Nº d'ordre: 35-2012 Année 2012

Thèse

Présentée

devant l'Université Claude Bernard - Lyon 1

pour l'obtention

du Diplôme de Doctorat

(arrêté du 7 août 2006)

et soutenue publiquement le
14 Mars 2012

par

Patrícia Martins Simões

Diversity and dynamics of Wolbachia-host associations in arthropods from the Society archipelago, French Polynesia

---

Directeur de thèse: Marie-France Sagot
Co-Directeur: Sylvain Charlat

Jury:	Christian Gautier,	Examinateur
	Claudio Bandi,	Rapporteur
	Didier Bouchon,	Rapporteur
	Olivier Duron,	Examinateur

UNIVERSITÉ CLAUDE BERNARD-LYON 1

Président de l'Université	**M. le Professeur M. A. BONMARTIN**
Vice-Président du Conseil d'Administration	M. le Professeur G. ANNAT
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. le Professeur D. SIMON
Vice-Président du Conseil Scientifique	M. le Professeur J-F. MORNEX
Secrétaire Général	M. G. GAY

SECTEUR SANTÉ

Composantes

Faculté de Médecin Lyon-Est - Claude Bernard	Directeur: M. le Professeur J. ETIENNE
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur: M. le Professeur F-N. GILLY
UFR d'Ontologie	Directeur: D. BOURGEOIS
Institut des Sciences Pharmaceutiques et Biologiques	Directeur: M. le Professeur F. LOCHER
Institut Techniques de Réadaptation	Directeur: M. le Professeur MATILLON
Département de Formation et Centre de Recherche en Biologie Humaine	Directeur: M. le Professeur P. FARGE

SECTEUR SCIENCES

Composantes

Faculté des Sciences et Technologies	Directeur: M. le Professeur S. De MARCHI
Département Biologie	Directeur: M. le Professeur F. FLEURY
Département Chimie Biochimie	Directeur: Mme. le Professeur H. PARROT
Département Génie Electrique et des Procédés	Directeur: M. N. SIAUVE
Département Informatique	Directeur: M. le Professeur S. AKKOUCHE
Département Mathématiques	Directeur: M. le Professeur A. GOLDMAN
Département Mécanique	Directeur: M. le Professeur H. BEN HADID
Département Physique	Directeur: M. le Professeur S. FLECK
Département Sciences de la Terre des Activités Physiques et Sportives	Directeur: M. le Professeur I. DANIEL
UFR Sciences et Techniques	Directeur: M. C. COLLIGNON
Observatoire de Lyon	Directeur: M. B. GUIDERDONI
Ecole Polytechnique Universitaire de Lyon 1	Directeur: M. P. FOURNIER
Ecole Supérieure de Chimie Physique Electronique	Directeur: M. G. PIGNAULT
Institut Universitaire de Technologie de Lyon 1	Directeur: M. le Professeur C. COULET
Institut de Science Financière et d'Assurances	Directeur: M. le Professeur J. C. AUGROS

Résumé

Certains symbiotes intracellulaires résident dans le cytoplasme des cellules et manipulent le système reproductif de leurs hôtes. Du fait de leur transmission maternelle, ces parasites sont sélectionnés pour optimiser la survie et la reproduction de leur hôtes femelles. Chez les arthropodes, la bactérie *Wolbachia* infecte au moins 66% des espèces d'insectes mais peuvent aussi infecter des nématodes. Cette large distribution dans les populations hôtes confère à *Wolbachia* un potentiel important en tant que moteur d'évolution. En particulier, elle pourrait être utilisée comme vecteur transgène dans les espèces nuisibles. Mais la dynamique évolutive des infections à l'échelle des communautés est mal connue, en particulier la fréquence des transferts de parasites entre hôtes de différentes espèces et la stabilité évolutive des associations.

Mon travail de thèse a porté sur la détection et dynamique des infections de *Wolbachia* à une échelle microevolutive, c'est-à-dire, dans des communautés d'arthropodes avec moins de 5 My. L'objectif de ce travail était à la fois la characterisation des communautés géographiques d'arthropodes et celle des infections par *Wolbachia* de ces communautés. Nous avons également éxaminé l'existence de transferts horizontaux récents de ces symbiotes entre des taxa distantes ainsi que les routes écologiques potentielles pour ces transmissions.

Abstract

Sexual parasites are intracellular symbionts capable of manipulating the reproduction of their hosts. They are widespread in Arthropods where they display a wide range of reproductive manipulations; these can be potentially involved in the evolution of mating systems, speciation, gene acquisition and sex determination. In particular, *Wolbachia* is thought to infect more than 66% of insect species and is also found in nematodes. However, little is known about the dynamics of *Wolbachia* infections at the community level. Although at the intra-population level, invasion dynamics have been extensively studied, the same is not true at the community level, where the turnover of infections remains largely uncharacterised.

How often are new infections acquired through horizontal transfers between distantly related hosts? What ecological routes are behind the movement of these parasites between hosts? These questions, crucial to assess the impact of the infections, remain unanswered. Moreover, as *Wolbachia* is seen as a good candidate for a transgenic vector against pests, understanding its dynamic at the community level is crucial.

We propose to address them by performing an exhaustive characterisation of sexual parasites in simplified systems, using the opportunity offered by small arthropod communities in isolated islands. Our work aims at characterising geographical communities of arthropods collected from 4 young, isolated islands (less than 5 My old) and the *Wolbachia* infections found in these communities. Furthermore, we investigate the existing evidence for recent horizontal transfers of these symbionts between distant taxa and potential ecological routes facilitating these transmissions.

Acknowledgements

First of all I would like to thank my supervisors Marie-France Sagot and Sylvain Charlat for accepting to be my mentors in this long journey. With them I had the opportunity to work on this thesis, the chance to learn and grow as a person and as researcher.

I am also very grateful to Christian Gautier for his ever present support throughout my thesis. Both professionally with his statistical expertise and discussions ranging from ecology, evolution and science in genera and personally.

Many thanks to the pôle administratif: Nathalie Arbasetti, Eloïse Clair, Isabelle Ravis, Aurélie Wurtz and Laetitia Mangeot and the pôle informatique: Stephane Delmotte, Lionel Humblot, Simon Penel and Bruno Spataro.

I would also like to thank all past and present members of both the BAOBAB and GEIHP teams.

And, even though I fear that I am leaving someone out, I would like to thank some very special people. These have been, together with the scientific side, the most important enrichment in my life throughout these years in Lyon.

They are my friends and family.

Without them this project would not have arrived to its successful end. My life would not be as rich and complete as it is now. To Janice Kielbassa, Paulo Milreu, Vincent Lacroix, Camille Degryse, Leonor Palmeira, Linda Artmann, Marie-France Sagot, Christian Gautier, Lauranne Duquenne, Florence Bouheddi, Martin Wannagat, Sophie Abby, Thomas Kämpfe, Adriana Milreu, Carole Forfait, Cécile Gotteland, Émilie Vautrin, Matteo Brilli, Erika Kvikstad, Joanna Parmley, Christian Baudet, Cecília Klein, Vicente Acuña, Paulo Fonseca, Marília Braga, Augusto Vellozo, Pierre-Antoine Farnier, Gwen Gueguen, Natacha Kremer, Hélène Henry. Closer to home there were João Dias, Isabel Zilhão, Ana Madeira, Catarina Raposo, Nuno Sepúlveda, Teresa Maia, Catarina Lemos Varela and Bruno Martins. To my family: my parents, my grandparents, my aunts and uncles and all my cousins. There are no words to thank you for always being there. For always believing in me, for your unconditional love.

The geographical distance was never a barrier. My heart carries you all.

And finally, the constant throughout my life, my first best friend, my confident, my light, my model to follow, my sister Sofia.

" no regresso encontrei aqueles
que haviam estendido o sedento corpo
sobre infindáveis areias

tinham os gestos lentos das feras amansadas
e o mar iluminava-lhes as máscaras
esculpidas pelo dedo errante da noite

prendiam sóis nos cabelos entrançados
lentamente
moldavam o rosto lívido como um osso
mas estavam vivos quando lhes toquei
depois
a solidão transformou-se de novo em dor
e nenhum quis pernoitar na respiração
do lume

ofereci-lhes mel e ensinei-os a escutar
a flor que murcha no estremecer da luz
levei-os comigo
até
onde o perfume insensato de um poema
(...)"

in Os Amigos
Al Berto

Contents

1 Introduction and Motivation **1**

1.1 Sexual Parasites: a brief overview 1

1.2 DNA Barcoding 7

1.3 Questions to be addressed in my PhD project 9

2 Data Acquisition **11**

2.1 Collection, sorting and DNA extractions 12

2.1.1 Collecting sites and specimens 12

2.1.2 Classification and sorting 17

2.1.3 DNA extractions 19

2.2 Sequencing of a molecular marker for the hosts: Cytochrome Oxydase I (COI) 21

2.3 Screening of *Wolbachia* infections 22

2.4 Selection of a marker for *Wolbachia*: Fructose Bisphosphate Aldolase (FbpA 30

2.5 Sequencing and sequence editing 31

2.5.1 Sequencing 31

2.5.2 Sequence, assembly and editing 31

2.5.3 Submission to the Moorea Biocode Project 33

2.6 Alignments and phylogenies 33

2.6.1 Alignments 33

2.6.2 Haplotype groups definition 33

2.6.3 Tree reconstruction method(s) 34

2.6.4 Cluster definition 35

3 Results and Discussion **37**

3.1 Diversity and biogeography of the Host mitochondria 38

3.1.1 Mitochondrial diversity in the Society Archipelago 39

3.1.2 Biogeography of Mitochondrial in the Society Archipelago . . . 48
3.1.3 Discussion 60
3.2 Incidence and diversity of *Wolbachia* infections 65
3.2.1 Incidence 66
3.2.2 Diversity of *Wolbachia* infections as inferred using the FbpA marker 71
3.2.3 Discussion 81
3.3 Horizontal transfers of *Wolbachia* infections 85
3.3.1 Defining groups of closely related *Wolbachia* strains 85
3.3.2 Proportion of *Wolbachia* clades being horizontally transferred between different arthropod orders 87
3.3.3 Host specificity of horizontally transferred *Wolbachia* clades . . 88
3.3.4 Investigating preferential transfers between Orders 89
3.3.5 Exploring the ecological routes of Wolbachia horizontal transfer 94
3.3.6 Discussion 98

4 Conclusions and Perspectives 103
4.1 Diversity and biogeography of the Host mitochondria 103
4.2 Incidence and diversity of *Wolbachia* infections 104
4.3 Horizontal transfers of *Wolbachia* infections 105

Bibliography 107

Appendix 121

Chapter 1

Introduction and Motivation

Contents

1.1 Sexual Parasites: a brief overview 1

1.2 DNA Barcoding 7

1.3 Questions to be addressed in my PhD project 9

1.1 Sexual Parasites: a brief overview

Symbiosis, as defined originally by de Bary (1879), refers to the living together of distinct organisms [1]. Symbiotic relations are widespread in nature, particularly those involving micro-organisms such as bacteria. In such associations, the bacteria form persistent infections with their hosts and can affect, amongst others, their nutrition, reproduction, defense against natural predators and immunity (for a review see [2]). The symbionts can be transmitted from host to host in an infectious (horizontal) way, in a heritable (vertical) way, or by mechanisms resulting from a combination of both.

In animals and insects, most of these symbionts are intracellular – they are then called endosymbionts – and are predominantly transmitted vertically, i.e., they are heritable organisms passed from parent to offspring. The transmission is maternal, usually via the egg's cytoplasm. In insects, endosymbionts are traditionally divided into primary or obligate symbionts, and secondary or facultative symbionts. The former are typically nutritional mutualists occurring in insects with unbalanced diets (an example is *Buchnera aphidicola* that lives inside the cells of aphids [Hemiptera]) while the latter can range from parasites to mutualists [3].

Within the maternally transmitted facultative endosymbionts, we find a group of bacteria which are able to manipulate their host's reproduction to increase their own

fitness. Since they are transmitted via the egg cytoplasm, these microorganisms – also known as reproductive parasites – are positively selected if they have beneficial effects towards the infected females, i.e., the host sex responsible for their transmission. Thus, selection favours reproductive parasites which increase the fitness of the infected females and/or their daughters. Inversely, the phenotypes induced by these parasites can be (very) detrimental to the hosts not involved in their transmission, namely males and uninfected females [4]. In arthropods, reproductive parasites are known to induce diverse phenotypes [1] described next.

1. Cytoplasmic incompatibility (CI): The fitness of uninfected females is decreased due to sperm induced sterility when they cross with infected males.

2. Male-killing (MK): In this case, males (larvae or embryos) are killed because they are dead-ends from the parasite's point of view. The death of males can either facilitate the infectious transmission of the symbiont or the survival (feeding) of their infected sisters, thus increasing the fitness of infected females.

3. Feminisation (F): This leads to an increase in the proportion of infected females because genetic males harbouring the parasite are turned into functional infected females.

4. (Thelytokous) Parthenogenesis (P): This is observed in haplodiploid species where the presence of the parasite in unfertilised eggs (which normally give origin to males) induces the duplication of chromosomes and, consequently, the assexual production of infected females.

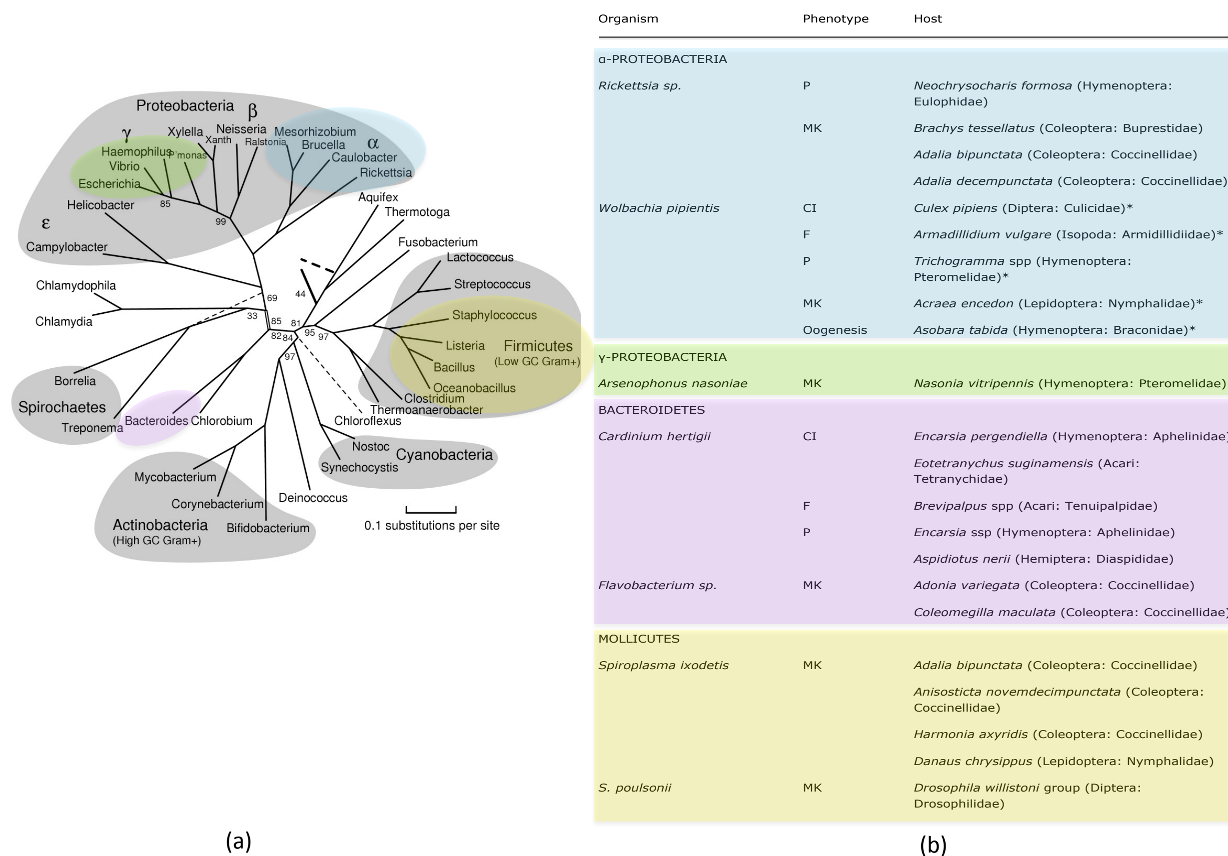
The last three phenotypes (MK, F, and P) all induce a female-sex bias, the most extreme case of which is male-killing (MK).

Bacterial reproductive parasites are known from highly diverse classes and phylum such as α-proteobacteria [5, 6, 7], γ-proteobacteria [8, 9], Mollicutes [10, 11, 12, 9] and Bacteroidetes [13, 14, 9], as can be seen in Figure 1.1. They are not only diverse but they seem also to be widely distributed in arthropods.

Screening studies of *Wolbachia*, an α-proteobacteria from the *Rickettsiaceae* family [1], typically performed with a single (or a few) individual(s) per species, have led to estimates of $\sim 20\%$ of insect species being infected with this parasite, while a more recent meta-analysis [16] has proposed that the global incidence – here seen as the proportion of infected species – might be as high as 67%.

Cases of high prevalence are known in populations of butterflies [17, 18], isopods [19], mosquitos [20], flies [21], beetles [22], ants [23] and parasitic wasps [24, 25]. More-

Figure 1.1: Distribution of known bacterial reproductive parasites in the bacteria phylogeny as adapted from [15]. In (a) and in (b) a summary of the arthropod hosts and induced phenotypes for some of these endosymbionts as adapted from [9].



Organism	Phenotype	Host
α-PROTEOBACTERIA		
Rickettsia sp.	P	*Neochrysocharis formosa* (Hymenoptera: Eulophidae)
	MK	*Brachys tessellatus* (Coleoptera: Buprestidae)
		Adalia bipunctata (Coleoptera: Coccinellidae)
		Adalia decempunctata (Coleoptera: Coccinellidae)
Wolbachia pipientis	CI	*Culex pipiens* (Diptera: Culicidae)*
	F	*Armadillidium vulgare* (Isopoda: Armidillidiidae)*
	P	*Trichogramma* spp (Hymenoptera: Pteromelidae)*
	MK	*Acraea encedon* (Lepidoptera: Nymphalidae)*
	Oogenesis	*Asobara tabida* (Hymenoptera: Braconidae)*
γ-PROTEOBACTERIA		
Arsenophonus nasoniae	MK	*Nasonia vitripennis* (Hymenoptera: Pteromelidae)
BACTEROIDETES		
Cardinium hertigii	CI	*Encarsia pergendiella* (Hymenoptera: Aphelinidae)
		Eotetranychus suginamensis (Acari: Tetranychidae)
	F	*Brevipalpus* spp (Acari: Tenuipalpidae)
	P	*Encarsia* ssp (Hymenoptera: Aphelinidae)
		Aspidiotus nerii (Hemiptera: Diaspididae)
Flavobacterium sp.	MK	*Adonia variegata* (Coleoptera: Coccinellidae)
		Coleomegilla maculata (Coleoptera: Coccinellidae)
MOLLICUTES		
Spiroplasma ixodetis	MK	*Adalia bipunctata* (Coleoptera: Coccinellidae)
		Anisosticta novemdecimpunctata (Coleoptera: Coccinellidae)
		Harmonia axyridis (Coleoptera: Coccinellidae)
		Danaus chrysippus (Lepidoptera: Nymphalidae)
S. poulsonii	MK	*Drosophila willistoni* group (Diptera: Drosophilidae)

(b)

over, *Wolbachia* has been found to infect almost all arthropod Orders, and also nematodes where they are restricted to the *Onchocercidade* family (encompassing the agents of filariases [26]) and are mutualistic symbionts.

The invasive capacity offered by the different induced-phenotypes (CI, MK, P and F) and seen in various studies on *Wolbachia*, indicate that this wide range of reproductive manipulation systems enables *Wolbachia* to be potentially involved in the evolution of mating systems, speciation, gene acquisition and sex determination (for a review see [27]). As a consequence, sexual parasites can be seen as an important evolutionary force [28, 27, 29, 17]. Until recently, work in the field of reproductive parasitism was biased towards *Wolbachia* which has led to estimations of a low incidence (4-7%) for the remaining endosymbionts [9]. The potential evolutionary impact of these symbionts should however not be underestimated.

Diversity and phylogeny of the Wolbachia genus The bacterium identified as *Wolbachia pipientis* was first described by Hertig (1936) in the mosquito *Culex pipiens*, where it was mainly found in the cytoplasm of the cells present in reproductive organs. In the following years, various studies on arthropods showed that these endosymbionts have a preferential tropism for testes or ovaries (reproductive tissues) although in some associations they have been found in other somatic tissues such as the nervous system or the hemolymph [30, 31]. In filarial nematodes where only mutualistic associations have been characterised, *Wolbachia* strains are typically found in the female germline and in the hypodermal lateral chords, although recently infections have also been reported in other somatic tissues [32].

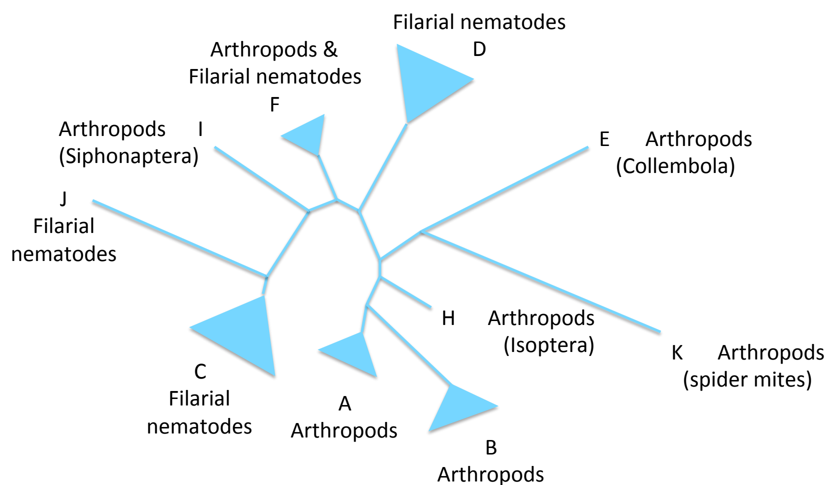
The earliest phylogenies – based on the 16S rDNA [1] sequences – suggested that *W. pipientis* formed a monophyletic clade within the α-proteobacteria, with the genera *Anaplasma*, *Rickettsia*, *Orientia* and *Ehrlichia* as closest relatives (sister-groups). In the last decades, many bacteria phylogenetically very closely related to *W. pipientis* have been found – both in arthropods and in some nematodes – that have further confirmed the monophyly of the *Wolbachia* genus. Despite the fact that the intra-genus divergence at the level of the 16S rDNA *locus* is superior to 3%, the threshold usually accepted as the maximum divergence within bacterial species, no new species have been recognised in the *Wolbachia* genus so far.

Other phylogenetic studies using the more variable *ftsZ* gene (cell division protein) revealed distinct bacterial lineages both within and among arthropods and nematodes. Two divergent groups of *Wolbachia* reproductive parasites were identified in arthropods – and called the A and B supergroups [33] – while two other divergent groups of mutualistic *Wolbachia* infections were found in nematodes – and called the C and D supergroups [26]. The divergence of the A and B groups was then estimated as having occurred $\sim$ 60 My ago [33] and their separation from the mutualistic groups in nematodes as being $\sim$ 100 My [26] old. Since these first descriptions of *Wolbachia* supergroups, the interest in these symbionts has led to the investigation of a wider range of host taxa, also with improved PCR based detection techniques and sequencing. More recently, other supergroups have thus been recognised using data from multiple genes [34, 35, 36, 37].

In total, 10 *Wolbachia* supergroups are validated today – ranging from A to K, with the exception of the previously detected G supergroup – with most found either in arthropods or nematodes [32], see Figure 1.2. In arthropods, the presence of the A and B supergroups has been confirmed in almost all insect taxa, spiders, mites, scorpions and terrestrial crustaceans, while additional supergroups have been found

in more restricted host ranges. These include the supergroup E found in springtails (Collembola), the supergroup H found in termites (Isoptera), the supergroup I found in fleas (Siphonaptera), and the supergroup K found in spider mites (Tetranychidae) (see [32, 36] and references therein). In filarial nematodes, only one new supergroup has been recognised – the J supergroup (Spirulida) – although a further supergroup may have been identified for the first time in a plant parasitic nematode [38]. Only one supergroup has been found to infect both host taxa, this is the supergroup F. Initially discovered in termites, it is now known to be present in other insect orders – such as Blattaria, Coleoptera, Dipetra, Hemiptera, Orthorptera and Phthiraptera –, in scorpions and in human filarial of the *Mansonella* genus (see [36] and references therein).

Figure 1.2: Schematic representation of *Wolbachia* phylogeny with the 10 validated supergroups, adapted from [36].



Horizontal transfer of Wolbachia Until recently, both horizontal transfer (HT) and recombination events have been considered to be absent in or between nematode-infecting *Wolbachia* clades. Although a recent work by Ros and colleagues [36] has suggested a recombination between the C and D supergroups, it has been argued that this is likely due to a hybrid sequence [39]. So far, no other studies have been able to detect recombination in filarial nematodes [40, 41, 42]. Phylogenies of the nematode hosts and their associated *Wolbachia* are very congruent, suggesting they have been co-evolving with long, stable associations [43]. Notwithstanding, evidence for a recent HT event of an F supergroup infection has been reported in filariae [32]. Interestingly, this is, as mentioned, the sole supergroup whose host range encompasses both filarial

nematodes and arthropods.

Inversely, evidence for horizontal transfer and recombination between the arthropod-infecting supergroups A and B are common [44, 45, 46]. Recombination within and between these supergroups has been reported with housekeeping genes, surface protein coding genes (*wsp*), phages and transposable elements (IS) (see [45] and references therein). Moreover, a number of arthropod species have been shown to harbour multiple infections, in some cases from the same supergroup or from both the A and B supergroups [47, 48, 49, 25], providing a means of gene flow. The *Wolbachia* genus, specifically the arthropod supergroups which are associated with reproductive manipulation, seems to be an exception within vertically transmitted endosymbionts, with extensive recombination occurring both intra- and inter-genically. It has been suggested that recombination may be very important for the survival and adaptation of *Wolbachia* to a wide range of arthropod hosts, i.e., for their being extremely generalist in host usage and establishing a predominant parasitic relation with their hosts. A multilocus sequencing typing (MLST) protocol has been recently proposed – using 5 housekeeping genes: *ftsZ, gatB, fbpA, coxA* and *hcpA* – to assess strain diversity within host species and identify closely related *Wolbachia* strains in distinct arthropod hosts [44].

Despite frequent recombination and the confounding effect it has on phylogenetic reconstructions, evidence for horizontal transfer of the supergroups A and B strains is abundant in the literature. This can be seen through the incongruence between host and parasite phylogenies with distantly related arthropod host species harbouring very similar *Wolbachia* infections [50, 51, 52, 44, 53, 54], and close host species harbouring parasites from different supergroups [55, 56, 49, 57]. Some studies have started to focus on the ecological routes for *Wolbachia* host shifting. Interactions such as host-parasitoid [51, 58, 59], predator-prey [52], sharing of common food sources [60, 53, 54], social parasitism [25] and exchange of haemolymph [61] are thought to mediate the movement of *Wolbachia* parasites between distant hosts with similar ecology.

However, it has also been shown in various arthropod groups – such as *Agenelopsis* spiders [62], *Acraea* butterflies [40], *Armadillidium* isopods [63, 52] and in the parasitic wasp *Trichogramma* [64] – that closely related *Wolbachia* strains are often found in closely related host species. Once more, based on the incongruence between host and parasite phylogenies, this specialisation is believed to be due to horizontal transfers occurring preferentially (or more frequently) between closely related species rather than more distantly related ones. It is thought that this preferential movement between close hosts may be due to some degree of physiological specialisation exhibited by the

bacteria.

It remains unclear what is the extent to which the ecological component counterbalances the phylogenetic relatedness of hosts in the horizontal transfer of *Wolbachia*.

1.2 DNA Barcoding

In animals, the mitochondrial genome (mtDNA) is a small ($\sim$ 1.6x10^4 bp) haploid, circular DNA molecule. It contains few genes (37) with no known introns or repetitive sequences and only a few noncoding regions [65]. Gene transfers from the mtDNA to the nuclear genome (ncDNA) – nuclear mitochondrial (numt) – have been identified in many species (for a review, see [66]) and are believed to be part of an ongoing process. It is thought that a major factor preventing further functional gene transfer in animals is probably the non-standard genetic code of mtDNA [67]. Due to its cytoplasmic location, mtDNA is almost exclusively maternally inherited, with no or almost no recombination [68]. Thus, the effective population size is a quarter of that of the nuclear autosomal DNA, and natural selection is thought to be weaker, as it acts on a single selective unit. Mutation rate in the mtDNA is on average higher than that observed in ncDNA [69]. In Drosophila, the synonymous substitution rate in mitochondrial genes is 4.5 to 9.0 times higher than that of average nuclear genes [70].

These unique characteristics of the mtDNA – high mutation rate, haploidy, lack of recombination and almost neutral selection – have been crucial for the establishment of mtDNA as a popular genetic marker for the study of species evolution and population genetics. Though some issues concerning problems of non-neutrality [71], recombination and introgression [72] have started to appear recently, mitochondrial (mt) protein coding genes continue to be extensively used as good markers to identify species.

In 2003, Herbert and colleagues [73] proposed a DNA Barcoding approach whose aim was to develop a standardised, rapid and inexpensive animal species identification. This approach considers the use of one (or of a few) gene(s) to identify specimens to the species level as well as the discovery of new species and assignation of unknown specimens to species groups thanks to the development of comprehensive sequence databases [73, 74, 75].

Due to its unique characteristics, mt protein coding genes have become the "ideal" candidate markers for DNA Barcoding. In animals, DNA barcode itself refers to a 648 bp region from the 5'-end of the cytochrome c oxidase 1 (COI) gene – the *cox locus* – and its interest is based on two assumptions: 1) it is very likely that every species will have a unique DNA barcode; 2) genetic variation between species exceeds the variation

within a species [73, 76].

To date, the COI gene has proved to be suitable for the identification of a large range of animal taxa (see [76] and references therein). The DNA Barcode of Life Data System (BOLD, http://www.boldsystems.org) has been officially established in 2007 and is a data system enabling the acquisition, storage, analysis and publication of DNA barcode records. By early 2012, $\sim 1.48\text{x}10^5$ species have been formally described and $\sim 1.5\text{x}10^6$ barcodes have been deposited in this database.

In the last years, various more focused databases have appeared connected with BOLD. These smaller DNA barcode databases are focused on specific taxa – e.g. fish (Fishbol), ants (formicidaebol) or butterflies and moths (lepbarcoding) – or ecosystems – e.g. the Moorea tropical islands (mooreabiocode) and the polar regions (polarbarcoding).

Notwithstanding the promises of this global barcoding project, there are some crucial pitfalls that should be kept in mind (for an in-depth discussion and references see [76]).

1. Sampling shortage across taxa: robust specimens assignation is only possible for clades whose taxonomy is well understood and representative specimens thoroughly sampled.

2. Mitochondrial (maternal) inheritance and sexual parasites: indirect selection on mitochondrial DNA arises from linkage disequilibria with endosymbionts and this can be very problematic in arthropods where (as seen above) the presence of *Wolbachia* is estimated to be wide. Moreover, other symbionts are known to interfere with arthropod reproduction and their incidence is only now starting to be assessed.

3. Nuclear copies of the COI gene: these can act as confounding factors in the identification of specimens. This problem has already been reported in primates and some marine zooplankton.

4. Heterogeneous rate of evolution of the *cox* locus: this represents another confounding factor leading to a lack of resolution for some taxa. Reports of different evolutionary rates are known for molluscs, cnidarians and six dermapteran species.

5. Geographical structure: high rates of intraspecific divergence can derive from geographically isolated populations, as indicated in [73].

1.3 Questions to be addressed in my PhD project

Islands are isolated systems, i.e., they are biological systems where contact with other communities is either inexistent or very limited for an extended period of time. Such systems can be actual islands, small areas of land completely surrounded by water, or any ecosystem or favourable habitat surrounded by different ecosystems or unsuitable habitats [77].

Remote island systems that are both oceanic and volcanic – such as the Society Archipelago (French Polynesia) or the Hawaiian Islands – are an example of *de novo* islands with a high degree of isolation. They were formed as the Pacific plate moved over a "hot spot" – corresponding to fixed intrusions in the earth's mantle where hot rock from the core-mantle boundary is expelled from the crust surface [78]. These systems have been viewed as "natural laboratories" for the study of evolution owing to some particular geographical and ecological characteristics [79]: 1) they constitute well-defined, discrete geographic entities; 2) gene flow between islands is reduced due to the oceanic boundaries; 3) often, they are smaller land masses allowing for a more extensive characterisation of their flora and fauna; 4) these systems are both enriched in terms of habitats (novel and very specific ecological niches) and species diversity (high endemism and species radiation).

In 2000, Werren and Windsor proposed [47] that *Wolbachia* infections might be at a global equilibrium. Although extensive studies on the invasion dynamics of sexual parasites exist at an intra-populational level, much remains to be known about their dynamics at a micro-evolutionary time-scale where communities are concerned. Recently, more studies have started to focus on the transfer of *Wolbachia* in specific communities, e.g., of Drosophila-hymenopteran parasitoids [51], mushroom-feeding dipterans [54], rice-field insects [60], and pumpkin-leaf feeding arthropods [53]. However, few or no studies have focused on insular ecosystems.

Using an extensive sampling of arthropods performed in four islands – less than 3My old – of the Society Archipelago (French Polynesia), we were interested in characterising both the diversity of arthropods found in these islands and their *Wolbachia* infections using the CO1 gene for the hosts and the FbpA (Fructose Biphosphate Aldolase) marker for *Wolbachia*. We wanted also to assess the distribution of these infections in major taxa and look for evidence of horizontal transmission of *Wolbachia* between distantly related arthropod hosts using a "top-down" approach, i.e., working not with one ecological community – classically defined as a group of populations from different species which co-exist (and interact) in the same space – but with an ensemble of arthropod communities collected in different geographical locations. Finally, we wished

to infer possible ecological connections, i.e., possible routes for the transmission of *Wolbachia* infections.

Chapter 2

Data Acquisition

Contents

2.1 Collection, sorting and DNA extractions 12

2.1.1 Collecting sites and specimens 12

2.1.2 Classification and sorting 17

2.1.3 DNA extractions 19

2.2 Sequencing of a molecular marker for the hosts: Cytochrome Oxydase I (COI) 21

2.3 Screening of *Wolbachia* infections 22

2.4 Selection of a marker for *Wolbachia*: Fructose Bisphosphate Aldolase (FbpA 30

2.5 Sequencing and sequence editing 31

2.5.1 Sequencing 31

2.5.2 Sequence, assembly and editing 31

2.5.3 Submission to the Moorea Biocode Project 33

2.6 Alignments and phylogenies 33

2.6.1 Alignments 33

2.6.2 Haplotype groups definition 33

2.6.3 Tree reconstruction method(s) 34

2.6.4 Cluster definition 35

2.1 Collection, sorting and DNA extractions

The data collected for the present study is the result of an extensive fieldwork performed during the period from March 2005 to July 2007 in the Society Archipelago (French Polynesia) as part of the Moorea Biocode project. This is an international project that aims to make a comprehensive assessment of non-microbial life in a tropical ecosystem. In this chapter, we present the characteristics of our collecting sites, specimens acquisition and processing. We would like to thank S. Charlat, M. Saillan, A. Duplouy and C. Vermenot who performed the collecting work and, also S. Martinez (sorting), G. Mialdea (sorting and DNA extraction) and D. Reiss (DNA extraction) for their help during this phase. Finally, this project also benefited from the collaboration of the entomological group of Paopao's College.

2.1.1 Collecting sites and specimens

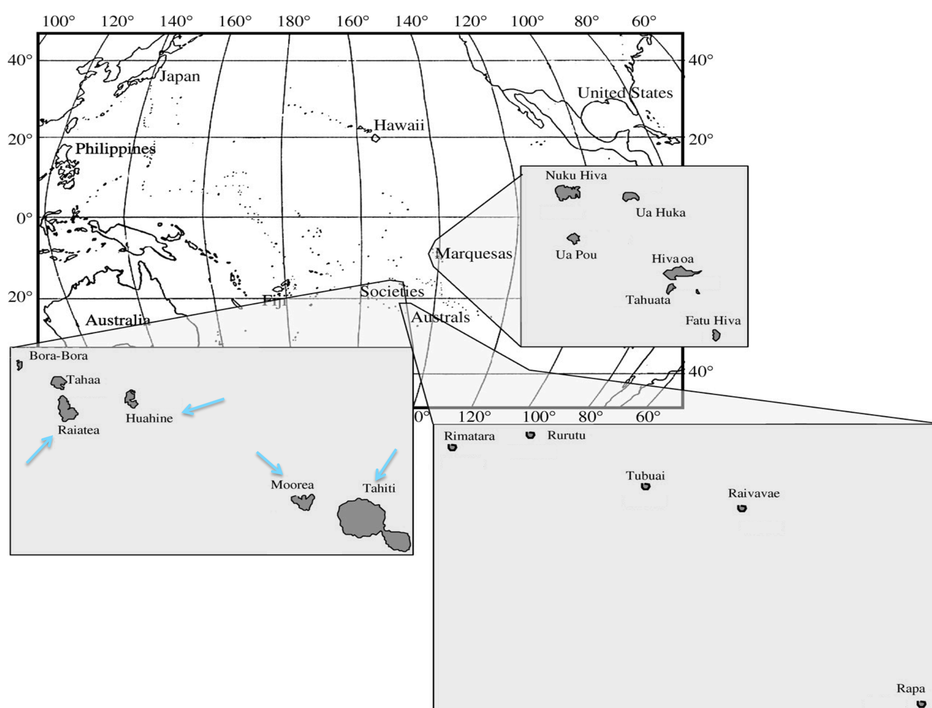
The Society Archipelago, thought to have originated from the Tahiti/Society hotspot [78] on the South Pacific plate, is formed by a chain of 14 islands arranged into 2 geographical groups: Winward and Leeward. The islands exhibit a westwards age progression with the oldest, Maupiti (Winward group), estimated to have been formed less than 5 My [80]. Arthropod communities used in the present study were sampled from the two major islands of both groups: Raiatea and Huahine, from the western Winward group; Moorea and Tahiti from the eastern Leeward group (see Figure 2.1).

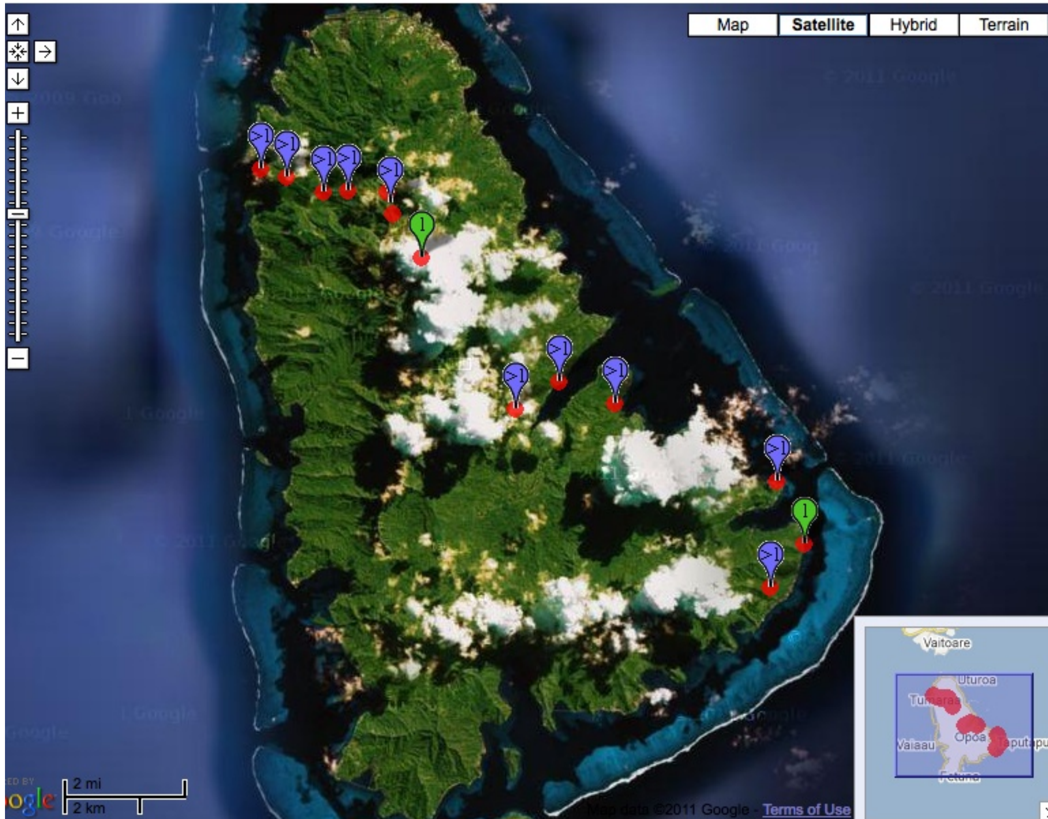
Collecting sites within the islands, although influenced by accessibility, included both low and high altitude positions so that individuals from environments that are, respectively, more disturbed notably by human presence and activities, and others that are less disturbed and where potentially endemic species may be found, could be represented in our dataset (see Figure 2.2 for the collecting sites distribution within islands).

To maximise the number and diversity of the arthropods collected, a variety of methods (see Table 2.1) were used which consisted in:

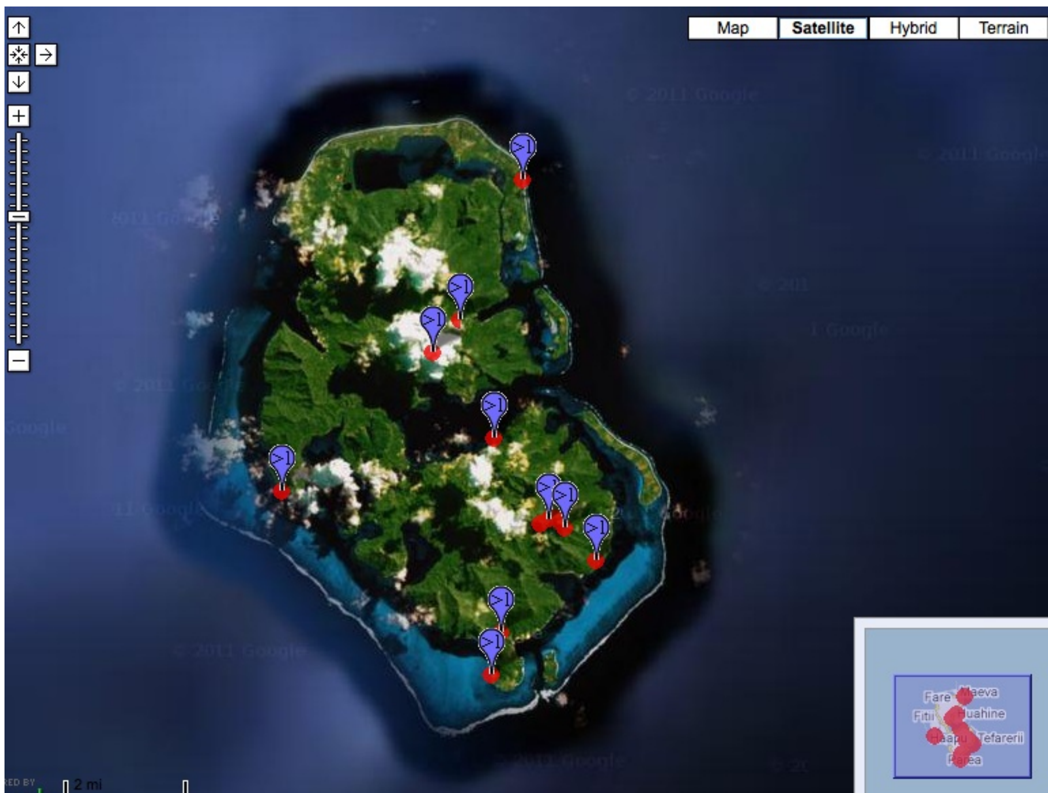
1. Sweeping nets, which are particularly successful with flies (Diptera), grasshoppers (Orthopetra), beetles and bugs (Coleoptera), and Hymenopterans which sit high on vegetation or do not fall off when disturbed and may be poor fliers [82], see Figure 2.3.

Figure 2.1: The main high archipelagos in French Polynesia: Australs, Marquesas and Society (image adapted from [81]). The four islands from the Society archipelago used in this study are indicated by means of a blue arrow: Raiatea, Huahine, Moorea and Tahiti.

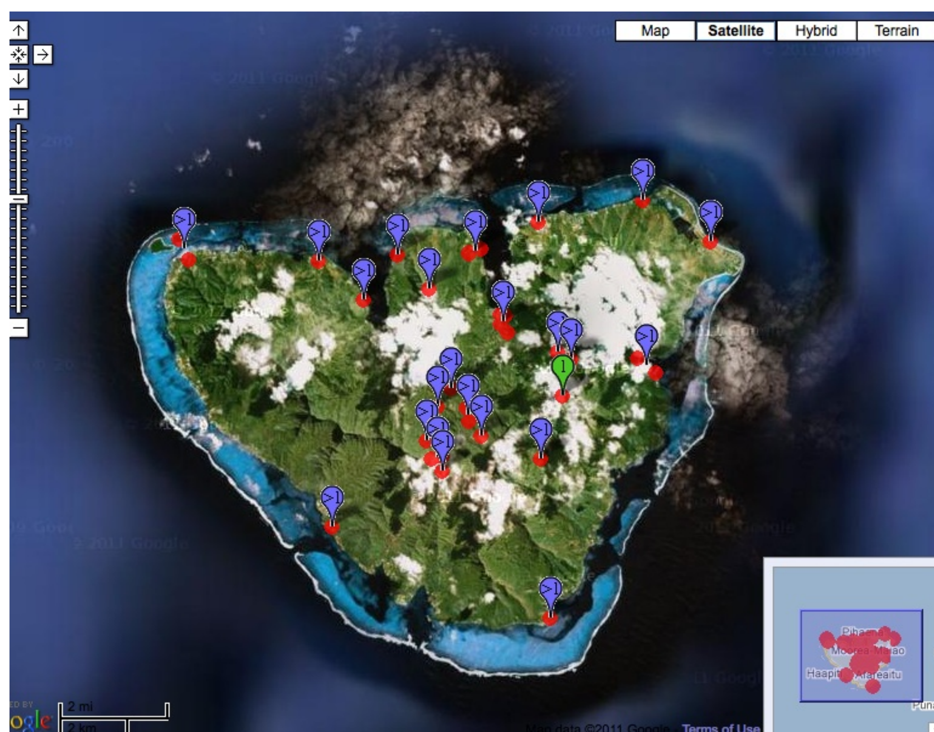




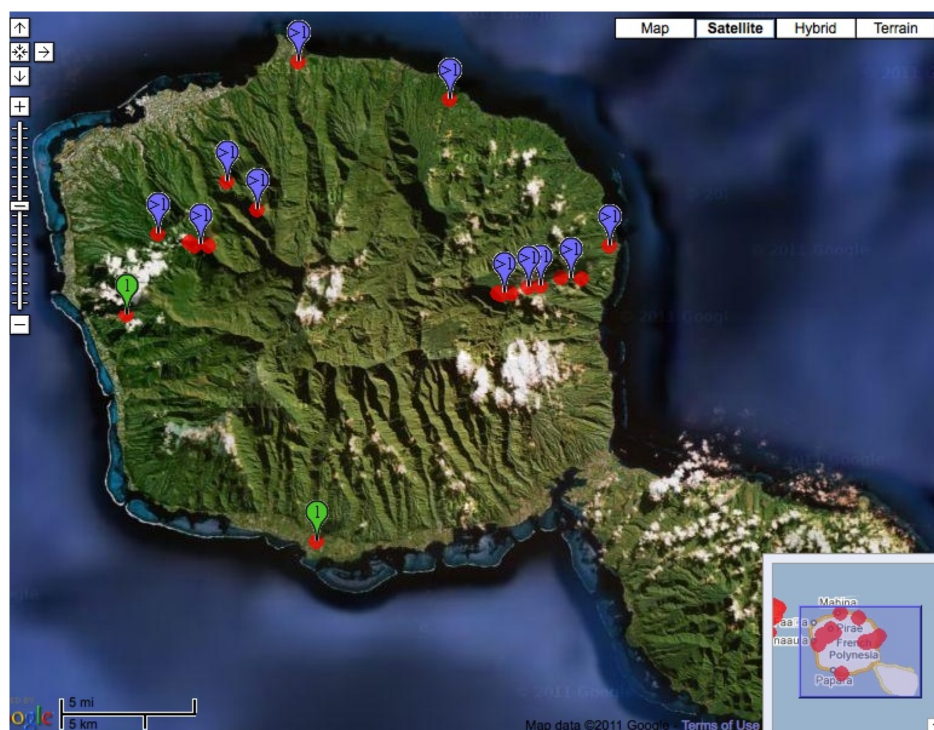
(a) Map of the 17 collecting sites in Raiatea covered in a total of 7 days.



(b) Map of the 12 collecting sites in Huahine covered in a total of 6 days.



(c) Map of the 37 collecting sites in Moorea covered in a total of 87 days



(d) Map of the 20 collecting sites in Tahiti covered in a total of 21 days

Figure 2.2: Collecting sites on the four Society Archipelago islands used in our study: Raiatea, Huahine, Moorea and Tahiti. Legend: blue balloons - more than 1 specimen was collected in the site, green balloons - only 1 specimen collected in the site. Images were extracted from the Moorea Biocode Project database where information regarding all collected specimens in this project has been deposited (http://biocode.berkeley.edu/).

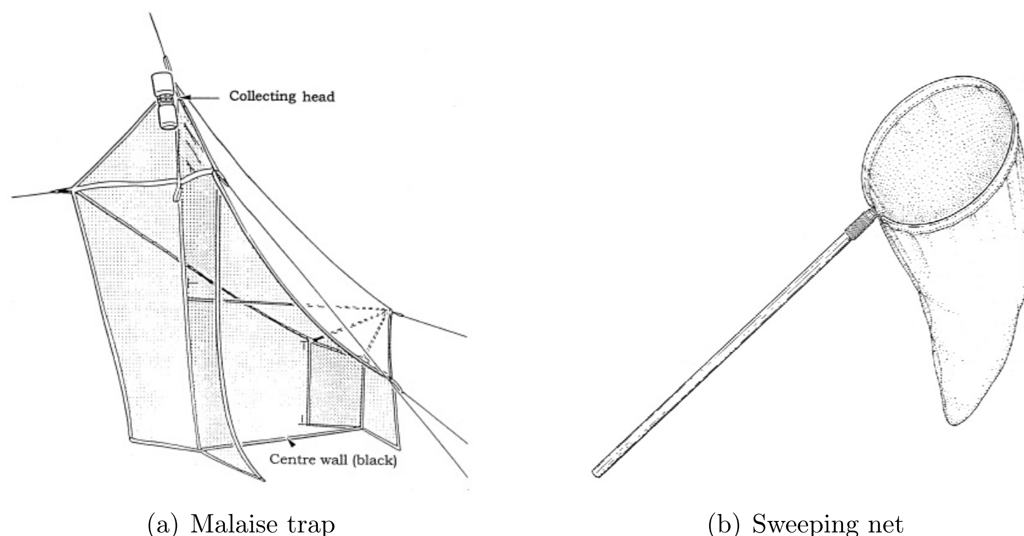
2. Malaise traps, which work on the principle that insects once inside the trap will fly to the highest and brightest point where they fall on a collecting vessel with alcohol. These traps are better suited for small flying insects and are most efficient to collect some Dipterans and Hymenopterans. Other arthropods such as spiders and mites can also be collected using these traps [82], see Figure 2.3.

3. Light traps, including UV light, which are best suited for night-flying insects, particularly moths (Lepidoptera) [82].

4. Pan traps, bright coloured collecting vessels filled with liquid (soap) that are placed in the ground. They mainly serve to capture ground-dwelling and low flying arthropods, including Coleoptera, Hymenoptera, Diptera, Apterytotes and Orthoptera [82].

5. Nets, which are used for insects that spend most of their time flying or resting for short time periods on vegetation. The main groups successfully captured by this method include dragonflies (Odonata), butterflies and moths (Lepidoptera), flies (Diptera) and Hymenopterans [82].

A total of fourteen Arthropod orders were sampled in the four islands with the various methods (see Table 2.1).



(a) Malaise trap (b) Sweeping net

Figure 2.3: Two of the collecting methods generally used for collecting Arthropods.

Once individuals were collected, they were preserved in 96% ethanol without refrigeration. Possible cross-specimens contamination with *Wolbachia* DNA between individuals stored in ethanol was tested and ruled out [83]. As such, mass collections

of specimens were preserved in 96% ethanol, at room temperature and protected from light, for any length of time.

An excel worksheet with all information concerning the collection events and conforming to the requirements of the Moorea Biocode Project FIMS database (http://mooreabiocode.org) was produced on site. This data include identification of collecting event, collector, date of collecting event, GPS coordinate of collecting site, altitude, specimen unique identification number, available taxonomic information, sex, life stage, etc. Due to the number of collectors (including pupils from the Paopao's College entomological group), information was sometimes missing (see missing data in Table 2.1), particularly regarding collecting methods (2049 entries with unregistered collecting method in 11022 specimens sorted, i.e., corresponding to 18% of the specimens).

Table 2.1: Arthropod orders present in all four islands: Huahine, Moorea, Raiatea and Tahiti, and the collecting methods used in each island per Order. Legend: n=net, l=light trap, m=malaise trap, md=missing data, s=sweeping net, uv=UV trap and p=pan trap (soap); in green: Orders more represented in our dataset; in parenthesis: total number of specimens collected.

	Island			
Order	Moorea	Tahiti	Huahine	Raiatea
ARANEAE	md,m,s (315)	md,s,m (147)	uv,s (283)	s (240)
BLATTARIA	md,m,n (74)	md,s,m (18)	s (1)	s (6)
COLEOPTERA	md,l,m,s,n (502)	md,s,l,m (108)	s,uv (71)	s (56)
COLLEMBOLA	md (24)	s,m (94)	s (59)	s (5)
DIPTERA	md,p,m,s,l (874)	md,s,l,m (1045)	s,uv (707)	s (271)
HEMIPTERA	md,p,s,m,l,n (683)	s,md,l,m (334)	s,uv (381)	s (123)
HYMENOPTERA	md,s,p,m,n,l (702)	s,md,l,m (289)	s,uv (394)	s (274)
ISOPODA	md,n (38)	md (4)	n (6)	n (7)
ISOPTERA	l,md,m (34)	m (2)	n (3)	n (1)
JULIDA	md,m (23)	m (3)	md (9)	md (7)
LEPIDOPTERA	md,l,m,s,n (842)	md,s,l,m,n (570)	s,uv (230)	s (102)
NEUROPTERA	md,m (15)	s,m (15)	s (8)	s (7)
ODONATA	md,m,s (42)	md,s (14)	s (7)	s (5)
ORIBATIDA	s (7)	md (3)	s (6)	s (6)
ORTHOPTERA	md,s,m (136)	md,s,m (144)	s (119)	s (54)
PSOCOPTERA	md,m,s (62)	md,s,m,l (70)	s (42)	s (20)
SPIROBOLIDA	md,m(18)	md (1)	md (10)	md (5)
THYSANOPTERA	md,m,s (11)	md (3)	s (20)	s (4)

2.1.2 Classification and sorting

After September 2007, all specimens had been shipped and received in Lyon. Initially, storage was made at room temperature and protected from sunlight. Sorting of the

specimens was performed during this period using the following protocol:

1. The content of mass storage tubes (50ml falcon tubes) was cleaned (removal of vegetation) and sorted to order-level groups, occasionally classified to the family or subfamily level.

2. Each specimen was photographed (Nikon D200, lens Tamron 90mm macro) to keep a record of the morphological characteristics of the specimen. A single photo was taken from a lateral view, and when needed additional photographs were taken from dorsal and/or ventral views or zooming in on details (e.g., of wing patterns, mouth appendages, dorsal hairs). All photos were taken using as background millimetric paper to allow for the dimensions of the specimens to be registered. Most specimens were photographed in an ethanol solution, as this provided the most natural and expanded view.

3. With the aid of a magnifying glass, each specimen was carefully observed and attributed a clustering number, i.e., a morpho-cluster. This number was meant to identify a group of individuals grouped together based on shared morphological characteristics that led us to consider them as belonging to a same "species". Our aim was to produce a classification where a maximum of the observed phenotypic diversity would be preserved. For each morpho-cluster, a type specimen was chosen and used for comparison when classifying new specimens.

4. Each specimen was, afterwards, individually stored at room temperature and, at a later time, at -20°C. Before transferring specimens from room temperature to -20°C, storage ethanol was changed once.

5. When more than 10 specimens were available in one morpho-cluster, these were not stored in individual tubes but in mass, at -20°C, for future work.

While in this phase, we collaborated with experts on the Coleoptera (Roland Allemand), Araneae (Michael Dierkens) and Odonata (Daniel Grand) Orders producing a more accurate morphological classification of these taxonomic groups. We were able to obtain accurate morpho-species for the spiders order and to access how much of the true diversity we were catching in our sampling. A total of 70 species had already been described in this archipelago, and in our 68 Araneae morpho-clusters, we identified 50 corresponding to previously described spider species from the Society Archipelago and 18 corresponding to first reports in this area [84]. For this Order, our morpho-clusters corresponded to true-taxonomically identified species. As such, we estimate that 71%

(50 in 70) of the known species diversity for this group has been covered in our sampling. Although this cannot be taken as a measure of the overall diversity represented in our sampling, the result with the Araneae order is encouraging as it refers to a group which had not been extensively sampled before.

A total of 1337 morpho-clusters were formed from the 11022 specimens sorted during this process.

2.1.3 DNA extractions

A selection of specimens was made, previous to extractions, to minimise the number of individuals that we would need to sequence without compromising the genetic diversity of our data. This selection was applied to each morpho-cluster using the following criteria (see also Figure 2.4):

1. If the the sex of specimens was known, then two females and two males were selected per island.

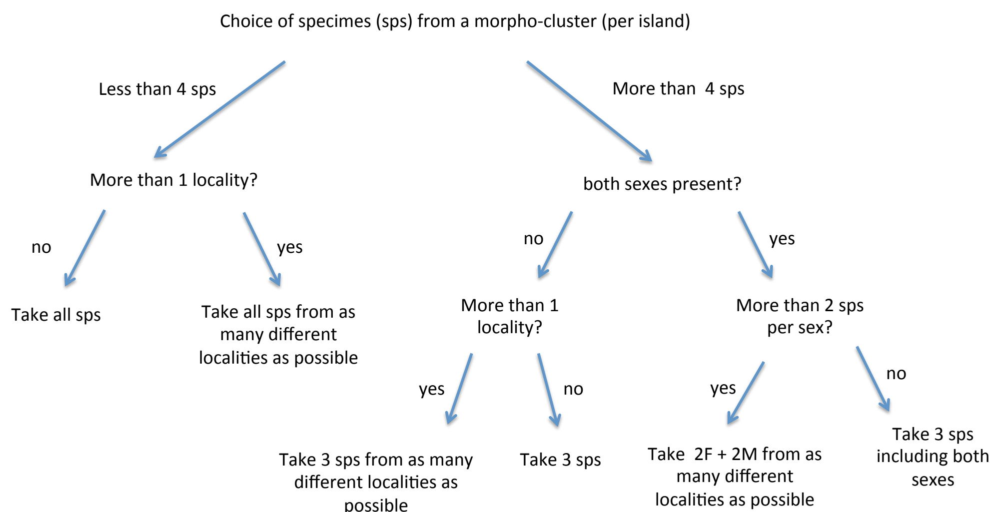
2. When the sex was undetermined, we selected only three specimens per island and according to their location. If all individuals had been sampled on the same location, we would collect three of these individuals for extractions. If individuals had been collected from two different locations within one island, then one individual would be selected from each location and the third one would be randomly picked from one of the locations. If specimens had been collected in three distinct intra-island locations, then one individual would be selected from each possible location. If more than $m > 3$ locations had been sampled within one island, then three distinct locations would be randomly picked from the m and one individual would be sampled from each one of these three locations.

3. If less than three (or four) specimens were available per island, all extant individuals were selected.

DNA extractions were thus performed on a total of 4671 specimens (representing 43% of the 11.022 specimens collected and sorted) according to the following protocol:

1. *Tissue preparation*: specimens were removed from -20°C and left to dry on tissue, at room temperature, until all ethanol had evaporated. If specimens were less than 5mm long, the whole body was used for DNA extraction, otherwise only abdominal tissue was used for the extraction. In the latter case, tissue was cut from the mid-lower abdomen including the gonads but leaving the distal part of

Figure 2.4: Criteria used to select the specimens to be sequenced. Legend: F= females; M=males.



the abdomen (carrying the genitalia) intact. To optimise the efficiency of extraction, all tissues were quickly frozen (15min at -80°C) before tissue disruption and homogenisation (20sec at 20Hz) using a TissueLyser (QIAGEN). Ninety six well plates were prepared with tissue for extraction taking into account:

(a) Individuals from the same morpho-cluster were distributed into different plates;

(b) Individuals from morpho-clusters encompassing various arthropod orders were distributed in the same plate;

(c) All islands should be represented in each plate, whenever possible. Where 1-individual morpho-clusters were concerned, these clusters were distributed in various plates.

With this plan for plate distribution, we expected to minimise the impact of experimental problems (PCR reactions, sequencing, etc) that might occur and could have a confounding effect in the results of both host and parasite markers amplification.

2. *DNA extraction*: all extractions were performed with Nucleospin® 96 Tissue kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany), according to the manufacturer's instructions except on: (a) the pre-lysis step where incubation

with proteinase K was made overnight and (b) the elution step which was made in 2 sub-steps (50 +50 μl of water).

After elution, 96 well plates with 20μl of extracted DNA were stored at -20°C and the remaining 80 μl were stored in plates at -80°C (long term DNA stock).

2.2 Sequencing of a molecular marker for the hosts: Cytochrome Oxydase I (COI)

DNA barcoding is generally viewed as a large-scale effort to produce an efficient, standardised and "universal" protocol to perform molecular typing of any organism in order to assign their position in an extant taxonomic classification [73, 85].

Protein coding mitochondrial genes have, to date, been the favourite choice for barcoding molecular markers. This choice is mainly motivated by some characteristics of the mitochondrial (*mt*) genome such as: hypothesis that the *mt* genome is under neutral selection, haploidy, protein coding genes evolve faster than nuclear ones and rarely have indels (as most lead to shifts in the reading-frame [73]). In animals, the most used gene has been *cox*1 coding for the *mt* cytochrome oxydase subunit I (COI), the major subunit of the transmembrane protein cytochrome C. Due to the richness of information already available for this gene and the lack of a better alternative (see the review by Pons and Vogler [86]), we chose this gene as the molecular marker for the hosts. Not only does it allow us to profit from the existing databases to perform a more accurate identification of specimens but, most importantly, its maternal inheritance is of major importance for us as we are interested in a fast evolving marker with the same mode of inheritance as *Wolbachia*. This allows us to follow the host's maternal lineages with a high resolution at recent divergence events.

PCR reactions The mitochondrial gene Cytochrome Oxidase I (COI) was amplified with the primers LCO1490 5'-GGTCAACAAATCATAAAGATATTGG-3' and HCO2198 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' [87], with expected amplicons of size $\sim$ 658 bp. PCRs were performed in a total volume of 30μl with 1.5 mM of $MgCl2$, 2 mM of all four dNTPS, 0.2μM of each primer, 0.2 Units/μl L EuroTaq® DNA polymerase (EUROBIO, Les Ulis, France). The temperature profile was as follows: initial denaturation at 95°C for 120 seconds (sec); 35 cycles of 94°C for 30 sec, 47°C for 30 sec and 72°C for 90 sec; and a final extension at 72°C for 600 sec.

All reactions took place in a Tetrad®Thermocycler (Bio-Rad, Hercules, CA, USA)

with 2μl of DNA template, except in the negative PCR control where 2μl of water was added instead of DNA. Qualitative analysis of PCR success was performed by gel electrophoresis. A total of 8μl (5μl of PCR product and 3μl of Bromophenol Blue) were separated in a 1% agarose gel with a run of 25 min, at 100V. The gels were stained with Ethidium Bromide, and visualised under UV light. A quantitative criterion based on band intensity was used for the selection of positive COI amplicons to be sent for sequencing.

2.3 Screening of *Wolbachia* infections

Wolbachia is a monophyletic genus including deeply diverging lineages. Its detection is traditionally performed using PCR. *Wolbachia*'s large impact and diversity, and the diversity of our Arthropod (host) dataset made it urgent to assess the efficiency of existing screening protocols and, if possible, to improve the detection of infection by *Wolbachia*. The sensitivity and range of commonly used PCR primers and of a new set of 16S primers were then evaluated with multiple strains encompassing the *Wolbachia* genus diversity and, concomitantly, the host's diversity. We showed that some primer sets are significantly more efficient than others, but that no single protocol can ensure the specific detection of all known *Wolbachia* infections. Based on our analysis, we used a new pair of 16S primers to perform the detection of *Wolbachia* infections. The results of these tests were summarised in the following paper, published in *Molecular Ecology Resources* in 2011.

PCR reactions The 16S rRNA gene was amplified with primers 553F_W (5'-ATACGGAGAGGGCTAGCGTTA-3') and 1334R_W (5'-CTTCATRYACTCGAGTT-GCWGAGT-3'). PCRs were performed in a total volume of 15 μl with 1.5 mM of MgC2, 2 mM of all four dNTPs, 0.2 μM of the forward primer (553F_ W) and 2 μM of the degenerate reverse primer (1334R_ W), 0.4 Units/μL EuroTaq® DNA polymerase (EUROBIO, Les Ulis, France). The temperature profile was as follows: initial denaturation at 94°C for 2 minutes (min); 35 cycles of 94°C for 1 min, 62°C for 1 min and 72°C for 1 min; and a final extension at 72°C for 10 min. All reactions took place in a Tetrad®Thermocycler (Bio-Rad, Hercules, CA, USA) with 2μl of DNA template, except in the negative PCR control where 2μl of water were added instead of DNA.

Qualitative analysis of PCR success was performed by gel electrophoresis. A total of 8μl (5μl of PCR product and 3μl of Bromophenol Blue) were separated in a 1% agarose gel with a run of 25 min, at 100V. The gels were stained with Ethidium Bromide, and

visualised under UV light. DNA from all specimens positive for *Wolbachia* infection was pooled in new 96 well plates to further amplify and sequence a *Wolbachia* protein-coding gene.

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 567–572 doi: 10.1111/j.1755-0998.2010.02955.x

MOLECULAR DIAGNOSTICS AND DNA TAXONOMY

Wolbachia detection: an assessment of standard PCR Protocols

P. M. SIMÕES,*† G. MIALDEA,† D. REISS,† M.-F. SAGOT* and S. CHARLAT†
**Équipe BAMBOO (INRIA) and †Equipe Génétique et Evolution des interactions Hôtes-Parasites, Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université Claude Bernard Lyon 1, 43 bd du 11 Novembre 1918, 69100 Villeurbanne, France*

Abstract

***Wolbachia* is a large monophyletic genus of intracellular bacteria, traditionally detected using PCR assays. Its considerable phylogenetic diversity and impact on arthropods and nematodes make it urgent to assess the efficiency of these screening protocols. The sensitivity and range of commonly used PCR primers and of a new set of 16S primers were evaluated on a wide range of hosts and *Wolbachia* strains. We show that certain primer sets are significantly more efficient than others but that no single protocol can ensure the specific detection of all known *Wolbachia* infections.**

Keywords: 16S rRNA gene, MLST, PCR methods, specificity, *Wolbachia* screening

Received 9 August 2010; revision received 12 November 2010; accepted 18 November 2010

The intracellular bacterium *Wolbachia* is currently considered the most abundant endosymbiont in arthropods and shows a global geographic distribution (Werren & Windsor 2000; Werren *et al.* 2008); in insects, it is estimated that over 65% of the species carry members of the *Wolbachia* clade (Hilgenboecker *et al.* 2008). *Wolbachia* is also present in parasitic filarial nematodes that are causative agents of river blindness and elephantiasis (Taylor & Hoerauf 1999; Fenn & Blaxter 2004).

In addition to its abundant presence in arthropods and nematodes, the increasing interest in *Wolbachia* is motivated by the diversity of its phenotypic effects. In arthropods, it can manipulate the host reproduction through male-killing (Jiggins *et al.* 2001), cytoplasmic incompatibility (Poinsot *et al.* 2003), parthenogenesis induction (Stouthamer *et al.* 1999) and feminization of genetic males (Hiroki *et al.* 2002; Vandekerckhove *et al.* 2003) with large impact on host ecology and evolution (Jiggins *et al.* 2000; Bordenstein *et al.* 2001; Charlat *et al.* 2003, 2007; Kremer *et al.* 2009). Additionally, *Wolbachia* is known both to play a protective role against some RNA viral infections in *Drosophila* (Browlie and Johnson, 2009) and to establish obligate associations, as seen in all infected nematodes (Fenn & Blaxter 2004; Fenn *et al.* 2006), and more rarely in arthropods (Dedeine *et al.* 2005; Hosokawa *et al.* 2010). This diversity of effects offers promising applications to the fields of pest and disease vector control (Kambris *et al.* 2009; Moreira *et al.* 2009).

Correspondence: Patrícia M. Simões, Fax: (+33) (0) 4 72 43 13 88; E-mail: simoes.m.pat@gmail.com

The *Wolbachia* genus encompasses a large phylogenetic diversity, with deeply diverging 'supergroups' and a root situated approximately 100 Ma (Bandi *et al.* 1998). These supergroups show an asymmetric distribution within the host landscape: supergroups A and B are commonly found in arthropods; supergroups C and D are restricted to filarial nematodes; supergroup E is exclusive to springtails; supergroup F is present both in arthropods and in nematodes; supergroup H is found in a single genus of termites; and supergroup K in one spider mite species. Other *Wolbachia* strains have been detected, and their clustering into supergroups is either controversial (supergroup G) or as yet unclear (Lo & Evans 2007; Ros *et al.* 2009).

Detection of *Wolbachia* has been traditionally performed using PCR assays, targeting the 16S rRNA gene or protein-coding genes such as *wsp* and *ftsZ*. The wide diversity and large impact of *Wolbachia* make it urgent to assess the efficiency of these detection protocols. Here, we use an extensive collection of insects of both known and unknown *Wolbachia* infection status to compare the sensitivity and range of several PCR primers: those commonly used as screening PCR primers, a new set of 16S primers and the primers integrated in the *Wolbachia* MLST (Multi Locus Sequencing Typing) system (Baldo *et al.* 2006).

The samples used include representatives from most of the known *Wolbachia* supergroups (A, B, C, D, F, H and K) as well as two new groups: M (Jack Werren, personal communication) and I (Haegeman *et al.* 2009). Additionally, we included arthropod specimens of unknown infection status, collected in the French Polynesia

© 2011 Blackwell Publishing Ltd

between March 2005 and July 2007 (see Table S1 for further details on the samples used).

All tissue samples were conserved in 96% ethanol at −20 °C. DNA extraction was performed with the Nucleospin® Tissue kit (MACHEREY-NAGEL GmbH & Co. KG, Düren, Germany), according to the manufacturer's instructions, with elution in 150 μL of water. For small specimens (<5 mm long), extraction was performed on the entire body while abdominal tissue was used for larger specimens. Samples received from colleagues as dehydrated DNA were re-suspended at room temperature for 2–3 h in 90 μL of water. All samples were stored at −20 °C. The quality of the samples that failed to amplify *Wolbachia* DNA was assessed using the metazoan CO1 primers LCO-HCO (Folmer *et al*. 1994); samples that failed to amplify the CO1 locus were further tested using eukaryotic 18S and arthropod 18S primers (Medlin *et al*. 1988; Halanych *et al*. 2001; Duron *et al*. 2008).

Thirteen different *Wolbachia* PCR primers, targeting six loci, were evaluated in this study, hereafter denominated using the following short names: 16S-1 to 16S-6 (O'Neill *et al*. 1992; Werren & Windsor 2000; Casiraghi *et al*. 2001; Sakamoto *et al*. 2006; this study); wsp (Zhou *et al*. 1998), GatB, CoxA, HcpA, Fbp, FtsZ-1 (Baldo *et al*. 2006) and FtsZ-2 (Casiraghi *et al*. 2001). Primer sequences and expected amplicon size are provided in Table S2. PCR protocols (Table S3 and S4) followed recommendations from the original publications, except for the 16S-6 pair, designed in this study (see Data S1 for details on primer design). All reactions used EuroTaq® DNA polymerase (EUROBIO, Les Ulis, France) and took place in a Tetrad®Thermocycler (Bio-Rad, Hercules, CA, USA) with 1 μL of template. Fragments were separated by 1% agarose gel electrophoresis, stained with Ethidium Bromide and visualized under UV light.

We first performed a preliminary assessment of 12 existing pairs of primers using a set of 12 DNA extracts, including arthropods and nematodes and at least one representative of the majority of the known supergroups: A (2 samples), B (2 samples), C (1 sample), D (1 sample), F (3 samples), I (1 sample), H (1 sample) and M (1 sample) (Table S1). Our results show that the different primers can produce very distinct patterns (Table 1 and Fig. S1). The primers 16S-2, 16S-5, wsp, FtsZ-2 and the 5 MLST primers detected 10 or more infections of 12 in total. Discrimination among these pairs is mainly based on two samples, which harbour infections from supergroups F (*Mansonella pertans*) and I (*Rhadopholus similes*). Overall, only two pairs of primers (16S-2 and CoxA) produced identical results and only the FbpA primers detected the 12 infections, although one sample (F infection from *M. pertans*) produced a very faint band.

Next, we assessed the specificity (detection of *Wolbachia* infections only) of all the primer pairs that detected at least 10 of the 12 *Wolbachia* infections (Table 1 and Fig. S2). To this purpose, we used DNA templates from six close relatives of *Wolbachia*: *Rickettsia conovii*, *Rickettsia*

Table 1 Preliminary assessment of PCR protocols: primers are denominated using short names (see text)

	Host	Unknown arthropod						*Bm*	*Ov*	*Cp.*	*Mp.*	*Rm*	*Zsp*	Number						
Primers	Supergroup	A	F	A	B	B	M	D	C	F	F	I	H	of positives	*Rc*	*Ap*	*Eca*	*Ech*	*Rt*	*Rb*
16S-1		2	0	0	3*	2	0	0	0	0	*	0	0	3	na	na	na	na	na	na
16S-2		3	3	3	3	3	3	3	1	3	0	0	3	9 + (1)	0	*	0	0	0	0
16S-3		0	*	0	1	3*	0	1	0	*	0	0	*	1 + (2)	na	na	na	na	na	na
16S-4		3*	2	3	3	2	3	3	*	2*	*	0	3*	9	na	na	na	na	na	na
16S-5		3	3	3	3	3	3	3	2	3*	*	2	3	11	*	2*	2*	*	*	*
16S-6		3	3	3	3	3	3	3	3	3	0	2*	3	11	0	0	0	0	0	0
FbpA		3	3	3	3	3	3	3	3	3	1*	3*	3	11 + (1)	1*	*	*	*	1*	1*
FtsZ-1		3	3	3	3*	3	3	3	3	3*	0	*	2*	10	1*	*	*	*	*	0
CoxA		3	3	3	3	3	3	2	3	3	0	0	3	10	*	*	*	0	1*	0
GatB		3	3	2*	3	3*	2	2	3	1*	*	2*	2	10 + (1)	3*	3*	3*	3*	3*	1*
HcpA		3	3	3	3	3	3	3*	3*	2*	1*	*	3	10 + (1)	*	*	2*	*	2*	*
wsp		3*	3*	3	3*	3*	2	2	3	1*	*	*	3*	9 + (1)	2*	1*	*	2*	*	0
FtsZ-2		3	3*	3	3	3*	3*	3*	3	3*	1*	*	3*	10 + (1)	1*	*	*	3*	*	*

Results are coded as follows: '0' = no amplification; '1' = very weak band of correct size; '2' = faint band of correct size; '3' = strong band of correct size.

*Amplification(s) of incorrect size (false bands and 'na' = not applicable. Abbreviations: *Bm*, *Brugia malayi; Ov*, *Onchocercus volvulus*; *Cp*, *Chorthippus paralelus*; *Mp*, *Mansonella pertans*; *Rs*, *Rhadopholus similes*; *Zsp*, *Zootermopsis* spp; Rc, *Rickettsia conovii*; *Ap*, *Anaplasma phagocytophylum*; Eca, *Ehrlichia cannis*; Ech, *Ehrlichia chaffensis*; Rt, R*ickettsia typhi* and Rb, *Rickettsia bellii*.

© 2011 Blackwell Publishing Ltd

Table 2 Assessment of detection protocols on a large sample of known *Wolbachia* results is coded as in Table 1. All samples are from unknown arthropods unless otherwise stated. Abbreviations are as in Table 1 except Pch (ld): *Pityogenes chalcographus*

	Supergroup							
Protocol	A	B	B	F	F	A	A	A**
16S-2	3	3	3	3	0	3	3	3
16S-6	3	3	3	3	0	3	3	3
FbpA	3	3	3	3	1	3	3	3
	A	F**	A	B	F	B	A	B
16S-2	3	3	3	3	3	3	2	3
16S-6	3	3	3	3	3	3	3	3
FbpA	3	3	3	3	3	3	3	3
	B	A	A**	B**	A	F	A	B
16S-2	3	3	3	3	3	3	3	3
16S-6	3	3	3	3	3	3	3	3
FbpA	3	3	3	3	3	3	3	3
	B	B	B	F	B	B**	B	B
16S-2	3	3	3	3	3	3	3	3
16S-6	3	3	3	3	3	3	3	2
FbpA	3	3	3	2	3	3	3	3
	A	A	B	F	A	M**	B	F
16S-2	3	3	3	3	3	3	3	3
16S-6	3	3	3	3	3	3	3	3
FbpA	3	3	3	3	3	3	3	3
	D** (Bm)	D (Bm)	D (Bm)	D (Bm)	C** (Oc)	C (Oc)	C (Oc)	C (Oc)
16S-2	2	3	1	na	3	3	3	3
16S-6	3	3	3	na	3	0	0	0
FbpA	3	3	2	na	3	3	3	3
	F** (Cp)	F (Cp)	F (Cp)	F** (Mp)	A (Pch (ld))	A (Pch (ld))	A (Pch (ld))	A (Pch (ld))
16S-2	3	3	3	0	0	0	0	0
16S-6	3	3	*	0	0	0	0	0
FbpA	3	3	2	1*	*	0	2	0
	I** (Rs)	H** (Zsp)	H (Zsp)	K (Bsp)				
16S-2	0	3	3	3				
16S-6	1	3	2	3				
FbpA	3*	3	3	3				

typhi, *Anaplasma phagocytophilum*, *Ehrlichia canis*, *Ehrlichia chaffensis* and *Rickettsia bellii*. Only the 16S-2 pair exhibited complete specificity to *Wolbachia* infections.

Following the sensitivity and specificity tests, we decided to analyse more exhaustively the most satisfactory sets of primers only. Our selection was based on the

© 2011 Blackwell Publishing Ltd

following criteria: (i) sensitivity (detection of the maximum diversity of *Wolbachia* strains) and (ii) specificity and, preferably, absence of false (wrong size) amplifications. Only the 16S-2 primers thus qualified for a more extensive test. However, this pair of primers does not detect all *Wolbachia* infections. We therefore designed new 16S primers (16S-6) in an attempt to improve the detection of *Wolbachia* without compromising specificity and tested them on the above-mentioned 12 *Wolbachia* and 6 non-*Wolbachia* templates (see Data S1 for details, Fig. S3, Table S5). This resulted in the amplification of 11 *Wolbachia* infections of 12 and no amplification of closely related groups (Table 1 and Fig. S4). Therefore, we tested the most promising pairs of primers, namely 16S-2 and 16S-6, on a more extensive set of samples (59 DNA extracts) of known infection status including a sample from supergroup K, more samples from the A, B, C, D, F and I supergroups (including isolates from different hosts) and four low-density infections (A supergroup). Notably, the large set of A and B infections include most of the known diversity within these groups based on a MLST characterization (Jack Werren, personal communication). The 16S-2 primers produced 15% (7 of 59) false negatives while the 16S-6 pair produced 17% (10 of 59) false negatives (contingency test, $P = 0.6$) (Tables 2 and S6). Both pairs seem equally sensitive to most supergroups, in particular the widespread A and B infections. The 16S-2 pair seems more efficient for the C group while only the 16S-6 pair amplifies the I group (albeit very weakly). The two pairs detected the K and H infections, but the 16S-2 primers produced stronger signals for the H group (Fig. S5).

This extensive test also included the FbpA primers, despite their lack of specificity, since they were the only ones detecting the 12 *Wolbachia* infections in the preliminary assessment (although one signal was very weak). They were the only pair to detect both an F infection (from an unknown Arthropod host) and a low-density A infection, producing 9% of false negatives. Overall, the false negative rate did not vary among the three pairs of primers tested (contingency test, $P = 0.12$) and none of the three pairs (16S-2, 16S-6 or FbpA) was able to detect all the infections in this extended test. As one cannot rule out the detection of non-*Wolbachia* strains using the FbpA primers, in particular the detection of *R. bellii*, the assessment of this primer pair was ended at this stage.

A final comparison between the 16S-2 and 16S-6 pairs was conducted with a set of 90 arthropods of unknown infection status. As expected from the previous tests, the results differed slightly among primer pairs (Table S7 and Fig. S6). Quantitatively, the 16S-6 pair detected 26 infections, three of which were exclusive to these primers; the 16S-2 pair detected 24 infections, of which a single amplification was exclusive to this pair. Overall, the two pairs of primers agreed in 23 of 27 (85%) positive samples. The 16S-2 pair produced, in general, stronger signals, and the detection rate did not differ significantly between the two pairs (contingency test, $P = 0.87$).

Standard PCR can be implemented on large samples and remains at present the most effective method to screen *Wolbachia* infections. Our results highlight one constraint associated with these methods, that is, the trade-off between sensitivity and specificity. We observed that two sets of 16S primers (including the ones designed in this study) amplify only *Wolbachia* infections. However, these showed some level of variation in terms of sensitivity. The FbpA primers appeared slightly more sensitive but lacked specificity. In an attempt to satisfy the sensitivity and specificity criteria in a single reaction, we evaluated the feasibility of a multiplex PCR (see Data S1 for details). Annealing temperature constraints suggested that the 16S-6 and FbpA pairs could be combined. However, we observed a reduced efficiency of each pair when used in combination.

Projects involving PCR followed by sequencing may give priority to sensitivity since non-*Wolbachia* DNAs will be identified at the sequencing stage. In addition, a protein-coding locus would be more informative than an rRNA locus if the sequences are to be used for typing or phylogenetic inference. The FbpA primers would then represent the best candidates. The additional cost incurred will depend on the actual incidence of close relatives to *Wolbachia*, in particular *R. bellii*. In contrast, PCR-based projects aimed at filtering *Wolbachia* infections only should give priority to specificity. The 16S-2 and 16-6 primers would then be most appropriate. Independent reactions involving these different primer pairs would ensure the lowest false positive and false negative rates.

Acknowledgements

This project was supported by Fundação para a Ciência e Tecnologia (FCT) with a grant to P. Simões (SFRH/BD/28235/2006). We are grateful to the CNRS (grant ATIP), to an ANR Blanc MIRI (project BLAN08-1_335497) and the ERC grant AdG Sisyphe (M.-F. Sagot) for provision of the funds.

We thank the following colleagues for kindly providing samples: J. Werren (University of Rochester), S. Bordenstein (Vanderbilt University); P. Martinez (University of Madrid); A. Haegeman (University of Gent); T. Nutman (National Institute of Health); V. Ros and H. Breeuwer (University of Pennsylvania); L. Ford (Liverpool School of Tropical Medicine); C. Stauffer (Institute of Forest Entomology) and D. Raoult (Faculty of Medicine of Marseille). We thank the Geneious developer R. Moir, M. Gouy for adjustments in SeaView and R. Allemand for help with the arthropod. We would also like to thank the three anonymous reviewers for their critical assessment of our work.

© 2011 Blackwell Publishing Ltd

References

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

Bandi C, Anderson JC, Genchi C, Blazter ML (1998) Phylogeny of Wolbachia in filarial nematodes. *Proceedings of the Royal Society of London Series B*, **265**, 2401–2413.

Bordenstein SR, O'Hara FP, Werren JH (2001) *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in Nasonia. *Nature*, **409**, 707–710.

Brownlie J, Johnson K (2009) Symbiont-mediated protection in insect hosts. *Trends in Microbiology*, **17**, 348–354.

Casiraghi M, Anderson TJ, Bandi C, Bazzocchi C, Genchi C (2001) A phylogenetic analysis of filarial nematodes: comparison with the phylogeny of Wolbachia endosymbionts. *Parasitology*, **122**, 93–103.

Charlat S, Hurst GDD, Mercot H (2003) Evolutionary consequences of *Wolbachia* infections. *Trends in Genetics*, **19**, 217–223.

Charlat S, Reuter M, Dyson EA *et al.* (2007) Male-killing bacteria trigger a cycle of increasing male fatigue and female promiscuity. *Current Biology*, **17**, 273–277.

Dedeine F, Bouletreau M, Vavre F (2005) *Wolbachia* requirement for oogenesis: occurrence within the genus Asobara (Hymenoptera, Braconidae) and evidence for intraspecific variation in A. tabida. *Heredity*, **95**, 394–400.

Duron O, Bouchon D, Boutin S *et al.* (2008) The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biology*, **6**, 27.

Fenn K, Blaxter M (2004) Are filarial nematode *Wolbachia* obligate mutualist symbionts? *Trends in Ecology & Evolution*, **19**, 163–166.

Fenn K, Conlon C, Jones M *et al.* (2006) Phylogenetic relationships of the *Wolbachia* of nematodes and arthropods. *PLoS Pathogens*, **2**, e94.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Haegeman A, Vanholme B, Jacob J *et al.* (2009) An endosymbiotic bacterium in a plant-parasitic nematode: member of a new *Wolbachia* supergroup. *International Journal for Parasitology*, **39**, 1045–1054.

Halanych K, Feldman R, Vrijenhoek R (2001) Molecular evidence that *Sclerolinum brattstromi* is closely related to Vestimentiferans, not to Frenulate Pogonophorans (Siboglinidae, Annelida). *The Biological Bulletin*, **201**, 65–75.

Hilgenboecker K, Hammerstein P, Telschow A, Werren JH (2008) How many species are infected with *Wolbachia*? A statistical analysis of current data. *FEMS Microbiology Letters*, **281**, 215–220.

Hiroki M, Kato Y, Kamito T, Miura K (2002) Feminization of genetic males by a symbiotic bacterium in a butterfly, Eurema hecabe (Lepidoptera: Pieridae). *Naturwissenschaften*, **89**, 167–170.

Hosokawa T, Koga R, Kikuchi Y, Meng X-Y, Fukatsu T (2010) *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 769–774.

Jiggins FM, Hurst GDD, Majerus ME (2000) Sex-ratio-distorting *Wolbachia* causes sex-role reversal in its butterfly host'. *Proceedings of the Royal Society B Biological Sciences*, **267**, 69–73.

Jiggins FM, von Der Schulenburg JH, Hurst GDD, Majerus ME (2001) Recombination confounds interpretations of *Wolbachia* evolution. *Proceedings of Biological Sciences*, **268**, 1423–1427.

Kambris Z, Cook PE, Phuc HK, Sinkins SP (2009) Immune activation by life-shortening *Wolbachia* and reduced filarial competence in mosquitoes. *Science*, **326**, 134–136.

Kremer N, Charif D, Henri H *et al.* (2009) A new case of *Wolbachia* dependence in the genus Asobara: evidence for parthenogenesis induction in Asobara japonica. *Heredity*, **103**, 248–256.

Lo N, Evans TA (2007) Phylogenetic diversity of the intracellular symbiont *Wolbachia* in termites. *Molecular Phylogenetics and Evolution*, **44**, 461–466.

Medlin L, Elwood H, Stickel S, Sogin M (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene*, **71**, 491–499.

Moreira LA, Iturbe-Ormaetxe I, Jeffery JA *et al.* (2009) A *Wolbachia* symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium. *Cell*, **139**, 1268–1278.

O'Neill SL, Giordano R, Colbert AM, Karr TL, Robertson HM (1992) 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proceedings of the National Academy of Sciences*, **89**, 2699–2702.

Poinsot D, Charlat S, Merçot H (2003) On the mechanism of *Wolbachia*-induced cytoplasmic incompatibility: confronting the models with the facts. *Bioessays*, **25**, 259–265.

Ros VID, Fleming VM, Feil EJ, Breeuwer JAJ (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**, 1036–1043.

Sakamoto JM, Feinstein J, Rasgon JL (2006) Wolbachia infections in the Cimicidae: museum specimens as an untapped resource for endosymbiont surveys. *Applied and Environmental Microbiology*, **72**, 3161–3167.

Stouthamer R, Breeuwer JA, Hurst GDD (1999) *Wolbachia* pipientis: microbial manipulator of arthropod reproduction. *Annual Review of Microbiology*, **53**, 71–102.

Taylor MJ, Hoerauf A (1999) *Wolbachia* bacteria of filarial nematodes. *Parasitology Today*, **15**, 437–442.

Vandekerckhove T, Watteyne S, Bonne W (2003) Evolutionary trends in feminization and intersexuality in woodlice (Crustacea, Isopoda) infected with Wolbachia pipientis (alpha-Proteobacteria). *Belgian Journal of Zoology*, **133**, 61–69.

Werren JH, Windsor DM (2000) *Wolbachia* infection frequencies in insects: evidence of a global equilibrium?. *Proceedings of Biological Sciences*, **267**, 1277–1285.

Werren JH, Baldo L, Clark ME (2008) *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, **6**, 741–751.

Zhou W, Rousset F, O'Neil S (1998) Phylogeny and PCR-based classification of Wolbachia strains using wsp gene sequences. *Proceedings Biological Sciences*, **265**, 509–515.

Supporting Information

Additional supporting information may be found in the online version of this article.

Data S1 Design of new primers.

Fig. S1 Preliminary assessment of existing *Wolbachia* detection protocols.

Fig. S2 Assessment of the specificity of existing protocols.

Fig. S3 The figure shows the 16S primers aligned with the following sequences.

Fig. S4 Preliminary assessment of a new *Wolbachia* detection protocol (16S-6 primers).

Fig. S5 Assessment of detection protocols on a large sample of *Wolbachia*-infected hosts.

Fig. S6 Blind test on a random sample of arthropods.

Table S1 Tissue and DNA material used.

Table S2 Primers assessed in the present study.

Table S3 PCR mix concentrations.

© 2011 Blackwell Publishing Ltd

Table S4 PCR programs.

Table S5 The 11 *Wolbachia* sequences chosen as representatives of the known supergroups.

Table S6 Comparison among 16S-6, 16S-2 and FbpA.

Table S7 DNA material used in the blind test on a random sample of arthropods.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2011 Blackwell Publishing Ltd

2.4 Selection of a marker for *Wolbachia*: Fructose Bisphosphate Aldolase (FbpA

We were interested in a fast evolving gene, conserved throughout the *Wolbachia* genus (both in nematode and arthropod infections).

Five housekeeping genes (*coxA, fbpA, ftsZ, gatB* and *hcpA*) have been identified that are present in a single copy, scattered throughout the *Drosophila melanogaster*'s *Wolbachia* strain (*w*Mel) genome and under strong stabilising selection within the *Wolbachia* genus [44]. Since 2006, they have been used together in the characterisation of *Wolbachia* strains via what is currently known as the Multi Locus Sequencing Typing (MLST) of *Wolbachia*.

The *fbpA* gene was shown to have the highest nucleotide diversity (per site) of all 5 MLST genes, even though it exhibited intragenic recombination both within and between supergroups A and B [44].

Of the most commonly used genes to study *Wolbachia*'s diversity, the gene coding for a surface protein (*wsp*) seems to be the most rapidly evolving one but it also evidences extensive recombination [88] and strong diversifying selection [40]. Only the *ftsZ* gene ("Filamenting temperature-sensitive mutant Z"; also part of the MLST genes) has not shown evidence of recombination. This gene, coding for a protein involved in cell division, is also the slowest evolving of the 5 MLST genes.

As we were attempting to evaluate the turnover of infections at a micro-evolutionary time scale, we needed a more fast evolving gene. We opted then for FbpA as the molecular marker of *Wolbachia*'s diversity.

PCR reactions The *Wolbachia* fructose-bisphosphate aldolase gene was amplified with the primers FbpA-R1 5'-GCTGCTCCRCTTGGYWTGAT-3' and FbpA-R1 5'-CCRCCAGARAAAAYYACTATTC-3' [44], with amplicons of expected size varying between 429-423 bp. PCRs were performed in a total volume of 30μl with 1.5 mM of MgCl2, 0.2 mM of all four dNTPS, 1.0μl M of each primer, 0.5 Units/μl EuroTaq® DNA polymerase (EUROBIO, Les Ulis, France). The temperature profile was as follows: initial denaturation at 94ºC for 120 seconds (sec); 36 cycles of 94ºC for 30 sec, 56ºC for 45 sec and 72ºC for 90 sec; and a final extension at 72ºC for 600 sec. All reactions took place in a Tetrad®Thermocycler (Bio-Rad, Hercules, CA, USA) with 2μl of DNA template, except in the negative PCR control where 2μl of water. Qualitative analysis of PCR success was performed by gel electrophoresis. A total of 8μl (5μl of PCR product and 3μl of Bromophenol Blue) were separated in a 1% agarose gel with a

run of 22 min, at 100V. The gels were stained with Ethidium Bromide, and visualised under UV light.

2.5 Sequencing and sequence editing

2.5.1 Sequencing

A total of 20-25μl of positive PCR product was used for sequencing of both strands of the COI and FbpA amplicons. Sequencing was made at the French Centre National de Séquençage (Genoscope) as part of the ASpeedID project, a large collective barcoding project headed by Jean-Yves Rasplus and of which S. Charlat is part.

2.5.2 Sequence, assembly and editing

Following the sequencing process, a trace chromatogram was obtained for the positive and the negative strands of each PCR product. Chromatograms were then cleaned, edited, and a consensus sequence from complementary strands was obtained as specified below. All steps were performed using GENEIOUS v5.4.0 (Biomatters) [89].

1. Trace chromatograms were paired-grouped according to specimen identity. For each pair, we started by removing low quality end sequences of both the positive and negative strands. As the 5' end of a sequence is more prone to sequencing errors (end of the amplification), we opted to always remove at least the first 45 bp. Additionally, we chose to remove low-quality regions from the ends until the remaining sequence had a probability lower than 0.1% of having an error in base. This trimming procedure was automatically performed before the alignment step. Cleaner sequences for each strand were then used to produce the consensus.

2. After trimming, both strands were aligned using the assembly alignment algorithm of GENEIOUS with the "Medium Sensitivity/Fast" option. A consensus sequence was then produced with quality scores. In case of disagreement/conflict between the 2 strands, we used the "Highest Quality" option to automatically choose the base with highest quality. In this case, the final quality score of the base call was calculated as the difference between the 2 conflicting base call quality scores. If the final base quality had a score smaller than 20 (probability of incorrect base call superior to 1%), then an ambiguity (N) was attributed to that base call in the consensus sequence. Although this correction to base calling produced slightly more ambiguities in the final consensus sequence, overall these

sequences were cleaner in the sense they had very good quality where a base was called.

3. Sequences that were not assembled into a consensus (unique reads) were re-evaluated with a new base call quality procedure. This new base call quality score was performed using the "find heterozygotes" option, to detect double peaks, imposing the condition that the minimum confidence score to change a base call to an ambiguity (N) should be higher than 25. This procedure allowed us to recover a sequence with good quality from a single read, as sometimes one of the reads was bad but the other had a good chromatogram. If both reads were recovered (see below the minimum length requirement for sequences), then we opted to keep the 5' end (the LCO1490 primer for the *cox* gene and the FbpA_F primer for the *fbpA* gene) read.

4. Both consensus and single-read sequences were then selected based on their length and number of ambiguities. For COI, the dataset is composed of sequences with a minimum of 450 bp (68% of the expected total size) and a maximum of 1% ambiguities. For FbpA, the dataset consists of sequences with a minimum of 350 bp and a maximum of 1% ambiguities.

5. For each gene, a reference sequence was produced with the complete length of the amplicon without the flanking primers. All sequences within each dataset were then aligned with the respective reference so as to remove the primer regions from our sequences. For COI, we used the trimmed sequence of the *Hypolimna bolina* cytochrome oxidase subunit I (GenBank: EF683668.1) and for FbpA the trimmed sequence of the *Drosophila melanogaster* putative fructose-bisphosphate aldolase (accession number AE017196.1) to remove the sequence ends outside the reference sequence. Again, only sequences longer than 450 bp and 350 bp were kept in the CO1 and FbpA sequence datasets, respectively.

6. A last quality check was performed using GetOrf (EMBOSS 6.3.1) to find open reading frames (ORF) in nucleotide sequences. As the region(s) amplified both on COI and FbpA correspond to a coding region, we searched for the longest nucleotide region between stop codons, restricting the maximum length difference between the nucleotide sequence and the ORF to 4 nucleotides. All 6 reading frames were searched for, and only sequences with no stop codon(s) were kept (with the sequences for shorter ORFs of both CO1 and FbpA being removed from the initial sequence datasets). All ORFs were put into frame by adding 1 or 2

N's (to sequences starting in the third and second codon position, respectively) at the beginning of the sequences.

2.5.3 Submission to the Moorea Biocode Project

Information related to specimens (collecting information, taxonomy, photos) as well as CO1 sequences were deposited in the Moorea Biocode database, accessible through the web at the URL http://mooreabiocode.org/. This database includes all specimens collected in the general frame of the Moorea Biocode project. Specimens used in the present project can be specifically accessed by choosing "Laboratoire Biométrie & Biologie Evolutive" as holding institution. The database also allows localisation of the specimens on a google earth interface. Sequence data has been deposited through the biocode plugin of the software geneious, providing access to the original trace files as well as to the final sequences in fasta format. At the moment, only members of the biocode project can access sequence data, which will be automatically deposited on public databases (GenBank and BOLD) upon publication.

2.6 Alignments and phylogenies

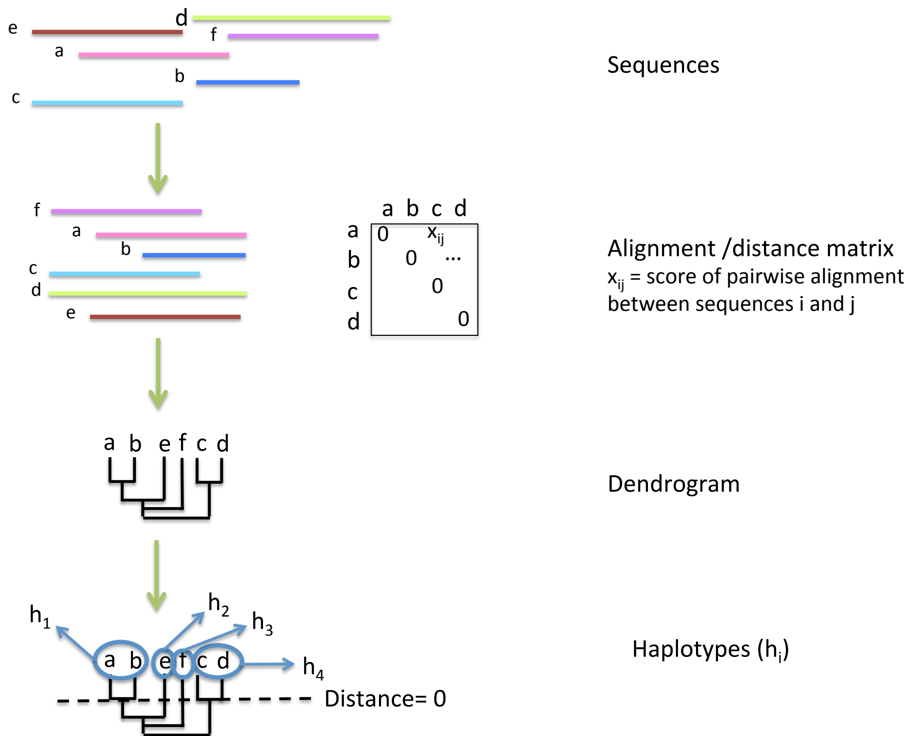
2.6.1 Alignments

Both the COI and the FbpA sequences were aligned using the MUSCLE algorithm, as implemented in SEAVIEW [90], based on the nucleotide translated sequences. For the COI sequences, the invertebrates *mt* genetic code was used, and for FbpA, the universal one for protein translation.

2.6.2 Haplotype groups definition

As expected from our dataset, sequences obtained for a given marker showed various degrees of variation due to polymorphism. Unique sequences were used to define different haplotypes (see also Figure 2.5). Whenever several identical sequences were identified, these were grouped into a single haplotype. The grouping of sequences into haplotypes was performed by calculating the uncorrected pairwise distance matrix (taking into account gaps) to produce a dendrogram. We then hierarchically grouped sequences setting the maximum diameter of a cluster to zero and using as distance measure the distance between 2 elements, i.e., sequences. All operations were performed using R and the package *ape*. A total of 1548 COI and 247 FbpA haplotypes were identified.

Figure 2.5: Grouping of sequences into haplotypes. A distance matrix was calculated using uncorrected pairwise distance and used to produce the dendrogram. The maximum distance between elements/sequences within the same cluster was then set to zero and an hierarchical clustering of sequences was performed using as distance measure the maximum distance between 2 elements.



2.6.3 Tree reconstruction method(s)

To reconstruct the evolutionary history of our sequences we constructed maximum likelihood ML phylogenies for each marker.

Maximum likelihood tree building methods are statistical methods that search for the tree with the topology and branch lengths that have the higher probability of producing the observed dataset. This probability is conditional to the observed sequences and the model of sequence evolution used [91, 92]. Due to the extensive and heavy calculations required to calculate the probabilities and likelihoods, ML phylogenies are constructed using heuristics.

We used the PhyML v3.0 algorithm, as implemented in SEAVIEW, to produce phylogenies for both the COI and FbpA markers. As both the COI and FbpA markers code for functional proteins, a general time reversible model (GTR) with a proportion of invariant sites (I) and across site rate variation (Γ) was used for sequence evolution

(GTR+Γ+I model). Both the proportion of invariant sites and the across site rate variation were estimated from the observed sequence dataset. In the latter case, we used 4 categories of sites rate variation to estimate the gamma shape parameter.

The BioNJ tree was chosen as the starting tree and the heuristic search in the tree-space was performed using the operations of NNI (Nearest Neighbour Interchange) and SPR (Subtree Prunning and Regrafting) to search for the best tree topology. The NNI search corresponds to a swap between two adjacent branches in a tree. The gain of likelihood is evaluated for each move (neighbouring tree) and the first improvement found is accepted. The SPR search consists in removing a branch (internal or leading to a leaf) with the attached subtree and in reinserting it in the remaining tree in all possible places.

Branch support was calculated with the approximate likelihood ratio test (aLRT) measure (Shimodaira-Hasegawa (SH)-like procedure), which uses the test statistic 2(L1-L2) where L1 is the log-likelihood of the current tree and L2 is the log-likelihood of the second best NNI configuration around the branch of interest. It measures how much a given branch improves the likelihood of the tree by comparison with the likelihood of the tree obtained when collapsing that branch solely and maintaining the remaining of the tree topology. In general, a good concordance between bootstrap proportions and aLRT values is obtained.

2.6.4 Cluster definition

A monophyletic group is a group of organisms including all descendants from the most recent common ancestor of that group. Organisms in this monophyletic group, i.e. clade, are in general characterised by shared derived characteristic, also called synapomorphies.

In our study, monophyletic clusters of closely related sequences were defined for both markers COI and FbpA. These clusters were calculated using the distance matrix obtained by adding the branch lengths in the ML phylogenies (patristic distances).

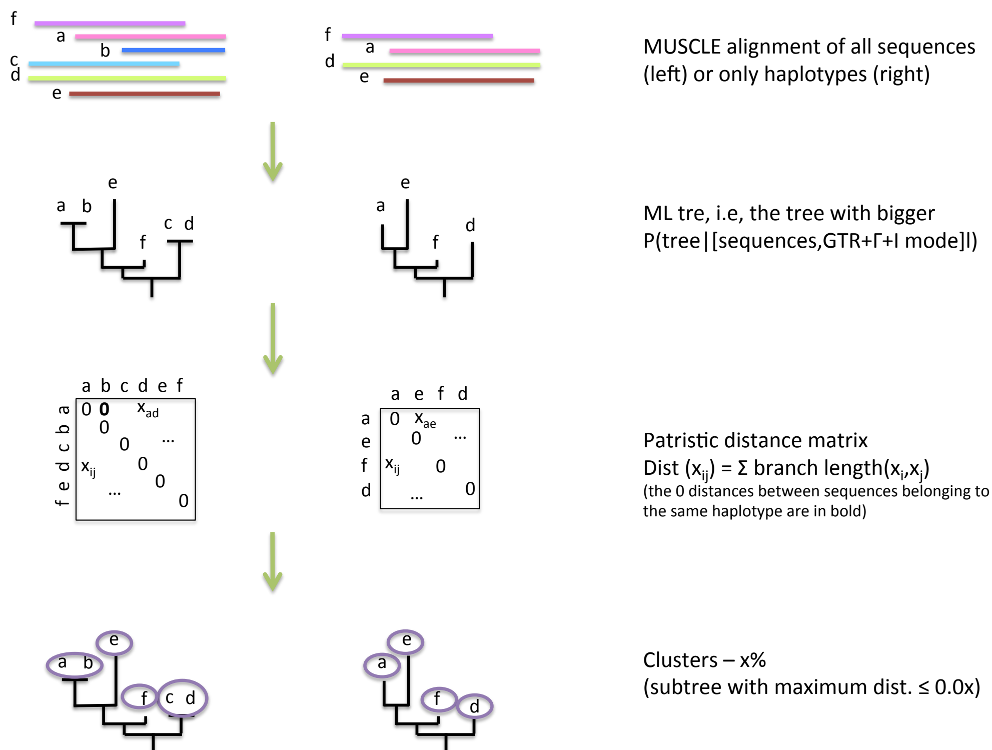
A maximum distance between leaves, i.e., a threshold was defined. Next, to compute the clusters, we used a recursive algorithm that starts at the root u of the tree and calculates the maximum distance between any pair of leaves in the subtree rooted at v. If this distance is bigger than the threshold, then the algorithm proceeds to the nodes that are children of the root and so on recursively until a node v' is found for which the maximum distance between any two leaves in the subtree rooted at v' is lower or equal to the threshold. Once such a node has been found, all its descendent leaves form a cluster, and the algorithm then proceeds to the next not yet considered sibling of v', if

any exists. The recursion ends when each leaf has thus been grouped into a cluster.

This method allowed us to obtain monophyletic clusters for which the maximum distance between leaves is no more than the threshold (see Figure 2.6). In our study, monophyletic clusters were defined using either all sequences (see the section on Biogeography) or only representatives from each haplotype group (see the Section on host and parasite evolution).

Various thresholds were used throughout our study. For COI, we used a 3% and 11% patristic distance, while for FbpA we used a 2% and 5% distance.

Figure 2.6: Defining clusters-x%. Either all sequences or representatives from each haplotype were aligned using the MUSCLE algorithm. A maximum likelihood tree was obtained using PhyML and the patristic distance matrix was calculated and used to search the tree to obtain all subtrees with a maximum distance between 2 leaves $\leq 0.0x$. The leaves of one such subtree formed a cluster-x%.



Chapter 3

Results and Discussion

Contents

3.1 Diversity and biogeography of the Host mitochondria . . 38

3.1.1 Mitochondrial diversity in the Society Archipelago 39

3.1.2 Biogeography of Mitochondrial in the Society Archipelago . . 48

3.1.3 Discussion 60

3.2 Incidence and diversity of *Wolbachia* infections 65

3.2.1 Incidence 66

3.2.2 Diversity of *Wolbachia* infections as inferred using the FbpA marker 71

3.2.3 Discussion 81

3.3 Horizontal transfers of *Wolbachia* infections 85

3.3.1 Defining groups of closely related *Wolbachia* strains 85

3.3.2 Proportion of *Wolbachia* clades being horizontally transferred between different arthropod orders 87

3.3.3 Host specificity of horizontally transferred *Wolbachia* clades . 88

3.3.4 Investigating preferential transfers between Orders 89

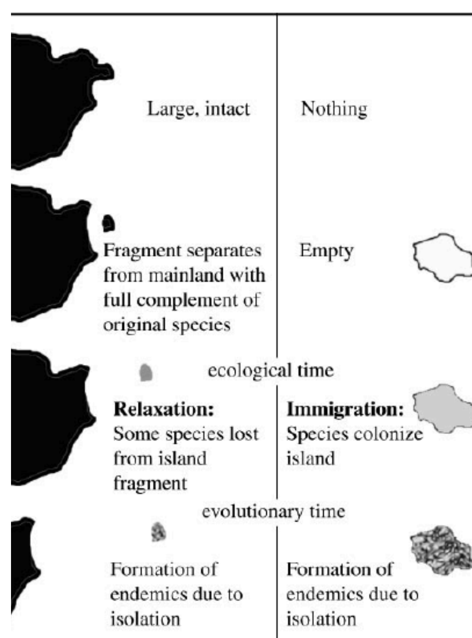
3.3.5 Exploring the ecological routes of Wolbachia horizontal transfer 94

3.3.6 Discussion 98

3.1 Diversity and biogeography of the Host mitochondria

Islands are isolated systems, i.e., they are biological systems where contact with other communities is either inexistent or very limited for an extended period of time. Such systems can be actual islands, small areas of land completely surrounded by water, or any ecosystem or favourable habitat surrounded by different ecosystems or unsuitable habitats [77]. We can consider two main categories of islands based on their history [93]: (1) darwinian or *de novo*-formed islands, which have never been in contact with any source of colonialism and are thus full of empty or available ecological niches; and (2) fragment islands, which are patches of favourable habitats that were previously part of a larger ecosystem and have since become surrounded by unsuitable ones, meaning their ecological niches are already largely filled (Figure 3.1).

Figure 3.1: Historical formation and distinction between fragment and darwinian islands, adapted from [93].



This broad division reflects the very distinct patterns of biodiversity, speciation and even species conservation problems that can emerge in islands (for a detailed discussion, see [93]). In fragment islands, the ecological space is already filled from the start. We thus expect a decrease in the species richness (relaxation or loss of species from the island) due to the smaller area as well as, along evolutionary time, the appearance of paleo-endemics by relictualism. The latter is due to divergent evolution between the

mainland population and the island populations of a previously widespread species. In *de novo*-formed islands instead, the species richness is due to the immigration of colonisers and as such tends to increase in the beginning. Endemics can arise after a radiation event followed by adaptation to new ecological niches (called neo-endemics). When looking at neo-endemics, we can thus consider evolutionary changes occurring at the time scale of the ages of the islands.

Remote island systems that are both oceanic and volcanic, such as the Society Archipelago used in our study, are an example of *de novo* islands with a high degree of isolation. Various evolutionary studies in such systems have suggested that the observed biological diversity is exclusively due to colonisation events by a few colonisers followed by within and between islands diversification (high endemism). This leads to a suggested model in which immigration rate is very slow and it is speciation, via local adaptations, that is mainly responsible for the observed species diversity [77, 93].

Nevertheless, in the last decade, a number of studies have shown that even in these remote systems, various factors can influence within-island diversity: (1) adaptive radiation; (2) multiple successive colonisations by either migrants from neighbouring islands and/or distant continents (long-distance dispersal); (3) vicariant speciation; and (4) successive bottlenecks and founder events.

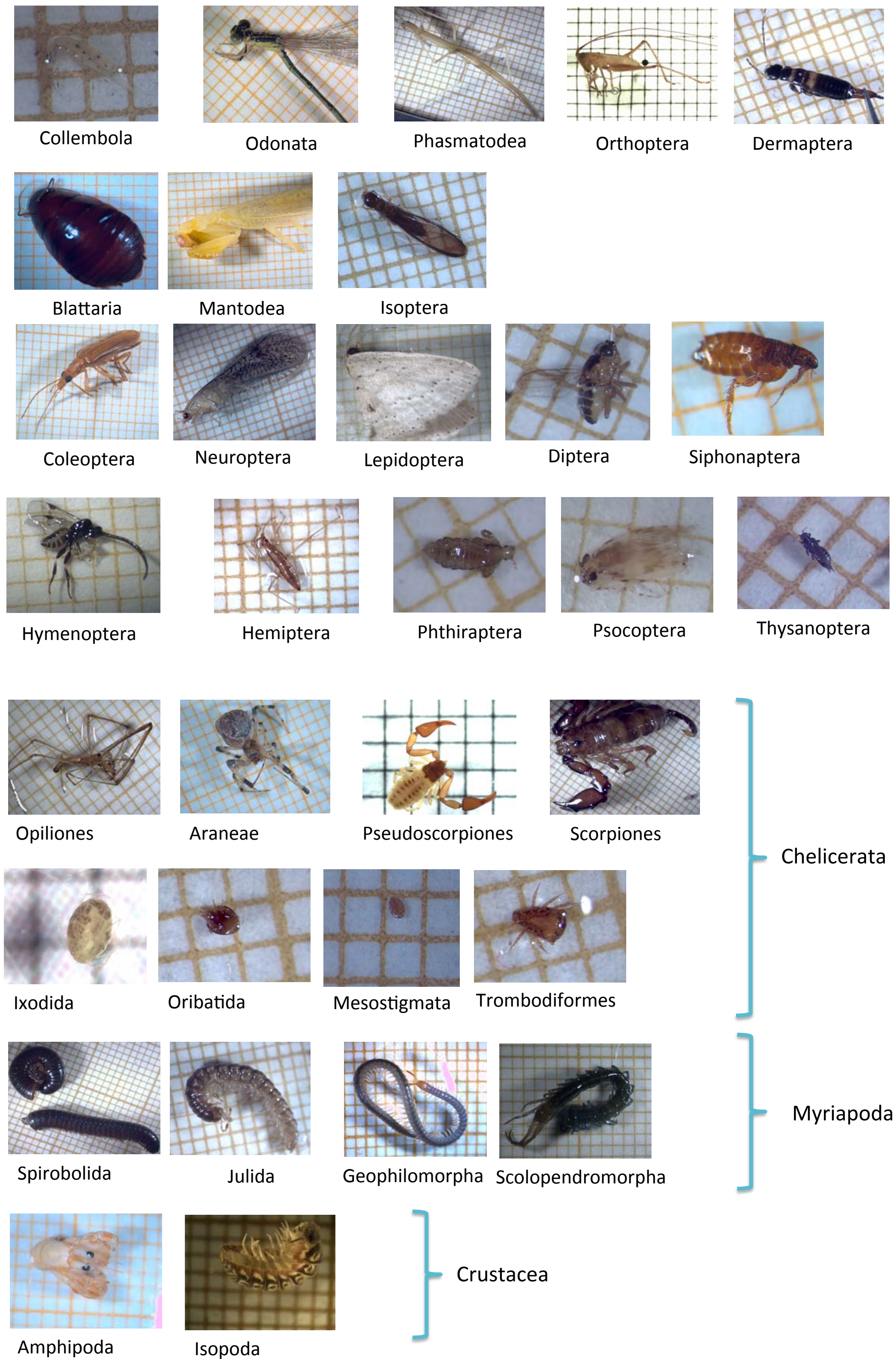
Previous studies focusing on specific groups - such as the blackfly genus *Simulum* [94], the spider genus *Tetragnatha* [95], the broad-nosed weevil genus *Rhyncogonus* [81] - have shown a high degree of genetic diversity in these genera with new endemic species being found, particularly, in the youngest island Tahiti. In general, more diversity is found in younger islands than in older ones. It is thought that this results from erosion processes and, consequently, from a loss of habitat diversity (including running water habitats) on the older islands [81].

In this section, we shall be interested in describing the diversity of hosts collected in our dataset and in studying the presence of a geographical structure in the aforementioned dataset.

3.1.1 Mitochondrial diversity in the Society Archipelago

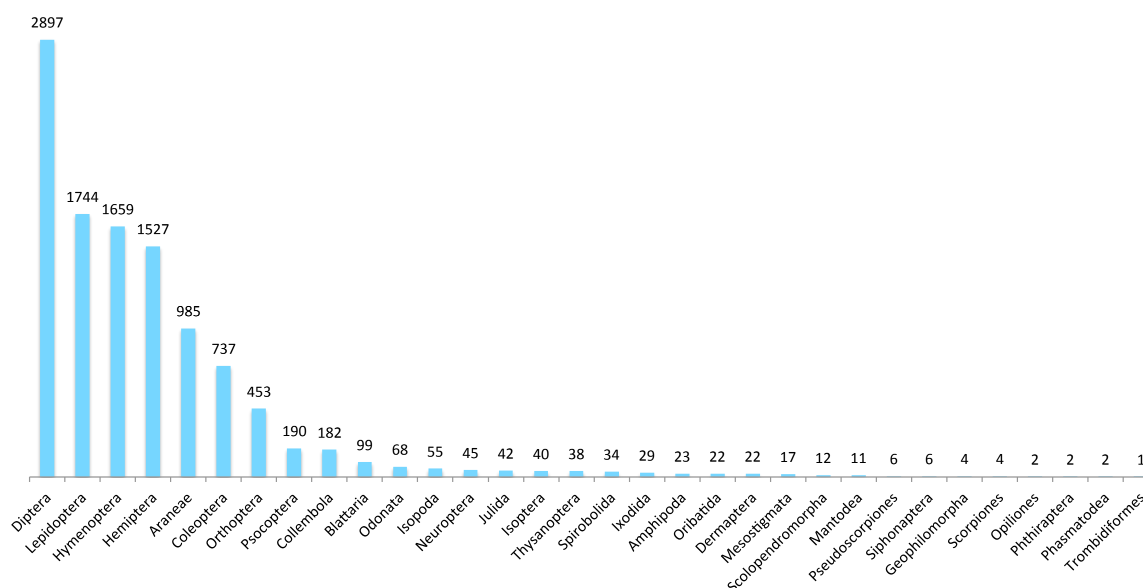
As mentioned previously (Chapter 2), 11022 individuals collected in the 4 islands – Raiatea, Huahine, Moorea and Tahiti – were examined and taxonomically characterised (at least, to the Order level), photographed and stored. These specimens included a total of 32 Orders, as seen in Figure 3.2, covering both terrestrial and aquatic arthropods.

Figure 3.2: All 32 Orders sampled in the ensemble of the 4 Society islands used in our study.



As can be seen in Figure 3.3, some Orders were more successfully sampled than others. This is known to be influenced by various factors such as: the species richness of each Order, the collecting methods, the ecological niches/habitats sampled (strongly correlated with the distribution of species in the extant habitats and their abundance). A variety of collecting methods was used to maximise the number and diversity of species collected, thereby minimising the impact of this variable (for further details see Table 2.1 in Chapter 2). An effort was also made to sample as many distinct habitats as possible although this was conditioned by access to the collecting sites. A total of 18 of the Orders, i.e. 50%, were sampled in all 4 islands (full description in Table 2.1 in Chapter 2).

Figure 3.3: Distribution of the 11022 individuals sorted according to the Order. Indicated on top of each Order bar is the total number of specimens sorted for that Order.

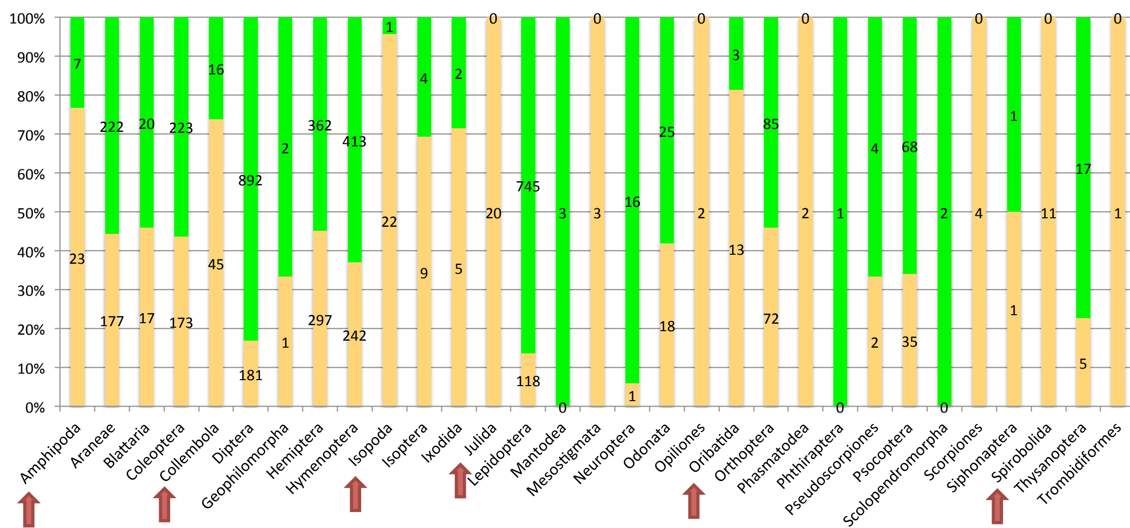


In order to better characterise the diversity of hosts collected and to infer the recent evolutionary history of cytoplasmic lineages, we chose to sequence the host COI marker with the commonly used LCO-HCO primers [87]. As mentioned previously, the choice of the *cox locus* was based on the maternal inheritance of this marker, the richness of molecular data already available for this locus and, consequently, the contribution our data will bring to the inventory of non-microbial life in an insular tropical ecosystem (via the submission of our COI sequences to the Moorea Biocode Project, http://mooreabiocode.org/). As described in Chapter 2, the number of host specimens to be

sequenced was reduced to 4671 without compromising the genetic diversity collected in the field. This reduction was based on a morphological classification of the individuals into morpho-clusters, i.e., into groups of morphologically very similar individuals.

We were able to obtain a successful sequencing reaction for $\sim 67\%$ of the specimens, i.e. for 3135 individuals. Sequences were obtained from the consensus of both DNA strands, with less than 1% ambiguities (for further details on sequence assembly see Chapter 2). As can be seen in Figure 3.4, the amplification success varied depending on the Order. For some Orders, we were not able to obtain any COI sequence. Such was

Figure 3.4: Proportion of individuals successfully characterised by the COI marker according to their Order. The green bars represent the proportion of individuals for which a COI sequence was obtained while the orange bars correspond to the proportion of individuals for which no COI molecular marker was obtained. The number of individuals with and without marker are indicated, respectively, in the center of the green and orange bars. Orders for which a low ratio of amplification success was obtained ($< 30\%$) and for which we had more than 10 individuals tested are indicated with a red arrow.



the case of the mite Orders Thrombidiformes and Mesostigmata, and of the Scorpiones and Opiliones, all belonging to the Class of Arachnida. Could this be due to some shared mutations, in the *cox locus* primers target region , in this Class? This is possible although we did not test enough representatives of each of these Orders to be able to distinguish between this scenario and experimental random errors. Interestingly, we had no similar problems with another arachnid Order - Order Araneae (spiders) - for which the success ratio of the COI amplification was $\sim 56\%$ and 396 specimens were tested. An additional problem, in the case of the mite Orders, was the very small size of the specimens. The negative amplifications, and subsequent sequencing, may be due

to the small quantities of these mites' template genomic DNA. This same size problem may also explain the lack of COI characterisation of another mite Order – Oribatida – for which however more specimens were available.

Within the subphylum Hexapoda, we had problems obtaining COI sequences for two Orders only: Collembola (springtails) and Phasmatodea (stick insects). In the latter case, again only 2 individuals were available, so that random failure cannot be excluded. As for the Order Collembola, it is more likely that the lack of success is due to divergence in the primers target region. Another explanation might be the presence of PCR inhibitors specific to this group. We checked the plates in which the samples were distributed and there was no apparent problem with both the plates used for the PCR amplification nor with the plates sent for sequencing. As mentioned previously, specimens were randomly distributed in different plates to eliminate a confounding effect.

Collembola and Oribitida that we saw above are two among the six Orders for which a low ratio of sequencing success was obtained (< 30%) and for which we had more than 10 individuals tested (indicated with a red arrow in Figure 3.4). The others included two Orders from the Class Malacostraca: Amphipoda and Isopoda, plus two Orders (both millipedes) from the Class Diplopoda: Julida and Spirobolida. Due to the number of individuals tested and, again, based on the rate of sequencing success in the plates where these specimens were distributed, we think that the most likely scenario is the inadequacy of the LCO-HCO primers to amplify the *cox locus* in these groups of arthropods.

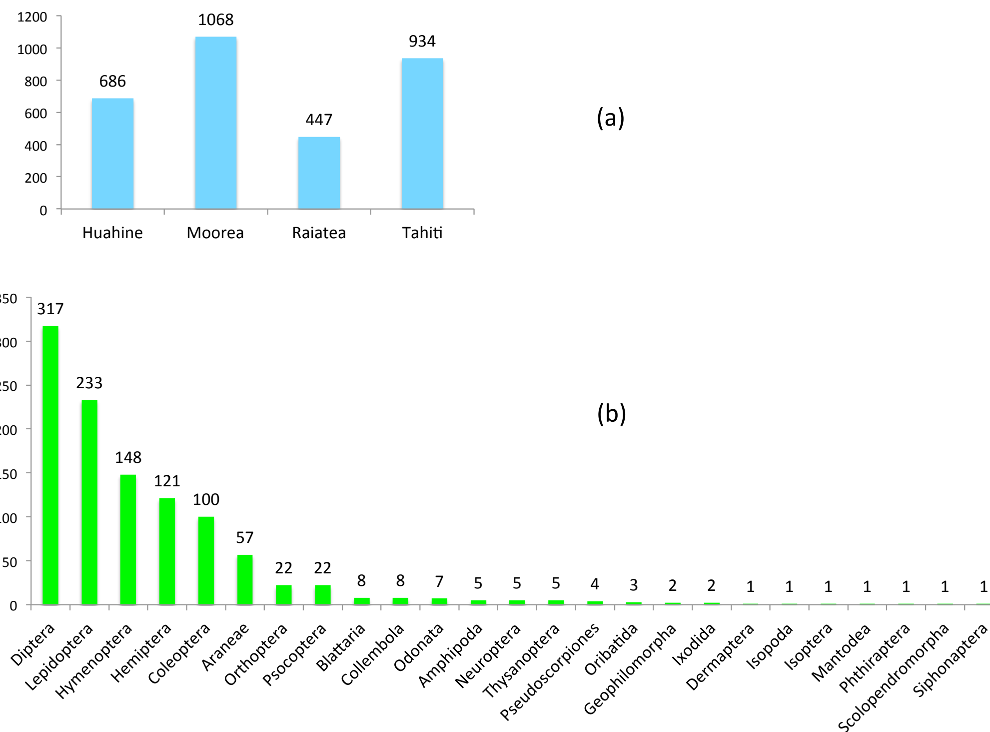
For the 3135 specimens for which we had a COI marker sequence, we were able to perform a more in-depth taxonomic classification, i.e., to determine the families for most of our specimens.

We started by clustering the COI sequences into groups with very low sequence divergence. We chose to use as cut-off for the maximum divergence between sequences within a cluster the value of 3%. As explained in the previous chapter, divergence was measured as the sum of the branch lengths between a pair of sequences. This cutoff value was chosen as it is frequently accepted as a threshold between the maximum intra-species divergence among COI lineages and the divergence observed among inter-species COI lineages. Even though this "universal" threshold for species delimitation is not without caveats (in particular, it assumes the existence of an arthropod molecular clock), it seemed a good choice for our type of data, i.e., for a dataset as large and diversified as ours. We are aware that these groups, with up to 3% COI divergence, will (most likely) not all correspond to true taxonomic species. However, our interest is

in approximately estimating the species richness, i.e., the total number of species that might be present in our dataset, and in producing a better taxonomic classification for specimens using only one individual per group of 3% divergence (henceforward referred to as clusters-3%).

A total of 1074 clusters-3% were obtained. Thus, we roughly estimate that 1074 species were collected in our sampling. This is one of the most extensive samplings performed to date in such a system. As can be seen in Figure 3.5, the clusters-3% are more or less homogeneously distributed among islands, except for Raiatea where less clusters (and less individuals) were collected, due to unfavourable weather conditions.

Figure 3.5: Distribution of arthropod clusters-3% in (a) the 4 islands sampled and (b) within each Order. All 3135 specimens for which a COI sequence was obtained with the HCO-LCO primers [87] were taken into account.



Not surprisingly, the more species-rich Orders are those for which we obtained more clusters-3% - Diptera, Lepidoptera, Hymenoptera, Hemiptera and Coleoptera. And Araneae for which we estimate to have covered $\sim$ 70% of their known species diversity in the Society Archipelago (further details in Chapter 2). Our sampling was in fact extensive and, although it does not represent the full diversity of species present in these islands, it allowed us to capture a large part of the diversity present.

Once we had the clusters-3%, we were interested in using both the photographs taken and the COI sequence information to better characterise taxonomically the specimens collected. The combined result of the BLAST searches, photograph analyses and help from experts – T. Ramage, R. Allemand, M. Dierkens – allowed us to determine a total of 140 families and their feeding habits (see Tables 3.1 and 3.2).

For some Orders, such as Diptera and Lepidoptera which are notoriously difficult to characterise taxonomically, we were able to obtain a better taxonomic classification for only $\sim 52 - 54\%$ of the specimens. In the case of latter Order, this was mainly due to the degradation of the wings (and associated patterns) conserved in alcohol. Nonetheless, as can be seen in Table 3.1, more than 20 families were identified for the most represented Orders in our dataset, i.e., Coleoptera, Diptera, Hemiptera, and Hymenoptera. The sole exception was Lepidoptera. In the case of spiders, 13 families were identified with the 57 COI clusters-3%. Overall, we were able to identify most of the specimens to the family level and the number of families found, that we estimate to be ≥ 140, once more highlights the success of our sampling effort. Particularly in Orders where more individuals were collected, we see that some families are more represented than others. Not surprisingly, these families are, in general, those that are known to be extremely species-rich. We have, for example, (1) the Curculionidae, Scarabidae, Staphylinidae and Carabidae families in Coleoptera; (2) the Noctuoidea, Geometroidea (both moths), Nymphalidaed and Lycaenidade families in Lepidoptera; (3) the Culicidae (mosquitoes), Tephritidae (fruit flies), Drosophilidae, Sarcophagidae and Tacchinidae families in Diptera; (4) many parasitoid wasps (comprising the Braconidae, Chalcididae, Trichogrammatidae, Dryiniidae or Ichneumonidae families), ants (Formicidae) and wasps (Vespoidea) in Hymenoptera.

On the long term, we expect to have more taxonomic resolution based on our photos and in collaboration with the UC Berkeley Moorea Biocode experts. Our samples will then provide a significant contribution to the ongoing barcoding of life project, which requires both a morphological and a COI marker characterisation of specimens.

Table 3.1: Families and number of COI clusters-3% identified. Only Orders for which one or more families were identified are represented in the Table. Legend: nbF = total number of families identified; spF = number of COI clusters-3% per family; ratio = proportion of species for which a successful family id was achieved. The total number of clusters-3% or "species" used for the ratio calculation is as indicated in Figure 3.5 (b).

Order/Family	nbF	spF	ratio	Order/Family	nbF	spF	ratio	Order/Family	nbF	spF	ratio	Order/Family	nbF	spF	ratio
Araneae	13	57	1.00	Rhizophagidae	-	1	-	Gerridae	-	1	-	Arctiidae	-	4	-
Anyphaenidae	-	2	-	Scarabaeidae	-	2	-	Issidae	-	2	-	Cosmopterigidae	-	2	-
Araneidae	-	8	-	Scolytidae	-	12	-	Lygaeidae	-	3	-	Crambidae	-	15	-
Clubionidae	-	2	-	Silvanidae	-	1	-	Miridae	-	12	-	Geometridae	-	10	-
Dictynidae	-	1	-	Staphylinidae	-	11	-	Nabidae	-	2	-	Immidae	-	5	-
Linyphiidae	-	1	-	**Collembola**	2	4	0.5	Pentatomidae	-	2	-	Lycaenidae	-	4	-
Miturgidae	-	1	-	Campodeidae	-	1	-	Plataspididae	-	1	-	Noctuidae	-	47	-
Pholcidae	-	2	-	Entomobryidae	-	3	-	Psyllidae	-	1	-	Nymphalidae	-	3	-
Salticidae	-	12	-	**Diptera**	20	164	0.52	Rhyparochromidae	-	6	-	Pterophoridae	-	1	-
Sparassidae	-	1	-	Agromyzidae	-	3	-	Scutelleridae	-	2	-	Pyralidae	-	8	-
Tetragnathidae	-	11	-	Bombyliidae	-	1	-	Tingidae	-	1	-	Sphingidae	-	4	-
Theridiidae	-	12	-	Calliphoridae	-	9	-	Tropiduchidae	-	1	-	tineidae	-	1	-
Thomisidae	-	3	-	Culicidae	-	6	-	Veliidae	-	2	-	Tortricidae	-	21	-
Uloboridae	-	1	-	Dolichopodidae	-	12	-	**Hymenoptera**	21	142	0.96	**Mantodea**	1	1	1.00
Blattaria	3	8	1.00	Drosophilidae	-	58	-	Apidae	-	4	-	Mantidae	-	5	-
Blaberidae	-	2	-	Limoniidae	-	9	-	Bethylidae	-	3	-	**Neuroptera**	1	5	1.00
Blattellidae	-	5	-	Muscidae	-	5	-	Braconidae	-	25	-	Chrysopidae	-	5	-
Blattidae	-	1	-	Mycetophilidae	-	4	-	Chalcididae	-	2	-	**Odonata**	2	7	1.00
Coleoptera	25	100	1.00	Neriidae	-	4	-	Chrysididae	-	1	-	Coenagrionidae	-	4	-
Aderidae	-	1	-	Phoridae	-	5	-	Crabronidae	-	9	-	Libellulidae	-	3	-
Anthribidae	-	8	-	Pipunculidae	-	1	-	Diapriidae	-	1	-	**Orthoptera**	3	22	1.00
Aphodiidae	-	1	-	Platystomatidae	-	3	-	Dryinidae	-	1	-	Gryllidae	-	18	-
Buprestidae	-	1	-	Psychodidae	-	5	-	Eulophidae	-	3	-	Mogoplistidae	-	1	-
Cantharidae	-	1	-	Sarcophagidae	-	5	-	Evaniidae	-	1	-	Tettigoniidae	-	3	-
Carabidae	-	7	-	Stratiomyidae	-	5	-	Figitidae	-	4	-	**Psocoptera**	8	18	0.82
Cerambycidae	-	5	-	Syrphidae	-	9	-	Formicidae	-	24	-	Caeciliusidae	-	3	-
Chrysomelidae	-	2	-	Tachinidae	-	6	-	Ichneumonidae	-	18	-	Mesopsocidae	-	1	-
Clavicornea	-	5	-	Tephritidae	-	6	-	Megachilidae	-	2	-	Myopsocidade	-	1	-
Coccinellidae	-	9	-	Tipulidae	-	8	-	Mymaridae	-	2	-	Peripsocidae	-	6	-
Corylophidae	-	1	-	**Hemiptera**	23	110	0.91	Platygastridae	-	6	-	Philotarsidae	-	1	-
Cucujidae	-	1	-	Alydidae	-	2	-	Scelionidae	-	9	-	Psocidae	-	3	-
Curculionidae	-	9	-	Anthocoridae	-	3	-	Sphecidae	-	3	-	Stenopsocidae	-	2	-
Elateridae	-	4	-	Aphididae	-	11	-	Torymidae	-	1	-	Trichopsocidae	-	1	-
Endomychidae	-	1	-	Aradidae	-	2	-	Trichogrammatidae		15	-	**Scolopendromorpha**	1	1	1.00
Hydrophilidae	-	1	-	Cercopidae	-	7	-	Vespidae	-	8	-	Scolopendridae	-	-	-
Mycetophagidae	-	1	-	Cicadellidae	-	37	-	**Isoptera**	1	1	1.00	**Siphonaptera**	1	1	1.00
Nitidulidae	-	5	-	Coreidae	-	1	-	Kalotermitidae	-	-	-	Pulicidae	-	-	-
Oedemeridae	-	5	-	Cydnidae	-	2	-	**Ixodida**	1	2	1.00	**Thysanoptera**	1	1	0.2
Platypodidae	-	4	-	Delphacidae	-	7	-	Ixodidae	-	-	-	Thripidae	-	-	-
Ptilodactylidae	-	1	-	Derbidae	-	2	-	**Lepidoptera**	13	125	0.54				

Table 3.2: Feeding habits of the families identified in our dataset.

Araneae	*Phytophagous/Saproxylophagous*	*Nectarivore/Predator*	*Predator*	Cosmopterigidae
Predator	Elateridae	Syrphidae	Gerridae	Crambidae
Anyphaenidae	*Predator*	*Nectarivore/Predator-/Parasitoide*	Nabidae	Geometridae
Araneidae	Carabidae	Bombyliidae	Veliidae	Immidae
Clubionidae	Coccinelidae	*Parasite*	*Predator/Phytophagous-/Granivorous*	Lycaenidae
Dictynidae	Coccinellidae	Pipunculidae	Geocoridae	Noctuidae
Linyphiidae	Staphylinidae	*Phytophagous*	**Hymenoptera**	Nymphalidae
Miturgidae	*Saprophyte*	Agromyzidae	*Nectarivore*	Pterophoridae
Pholcidae	Aderidae	*Phytophagous/Saprophyte*	Apidae	Pyralidae
Salticidae	*Saprophyte/Carpophagous*	Platystomatidae	Megachilidae	Sphingidae
Scytodidae	Nitidulidae	*Predator*	*Parasitoide*	Tortricidae
Sparassidae	*Saprophyte/Coprophagous*	Dolichopodidae	Bethylidae	**Mantodea**
Tetragnathidae	Scarabaeidae	*Saprophyte/Phytophagous*	Braconidae	*Predator*
Theridiidae	*Saprophyte/Mycetophagous-/Saproxylophagous*	Limoniidae	Chalcididae	Mantidae
Thomisidae	Anthribidae	*Saprophyte/Predator*	Chrysididae	**Neuroptera**
Uloboridae	*Saprophyte*	Stratiomyidae	Diapriidae	*Predator*
Blattaria	Ptilodactylidae	*Mycetophagous/Phytophagous-/Hematophagous*	Dryinidae	Chrysopidae
Omnivore	*Saproxylophagous*	Psychodidae	Eulophidae	**Odonata**
Blaberidae	Buprestidae	**Hemiptera**	Evaniidae	*Predator*
Blattellidae	Cantharidae	*Granivorous/Mycetophagous-/Phytophagous*	Figitidae	Coenagrionidae
Blattidae	Cerambycidae	Lygaeidae	Ichneumonidae	Libellulidae
Coleoptera	*Xylophagous*	Rhyparochromidae	Platygastridae	**Orthoptera**
Carpophagous/Xylophagous	Scolytidae	Mycetophagous	Scelionidae	*Omnivore*
Oedemeridae	**Collembola**	Aradidae	Sphecidae	Gryllidae
Granivorous	*Saprophyte*	*Phytophagous*	Torymidae	Mogoplistidae
Rhizophagidae	Campodeidae	Aphididae	*Parasitoide/Predator*	*Phytophagous/Predator*
Coprophagous	Entomobryidae	Cercopidae	Crabronidae	Gryllidae
Aphodiidae	**Diptera**	Cicadellidae	*Predator*	Tettigoniidae
Coprophagous/Saprophyte-/Predator	*Carpophagous*	Coreidae	Formicidae	**Scolopendromorpha**
Hydrophilidae	Drosophilidae	Cydnidae	*Predator/Nectarivore*	*Predator*
Mycetophagous	Neriidae	Delphacidae	Vespidae	Scolopendridae
Corylophidae	Tephritidae	Derbidae	**Isoptera**	**Thysanoptera**
Endomychidae	*Mycetophagous*	Membracidae	*Xylophagous*	*Phytophagous/Predator*
Mycetophagidae	Mycetophilidae	Pentatomidae	Kalotermitidae	Phlaeothripidae
Platypodidae	*Necrophagous/Coprophagous*	Plataspididae	**Ixodida**	-
Mycetophagous/Granivorous	Calliphoridae	Scutelleridae	*Hematophagous*	-
Silvanidae	*Necrophagous/Parasite*	Tingidae	Ixodidae	-
Phytophagous	*Sarcophagidae*	Tropiduchidae	**Lepidoptera**	-
Chrysomelidae	*Nectarivore/Hematophagous-/Phytophagous*	*Phytophagous/Predator*	*Phytophagous/Nectarivore*	-
Curculionidae	Culicidae	Miridae	Arctiidae	-

3.1.2 Biogeography of Mitochondrial in the Society Archipelago

Another question that we might ask is how are the COI lineages distributed geographically. Can we find spatial patterns that might tell us how colonisation(s) occurred in these islands or give an indication on the frequency of migration events? Before addressing these questions, we need to recall how this archipelago was formed.

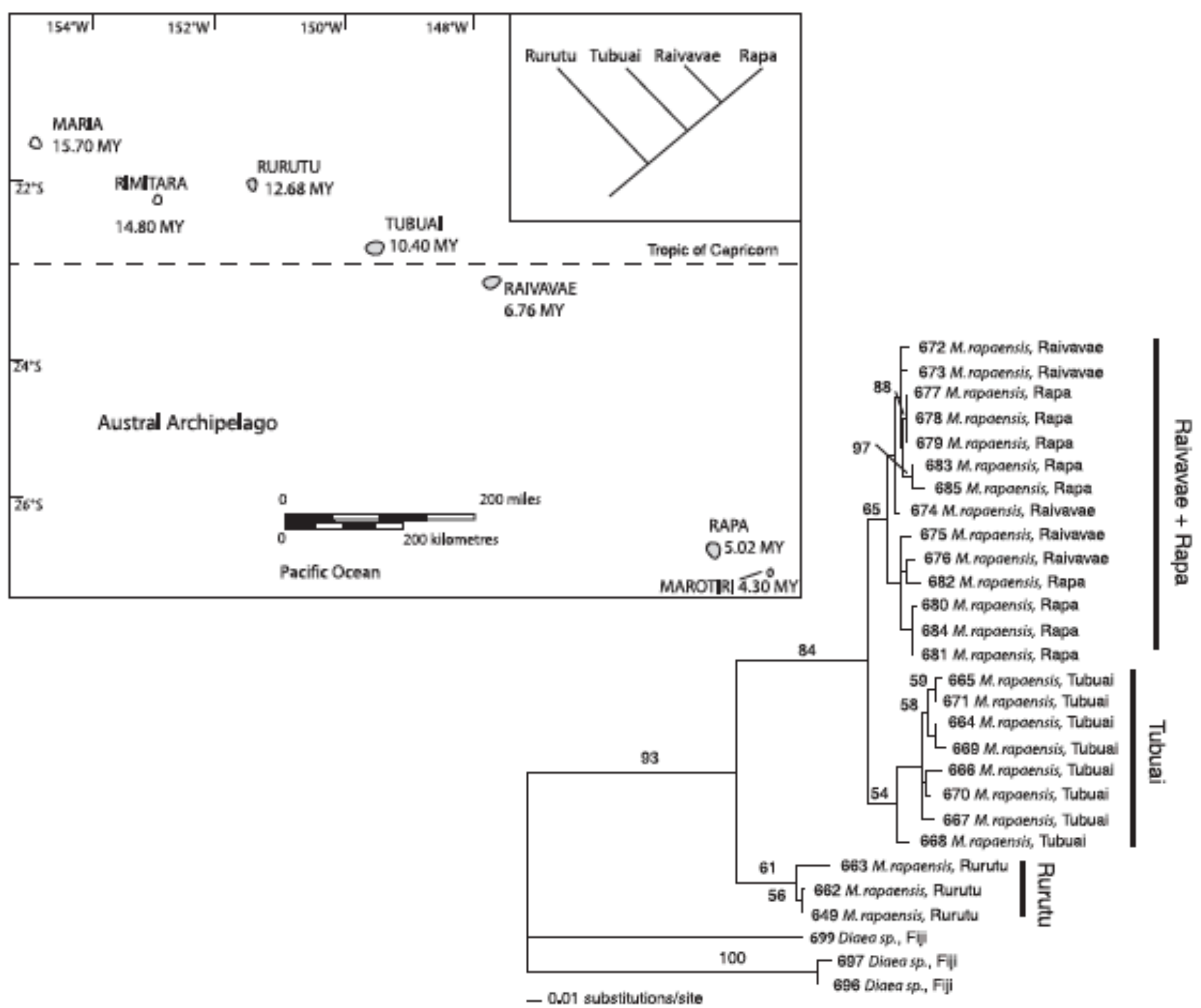
The Pacific plate moves steadily north-westwards and most of the central Pacific islands are formed by chains of volcanoes, each sequentially younger than the preceding one, following the movement of the plates over stationary hot spots [96]. As a consequence, until recently, the prevailing paradigm in studies of the Pacific biogeography was that colonisation of the central Pacific archipelagos was mainly made by oceanic dispersal of colonisers originating from the western Pacific Rim, either large continental masses such as Australia and southeast Asia, or from continental islands such as New Guinea [81]. According to this hypothesis, colonisation was then made from the older (western) archipelagos, these serving as stepping stones for the colonisation of the more remote (eastern) ones. The same "progression rule" would then be seen within the archipelago, with colonisation generally from the older hot spot islands to the younger ones (Figure 3.6).

While work with some plants [98, 99], *Simulium* black flies, *Rhyncogonus* weevils and *Partula* snails [81] in the Hawai and French Polynesia (Society, Marquesas and Austral) archipelagos support this hypothesis of a western colonisation in a stepping-stone-like manner through the central Pacific islands, more recent works with plants and spiders (jumping and crab spiders) point to stronger affinities with American (eastern) lineages [81]. Moreover, evidence for both single colonisations and for multiple or repeated colonisations within archipelagos exist for the Hawai, Society and Marquesas archipelagos [81, 96]. In general, evidence from hot spot island chains, both in archipelagos organised in a linear succession of islands (e.g., Hawai [100, 77], Marquesas and Society [101] in the Pacific Ocean) or in clusters (e.g., Canaries in the Atlantic Ocean and Galapagos in the Pacific Ocean [93]), strongly suggest directional colonisation (founding) events from the oldest to the youngest islands within archipelagos. However, cases of back-dispersal/colonisation [94] and vicariant radiation [102] are also known.

Defining the statistics for spatial structuring

In order to study biogeographical patterns and migration rates with our dataset, we first need to test if there is a spatial structure in the data.

Figure 3.6: The progression rule pattern of islands colonisation within a linear chain archipelago hypothesis as inferred for crab spiders *M. rapiensis* from the Austral archipelago (FP). The Austral islands distribution (with estimated ages), the progression rule hypothesis (left) and an ML tree obtained from COI data, with values above the branches referring to bootstrap values from 100 replicates (right) supporting the successive colonisation from older to younger islands. Adapted from [97].



We predict that if there is such a structure, then the haplotypes present within an island should be less divergent between them than to the haplotypes present in other islands, i.e., intra-island divergence should be smaller than inter-island divergence. Our null hypothesis H_0 is, thus: "The spatial distribution of (phylogenetically) close COI lineages is independent of the island where they were sampled".

It is worth pointing out here that we are testing whether divergence is lower within islands than between islands. If the null hypothesis H_0 is not rejected, this means that our dataset supports the biological explanations of (1) high migration rates between the islands which will have erased any noticeable signal of geographical structure from the first colonisation events, and/or (2) very recent colonisation events and not enough time having passed for divergence to occur and accumulate between populations of the different islands. However, if we reject H_0, we are not able to know if the detectable signal supports a successive colonisation model or a physical distance-dependent model. We are only accepting the hypothesis that, within groups of closely related COI lineages sampled in all islands, the intra-island divergence accumulated between lineages has mainly occurred after the island colonisation (e.g: radiation events). This would mean that lineages from each island would group preferentially together and that we could then search for patterns of colonisation by looking at the phylogenetic relationship between these groups (see Figure 3.6).

To test for this, we needed to: (1) define groups of closely related COI lineages allowing us to address events occurring within the time frame of the islands formation, and (2) look at the distances between haplotypes from *(i)* clusters containing samples present in all 4 islands, and *(ii)* clusters with samples collected in 3 of the 4 islands and 3 only. We would like to stress that these two sets of clusters will be defined so as to be completely independent from one another. We can thus treat the lack of individuals from the fourth island as missing data. We only focused on the cases of 4 and 3 islands respectively (and not on clusters including 2 islands) because we were interested in determining migration frequencies but also biogeographical patterns if possible.

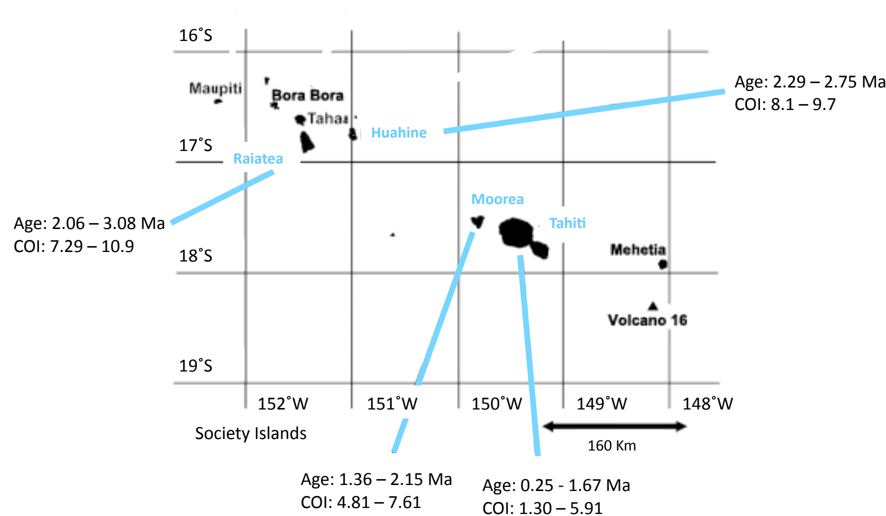
A threshold for COI clusters divergence in the time-frame of the islands formation

One important issue is to establish at which depth we should define the groups of phylogenetically close individuals to look for spatial patterns that are compatible with the formation of the islands. If we accept the strong assumption of a molecular clock in the phylogeny of the Arthropods, and if we use the age estimated for the formation of Raiatea (the oldest of the four islands) as the time of divergence between 2 sequences,

we can then make a rough estimation of the maximum divergence we would expect to have occurred between COI sequences after their most recent common ancestor (*mrca*) colonised Raiatea. We are also assuming that the colonisation of Raiatea occurred soon after the island formation.

Using the value of 3.54%.My^{-1} – as estimated for Insects in [103] – for the *cox1* gene rate of divergence (2*substitution rate) and the maximum estimate for the age of Raiatea as the time of divergence (3.08 My, see also Figure 3.7), we establish at $\sim$ 11% of divergence, measured as the average number of substitutions per site, the maximum depth at which we should be looking within the phylogeny of each Order.

Figure 3.7: Age estimates of the four Society islands in our study [104], in My, and corresponding % of divergence (calculated as the average number of substitutions per site) when we use the age of the islands as time of divergence and a molecular clock hypothesis with a rate of divergence of 3.54%.My^{-1} [103]. Legend: Age - lowest and biggest K-Ar (Potassium-Argon) estimates for island age formation, COI - % of divergence estimated using the lowest and biggest estimates of island age formation as time of divergence.



Before establishing which depth best allows us to address events occurring at the time of the formation of Raiatea or posterior to this event, we started by assessing the number of COI monophyletic clusters for different divergence thresholds (cutoff values). For each Order, the monophyletic clusters were produced by varying the cutoff value from 0 to the maximum COI divergence observed in the tree, by increments of 0.01. We expect that, as the COI divergence increases, the number of clusters decreases due to a fusion of some of the clusters formed when using a lower COI divergence as cutoff value and, consequently, the diversity and number of haplotypes within a cluster increases.

The results obtained are summarised in Table 3.3 for all Orders with a minimum of 4 sequences. In Table 3.3, the distribution of the number of clusters using as cutoff 2, 3, 5, 9 and 10% of divergence shows us that the Orders with more specimens present a bigger diversity of COI haplotypes, as would be expected.

Table 3.3: Number of COI sequences obtained per Order and number of clusters obtained with, respectively, less than 2, 3 ,5, 9 and 11% of COI divergence as measured by the branch-length in the ML tree obtained, for each Order, with the COI sequences amplified using primers LCO-HCO. Legend: haplos - distinct COI alleles found within an Order, Cluster2, Cluster3, Cluster5, Cluster9, Cluster10, Cluster11 - number of clusters obtained with, respectively, maximum 2%, 3%, 5%, 9% and 11% COI divergence between leaves.

Order	haplos	Cluster2	Cluster3	Cluster5	Cluster9	Cluster11
Diptera	497	331	317	292	269	257
Lepidoptera	341	242	233	219	204	201
Hymenoptera	171	149	148	145	135	135
Hemiptera	164	122	121	118	114	110
Coleoptera	123	101	100	98	95	95
Araneae	102	60	57	51	45	44
Orthoptera	40	26	22	20	16	13
Psocoptera	26	22	22	22	21	21
Odonata	14	7	7	7	7	6
Blattaria	11	9	9	9	9	7
Neuroptera	12	5	5	5	5	5
Thysanoptera	10	5	5	5	5	5
Collembola	10	9	8	8	8	8
Isoptera	4	1	1	1	1	1
Pseudoscorpiones	4	4	4	4	4	4

This diversity is seen at the level of 2 - 10% of divergence with a decrease in the number of clusters with increasing depth. For clusters with less than 40 sequences, the number of clusters is almost stable. Specimens sampled are either very closely related (identical or near identical COI sequences), or are phylogenetically very distant.

Taking into account the maximum divergence that we expect to be explained by the biogeography of the islands, as well as the distribution of the number of clusters with COI divergence for the 6 most represented Orders, we decided to use clusters at 9% and 11% of divergence. The former cutoff value is the percentage of divergence estimated when we use as time of divergence the mean between the minimum and maximum age estimates for Raiatea (2.57 My corresponding to $\sim$ 9.1% divergence). For the latter cutoff, as the percentage of divergence corresponding to the maximum age for the formation of Raiatea (3.08 My) is estimated to be of 10.9%, we used 11% as the maximum divergence observed between sequences of the same clusters that could be explained by a biogeographical pattern.

The clusters with a maximum of 9 and 11% COI divergence with sequences from all 4 islands concerned the same Orders, namely Diptera, Lepidoptera, Hymenoptera, Hemiptera, Coleoptera, Araneae and Odonata. These are the same Orders also for the clusters with sequences in only 3 islands, to which are added the Orders Blattaria, Isoptera, Neuroptera, Orthoptera and Psocoptera. In both cases (3 and 4 islands), the number of clusters does not depend on the cutoff used (9 or 11% divergence, see also Table 3.4).

Table 3.4: Orders with clusters sampled in all 4 islands and in only 3 islands. The number of clusters obtained using as cutoff 9% (cut9) and 11% (cut11) of divergence are indicated for both the 4 and the 3 islands cases. The number of clusters-11% with ≥ 8 individuals for the 4 islands case and with ≥ 6 individuals for the 3 island case are indicated, respectively, as size8 and size6.

	4 islands			3 islands		
Order	cut9	cut11	size8	cut9	cut11	size6
Araneae	4	4	3	8	8	8
Blattaria	0	0	0	1	1	0
Coleoptera	4	4	1	3	3	3
Diptera	11	11	11	22	22	14
Hemiptera	6	6	4	16	16	9
Hymenoptera	11	11	7	12	12	7
Isoptera	0	0	0	1	1	0
Lepidoptera	3	3	3	28	28	18
Neuroptera	0	0	0	1	1	0
Odonata	1	1	0	1	2	2
Orthoptera	2	2	0	1	1	0
Psocoptera	0	0	0	3	3	3
Total	43	43	29	97	97	64

Moran's Index and Joint Count

In order to test whether our null hypothesis holds, i.e., whether phylogenetic distances within a (sub)tree are not dependent on the island where the specimens were collected, we needed to choose a good statistic. A classical test for spatial structure is Moran's Index (I) autocorrelation coefficient, which has been extensively used to study the strength and direction of spatial dependence in phylogenetic data for quantitative traits [105, 106, 107, 108].

The spatial autocorrelation of a trait is obtained by looking at pairwise measures of that trait for pairs of specimens (represented by their locations in space) [105, 109]. For each point or location, we have a value of the trait and points are considered to be neighbours if they are connected. The strength of the connections depends on the distance between points and is taken into account in a *weight matrix* - w_{ij}.

In the case of phylogenies, all specimens are connected to all others - we have a fully connected graph - and the distance between specimens is now the divergence between them, as measured by the patristic distance. To transform divergence into weights, Gittleman & Kot [107] proposed the following function:

$$w_{ij} = \frac{1}{(d_{ij})^b} \tag{3.1}$$

where w_{ij} is the weight between species i and j, d_{ij} is the patristic distance between i and j and b is a coefficient parameter. Notice that the d_{ij} are strictly positive.

However, Moran's I index is used for continuous traits, while instead in our case we have a qualitative trait, that is, the island where specimens were collected. This variable takes four values: Raiatea, Huahine, Moorea and Tahiti.

We therefore used the Join Counts statistic, derived from the Moran's Index, for discrete variables. This statistics counts 3 quantities: (1) the number of joins/connections between points of the same state (J_{BB}):

$$J_{BB} = \frac{1}{2} \sum_{ij} w_{ij} (BB)_{ij}, \tag{3.2}$$

with $(BB)_{ij}$ being the connection between species i and j; (2) the number of joins between points of different states (J_{WB}):

$$J_{WB} = \frac{1}{2} \sum_{ij} w_{ij} (WB)_{ij}, \tag{3.3}$$

with $(WB)_{ij}$ being the connection between species i and j; and (3) the total number of joins between points of different states (J_{tot}):

$$J_{tot} = \frac{1}{2} \underbrace{\sum_{B=1}^{k-1} \sum_{W=B+1}^{k}}_{\text{all possible pairs WB}} \sum_{ij} w_{ij} (WB)_{ij}, \tag{3.4}$$

with k being the total number of states of the categorical value. The observed results are then compared to the expected values (μ_1) and variances (μ_2) for each quantity (for a detailed discussion, see [105] and references cited within). This statistics is asymptotically normal and allows for the standard normal deviate (SND) to be calculated. If we use the the cumulative J_{tot} statistic to test for the existence of a spatial structure in the phylogenetic distances, we can therefore reject the null hypothesis if J_{tot} has either

high values or very low values, as this means that individuals collected in different islands are generally phylogenetically closer than those collected in the same island. Due to the normal approximation, we can then, for a level of 5% significance, reject the H_0 if J_{tot} is lower than -1.96 or bigger than 1.96.

One remaining point to define, concerns the weight function (w_{ij}). One characteristic of our dataset is the presence of a reasonable number of very short and null branches in the phylogenies for all Orders. Thus, to avoid a null denominator, we calculate weights using an adaptation of equation (3.1):

$$w_{ij} = \frac{1}{(d_{ij} + a)^b} \quad (3.5)$$

where w_{ij} and d_{ij} are as defined previously in (3.1), and a and b are coefficient parameters that allow us to, respectively, use null distances and smooth the variation of w for small values of d_{ij}.

The choice for the parameters a and b is critical as accepting or rejecting the H_0 hypothesis will depend greatly on the weights chosen, as can be seen in Figure 3.8. We opted to start by using the set of values (0.05, 0.1), respectively, for parameters a and b because they allowed a near linear variation of w for very small distances. To test the robustness of our statistic, we then used the other 3 combinations of (a, b) values shown in Figure 3.8.

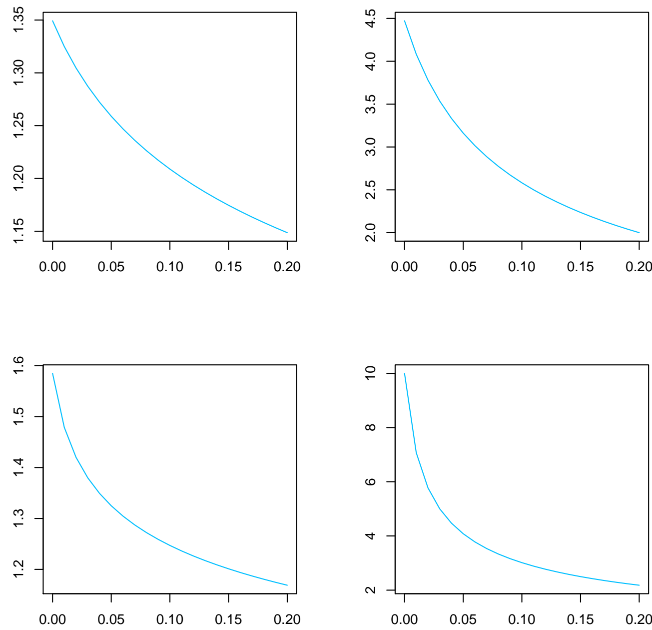
Analysis of a potential spatial structure in the four islands

We started by determining if our experimental set-up (e.g., collecting with a net) had an impact on the geographical distribution of the COI haplotypes. Using the COI clusters-3% (here seen as similar to species) for which: 1) two or more COI alleles were sequenced and 2) these alleles were obtained from specimens collected on two or more sites, we used the Morgan Index (I) to determine if identical alleles, within a species, were preferentially found on the same location. Although we had only 23 species on which this test could be performed, we did not obtain a signal for spatial structure. The probability of obtaining an I value at least equal to the observed was never lower than 0.20 (see Table 4.1 in Appendix). Thus, we have no indication that our experimental set-up might have introduced a confounding effect on the spatial structure of genotypes. We therefore could search for spatial structure in the four islands.

The Join Count with a 4-states statistics for autocorrelation testing was implemented using the R package *spdep*.

Although 43 and 97 clusters-11% were obtained, respectively, with individuals col-

Figure 3.8: Weight function w_{ij} calculated with 4 different combinations of values for the (a, b) parameters. In (a) w_{ij} with parameters (a,b)=(0.05,0.1), in (b) w_{ij} with parameters (a,b)=(0.05,0.5), in (c) w_{ij} with parameters (a,b)=(0.01,0.1), and in (d) w_{ij} with parameters (a,b)=(0.01,0.5). Legend: in the x axis we have d_{ij} and in the y-axis we have the values of the w_{ij} function.



lected in 4 and 3 islands (see Table 3.4), these clusters are predominantly composed by few individuals. As mentioned previously, the Joint Count test and Moran's I index are, like any statistics, sensitive to low sample sizes. In order to keep as much information as possible from the original dataset, we opted to use only the clusters-11% for which the mean number of individuals per island was ≥ 2 (see Table 3.4). It should be noticed that although we used clusters which have, on average, 2 or more individuals per island, this is not always true. Frequently the clusters present an uneven coverage of the islands.

Notwithstanding, we focused our study on clusters with at least 8 individuals collected throughout the 4 islands, as well as on clusters with 6 individuals sampled in only 3 islands.

This led us to a total of 29 and 64 COI clusters-11%, respectively, for the 4 and 3 islands cases. These clusters belonged to 6 Orders: Araneae, Coleoptera, Diptera, Hemiptera, Hymenoptera and Lepidoptera (see Table 3.4). In the case of 3 islands, we had also 2 clusters-11% from the Order Odonata.

For each group, we measured J_{tot}. If this value was inferior to *-1.96*, meaning a

strong negative correlation, then we rejected our H_0 hypothesis of no spatial autocorrelation between phylogenetic distances with a level of 5% significance and accepted that differently-coloured joins are less frequent than expected by chance. Using the J_{tot} statistics allows us to avoid the problems of: (1) oversampling when looking individually at single combinations of differently-coloured joins, and (2) the small sample size (n) which can highly affect the J_{BB}, with simulations strongly advised when n <10 (for further discussion, see [105]).

We started by looking at the clusters represented in all 4 islands. A first overall test of spatial structure was made using all 29 groups together. For this, we calculated:

$$J' = \frac{\Sigma J_{tot}}{\sqrt{n}} \tag{3.6}$$

where n is the number of clusters under analysis. We obtain J' equal to -7.57 which implies that, at a significance level of 5%, we reject the H_0 hypothesis of no spatial autocorrelation of phylogenetic distances. This strong spatial correlation was further confirmed with the 3 additional pairs of values for parameters (a,b), as described in Figure 3.8, with $J' \leq -7.57$ in all situations. But is this spatial signal given equally by all clusters-11% or, inversely, only a few of them show a very strong signal? To further answer this question, we looked at each cluster and calculated the J_{tot} statistics for each case. We determined that 23 clusters have negative J_{tot} statistics, i.e., $\sim$ 80% of the clusters, with 7 of these presenting strong spatial structure for at least 3 of the J_{tot} statistics, as can be seen in Table 3.5.

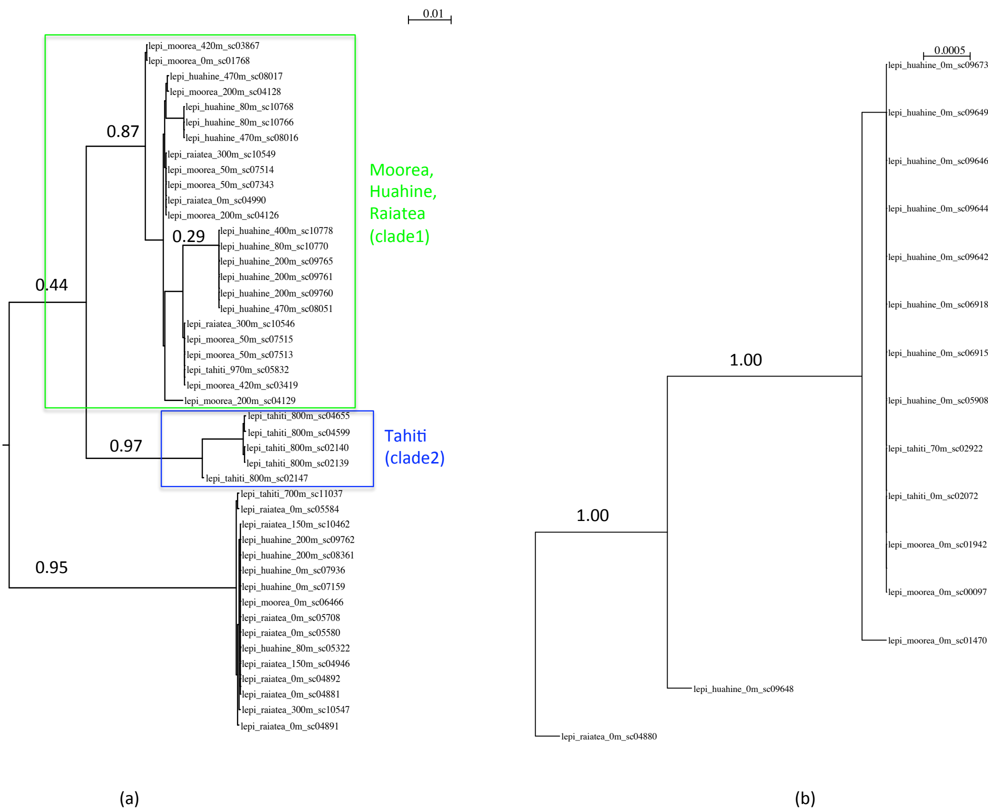
Table 3.5: J_{tot} obtained for the 4 islands COI clusters-11%, with $\geq$ 8 leaves, presenting a signal of spatial structure. Legend: Id - cluster id, Size - number of leaves in the cluster, Jtot1 - J_{tot} z-scores with parameters (a,b)=(0.05,0.1) for the weight-matrix, Jtot2 - J_{tot} z-scores with parameters (a,b)=(0.05,0.5), Jtot3 - J_{tot} z-scores with parameters (a,b)=(0.01,0.1), Jtot4 - J_{tot} z-scores with parameters (a,b)=(0.01,0.5). The * is used to indicate clusters with only 3 J_{tot} statistics indicating spatial structure. When the taxonomic family is not known we use a '-'.

id	Order	Family	Size	J_{tot1}	J_{tot2}	J_{tot3}	J_{tot4}
697	Lepidoptera	Crambidae	45	-6.71	-6.69	-6.63	-6.53
312	Diptera	Culicidae	24	-3.76	-3.76	-3.77	-3.77
278	Diptera	Neriidae	11	-3.22	-3.22	-3.19	-3.17
288	Diptera	-	13	-3.05	-3.07	-3.24	-3.33
492	Hemiptera	Cicadellidae	13	-2.93	-2.99	-3.33	-3.51
647	Hymenoptera*	Formicidae	26	-2.04	-2.03	-1.98	-1.94
714	Lepidoptera	Crambidae	15	-2.03	-2.03	-2.05	-2.07

These include clusters from the Lepidoptera, Diptera, Hemiptera and Hymenoptera Orders. In these clusters, the intra-island diversity is lower than the inter-island one,

based on the Join Count statistics. Thus, the overall frequency of migration seems to be lower in these groups. To determine migration patterns between the various islands, we decided to consider the tree topologies. When we looked at such tree topologies of each cluster and their node support - aLRT value - we found that only the 2 Crambidae clusters had good overall node support. In the remaining 5 clusters, we observed very low branch support. This can be due to sequences often differing by ambiguities and, consequently, leading to equally probable alternative topologies. Moreover, our method seems to be extremely sensitive to very small values. We indeed saw that some of our patristic distances can be extremely small, i.e., ≤ 0.001 in these clusters. Consequently, we opted to discuss here only the groups with good topological support (see Figure 3.9).

Figure 3.9: Phylogenetic trees of the Crambidade family clusters with strong spatial signal in the 4 island case. In (a) we find the subtree for the Crambidade cluster 697. The tahitian COI monophyletic group of COI alleles is shown in the blue box (clade 2). Clade 1 is the sister monophyletic group of clade 2. and in (b) we find the subtree for the Crambidade cluster 714. The aLRT branch support values are indicated. The islands where specimens were collected are indicated in the id used for specimens.



In the case of cluster 697, we see that the COI haplotypes from Tahiti seem to

be more divergent from the alleles in the other islands. In particular, we find a clade of Tahitian COI alleles collected in a same location (GPS: S 17 37 52.3, W 149 21 05.6) that is sister to another clade, this one of more closely related alleles distributed among the other 3 islands: Huahine, Moorea and Raiatea (clade1 in Figure 3.9). This could suggest that geographical isolation of this Tahitian population has occurred after the island colonisation, while migration between Raiatea, Huahine and Moorea occurs (more) frequently. With the second cluster of Crambidae moths, the COI allele found in the Raiatea specimen is more divergent in comparison to the remaining alleles. This would fit a scenario of migration between Huahine, Moorea and Tahiti being frequent for this group of closely related COI lineages and of low migration in Raiatea.

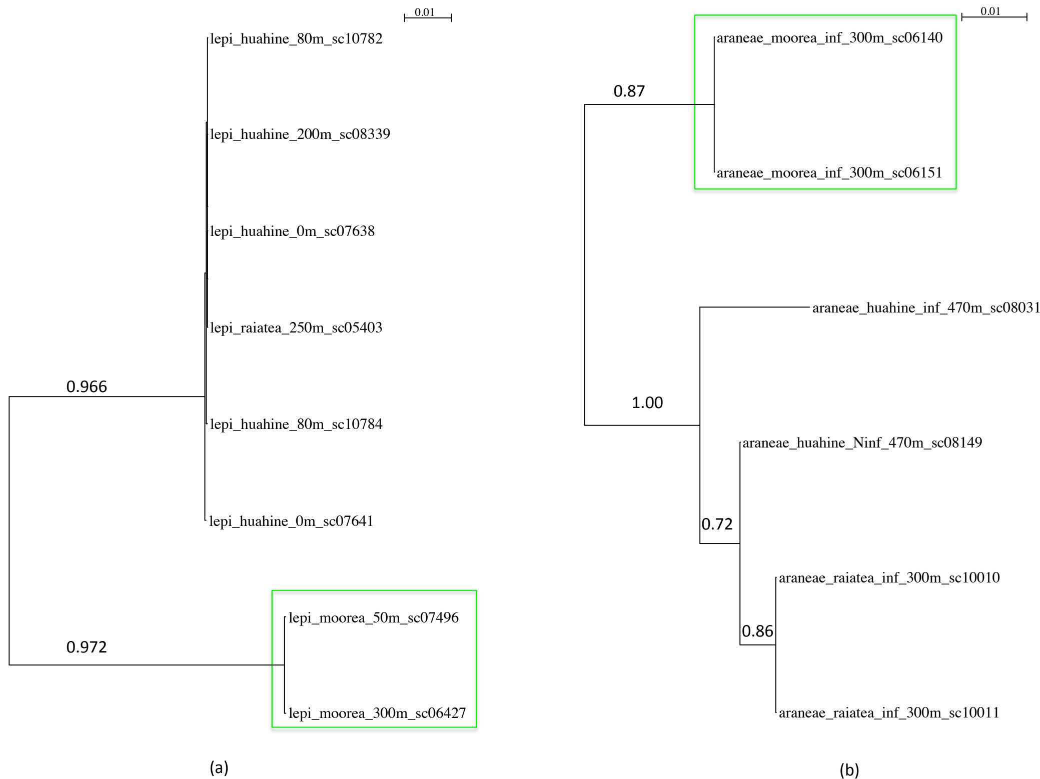
When we look at the 64 clusters with individuals from only 3 islands we also find an overall signal for spatial structure with $J' \leq -4.72$, with all 4 combinations of parameters (a,b). When we analyse each cluster individually, we find 7 clusters with strong spatial signal (see Table 3.6).

Table 3.6: J_{tot} obtained for the 3 islands COI clusters-11%, with $\geq$ 6 leaves, presenting signal of spatial structure. Legend: Size - number of leaves in the cluster, Jtot1 - J_{tot} z-scores with parameters (a,b)=(0.05,0.1) for the weight-matrix, Jtot2 - J_{tot} z-scores with parameters (a,b)=(0.05,0.5), Jtot3 - J_{tot} z-scores with parameters (a,b)=(0.01,0.1), Jtot4 - J_{tot} z-scores with parameters (a,b)=(0.01,0.5). When the taxonomic family is not known we use a -.

Id	Order	Family	Size	J_{tot1}	J_{tot2}	J_{tot3}	J_{tot4}
11	Araneae	Clubionidae	12	-3.67	-3.70	-3.85	-3.91
803	Lepidoptera	Noctuidae	11	-3.09	-3.11	-3.29	-3.37
877	Odonata	Libellulidae	9	-2.01	-2.01	-3.29	-3.28
739	Lepidoptera	-	8	-2.38	-2.38	-2.41	-2.44
282	Diptera	Syrphidae	7	-2.65	-2.65	-2.65	-2.65
805	Lepidoptera	Noctuidae	6	-2.42	-2.43	-2.47	-2.48
25	Araneae	Theridiidae	6	-2.16	-2.16	-2.19	-2.22

Curiously, after looking at the tree topologies and node support for these groups, we found 2 clusters with supported topological patterns. Despite the fact that they belonged to distinct Orders - Araneae (cluster 25) and Lepidoptera (cluster 739) - they both show a closer proximity between the Huahine and Raiatea COI alleles when compared to Moorea (see Figure 3.10). This fits a scenario where migration between Huahine and Raiatea is more frequent. As these older islands are nearer ($\sim 50Km$) to one another than to Moorea (more than $200Km$), this seems to suggest that in some clusters, migration may be distance-dependent.

Figure 3.10: Phylogenetic tree of the Lepidoptera cluster 805 and Araneae cluster 25 with strong spatial signal, in the 3 islands case. In (a) we find the Lepidoptera cluster phylogram and in (b) the Araneae cluster phylogram. The aLRT branch support values are indicated.



3.1.3 Discussion

A total of 11.022 specimens, from those collected in the field, were sorted and photographed. They encompassed 32 arthropod Orders, 4 subphylum – Chelicerata (spiders, scorpions, mites and ticks), Crustacea (aquatic arthropods), Myriapoda (millipedes, centipedes) and Hexapoda (springtails and insects) – and 4 islands.

To better characterise the diversity of arthropods collected, we used the *cox locus* (COI) as molecular marker. However, previous to the amplification and sequencing procedures, we reduced the number of specimens to be characterised. This first reduction of the dataset was based on 2 factors: (1) the fact that during the sorting process we found many specimens to be almost identical, i.e., to share many morphological characteristics which led us to believe that they were individuals from a same species, and as such they were grouped into a morpho-cluster; (2) the experimental (and financial) cost of amplifying and sequencing $\sim$ 11.000 individuals. Based on the distribution of specimens into morpho-clusters, we were able to reduce to 4671 specimens ($\sim$ 43%)

to be molecularly characterised (for details on this procedure please check Chapter 2).

As mentioned in Chapter 2, the choice of this mitochondrial *locus* was based on its maternal inheritance, fast evolving sequence and, particularly for this chapter, its standard use for species characterisation and the extensive knowledge already available for this *locus* (e.g.: in databases such as BOLD and the Moorea Biocode Project). We opted to use a pair of "universal" primers to amplify the *cox locus* – Folmer's LCO-HCO primers [87] – and we were able to have a 67% amplification and sequencing success.

As seen in Figure 3.4, some Orders were less successfully amplified than others. Extreme cases, i.e., cases for which we were not able to obtain a single COI sequence included, 2 myriapod Orders (Julida and Spirobolida) and chelicerates such as Mesostigmata and Thrombidiformes (mite Orders), scorpiones and oribatidians. In the case of the myriapod Orders, the lack of amplification and the number of individuals tested seems to suggest that the universal primers are not adequate for these groups. This might be due to mutation in the primers target region. In the case of the mite Orders, it is also possible that the extremely small size of the specimens tested ($\leq 1mm$) might be responsible for the lack of amplification success. Other cases of low sequencing success, i.e., $\leq$ 25%, include the Amphipoda, Isopoda and Collembola Orders. The former 2 belong to the crustacea subphylum and the later to the Hexapoda one. Again it could be a problem with the primers target sequence, as these are very divergent lineages to Class Insecta.

Overall, we were able to obtain COI data for 24 Orders, i.e., for 75% of the Orders sampled.

Once we had the COI marker information, we were able to construct phylogenies for each Order and, with the help of the previously taken photographs and BLAST results, to obtain a better classification of our specimens. We could characterise most individuals with the COI marker to the family level. The sole exception were the Diptera, Lepidoptera and Collembola Orders for which $\sim$ 50% of the species were attributed to a family. Notwithstanding, we could identify 140 distinct families in our set of individuals with a COI sequence, i.e., in 3135 specimens. Within these families, we find species with feeding habits that comprise omnivorous, predator, parasite, parasitoide, hematophagous, phytophagous, nectarivore, xylophagous, saprophyte, mycetophagous, coprophagous, granivorous and carbophagous. This covers almost all types of feeding habits and different habitats.

To estimate the number of species captured, we used the traditional barcoding cut-off of 3% divergence to define the clusters-3% as approximately equivalent to species.

Using only the 3135 individuals with a COI sequence, we thus estimate that we sampled $\sim$ 1074 species although we would like to insist that this estimation is but an approximation. Although we performed controls for heteroplasmy and nuclear mitochondrial pseudogenes (by controlling the chromatograms quality and checking for the presence of unexpected *stop* codons in the translated COI sequence), it is likely that in such a diverse group of animals we have variations in the substitution rate of the COI *locus* for different groups. Moreover, it is well documented that bacterial parasites such as *Wolbachia* can bias the mtDNA diversity in their hosts populations. This makes the use of barcoding for species delimitation a controversial topic as the number of species can be overestimated in some cases and underestimated in others.

Once we had a better classification of the diversity of arthropods with the COI marker, we wanted to see how this mitochondrial diversity was distributed in the 4 islands sampled. We were interested in detecting migration and infer possible biogeographical patterns in our dataset. For this, we used COI clusters with a maximum of 11% divergence, as this is the maximum divergence estimated with the oldest island (Raiatea) formation, under the assumption of an arthropod molecular clock calibrated to 3.54% divergence per My. We further focused our interest on the clusters-11% found in all 4 islands or in exactly 3 of them.

For each case separately, we started by assessing the presence of a spatial structure in the set of clusters found, that is we assessed whether the COI alleles of a given cluster were less divergent within an island than in relation to all the remaining islands. A strong signal for spatial autocorrelation was obtained. This suggests that, in general, we have low levels of arthropods migration between these islands. When we looked further in detail at each cluster, we confirmed that the majority of the clusters had a low signal (non-significant at 5%) but the cumulative effect of all clusters was sufficient to give a general pattern. However, a few clusters (7), for both the 4 and 3 islands cases, presented strong signals also at the individual level. But one has to keep in mind that our statistics, even for individual clusters, is itself a cumulative measure of all the connections between specimens in different islands. Thus, it does not allow us to see migration patterns and frequency of migration between the different islands. To determine migration and other biogeographical patterns, we need, once we have an indication that there is an overall spatial structure, to look at the tree topologies. In both cases (of 4 and 3 islands), we found only 2 clusters for which the tree topology was strongly supported and we could speculate about possible biological explanations. In the clusters with specimens in exactly 3 islands, migration seems to be more frequent between the 2 closest islands – Raiatea and Huahine – than with Moorea. This

means that migration in these cases may be distance-dependent. In the clusters with specimens collected in all 4 islands, we found 2 groups of Crambidae moths suggesting different migration patterns: (1) suggested that one population in Tahiti (the youngest island) may be geographically isolated and the migration is more frequent between the remaining 3 islands; (2) migration between Huahine, Moorea and Tahiti is frequent and less so in Raiatea.

In both the 4 and 3 islands cases, we found more clusters with a cumulative signal for spatial structure. But when we looked at the trees, the branch support was not very good and no clear topological pattern was observed. As mentioned previously, often our distances are very small and the sequences may diverge by ambiguities. One may then wonder on the meaning of divergences as small as 0.05%. This could be an artefact of the tree reconstruction method, particularly when we use sequences with varying length (from 450bp to 658bp). The use of evolutionary models which take into account multiple substitutions per site may also influence the final optimisation of branch lengths. Additionally, our method may be very sensitive to very small distances, although we did use 4 distinct versions of our statistics to test for robustness (by varying the parameters in the transformation of phylogenetic distances into weighted connections between specimens). Consequently, the signal from these clusters has to be taken with caution and to be further investigated.

Overall, it seems that some clusters present a strong signal for spatial structure in these islands. This is not surprising; on the contrary, this is to be expected in isolated systems such as islands. Although we have indications of possible patterns, this needs to be further tested to assess migration and possibly colonisation patterns.

What can we infer from the overall signal for spatial structure found in both the 4 and 3 islands cases? Indeed, if we consider the clusters which did not have a significant spatial signal, we see that 76% (22 out of 29) of the clusters present in all 4 islands and 89% (57 out of 64) of those present in only 3 islands do not reject the null hypothesis of "the spatial distribution of (phylogenetically) close COI lineages is independent of the island where they were sampled". One likely scenario is that migration between the islands is, in general, quite frequent and independent of the geographic distance between the islands. This could be due to human movements among the islands rather than to the dispersal capacity of the species in question. One should however not forget that we may be loosing signal in these groups because we have caught species with high dispersion rates and failed to capture enough endemics. This could have been further enhanced when we applied our filtering procedure to select the specimens to be sequenced. As described in Chapter 2, when selecting for representatives from the

morpho-clusters to be sequenced, we randomly picked their location within an island. We may thus be over-estimating the frequency of migration between these islands, by neglecting to use specimens representing indigenous species (e.g: those collected on higher altitudes).

We think that our dataset can become potentially more interesting to address these questions, if we are able to (1) further characterise the specimens at a taxonomical level and, concomitantly, (2) complement some of the groups to be further studied with more endemic species. This would imply, in a first step, a new selection of specimens for sequencing based on the improved taxonomy and, also, their collecting site. This time, we should focus on individuals collected preferentially at higher altitudes because these are less disturbed by humans. In a second step, we would ideally need to return to the Society archipelago for a second field work.

3.2 Incidence and diversity of *Wolbachia* infections

Wolbachia infections are known in many Arthropod Orders: Acari [110, 36], Amphipoda [111], Araneae [62, 112], Coleoptera [113, 22, 114], Diptera [115, 116], Hemiptera [117, 118], Hymenoptera [119, 64, 120], Isopoda [121], Isoptera [122], Lepidoptera [17, 123], Odonata [124], Orthoptera [114], Psocoptera [125], Pseudoscorpionida [126], Siphonaptera [127], and Thysanoptera [128]. These infections were often detected in studies of *Wolbachia*-induced phenotypes in specific groups of hosts with economical, agricultural or medical impact, or through more general PCR surveys. Such surveys have shown the overall frequency of *Wolbachia*-infected insect species from neotropical Panama to be of 16.9% [55], in North America of 19.3% [47, 62], in Britain of 21.7 - 22.8% [129, 9], in Japan of 35.1%, and in rice-field insect communities in Thailand of 23.4% [60, 118, 130]. Similar observations in different regions have led Werren and Windsor [47] to suggest that incidence, that is the overall frequency of infected species in arthropods, might have reached a worldwide equilibrium, meaning that the rate at which new arthropod species are infected is balanced by the rate at which already infected species lose their infections. This frequency was estimated in [47] to vary between 16 and 22%, based on observations made in three different major geographical regions (North America, Central America and the UK). Other studies showed that the presence of this symbiont is rather ubiquitous in the phylum Arthropoda with not only insects harbouring *Wolbachia* at high incidence. Infections were indeed found in fresh water crustaceans (Order Isopoda) [63], with an estimated 26% of the species infected, and more recent studies have established the widespread presence of *Wolbachia* in arachnids (class Arachnida) with infected species concerning spiders, mites, scorpions and pseudoscorpions [131, 132, 112, 133, 62]. In these studies, the proportion of infected species varied between 16.7% (in mites) [131] to up to 52.7% (Australian Araneae survey [132]).

Thus, *Wolbachia* infections are not only ubiquitous within arthropods but they also seem to be present in a high number of species within different Orders.

However, estimations have frequently been performed with one or a few individuals per species [55, 129, 60, 131]. This impacts the estimation of incidence mainly by underestimating its true value. Indeed, detection of an infection within a species is dependent on the frequency of infected individuals inside the species, i.e., on the symbiont's prevalence. Consequently, in studies with a low sample size per species, it is likely that incidence is underestimated. Additionally, false PCR negatives will have

a greater impact when only one specimen is tested. This will further contribute to a lower estimation of incidence.

In an attempt to account for both the underestimation induced by a low small sample size per species and the known variation in prevalence (that is, the intraspecific frequency), Hilgenboecker and colleagues [16] performed a meta-analysis of the *Wolbachia* screenings done in the last decade and estimated that up to 66% of the world's arthropod species should be expected to harbour infections by *Wolbachia* (notably, this analysis defines a species as infected if it carries infection with a prevalence as low as 1%). This work put a new, more realistic estimate on the (global equilibrium) frequency and suggested that *Wolbachia* is one of the most abundant intracellular bacterial genera known to date ($\sim 10^6$ insect species alone are thus estimated to be infected).

With this first extensive sampling of arthropod communities from young, isolated insular systems, we were interested in assessing the incidence of *Wolbachia* infections in these communities and seeing if it is in agreement with previous estimations and with the global equilibrium hypothesis. We also wanted to characterise the diversity of *Wolbachia* infections found in our system and see how this diversity is distributed between and within Orders, and then to compare this with what is already known from mainland studies.

One thing that should be kept in mind throughout is that our dataset is expected to be highly rich in species content (due to the extensive field sampling that was done) but with few individuals per species (due to the filtering process used with the morpho-cluster definition, see Chapter 2 for further details). Incidence estimations are therefore expected to be near the values obtained in previous screening studies [55, 129, 47, 131, 62, 9].

3.2.1 Incidence

We tested the presence of *Wolbachia* in the 4860 individuals after the initial filtering described in Chapter 2.

For this screening, we used the new 16S rDNA primers 553F_W/1334R_W (for further details see [134] in Chapter 2).

As mentioned in the previous section, we were able to sequence the mitochondrial COI *locus* for 3135 specimens, corresponding to 1547 distinct COI haplotypes. Concomitantly, in a first PCR using the 553F_W/1334R_W primers, we detected *Wolbachia* in 1177 hosts (of the 3135 specimens). A second PCR was carried out using the FbpA-F1/FbpA-R1 primers [44] to (1) confirm the presence of Wolbachia using this

second independent PCR and (2) to characterise the *Wolbachia* strains detected. This resulted in 1001 infections being confirmed by both PCR reactions. This result is in agreement with our previous results [134] of $\sim 85\%$ of known infections being detected by both primers. It should be noted that this second PCR was only performed on the 1177 individuals previously tested positive with the 16S rDNA primers. Thus, only infections confirmed by both screening protocols were used to estimate the symbiont's frequency of infections in arthropods.

Variation in the FbpA-PCR success rate among host orders The success of the second PCR seems to vary with the Order being tested, as can be seen in Table 3.7 (column % Confirmed infections). Coleopterans seem to have a lower success with the second PCR when compared to other Orders such as Diptera, Lepidoptera, Hemiptera, Hymenoptera, Araneae, Psocoptera and Orthoptera. Observe that only Orders for which 20 or more distinct host COI lineages tested positive for infection, with both PCR reactions, were kept for further analysis, while the remaining Orders were excluded due to a very small sample size. A $\chi 2$ test (contingency table, p<<0.01) confirms that the success of the FbpA PCR depends on the Order being tested. These results seem to indicate that, although we are using primers specific to the symbiont, we might have a host effect on the success of its detection. Coleopteran infections appear thus to be less successfully detected with the FbpA primers. Notably, specimens for which the second PCR did not confirm the infection status were randomly distributed in various 96-well plates, and the same conditions were used for all specimens, so that experimental errors do not represent a likely explanation for this observation. Another possible explanation is that our 16S primers have a higher false positive rate in this group of insects. This could be due to the presence in the host micro-flora of other bacteria which are phylogenetically close to the *Wolbachia* genus. Although we tested for *Wolbachia* specificity with our 16S primers (for further details see [134]), we cannot overrule this hypothesis. Alternatively, we can have more divergent *Wolbachia* infections in Coleoptera, diverging at the FbpA primer target regions. Our results would then point to a scenario where the diversity of *Wolbachia* may be dependent on the host Order.

Sequencing of the 16S products for the unconfirmed infections and detection with other *Wolbachia* specific primers (e.g., for the *wsp* and *ftsZ loci*) might shed some light into the reasons for the observed variation.

Table 3.7: Detection of COI lineages infected by *Wolbachia*. The success in *Wolbachia* detection with 16S and FbpA primers is indicated per Order. The percentage of 16S positive infections confirmed by FbpA and the proportion of COI lineages found to be infected are also indicated. Note that only Orders above the single horizontal line, i.e, with more than 20 distinct host COI lineages confirmed *Wolbachia* infections (positive for both primers), were included in the statistical analysis to test an Order effect in the success of *Wolbachia* detection. Exclusion of the remaining Orders was due to small sample sizes. Legend: PCR1- number of distinct COI lineages with positive infections as detected with the 16S primers, PCR2- number of distinct COI lineages with positive infection status as detected with both the 16S and the FbpA primers, COI - number of host distinct COI lineages. An horizontal line was used to separate the Orders with more than 10 positive infections, as confirmed by both primers.

Orders	PCR1	PCR2	% confirmed infections	COI	% infected COI lineages
Diptera	219	179	0.82	496	0.36
Lepidoptera	131	106	0.81	341	0.31
Hymenoptera	73	60	0.82	171	0.35
Hemiptera	101	91	0.90	164	0.55
Coleoptera	28	16	0.57	116	0.14
Araneae	38	32	0.84	102	0.31
Orthoptera	16	15	0.94	40	0.38
Psocoptera	21	21	1.00	26	0.81
Odonata	5	4	0.80	14	0.29
Neuroptera	1	1	1.00	12	0.08
Blattaria	3	2	0.67	11	0.18
Thysanoptera	3	3	1.00	10	0.30
Collembola	5	4	080	10	0.20
Amphipoda	1	1	1.00	5	0.20
Isoptera	3	2	0.67	4	0.50
Pseudoscorpionida	3	2	0.67	4	0.50
Oribatida	1	1	1.00	3	0.33
Scolopendromorpha	1	0	0.00	2	0.00
Dermaptera	1	1	1.00	1	1.00
Siphonaptera	1	1	1.00	1	1.00
Total	654	543	nap	1548	nap

Variation in Wolbachia incidence between orders Given this, only *Wolbachia* infections confirmed by both screening protocols were used to estimate the symbiont's frequency of infections in arthropods. As mentioned, 1001 hosts, corresponding to 543 distinct COI haplotypes, were found to be harbouring *Wolbachia* symbionts. Thus, considering each haplotype as a distinct maternal lineage, we estimated that 35% (543 out of 1548) of such lineages carry Wolbachia.

When we look at the infection frequency at the Order level in Table 3.7, Coleoptera has very few infected COI lineages (14%) while Psocoptera (81%) and Heminoptera (55%) have a high number of infected COI lineages. We then proceeded to test if the frequency of infections varied significantly between Orders, For this we assumed that the infection status within Arthropods (and, consequently, within Orders) did not

depend on the host species maternal lineage, i.e., that all arthropod COI lineages are equally likely to be infected by *Wolbachia* symbionts. A χ^2 test rejects ($p<0.0005$) the hypothesis that the frequency of infected host lineages is independent of the Order being studied. Furthermore, pairwise comparisons using 2x2 contingency tables and Bonferroni corrections for multiple hits confirmed that the Order Psocoptera is significantly more infected than the remaining Orders except for Hemiptera (adjusted p-value = 0.015), and that Coleoptera is significantly less infected than the remaining Orders (adjusted p-value < 0.002). A correction for identical maternal lineages testing positive for infection was made whenever these cases arose. Only one haplotype was considered for the estimation of the frequency of infected mitochondrial lineages both within Orders and in the overall arthropod group.

Although there seems to be a difference in the rate of infection between different Orders, this could be partially explained by various individuals with nearly identical COI haplotypes testing positive for *Wolbachia* infection. In such cases, polymorphism of the infection frequency at the intra-specific level could interfere with the inter-species estimation of *Wolbachia* frequency. A simple scenario where this could be seen is the following. Let us assume a host population with: 1) polymorphism in the COI sequence (*cox locus*), 2) a fixed (or nearly fixed) *Wolbachia* infection in this population, with 3) no linkage disequilibrium observed between the *Wolbachia* infection and the host COI lineages. Then various individuals carrying distinct COI haplotypes are sampled from this population and found to be infected. If incidence is estimated based solely on the frequency of different COI haplotypes associated with the parasite, then a bias towards an overestimation of the across-species frequency of *Wolbachia* infection is created when we use individuals from this population. In other words, a unique COI haplotype (maternal lineage) cannot be seen as strictly equivalent to a host species.

To avoid this intermingling of intra- and inter-species levels and further study incidence, we needed to better approximate species in our dataset, i.e., we needed to define and apply a measure allowing us to estimate the number of species that had been collected (based on COI divergence).

Investigating incidence of infected species Until now, we had been working with COI sequences obtained from individuals initially classified into morpho-clusters. As discussed previously (see Chapter 2), this clustering was based on shared morphological traits. Once the COI marker information was available, we opted to use the host's molecular information to obtain a measure of the diversity of host species sampled within each Order. Although this would not give an unambiguous species classification,

it allowed us to have an objective and unambiguous criterion (phylogenetic distance between host COI sequences) that could be applied to all Orders.

Again, we used the COI clusters with a maximum of 3% divergence (from now on referred to as *clusters-3%*) defined in the previous chapter.

Once this measure for species delimitation had been defined, we were able to assess more accurately the question of the overall incidence of *Wolbachia* infections in our arthropod community and whether incidence seemed to vary among Orders. We found 421 of the 1074 clusters-3% to have at least one infected individual. Due to the frequency in our data of clusters made of a single specimen ($\sim$ 28% of the infected clusters-3%), we opted to use the presence of one infected individual as sufficient evidence for infection detection, even though this does not allow us to have an extra control for false positives due to possible contaminations at the DNA extraction and PCR steps.

We could therefore estimate that approximately 39% of the insular arthropod "species" harbour an infection by *Wolbachia*. This is a slightly higher estimate than those obtained in previous widespread surveys (16.6% and 19.3%, [47] or 22.8%, [9]). A χ^2 test rejects (p« 0.0001) the hypothesis of similar infection frequencies in these studies and ours. Can we deduce from this that insular arthropod communities are more infected than their mainland counterparts? It is possible although we should be cautious with this result. As can be seen in Table 3.8, we have many cases of single infections being detected within our COI clusters-3% (and also many cases of one-specimen clusters). In total, only $\sim$ 40% of these host clusters have 2 or more individuals testing positive for infections. Thus, we cannot overrule the possibility that some false positives are being taken into account in our estimate. These single infections need to be confirmed with sequencing and, if possible, screening of more individuals from the same morpho-species.

In Table 3.8, we can see that the estimated incidence varies among the Orders from 82% in Psocoptera to as low as 15% in Coleoptera. Except for Psocoptera and Hemiptera, all other Orders with more than 20 clusters-3% have a slightly higher incidence than was previously reported from mainland studies [47, 9] but remain concordant with them. A χ^2 test ($p<0.0005$, simulated p-value with 2000 replicates) rejected the independence of the number of infected clusters-3% from the arthropod Order being tested. Once more, Orders with small sample sizes, i.e., all Orders with 20 or less clusters were removed from this analysis.

Two interesting cases are the Orders Psocoptera and Hemiptera for which incidence values, in our system, were very high, particularly for Psocoptera. Various factors

Table 3.8: Comparison between the rate of infected COI haplotypes and the incidence, i.e., the proportion of infected species estimated using COI clusters with a maximum of 3% divergence (clusters-3%) as proxy to the number of species present in our dataset. Here we show only Orders for which more than 10 COI sequences are available. Observe that only Orders above the single horizontal line, i.e., with more than 20 clusters-3%, were included in the statistical analysis. Legend: size1 - number of infected clusters-3% with only one infected individual and in parenthesis the number of these clusters for which only one individual was sampled. Abbreviations: nb = number, clusters = clusters-3%

Order	nb clusters	nb infected clusters	% infected COI	% infected clusters	size1
Diptera	317	122	0.36	0.38	72(35)
Lepidoptera	233	81	0.31	0.35	46(22)
Hymenoptera	150	58	0.35	0.39	38(26)
Hemiptera	121	72	0.55	0.6	34(21)
Coleoptera	100	15	0.14	0.15	10(5)
Araneae	57	25	0.31	0.44	11(5)
Orthoptera	22	11	0.38	0.5	5(3)
Psocoptera	22	18	0.81	0.82	7(6)
Odonata	7	3	0.29	0.43	2(1)
Neuroptera	5	1	0.08	0.2	1(1)
Blattaria	8	2	0.18	0.25	1(1)
Thysanoptera	5	3	0.30	0.6	1(1)
Collembola	8	4	0.20	0.5	4(3)
Pseudoscorpiones	4	2	0.5	0.5	2(2)

can contribute to these high values, among them are the phylogenetic closeness of the Orders (sister clades) and another is the ecology of the species collected. In Hemiptera, we found representatives from families whose feeding habits range from phytophagous to predators and saprophytes (scavengers), while psocopterans are generally saprohytes. Could these very high incidences be due to the fact of being part of similar ecological networks where the horizontal transfer of *Wolbachia* infections may be frequent?

3.2.2 Diversity of *Wolbachia* infections as inferred using the FbpA marker

We estimated previously that as many as 39% of the host clusters-3% were infected with *Wolbachia*. We are now interested in assessing the diversity of these infections in the arthropod communities sampled and in seeing how this diversity of symbionts appears distributed within the Orders. To address these questions, we proceeded to sequence the *fbpA locus* of the 1001 individuals tested positive in both PCR essays.

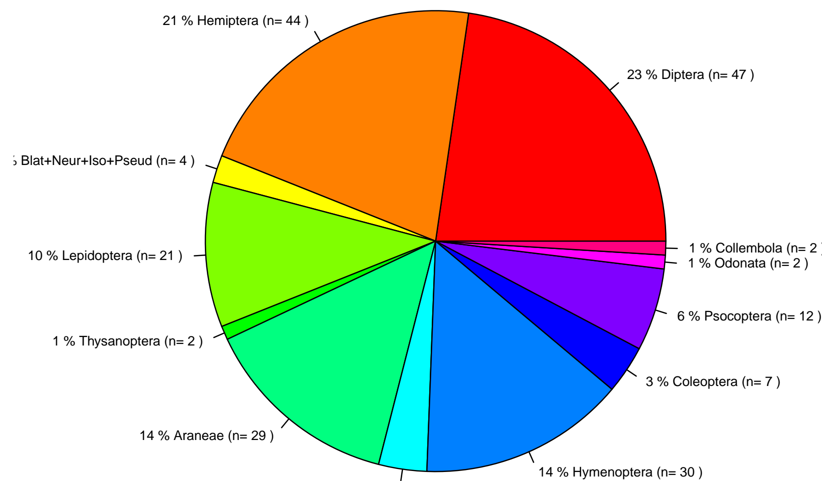
After sequencing, we were able to characterise a total of 616 *Wolbachia* infections with this *locus*, i.e., approximately 60% of the positive infections. Two important factors can explain the 40% failure: 1) some infected specimens yielded a poorly con-

centrated FbpA PCR product, so that the sequencing reaction failed and 2) there could have been co-infection of individuals, i.e., more than one *Wolbachia* infecting the same host. While the former factor cannot be overruled, we think the latter is more plausible because the presence of co-infections (double or triple) with *Wolbachia* parasites has been reported in various species comprising flies and mosquitoes (Diptera) [135, 47, 136, 137], parasitic wasps and other parasitoids (Hymenoptera) [48, 138], ants (Hymenoptera) [23, 25, 139], beetles and weevils (Coleoptera) [22, 140, 141], planthoppers and whiteflies (Hemiptera) [117], spiders (Araneae), moths and butterflies (Lepidoptera) [142, 143, 144]. When more than one symbiont is present, chromatograms become difficult to read as multiple peaks appear for the same position, and attributing a base (base calling) becomes often hard. When both complementary strands are sequenced (as in our case), the consensus sequence produced has very poor quality. This type of chromatogram was very frequent in our raw sequence dataset. For the above reasons, we believe the presence of co-infections in our host dataset might be the major factor contributing for only 60% of the strains being molecularly characterised. We shall come back to this later.

Distribution of the supergroups of Wolbachia in arthropods A total of 178 distinct FbpA sequences (hereafter referred to as FbpA haplotypes) were obtained from these 616 infections. The better sampled Orders accounted for 82% of the diversity of the haplotypes found (see Figure 3.11 and Table 3.9). In itself, this result was to be expected as the more individuals and the more diverse groups of species are collected, the more likely it is to find different *Wolbachia* lineages, not only because the prevalence may vary in the different host populations, but also because it becomes more likely to detect rare infections.

Curiously, within these Orders, although more Lepidoptera infections were characterised (147 host sequences) than Araneae ones (43 host sequences), Araneae presented a slightly bigger number of distinct FbpA haplotypes. The question arises though of how closely related (or diverse) are the infections found in Araneae. To address this and the more general question of how diverse were the infections found in insular arthropod communities, we started by doing a BLAST search of all FbpA sequences against the MLST FbpA allele database. Except for 3 Araneae, we were able to always have a single best hit, and only 9 of the 616 infections (1.5%) had less than 423 bp. Of these smaller sequences, only the pseudoscorpion infection (the best hit being allele-167) and one spider infection (best hit to both allele-26 and allele-86) should be looked at with caution as the overlap between paired sequences was short (less than 370 bp). Of

Figure 3.11: Distribution of the observed FbpA haplotype diversity in each Order. The total number of haplotypes found in each Order is indicated in brackets after the Order name. It should be noted that as some FbpA haplotypes were found in more than one Order, the sum of the haplotypes detected in all Orders is bigger than 178.



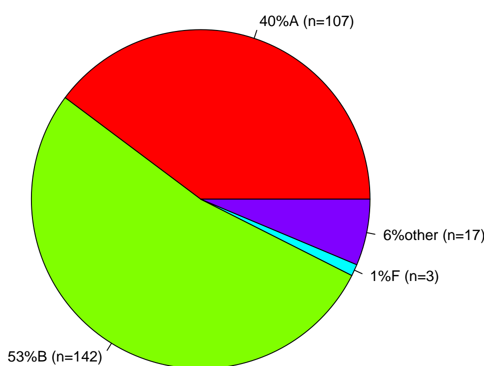
the remainder infections, more than 94% had less than 10 mismatches to their best hit (less than 2% uncorrected pairwise distance to the best hit) and the minimum sequence identity obtained was 86.52% (i.e., 57 mismatches in 423 bp) and corresponded to a highly divergent *Wolbachia* infection found in a parasitoid wasp (Hymenoptera: Trichogrammatidae). We then built an ML phylogeny (GRT+I+Γ model) with all our FbpA haplotypes and all 225 currently available FbpA alleles to help us determine to which supergroup the FbpA sequences belonged. We used the sole D supergroup FbpA allele to root the FbpA tree and, with the help of the supergroup annotation available in the MLST database, our phylogeny does not reject the monophyly of supergroups A, B, D, F and H (see Figure 4.1 in Appendix). The FbpA alleles were thus identified as belonging to a known supergroup if they were found within one of these clusters. In Table 3.9, we present a summary of the *Wolbachia* infections found within each Order, with the supergroups assigned.

Table 3.9: Characterisation of the FbpA sequences obtained for the 616 molecularly characterised hosts. Legend: W haplo - number of distinct unique FbpA sequences found per Order; Supergr - *Wolbachia* supergroups to which our FbpA alleles seem to belong, based on an ML phylogeny built with all the FbpA MLST alleles and all the FbpA haplotypes, and rooted with the sole FbpA sequence for a D supergroup (the tree can be found in Appendix); nb clusters-3% - number of host clusters-3% harbouring at least one infected COI sequence; MLST best hit - FbpA MLST alleles found to be the best match to at least one FbpA sequence within our dataset, with minimum identity $\geq$ 95%. Whenever the best hit did not agree with the closest allele in the phylogeny or the minimum identity was lower than < 95%, we used a "u". Whenever one or more alleles are found to be the best match, they are separated by a "-". Short sequences with good hits but based on a small alignment overlap are indicated with a *.

Order	W haplo	Supergr	nb clusters-3%	MLST best hit
Dipt	47	A	38	1,3,8,15,17,23,36,44,59,61,74,96,97,98
		B	24	4,6,7,9,10,14,27,132
		other	7	124,148,166
Hemi	44	A	6	8,17,39,74,96,114
		B	44	4,6,7,9,27,80,89,91,132,142,162,187,203
		other	2	166, u
Lepi	21	A	18	3,8,23,130
		B	31	4,6,9,10,27,162,204
		other	1	148
Hyme	30	A	27	8,17,26,36,61,112,118,120,134
		B	12	6,9,27,162
		F	1	125
		other	6	u
Aran	29	A	15	19,20,26,61,97,101,179,26_86
		B	1	u
		other	1	u
Psoco	12	B	17	6,9,136,162,204
Coleo	7	A	6	49
		B	1	4,6,27
		other	1	u
Orth	7	A	1	102
		B	3	88,142
Thys	2	B	2	9,u
		other	1	148
Isop	1	A	1	102
Odon	2	B	1	9
		F	1	125
Blat	1	F	1	125
Coll	2	other	2	227, 124
Neur	1	B	1	10
Pseu	1	other	1	167*

Characterisation and frequency of the different* Wolbachia *supergroups detected Based on the BLAST search and the phylogenetic relationship of our alleles and the MLST FbpA alleles, we detected infections grouping with three known supergroups of *Wolbachia*, namely the supergroups A, B, and F, as well as other infections clustering with more divergent FbpA alleles. Although in the latter case, no specific supergroup could be determined, we can consider that two groups of divergent clusters were identified (see the tree in Appendix): (1) a group of FbpA sequences (including those closely related to FbpA alleles 166 and 167) which, together with the F supergroup, form a sister group to A, B and H clusters; (2) a cluster of sequences (closely related with FbpA alleles 124, 148, 223-225 and 227) which form a sister group to B and (A,H) infections. As can be seen in Figure 3.12, the A and B supergroups represent

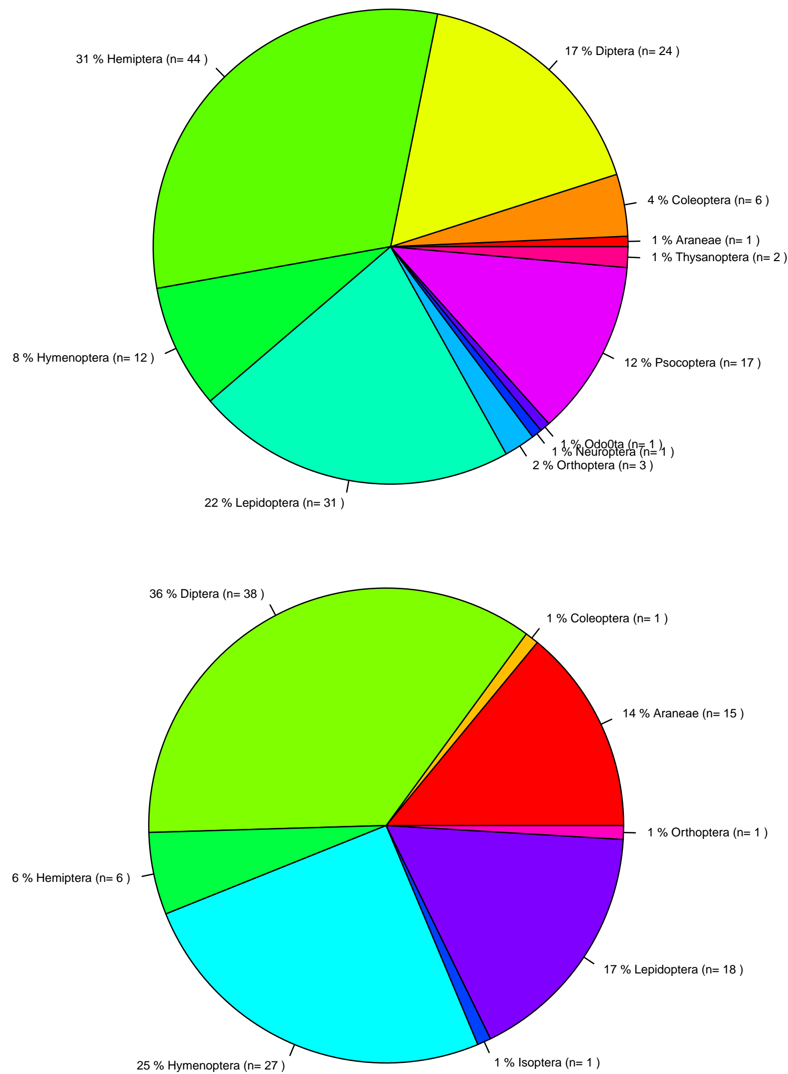
Figure 3.12: Infection frequency of *Wolbachia* supergroups present in our arthropod communities. The number of host cluster-3% found to be infected with a given supergroup is indicated between brackets following the group identification.



89% of the infections found in our arthropod clusters-3%, with the supergroup B being more frequent (53%) than the A (40%). Again, our results are in agreement with what is known from other geographical areas, i.e., the A and B supergroups are common and widely distributed in insects (where they have been mostly studied) but also in other groups such as Araneae.

Distribution of the supergroups among the Orders In Figure 3.13, we can see how infections from these 2 supergroups are distributed among host Orders. The B infections were found in 11 of the 15 Orders and are more frequent in Hemiptera and Lepidoptera (Figure 3.13 top). In contrast, the A infections were detected in 8 Orders and are more frequent in Araneae, Diptera and Hymenoptera (Figure 3.13 bottom).

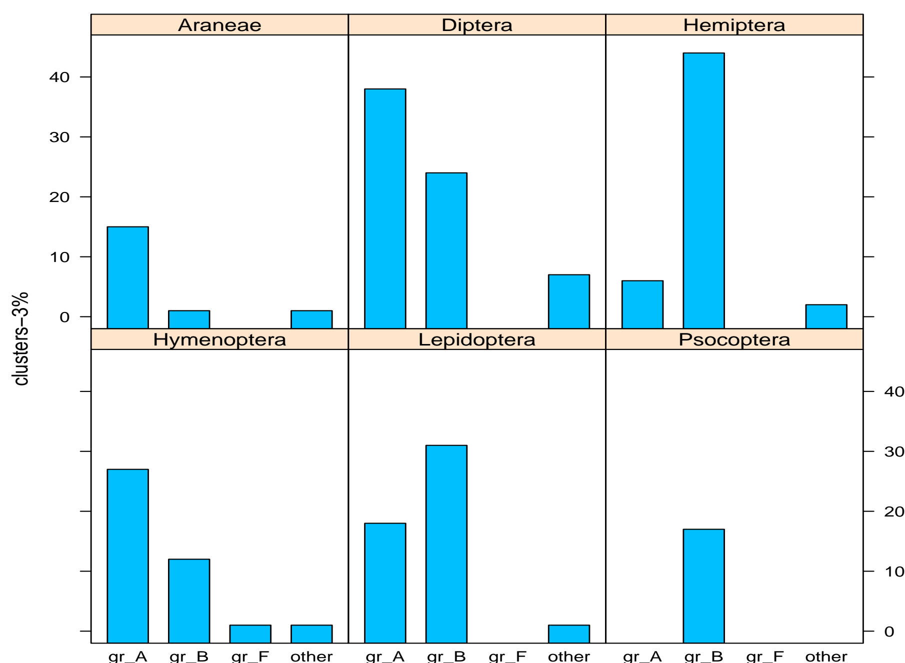
Figure 3.13: Distribution of B (top) and A (bottom) *Wolbachia* infections in the arthropod Orders found to be infected. In both cases the number of host clusters-3% found to be infected is indicated within brackets following the Order's name.



This difference in the number of Orders infected by the supergroups A and B is mainly due to lowly sampled Orders where only 1 or 2 infections were detected. The sole exception is the Order Psocoptera for which we found all of the 17 infected host clusters-3% whose *Wolbachia* we were able to characterise, to harbour the B strains exclusively. This result, taken together with the infection frequency estimated for this Order (82%), seems to suggest that the B infections may be highly successful in Psocoptera. As concerns the F infections, only one infected individual was found in the Orders Odonata (genus *Tholymis*) and Blattaria. In Hymenoptera, 2 ants from the genus *Tetramorium* harboured an F infection. Although no F infections were detected in Isoptera (termites) where they were first described, this is most likely due to very few individuals (4 in total) having been tested.

In previous surveys [47, 9, 145], it was reported that hymenopterans seem to be more frequently infected with A infections, while the B infections are more frequent in lepidopterans. As can be seen in Figure 3.14, for Orders with more than 10 infected COI clusters-3%, this same trend is observed in our data. While in Lepidoptera, approxi-

Figure 3.14: Distribution of the different *Wolbachia* supergroups within Orders. The number of infected host clusters-3% COI clusters was used as *proxy* for the number of infected species. Are indicated here Orders with 10 or more infected host clusters.



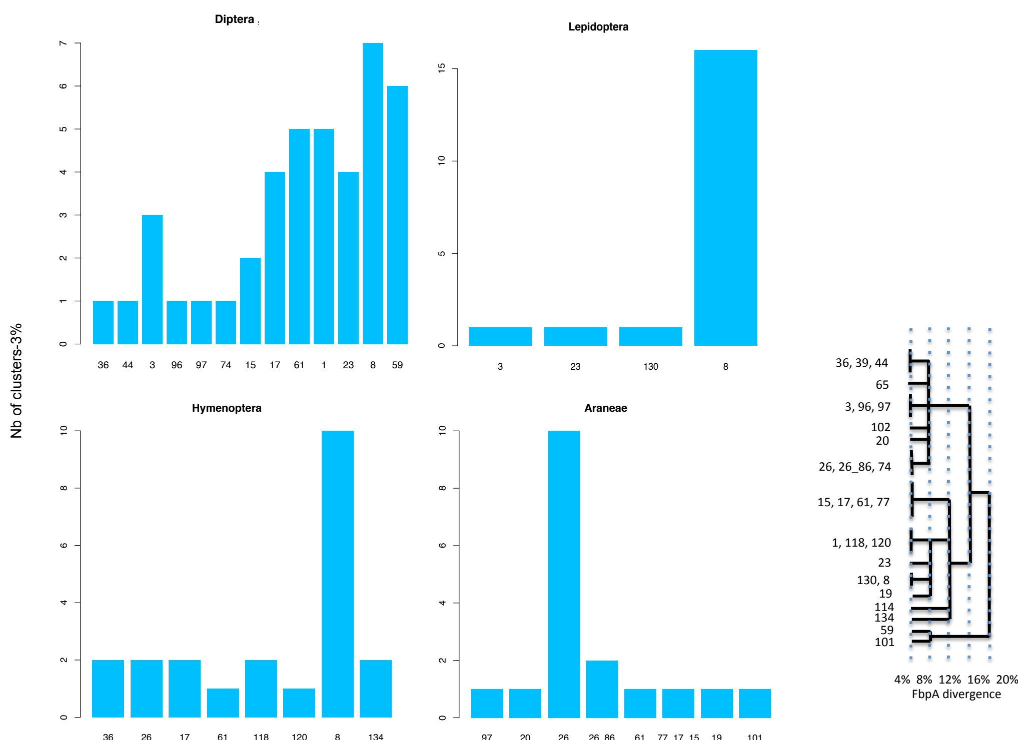
mately two thirds of the infected host clusters harbour B infections, in Hymenoptera a similar proportion of hosts harbours A infections. In this latter Order, there were

also 2 host clusters infected with F strains, both of ants (Formicidae). In Araneae, A infections were very frequent with 88% of the infected host clusters (15 out of 17), spanning 8 genera from 3 families (Tetragnathidae, Theridiidae and Uloboridae), harbouring infections from this group. Inversely, 85% (44 out of 52) of the Hemiptera host clusters harbour B infections. These results seem to suggest that different *Wolbachia* supergroups may have spread widely within these Orders.

Distribution of the different* Wolbachia *strains Until now, we have been discussing the diversity of *Wolbachia* infections by looking at the frequency of different supergroups. We thus looked at the presence of groups of distantly related symbionts. However, as can be seen in Table 3.9 (with the different FbpA alleles obtained as best hit in a BLAST search) and the phylogeny with all FbpA alleles in Appendix, within each of these groups we also find different infections. As mentioned previously, more than 94% of our sequences have less than 2% uncorrected-pairwise distance to an FbpA allele in the MSLT database. Our strains are therefore very closely related to the ones in the MLST database (at least, at the level of the FbpA marker). We then opted to represent the diversity of infections found within a *Wolbachia* supergroup by counting the number of host clusters which harboured strains mapping to the same FbpA allele. The diversity of the A and B infections is presented in Figures 3.15 and 3.16, respectively, for Orders with more than 2 distinct strains in the same supergroup. In both figures, a simplified cladogram, based on the patristic distance between groups of FbpA alleles (from the MLST database) at intervals of 4% divergence, is presented to illustrate the phylogenetic relationship between the strains found in the Orders presented.

(a) Diversity of the infections from supergroup A As concerns the A infections, all Orders except Diptera seem to have a group of very closely related infections that are widespread within the Order. In Araneae, 6 of the 8 genera infected with symbionts of the A supergroup are more specifically infected by a group of *Wolbachia* very closely related to strains harbouring the FbpA allele-26. They have not yet been described in Araneae but are known to infect Coleoptera (beetles) and Diptera (flies). In both Hymenoptera and Lepidoptera, we found that infections very closely related to strains harbouring the FbpA allele-8 are pervasive. Such infections are known to affect lepidopetrans such as the Mediterranean flour moth *Ephestis kuehniella* and a CI inducing *Wolbachia* – *wBol2* – which has already been described in populations of *Hypolimnas bolina* from our archipelago. In agreement with these previous results,

Figure 3.15: Infection frequency of A infections found within arthropod Orders. Groups of very closely related FbpA haplotypes (less than 2% uncorrected-pairwise distance) found in our study are represented by the FbpA allele found to be the best hit in a BLAST search against the MLST database of FbpA alleles. Are indicated here Orders for which 2 or more FbpA alleles were identified in the BLAST search. Each bar represents the number of host clusters-3% found to be infected by a *Wolbachia* group. FbpA alleles associated with well known A infections which are present in our dataset are: allele 1- *Drosophila melanogaster*'s strain *wMel*, allele 3 - *Aedus albopictus*' strain, allele 8 - *Hypolimnas bolina*'s strain *wBol2*, allele 17 - *Nasonia giraulti*'s strain, allele 23 - *D. simulans*' strain *wRi*.



specimens of this species present in our dataset were found to be infected with strains harbouring an FbpA haplotype identical to *wBol2*. In Diptera, we detected the highest diversity of infections from the supergroup A and some more distantly related sequences had similar frequencies, e.g. infections closely related to *wBol2*, *wMel* (allele-1), *wRi* (allele-23) and to strains harbouring the allele-59.

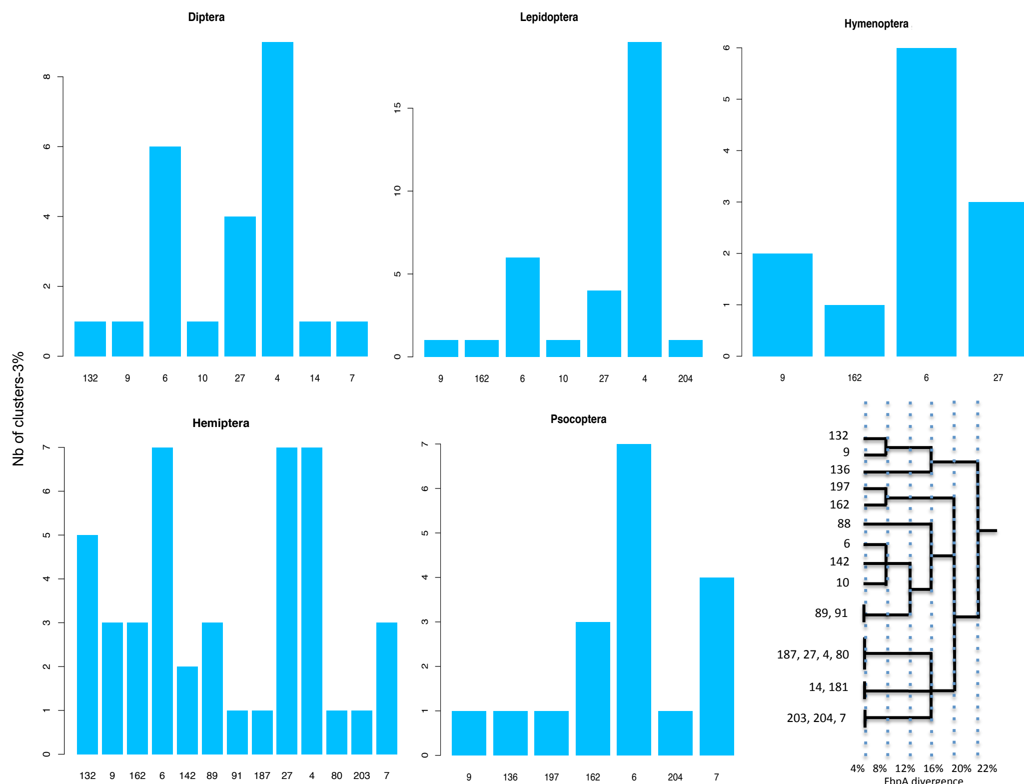
(b) Diversity of the infections from supergroup B We observed the inverse with the B strains. Only in Lepidoptera did we detect a group of very closely related *Wolbachia* (allele-4) that seemed to be widespread in the Order. All Orders are infected with distantly related B strains, and some groups of closely related *Wolbachias* seem

to be widespread in arthropods. Such was the case of the infections mapping to *wMa* (allele-6) and *wPip* (allele-4) which were found, respectively, in 5 and 4 Orders (in the latter case, absent in Psocoptera), as can be seen in Figure 3.16. Inversely to what was seen in Araneae with A infections, in the 2 Orders where B infections were pervasive, i.e., Hemiptera and Psocoptera, no group of B strains was more frequent than the others. Instead, we found the maximum number of diversity among the B groups in Hemiptera. In both Orders, we found infections from very distant B strains (more than 16% divergence). Notwithstanding, at least 4 distinct groups of very closely related B infections are shared between these 2 Orders. This supports our hypothesis that the high incidence observed in these Orders may be the result of a frequent horizontal transmission of *Wolbachia* symbionts due to a sharing of the same habitats and possible feeding resources as discussed in more detail in the following section.

(c) *Wolbachia* infection in less well sampled Orders In Neuroptera, we found one specimen harbouring a B infection. Previously, an F and an unassigned supergroup infections had already been detected in this host taxa [146]. In Pseudoscorpiones, the sole specimen found to be infected harbours a highly diverging *Wolbachia* closely related to a MK strain (*wCsc1*) already identified in this Order [126]. In Odonata and Blattaria, we detected very closely related F strains (allele-125), while in Collembola (springtails), we found 2 divergent infections of uncertain phylogenetic position. The thysanopteran infection is also detected within this cluster.

(d) *Wolbachia* infection in the species-richest Order Although well sampled and the species-richest insect Order ($\sim$ 350.000 species), Coleoptera presented the lowest percentage of infected hosts. This result was not surprising in light of previous surveys [47, 129, 9, 142] and studies which focused on more specific groups (for a recent review see [145]). Nevertheless, very diverse *Wolbachia* infections were found in this Order. These comprised the A, B and other more divergent infections and encompassed individuals found in a range of habitats and of feeding habits from phytophagous (Curculiniodae), predators of insects and other invertebrates (Coccinelidae, Staphylinidae), saprophytes (Nitiulidae, Scolytidae, Staphylinidae) parasitoids (Staphylinidae) and carbophagous/xylophagous insects (Oedemeridae).

Figure 3.16: Infection frequency of B infections found within arthropod Orders. Groups of very closely related FbpA haplotypes (less than 2% uncorrected-pairwise distance), found in our study, are represented by the FbpA allele found to be the best hit in a BLAST search against the MLST database of FbpA alleles. Represented are Orders for which 2 or more FbpA alleles were identified in the BLAST search. Each bar represents the number of host clusters-3% found to be infected by each *Wolbachia* group. FbpA alleles associated with well known B infections which are present in our dataset are: allele 4 - *Culex pipiens*'s strain *wPip*, allele 6 - *Drosophila simulans*' strain *wMa*, allele 9 - *Nasonia vitripennis* strain,



3.2.3 Discussion

To detect the presence of *Wolbachia* in our collection, we used the standard PCR technique with a new set of *Wolbachia*-specific primers [134]. Based on this screening, we detected 1678 infected specimens out of the 4860 arthropod individuals used in our study. Infected individuals were found in 24 out of 32 arthropod Orders. Within the uninfected groups were the Orders Isopoda, Opiliones, Phthiraptera (parasitic lices) and Scorpiones, for which infections have been well documented. We believe that lack of detection in these Orders is most likely due to the small number of hosts tested (4 specimens or less). The sole exception is Order Isopoda for which 23 individuals, encompassing all 4 islands, were tested and none was found to harbour the *Wolbachia*

parasite. Nevertheless, one cannot conclude from this lack of detection that *Wolbachia* infections are absent in terrestrial isopods from our island system. Three main reasons can explain this lack of success: (1) *Wolbachia* symbionts may be present in the isopod species tested but at low frequencies, which would require a stronger sampling effort to be able to detect their presence; (2) a failure of the 16S primers to detect the symbiont which could be due to improper binding of the primers to the target, or to the presence of the symbiont at low density in the host; (3) DNA quality.

With the intent to characterise the infections found, we performed a second PCR targeting the *fbpa locus* [44] which is more polymorphic than the 16S *rRNA locus*. Infection was detected by the 16S screening but not confirmed by the FbpA PCR in the Orders Scolopendromorpha (centipedes), Ixodida (ticks), Julida (millipedes) and Mesostigmata (mites). However, it should not be excluded that *Wolbachia* infections are also present in these Orders, particularly as they have already been described in isolates from other geographical regions [147, 148, 47].

We excluded from our analysis hosts for which no COI information was obtained as this was crucial for our analysis. This had as a "side-effect": the exclusion of infections, confirmed by both independent PCR essays found in individuals for which no COI sequence was available. In total, we excluded 254 *Wolbachia* infections for which we had been able to obtain an FbpA sequence. Within these infections were the strains detected in one dermapteran and one acarian (Order Oribatida). In the dermapteran specimen, we found an F strain (allele-125) that is very closely related to other infections found in Blattaria, Hymenoptera and Odonata. In contrast, the acarian specimen harboured a parasite with a highly divergent FbpA sequence (allele-225). Again, we found very closely related infections in 2 other Orders: Hemiptera and Hymenoptera.

Overall, our results are quite in agreement with what was already known about the dynamics of *Wolbachia* infections in arthropods from other geographical regions (South and North America, Europe and Asia). *Wolbachia* infections are ubiquitously distributed in the insular arthropod communities sampled. Based on our *proxy* for species number, we indeed estimate that as many as 39% of the arthropod species collected are infected. This value is higher than the 16-23% previously reported in extensive surveys [47, 9]. However, it is possible that this is in large part explained by the difference in sample size per species. Indeed, the mean sample size in our survey was 3.3, which is higher than in most earlier global surveys – based in one or two specimens per species. One should also keep in mind that the 3% COI clusters are a rough *proxy* for species. It is possible that by using these clusters we are grouping various species

in the same cluster, as various cases of very low COI divergence between close species are known (e.g: within the *Drosophila genus*). Moreover, as we considered as positive infections cases where only one individual was found to be infected, if some of these clusters correspond to an assemblage of closely related species and only one individual was infected, then, an artifactual over-estimation of incidence will be produced. Finally, single infections should be further confirmed by PCRs of more individuals of the same species (when and if we are able to obtain a better taxonomical classification for our specimens).

According to our estimates, hemipterans and psocopterans seem to be highly infected in these insular communities. Moreover, B infections seem to be very successful and at least 4 groups of distinct B infections are shared between these 2 Orders. As species from these Orders are mainly phytophagous and scavengers, this seems to point to high frequencies of horizontal transfer of this parasite occurring between these Orders. Moreover, closely related symbionts are found in other Orders too. Indeed, not only B infections but also A and other more divergent strains seem to be transferred between distantly related hosts. Could this be a characteristic of our insular system? This is possible but it could also be that we are detecting more transfers because we performed a more extensive sampling. In our study we have sampled various communities (from 4 islands) with approximately 1074 species being collected. This represents the largest collecting effort to date, covering a huge diversity of distantly related hosts.

As expected, given the approach used, a number of infections confirmed by both *Wolbachia*-specific PCRs could not be characterised at the sequence level. As mentioned previously, one likely explanation is the presence of multiple infections, which makes the Sanger sequencing inappropriate. Specifically in our case, 1001 hosts provided a strong FbpA PCR amplicon; out of those, 870 led to a sequence, resulting in a $\sim$ 87% success rate. If in fact this lack of success is mainly due to co-infections, then our results suggest that the proportion of multiple infections in our dataset ($\sim$ 13%) is not significantly different from what has been observed in samples from temperate and neotropical regions ($7 - 34\%$ estimated by Werren and colleagues [47]).

Because multiple infections cannot be detected, it is important to keep in mind that some cases of horizontal transfer will remain hidden in our sample due to our partial description of *Wolbachia* infections. This problem seems to particularly affect species belonging to the families Pholcidae (Araneae), Blattelidae (Blattaria), Sarcophagidae (Diptera), Nabidae (Hemiptera), Formicidae (Hymenoptera), Crambidae and Nymphalidae (the latter two from the Lepidoptera Order). In all these families, we saw more than 60% of the *Wolbachia* infections failing to be sequenced. These

families might therefore in the future be further characterised for co-infections. For this purpose, the use of parallel sequencing of many tagged PCR products in a single 454 sequencing reaction seems to be quite appropriate [149].

3.3 Horizontal transfers of *Wolbachia* infections

As mentioned before, horizontal transfer (from now on denoted by HT) of *Wolbachia* infections in arthropods is well documented, mostly via the observed incongruence between the phylogenies of the symbionts and of their hosts [55, 50, 1, 62, 150].

A pre-requisite for HT seems to be the need for intimate contact between the parasite's donor and recipient hosts as *Wolbachia* is an obligatory intracellular symbiont, although it was recently shown that some laboratory *Wolbachia* strains can survive outside the host cells and/or the death of the cells for short periods of time [151].

In the last decade, some studies have also focused on the ecological routes of Wolbachia HT. Transfers from host to parasitoids were demonstrated in parasitoid wasps [64, 51, 58, 59]. Host-parasitoid interactions have thus been proposed as a potential route for inter-taxon exchanges of *Wolbachia*, where the parasitoids can also function as a vector for the transmission between different hosts when hosts occasionally resist to the parasitoids. Sharing of the same food source has been suggested as another such possible route, in this case between communities of pumpkin-leaf feeding arthropods [53], mushroom feeding (mycetophagous) dipterans [54], or rice-field insects [60], or between uninfected and infected parasitoid wasps [59]. Other possible routes for HT are the exchange of haemolymph between infected and uninfected individuals which has been shown to occur in the isopod *Armadillium vulgare* [61], social parasitism as seen in the fire ants *Solenopsis spp* [25], and predator-prey interactions [52].

Sharing food resources, predators, parasites and/or parasitoids may play an extremely important role for the non-vertical transmission of *Wolbachia*.

In the present section, we are interested in detecting and characterising cases of *Wolbachia* horizontal transfer between distantly related taxa. To this purpose, we are going to use a conservative approach, that is, to consider as valid transfers cases where highly similar *Wolbachia* strains are found in different arthropod orders. Having identified cases of horizontally transferred symbionts, we are going to describe the hosts involved in these transfers and assess if some taxa are more involved in HT events than others and what may be the ecological routes behind the parasite's transfers.

3.3.1 Defining groups of closely related *Wolbachia* strains

When characterising the *Wolbachia* infections associated with a host COI lineage, we saw that very closely related FbpA sequences, i.e., sequences that group with the same MLST allele, were found in hosts belonging to more than one Order. This grouping procedure provided us with a good measure to characterise the infections found in

our insular system as concerns their diversity and frequency. However the maximum divergence among the elements in these groups was not the same all over. In order to address more accurately HT events and to eliminate some of the uncertainty introduced by the ambiguities in our FbpA sequences, we opted to continue looking at clades of *Wolbachia* infections shared among Orders. This required redefining such clades.

The new clades/clusters of FbpA sequences were defined as previously with the host's COI sequences, i.e., as monophyletic clusters with a maximum divergence of $x\%$. It should be noted that the divergence used to define these groups affects the number of horizontal transfers inferred. For smaller divergence values, only recent transfers can be detected. Inversely, if clusters are too large, i.e., if they contain very divergent FbpA sequences, then the observation that two arthropod orders contain one same cluster cannot be taken as evidence for horizontal transfer.

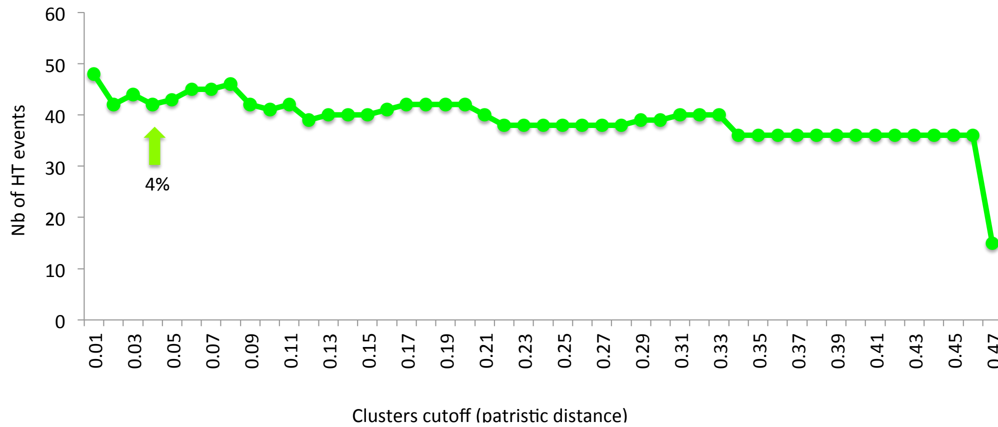
Since we were interested in addressing recent HT events, we chose a threshold corresponding to a lower divergence.

To determine which divergence value should be used as cutoff for the *Wolbachia* clusters, we started by varying this cutoff value between 1% and the maximum patristic distance observed in the phylogeny built with the FbpA sequences, by increments of 1%. For each value, we calculated both the number of clusters obtained and the minimum number of HT events estimated with the clusters produced. This number of HT was calculated, for each clade i shared among Orders, as equal to the number of Orders (n_i) minus one. The total number of HTs inferred was thus considered to be equal to $\sum_i(n_i - 1)$. It is important to remember that this calculation is a clear underestimation and is not meant as a quantification of the number of parasite jumps that have occurred in these hosts. Instead it is a simple measure that allows us to see how the choice of cutoff for our clusters can impact on the detection of the HTs.

As can be seen in Figure 3.18, the number of HT events varies slightly between 48 and 40, for cutoff values between 1 and 36%. Overall, the number of HTs tends to decrease very slowly between these cutoff values and then to stabilise at 36 HT events. Interestingly, clusters formed with a maximum of 4% patristic distance (henceforward referred to as *clusters-4%*) are the closest in composition to those obtained in the previous section when characterising the *Wolbachia* infections. Moreover, if we use as rate of divergence for the FbpA locus the synonymous substitution rate determined by Ochman [152] for protein coding genes in bacteria – $\sim 0.9\%$ per My – then we can estimate that our clusters-4% are composed of sequences that have started diverging ~ 2.2 My ago.

Thus, by using 4% of divergence, we work with groups of closely related parasites

Figure 3.17: Minimum number of HT events estimated when FbpA sequences are grouped in clusters with maximum patristic distance of x%, i.e., a cutoff value of x%. In the x axis, we have the clusters cutoff values, varying from 0.01 to 0.47 (maximum distance between FbpA sequences in our dataset) by increments of 1%. Indicated with an arrow is the minimum number of HTs between Orders (42) estimated for clusters with maximum patristic distance of 0.04. These are the clusters whose sequence composition is closer to the groups produced with sequences mapping to a single FbpA MLST allele. The number of HT events estimated was calculated as the value of $\sum_i (n_i - 1)$, where n_i is the number of Orders found to be infected by cluster i.



whose most recent common ancestor is $\sim$ 2.2 My old and, therefore, has clearly jumped between host taxa (arthropod Orders) that are estimated to have diverged more than 200 My ago. Additionally, these ancestral symbionts have started diverging approximately at the time of formation of the older islands (Raiatea and Huahine).

We therefore opted to use this maximum of 4% divergence for our *Wolbachia* clades. This led to 52 FbpA clusters-4% that will be considered in the analyses below.

3.3.2 Proportion of *Wolbachia* clades being horizontally transferred between different arthropod orders

Once we had defined a threshold for our parasite clades, we wanted to know how much of the *Wolbachia* diversity found in these islands was being shared between distant taxa. To this purpose, we chose to consider as valid transfers only the cases where at least two individuals from the same species had been found to be infected. This would allow us to eliminate false positive HT events due to possible contaminations. We shall come back to this in the discussion.

Based on the FbpA clusters-4%, we have 28.9% (15 out of 52 clusters) of the insular *Wolbachia* clades involved in HT events. This suggests that horizontal transfer between distant host taxa involves many different *Wolbachia* groups. Clearly, we are looking

at recent transfers only and we are not estimating a number of transfers. However, we can conclude from this conservative approach that at least 30%, i.e., a third of the *Wolbachia* clades as defined in the present study have transferred at least once within the last 2My. This seems to point to a high frequency of HT events even when the hosts are very distantly related. It also suggests that a considerable proportion of the *Wolbachia* symbionts may not require host specific physiological characteristics to invade a new host population but, on the opposite, may only need the (ecological) opportunity to jump from one host to another.

Although a number of previous works have shown that the phylogenies for the host and for *Wolbachia* are highly incongruent, none had yet focused on so many *Wolbachia* clades being exchanged between hosts that are so distant. Our dataset thus allowed us to better address this question as we were able to collect $\sim$ 1074 species distributed among 20 Orders.

3.3.3 Host specificity of horizontally transferred *Wolbachia* clades

Next, we were interested in seeing if *Wolbachia* clades shared by different Orders exhibited differential specificity, i.e., how their host range varied. In Table 3.10, we give a list of the 15 FbpA clusters-4% found to be infecting 2 or more Orders, as well as the *Wolbachia* supergroups to which these clusters of symbionts belong. We can see that both the A and B infections are frequently shared between distant hosts. In the case of the A infections, we have $\sim$ 33% of the clades (7 out of 21) being transferred while $\sim$ 35% (7 out of 20) of the B clades are jumping between Orders. Although most of the horizontally transferred *Wolbachia* belong to the A and B clades (as expected from the fact that they represent 89% of the parasite diversity found), we also observed the F clade being transferred between Orders.

Although we have the same number of A and B clades involved in the horizontal transmission of *Wolbachia*, as shown in Figure 3.18, when we look at the total number of Orders infected by each supergroup, the A group seems to have a narrower host range (7 Orders) than the B group (10 Orders). It should be noted that in Figure 3.18, we opted to represent, for each parasite clade, all possible transfers between the Orders it infects. For example, if 4 Orders are found to share a *Wolbachia* cluster, then each Order is connected by an edge to the other 3. This allows us to count the edges between any 2 Orders to know how many clusters (A and/or B) are shared in total between them. Thus, when considering the two most frequent supergroups, it seems that the B infections are being horizontally transferred among a broader range of hosts.

Table 3.10: FbpA clusters-4% horizontally transferred between Arthropod Orders. All clusters have a maximum 4 % of patristic distance. The distances were calculated from the ML tree obtained for the 616 FbpA sequences using the GTR+Γ+I model of sequence evolution with both the gamma and the percentage of invariant sites being estimated from the data. Groups for which we found evidence of horizontal transfer, but in some of the Orders only one individual has been found to be infected are indicated with an *.

Supergr	Host Order
A	Orthoptera, Isoptera
A	Diptera, Lepidoptera
A	Diptera, Hymenoptera
A	Araneae, Diptera, Hemiptera, Hymenoptera
A	Diptera, Hemiptera, Hymenoptera, Lepidoptera
A	Diptera, Hymenoptera, Lepidoptera*
A	Diptera, Hymenoptera
B	Hemiptera, Hymenoptera*, Lepidoptera*, Psocoptera
B	ortho,hemi
B	Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Psocoptera
B	Diptera*, Neuroptera, Lepidoptera
B	Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera
B	Lepidoptera, Psocoptera
B	Diptera, Hemiptera, Hymenoptera, Lepidoptera, Odonata, Psocoptera, Thysanoptera
F	Blattaria, Odonata, Hymenoptera

When we consider the F clade that is horizontally transferred, it is interesting to notice that less well sampled Orders, such as Blattaria and Odonata, are involved in these HT events.

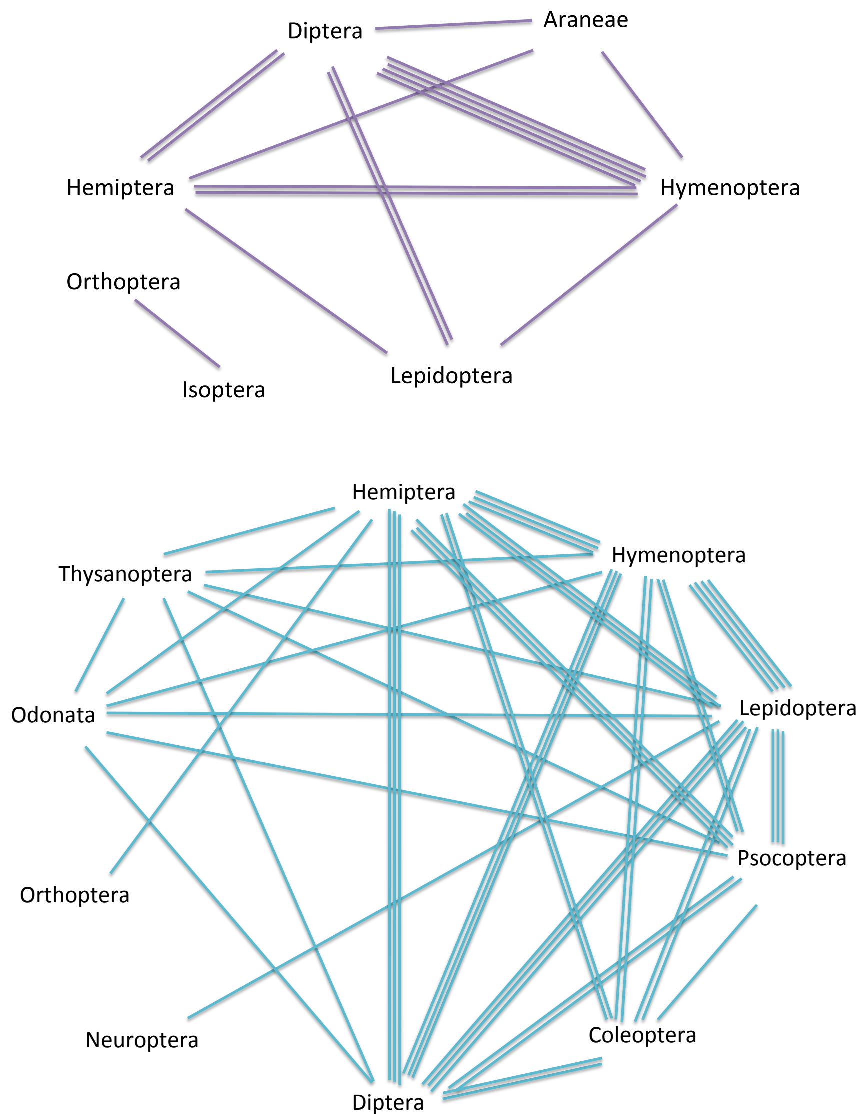
3.3.4 Investigating preferential transfers between Orders

We are now interested in seeing how HT events connect the host Orders. The first observation is that the Orders are not equally involved in horizontal transfer(s), i.e., some Orders share *Wolbachia* symbionts with many other host taxa, as can be seen in Table 3.11.

Table 3.11: Involvement of the different Orders in horizontal transfer events. The first line indicates the number of clusters-3% (sp) sampled in each Order and in the second line the number of Orders with which an Order shares *Wolbachia* clades (nbO).

Order	Hyme	Dipt	Hei	Lepi	Odon	Psoc	Thys	Coleo	Aran	Orth	Blat	Isop	Neur
Sp	150	317	121	233	7	22	5	100	57	22	8	1	5
nbO	9	8	8	8	7	7	6	5	3	2	1	1	1

Figure 3.18: Dynamics of *Wolbachia* jumps between Orders. In lilac we have the dynamics for the supergroup A clusters-4% and in blue the dynamics for the B clusters-4%. In both networks, we have drawn all the possible transfers that might have occurred between Orders. Consequently, the sum of the edges is bigger than 42 (the minimum # of HT events estimated based on the clusters cutoff value). Only clusters whose supergroups assignation was certain were used in these graphs.



Two non mutually exclusive factors can explain this observation: 1) some Orders may indeed have a narrower range of other Orders with which they exchange *Wolbachia* infections, as opposed to Orders carrying more generalist *Wolbachia* clades; 2) what we observe might reflect an artefact related to the diversity of each Order and sampling effort, i.e., better sampled Orders (which are also the most abundant and diverse) are more likely to be observed as sharing infections with other Orders. If we look at less well sampled Orders, we see that Thysanoptera and Odonata, though less represented in our dataset, were found to share a *Wolbachia* clade with 5 other Orders, while Orthoptera, Neuroptera and Isoptera share infections with 1-2 other host groups. In the case of the better sampled host taxa, we see that while the Orders Diptera, Hemiptera, Hymenoptera and Lepidoptera are widely connected (they share infections with 7-9 distinct Orders), spiders and coleopterans appear to have a narrower range of host taxa with which they exchange their symbionts (3 and 5 Orders, respectively). Interestingly, psocopterans – with only 22 host species and the highest incidence observed – are also one of the Orders with a wider range of host taxa with which they share infections (7 Orders). We tested if the sampling effort could be responsible for the observed number of connections – seen here as the number of Orders with which *Wolbachia* infections are shared – and we reject the null hypothesis of no correlation (Spearman's rank correlation rho=0.75, p-value=0.002) between both. Thus, the sampling effort – that is, the number of species per Order – could explain, at least in part, the observed variation between Orders.

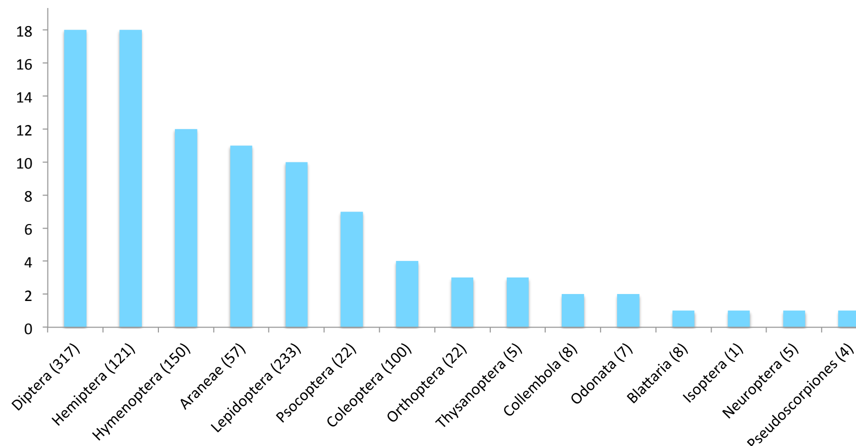
One can ask whether some host taxa might be exchanging preferentially their infections with some of the other Orders. This seems to be the case with the better sampled Orders in general (see Figure 3.18). However, this is not surprising in view of the previous results (sampling effort) and as more *Wolbachia* clades were found to infect these Orders (Figure 3.19).

We thus needed to define a measure that would allow us to see how closely related are the Orders based on the infections they share but, at the same time, to eliminate the effect of an uneven sampling effort.

Again, our analysis will only focus on Orders for which we have sampled 20 or more host COI clusters-3%, i.e., our next results concern only Araneae, Coleoptera, Diptera, Hemiptera, Hymenoptera, Lepidoptera, Orthoptera and Psocoptera.

If some Orders are preferentially exchanging their infections among themselves, then we would expect them to have similar infections. The more infections are thus shared between two Orders, the closer they should be when we measure proximity based on the diversity of the *Wolbachia* infections found in our data. Therefore, we measure

Figure 3.19: Number of *Wolbachia* clades found per Order. the sample size, i.e, number of clusters-3%, of an Order is indicated in brackets after the Order name.



the similarity between the infections found in each pair of Orders as the proportion of shared *Wolbachia* clades in the set representing the union of the clades found in both Orders. In practice, this measure is known as the Jaccard Index (JI).

$$JI_{(i,j)} = \frac{shared_{i,j}}{tot_{i,j}}, \tag{3.7}$$

where $shared_{i,j}$ is the number of *Wolbachia* clusters shared by Orders i and j, and $tot_{i,j}$ is the total number of it Wolbachia clusters found to infect both Orders.

To correct for the sampling effort bias, we performed 10.000 simulations of randomly sampling the same number of clusters-3% (species) per Order. Species were selected from the set of infected and uninfected ones, and the number of species was fixed at 22 per Order (number of species found in the less sampled Orders – Psocoptera and Orthoptera, see Table 3.8 for details). For each sampling procedure, we calculated the Jaccard Index (JI) for each pair of Orders, and the proximity between pairs of Orders was calculated as the average of the 10.000 JI. The results are shown in Table 3.12 and the most interesting observations are highlighted next.

- Araneae: Spider infections seem to be more closely related with those found in Diptera and Hymenoptera. This result is in agreement with what could already be seen in Figure 3.15.

- Coleoptera: These insects seem to frequently share infections with lepidopterans as they are the taxa with more closely related infections. We also see similarities with the dipteran and hemipteran *Wolbachia* clades but to a lesser extent.

Table 3.12: Proximity of arthropod Orders based on the sharing of *Wolbachia* infections. Represented is the mean similarity measure between Orders based on 10000 sampling procedures.

	Araneae	Coleoptera	Diptera	Hemiptera	Hymenoptera	Lepidoptera	Orthoptera	Psocoptera
Araneae	1.00	0.00	**0.07**	0.03	**0.06**	0.00	0.00	0.00
Coleoptera	0.00	1.00	**0.12**	**0.15**	0.09	**0.23**	0.00	0.05
Diptera	0.07	0.12	1.00	0.13	**0.22**	**0.28**	0.00	0.05
Hemiptera	0.03	0.15	0.13	1.00	0.15	**0.20**	0.04	**0.20**
Hymenoptera	0.06	0.09	**0.22**	0.15	1.00	**0.26**	0.00	0.10
Lepidoptera	0.00	**0.23**	**0.28**	**0.20**	**0.26**	1.00	0.00	0.08
Orthoptera	0.00	0.00	0.00	**0.04**	0.00	0.00	1.00	0.00
Psocoptera	0.00	0.05	0.05	0.20	**0.10**	**0.08**	0.00	1.00

- Diptera: In this group, the *Wolbachia* parasites found are mostly similar to those identified in Lepidoptera and Hymenoptera.
- Hemiptera: This Order appears to share more infections with Lepidoptera and Psocoptera, as well as, to a slightly lesser extent, with Coleoptera and Hymenoptera.
- Hymenoptera: In this case, the infections are preferentially found to be shared with Diptera and Lepidoptera. This result is in agreement with what is seen in Diptera.
- Lepidoptera: This is the Order which seems to have a wider range of taxa with which it preferentially shares infections. Among the taxa, we find Diptera, Hymenoptera, Hemiptera and Coleoptera.
- Orthoptera: Because this Order has so few infections and only one of the shared *Wolbachia* clades is represented in this analysis, we see that the Order shares infections preferentially with Hemiptera. As for spiders, the values are low.
- Psocoptera: As indicated in Figure 3.16, psocopterans and hemipterans share various symbionts. It seems that Psocoptera does indeed exchange infections preferentially with hemipterans. This tendency appears to be reciprocal as seen above for hemipterans.

Although our measure is only qualitative, it does allow us to have some indications on how infections may be preferentially transferred between some Orders. One clear example concerns the frequent transfers occurring between Diptera, Hymenoptera and Lepidoptera, even if in Lepidoptera the range of Orders involved is somewhat larger. Another example concerns the infections found in Psocoptera and Hemiptera. Already

when we characterised the *Wolbachia* infections, our results seemed to suggest that these two Orders could be frequently exchanging infections. We were now able to confirm this initial observation after the removal of the important confounding effect represented by an uneven sampling effort of the Orders.

As seen in the previous section, incidence – defined as the proportion of infected host COI clusters-3% – varied significantly between Orders (82% in Psocoptera to 15% in Coleoptera). If an Order has most of its species infected, then we will frequently sample infected species. Moreover, the same infected species can be re-sampled frequently. Thus, it is possible that Orders with high incidence will seem to share more infections because we re-sample more often infected species sharing the *Wolbachia* clades. To check for this confounding effect, we also tested the effect of incidence on our proximity measure by doing again the 10.000 samplings but this time sampling only clusters from the pool of infected species. The results did not change which seems to suggest that our measure is not greatly impacted by differences in incidence.

Overall, our data appears to show that arthropod Orders do not randomly share infections, but instead that some taxa exchange more frequently among them than with other Orders.

This led us then to address the question of the importance of the ecological connections in these transfer events as a possible explanation for this observed pattern of preferential HT. We are therefore now going to focus our attention on the characterisation of the routes of transfer that might be involved in the various HT events detected.

3.3.5 Exploring the ecological routes of Wolbachia horizontal transfer

We presented in Table 3.10 the *Wolbachia* clades found to be shared between distinct arthropod Orders. We also enumerated the Orders infected by each clade. We are now interested in detailing more these transfers, i.e., based on the taxonomic and ecological information (mostly feeding habits) available for each host, we are interested in trying to infer possible routes of horizontal transmission.

Although we do not have an accurate description of the habitats where the specimens were found, we have been able to classify most of the 616 hosts specimens, with known infections, up to the family level. We are then going to use the information available in Table 3.2 to provide a more detailed description of the 20 cases of horizontal transmission of *Wolbachia* among the Orders.

As shown in Figure 3.20 which presents all the clusters of *Wolbachia* from the supergroup A, we are not able to characterise all families involved in the various HT events. We are then going to describe only a few cases where we can clearly identify possible routes of transmission of *Wolbachia*.

Figure 3.20: Supergroup A infections horizontally transferred between families belonging to different arthropod Orders. *Wolbachia* clades for which evidence of horizontal transfer was found to be validated between some Orders but not all are indicated. In these cases, Orders for which only one individual was found to harbour the parasite are represented in red with an * and the possible transfer with interrupted lines.

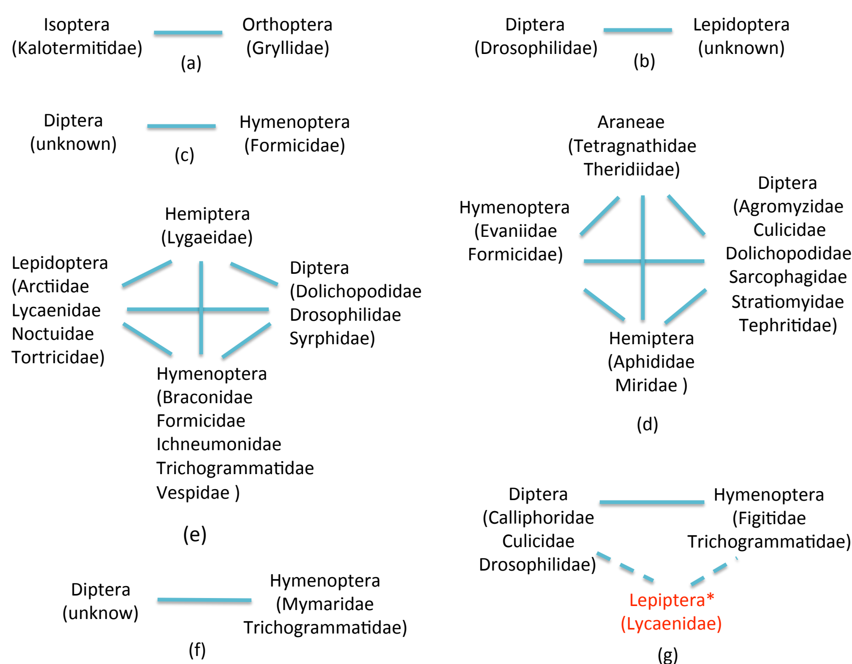


Figure 3.20 (a) Kalotermitidae is a family of termites (Order Isoptera), thus xylophagous, i.e., feeding on dry wood. On the other hand, Gryllidae is a family of species mostly phytophagous although some predation is also observed (but mostly between species within the family). Some cases of commensalism between crickets (Myrmecophilinae) and termites are however known ([153]), thus it is possible that a route of transmission in this case is the sharing of food sources.

Figure 3.20 (d) This is a clade which is not only present in various Orders (4) but which has also been found to infect at least two different families within each Order. It then becomes interesting to determine if one or more routes for transmission of *Wolbachia* are involved in this case. Spiders (Order Araneae) are known to be predators which feed on small or even medium sized insects.

They are generalist predators of other insects. Thus, their connection with the other 3 Orders is very likely one of predator/prey. In Hymenoptera, we find two families of wasps (Evaniidae and Pteromalidade) which parasite the eggs of many insects of different Orders (including other hymenopterans, hemipterans and dipterans) and a family of predators (Formicidade) which can feed on any dead insect. This suggests predator/prey and host/parasitoid (parasitism) interactions may very likely be the ecological connections involved in the symbiont's transfer in this Order. In the Order Hemiptera, the two families found to harbour this shared symbiont are both phytophagous although members of the Miridae family can also be predators. The highest diversity of infected families and also of feeding habits is found in Diptera. We thus identified families of phytophagous (Agromyzidae and Culicidade), hematophagous (Culicidae), saprophytic (Stratiomyidae), predator (Dolichopodidae, Sarcophagidae) and carbophagous (Tephritidae) dipterans. Consequently, predator/prey interactions seem to connect both Diptera and Hemiptera with the remaining Orders and, additionally, sharing of food resources (in the case of phytophagous hemipterans and dipterans) may also be a route for transfers between these two Orders. In summary, it seems that a complex network of ecological connections may be involved in these *Wolbachia* clade transfers. Predation appears to be one of the most likely routes for transmission, as we were able to find in each Order families which can prey on species from the other taxa. However, parasitoid/predator associations and sharing of food sources/pools may also play a role as a vehicle for horizontal transmission between certain Orders. Clearly, one or more routes may explain the horizontal transmission of the parasite between families from both different Orders and intra-Order. A more in-depth characterisation of the species found in each family and their habitats and ecology is needed to further elucidate which is really involved. In addition, a more accurate description of events of transfer would be needed to identify species (and not only families and Orders), between which transfers have taken place.

We are now going to focus our attention on the B and F parasite clades being transferred (see Figure 3.21). Again we detail only a few cases for which the ecological connections can be inferred.

Figure 3.21: Supergroup B and F infections horizontally transferred between families belonging to different arthropod Orders. *Wolbachia* clades for which evidence of horizontal transfer was found to be validated between some Orders but not all also indicated. In these cases, Orders for which only one individual was found to harbour the parasite are represented in red with an * and the possible transfer with interrupted lines.

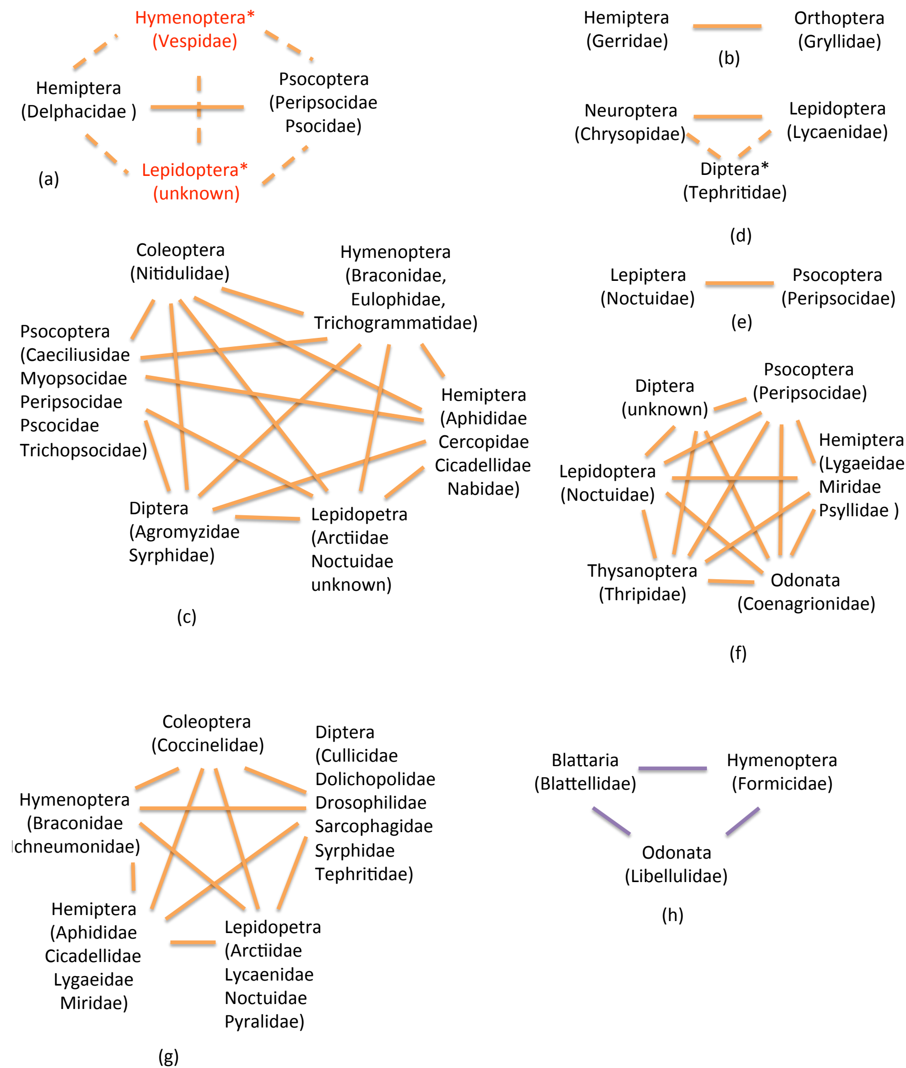


Figure 3.21 (c) Here we see a B clade that is shared among six Orders and involves various families within these taxa. In Psocoptera, we find 5 families infected with this clade of symbionts. All psocopterans families are thought to be saprophytes, i.e., individuals feeding on dead or decomposing material. We also find saprophytes in the infected colepoteran family (Nitidulidae). This seems to suggest that transfers may occur via sharing of common food pools, at least between these two Orders. In Lepidoptera and Hemiptera, we find mainly phytophagous families. All lepidopteran families are known to have phytophagous larvae and adults

which feed on nectar (nectivore), while in Hemiptera, the members of the Aphididae, Cercopidae and Cicadellidae families are also phytophagous. In addition, we also found in Diptera one family of phytophagous species (Agromyzidade) and another (Syrphidae) whose members are mainly nectarivores. Consequently, once more the sharing of common food sources seems to be a likely candidate for providing a mechanism of HT among these 3 Orders. However, the hymenopteran families found to be infected are all mainly composed of parasitoid wasps. This suggests that host/parasitoid interactions may also represent an important contribution. Moreover, it is possible that predator/prey interactions may be at play as both the hemiptera family Nabidae and the diptera family Syrphidade have predator species. Again, various routes for transfer may be involved in the HT of symbionts from this B clade: sharing of common food pools, parasitism and predator/prey interactions.

Figure 3.21 (h) This is the sole F clade being shared among the clades detected in our system. Predator/prey interactions represent the most likely ecological connection between the three Orders involved as both the ants (Hymenoptera: Formicidae) and the Libellulidae (Odonata) families are active predators of other insects. The connection with cockroaches (Blattelidae family) is more difficult to understand. These animals are well known scavengers feeding on both dead plant materials, decaying fruits and other dead insects. It could be that they feed on common dead insects, found in the ground, with ants.

From the few cases described above, it seems that some ecological relations may represent an important means for the horizontal transfer of *Wolbachia*. This includes parasitism (parasitoid/prey associations), predation (predator/prey associations) and sharing of common food sources. Moreover, these routes appear common to *Wolbachia* symbionts in general as shared infections from all the three supergroups A, B and F seem to be using the same mechanisms of transfer.

3.3.6 Discussion

In this section, we were interested in better characterising the dynamics of *Wolbachia*'s horizontal transfers and suggest possible ecological routes of transfer. To this purpose, we focused our attention on transfers occurring between very distantly related host taxa - arthropod Orders - as the ecology of these groups and their interactions will be one of the major (if not the major) factor determining the success of *Wolbachia* symbionts invading new hosts.

To address the question of horizontal transfer (HT) events we first needed to establish what types of transfers we wanted to detect, i.e, whether we wanted to detect ancient transfers or more recent ones. As one of the purposes motivating our choice of an island system was the access to recent evolutionary events, we were naturally interested in assessing recent horizontal jumps of our symbionts. We thus decided to study the dynamics of infections estimated to have occurred in the last ~ 2 My. This was well within the time frame of the islands formation.

Clades of FbpA sequences were then defined using the synonymous substitution rate determined by Ochman and colleagues [152] for bacterial protein coding genes. We obtained 52 clades with a maximum of 4% divergence between the FbpA sequences. We identified cases of HT by detecting *Wolbachia* clades infecting two or more Orders. In a first analysis, we determined 20 infections being shared between distant host taxa. However, a more detailed analysis revealed that in 5 of these cases, only one individual had been found to be infected in one of the Orders. We decided to adopt a conservative approach and consider as valid horizontal transfers only cases where two or more individuals from the same host clusters-3% were found to be infected. This choice was mainly motivated by the wish to eliminate false positives due to contaminations.

Clearly, it is possible that some of these cases of single infections may be true infections. One cannot exclude the fact that the number of hosts screened per species was in general low, and that prevalence most assuredly varied depending on the host species. Thus, one single infected individual can be due to a contamination but also to a low sampling of host populations with a medium to low prevalence of *Wolbachia* infection. This needs to be further verified by a repetition of both amplification and sequencing of the FbpA marker for the plates where these single specimens are located, and, if these are available in our remaining sorted specimens, by the screening of more individuals from the same species (or morpho-species).

As a consequence, we considered that only 15 infections have been horizontally transferred between Orders. We thus estimated that $\sim 30\%$ of these clades have been successfully jumping between very distant hosts in the last 2 My. This seems to suggest that horizontal transfers may be quite frequent even between physiologically and molecularly very distinct hosts.

The set of HT infections comprises seven A, seven B and one F infection. Once we had identified the infections being shared and the taxa sharing them, we were interested in assessing if there were differences in host specificity. Overall, B infections seem to have a broader range when compared to A infections. In both cases, we detected seven HT infections and while the A infections are distributed within a maximum range of

eight Orders, the B clusters span eleven Orders. Interestingly, the sole F clade detected in our dataset is also shared between Orders. Its host range is more reduced with only Blattaria, Odonata and Hymenoptera (ants) sharing this infection.

Next, we looked if there were some signals for preferential transfers occurring between certain Orders. Because we had different sampling efforts for the various Orders, with some being well sampled while others were not, we needed to eliminate this confounding factor. We performed simulations using only the Orders for which we had at least 22 host species (i.e., at least 22 COI clusters-3%)where we randomly picked 22 species. For each random sampling, we determined the Jaccard Index (Equation 3.7) which measures the similarity between 2 Orders as the ratio between the *Wolbachia* infections shared between the Orders and the total number of distinct infections found in both Orders. We then calculated the average of the 10000 measures obtained for each pair of host taxa. Our results seem to indicate that 2 groups of Orders preferentially exchange infections: (1) Psocoptera and Hemiptera (B clades) on one hand, (2) Lepidoptera, Dipetra and Hymenoptera on the other. Clearly, our measure is only qualitative and does not directly measure the number of transfers occurring between the Orders. It looks instead at the similarity of the infections present in two Orders: the greater the number of shared infections these Orders have, the more similar their overall infections will be. This is an indirect way of measuring preferential transfers as we are assuming that infections are transmitted one at a time, and then, the more infections are shared, the more transfers will have occurred. Nonetheless, this measure gives us a first indication that horizontal transfers between Orders may not be random and some Orders may actually exchange their infections more frequently with others.

Finally, we looked at what could be the mechanisms or routes of transfer for these symbionts. Based on our taxonomic knowledge of the families infected and their feeding habits, we could speculate on the possible ecological connections between the various Orders.

Shared infections from all 3 supergroups - A, B and F - seem to point to some common interactions: (1) predator/prey; (2) sharing of common food pools and (3) parasitism (host/parasitoid).

This is not unexpected as any mechanism for horizontal transmission is assumed to need the close contact between the donor and receiver hosts. As some recent studies have shown, these routes are in particular used in some systems [51, 58, 59, 53, 54]

Clearly, in our study, we can only speculate that these ecological connections might be playing a role in the HT events observed. Families comprise large sets of species that can have a wide range of habitats, feeding habits and behaviours. Not only we do not

have access to a species characterisation of the individuals involved in these transfers as, even if we had that taxonomic information, we would need an accurate description of their habitats and ecology (feeding habits, behaviour, etc). Notwithstanding, our study seems to suggest that the same mechanisms observed in specific mainland biological systems are involved in the transfer of *Wolbachia* in insular arthropod communities. In the future, with better taxonomic and ecological characterisations, we may perhaps be able to better elucidate whether these mechanisms are really behind the parasite's transfers.

Chapter 4

Conclusions and Perspectives

Contents

4.1 Diversity and biogeography of the Host mitochondria . . 103

4.2 Incidence and diversity of *Wolbachia* infections 104

4.3 Horizontal transfers of *Wolbachia* infections 105

In this PhD, we performed a study at what we called a "geographical community" level, i.e., we followed a top-down approach using a mixture of arthropod communities collected in four geographical locations to assess the presence and dynamics of horizontal transfers of the sexual parasite *Wolbachia* in these communities. The term community here should not be understood in the strict, classical ecological sense – defined as a group of populations from different species which co-exist (and interact) in the same space. Instead, we worked with an assemblage of communities collected in various locations of four islands of the Society archipelago - Raiatea, Huahine, Moorea and Tahiti.

4.1 Diversity and biogeography of the Host mitochondria

The first important observation concerning the results on biogeography is that there is no major effect of the experimental set-up on the geographical distribution of genotypes (see page 55). This allowed to discuss this distribution.

As we studied communities with a potentially high degree of isolation (island systems), our first main result is related to the distribution of the mitochondrial diversity in these islands. By first clustering individuals so that the common ancestor of a group

is not older than the creation date of the oldest island, we showed that in some cases, individuals of a same group within an island are genetically closer to each other than to individuals from the other islands. Inversely, in most cases we did not observe this. This is not always due to the fact that the number of individuals in a cluster does not allow a powerful test. Even so, re-examination of the initial data collected might allow us to perform a targeted sequencing of individuals for groups with a small sample size. This approach however requires a preliminary effort of classification of individuals based on their morphology.

Even if current data do not enable us to conclude, we can recognise the existence of factors which could homogenise the genotype patterns among the four islands, namely the over-abundance in our dataset of invasive species and the strong impact of human movements.

4.2 Incidence and diversity of *Wolbachia* infections

We start by observing that this is the first study that examined the incidence – defined as the proportion of infected species – in such a large spectrum of Orders.

The first result obtained is that incidence depends strongly on the Order (it ranges from 82% in Psocoptera to 15% in Coleoptera). This fact was already known. What is new is that the incidence observed in our dataset is higher in the case of some Orders, notably Psocoptera and Hemiptera. This result of course depends on the threshold chosen to define the species.

Thanks to the characterisation of 616 infection status, we could confirm that in insular systems, we also observe a high diversity of *Wolbachia* infections. Indeed, we detected different infections that cover the A, B and F supergroups as well as other *Wolbachia* symbionts with more divergent FbpA sequences. Our findings corroborate the fact that the A and B infections seem to be very frequent, with almost 90% of the observed FbpA diversity belonging to one of these groups, and that they are ubiquitous in the various Orders sampled. An interesting finding was the presence of exclusively B infections in Psocoptera. This further hints that something particular may be happening with this Order in these communities, and it would be interesting to investigate it more in depth. Multiple infections were also detected, despite the use of general primers. We think that most of the unsuccessfully assembled sequences were due to the presence of more than one *Wolbachia* symbiont in the host. We thus estimate that approximately 13% of the hosts in our dataset may be harbouring multiple infections. Again, our results are in agreement with previous observations made in other commu-

nities from temperate and neotropical regions. In our study, infection polymorphism seems to particularly affect some families of Araneae, Blattaria, Diptera, Hemiptera, Hymenoptera and Lepidoptera.

4.3 Horizontal transfers of *Wolbachia* infections

We showed that in the last 3 My, there have been many transfers of *Wolbachia* between different arthropod Orders. Indeed, one third of the *Wolbachia* clusters have been involved in horizontal transfer (HT). All supergroups clearly identified by us, namely, the supergroups A, B, and F, appear to be concerned by these long distance transfers. This might be the case also for another group of *Wolbachia* with FbpA sequences strongly divergent from the supergroups A, B, and F. However, this needs to be further confirmed to exclude a possible contamination effect. Indeed the case observed is based on a single infected individual, and is, as a consequence, only very preliminary. A way to do this confirmation might be to both sequence more individuals (if they are available in our specimens stock) and to perform further random PCRs of the individuals we have already to confirm the signals with a better control for possible contaminations.

Although no evaluation of the number of transfers was performed, our results are compatible with the importance of ecological connections among species sharing a same environment. Such connections had already been mentioned in the literature, for instance via a host parasitoid relation, sharing of common food sources or predator/prey interactions.

Although we worked with data collected in an extensive sampling of the four islands, we lacked a better description of the habitats in which the sampled organisms lived as well as a better classification of the species collected. This meant that we could not reach a very fine level of characterisation of the routes underlying each inferred transfer. In particular, we might have missed completely information on unique or very rare ecological niches and endemic species.

In order to be able in the future to go from the qualitative approach we were able to follow in this PhD to a more quantitative one, we would need, first to work with more than one marker in particular for *Wolbachia*, and second, to be able to develop new algorithms capable to handle big trees such as the ones we had.

Bibliography

[1] ONeill, S. L. and Hoffman, A. A. and Werren, J. H., editor. *Influencial Passengers Inherited Microorganisms and Arthropo Reproduction.* Oxfor University Press, Oxford, UK, 1997.

[2] Dale, C. and Moran, N.A. Molecuar Interactions between Bacterial Symbionts and their Hosts. *Cell*, 126:pp 453 – 465, 2006.

[3] Ferrari, J. and Vavre, F. Bacterial symbionts in insects or the story of communities affecting communities. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 366:pp 1389– 1400, 2011.

[4] Bandi, C. and Dunn, A. M. and Hurst, G. D. D. and Rigaud, T. Inherited microorganisms, sex-specific virulence and reproductive parasitism. *Trends in Parasitology*, 17:pp 88 – 94, 2001.

[5] Yen, J.H. and Barr. A.R. The etiological agent of cytoplasmic incompatibility in Culex pipiens. *Journal of Invertebrate Pathology*, 22(2):pp 242 – 250, 1973.

[6] Rousset, F. and Bouchon, D. and Pintureau, B. and Juchault, P. and Solignac, M. Wolbachia Endosymbionts Responsible for Various Alterations of Sexuality in Arthropods. *Proc. Royal Soc.of Lon.B Biol. Sci.*, 250(1328):pp 91–98, 1992.

[7] Stouthamer, R. and Breeuwer, J.A. and Luck, R.F. and Werren, J.H. Molecular identification of microorganisms associated with parthenogenesis. *Nature*, 361:pp 66 – 68, 1993.

[8] Gherna, R.L. and Werren, J.H. and Weisburg, W. and Cote, R. and Woese, C.R. and Mandelco, L. and Brenner, D.J. Arsenophonus nasoniae gen. nov., sp. nov., the causative agent of the Son-killer trait in the parasitic wasp Nasonia vitripennis. *Int J Syst Bacteriol*, 41:pp 563 – 565, 1991.

[9] Duron, O and Bouchon, D and Boutin, S. and Bellamy, L. and Zhou, L. and Engelstadter, J. and Hurst, G. D. The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone. *BMC Biol.*, pages pp. 6 – 28, 2008.

[10] Hurst, G. D. D. and Jiggins, F. M. and von der Schulenburg, J. H. G. and Bertrand, D. and West, S. A. and Goriacheva, I. I. and Zakharov, I. A. and Werren, J. H. and Stouthamer, R. and Majerus, M. E. N. Male killing Wolbachia in two species of insect. *Proc. Royal Soc.of Lon.B Biol. Sci.*, 266(1420):pp 735–740, 1999.

[11] Williamson, D. L. and Sakaguchi, B. and Hackett, K. J. and Whitcomb, R. F. and Tully, J. G. and Carle, P. and Bove, J. M. and Adams, J. R. and Konai, M. and Henegar, R. B. Spiroplasma poulsonii sp. nov., a new species associated with male-lethality in Drosophila willistoni, a neotropical species of fruit fly. *International Journal of Systematic Bacteriology*, 49(2):pp 611–618, 1999.

[12] Tinsley, M. C. and Majerus, M. E. N. A new male-killing parasitism: Spiroplasma bacteria infect the ladybird beetle Anisosticta novemdecimpunctata (Coleoptera: Coccinellidae). *Parasitology*, 132(06):pp 757–765, 2006.

[13] Hunter, M. S. and Perlman, S. J. and Kelly, S. E. A bacterial symbiont in the Bacteroidetes induces cytoplasmic incompatibility in the parasitoid wasp Encarsia pergandiella. *Proc. Royal Soc.of Lon.B Biol. Sci.*, 270(1529):pp 2185 – 2190, 2003.

[14] Zchori-Fein, E. and Perlman, S. J. Distribution of the bacterial symbiont Cardinium in arthropods. *Mol. Ecol.*, 13(7):pp 2009 – 2016, 2004.

[15] Bern, M. and Goldberg, D. Automatic selection of representative proteins for bacterial phylogeny. *BMC Evolutionary Biology*, 5(1):34, 2005.

[16] Hilgenboecker, K.,and Hammerstein, P., and chlattmann, P. ,Tand lschow, A. and Werren, J. How many species are infected with Wolbachia? a statistical analysis of current data. *FEMS Microbiol. Lett.*, 281:215 – 220, 2008.

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

[18] Jiggins, F.M. and Hurst, G.D.D. and Majerus, M.E.N. High prevalence of male-killing Wolbachia in the butterfly host Acraeaencedana. *Journal of Evolutionary Biology*, 13:pp 495 – 501, 2000.

[19] T. Rigaud. *Inherited microoraganisms and sex determination of arthropod hosts. In: Influential Passengers: Inherited Microorgan isms and Arthropod Reproduction* . Oxfor University Press, Oxford, UK, 1997.

[20] Kitrayapong, P. and Baimai, V . and O'Neill, S.L. Field prevalence of Wolbachia in the mosquito vector Aedes albopictus. *The American journal of tropical medicine and hygiene.*, 66:pp 108 – 111, 2002.

[21] Cheng, Q. and Ruel, T.D. and Zhou, W. and Moloo, S.K. and Majiwa, P. and O'Neill S.L. and Aksoy, S. Tissue distribution and prevalence of Wolbachia infections in tsetse flies, Glossina spp. *Medical and veterinary entomology.*, 14:pp 44 – 50, 2000.

[22] Kondo, N. and Ijichi, N. and Shimada, M. and Fukatsu, T. Prevailing triple infection with Wolbachia in Callosobruchus chinensis (Coleoptera: Bruchidae). *Mol. Eco.*, 11(2):pp. 167 – 180, 2002.

[23] Reuter, M. and Keller, L. High Levels of Multiple *Wolbachia* Infection and Recombination in the Ant Formica exsecta. *Mol. Biol. Evol.*, 20:pp. 1428 – 1433, 2003.

[24] R. Stouthamer. *Wolbachia-induced parthenogenesis. In: Influential Passengers: Inherited Microorganisms and Arthropod Reproduction* . Oxfor University Press, Oxford, UK, 1997.

[25] Dedeine, F. and Ahrens, M. and Calcaterra, L. and Shoemaker, D. D. Social parasitism in fire ants (Solenopsis spp.): a potential mechanism for interspecies transfer of Wolbachia. *Mol. Eco.*, 14(5):1543–1548, 2005.

[26] Bandi, C. and Anderson, T.J.C. and Genchi, C. and Blaxter, M.L. . Phylogeny of W. pipientis in filarial nematodes. *Proc R Soc Lond B*, 265:pp 2407 – 2413., 1998.

[27] Charlat, S., Hurst, G. D. D. & Mercot, H. Evolutionary consequences of Wolbachia infections. *Trends Gen.*, 17:pp 219 – 223, 2003.

[28] Bordenstein, S.R. and O'Hara, F. P. and Werren, J. H. Wolbachia-induced incompatibility precedes other hybrid incompatibilities in Nasonia. *Nature*, 409:pp 707 – 710, 2001.

[29] Hotopp, J. and Clark ,M, and Oliveira, D. and Foster, J. and Fischer, P. and Torres, M. and Giebel, J. and Kumar, N. and Ishmael, N. and Wang, S. and Ingram, J. and Nene, R. and Shepard, J. and Tomkins, J. and Stephen, R.S. and Spiro, D. and Ghedin, E. and Slatko, B. and Tettelin, H. and Werren, J.H. Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes. *Science*, 317:pp 1753, 2007.

[30] Dobson, S.L and Bourtzis, K. and Braig, H.R and Jones, B. F. and Zhou, W. and Rousset, F. and O'Neill, S. L. Wolbachia infections are distributed throughout insect somatic and germ line tissues. *Insect Biochem. Mol. Biol.*, 29(2):pp 153 – 160, 1999.

[31] Stouthamer, R. and Breeuwer, J.A.J and Hurst, G.D.D. *Wolbachia pipientis*: Microbial Manipulator of Arthropod Reproduction. *Annu. Rev. Microbiol.*, 53:pp 71 –102, 1999.

[32] Ferri, E. and Bain, O. and Barbuto, M. and Martin, C. andD Lo, N. and Uni, S. and Landmann, F. and Baccei, S.G and Guerrero, R. and de Souza Lima, S. and Bandi, C. and Wanji, S. and Diagne, M. and Casiraghi, M. New Insights into the Evolution of *Wolbachia* Infections in Filarial Nematodes Inferred from a Large Range of Screened Species. *PLoS ONE*, 6(6):e20843, 06 2011.

[33] Werren, J.H and Zhang, W. and Guo, L.R. Evolution and Phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc Biol Sci*, 261:pp 55 – 63, 1995.

[34] Casiraghi, M. and Bordenstein, S. R. and Baldo, L. and Lo, N. and Beninati, T. and Wernegreen, J. J. and Werren, J. H. and Bandi, C. Phylogeny of Wolbachia pipientis based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the Wolbachia tree. *Microbiology*, 151(12):pp 4015–4022, 2005.

[35] Lo, N. and Paraskevopoulos, C. and Bourtzis, K. and ONeill, S. L. and Werren, J. H. and Bordenstein, S. R. and Bandi, C. Taxonomic status of the intracellular bacterium *Wolbachia pipientis*. *Int. Jour. Sys. Evol. Microbiol.*, 57(3):pp. 654 – 657, 2007.

[36] Ros, Vera I. D. and Fleming, Vicki M. and Feil, Edward J. and Breeuwer, Johannes A. J. How Diverse Is the Genus Wolbachia? Multiple-Gene Sequencing

Reveals a Putatively New Wolbachia Supergroup Recovered from Spider Mites (Acari: Tetranychidae). *Appl. Environ. Microbiol.*, 75(4):1036–1043, 2009.

[37] Bordenstein, S. R. and Paraskevopoulos, C. and Dunning Hotopp, J. C. and Sapountzis, P. and Lo, N. and Bandi, C. and Tettelin, H. and Werren, J. H. and Bourtzis, K. Parasitism and Mutualism in Wolbachia: What the Phylogenomic Trees Can and Cannot Say. *Mol. Biol. Evol.*, 26(1):pp 231–241, 2009.

[38] Haegeman, A. and Vanholme, B. and Jacob, J. and Vandekerckhove , T.T.M. and Claeys, M. and Borgonie, G. and Gheysen, G. An endosymbiotic bacterium in a plant-parasitic nematode: Member of a new Wolbachia supergroup. *International Journal for Parasitology*, 39(9):pp 1045 – 1054, 2009.

[39] Foster, J. and Slatko, B. and Bandi, C. and Kumar, S. Recombination in Wolbachia Endosymbionts of Filarial Nematodes? *Applied and Environmental Microbiology*, 77:pp 1921 –1922, 2011.

[40] Jiggins, F. M. and Hurst, G. D. D. and Yang, Z. Host-symbiont conflicts: Positive selection on an outer membrane protein of parasitic but not mutualistic rickettsiaceae. *Mol. Biol. Evol.*, 19(8):1341–1349, 2002.

[41] Casiraghi, M., J. H. Werren, C. Bazzocchi, A. Biserni, and C. Bandi. dnaA gene sequences from Wolbachia pipientis support subdivision into supergroups and provide no evidence for recombination in the lineages infecting nematodes. *Parasitologia*, 45:pp 13 – 18, 2003.

[42] Bordenstein, S. R. and Paraskevopoulos, C. and Dunning Hotopp, J. C. and Sapountzis, P. and Lo, N. and Bandi, C. and Tettelin, H. and Werren, J. H. and Bourtzis, K. Parasitism and Mutualism in Wolbachia: What the Phylogenomic Trees Can and Cannot Say. *Mol. Biol. Evol.*, 26(1):pp 231–241, 2009.

[43] Casiraghi, M., T. J. Anderson, C. Bandi, C. Bazzocchi, and C. Genchi. A phylogenetic analysis of filarial nematodes: comparison with the phylogeny of Wolbachia endosymbionts. *Parasitology*, 122:pp 93 –103, 2001.

[44] Baldo, L. and Dunning Hotopp, J. C. and Jolley, K. A. and Bordenstein, S. R. and Biber, S. A. and Choudhury, R. R. and Hayashi, C.l and Maiden, M. C. J. and Tettelin, H. and Werren, J. H. Multilocus Sequence Typing System for the Endosymbiont Wolbachia pipientis. *Appl. Environ. Microbiol.*, 72(11):7098–7110, 2006.

[45] Baldo, L. and Bordenstein, S. and Wernegreen, J.J. and Werren, J.H. Widespread Recombination Throughout Wolbachia Genomes. *Mol. Biol. Evol.*, 23(2):pp 437 – 449, 2006.

[46] Jiggins, Francis M. The Rate of Recombination in Wolbachia Bacteria. *Mol. Biol. Evol.*, 19(9):pp 1640 – 1643, 2002.

[47] Werren, J.H and Windsor, D.M. *Wolbachia* infection frequencies in insects: evidence for a global equilibirum? *Proc. R. Soc. B*, 267:pp. 1277 – 1285, 2000.

[48] Vavre, F. and Dedeine, F. and Quillon, M. and Fouillet, P. and Fleury,, F. adn Bouletreau, M. Within-species diversity of Wolbachia-induced cytoplasmic incompatibility in haplodiploid insects. *Evolution*, 55:pp. 1710 – 1714, 2001.

[49] Charlat, S. and Engelstadter, J. and Dyson, E.A. and Hornett, E.A. and Duplouy, A. and Tortosa, P. and Davies, N. and Roderick, G.K. and Wedell, N. and Hurst, G.D.D. Competing Selfish Genetic Elements in the Butterfly Hypolimnas bolina. *Curr. Biol.*, 16(24):pp 2453 – 2458, 2006.

[50] Werren, John H. Biology of *Wolbachia*. *Annu. Rev. Entomol.*, 42(1):587–609, 1997.

[51] Vavre, F and Fleury, F and Lepetit, D and Fouillet, P and Bouletreau, M. Phylogenetic evidence for horizontal transmission of Wolbachia in host-parasitoid associations. *Mol. Biol. Evol.*, 16(12):pp. 1711 – 1723, 1999.

[52] Cordaux, R. and Michel-Salzat, A. and Bouchon, D. Wolbachia infection in crustaceans: novel hosts and potential routes for horizontal transmission. *J. Evol. Biol.*, 14(2):237–243, 2001.

[53] Sintupachee, S. and Milne, J. and Poonchaisri, S. and Baimai, V. and Kittayapong, P. Closely Related *Wolbachia*Strains within the Pumpkin Arthropod Community and the Potential for Horizontal Transmission via the Plant. *Micro. Eco.*, 51:pp. 294 – 301, 2006. 10.1007/s00248-006-9036-x.

[54] Stahlhult, J. K. and Desjardins, C. A. and Clark, MI. E. and Baldo, LA. and Russell Jacob, A. and Werren, J. H. and Jaenike, J. The mushroom habitat as an ecological arena for global exchange of Wolbachia. *Mol. Eco.*, 19(9):pp. 1940 – 1952, 2010.

[55] Werren, J.H and Windsor, D. and Guo, L. Distribution of *wolbachia* among Neotropical arthropods. *Proc. R. Soc. B*, 262:pp. 55 – 63, 1995.

[56] Baudry, E., J. Bartos, K. Emerson, T. Whitworth, and Werren, J. H. Wolbachia and genetic variability in the birdnest blowfly Protocalliphora sialia. *Mol. Ecol.*, 12:pp 1843 – 1854, 2003.

[57] Arthofer, W and Riegler, M and Schneider, D. and Krammer, M. and Miller, W.J. and Stauffer, C. Hidden Wolbachia diversity in field populations of the European cherry fruit fly, Rhagoletis cerasi (Diptera, Tephritidae). *Mol. Ecol.*, 18(18):3816–3830, 2009.

[58] Heath, B.D. and Butcher, R.D.J. and Whitfield, W.G.F. and Hubbard, S.F. Horizontal transfer of *Wolbachia* between phylogenetically distant insect species by a naturally occurring mechanism . *Current Biology*, 9:pp. 313–316, 1999.

[59] Huigens, M.E. and Luck, R.F. and Klaassen, R.H. and Timmermans, M.J. and Stones-Havas, Stouhamer, R. Infectious Parthenogenesis. *Nature*, 405:pp. 178–179, 2000.

[60] Kittayapong, P. and Jamnongluk, W. and Thipaksorn, A. and Milne, J. R. and Sindhusake, C. *Wolbachia* infection complexity among insects in the tropical rice-field community. *Mol. Eco.*, 12(4):1049–1060, 2003.

[61] Rigaud, T. and Juchault, P. Success and failure of horizontal transfers of feminizing Wolbachia endosymbionts in woodlice. *Journal of Evolutionary Biology*, 8(2):pp. 249 – 255, 1995.

[62] Baldo, L. and Ayoub, N. A. and Hayashi, C.Y. and Russel, J. A. and Stahult, J .K and Werre, J.H. Insight into the routes of Wolbachia invasion: high levels of horizontal transfer in the spider genus Agelenopsis revealed by Wolbachia strain and mitochondrial DNA diversity. *Mol. Eco.*, 17(2):pp. 557 – 569, 2008.

[63] P. Bouchon,D. and Rigaud, T. and Juchault. Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proc. Biol. Sci.*, 265:pp. 1081 – 1090, 1998.

[64] Schilthuizen M, Stouthamer R. Horizontal transmission of parthenogenesis-inducing microbes in Trichogramma wasps. *Proc. Biol. Sci.*, 264:pp. 361 – 366, 1997.

[65] David C. Chan. Mitochondria: Dynamic Organelles in Disease, Aging, and Development. *Cell*, 125(7):pp 1241 – 1252, 2006.

[66] Perna, N.T. and Kocher, T.D. Mitochondrial DNA: Molecular fossils in the nucleus. *Curr. Biol.*, 6(2):pp 128 – 129, 1996.

[67] Adams, K.L and Palmer, J.D. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol. Phylog. Evo*, 29(3):pp 380 – 395, 2003. <ce:title>Plant Molecular Evolution</ce:title>.

[68] Ballard, J.W.O and Rand, D.M. The Population Biology of Mitochondrial DNA and Its Phylogenetic Implications. *Annu. Rev. Ecol. Evol. Syst.*, 36:pp 621 – 642, 2005.

[69] Brown, W.M. and George, M. and Wilson, A.C. Rapid evolution of animal mitochondrial DNA. *PNAS*, 76(4):pp 1967 – 1971, 1979.

[70] Moriyama, E. and Powell, J. Synonymous substitution rates in Drosophila: Mitochondrial versus nuclear genes. *J. Mol. Evol.*, 45:378–391, 1997. 10.1007/PL00006243.

[71] Galtier, N. and Nabholz, B. and Glamin, S. and Hurst, G. D. D. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.*, 18(22):4541–4550, 2009.

[72] Hurst, G. D.D and Jiggins, F. M. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc. R. Soc. B*, 272(1572):1525– 1534, 2005.

[73] Hebert, P. D. N. and Cywinska, A. and Ball, S. L. and deWaard, J. R. Biological identifications through dna barcodes. *Proc. R. Soc. B*, 270(1512):313–321, 2003.

[74] Hebert, .l D. N AND Stoeckle, M. Y AND Zemlak, T. S AND Francis, C. M. Identification of birds through dna barcodes. *PLoS Biol.*, 2(10):e312, 09 2004.

[75] Kress, W. J. and Erickson, D. L. Dna barcodes: Genes, genomics, and bioinformatics. *PNAS*, 105(8):2761–2762, 2008.

[76] Frezal, L. and Leblois, R. Four years of DNA barcoding: Current advances and prospects. *Infection, Genetics and Evolution*, 8(5):727 – 736, 2008.

[77] Whittaker, R.J and Fernandez-Palaciso, J.M. *Island Biogeography: Ecology, Evolution and Conservation.* Oxfor University Press, Oxford, UK, 2007.

[78] Courtillot, V. and Davaille. A. and Besse, J. and Stock, J. Three distinct types of hotspots in the earth's mantle. *Earth Planet. Sci. Lett.*, 205(3-4):295 – 308, 2003.

[79] Emerson, B. C. Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Mol Ecol*, 11(6):pp 951 – 966, 2002.

[80] Neall, V. E and Trewick, S. A. The age and origin of the pacific islands: a geological overview. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 363(1508):3293–3308, 2008.

[81] Gillespie, R. G and Claridge, E. M and Goodacre, S. L. Biogeography of the fauna of French Polynesia: diversification within and between a series of hot spot archipelagos. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 363:3335– 3346, 2008.

[82] Michael Chinery. *Field Guide Inesects of Britain and Northern Europe.* Collins, 3rd edition edition, 1993.

[83] Duplouy, A. and Vermenot, C. and Davies, N. and Roderick, G. and Hurst, G. and Charlat, S. Assessing risks of Wolbachia DNA cross-specimen contamination following mass collection and ethanol storage. *Mol. Ecol. Res.*, 9(1):46–50, 2009.

[84] Dierkens, M. and Charlat, S. Contribuition ‡ la connaissance des araignÈes de Óles de la SociÈtÈ (PolynÈsie FranÁaise). *Rev. Arachnol.*, 17:63–81, 2009.

[85] Moritz, C. AND Cicero, . Dna barcoding: Promise and pitfalls. *PLoS Biol*, 2(10):e354, 09 2004.

[86] Pons, J. and Vogler, A. P. Complex pattern of coalescence and fast evolution of a mitochondrial rrna pseudogene in a recent radiation of tiger beetles. *Mol. Biol. Evol.*, 22(4):991–1000, 2005.

[87] Folmer, O. and Black, M. and Hoeh, W. and Lutz, R. and Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Mar Biol Biotechnol.*, 3(5):294– 299, October 1994.

[88] Baldo, L. and Lo, N. and Werren, John H. Mosaic Nature of the Wolbachia Surface Protein. *J. Bacteriol.*, 187(15):5406–5418, 2005.

[89] A.J. Drummond, B. Ashton, S. Buxton, M. Cheung, A. Cooper, C. Duran, M. Field, J. Heled, M. Kearse, S. Markowitz, R. Moir, S. Stones-Havas, S. Sturrock, T. Thierer, and Wilson Ashton. Geneious v5.4, Available from http://www.geneious.com/. -, -:–, 2011.

[90] Gouy, M. and Guindon, S. and Gascuel, O. Seaview version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.*, 27(2):221–224, 2010.

[91] Joseph Feselstein. *Inferring Phylogenies.* Sinauer Associates, Inc, 2004.

[92] Guindon, S. and Gascuel, O. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst. Biol.*, 52(5):696–704, 2003.

[93] Gillespie, R.G and Roderick, G.K. Arthropods in Islands: Colonization, Speciation and Conservation. *Annu. Rev. Entomol.*, 47:pp. 595 – 632, 2002.

[94] Craig, D.A. Geomorphology, development of running water habitats, and evolution of Black flies on Polynesiian islands. *Bioscience*, 53:pp. 1079 – 1093, 2003.

[95] Gillespie, R.G. Biogeography of spiders on remote oceanic islands of the Pacific: archipelagos as stepping stones? *J. Biogeography*, 29:pp. 655 – 662, 2002.

[96] Cowie, .t H. and Holland, Br.S. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *J. Biogeography*, 33(2):pp. 193 – 198, 2006.

[97] Garb J.E and Gillespie, R.G. Island hopping across the central Pacific: mitochondrial DNA detects sequential colonization of the Austral Islands by crab spiders (Araneae: Thomisidae). *J. Biogeography*, 33:pp. 201 – 220, 2006.

[98] Wright, S.D and Yong, C.G and Dawson, J.W and Whittaker, D.J and Gardner , R.C. Riding the ice age El Nino? PAcific biogeography and evolution of *Metrosideros* subg.*Metrosideros* (Mystaceae) inferred from nuclear ribosomal DNA. *Proc. Natl. Acad. Sci. U. S. A.*, 97:pp. 4118 – 4123, 2000.

[99] Gemill, C.E and Allan, G.J and Wagner, W.L and Zimmer, E.A. Evolution of insular Pacific *Pittosporum* (Pittosporaceae): origin of the Hawaiian radiation. *Mol. Phylogenet. Evol.*, 22:pp. 31 – 42, 2001.

[100] Carson, H.L. Tracing ancestry with chromosomal sequences. *Trends Ecol. Evolut.*, 2:pp.203 –207, 1987.

[101] Goodacre, S L. Population structure, history and gene flow in a group of closely related land snails: genetic variation in Partula from the Society Islands of the Pacific. *Mol. Eco.*, 11(1):pp. 55 – 68, 2002.

[102] Roderick, G. K. and Gillespie, R. G. Speciation and phylogeography of Hawaiian terrestrial arthropods. *Mol. Eco.*, 7(4):pp. 519 – 531, 1998.

[103] Papadopoulou, A. and Anastasiou, I. and Vogler, A. P. Revisiting the Insect Mitochondrial Molecular Clock: The Mid-Aegean Trench Calibration. *Mol. Biol. Evol.*, 27(7):1659–1672, 2010.

[104] Uto K. and Yamamoto Y., and Sudo M. and Uchiumi S. and Ishizuka O. and Kogiso T. and Tsunakawa H. New K-Ar ages of the Society Islands, French Polynesia, and implication for the Society hotspot feature. *Earth Planets Space*, 59:pp. 979 – 746, 2007.

[105] Sokal, R. and Oden, N. Spatial autocorrelation in biology: 1. Methodology. *Biol. J. Linn. Soc. Lond.*, 10:pp. – 199 – 228, 1978.

[106] Sokal, R. and Oden, N. Spatial autocorrelation in biology: 2.Some biological implications and four applications of evolutionary and ecological interest. *Biol. J. Linn. Soc. Lond.*, 10:pp. 229 –249, 1978.

[107] Gittleman, John L. and Kot, Mark. Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects. *Syst. Biol.*, 39(3):227–241, 1990.

[108] Flecketon, R. and Jetz, W. Space versus Phylogeny: disentangling phylogenetic and spatial signals in comparative data. *Proc. R. Soc. B*, 276:pp. 21 –30, 2009.

[109] Epperson, B. Covariances among join-count spatial autocorrelation measures. *Theor. Popul. Biol.*, 64:pp. 81 –87, 2003.

[110] Gotoh, Tetsuo and Sugasawa, Jun and Noda, Hiroaki and Kitashima, Yasuki. *Wolbachia*-induced cytoplasmic incompatibility in Japanese populations of *Tetranychus urticae* (Acari: Tetranychidae). *Exp. Appl. Acarol.*, 42:pp. 1–16, 2007. 10.1007/s10493-007-9072-3.

[111] Rodgers-Gray, T.P and Smith , J.E and Ashcroft , E.A and Isaac , R.E and Dunn, A.M. Mechanisms of parasite-induced sex reversal in Gammarus duebeni. *Int. J. Parasitol.*, 34(6):pp. 747 – 753, 2004.

[112] Goodacre, S. L. and Martin, O.Y. and Thomas, C. F.G and Hewitt , G.M. Wolbachia and other endosymbiont infections in spiders. *Mol. Eco.*, 15(2):pp. 517 – 527, 2006.

[113] v. d. Schulenburg, J. H.G. and Hurst, G.D.D. and Tetzlaff, Da. nd Booth, G.E and Zakharov, I.A. and Majerus, M.E.N. History of Infection With Different Male-Killing Bacteria in the Two-Spot Ladybird Beetle Adalia bipunctata Revealed Through Mitochondrial DNA Sequence Analysis. *Genetics*, 160:pp. 1075 – 1086, 2002.

[114] Giordano, R. and Jackson, J.J. and Robertson, H.M. The role of Wolbachia bacteria in reproductive incompatibilities and hybrid zones of Diabrotica beetles and gryllus crickets. *PNAS*, 94(21):pp. 11439 – 11444, 1997.

[115] Braig HR, Guzman H, Tesh RB, O'Neill SL. Replacement of the natural Wolbachia symbiont of Drosophila simulans with a mosquito counterpart. *Nature*, 367:pp. 453 – 455, 1994.

[116] Sinkins, S.P. and Braig,H.R. and Oneill, S. Wolbachia pipientis: Bacterial Density and Unidirectional Cytoplasmic Incompatibility between Infected Populations of Aedes albopictus. *Exp. Parasitol.*, 81(3):pp. 284 – 291, 1995.

[117] Li, Z-X and Lin, H-Z and Guo, X-P. Prevalence of *Wolbachia* Infection in *;Bemisia tabaci*. *Curr. Microbiol.*, 54:pp. 467 – 471, 2007. 10.1007/s00284-007-0011-7.

[118] Kikuchi, Y. and Fukatsu, T. Diversity of Wolbachia Endosymbionts in Heteropteran Bugs. *Appl. Environ. Microbiol.*, 69(10):pp. 6082 – 6090, 2003.

[119] Perrot-Minnot, M. J. and Guo, L. R. and Werren, J. H. Single and Double Infections with Wolbachia in the Parasitic Wasp Nasonia vitripennis: Effects on Compatibility. *Genetics*, 143(2):pp. 961 – 972, 1996.

[120] Wenseleers T, Ito F, Van Borm S, Huybrechts R, Volckaert F, Billen J. Widespread occurrence of the micro-organism Wolbachia in ants. *Proc. Biol. Sci.*, 265:pp. 1447 – 1452, 1998.

[121] Grandjean F, Rigaud T, Raimond R, Juchault P, Souty-Grosset C. Mitochondrial DNA polymorphism and feminizing sex factors dynamics in a natural population of Armadillidium vulgare (Crustacea, Isopoda). *Genetica*, 92:pp. 55 – 60, 1993.

[122] Bordenstein, S. and Rosengaus, R. Discovery of a Novel *Wolbachia* Supergroup in Isoptera. *Curr. Microbiol.*, 51:pp. 393 – 398, 2005. 10.1007/s00284-005-0084-0.

[123] Kumpulainen T, Grapputo A, Mappes J. Parasites and sexual reproduction in psychid moths. *Evolution*, 58:pp. 1511 – 1520, 2004.

[124] Thipaksorn, A. and Jamnongluk, W. and Kittayapong, P. Molecular Evidence of *Wolbachia* Infection in Natural Populations of Tropical Odonates. *Curr. Microbiol.*, 47:pp. 0314 – 0318, 2003. 10.1007/s00284-002-4010-4.

[125] Dong, P. and Wang, J.J and Hu, F. and Jia, F.X. Influence of wolbachia infection on the fitness of the stored-product pest Liposcelis tricolor (Psocoptera: Liposcelididae). *J. Econ. Entomol.*, 100:pp. 1476 – 1481, 2007.

[126] Zeh, D W and Zeh, J A and Bonilla, M M. Wolbachia, sex ratio bias and apparent male killing in the harlequin beetle riding pseudoscorpion. *Heredity*, 95(1):41–49, June 2005.

[127] Gorham, C.H. and Fang, Q.Q and Durden, L.A. Wolbachia endosymbionts in fleas (Siphanoptera). *J. Parasit.*, 89:pp. 283 – 289, 2003.

[128] Arakaki N. and Miyoshi T. and Noda, H. Wolbachia-mediated parthenogenesis in the predatory thrips Franklinothrips vespiformis (Thysanoptera: Insecta). *Proc. R. Soc. B*, 268:pp. 1011 – 1016, 2001.

[129] West, S.A and Cook, J.M and Qerren, J.H and Godfray, C.J. *Wolbachia* in two insect host-parasitoid communities. *Mol. Eco.*, 7:pp. 1457 – 1465, 1998.

[130] Tagami, Y. and Miura, K. Distribution and prevalence of *Wolbachia* in Japanese populations of Lepidoptera. *Insect. Mol. Biol.*, 13:pp. 359 – 364, 2004.

[131] Gotoh, T and Noda, H. and Hong,X.-Y. *Wolbachia* distribution and cytoplasmic incompatibility based on a survey of 42 spider mite species (Acari: Tetranychidae) in Japan. *Heredity*, 91:pp. 208 – 216, 2003.

[132] Rowley, S. and Raven, R.J. and McGraw, E. A. *Wolbachia pipientis*in Australian spiders. *Curr. Microbiol.*, 40:pp. 208 – 214, 2004.

[133] Baldo, L. and Prendini, L. and Corthals, A. and Werren, J. *Wolbachia* Are Present in Southern African Scorpions and Cluster with Supergroup F. *Curr. Microbiol.*, 55:pp. 367 – 373, 2007. 10.1007/s00284-007-9009-4.

[134] Simões, P. M. and Mialdea, G. and Reiss, D. and Sagot, M.-F. and Charlat, S. Wolbachia detection: an assessment of standard PCR Protocols. *Mol. Ecol. Resour.*, 11(3):567–572, 2011.

[135] Arthofer, W. and Riegler, M. and Schneider, D. and Krammer, M. and Miller, W. J. and Stauffer, C. Hidden *Wolbachia* diversity in field populations of the European cherry fruit fly, Rhagoletis cerasi (Diptera, Tephritidae). *Mol. Eco.*, 18(18):3816–3830, 2009.

[136] Sinkins, S.P. and Braig, H.R and ONeill, S.L. Wolbachia superinfections and the expresssion of cytoplasmic incompatibility. *Proc. Biol. Sci.*, 261:pp. 325 – 330, 1995.

[137] Dobson, S.L. and Marsland, E.J. and Rattanadechakul, W. Wolbachia-induced cytoplasmic incompatibility in single- and superinfected Aedes albopictus (Diptera: Culicidae). *J. Med. Entom.*, 38:pp. 382 – 387, 2001.

[138] Perrot-Minot, M.J. and Guo, L.R. and Werren, J.H. Single and double infections with *Wolbachia*in the parasitic wasp *Nasonia vitripennis*: effects on compatibility. *Genetics*, 143:pp. 661 – 972, 1996.

[139] Van Borm, S. and Wenseleers, T. and Billen,J. and Boomsma, J.J. Cloning and sequencing of wsp encoding gene fragments reveals a diversity of co-infecting Wolbachia strains in Acromyrmex leafcutter ants. *Mol. Phylogenet. Evol.*, 26:pp. 102 – 109, 2003.

[140] Malloch, G. and Fenton, B. and Butcher, R. D. J. Molecular evidence for multiple infections of a new subgroup of *Wolbachia* in the European raspberry beetle *Byturus tomentosus*. *Mol. Eco.*, 9(1):pp. 77 – 90, 2000.

[141] Zhang, X., and Luckhart, S., and Tu, Z., and Pfeiffer, D. G. Analysis of Wolbachia strains associated with Conotrachelus nenuphar (Coleoptera: Curculionidae) in the Eastern United States. *Environ. Entom.*, 39:pp. 396 – 405, 2010.

[142] Jeyaprakash, A. and Hoy, M. A. Long PCR improves Wolbachia DNA amplification: wsp sequences found in 76% of sixty-three arthropod species. *Insect. Mol. Biol.*, 9(4):393–405, 2000.

[143] Ikeda,T. and Ishikawa, H. and Sasaki. Regulation of Wolbachia density in the Mediterranean flour moth, Ephestia kuehniella, and the almond moth, Cadra cautella. *Zool. Sci.*, 20:pp. 153 – 157, 2003.

[144] Hiroki, M. and Tagami, Y. and Miura, K. and Kato, Y. Multiple infection with Wolbachia inducing different reproductive manipulations in the butterfly Eurema hecabe. *Proc. Biol. Sci.*, 271:pp. 1751 – 1755, 2004.

[145] Jacob A. Russell. The ants (Hymenoptera: Formicidae) are unique and enigmatic hosts of prevalent Wolbachia (Alphaproteobacteria) symbionts. *Myrmec. News*, 16:pp. 7 – 23, 2012.

[146] Dunn, A.K. and Stabb, E.V. Culture-Independent Characterization of the Microbiota of the Ant Lion Myrmeleon mobilis (Neuroptera: Myrmeleontidae). *Appl. Environ. Microbiol.*, 71(12):pp 8784 – 8794, 2005.

[147] Weeks, A. R. and Velten, R. and Stouthamer, R. Incidence of a new sex ratio distorting endosymbiotic bacterium among arthropods. *Proc. Royal Soc.of Lon.B Biol. Sci.*, 270(1526):1857–1865, 2003.

[148] Andreotti, R. and Perez de Leon, A. and Dowd, S. and Guerrero, F. and Bendele, K. and Scoles, G. Assessment of bacterial diversity in the cattle tick *Rhipicephalus (Boophilus)* microplus through tag-encoded pyrosequencing. *BMC Microbiology*, 11(1):6, 2011.

[149] Meyer, M. and Stenzel, U. and Hofreiter, M. Parallel tagged sequencing on the 454 platform. *Nature Protocols*, 3:pp 267 – 278, 2008.

[150] Raychoudhury, R. and Baldo, L. and Oliveira, De C S G and Werren, Jo H. Modes of acquisition of Wolbachia: horizontal transfer, hybrid introgression, and codivergence in the Nasonia species complex. *Evolution*, 63:pp. 165 – 183, 2009.

[151] Werren, J.H and Baldo, L. and Clark, M.E. Wolbachia: master manipulators of invertebrate biology. *Nat. Rev. Microbiol.*, 6:pp. 741 – 751, 2008.

[152] Ochman, H. and Elwyn, S. and Moran, N. A. Calibrating bacterial evolution. *PNAS*, 96(22):pp 12638–12643, 1999.

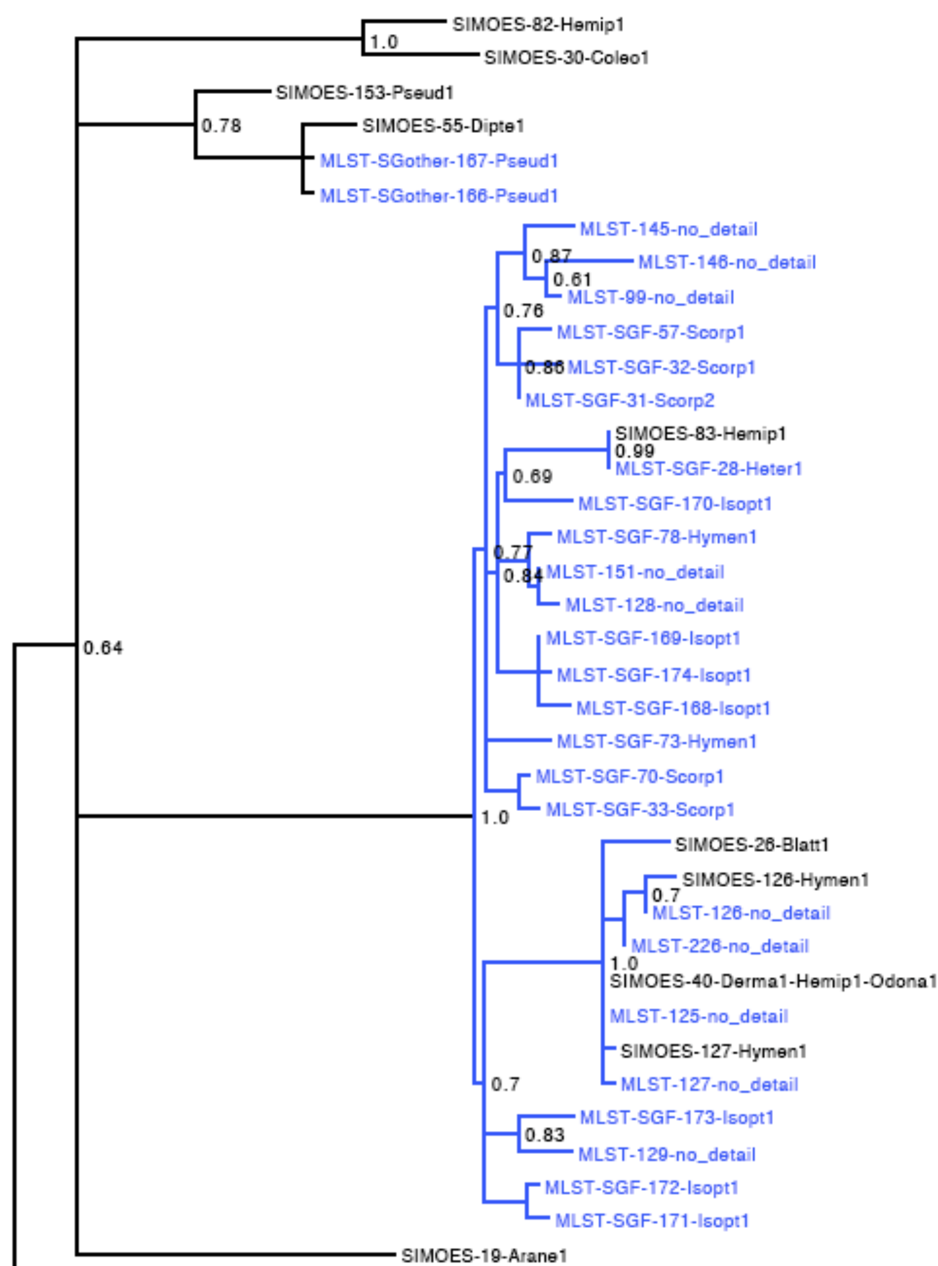
[153] Vincent H. Resh and Ring T. CardÈ, editor. *Encyclopedia of Insects(Second Edition)*. Academic Press, 2009.

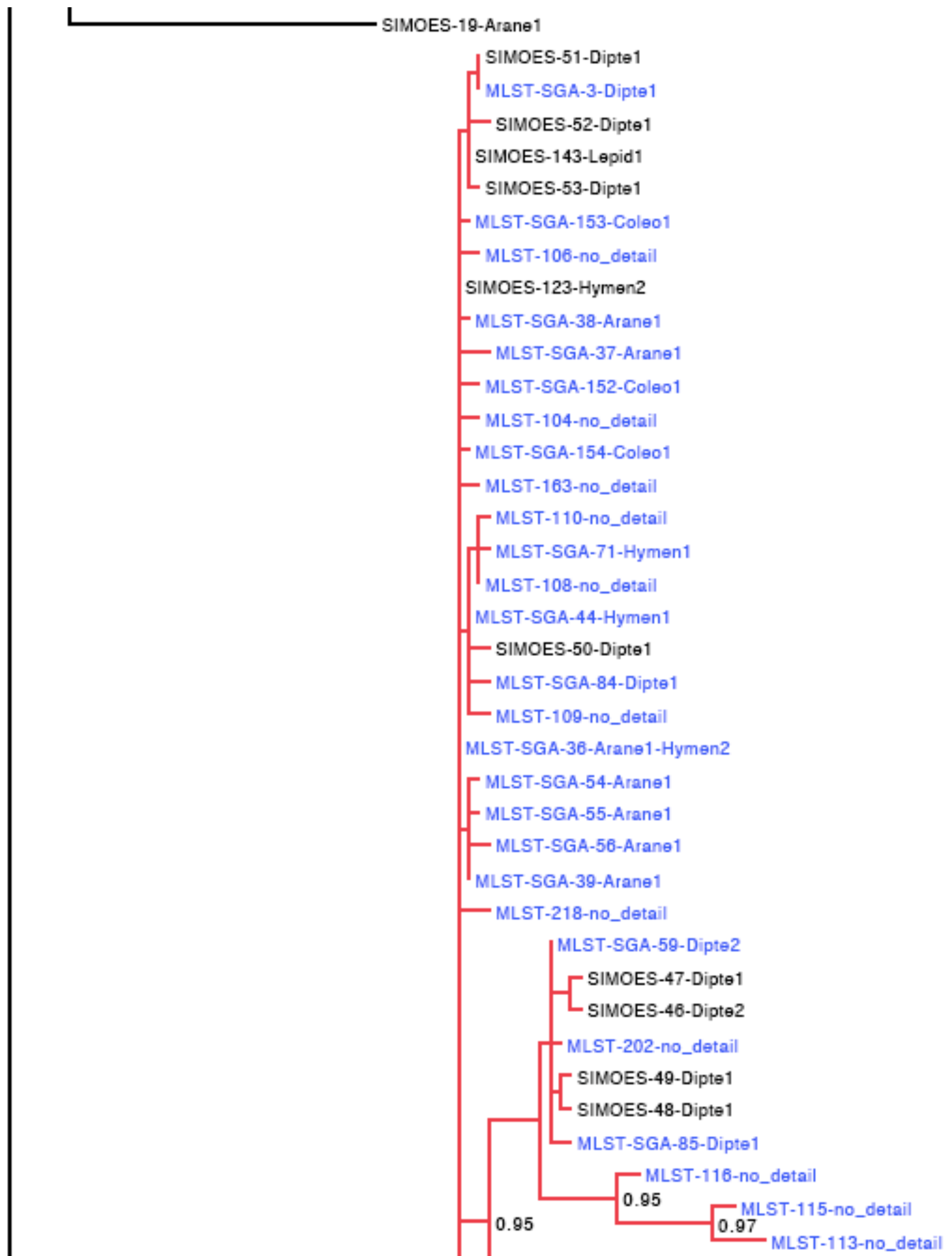
Appendix

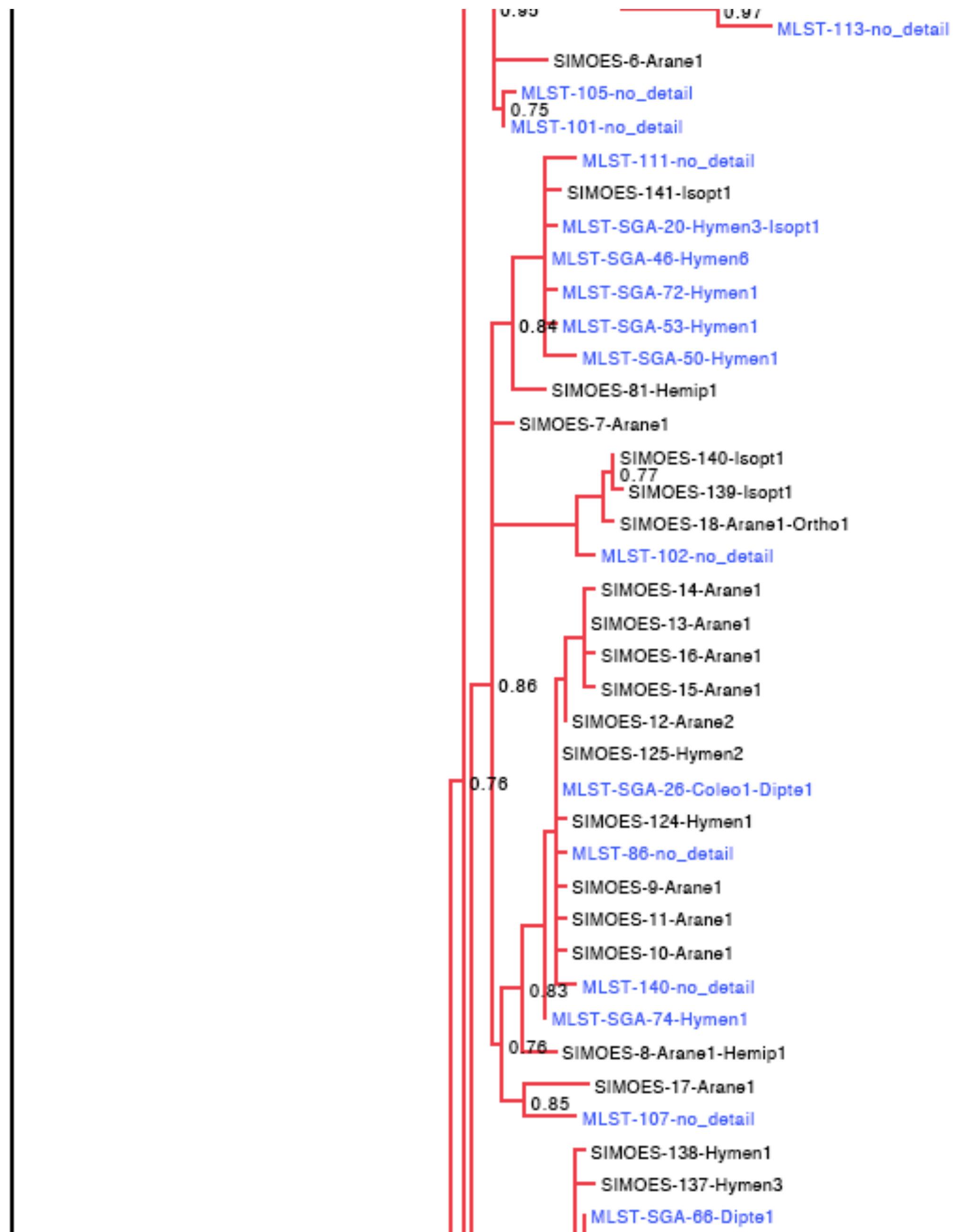
Table 4.1: Probabilities of Morgan's Index (I) for spatial structure in the distribution of COI haplotypes due to the experimental set-up. A total of 23 clusters-3% (treated here as species – sp) for which 2 or more distinct COI alleles were found in 2 or more collecting locations, in the same island, were used. Indicated are the Probabilities of the expected I being lower (Pinf), equal (Peq), superior (Psup) or $\geq$ to the observed I.

Island	sp	Pinf	Peq	Psup	P($\geq$)
Huahine	38	0.00	0.79	0.21	1.00
Tahiti	103	0.00	0.61	0.39	1.00
Tahiti	204	0.00	0.83	0.17	1.00
Tahiti	449	0.00	0.61	0.39	1.00
Tahiti	803	0.00	0.58	0.43	1.00
Tahiti	900	0.00	1.00	0.00	1.00
Tahiti	929	0.00	0.27	0.73	1.00
Raiatea	46	0.38	0.41	0.21	0.62
Raiatea	222	0.39	0.16	0.45	0.61
Tahiti	412	0.40	0.60	0.00	0.60
Raiatea	226	0.40	0.32	0.27	0.60
Tahiti	923	0.56	0.44	0.00	0.44
Tahiti	369	0.56	0.35	0.09	0.44
Huahine	341	0.61	0.39	0.00	0.39
Huahine	802	0.62	0.23	0.15	0.38
Huahine	519	0.62	0.30	0.08	0.38
Huahine	399	0.63	0.37	0.00	0.37
Huahine	1020	0.65	0.36	0.00	0.36
Tahiti	961	0.65	0.35	0.00	0.35
Huahine	1025	0.67	0.33	0.00	0.33
Tahiti	946	0.67	0.33	0.00	0.33
Raiatea	25	0.69	0.31	0.00	0.31
Moorea	237	0.80	0.12	0.08	0.20

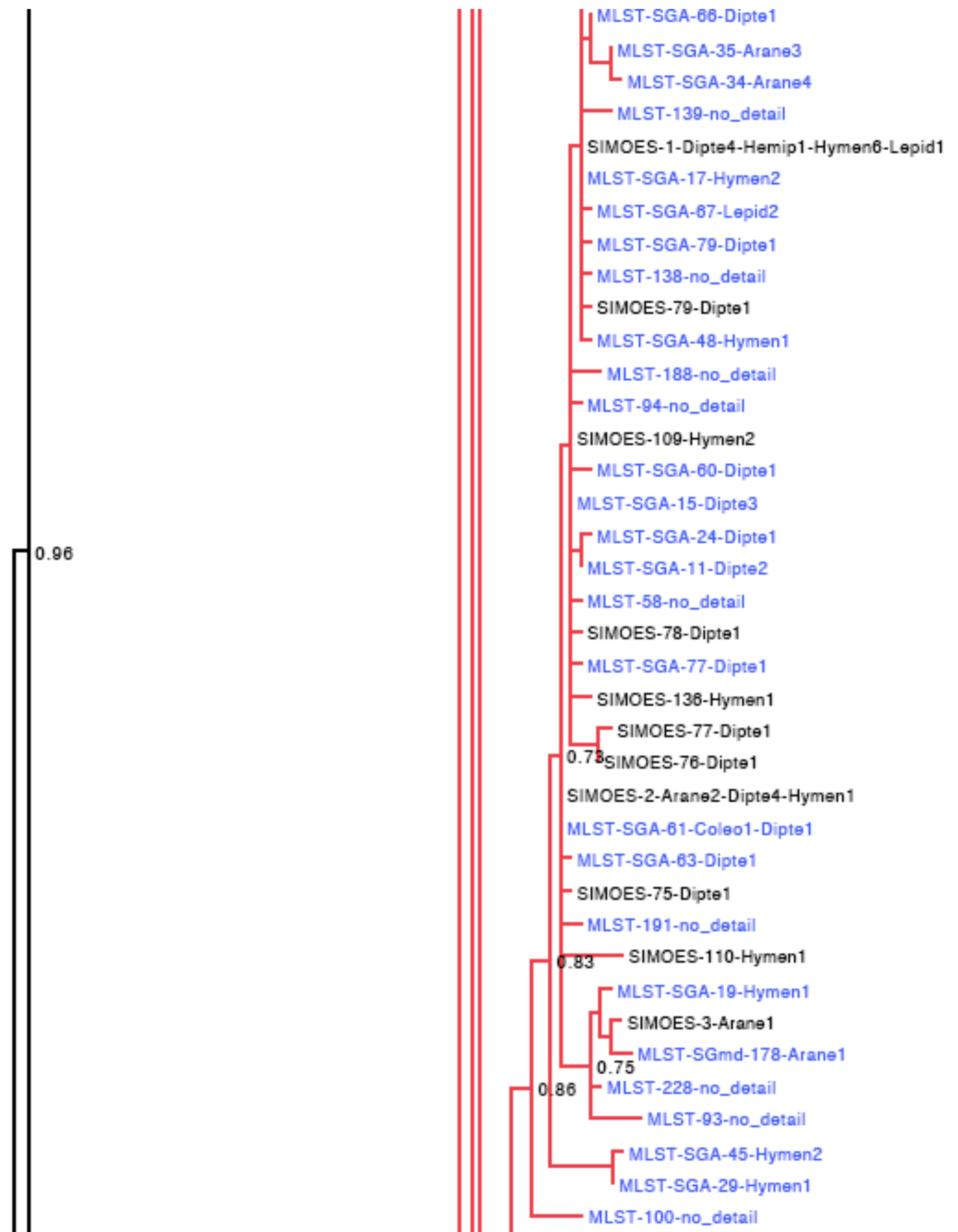
Figure 4.1: ML tree with FpA alleles from the MLST database and our FbpA haplotypes. All branches with less then 0.5 aLRT support were collapsed into multifurcations. FbpA alleles form the MLST database are identified as such in their respective sequences and supergroup assignation (SG*supergroup letter*), whenever this information was available in the database. The names of these alleles appear in blue while the FbpA alleles found in our work appear in black and are identified with SIMOES and the Order(s) where they were found. For simplicity, the branches of the monophyletic groups corresponding to distinct supergroups were coloured with different colours: in blue - supergroup F; in red -supergroup A and in green - supergroup B . Due to the size of the tree, we opted to cut it and include it in multiple pages.

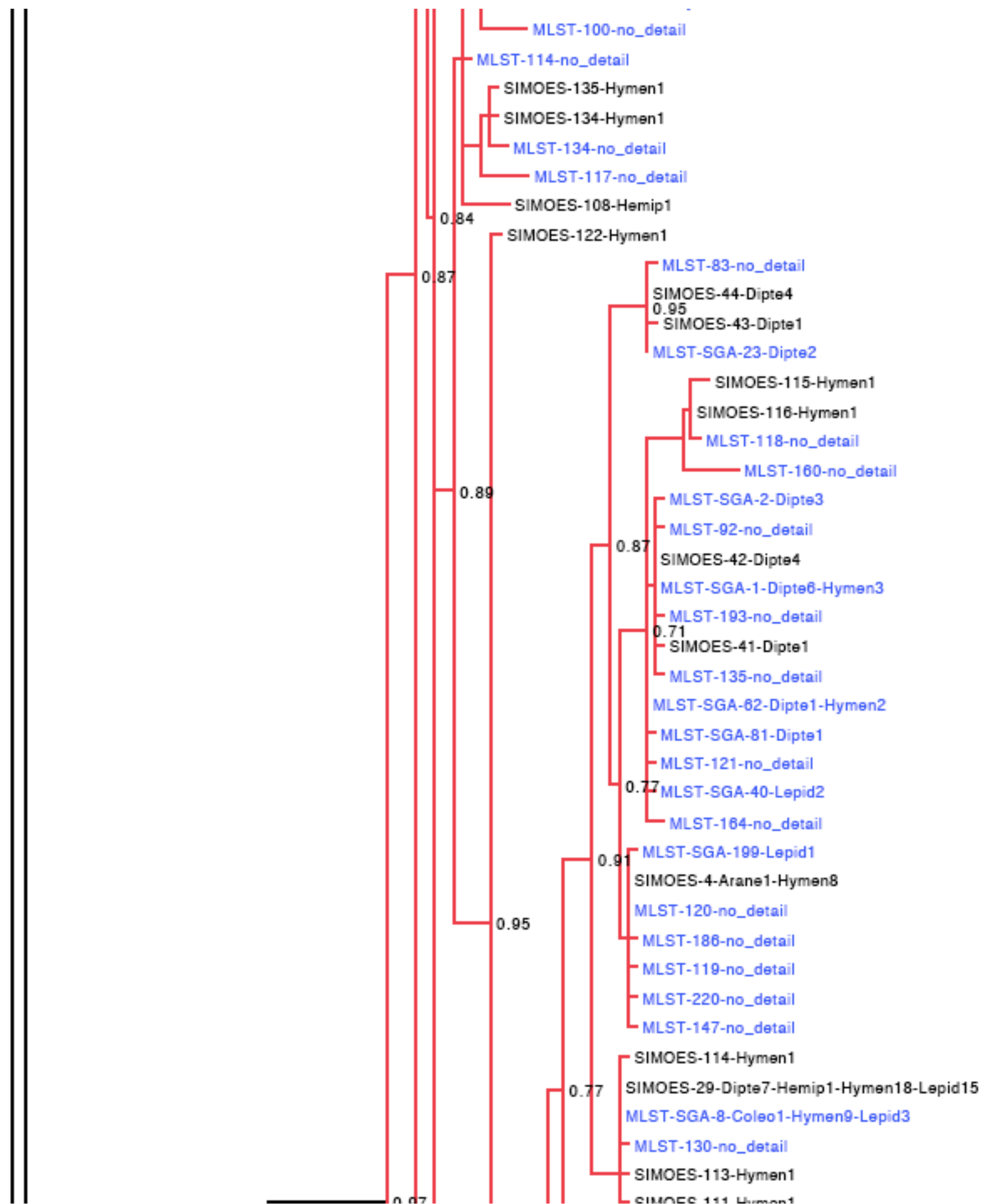


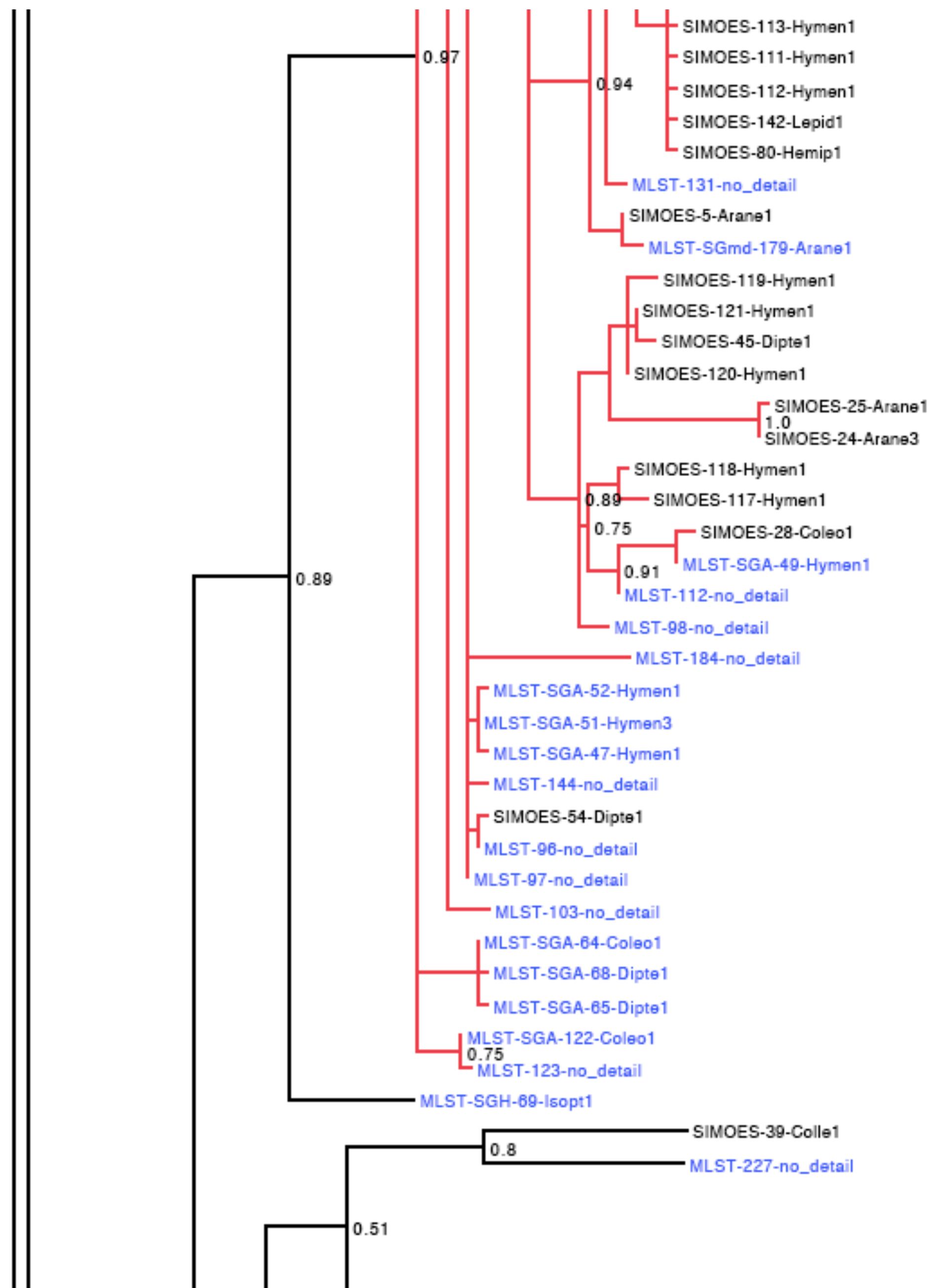

SIMOES-19-Arane1
SIMOES-51-Dipte1
MLST-SGA-3-Dipte1
SIMOES-52-Dipte1
SIMOES-143-Lepid1
SIMOES-53-Dipte1
MLST-SGA-153-Coleo1
MLST-106-no_detail
SIMOES-123-Hymen2
MLST-SGA-38-Arane1
MLST-SGA-37-Arane1
MLST-SGA-152-Coleo1
MLST-104-no_detail
MLST-SGA-154-Coleo1
MLST-163-no_detail
MLST-110-no_detail
MLST-SGA-71-Hymen1
MLST-108-no_detail
MLST-SGA-44-Hymen1
SIMOES-50-Dipte1
MLST-SGA-84-Dipte1
MLST-109-no_detail
MLST-SGA-36-Arane1-Hymen2
MLST-SGA-54-Arane1
MLST-SGA-55-Arane1
MLST-SGA-56-Arane1
MLST-SGA-39-Arane1
MLST-218-no_detail
MLST-SGA-59-Dipte2
SIMOES-47-Dipte1
SIMOES-46-Dipte2
MLST-202-no_detail
SIMOES-49-Dipte1
SIMOES-48-Dipte1
MLST-SGA-85-Dipte1
MLST-116-no_detail
0.95
MLST-115-no_detail
0.97
MLST-113-no_detail
0.95



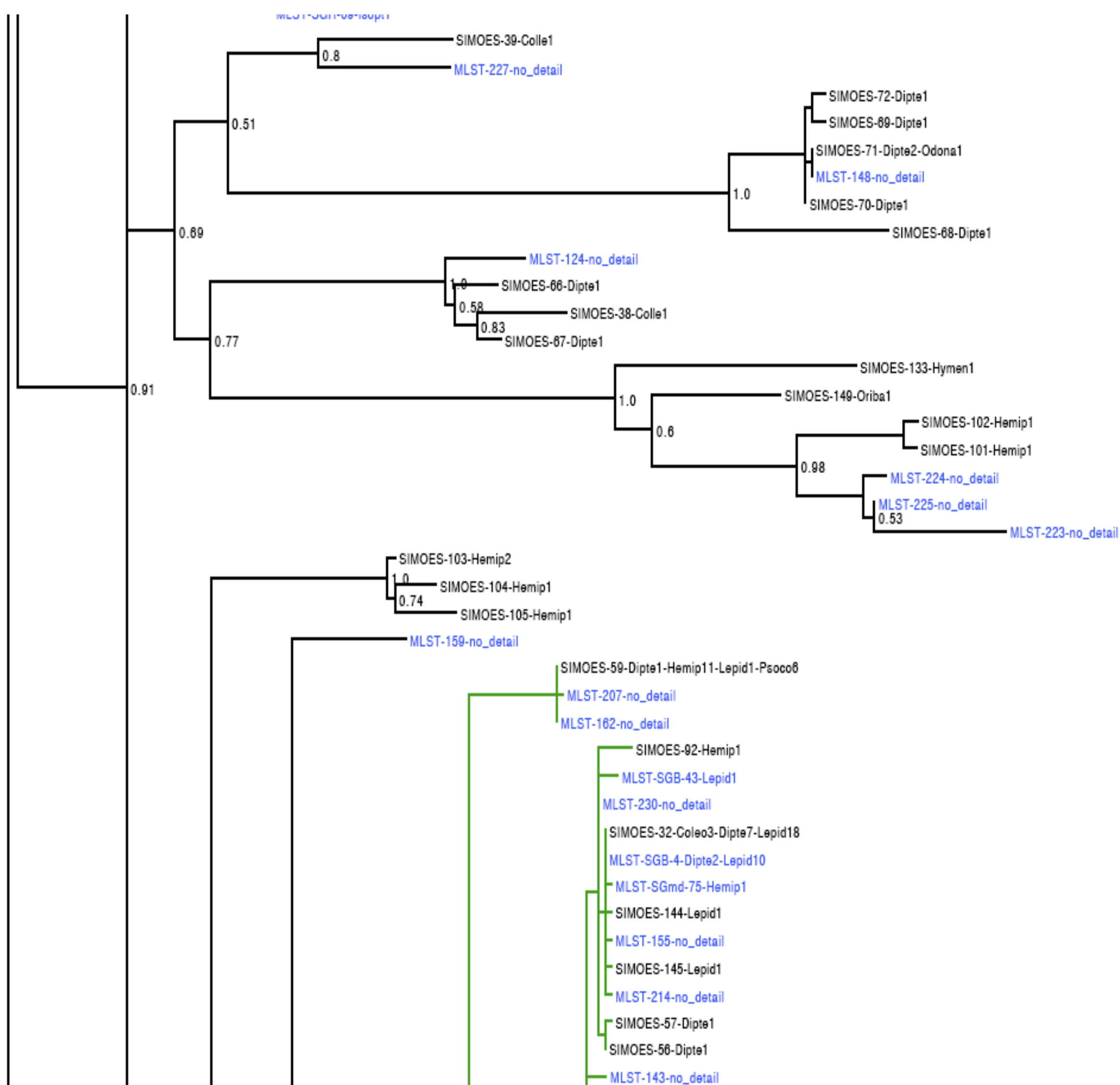
0.95
0.97
MLST-113-no_detail
SIMOES-6-Arane1
MLST-105-no_detail
0.75
MLST-101-no_detail
MLST-111-no_detail
SIMOES-141-Isopt1
MLST-SGA-20-Hymen3-Isopt1
MLST-SGA-46-Hymen6
MLST-SGA-72-Hymen1
0.84
MLST-SGA-53-Hymen1
MLST-SGA-50-Hymen1
SIMOES-81-Hemip1
SIMOES-7-Arane1
SIMOES-140-Isopt1
0.77
SIMOES-139-Isopt1
SIMOES-18-Arane1-Ortho1
MLST-102-no_detail
SIMOES-14-Arane1
SIMOES-13-Arane1
SIMOES-16-Arane1
0.86
SIMOES-15-Arane1
SIMOES-12-Arane2
SIMOES-125-Hymen2
0.76
MLST-SGA-26-Coleo1-Dipte1
SIMOES-124-Hymen1
MLST-86-no_detail
SIMOES-9-Arane1
SIMOES-11-Arane1
SIMOES-10-Arane1
0.83
MLST-140-no_detail
MLST-SGA-74-Hymen1
0.76
SIMOES-8-Arane1-Hemip1
SIMOES-17-Arane1
0.85
MLST-107-no_detail
SIMOES-138-Hymen1
SIMOES-137-Hymen3
MLST-SGA-66-Dipte1


0.96
MLST-SGA-66-Dipte1
MLST-SGA-35-Arane3
MLST-SGA-34-Arane4
MLST-139-no_detail
SIMOES-1-Dipte4-Hemip1-Hymen6-Lepid1
MLST-SGA-17-Hymen2
MLST-SGA-67-Lepid2
MLST-SGA-79-Dipte1
MLST-138-no_detail
SIMOES-79-Dipte1
MLST-SGA-48-Hymen1
MLST-188-no_detail
MLST-94-no_detail
SIMOES-109-Hymen2
MLST-SGA-60-Dipte1
MLST-SGA-15-Dipte3
MLST-SGA-24-Dipte1
MLST-SGA-11-Dipte2
MLST-58-no_detail
SIMOES-78-Dipte1
MLST-SGA-77-Dipte1
SIMOES-136-Hymen1
SIMOES-77-Dipte1
0.73
SIMOES-76-Dipte1
SIMOES-2-Arane2-Dipte4-Hymen1
MLST-SGA-61-Coleo1-Dipte1
MLST-SGA-63-Dipte1
SIMOES-75-Dipte1
MLST-191-no_detail
0.83
SIMOES-110-Hymen1
MLST-SGA-19-Hymen1
SIMOES-3-Arane1
MLST-SGmd-178-Arane1
0.75
0.86
MLST-228-no_detail
MLST-93-no_detail
MLST-SGA-45-Hymen2
MLST-SGA-29-Hymen1
MLST-100-no_detail

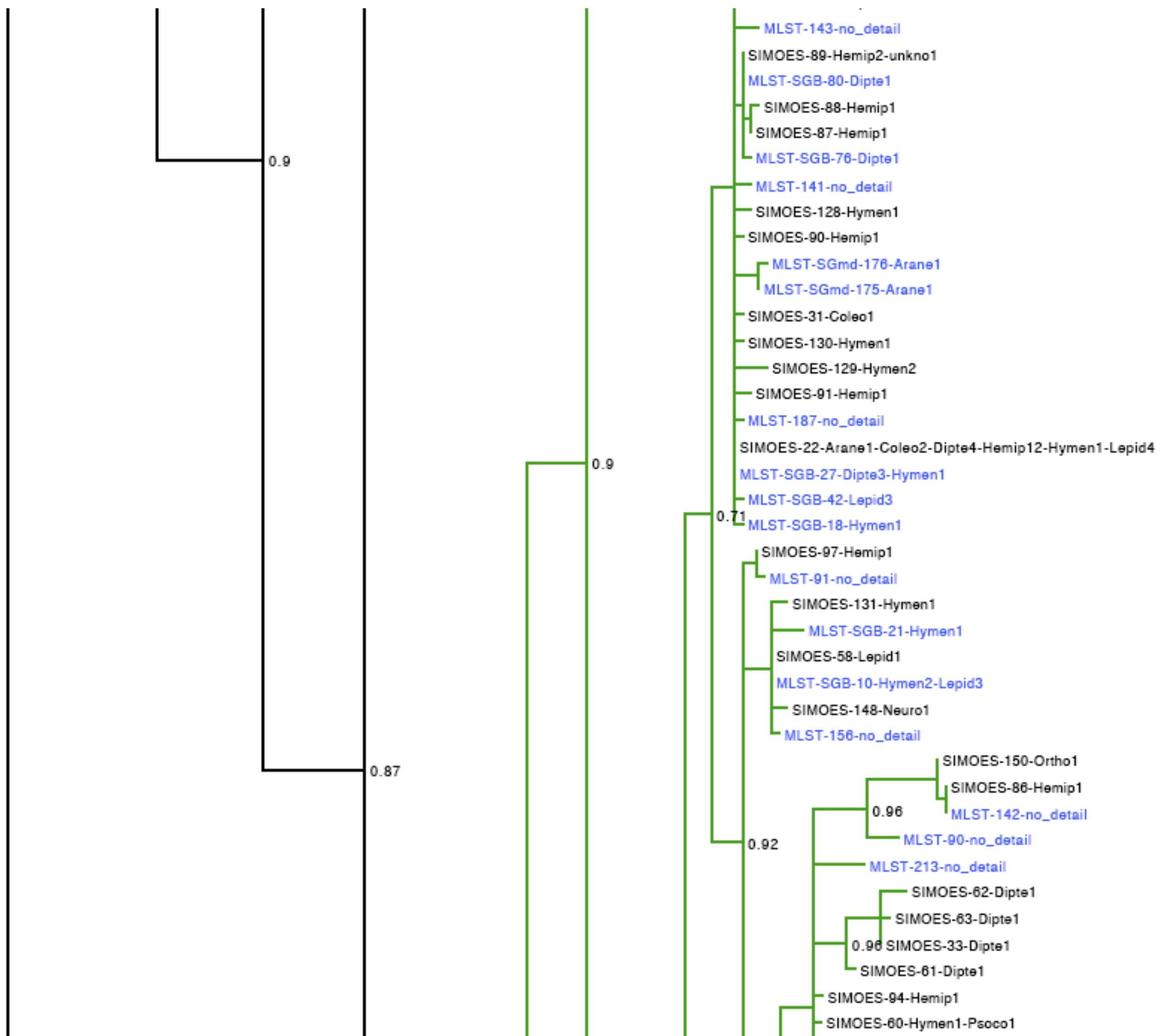

MLST-100-no_detail
MLST-114-no_detail
SIMOES-135-Hymen1
SIMOES-134-Hymen1
MLST-134-no_detail
MLST-117-no_detail
SIMOES-108-Hemip1
0.84
SIMOES-122-Hymen1
MLST-83-no_detail
0.87
SIMOES-44-Dipte4
0.95
SIMOES-43-Dipte1
MLST-SGA-23-Dipte2
SIMOES-115-Hymen1
SIMOES-116-Hymen1
MLST-118-no_detail
MLST-160-no_detail
MLST-SGA-2-Dipte3
0.89
MLST-92-no_detail
0.87
SIMOES-42-Dipte4
MLST-SGA-1-Dipte6-Hymen3
MLST-193-no_detail
0.71
SIMOES-41-Dipte1
MLST-135-no_detail
MLST-SGA-62-Dipte1-Hymen2
MLST-SGA-81-Dipte1
MLST-121-no_detail
0.77
MLST-SGA-40-Lepid2
MLST-164-no_detail
MLST-SGA-199-Lepid1
0.91
SIMOES-4-Arane1-Hymen8
MLST-120-no_detail
0.95
MLST-186-no_detail
MLST-119-no_detail
MLST-220-no_detail
MLST-147-no_detail
SIMOES-114-Hymen1
0.77
SIMOES-29-Dipte7-Hemip1-Hymen18-Lepid15
MLST-SGA-8-Coleo1-Hymen9-Lepid3
MLST-130-no_detail
SIMOES-113-Hymen1

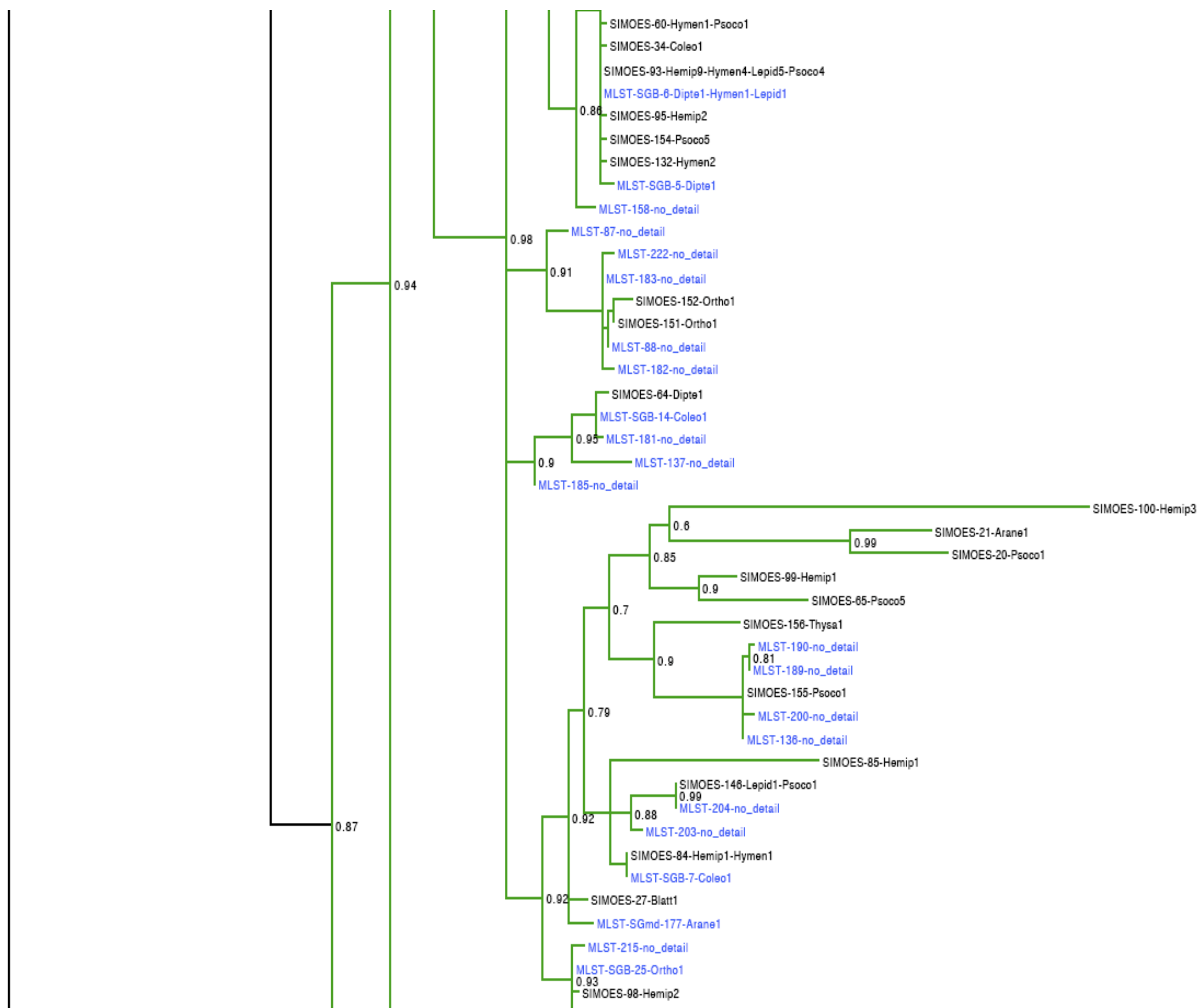


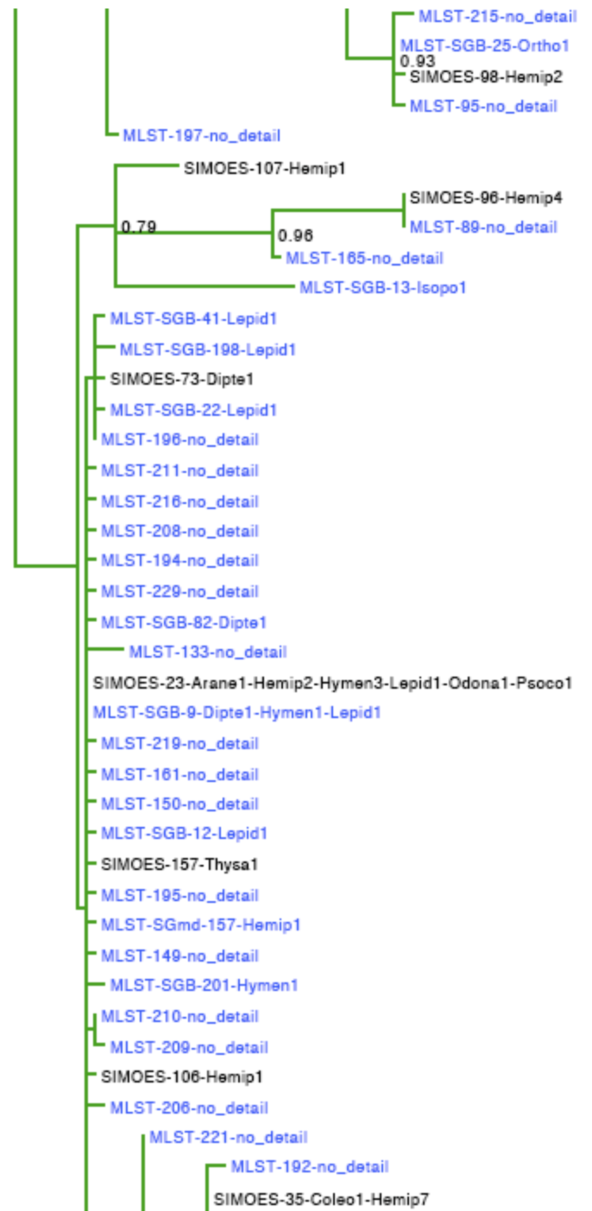
SIMOES-113-Hymen1
SIMOES-111-Hymen1
SIMOES-112-Hymen1
SIMOES-142-Lepid1
SIMOES-80-Hemip1
0.97
0.94
MLST-131-no_detail
SIMOES-5-Arane1
MLST-SGmd-179-Arane1
SIMOES-119-Hymen1
SIMOES-121-Hymen1
SIMOES-45-Dipte1
SIMOES-120-Hymen1
SIMOES-25-Arane1
1.0
SIMOES-24-Arane3
SIMOES-118-Hymen1
0.89
SIMOES-117-Hymen1
0.75
SIMOES-28-Coleo1
0.91
MLST-SGA-49-Hymen1
MLST-112-no_detail
MLST-98-no_detail
0.89
MLST-184-no_detail
MLST-SGA-52-Hymen1
MLST-SGA-51-Hymen3
MLST-SGA-47-Hymen1
MLST-144-no_detail
SIMOES-54-Dipte1
MLST-96-no_detail
MLST-97-no_detail
MLST-103-no_detail
MLST-SGA-64-Coleo1
MLST-SGA-68-Dipte1
MLST-SGA-65-Dipte1
MLST-SGA-122-Coleo1
0.75
MLST-123-no_detail
MLST-SGH-69-Isopt1
SIMOES-39-Colle1
0.8
MLST-227-no_detail
0.51

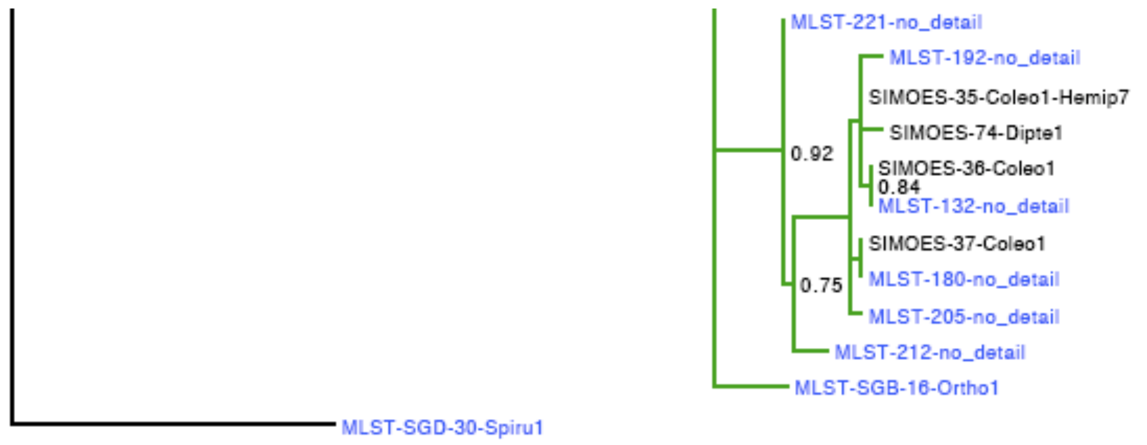


SIMOES-39-Colle1
0.8
MLST-227-no_detail
SIMOES-72-Dipte1
SIMOES-69-Dipte1
0.51
SIMOES-71-Dipte2-Odona1
MLST-148-no_detail
1.0
SIMOES-70-Dipte1
0.69
SIMOES-68-Dipte1
MLST-124-no_detail
SIMOES-66-Dipte1
0.58
SIMOES-38-Colle1
0.83
0.77
SIMOES-67-Dipte1
SIMOES-133-Hymen1
0.91
SIMOES-149-Oriba1
1.0
SIMOES-102-Hemip1
0.6
SIMOES-101-Hemip1
0.98
MLST-224-no_detail
MLST-225-no_detail
0.53
MLST-223-no_detail
SIMOES-103-Hemip2
SIMOES-104-Hemip1
0.74
SIMOES-105-Hemip1
MLST-159-no_detail
SIMOES-59-Dipte1-Hemip11-Lepid1-Psoco6
MLST-207-no_detail
MLST-162-no_detail
SIMOES-92-Hemip1
MLST-SGB-43-Lepid1
MLST-230-no_detail
SIMOES-32-Coleo3-Dipte7-Lepid18
MLST-SGB-4-Dipte2-Lepid10
MLST-SGmd-75-Hemip1
SIMOES-144-Lepid1
MLST-155-no_detail
SIMOES-145-Lepid1
MLST-214-no_detail
SIMOES-57-Dipte1
SIMOES-56-Dipte1
MLST-143-no_detail


MLST-143-no_detail
SIMOES-89-Hemip2-unkno1
MLST-SGB-80-Dipte1
SIMOES-88-Hemip1
SIMOES-87-Hemip1
MLST-SGB-76-Dipte1
0.9
MLST-141-no_detail
SIMOES-128-Hymen1
SIMOES-90-Hemip1
MLST-SGmd-176-Arane1
MLST-SGmd-175-Arane1
SIMOES-31-Coleo1
SIMOES-130-Hymen1
SIMOES-129-Hymen2
SIMOES-91-Hemip1
MLST-187-no_detail
SIMOES-22-Arane1-Coleo2-Dipte4-Hemip12-Hymen1-Lepid4
0.9
MLST-SGB-27-Dipte3-Hymen1
MLST-SGB-42-Lepid3
0.71
MLST-SGB-18-Hymen1
SIMOES-97-Hemip1
MLST-91-no_detail
SIMOES-131-Hymen1
MLST-SGB-21-Hymen1
SIMOES-58-Lepid1
MLST-SGB-10-Hymen2-Lepid3
SIMOES-148-Neuro1
MLST-156-no_detail
SIMOES-150-Ortho1
0.87
SIMOES-86-Hemip1
0.96
MLST-142-no_detail
MLST-90-no_detail
0.92
MLST-213-no_detail
SIMOES-62-Dipte1
SIMOES-63-Dipte1
0.96
SIMOES-33-Dipte1
SIMOES-61-Dipte1
SIMOES-94-Hemip1
SIMOES-60-Hymen1-Psoco1


SIMOES-60-Hymen1-Psoco1
SIMOES-34-Coleo1
SIMOES-93-Hemip9-Hymen4-Lepid5-Psoco4
MLST-SGB-6-Dipte1-Hymen1-Lepid1
0.86
SIMOES-95-Hemip2
SIMOES-154-Psoco5
SIMOES-132-Hymen2
MLST-SGB-5-Dipte1
MLST-158-no_detail
0.98
MLST-87-no_detail
MLST-222-no_detail
0.94
0.91
MLST-183-no_detail
SIMOES-152-Ortho1
SIMOES-151-Ortho1
MLST-88-no_detail
MLST-182-no_detail
SIMOES-64-Dipte1
MLST-SGB-14-Coleo1
0.95
MLST-181-no_detail
0.9
MLST-137-no_detail
MLST-185-no_detail
SIMOES-100-Hemip3
0.6
SIMOES-21-Arane1
0.99
SIMOES-20-Psoco1
0.85
SIMOES-99-Hemip1
0.9
SIMOES-65-Psoco5
0.7
SIMOES-156-Thysa1
MLST-190-no_detail
0.81
0.9
MLST-189-no_detail
SIMOES-155-Psoco1
MLST-200-no_detail
MLST-136-no_detail
0.79
SIMOES-85-Hemip1
SIMOES-146-Lepid1-Psoco1
0.99
MLST-204-no_detail
0.87
0.92
0.88
MLST-203-no_detail
SIMOES-84-Hemip1-Hymen1
MLST-SGB-7-Coleo1
0.92
SIMOES-27-Blatt1
MLST-SGmd-177-Arane1
MLST-215-no_detail
MLST-SGB-25-Ortho1
0.93
SIMOES-98-Hemip2


MLST-215-no_detail
MLST-SGB-25-Ortho1
0.93
SIMOES-98-Hemip2
MLST-95-no_detail
MLST-197-no_detail
SIMOES-107-Hemip1
SIMOES-96-Hemip4
0.79
MLST-89-no_detail
0.96
MLST-165-no_detail
MLST-SGB-13-Isopo1
MLST-SGB-41-Lepid1
MLST-SGB-198-Lepid1
SIMOES-73-Dipte1
MLST-SGB-22-Lepid1
MLST-196-no_detail
MLST-211-no_detail
MLST-216-no_detail
MLST-208-no_detail
MLST-194-no_detail
MLST-229-no_detail
MLST-SGB-82-Dipte1
MLST-133-no_detail
SIMOES-23-Arane1-Hemip2-Hymen3-Lepid1-Odona1-Psoco1
MLST-SGB-9-Dipte1-Hymen1-Lepid1
MLST-219-no_detail
MLST-161-no_detail
MLST-150-no_detail
MLST-SGB-12-Lepid1
SIMOES-157-Thysa1
MLST-195-no_detail
MLST-SGmd-157-Hemip1
MLST-149-no_detail
MLST-SGB-201-Hymen1
MLST-210-no_detail
MLST-209-no_detail
SIMOES-106-Hemip1
MLST-206-no_detail
MLST-221-no_detail
MLST-192-no_detail
SIMOES-35-Coleo1-Hemip7


MLST-221-no_detail
MLST-192-no_detail
SIMOES-35-Coleo1-Hemip7
SIMOES-74-Dipte1
SIMOES-36-Coleo1
0.84
MLST-132-no_detail
0.92
SIMOES-37-Coleo1
MLST-180-no_detail
0.75
MLST-205-no_detail
MLST-212-no_detail
MLST-SGB-16-Ortho1
MLST-SGD-30-Spiru1

EN FRANCAIS

Dynamique des infections de *Wolbachia*: Quelles leçons nous donnent les communautés insulaires des arthropodes?:

Certains symbiotes intracellulaires résident dans le cytoplasme des cellules et manipulent le système reproductif de leurs hôtes. Du fait de leur transmission maternelle, ces parasites sont sélectionnés pour optimiser la survie et la reproduction de leur hôtes femelles. Chez les arthropodes, la bactérie *Wolbachia* infecte au moins 66% des espèces d'insectes mais peuvent aussi infecter des nématodes. Cette large distribution dans les populations hôtes confère à *Wolbachia* un potentiel important en tant que moteur d'évolution. En particulier, elle pourrait être utilisée comme vecteur transgène dans les espèces nuisibles. Mais la dynamique évolutive des infections à l'échelle des communautés est mal connue, en particulier la fréquence des transferts de parasites entre hôtes de différentes espèces et la stabilité évolutive des associations. Mon travail de thèse a porté sur la détection et dynamique des infections de *Wolbachia* à une échelle microevolutive, c'est-à-dire, dans des communautés d'arthropodes avec moins de 5 My. L'objectif de ce travail était à la fois la characterisation des communautés géographiques d'arthropodes et celle des infections par *Wolbachia* de ces communautés. Nous avons également examiné l'existence de transferts horizontaux récents de ces symbiotes entre des taxa distantes ainsi que les routes écologiques potentielles pour ces transmissions.

MOTS-CLEFS: parasitisme sexuel; transfers horizontaux; *Wolbachia*; incidence; communautés d'arthropodes.

IN ENGLISH

Dynamics of *Wolbachia* infections: what insights from Insular Arthropod communities?

Sexual parasites are intracellular symbionts capable of manipulating the reproduction of their hosts. They are widespread in Arthropods where they display a wide range of reproductive manipulations; these can be potentially involved in the evolution of mating systems, speciation, gene acquisition and sex determination. In particular, *Wolbachia* is thought to infect more than 66% of insect species and is also found in nematodes. However, little is known about the dynamics of *Wolbachia* infections at the community level. Although at the intra-population level, invasion dynamics have been extensively studied, the same is not true at the community level where the turnover of infections remains largely uncharacterised. The question of how often are new infections acquired through horizontal transfers between distantly related hosts remains also open. Moreover, as *Wolbachia* is seen as a good candidate for a transgenic vector against pests, understanding its dynamic at the community level is crucial. We proposed to address them by performing an exhaustive characterisation of sexual parasites in simplified systems, using the opportunity offered by small arthropod communities in isolated islands.

KEYWORDS: sexual parasitism; horizontal transfers; *Wolbachia*; incidence; arthropod communities.