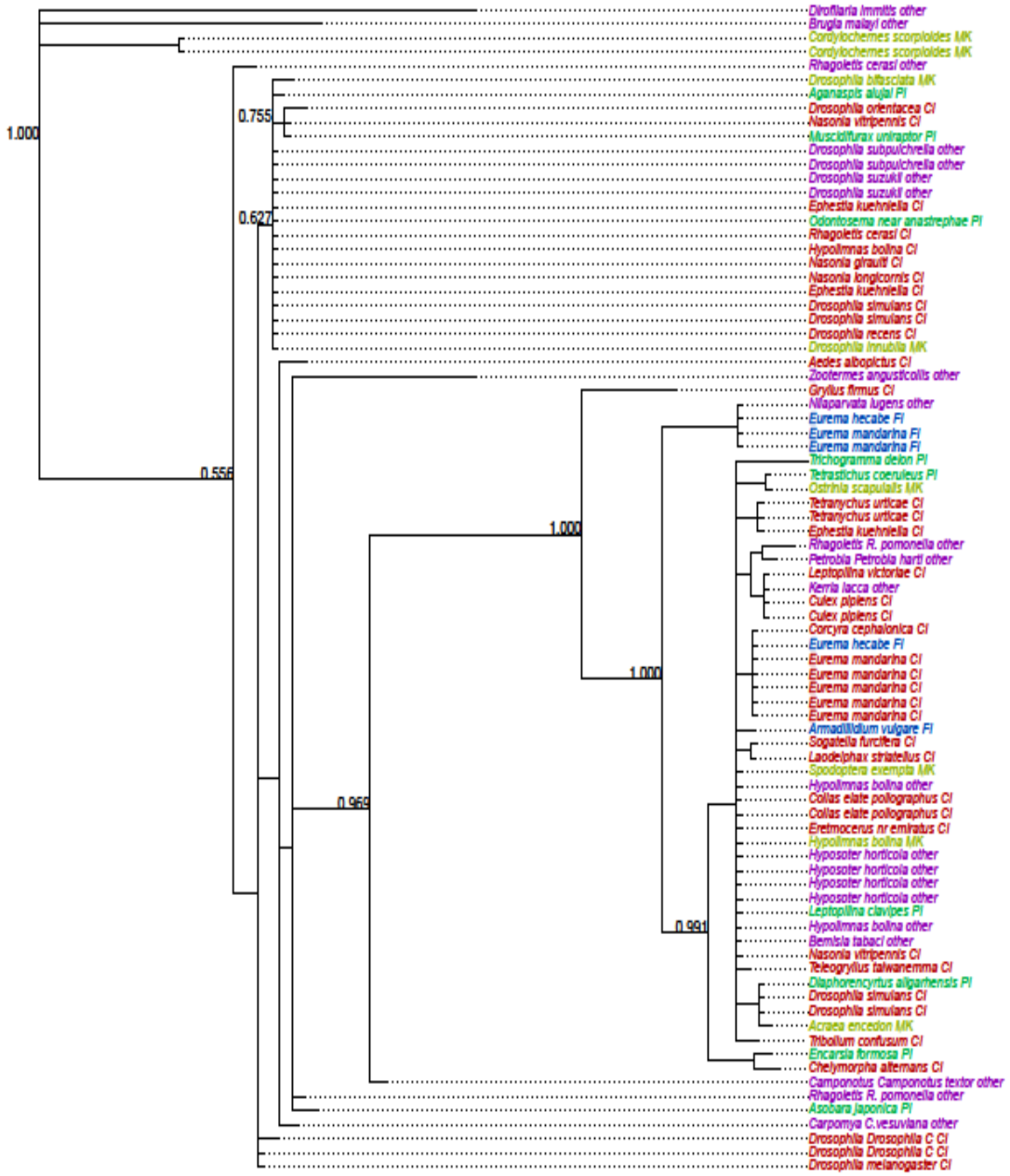

Figure 5: Bayesian MCMC inference phylogeny based on the *ftsZ* gene.

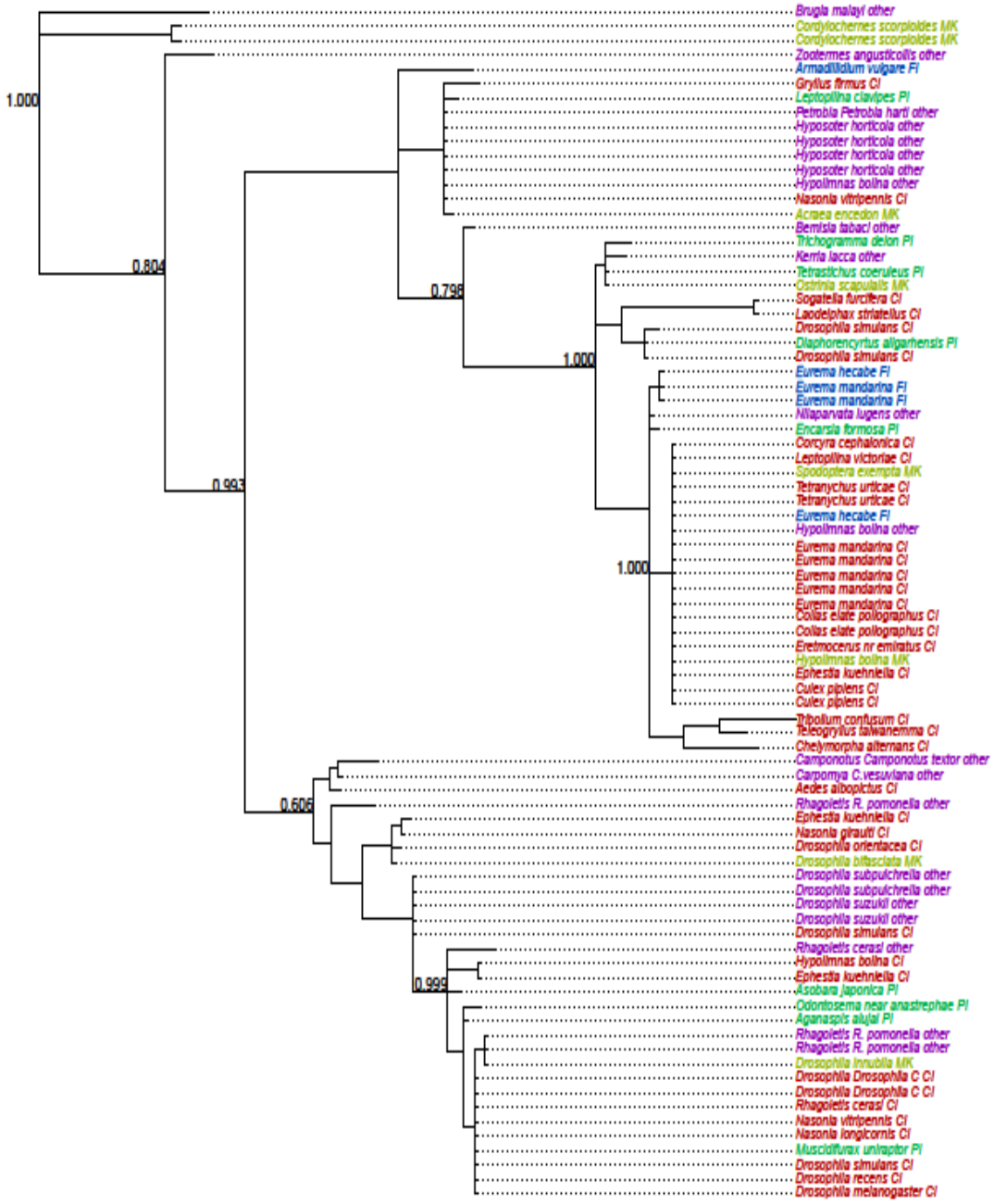

Figure 6: Bayesian MCMC inference phylogeny based on the *fbpA* gene.

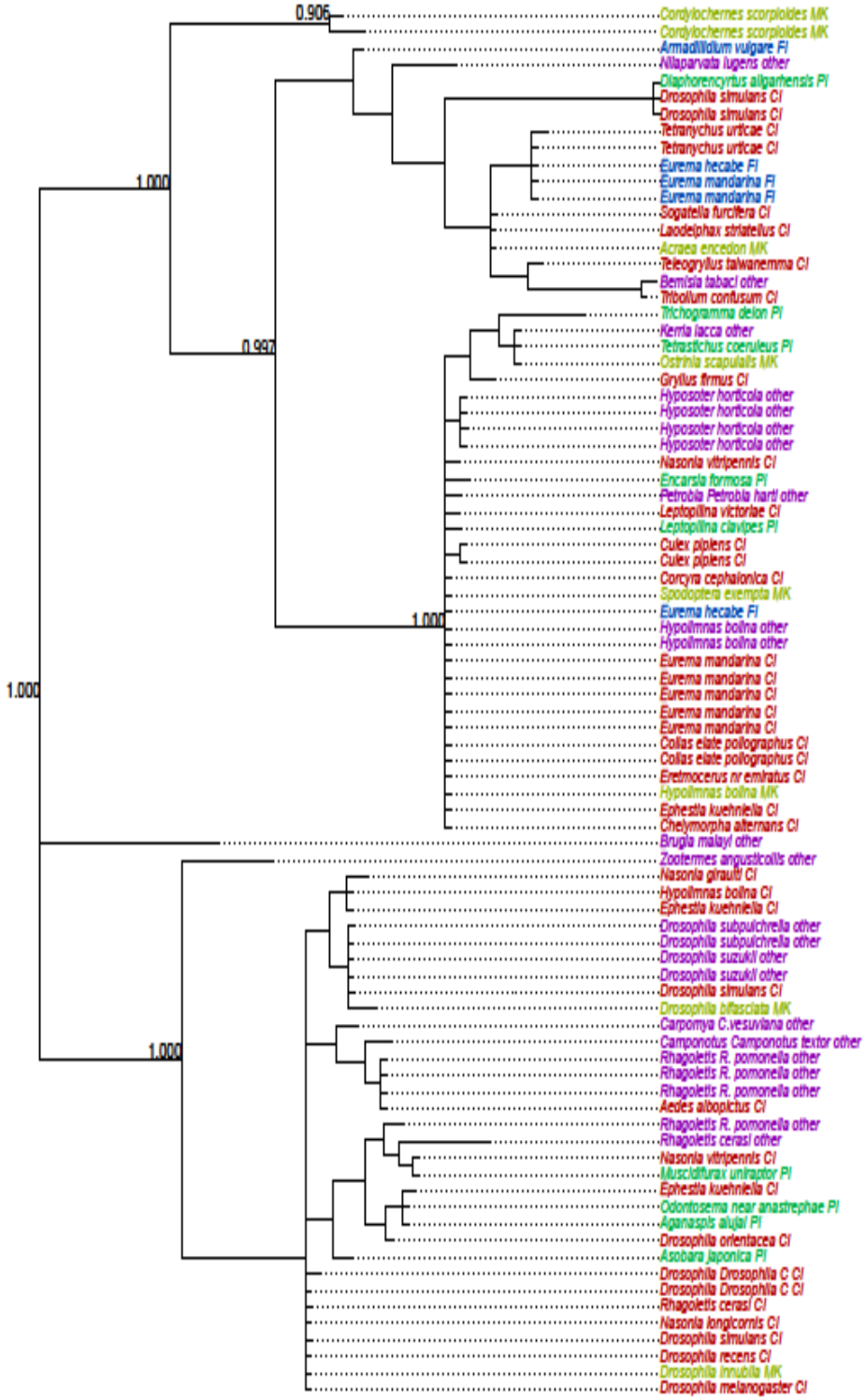

Figure 7: Bayesian MCMC inference phylogeny based on the *coxA* gene.

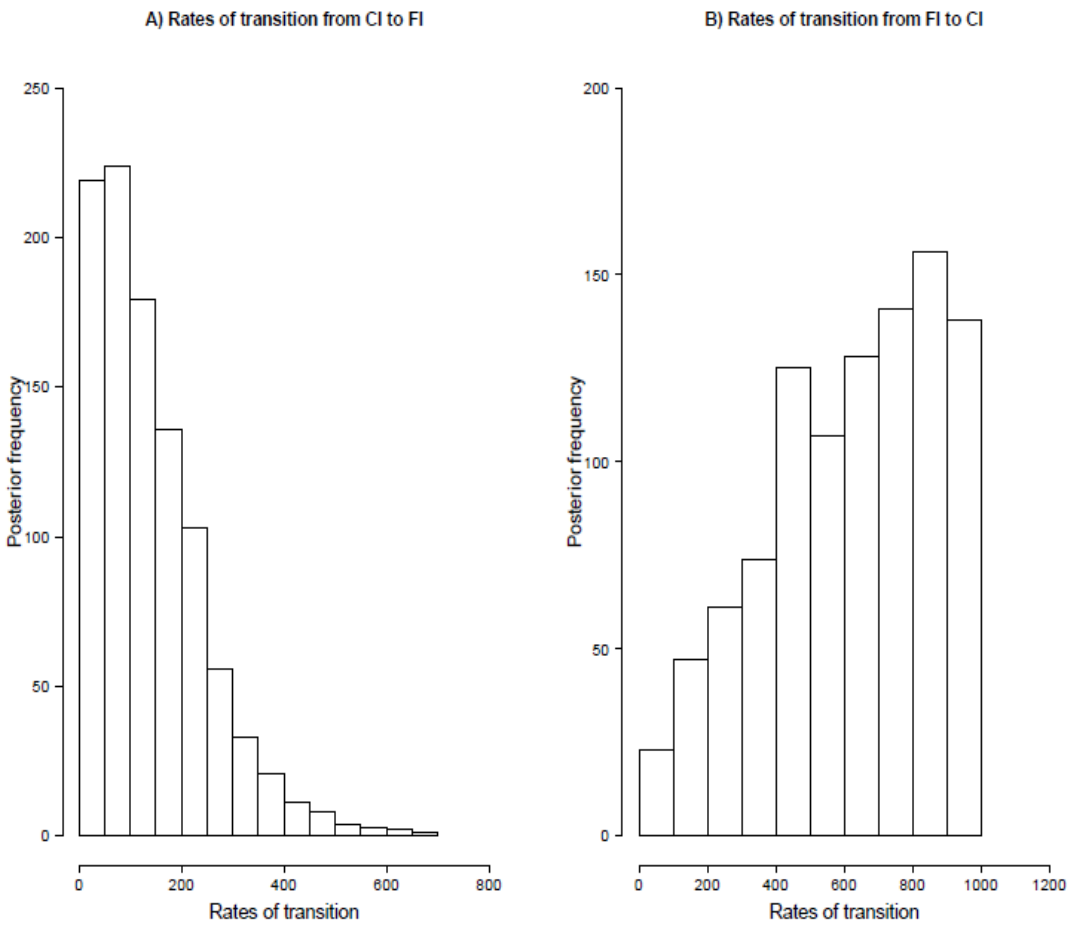

Figure 8: Frequency histograms showing posterior distribution of rates when an unconstrained model is fitted to the data. Probability distributions are plotted for the rates with A) lowest mean and B) highest mean.

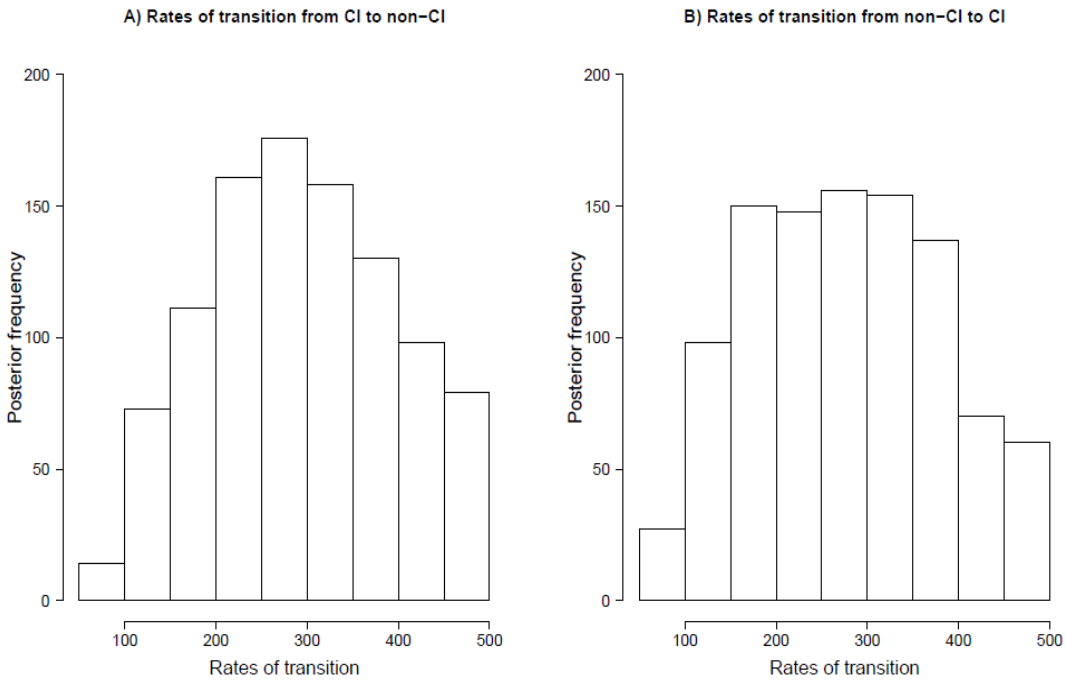

Figure 9: Frequency histograms showing posterior distribution of rates when an unconstrained model is fitted to the data and two states of the host phenotype trait (CI and non- CI) are considered for the analysis.

CHAPTER IV

DISCUSSION

This study investigated the patterns of evolution of *Wolbachia pipientis* in the propensity to induce reproductive phenotypes in arthropod hosts. Host phenotypes are clustered along the phylogenetic *Wolbachia* tree based on the concatenated sequences. The *Wolbachia* trees based on the alignment of single loci do not show a clear trait clustering pattern. A model fitting the transitions between host phenotype on a phylogenetic tree produced a higher likelihood when fitted to the actual tree than when fitted to a star tree with no phylogenetic structure. Hence, the genetic relatedness between different strains on the tree can explain transitions between induced host phenotypes. This implies that the host phenotype trait is determined to some extent by *Wolbachia* genetics. *Wolbachia* transition rates from and to a certain phenotype were estimated by two different approaches: Bayesian Monte Carlo Markov Chain estimation and maximum likelihood estimation. The former approach favors a model with no constraints imposed on the rates, whereas the latter approach favors a model where all transition rates between host phenotypes are the same.

Transition models fitted to phylogenetic trees produced the highest likelihood when fitted to a tree based on the concatenated sequences rather than trees based on individual loci. Also, the tree based on the concatenated sequences resolved better three major clades than the trees based on single loci. This validates the use of a multigene alignment to estimate the relatedness between *Wolbachia* strains rather than a single locus approach (Baldo et al., 2006).

Pagel's lambda is a tree transformation that assesses the phylogenetic signal within the studied trait (Pagel, 1999). Likelihood comparisons show that for our available data, the model based on the actual tree (i.e. lambda equals one) is more appropriate than the model based on a star tree (i.e. lambda equals zero). Hence, the induction of a reproductive phenotype can be viewed as a trait of *Wolbachia* that is more likely to differ between distantly related *Wolbachia* strains than closely related strains. However, exceptions are noticeable on the tree where *Wolbachia* strains with very close sequences induce different phenotypes in the host they infect (Figure 1). This suggests that the host genotype can sometimes influence the induced phenotype and these observed exceptions on the tree are due to genetic differences in the hosts. An example of these exceptions is the emergence of resistance in the infected host as is the case for the *Hypolimnas bolina* butterflies where *Wolbachia* can induce cytoplasmic incompatibility if male-killing is suppressed (Hornett et al., 2008).

Our analysis of the phylogenetic signal showed that host phenotypes contain a phylogenetic signal for the genes *gatB*, *hcpA*, *fbpA* and *wsp*. There are two possible explanations for our results. A gene with a phylogenetic signal for host phenotype could either be involved in the induction of a reproductive host phenotype or linked to genes involved in the induction of a host phenotype. The genes that show a phylogenetic signal in our analysis (*gatB*, *hcpA*, *fbpA* and *wsp*) have been described in the literature as house-keeping genes (Baldo et al., 2006; Braig et al.,1998; Hotopp et al., 2006; Foster et al., 2005; Tatusov et al., 2003; Tatusov et al., 1997). Previous findings suggested that the WSP protein undergoes positive selection in response to the host resistance, and hence is

involved in interactions with the host (Endo et al.,1996). However, none of the genes with phylogenetic signal of host phenotype has been reported to play a role in inducing a reproductive phenotype upon host infection. Therefore the most likely explanation for a stronger phylogenetic signal of host phenotype is that these genes are linked to genes involved in the induction of a phenotype.

Bayes factor values from our analysis suggest an asymmetry among the rates of switching between phenotypes and favor the most complex model, i.e. the model with the highest number of free parameters (number of free parameters = 20). This is conflict with the posterior distributions of transition rates which suggest for the binary case that the transition rates to and from CI are equal (Figure 9) and with results from AIC comparisons which favor a single-rate model for the binary and multi-state case (Table 3). Hence, our results suggest that the Bayes factors wrongly favor the most complex model. Previous studies have shown that the harmonic mean approximation for Bayes factors tends to overestimate the marginal likelihood and biases the results towards models with the highest number of parameters (Lartillot & Philippe, 2006). We therefore believe that the conflict between the Bayesian and the maximum likelihood analysis is due to inaccuracies in the harmonic mean approximation, and that the data do not show evidence for uneven transition rates between different host phenotypes.

The most commonly described phenotype induced by *Wolbachia* in arthropods is cytoplasmic incompatibility (Werren et al. 2008) and the same is true for the dataset analyzed here. Since the ability to induce cytoplasmic incompatibility is found in different *Wolbachia* lineages it has been suggested that this ability is ancestral (Rousset et al., 1992).

We estimated probabilities for ancestral states and found that in supergroup A cytoplasmic incompatibility is the most likely ancestral state but in supergroup B all phenotypes are equally likely ancestral states (Table 5). Hence, cytoplasmic incompatibility might not have been the ancestral state for all supergroups. More data are necessary to reconstruct the evolutionary history of host phenotype induction in different supergroups.

In conclusion, the phylogenetic signal of host phenotypes on *Wolbachia* phylogenies justifies a comparative analysis of host phenotypes as a trait of *Wolbachia*. This comparative analysis suggests that transitions between different host phenotypes occur at similar rates and that cytoplasmic incompatibility is the most likely ancestral state for supergroup A but not B. A more detailed reconstruction of the evolution of host phenotypes induced by *Wolbachia* will require a more systematic sampling regime that provides sequence and host phenotype information for a wider range of *Wolbachia* strains.

Bibliography

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, 267–281. http://doi.org/10.1016/j.econlet.2011.12.027

Atyame, C. M., Labb, P., Dumas, E., Milesi, P., Charlat, S., Fort, P., & Weill, M. (2014). Wolbachia divergence and the evolution of cytoplasmic incompatibility in Culex pipiens. *PLoS ONE*, *9*(1). http://doi.org/10.1371/journal.pone.0087336

Augustinos, A. A., Santos-Garcia, D., Dionyssopoulou, E., Moreira, M., Papapanagiotou, A., Scarvelakis, M., … Bourtzis, K. (2011). Detection and characterization of Wolbachia infections in natural populations of Aphids: Is the hidden diversity fully unraveled? *PLoS ONE*, *6*(12). http://doi.org/10.1371/journal.pone.0028695

Baldo, L., Hotopp, J. C. D., Jolley, K. A., Bordenstein, S. R., Biber, S. A., Choudhury, R. R., … Werren, J. H. (2006). Multilocus sequence typing system for the endosymbiont Wolbachia pipientis. *Applied and Environmental Microbiology*, *72*(11), 7098–7110. http://doi.org/10.1128/AEM.00731-06

Baldo, L., & Werren, J. H. (2007). Revisiting Wolbachia supergroup typing based on WSP: Spurious lineages and discordance with MLST. *Current Microbiology*, *55*(1), 81–87. http://doi.org/10.1007/s00284-007-0055-8

Bandi, C., Anderson, T. J., Genchi, C., & Blaxter, M. L. (1998). Phylogeny of Wolbachia in filarial nematodes. *Proceedings of The Royal Society London. Biological Sciences*, *265*(September), 2407–2413. http://doi.org/10.1098/rspb.1998.0591

Beukeboom, L. W. (2012). Microbial manipulation of host sex determination: Endosymbiotic bacteria can directly manipulate their host's sex determination towards the production of female offspring. *BioEssays*, *34*(6), 484–488. http://doi.org/10.1002/bies.201100192

Bing, X. L., Xia, W. Q., Gui, J. D., Yan, G. H., Wang, X. W., & Liu, S. S. (2014). Diversity and evolution of the Wolbachia endosymbionts of Bemisia (Hemiptera: Aleyrodidae) whiteflies. *Ecology and Evolution*, *4*(13), 2714–2737. http://doi.org/10.1002/ece3.1126

Boyle, L., O'Neill, S. L., Robertson, H. M., & Karr, T. L. (1993). Interspecific and intraspecific horizontal transfer of Wolbachia in Drosophila. *Science (New York, N.Y.)*, *260*(5115), 1796–1799. http://doi.org/10.1126/science.8511587

Braig, H. R., Guzman, H., Tesh, R. B., & O'Neill, S. L. (1994). Replacement of the natural Wolbachia symbiont of Drosophila simulans with a mosquito counterpart. *Nature*, *367*(6462), 453–5. http://doi.org/10.1038/367453a0

Braig, H. R., Zhou, W., Dobson, S. L., & O'Neill, S. L. (1998). Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont _Wolbachia pipientis_. *J Bacteriol*, *180*(9), 2373–2378. Retrieved from file:///Y:/16482.pdf

Charlat, S., Riegler, M., Baures, I., Poinsot, D., Stauffer, C., & Merçot, H. (2004). Incipient evolution of Wolbachia compatibility types. *Evolution; International Journal of Organic Evolution*, *58*(9), 1901–1908. http://doi.org/10.1111/j.0014-3820.2004.tb00478.x

Clancy, D. J., & Hoffmann, A. A. (1997). Behavior of Wolbachia endosymbionts from Drosophila simulans in Drosophila serrata, a novel host. *Am Nat*, *149*, 975–988. http://doi.org/10.1086/286033

Cordaux, R., Bouchon, D., & Grève, P. (2011). The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends in Genetics*. http://doi.org/10.1016/j.tig.2011.05.002

Dunning Hotopp, J. C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S. V., Eisen, J., … Tettelin, H. (2006). Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genetics*, *2*(2), 208–223. http://doi.org/10.1371/journal.pgen.0020021

Duron, O., Bouchon, D., Boutin, S. S. S., Bellamy, L., Zhou, L., Engelstadter, J., … Hurst, G. D. (2008). The diversity of reproductive parasites among arthropods: Wolbachia do not walk alone. *BMC Biology*, *6*, 27. http://doi.org/10.1186/1741-7007-6-27

Endo, T., Ikeo, K., & Gojobori, T. (1996). Large-scale search for genes on which positive selection may operate. *Molecular Biology and Evolution*, *13*(5), 685–690. http://doi.org/10.1093/oxfordjournals.molbev.a025629

Foster, J., Ganatra, M., Kamal, I., Ware, J., Makarova, K., Ivanova, N., … Slatko, B. (2005). The Wolbachia genome of Brugia malayi: Endosymbiont evolution within a human pathogenic nematode. *PLoS Biology*, *3*(4), 0599–0614. http://doi.org/10.1371/journal.pbio.0030121

Fukui, T., Kawamoto, M., Shoji, K., Kiuchi, T., Sugano, S., Shimada, T., … Katsuma, S. (2015). The Endosymbiotic Bacterium Wolbachia Selectively Kills Male Hosts by Targeting the Masculinizing Gene. *PLoS Pathogens*, *11*(7). http://doi.org/10.1371/journal.ppat.1005048

Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. *Contemporary Bayesian Econometrics and Statistics*. http://doi.org/10.1002/0471744735

Glowska, E., Dragun-Damian, A., Dabert, M., & Gerth, M. (2015). New Wolbachia supergroups detected in quill mites (Acari: Syringophilidae). *Infection, Genetics and Evolution*, *30*, 140–146. http://doi.org/10.1016/j.meegid.2014.12.019

Grenier, S., Pintureau, B., Heddi, A., Re, F. L., Jager, C., Louis, C., & Khatchadourian, C. (1998). Successful horizontal transfer of Wolbachia symbionts between Trichogramma wasps. *Proceedings of the Royal Society B-Biological Sciences*, *265*(1404), 1441–1445. http://doi.org/10.1098/rspb.1998.0455

Haegeman, A., Vanholme, B., Jacob, J., Vandekerckhove, T. T. M., Claeys, M., Borgonie, G., & Gheysen, G. (2009). An endosymbiotic bacterium in a plant-parasitic nematode: Member of a new Wolbachia supergroup. *International Journal for Parasitology*, *39*(9), 1045–1054. http://doi.org/10.1016/j.ijpara.2009.01.006

Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E., & Challenger, W. (2008). GEIGER: Investigating evolutionary radiations. *Bioinformatics*, *24*(1), 129–131. http://doi.org/10.1093/bioinformatics/btm538

Hatcher, M. J., Taneyhill, D. E., Dunn, A. M., & Tofts, C. (1999). Population dynamics under parasitic sex ratio distortion. *Theoretical Population Biology*, *56*(1), 11–28. http://doi.org/10.1006/tpbi.1998.1410

Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A., & Werren, J. H. (2008). How many species are infected with Wolbachia? - A statistical analysis of current data. *FEMS Microbiology Letters*, *281*(2), 215–220. http://doi.org/10.1111/j.1574-6968.2008.01110.x

Hoerauf, A., Mand, S., Fischer, K., Kruppa, T., Marfo-Debrekyei, Y., Debrah, A. Y., … Büttner, D. W. (2003). Doxycycline as a novel strategy against bancroftian filariasis - Depletion of Wolbachia endosymbionts from Wuchereria bancrofti and stop of microfilaria production. *Medical Microbiology and Immunology*, *192*(4), 211–216. http://doi.org/10.1007/s00430-002-0174-6

Hoffmann, A. A., Turelli, M., & Harshman, L. G. (1990). Factors affecting the distribution of cytoplasmic incompatibility in Drosophila simulans. *Genetics*, *126*(4), 933–948.

Hornett, E. A., Duplouy, A. M. R., Davies, N., Roderick, G. K., Wedell, N., Hurst, G. D. D., & Charlat, S. (2008). You can't keep a good parasite down: Evolution of a male-killer suppressor uncovers cytoplasmic incompatibility. *Evolution*, *62*(5), 1258–1263. http://doi.org/10.1111/j.1558-5646.2008.00353.x

Hu, S. (1987). Akaike information criterion statistics. *Mathematics and Computers in Simulation*, *29*(5), 452. http://doi.org/10.1016/0378-4754(87)90094-2

Hurst, G. D., & Werren, J. H. (2001). The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews. Genetics*, *2*(8), 597–606. http://doi.org/10.1038/35084545

Iturbe-Ormaetxe, I., Walker, T., & O' Neill, S. L. (2011). Wolbachia and the biological control of mosquito-borne disease. *EMBO Reports*, *12*(6), 508–518. http://doi.org/10.1038/embor.2011.84

Jaenike, J. (2007). Spontaneous emergence of a new Wolbachia phenotype. *Evolution*, *61*(9), 2244–2252. http://doi.org/10.1111/j.1558-5646.2007.00180.x

Jiggins, C., Bentley, J., Majerus, M., & Hurst, G. (2002). Recent changes in phenotype and patterns of host specialisation in *Wolbachia* bacteria, 1275–1283. Retrieved from http://eprints.ucl.ac.uk/23216/

Kageyama, D., Narita, S., & Watanabe, M. (2012). Insect sex determination manipulated by their endosymbionts: Incidences, mechanisms and implications. *Insects*. http://doi.org/10.3390/insects3010161

Kass, R., & Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*, 773–795.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. http://doi.org/10.1093/molbev/mst010

Lartillot, N., & Philippe, H. (2006). Computing Bayes Factors Using Thermodynamic Integration. *Systematic Biology*, *55*(2), 195–207. http://doi.org/10.1080/10635150500433722

Ma, W. J., Vavre, F., & Beukeboom, L. W. (2014). Manipulation of arthropod sex determination by endosymbionts: Diversity and molecular mechanisms. *Sexual Development*, *8*(1-3), 59–73. http://doi.org/10.1159/000357024

O ’neill, S. L., Giordano, R., Colbert, A. M. E., Karrf, T. L., & Robertson, H. M. (1992). 6S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Microbiology*, *89*(April), 2699–2702. http://doi.org/10.1073/pnas.89.7.2699

O’Neill, S. L., Giordano, R., Colbert, a M., Karr, T. L., & Robertson, H. M. (1992). 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(7), 2699–2702. http://doi.org/10.1073/pnas.89.7.2699

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. http://doi.org/10.1038/44766

Pagel, M., Meade, A., & Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology*, *53*(5), 673–684. http://doi.org/10.1080/10635150490522232

Partridge, L., & Hurst, L. D. (1998). Sex and conflict. *Science*, *281*, 2003–2008. http://doi.org/10.1126/science.281.5385.2003

Poinsot, D., Bourtzis, K., Markakis, G., Savakis, C., & Merçot, H. (1998). Wolbachia transfer from Drosophila melanogaster into D. simulans: Host effect and cytoplasmic incompatibility relationships. *Genetics*, *150*(1), 227–237.

R Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Revell, L. J., Harmon, L. J., & Collar, D. C. (2008). Phylogenetic signal, evolutionary process, and rate. *Systematic Biology*, *57*(4), 591–601. http://doi.org/10.1080/10635150802302427

Riegler, M., Charlat, S., Stauffer, C., & Merçot, H. (2004). Wolbachia Transfer from Rhagoletis cerasi to Drosophila simulans: Investigating the Outcomes of Host-Symbiont Coevolution. *Applied and Environmental Microbiology*, *70*(1), 273–279. http://doi.org/10.1128/AEM.70.1.273-279.2004

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., … Huelsenbeck, J. P. (2012). Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542. http://doi.org/10.1093/sysbio/sys029

Ros, V. I. D., Fleming, V. M., Feil, E. J., & Breeuwer, J. A. J. (2009). How diverse is the genus Wolbachia? Multiple-gene sequencing reveals a putatively new Wolbachia supergroup recovered from spider mites (Acari: Tetranychidae). *Applied and Environmental Microbiology*, *75*(4), 1036–1043. http://doi.org/10.1128/AEM.01109-08

Rousset, F., Bouchon, D., Pintureau, B., Juchault, P., & Solignac, M. (1992). Wolbachia endosymbionts responsible for various alterations of sexuality in arthropods. *Proceedings. Biological Sciences / The Royal Society*, *250*(1328), 91–98. http://doi.org/10.1098/rspb.1992.0135

Ruang-Areerate, T., & Kittayapong, P. (2006). Wolbachia transinfection in Aedes aegypti: a potential gene driver of dengue vectors. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(33), 12534–12539. http://doi.org/10.1073/pnas.0508879103

Slatko, B. E., Taylor, M. J., & Foster, J. M. (2010). The Wolbachia endosymbiont as an anti-filarial nematode target. *Symbiosis*, *51*(1), 55–65. http://doi.org/10.1007/s13199-010-0067-1

Stouthamer, R., Russell, J. E., Vavre, F., & Nunney, L. (2010). Intragenomic conflict in populations infected by Parthenogenesis Inducing Wolbachia ends with irreversible loss of sexual reproduction. *BMC Evol Biol*, *10*, 229. http://doi.org/10.1186/1471-2148-10-229

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V, … Natale, D. A. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, *4*, 41. http://doi.org/10.1186/1471-2105-4-41

Tatusov, R. L., Koonin, E. V, & Lipman, D. J. (1997). A genomic perspective on protein families. *Science*, *278*(5338), 631–637. http://doi.org/10.1126/science.278.5338.631

Wenseleers, T., & Billen, J. (2000). No evidence for Wolbachia-induced parthenogenesis in the social Hymenoptera. *Journal of Evolutionary Biology*, *13*(2), 277–280. http://doi.org/10.1046/j.1420-9101.2000.00168.x

Werren, J. H. (1997). Biology of *Wolbachia*. *Annual Review of Entomology*, *42*(1), 587–609. http://doi.org/10.1146/annurev.ento.42.1.587

Werren, J. H., Baldo, L., & Clark, M. E. (2008). Wolbachia: master manipulators of invertebrate biology. *Nature Rev Microbiol*, *6*(10), 741–751. http://doi.org/10.1038/nrmicro1969

Werren, J. H., Newton, I. C., Kent, M. E., Bordenstein, B. N.,Qu, S. R., Richards J., Kelkar S., Yogeshwar D. (2016). Comparative genomics of two closely related Wolbachia with different reproductive effects on hosts. *Genome Biol Evol*, evw096-. doi:10.1093/gbe/evw096

Werren, J. H., & Beukeboom, L. W. (1998). Sex determination, sex ratios, and genetic conflict. *Annual Review of Ecology and Systematics*, *29*, 233–261. http://doi.org/10.1146/annurev.ecolsys.29.1.233

Zabalou, S., Riegler, M., Theodorakopoulou, M., Stauffer, C., Savakis, C., & Bourtzis, K. (2004). Wolbachia-induced cytoplasmic incompatibility as a means for insect pest population control. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(42), 15042–15045. http://doi.org/10.1073/pnas.0403853101

Zug, R., & Hammerstein, P. (2012). Still a host of hosts for Wolbachia: Analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS ONE*, *7*(6), 7–9. http://doi.org/10.1371/journal.pone.0038544