THÈSE

Pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ DE POITIERS
UFR des sciences fondamentales et appliquées
Ecologie et biologie des interactions - EBI (Poitiers)
(Diplôme National - Arrêté du 7 août 2006)

École doctorale : Sciences pour l'environnement - Gay Lussac (La Rochelle)
Secteur de recherche : Biologie des organismes - Interactions symbiotiques

Présentée par :
Sandrine Geniez

Investigation of Wolbachia symbiosis in isopods and filarial nematodes by genomic and interactome studies

Directeur(s) de Thèse :
Pierre Grève, Barton Slatko

Soutenue le 27 septembre 2013 devant le jury

Jury :

Président	**Thierry Bergès**	Professeur des Universités, Université de Poitiers
Rapporteur	**Patrick Mavingui**	Directeur de recherche CNRS, Université de Lyon 1
Rapporteur	**Kenneth Pfarr**	Research Scientist, Universität Bonn, Deutschland
Membre	**Pierre Grève**	Professeur des Universités, Université de Poitiers
Membre	**Barton Slatko**	Senior Scientist, New England Biolabs, USA
Membre	**Elisabeth Herniou**	Chargé de recherche CNRS, Université de Tours

THÈSE

pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE POITIERS

FACULTE DES SCIENCES FONDAMENTALES ET APPLIQUEES

(Diplôme National – Arrêté du 7 août 2006)

École Doctorale **Sciences pour l'environnement Gay Lussac**

Spécialité : **Biologie des organismes – Interactions symbiotiques**

préparée au laboratoire **Écologie et Biologie des Interactions,**

équipe **Écologie Évolution Symbiose** – Poitiers – France

et à **New England Biolabs** – Ipswich – Etats-Unis

présentée le 27 septembre 2013 par

Sandrine GENIEZ

Investigation of *Wolbachia* Symbiosis in Isopods and Filarial Nematodes by Genomic and Interactome Studies

Directeurs de thèse : **Pierre GRÈVE**

Barton SLATKO

JURY

Dr. P. MAVINGUI	Directeur de recherche CNRS, Université de Lyon	*Rapporteur*
Dr. K. PFARR	Research Scientist, Université de Bonn	*Rapporteur*
Dr. E. HERNIOU	Chargée de recherche CNRS, Université de Tours	*Examinatrice*
Pr. T. BERGES	Professeur, Université de Poitiers	*Examinateur*
Pr. P. GREVE	Professeur, Université de Poitiers	*Examinateur*
Dr. B. SLATKO	Senior Scientist, New England Biolabs, Inc., Ipswich	*Examinateur*

Acknowledgments

First of all I would like to express my gratitude to my supervisors.
Dr. Barton Slatko, whose expertise and understanding added considerably to my experience, allowing me to develop my projects autonomously. I appreciate your vast knowledge and skills in many areas. I will never look at shellfishes the same way!
Pr. Pierre Grève qui a toujours su se montrer disponible même à distance et qui a su me montrer une autre facette de la recherche. J'ai eu grand plaisir à travailler avec toi, ta bonne humeur et des larges connaissances scientifiques ont toujours rendu nos discussions en franglais riches et intéressantes. Merci Pierre !

I would also like to thank Dr. Patrick Mavingui and Dr. Kenneth Pfarr for accepting to review this manuscript and Dr. Elisabeth Herniou and Pr. Thierry Bergès for being present for my PhD defense. Specialist's opinion is always valued.

Je souhaite aussi adresser un spécial merci au Pr. Didier Bouchon sans qui cette thèse quelque peu particulière de part sa bi-nationalité n'aurait pas été possible. Je tiens aussi à te remercier pour ton support statistique ainsi que pour nos discussions caféinées toujours très intéressantes.

I am also very grateful to Dr. Donald Comb and Jim Ellard for allowing me to perform my master degree internship and PhD studies in such a great environment and for their material support. I thank all the NEB family. This company is a real paradise and a well of knowledge and friendliness. Thanks for making me feel at home.

I'm also grateful to Dr. Tilde Carlow, Dr. George Tzertzinis and Dr. Christopher Noren for accepting to be part of my Ph. D committee at NEB and providing me so many helpful advices.

I also would like to thank all my colleagues, especially Dr. Jeremy Foster and the new commie Dr. Ashley Luck. Jeremy, you have been there since the beginning and your lab tricks and advices in and out of the lab have always been helpful. I can only hope that one day, I'll know the literature as you do! Ashley, even if we didn't work together for a long time, I really appreciate having you in the lab and I hope our roads will guide us to the same meetings ;). For the other members of the parasitology team, I'll remember the monthly division meetings where so many ideas arose. This team and all its members were wonderful, helpful and always supportive.
I also wish to thank the Sullivan lab at UC Santa Cruz for their hospitality and help with a particular thanks to Frederic Landmann. J'ai vraiment apprécié travailler à tes côtés et m'initier à ce monde merveilleux qu'est la biologie cellulaire.

I thank all the friends I made along the way in the US. I will always remember the wonderful time and parties we had at the gate house and the solidarity between all the frenchies. Sylvine, je te souhaite tout plein de bonheur pour ta nouvelle petite famille, je suis fière de pouvoir te compter parmi mes amis. Vanessa, je suis si heureuse pour toi et Jared, pour ce nouveau départ qui s'offre à vous, que je ne sais pas par où commencer. Je suis nostalgique de tous nos moments passés ensemble, de nos lunchs potins à nos rituels brunchs du dimanche midi. J'espère vraiment qu'on restera en contact malgré l'océan qui nous sépare. Alana, my dear, I really miss our animated discussion ;). I wish you all the happiness and I hope you'll finally decide to come visit us in France!
Thanks to my roommates and particularly to Renee and James. You guys are awesome. I really enjoyed living with you, we had so much fun and I really appreciated our long discussion about US history and politics comparison between France and US. Je n'oublierais pas de remercier ma seule french roomie « temporaire » en souvenir de nos vendredi-sushis ! ;) J'espère qu'on pourra se revoir bientôt maintenant qu'on est toutes les 2 rentrées au pays !

Merci également à l'ensemble de mes collègues français. Cela a toujours été un vrai plaisir de venir vous revoir durant ces 2 années à l'étranger. Je ne peux m'empêcher d'avoir une pensée particulière pour le staff technique, notamment Alexandra qui chouchoute si bien nos petites bêtes et Carine qui a bien mérité la bouteille de Champagne promise ;).Merci également à Joanne, pour le temps passé ensemble sur feu notre projet sur l'ER et pour ton soutien moral en toutes circonstances.

Il en va sans dire que je ne saurais oublier mes chers collègues doctorants, en particulier Gaël, Nicolas et Victorien qui ont si bien réussi à nous supporter sans passer par la fenêtre ;) Myriam et Bianca, merci pour ces très bons moments passés ensemble, et enfin Lise et Jessica, je suis très fière d'avoir pu partager ces 3 années avec vous, j'ai toujours eu grand plaisir à parler (en autres !) science avec vous et à taquiner ensemble notre Vico I national… J'ai vraiment apprécié cette entraide spontanée tout au long de nos thèses respectives.

Merci aux « Anonymous » pour votre soutien et votre bonne humeur ! Et, tout particulièrement à Bouziane, qui m'a initié (convertie ?) à la bioinformatique et a été d'un soutien sans faille durant ces derniers mois éprouvants, alors Bouziane un peu de bash pour toi :

```
[sgeniez@boubou03 ~]$ cd Anonymous_room/
[sgeniez@boubou03 ~]$ more memories
America
Beverly
France
DC.
Finistère
Abrest
Poitiers
Russie
Ibiza
Honolulu
Tunis
[sgeniez@boubou03 ~]$ awk '{print$2}' memories
[sgeniez@boubou03 ~]$ echo keep in touch !
```

J'ai tout de même une pensée toute paradoxale pour tous ceux que j'aurais oublié de remercier…

Mais il y en a un que je ne pourrais oublier de citer… Merci Fabien de m'avoir apporté tant de choses, pour ta patience et ton soutien, pour m'avoir encouragé à continuer dans les moments difficiles, et après ces longues années de distance, je ne peux que me réjouir d'être enfin à tes côtés.

Et enfin, je tiens à remercier tout particulièrement mes parents et Nicolas pour leur soutien sans faille tout au long de ma (longue ?) scolarité. Je n'en serais pas là aujourd'hui sans votre aide et votre support ; même si je sais que mes petites histoires de labo se rapprochent du chinois pour vous, vous avez toujours su m'accompagner et m'aider à prendre les bonnes décisions et directions….

Well, here I am.

RESUME

Les *Wolbachia* sont des alpha-proteobacteries présentes chez de nombreux arthropodes et nématodes filaires. Ces bactéries héritées maternellement induisent chez leurs hôtes des phénotypes allant du parasitisme au mutualisme, avec le long de ce continuum des phénotypes tels que la féminisation (F), l'incompatibilité cytoplasmique (IC) ou la mort des males. *Wolbachia* est ainsi un modèle particulièrement intéressant pour étudier les différents types de relations symbiotiques.

Chez *Brugia malayi*, comme pour les autres nématodes filaires, *Wolbachia* vit en symbiose obligatoire avec son hôte. L'élimination de la bactérie par des traitements antibiotiques entraîne une perte de fertilité voire la mort du nématode. Chez l'isopode terrestre *Armadillidium vulgare, Wolbachia* induit la féminisation des mâles génétiques en femelles fonctionnelles entraînant des biais de sex-ratio vers les femelles dans la descendance.

Pour comprendre les mécanismes impliqués dans ces deux symbioses, nous avons mis au point une nouvelle méthode de capture pour isoler l'ADN de *Wolbachia* et séquencer 7 souches de *Wolbachia* d'isopodes (F et IC). Une étude de génomique comparative a permis d'établir un premier pan-génome des bactéries du genre *Wolbachia* et d'identifier 2, 5 et 3 gènes présents seulement chez les souches mutualistes, féminisantes ou induisant la mort des males. L'expression des gènes potentiellement impliqués dans la féminisation ou le mutualisme a été étudiée au cours du développement de l'hôte. L'étude de l'interactome protéique bactérie-hôte a ensuite été initiée en utilisant comme appât des protéines bactériennes à domaines eucaryotes en vue d'identifier les cibles de *Wolbachia* chez l'hôte.

ABSTRACT

Bacteria of the genus *Wolbachia* are gram-negative alpha-proteobacteria present in many arthropods and filarial nematodes. These obligate intracellular bacteria are maternally inherited and induce a large number of phenotypes across the symbiosis continuum from mutualism to parasitism, including feminization (F), cytoplasmic incompatibility (CI) or male killing. Studying *Wolbachia* symbioses is therefore of particular interest in the investigation of symbiotic relationships.

In *Brugia malayi* and other filarial nematodes, they are obligate, leading to a loss of worm fertility, and eventual death upon their depletion with certain antibiotics. In arthropods, they are parasitic. In the isopod crustacean *Armadillidium vulgare* they cause feminization when present: genetic males develop as functional females leading to female biased sex-ratio progenies.

In order to understand the molecular mechanisms of these two symbioses, we set up a new procedure to capture *Wolbachia* DNA and perform whole-genome sequencing on 7 *Wolbachia* strains, symbionts of isopods (F & CI). Comparative genomics led to the establishment of the *Wolbachia* pan-genome as well as the identification of phenotype related gene patterns. We identified 2, 5 and 3 protein-coding sequences that are only found in mutualist, feminizing and male killing strains, respectively. Expression of genes potentially involved in feminization and mutualism were also analyzed throughout host post-embryonic development. A host-symbiont interactome approach was then initiated by protein-protein interaction studies using bacterial proteins with eukaryote like motifs as bait in order to identify *Wolbachia* host targets involved in symbiosis.

TABLE OF CONTENTS

TABLE OF ILLUSTRATIONS

Abbreviations

µg – Microgram

AA – Amino Acid

AH – Androgenic Hormone

ANK – Ankyrin

bp – Base pair

BLAST - Basic Local Alignment Search Tool

BLAT – BLAST Like Alignment Tool

BSA – Bovine Serum Albumin

cDNA – complementary DNA

CDS - Coding DNA Sequence

CI – Cytoplasmic Incompatibility

DMSO – Dimethyl sulphoxide

DNA – Deoxyribonucleic acid

DTT - Dithiothreitol

dNTP – deoxynucleoside triphosphate

EDTA – Ethylene Diamine Tetra Acetic acid

ELISA – Enzymes Linked ImmunoSorbent Assay

e.g. – Example

IS – Insertion Element

IPTG – IsoPropyl β-D-ThioGalactosidase

Kb – Kilobase

kDa – KiloDalton

LB – Luria Bertani medium

LF – Lymphatic Filariasis

MF – MicroFilariae

min – Minutes

mL – Millilitre

mM – MilliMolar

mRNA – Messenger RNA

mtRNA – Mitochondrial RNA

Mb – Megabase

MW – Molecular Weight

NGS – Next-Generation Sequencing

nm – Nanometer

OD – Optical Density

o/n – Overnight

ORF – Open Reading Frame

PAGE – PolyAcrylamide Gel Electrophoresis

PBS – Phosphate Buffered Saline

PBS-T – Phosphate Buffered Saline Tween

PCR – Polymerase Chain Reaction

RNA – Ribonucleic Acid

rRNA – Ribosomal RNA

qPCR – Quantitative Polymerase Chain Reaction

RT – Reverse Transcription

SDS – Sodium Dodecyl Sulfate

SDS-PAGE – Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis

sRNA – Small RNA

TBS – Tris Buffered Saline

T1SS – Type I Secretion System

T4SS – Type IV Secretion System

tmRNA – Transfer-messenger RNA

tRNA – Transfer RNA

WHO – World Health Organization

wsp – *Wolbachia* surface protein

X-gal – 5-bromo-4-chloro-3-indolyl-β-D-galactosidase

Glossary

Annotation – Adding biological information (structure and/or function) to genome sequence.

Assembly – In terms of genomics, the computational process of deciphering the sequence composition of a genome, using short sequences (reads) derived from different portions of input target DNA.

Bacteriophage – A virus that infects bacteria. A bacteriophage can either lyse a cell (known as the virulent phase of life cycle) or integrates its genome into host genome (known as temperate phase of life cycle).

Base coverage – The total number of nucleotides sequenced and aligned at a given reference position in the genomic sequence.

Basic Local Alignment Search Tool – (BLAST) An NCBI (National Center for Biotechnology Information, NIH, USA) bio-informatics homology search tool that searches the GENBANK sequence databases and reports the entries most similar to a query sequence (nucleotides or proteins).

BLAST Like Alignment Tool – An alignment tool like BLAST, but with a different structure in which the target is not GENBANK sequences (BLAST) but an index derived from the assembly of the entire genome.

Bootstrap analysis – A statistical analysis method used to assess the robustness of reconstructed phylogenic trees by assigning probability-like value for tree branches.

Cytoplasmic incompatibility - A post-fertilization defect in chromatin where mating between infected males and *Wolbachia*-free females (or females harboring a different *Wolbachia* strain than that in the male) result in high levels of embryonic death.

Coding DNA sequence – Genomic sequence that encodes for a protein sequence

Contigs – Contiguous sequences - A set of overlapping cloned or sequenced DNA segments derived from a single genetic source, which can be used to deduce the original DNA sequence of the source.

Core-genome – The pool of all genes shared by all the members of a coherent group of organisms (e.g. all the strains from a same bacterial species).

Dispensable-genome – The set of genes present in some, but not all, strains of the same bacterial species, implying they are not required for the species survival.

Genome coverage – The percentage of the reference genome that is successfully mapped after alignment.

Homologs - Genes with a relationship that arose from having the same ancestral origin.

Horizontal gene transfer – Transfer of genetic materials between evolutionary distinct organisms by non-sexual or non-asexual mechanisms.

Insertion sequence – The simplest type of transposable element in bacteria. It contains only the genes required for its own transposition.

Mapping – Alignment of raw sequencing reads on a reference sequence.

Mate pair data – Data from a pair of reads sequenced from the same circularized DNA fragment. The circularization steps allows for larger fragments to be used.

Male killing – the selective killing of male embryos.

Next-generation sequencing – A set of novel approaches to DNA sequencing that dispenses with the need to create libraries of cloned sequences in bacteria and holds the promise of providing faster and cheaper sequencing.

N50 – In computational biology, the N50 length is defined as the length for which the collection of all contigs of that length or longer contains at least half of the total of the lengths of the contig.

Orthologs – Genes with a relationship that arose from a speciation event.

Pan-genome – The global gene repertoire of a bacterial species: core-genome + dispensable-genome + strain specific genes.

Paired-end data – Data from a pair of reads sequenced from ends of the same DNA fragment. The genomic distance between the reads is approximately known and is used to constrain assembly solutions.

Paralogs – Genes with a relationship that arose from a duplication event.

Parthenogenesis – Asexual reproduction, embryos develop from unfertilized eggs.

Phylogenomics – Phylogeny at the whole-genome level.

Prophage – A bacterial virus residing in bacterial host in a passive state, not expressing many viral functions of producing progeny virions. It may be integrated into the bacterial DNA or free in the cytoplasm.

Pseudogene – Relic of an ancient functional gene that is no longer functional

Pyrosequencing – Sequencing DNA by observing the release of a pyrophosphate following incorporation of the next nucleotide into the growing chain.

Scaffold – A series of contigs that are in the right order but not necessarily connected in one contiguous stretch of sequence.

Synteny - The property of genes of being found on the same chromosome.

Transposable element - DNA sequence that can change its position within the genome, sometimes creating or reversing mutations.

Vertical gene transfer – Process in which genetic material is passed from parent to offspring or, more generally, from ancestor to descendent.

GENERAL INTRODUCTION

Symbiosis (biological) defines a close interaction between a least two partners, a host and a symbiont. This relationship definition was first used in 1877 by Albert Bernard Frank (Frank 1877) to describe the mutualistic symbiosis in lichens and clearly established in 1879 by Heinrich Anton de Bary (de Bary 1879) as "the living together of unlike organisms" (Richardson 1999). Currently, many types of symbiosis have been described and numerous scientists defend the theory that virtually every organism is in a symbiotic relationship with another one (Combes 1995; Thomas et al. 2007).

These symbiotic relationships can be divided into three main categories: mutualism, when both host and symbiont benefit from the relationship and depend on each other for survival; commensalism, when one partner benefits without affecting the other; and parasitism, when one partner benefits by affecting or harming the other (Combes 1995). The physical interaction between the host and the symbiont can be external where the symbiont lives on the host (ectosymbiosis or exosymbiosis); or internal, where the symbiont lives within the tissues of the host (endosymbiosis).

Symbiosis is a dynamic relationship: depending on the environment and the needs of the symbiotic partner, the symbiosis can be present along a continuum of mutualism to parasitism (Thrall et al. 2007; Fellous and Salvaudon 2009; Leigh 2010). Thus, symbiosis constitutes a major evolutionary force allowing many adaptations of the partners.

Particular attention has been paid to intra-cytoplasmic microorganisms because their survival and reproduction depend directly on the survival and reproductive success of their hosts (Bull and Rice 1991). This evolutionary concept was originally described by Van Valen in 1973 in response to the "Law of Extinction" and it is popularly known as the "Red Queen" hypothesis. This hypothesis proposed that species have to co-evolve with their predators and co-adapt in order to ensure their survival. Van Valen used the Red Queen, from Lewis Carroll's "through the Looking Glass" (Carroll 1871) to support his view that a species must keep evolving to keep up with its competitors and enemies who are also evolving. According to Carrols' Red Queen "… it takes all the running you can do, to stay in the same place. If you want to get somewhere else, you must run at least twice as fast as that!". Van Valen first described an "arms race" where two populations co-evolve in order to overcome each other's ability to overcome each other (VanValen 1973). This theory can be used to explain pathogen resistance and host-adaptation. Pathogens develop physical and molecular adaptations to escape host

immune system such as excretion of virulence factors. These bacterial factors, mostly proteins containing eukaryote-like motifs that mimic host proteins, are secreted via bacterial secretion systems (Durand et al. 2009). Stable symbiosis results from an equal reaction from the host immune systems and the bacterial virulence factors.
Intimate symbiotic relationships with symbionts as endocellular bacteria represent an important source of genetic variation to the host inducing major evolutionary adaptations (Fraune 2010, Gilbert et al 2010, McFall-Ngai et al 2013, Moran 2007)

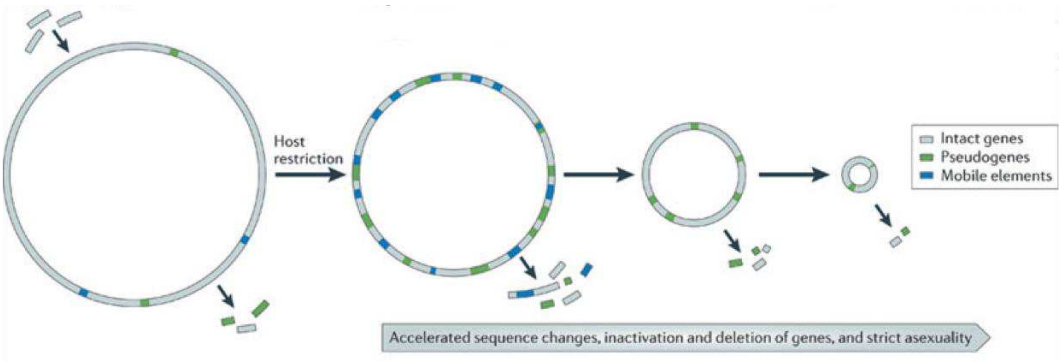


Figure 1: Genome reduction steps in host-restricted bacteria (McCutcheon 2012)

The development of the eukaryotic cell itself has been hypothesized as likely originated from endosymbioses between several free-living unicellular organisms. Sanger described in 1967 the endosymbiotic theory of cell evolution in which cell organelles such as chloroplasts and mitochondria result from the incorporation of photosynthetic and aerobic ancestors of existing Cyanobacteria and Alphaproteobacteria.

Similarly, the majority of mutualistic endosymbionts of insects have drastically reduced genome sizes (Figure 1) (McCutcheon 2012). This genome reduction has been hypothesized to result from a long time-scale host-symbiont co-evolution, highlighting endosymbiont adaptation to their host-dependent lifestyle. Accumulation of small deletions known as Muller's ratchet process ie accumulation of deleterious mutations in an irreversible manner leads to reduced-size genomes containing mainly the named "most essential genes". These so-called genes most frequently represent functions that are essential by the host such as genes related to nutrient provisioning (Akman et al 2002, McCutcheon 2012, Moran 2005).

As for symbionts, they ensure their proliferation, by taking advantage of their host reproduction success for their own transmission. Thus, any actions that tend to favor

host reproduction will be adaptive for the symbionts. This is particularly applicable for vertical transmission where symbionts invade the next generation by invading oocytes. This situation is well illustrated with *Wolbachia* species, which induce a variety of phenotypes to their host through the whole symbiosis continuum from parasitism to mutualism, but always favoring bacterial (symbiont) proliferation.

I. *Wolbachia*

A. *Wolbachia* biology

Bacteria of the genus *Wolbachia* are gram-negative alpha-proteobacteria present in many invertebrates such as insects, crustaceans, nematodes, mites, scorpions or spiders (Werren and Jaenike 1995; Werren et al. 2008). These intracellular bacteria were first reported within the reproductive tissues of the mosquito *Culex pipiens* by Hertig & Wolbach in 1924 (Hertig and Wolbach 1924) and these rickettsiae were later named *Wolbachia pipientis* (Hertig 1936). In 1970's, Yen & Barr established the association between the phenomenon of cytoplasmic incompatibly (CI) in *C. pipiens* and the presence of the rickettsial bacteria agent, which was eliminated through antibiotic curing (Yen and Barr 1971).

The advent of molecular techniques, in particular the polymerase chain reaction (PCR) rapidly accelerated intensive research on *Wolbachia* in the 1990s. Phylogenic studies using the 16S rDNA locus revealed that this alpha-proteobacteria is closely related to intracellular bacteria including *Ehrlichia, Anaplasma, Rickettsia and Neorickettsia* (Breeuwer 1992; O'Neill et al. 1992; Rousset 1992). *Wolbachia* is in the *Rickettsiales*, the earliest-branching of the alpha-proteobacterial orders (Williams et al. 2007) (Figure 2). *Wolbachia* are extremely common in arthropods and parasitic filarial nematodes including most of the agents of filariasis (Sironi et al. 1995; Bouchon et al. 1998; Taylor et al. 2005; Bouchon et al. 2008; Werren et al. 2008; Cordaux et al. 2012). It has been recently estimated that ~40 % of terrestrial arthropod species are infected with *Wolbachia* (Zug and Hammerstein 2012), making it the most widespread intracellular bacterial species. New *Wolbachia*-infected species are continually described within arthropods, as well as in nematodes as illustrated by the plant parasitic nematode *Radopholus similis* (Haegeman et al. 2009) showing that the estimated widespread of *Wolbachia* is probably underestimated.

Although horizontal transfers of *Wolbachia* can occur between different arthropod hosts (Heath et al. 1999; Schuler et al. 2013) and that recent studies showed that cannibalism of infested isopods leads to *Wolbachia* transmission (Le Clec'h et al. 2013a), vertical transmission from ascendants to descendants is the most common transmission path for *Wolbachia*, mainly via maternal inheritance.

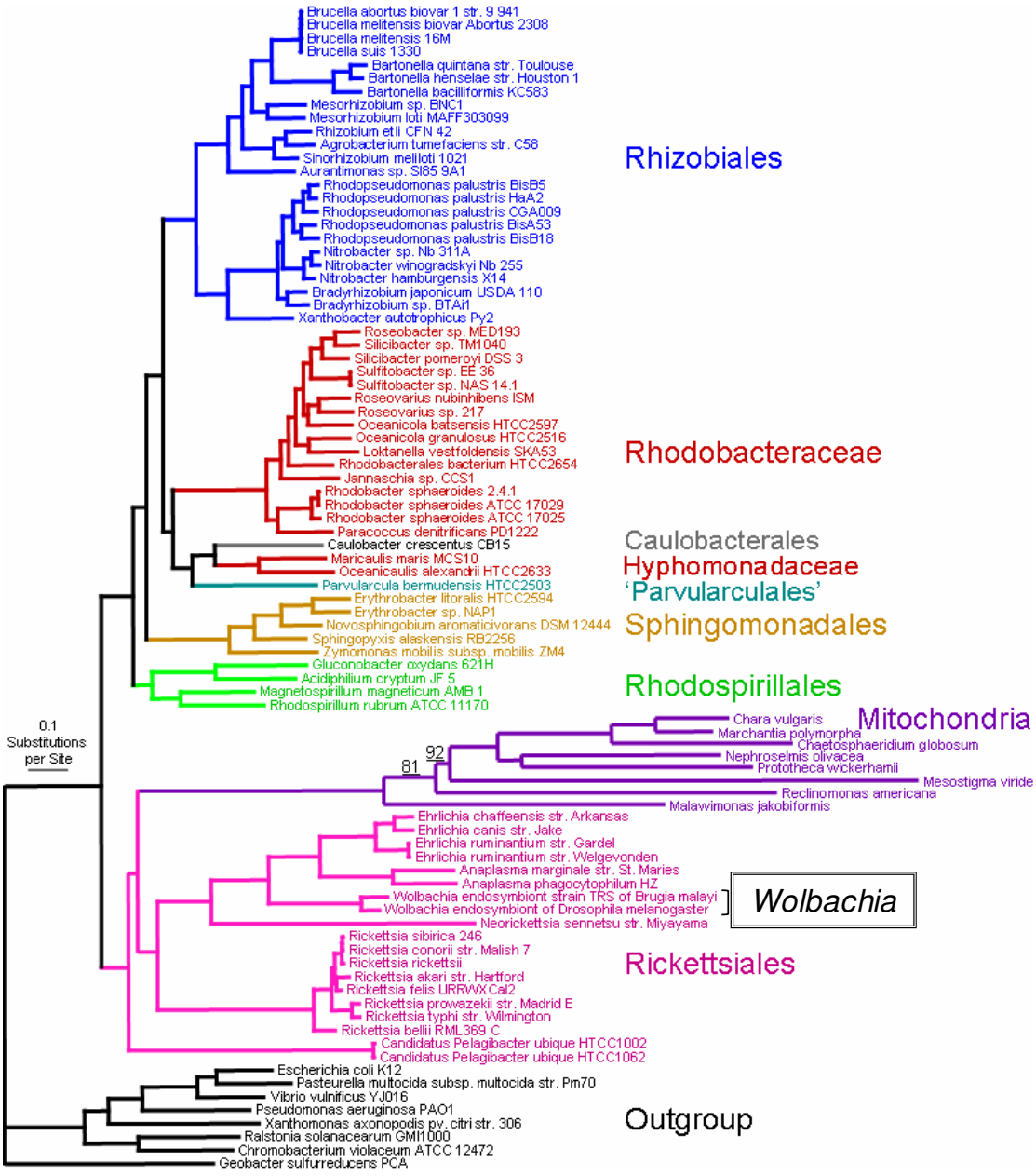


Figure 2: Phylogenic tree of the alpha-proteobacteria (Williams et al. 2007).

Thus, they are commonly more densely found in the reproductive tissues even if their host location depends on the host species.

In filarial nematodes such as *Brugia malayi, B. pahangi, Wuchereria bancrofti, Onchocerca volvulus,* at least one *Mansonella sp., Dirofilaria immitis and Litomosoides sigmodontis*, *Wolbachia* are located in the lateral cords of male and female worms as well as in oocytes but not in the male reproductive system. They are occasionally found extracellularly, close to the ovarian tissue (Landmann et al. 2010; Fischer et al. 2011). In insects, *Wolbachia* localization can be restricted to ovaries as well as being found in other tissues such as gut and fat tissues, salivary glands, immune system, nervous and muscular system and reproductive tissues (Figure 3) (Dobson et al. 1999; Ijichi et al. 2002; Saridaki and Bourtzis 2010). In the leaf-cutting ants *Acromyrmex octospinosus, Wolbachia* are found in fat body cells, gut tissue, muscle fibers but also extracellularly in the crop part of the gut (Andersen et al. 2012).

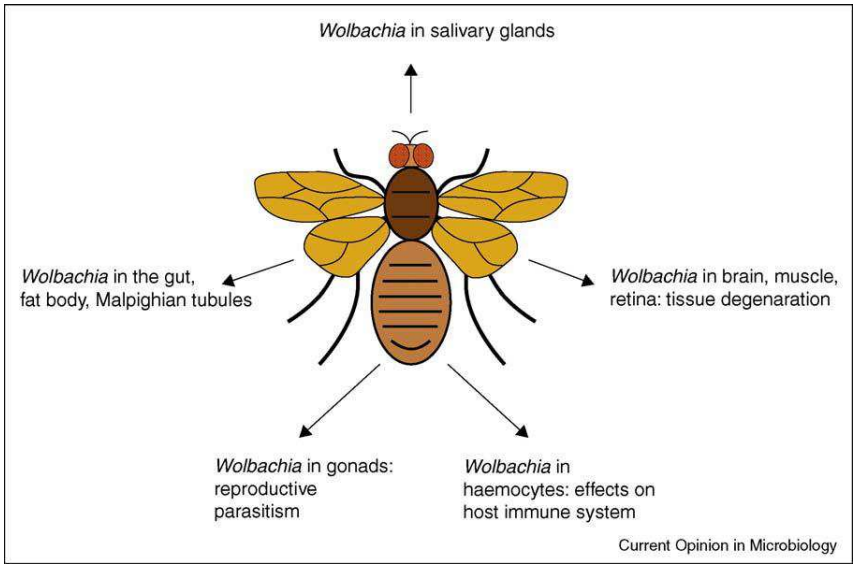


Figure 3: *Wolbachia* distribution in insect tissues and effects on host biology. (Saridaki and Bourtzis 2010)

In Isopod crustaceans, *Wolbachia* are found in all tissues investigated such as haemocytes and hematopoietic organs, digestive and muscular tissues and fat body with a higher density in the central nervous system and ovaries (Chevalier et al. 2011; Le Clec'h et al. 2012; Le Clec'h et al. 2013b; Genty et al. 2013) Dittmer et al. 2013, in prep). Immunostainning of *Armadillidium vulgare* ovaries localized *Wolbachia* in follicular cells as well as around the oocyte nucleus (Figure 4).

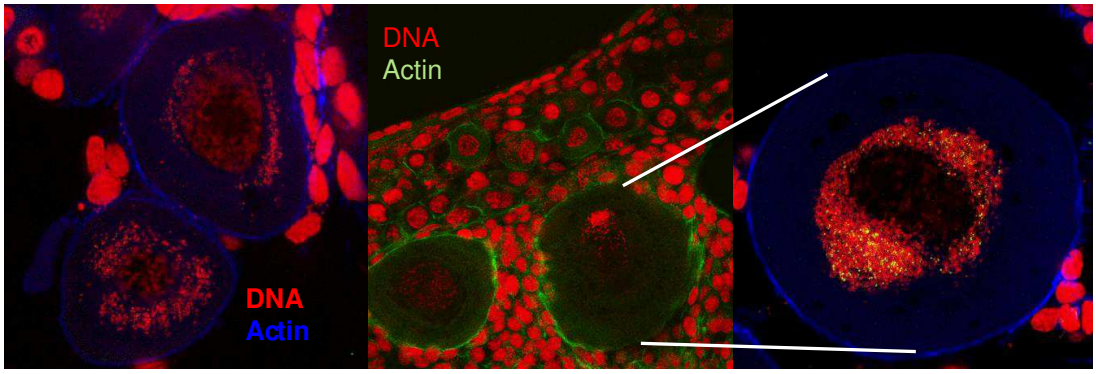


Figure 4: *Wolbachia* segregation in the ovaries. *Wolbachia* are concentrated in follicular cells and around the nucleus of the oocytes. Merges of confocal stack of *A. vulgare* ovaries stained for DNA (propidium iodide, red), *Wolbachia* (anti-hsp60, green in the rightest picture) and cortical actin (blue or green), in collaboration with F. Landmann (University of California, Santa Cruz).

B. Morphology and phylogeny of *Wolbachia*

Wolbachia have general morphological characteristics of *rickettsiae* (Hertig 1936). They are dimorphic, with irregularly formed rod like (0.5-1.3 µm in length) and coccoid forms (0.25-0.5 µm in diameter) (Kozek 2005). Usually, *Wolbachia* is present in a vacuole enveloped by three membrane layers. The outer layer is of host origin followed by the outer membrane of the bacteria, while the innermost layer consists of the plasma membrane of the bacteria (Martin et al. 1973; Fischer et al. 2011) (Figure 5). One of the most probable hypotheses is that the endobacteria could be inside the endoplasmic reticulum (Martin et al. 1973; Cho et al. 2011), an observation described for many intracellular bacteria. Wright proposed that this membrane may explain how the host controls the prokaryote (Wright 1979). More recently, in whiteflies, the presence of bacteriocytes, a type of differentiated host cell, where different bacterial species are grouped in the same insect cell has been illustrated. This intracellular bacterial ecosystem grouped in bacteriocytes establishes a stable regulation among the bacteria whose density may be regulated by the host cell or host bacteriophages but persist throughout the host life cycle via vertical transmission from mother to offspring (Gottlieb et al. 2008; Login et al. 2011; Login and Heddi 2013).

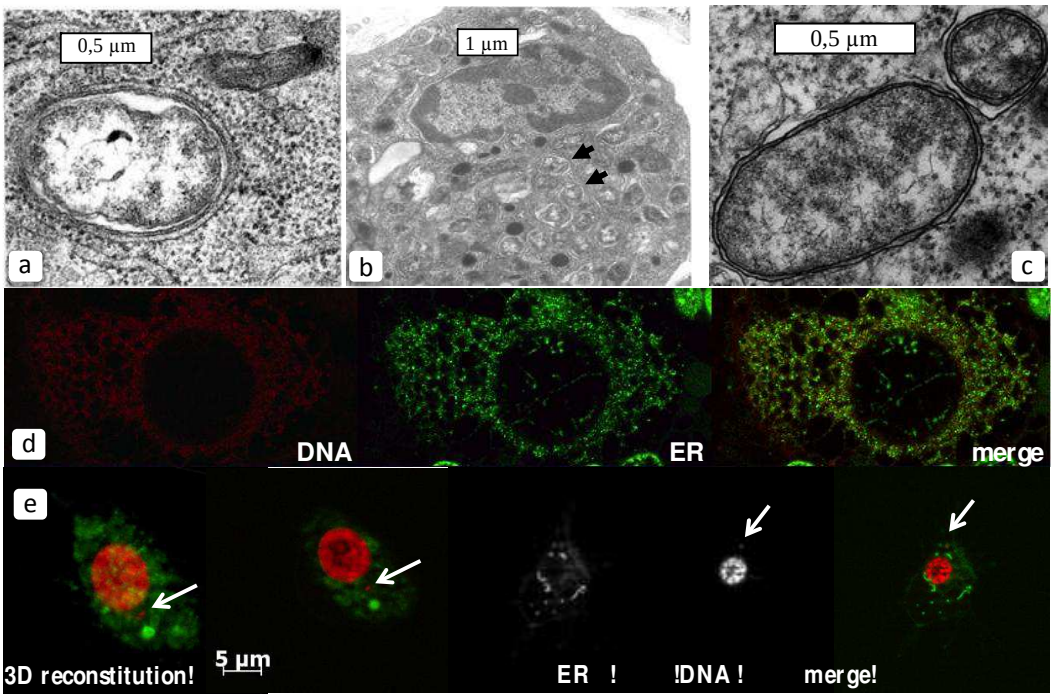


Figure 5: *Wolbachia* tissue localization in *A. vulgare* tissues. a, b, c: Transmission electron micrographs of *Wolbachia* within the androgenic gland, hemocytes and oocytes, respectively (Source: EES); d: Confocal stacks of oocytes stained for DNA (syto11, red) and endoplasmic reticulum (ER tracker, green), in collaboration with F. Landmann (University of California, Santa Cruz); e: Confocal stacks of haemocytes stained for DNA (syto11, red) and endoplasmic reticulum (ER tracker, green), in collaboration with L. Genty and J. Bertaux (EES). Arrows indicate *Wolbachia*.

Phylogenetic analyses of *Wolbachia* strains using single-gene or multilocus sequence typing (MLST) currently describe seven well-resolved *Wolbachia* groups, designated as supergroup lineages A to H, without a group G, along with a number of additional lineages (Lo et al. 2002; Casiraghi et al. 2003; Casiraghi et al. 2005; Lo and Evans 2007; Baldo and Werren 2007; Bordenstein et al. 2009; Haegeman et al. 2009)(Figure 6).

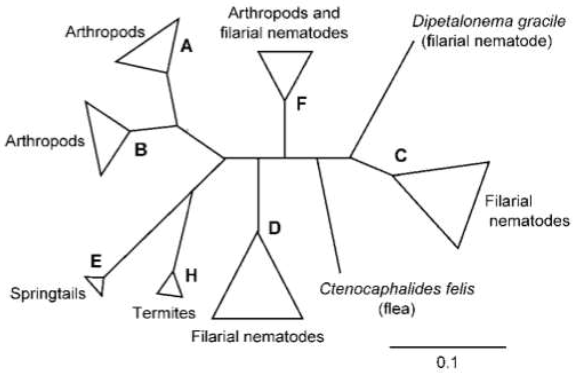


Figure 6: Schematic diagram of *W. pipientis* phylogeny supergroups based on various phylogenetic studies of the genes ftsZ, groEL, gltA and dnaA (modified from (Lo and Evans 2007). Host species are indicated next to each supergroup. Bar, 0.1 substitutions per site (an approximation based on a concatenated gene analysis of these four genes).

Supergroups A and B include most of the parasitic *Wolbachia* so far found in arthropods and supergroups C and D include the majority of the *Wolbachia* found in filarial nematodes although some nematodes have been found to lack *Wolbachia* (Bordenstein et al. 2003). The E supergroup consists of *Wolbachia* from primitive wingless insects, the springtails. The supergroup F contains *Wolbachia* bacteria of arthropods (termites) and the filarial parasite *M. ozzardi* and may be an intermediate group containing both arthropod and nematode *Wolbachia*. In addition, supergroup H encompasses *Wolbachia*, different from those in supergroup F, in termites. The *Wolbachia* from the filarial nematode *Dipetalonema gracile* and from the arthropod *Ctenocephalides felis* have not been designated to any of the existing supergroups (Gorham et al. 2003; Dittmar and Whiting 2004; Casiraghi et al. 2005). Based on a *wsp* gene phylogenetic analysis, several *Wolbachia* strains were recently identified in the flea *C. felis* belonging to either supergroup B or F (Tay 2013).

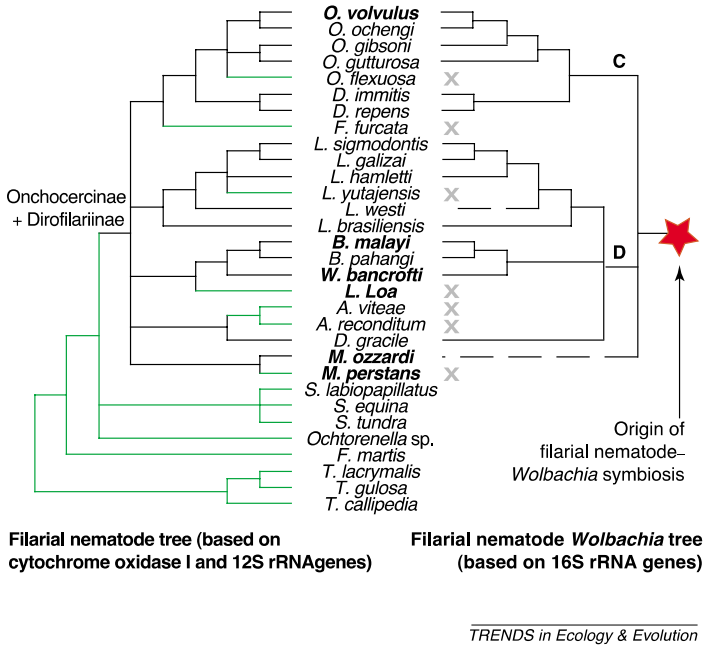


Figure 7: Filarial nematode and filarial nematode *Wolbachia* trees based on *cytochrome oxidase I* and *12S rRNA* genes and *16S RNA* genes, respectively (Fenn and Blaxter 2004a).

As opposed to arthropods where several horizontal transfers have been deduced from the absence of congruence between host and symbiont phylogenies (O'Neill et al. 1992; Vavre et al. 1999a; Huigens et al. 2004; Baldo et al. 2008; Schuler et al. 2013), *Wolbachia* is fixed within a nematode species (Bandi et al. 1998). Using *12S rRNA* and *cytochrome oxidase I (COI)* genes for host phylogeny and *16S rRNA* gene for *Wolbachia* phylogeny, a

close co-evolution between filarial nematodes and their endosymbiotic *Wolbachia* has been demonstrated (Figure 7) (Casiraghi et al. 2001; Casiraghi et al. 2004; Fenn and Blaxter 2004a).

Despite these evidences of a high diversity in the *Wolbachia* tree, all these bacteria have evolved to increase their spreading throughout host populations. *Wolbachia* use different mutualistic or parasitic host manipulations including cytoplasmic incompatibility, parthenogenesis, male killing or feminization. By favoring infected females in host populations, the scale of bacterial proliferation to next generations is enlarged.

C. Phenotypes associated with *Wolbachia* infection

In some cases, the presence of *Wolbachia* can be correlated with host survival or fecundity increase (Min and Benzer 1997; Vavre et al. 1999b; Dobson et al. 2004; Fry et al. 2004). The *Wolbachia* symbiosis can also evolve depending on environmental conditions. If *Wolbachia* has no effect on *Drosophila melanogaster* fecundity when reared with cornmeal diets, it provides a significant fecundity benefit to females when subjected to low or high iron environments; in nutritional stress, *Wolbachia* conferred a compensatory effect maintaining a normal reproduction level (Brownlie et al. 2009).

Similarly, in insects, *Wolbachia* infected populations are less susceptible to mortality induced by a range of viruses (Hedges et al. 2008; Teixeira et al. 2008; Osborne et al. 2009; Bian et al. 2010), inviting scientists to investigate *Wolbachia* symbiosis as a natural agricultural pest and disease vector control such as Dengue fever or Chikungunya infections (Moreira et al. 2009; Kambris et al. 2009; Saridaki and Bourtzis 2010; Hoffmann et al. 2011; Cook and McGraw 2010; Bian et al. 2013a; Bian et al. 2013b).

To date, the best-documented phenotypes associated with *Wolbachia* infection are related to the mutualism relationships and four facultative associations: cytoplasmic incompatibility, parthenogenesis, male killing and feminization. Even though there are other endosymbionts that induce some of these phenotypes, *Wolbachia* is the only known bacteria that induce this diversity of phenotype (Duron et al. 2008).

1 - *Wolbachia*, as an obligate mutualist

Wolbachia are found in 100% of filarial nematodes of species in which they are present, suggesting the importance of the symbiont in the host fertility and survival (Bandi et al. 1998; Taylor et al. 2000; Hoerauf et al. 1999; Taylor et al. 2005). In nematodes, attempts to remove them and maintain fertility and survival have failed (B. Slatko, personal communication). In rodent and human filarial nematodes, antibiotics such as tetracycline, doxycycline, rifampicin and azithromycin lead to *Wolbachia* depletion in the female reproductive tract, which eventually leads to degeneration of embryos. Detrimental effects on parasite growth development, fertility and viability induced by reduced levels of *Wolbachia* attest to the mutualistic relationship (Genchi et al. 1998; Bandi et al. 1999; Hoerauf et al. 2000a; Hoerauf et al. 2000b; Rao et al. 2002; Rao and Well 2002; Rajan 2004).

This mutualistic relationship is based on equilibrium between *Wolbachia* effects and host responses. *Wolbachia* presence activates host nematode autophagy, which regulates bacterial loads (Taylor et al. 2012; Voronin et al. 2012)

More recently, an obligate nutritional mutualist *Wolbachia* strain was identified in the bedbug *Cimex lectularius.* Elimination of this *Wolbachia* symbiont induced a delayed growth and sterility of the host, depression that was reversed by supplementation of B vitamins (Hosokawa et al. 2010; Li and Carlow 2012). This nutritional mutualism takes a particular place in the symbiosis evolution continuum between mutualism and parasitism. The symbiont is not essential for host survival but essential to host population: host reproduction is dependent on the bacteria presence. Another described strain provides an intermediate symbiosis pattern, at the intersection between mutualism and parasitism, the *Wolbachia* strain *w*Atab3 of the parasitic wasp *Asobara tabida Nees*. Aposymbiotic females obtained after antibiotic treatment were incapable of producing mature oocytes. The presence of *Wolbachia* became essential for host oocyte maturation by inhibiting programmed cell death processes. By regulating host apoptosis, *Wolbachia* evolved in a dependence relationship in-between parasitism and mutualism (Pannebakker et al. 2007). Thus, the bacterium is not necessary for females' survival, nor for males' fertility but without the bacteria, this species of wasp are not capable of reproduction, which has become symbiont dependent: this strain of *Wolbachia* is necessary for oogenesis (Dedeine et al. 2001; Dedeine et al. 2005; Kremer et al. 2010).

2 - *Wolbachia*, a reproduction manipulator

As *Wolbachia* is mainly maternally transmitted, it induces four main effects on host reproduction (CI, parthenogenesis, male killing and feminization) that favor infected females and thus increases its proliferation to next generations.

a) Cytoplasmic incompatibility

CI is the most phylogenetically diverse and studied *Wolbachia*-induced phenotype. CI is characterized by an embryonic mortality observed when infected males mate with uninfected females (unidirectional CI) or females infected with an incompatible *Wolbachia* strains (Bidirectional CI) (Tram and Sullivan 2002; Engelstadter and Telschow 2009) (Figure 8). While the evolutionary processes leading to the emergence of these incompatibilities are not known yet, it has been shown that the intensity of the CI response, from 50 to 100% is proportional to the bacterial load and male age (Charlat et al. 2001).

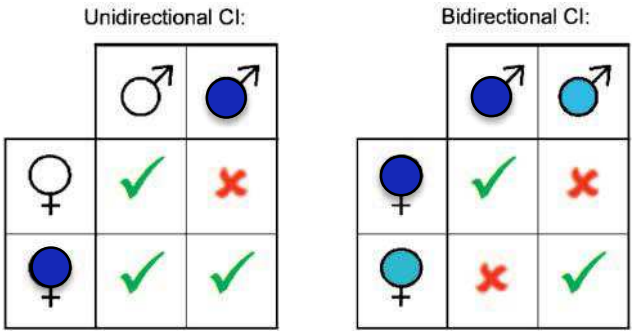


Figure 8: Uni- and bidirectional cytoplasmic incompatibility (CI). The two tables show success (green tick marks) or failure (red crosses) of offspring production of crosses between parents with different infection states. Empty symbols in the parent generation denote that these parents are uninfected, whereas the two shades of blue denote infection with two different strains of bacteria. With unidirectional CI, only crosses between infected males and uninfected females are incompatible. With bidirectional CI, crosses between males and females infected with different strains of CI-inducing bacteria are incompatible (Engelstadter and Telschow 2009).

This reproductive incompatibility has been mainly demonstrated in insects (Giordano et al. 1997; O'Neill et al. 1997; Bourtzis 1996) but also in three species of terrestrial crustaceans: *Porcellio dilatatus petiti* (Legrand et al. 1978), *P. d. dilatatus* (Sicard et al. submitted) and *Cylisticus convexus* (Moret et al. 2001).

This *Wolbachia*-induced phenotype has recently been used to exploit a natural way to control mosquitos' invasion particularly those hosting the Dengue disease or filarial nematodes. Releasing male mosquitos infected by a non-compatible *Wolbachia* strain

has been proven efficient to reduce the total number of mosquitos. O'Neill's lab showed significant results from the stable introduction of the *w*Mel *Wolbachia* strain from *D. melanogaster* to two natural mosquitos populations in Australia by releasing *w*Mel-infected *A. aegypti* adults. Because these *Wolbachia* infected populations became predominant due to the *Wolbachia*-induced CI and because *Wolbachia* is inhibiting the Dengue virus transmission, introductions of these infected population act as natural pest control. Dobson's lab used the same strategy at Maupiti in French polynesia to reduce the transmission of filarial worms by mosquitos, parasite responsible of lymphatic filariasis, by releasing males of a novel *Aedes polynesiensis* strain infected with an exogenous *Wolbachia* strain (Brelsfoard et al. 2009; Moreira et al. 2009; Popovici et al. 2010; Hoffmann et al. 2011; Rances et al. 2012; Baton et al. 2013; Chambers et al. 2011; O'Connor et al. 2012).

b) Parthenogenesis

Considering *Wolbachia* is mainly maternally transmitted, another strategy of host manipulation is to induce parthenogenesis, asexual reproduction resulting in all female broods. *Wolbachia*-induced parthenogenesis is currently known to be present in three arthropod orders: *Thysanoptera*, *Acari* and *Hymenoptera* (Stouthamer 1993; Bourtzis 2003; Kremer et al. 2009). For these arthropods, sex is usually regulated by the ploidy of the embryos; fertilized eggs (diploid) mature into female whereas unfertilized eggs (haploid) mature into male (Stouthamer 1993) (Figure 9).

Wolbachia-induced thelytokous parthenogenesis leads to the doubling of chromosome number of the unfertilized eggs. These diploid eggs mature into females, increasing the widespread of the endosymbiont. As a result, infected females produce up to twice the number of female as uninfected females (Schilthuizen and Stouthamer 1997).

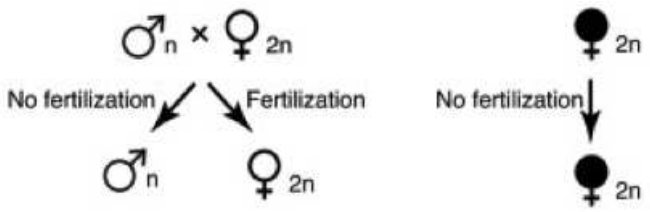


Figure 9: Parthenogenesis induction, characterized by sex-ratio distortion towards females through conversion of genetic (haploid) males into genetic (diploid) females. Black (white) coloration: individual carries (does not carry) endosymbionts. n/2n, haploid/diploid status of individual (Cordaux et al. 2011).

c) Male-killing

The *Wolbachia*-induced phenotype of male killing has been reported in four different arthropods orders: Diptera, Coleoptera, Lepidoptera and pseudoscorpiones (Hurst and Jiggins 2000; Jiggins et al. 2000). Male killing results in a selective killing of males due by *Wolbachia* (Figure 10). This adaptation to maternal transmission causes a sex-ratio distortion and a reduction in number of progeny, which give infected females a higher probability of survival and thus endosymbionts have a greater chance to spread in the population (Hurst and Randerson 2002).

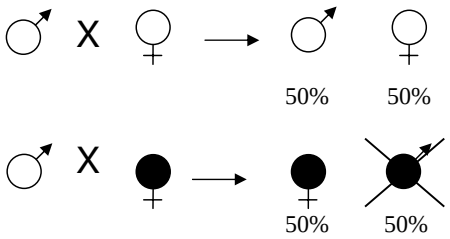


Figure 10: Male-killing illustration. Black (white) coloration: individual carries (does not carry) endosymbionts. The endosymbiont transmission rate from mother to offspring is assumed to be 100%.

d) Feminization

Feminization was first described in crustacean isopods (Vandel 1941; Martin et al. 1973; Bouchon et al. 2008). This *Wolbachia*-induced phenotype consists of development of infected genetic male offspring into functional females (Rigaud 1997b; Cordaux et al. 2011). As a result, the sex ratio of infected population favors females, which is a beneficial strategy for a maternally inherited bacterium, such as *Wolbachia* (Figure 10).

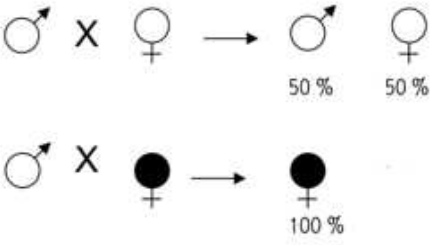


Figure 11: Feminization of genetic males, characterized by sex- ratio distortion towards females through conversion of genetic males into phenotypic females. Black (white) coloration: individual carries (does not carry) endosymbionts. The endosymbiont transmission rate from mother to offspring is assumed to be 100%.

This reproduction modification is mainly described in several species of crustacean isopods such as *A. vulgare*, *A. nasatum, Oniscus asellus* and *Porcellionides pruinosus*

(Bouchon et al. 1998; Rigaud et al. 2001; Michel-Salzat et al. 2001; Cordaux et al. 2004; Bouchon et al. 2008) but has also been described in two insect species: the butterfly *Eurema hecabe* (Hiroki et al. 2002) and the grass-dwelling leafhopper *Zyginidia pullula* (Negri et al. 2006).

II. *Wolbachia* in filarial nematodes

Human neglected diseases such as lymphatic filariasis and onchocerciasis are filarial infections by parasitic filarial worms of the phylum Nematoda. These parasites cause a wide range of clinical symptoms, including lymphedema, hydrocele, elephantiasis, dermatitis and blindness. They are mainly found in tropical countries and affect more than a million people with a billion people at risk. These are largely the "bottom billion" (1.25 billion people live on just $1.25/day), those who can least afford for even suboptimal care. For lymphatic filariasis and onchocerciasis, this leads to almost 5 million DALYs (Disability Adjusted Life Years), where 1 DALY corresponds to 1 lost year of "healthy life". They lead to severe morbidity and considerable socio-economic problems in these already poor areas. In many cases, treatment is nonexistent or insufficient due to the low drug-market potential to attract pharmaceutical companies, although several companies have donated anti-helminthics, such as GlaxoSmithKline (then SmithKline Beecham) and Merck which are collaborating with WHO to eliminate LF and provide several resources, notably albendazole. Other examples of neglected diseases caused by parasites are trypanosomiasis and leishmaniasis (Gulland 2013).

Filariasis occurs in human and animal infectious disease. So far, eight known filarial nematodes infect humans. Adult male and female nematodes live in different niche in the human body. *W. bancrofti*, *B. malayi*, and *B. timori* live in the lymphatic system, including the lymph nodes and cause the lymphatic filariasis and the elephantiasis, the most serious human filarial infections. *Loa Loa, M. streptocerca* and *O. volvulus* are located in the subcutaneous layer of the skin and are responsible for *Loa loa* blindness and the river blindness. *M. perstans* and *M.ozzard*i occupy the serous cavity of the abdomen and cause the serous cavity filariasis.

The World Health Organization (WHO) and the TDR, a Special Programme for Research and Training in Tropical Diseases, have worked together to develop scientific collaboration to help coordinate, support and influence global efforts to combat a portfolio of major diseases of the poor and disadvantaged. They have been involved with academic and industrial partners to provide a better prevention, detection and treatment.

A. Lymphatic Filariasis

Lymphatic filariasis (LF) is suspected to have affected humans since 2000 BC since an ancient Egyptian artifact, a statue of Pharaoh Mentuhotep II, shows possible elephantiasis symptoms (swollen limbs). The first written evidence of the LF appeared in ancient Greek and Roman literature with a differentiation of the LF symptoms from those of leprosy. In 1849, William Prout described for the first time a condition common to LF called chyluria in his book entitled On the Nature and Treatment of Stomach and Renal Diseases. The microfilariae were discovered in 1863 by the French surgeon Jean-Nicolas Demarquay while extracting fluid from a hydrocoele but it is Timothy Lewis who made the connection between these microfilariae and LF. Joseph Brancroft was the first to document adult worms in 1876 and give his name to the observed species (Cox 2002).

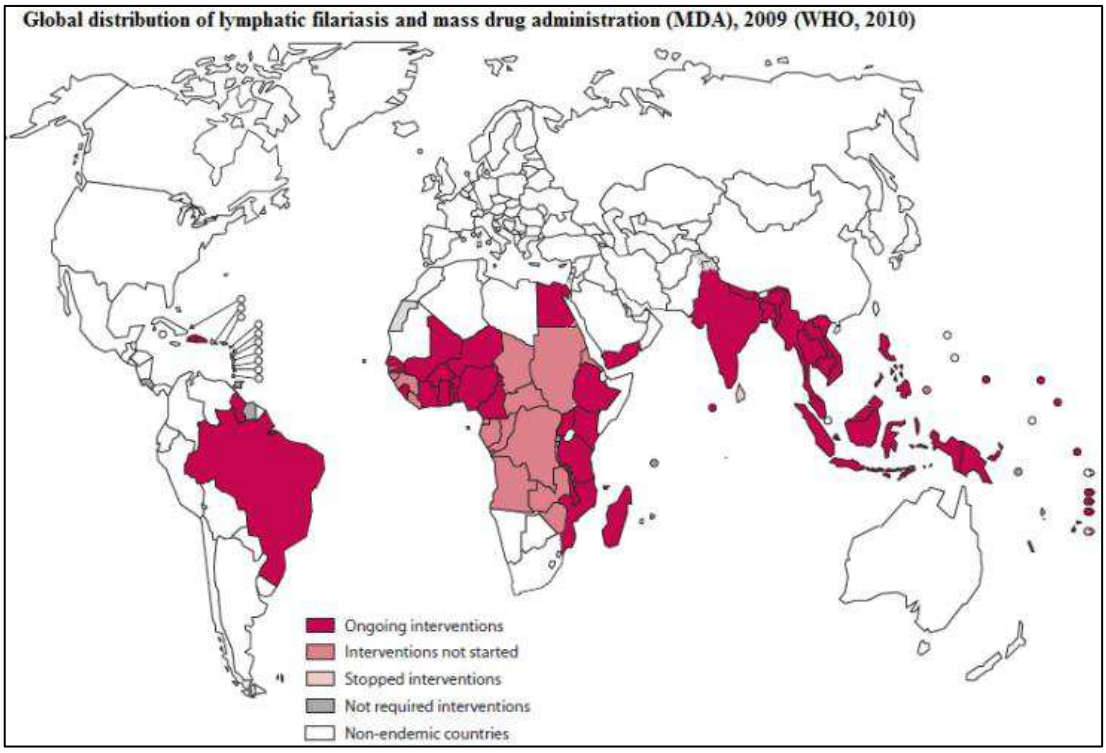


Figure 12: Distribution of the lymphatic filariasis in the world (Source: WHO, 2010).

Currently, 90% of human filariasis infections are caused by *W. bancrofti*, which exclusively infect humans, and most of the remainder by *B. malayi*, some of which can also infect some animal species such as felines and monkeys. These parasites are transmitted by mosquitoes of the genus *Culex* (in urban and semi-urban areas), *Anopheles* (in rural areas), *Aedes* (in islands of the Pacific) and *Mansonia*. The WHO

estimates that 120 million people in tropical and subtropical areas of the world (Figure 12) are infected with Lymphatic Filariasis. Almost 25 million men have genital disease (hydrocele) and almost 15 million, mostly women, have lymphedema or elephantiasis of the leg. Approximately 66% of the 1.4 billion people at risk live in the WHO South-East Asia Region and 33% in the African Region (Melrose, 2002; WHO, 2009).

LF is considered as one of the world's most disabling and disfiguring diseases (Melrose 2002; Leggat et al. 2004) mostly because of the gross swelling of the lower limbs and breast (lymphedema) and genitals (hydrocele), or swollen limbs with dramatically thickened, hard and fissured skin (elephantiasis)(Figure 13).

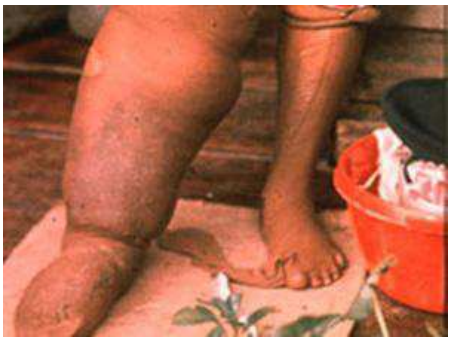


Figure 13: Elephantiasis of the lower leg. (Source: WHO)

1 - Life cycle of *Brugia malayi*

During a blood meal, a mosquito (mainly *Aedes*) containing *B. malayi* infects a new human host by releasing third-stage filarial larvae (L3) onto the skin, where they penetrate into the bite hole (Figure 14 #1). The larvae travel to the lymphatic system where they develop into adults (Figure 14 #2). Female worms measure about 43 to 55 mm in length by about 130 to 170 µm in width, and males measure about 13 to 23 mm in length by 70 to 80 µm in width (Sasa et al. 1976). Worms survive for more than a decade in the host during which time the female is constantly producing small microfilariae (measuring about 177 to 230 µm in length and about 5 to 7 µm in width.). Microfilariae can survive many months in the human host and have nocturnal periodicity: they sequester in tissues the day and migrate into lymph and enter the blood stream reaching the peripheral blood by night (Figure 14 #3). Following the ingestion by a mosquito during its nightly blood meal (Figure 14 #4), the microfilariae lose their sheaths and migrate through the wall of the proventriculus and cardiac portion of the midgut to reach the thoracic muscles (Figure 14 #5) where they develop into first-stage larvae (L1) (Figure 14 #6) then in third-stage larvae (L3) (Figure 14 #7)

in about 10 days. Infective L3 migrate through the hemocoel to the mosquito's proboscis (Figure 14 #8) and can infect another human during a blood meal (Figure 14 #1).

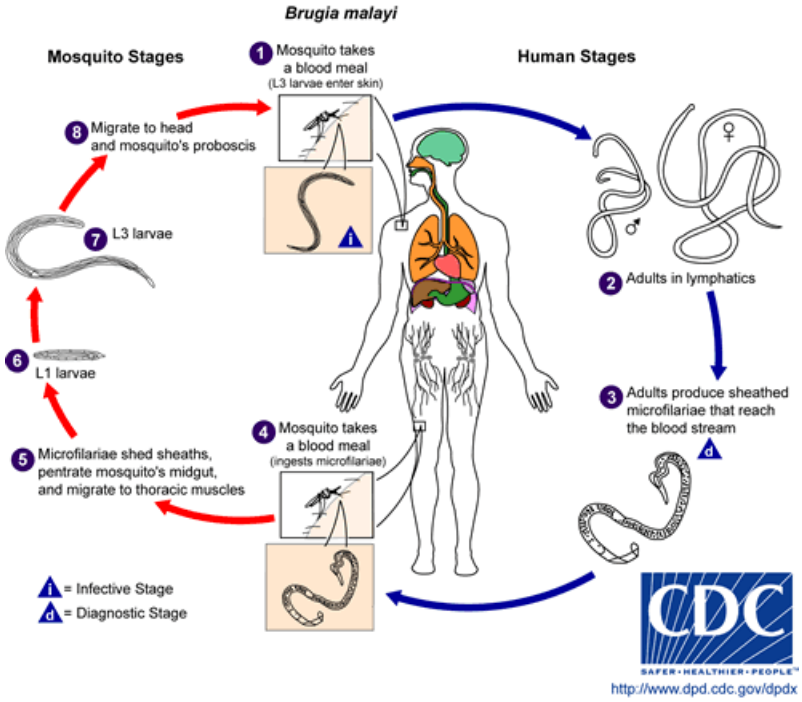


Figure 14: Life cycle of *Brugia malayi* (source: CDC website)

2 - Diagnostic and Treatment of Lymphatic Filariasis

Diagnosis of active LF infection in the absence of gross swelling is challenging. The commonly use diagnosis method consist in the detection of microfilariae in blood drawn at night (to coincide with the appearance of the microfilariae) by microscopic examination and determination of the nematode species by Giemsa staining of blood smears. Serologic techniques provide an alternative by detection of elevated levels of antifilarial IgG4 in the blood but lymphedema may develop many years after infection, these tests are most likely to be negative. Molecular-based diagnostic tools are considered more accurate and allow the actual identification of the infectious nematodes. These technics are mainly based on polymerase chain reaction (PCR), which have been used to amplify *B. malayi*, *B. timori* and *W. bancrofti* DNAs from blood samples or insect vector samples (Lizotte et al. 1994; Bockarie et al. 2000; Hoti et al. 2001; Kluber et al. 2001; Fischer et al. 2005; Rao et al. 2006).

More recently a new parasitic nematode detection procedure has been developed using loop-mediated isothermal amplification (Poole et al. 2012).

To achieve interruption of transmission in endemic regions, the WHO launched in 2000 the Global Programme to Eliminate Lymphatic Filariasis. This elimination strategy consisted in interrupting the transmission and controlling the morbidity. After mapping the Lymphatic Filariasis endemic region, mass drug administration of filaricidal drugs were provided with a recommendation of a yearly administration of Diethylcarbamazine citrate (DEC) and Albendazole for 5 years (or Ivermectin and Albendazole for areas also endemic for onchocerciasis). By 2010, 59 endemic countries had completed mapping and 53 countries had started implementing mass drug administration (Table1).

Programs for onchocerciasis control, (OCP, APOC and OEPA) together with lymphatic filariasis (GPELF) have been established to promote and sustain the application of MDA (Mass Drug Administration) to affected communities.

Table 1: The Worldwide Abundance, Burden of Disease, Distribution and Control/Elimination of Human Helminthiases (Lustigman et al. 2012).

Infection	Causal Agent	Region with Highest No. Infected	Number Infected (Millions)	DALYs (Millions)	Number of Deaths/ Year (Thousands)	Programmes Involved
Onchocerciasis	Onchocerca volvulus	SSA	37	1.5[a]	0.05 (in the OCP area)[b]	OCP, APOC, OEPA
Lymphatic filariasis	Wuchereria bancrofti; Brugia malayi	India, SEA, SSA	120	5.8	0.4	GPELF
Ascariasis	Ascaris lumbricoides	Asia, Africa, LA	1,221–1,472[c]	1.8–10.5[c]	3–60[c]	PPC, DtW, GPELF, SCI
Trichuriasis	Trichuris trichiura	Asia, Africa, LA	759–1,050[c]	1.0–6.4[c]	3–10[c]	PPC, DtW, GPELF, SCI
Hookworm infection	Necator americanus; Ancylostoma duodenale	Asia, Africa, LA	740–1,300[c]	0.1–22.1[c]	3–65[c]	PPC, DtW, GPELF, SCI
Schistosomiasis	S. mansoni S. haematobium S. japonicum	SSA, LA SSA China, SEA	207	1.7–4.5[c]	15–280[c]	SCI in SSA; national programmes elsewhere
Food-borne trematodiases	Clonorchis sinensis; Opisthorchis viverinni; Paragonimus spp.; Fasciolopsis buski; Fasciola hepatica	East Asia	56[d]	0.5–0.9[d]	7[d]	Large-scale control initiatives lacking
Cestode infections: cysticercosis	Taenia solium	SSA, Asia, LA	0.4 (LA only)	ND	ND	Large-scale control initiatives are lacking

Modified from references [2,3,10,14–16,36,77,120].
[a]From Remme et al. [10].
[b]From Little et al. [120].
[c]From Utzinger and Keiser [14].
[d]From Fürst et al. [36].
Abbreviations: SSA, sub-Saharan Africa; SEA, Southeast Asia; LA, Latin America; OCP, Onchocerciasis Control Programme in West Africa (1975–2002); APOC, African Programme for Onchocerciasis Control (1995–ongoing); OEPA, Onchocerciasis Elimination Program for the Americas (1993–ongoing); GPELF, Global Program to Eliminate Lymphatic Filariasis (2002–ongoing); PPC, Partners for Parasite Control (2001–ongoing); DtW, Deworm the World (2007–ongoing); SCI, Schistosomiasis Control Initiative (2002–ongoing).
doi:10.1371/journal.pntd.0001582.t001

These treatments reduce the microfilariae levels low enough to sustain transmission of filarial parasite to humans but they have a limited, if any, activity against adult worms (Noroes et al. 1997; Ramzy et al. 2002). Hygiene training of lymphoedema patients has also reduced the lymphoedema-associated morbidity (Kerketta et al. 2005) and prevents aggravation of the disease condition. Mosquito control is another measure used to suppress transmission with measures such as insecticide treatments.

B. *Wolbachia* as a drug target

Conventional strategies to combat filariasis have included vector control as a complement to antiparasitic drugs (Hougard et al. 2001; Richards et al. 2001). These treatments are only efficient in endemic areas with repeated administration to prevent new infections (Campbell 1991; Richards et al. 2001; Hoerauf et al. 2003). Mass drug administration can possibly lead to the development of drug resistance (Prichard 1994, 2001). This has required a search for new drugs for the control of filariasis especially targeting *Wolbachia*. The presence of *Wolbachia* in both male and female worms, as well as in juvenile stages has led to its being investigated as a potential drug target. Because *Wolbachia* is essential for worm development, fertility and survival (Hoerauf et al. 2000a; Taylor et al. 2005; Pfarr and Hoerauf 2007), antibiotic treatments such as tetracycline, have shown significant results. The depletion of *Wolbachia* induces the death of filarial worms (Taylor et al. 2005; Hoerauf 2006). *Wolbachia* are also known to contribute to the major adverse reactions observed after anti-parasitic treatment. Many reports suggest that antihelmintic drugs cause severe systemic inflammatory responses such as fever, headache, dizziness and enlargement of lymph nodes which could be associated to the release of *Wolbachia* from dead microfilariae (Boreham and Atwell 1983; Francis et al. 1985; Cross et al. 2001; Taylor et al. 2001; Supali et al. 2008). Thus, use of anti-*Wolbachia* chemotherapy against parasitic nematodes is being used. However, current antibiotic treatments are long and with high doses (ex: 6 weeks of 200 mg/ml doxycycline), difficult to maintain in endemic labs. In addition, anti-*Wolbachia* treatments are also potential adulticides: current chemotherapies reduce or eliminate the microfilariae stage, but have little effect upon the adults, which continue to survive up to 10 years in the mammalian host. Continual treatment is thus necessary to disrupt the life cycle.

These studies demonstrate the feasibility of developing treatments targeting the *Wolbachia* endosymbiont (Hoerauf et al. 2000b; Pfarr and Hoerauf 2006).

The A·WOL Consortium was founded in 2007 after a $23,000,000 grant award by The Bill & Melinda Gates Foundation (http://a-wol.com). The 63 scientists and staff working of the project collaborate with pharmaceutical companies to cover the entire drug discovery and development process that is required for new drugs to treat river blindness and elephantiasis by targeting *Wolbachia*. This project has led to the identification of several drug targets in *Wolbachia* (Taylor et al. 2013).

III. *Wolbachia* and sex determination

A. The biology of terrestrial crustacean

Sexual differentiation of crustaceans has been extensively studied since they represent the only group of invertebrates for which the existence of a proteinaceous sex hormone was demonstrated. Actually, the androgenic gland, which is responsible for male sexual character development, was discovered in 1954 in the amphipod *Orchestia gammarella* by Charniaux-Cotton (Charniaux-Cotton 1954) and further identified in males from several species of amphipods (Charniaux-Cotton et al. 1992; Hasegawa et al. 1993), decapods (Hoffmann 1969; Sagi et al. 1990; Lee T.H. 1994) and isopods (Katakura 1961; Juchault and Legrand 1964, 1966; Juchault 1977). This gland secrets the androgenic hormone, which controls the differentiation of primary and secondary sexual characters in males (Charniaux-Cotton 1960; Juchault and Legrand 1966; Charniaux-Cotton H. 1992; Sagi A. 1997). In females, the androgenic glands do not develop and gonads spontaneously evolve into ovaries (Katakura 1989). The androgenic hormone has been purified for the first time in the terrestrial isopod *A. vulgare* in 1987 (Hasegawa et al. 1987; Martin et al. 1990). The prohormone is composed of 123 amino acids whose structure is close to that of insulin although they present no sequence homologies. The complete amino acid sequence as well as the cDNA sequence of this hormone was only determined ten years later (Martin et al. 1999; Okuno et al. 1999). *A. vulgare* is one of the most studied models regarding sexual differentiation among crustaceans, providing a lot of data on its reproduction cycle, post-embryonic development and on the hormonal control of its sexual differentiation. Notably, sex-reversal experiments, carried out by implanting ectopic glands in young females (Suzuki and Yamasaki 1997; Martin and Juchault 1999) or by removing androgenic glands in males (Suzuki and Yamasaki 1991), favored the understanding of androgenic hormone action on the gonad differentiation during male development.

Following egg delivery in the marsupial pouch, embryos develop during four weeks before being released. The post-embryonic developmental stages have been precisely described by (Suzuki and Yamasaki 1995). Eight post-embryonic stages were defined within a period of ten to fifteen weeks (Table 2).

Table 2: Post-embryonic development stages of *A. vulgare*. Days (D) or weeks (W) after offspring are released from the female's brood pouch. Adapted from (Suzuki and Yamasaki 1995).

Stages	0	1	2	3	4	5	6	7	8
Age	1D	1W	2W	3W	4W	5W	6W	7W	8W
Size	1mm			2mm	2.5mm	2.9mm	3.5mm	4.4mm	5.2mm
Differentiation stage of gonads	Undifferentiated gonads				Differentiated gonads				
Differentiation stage of external sexual characters	No differentiation					Appearance of external sexual characters in males and females			

In the isopod *A. vulgare*, sexual differentiation is determined by the sexual chromosomes Z and W. Females are heterogametic ZW and males are homogametic ZZ. Embryonic gonads start to differentiate from the fourth stage into male or female gonads, according to the ZZ male genotype and to the WZ female genotype, respectively. Studying the molecular mechanism of the impact of *Wolbachia* on the feminization of males will focus on this fourth stage as well as the immediately preceding stages.

B. Feminization mechanism

In a regular mating, offspring are 50% male ZZ and 50% female ZW. But this 1:1 ratio can be female-biased when females are *Wolbachia*-infected (Martin et al. 1973; Bouchon et al. 1998; Bouchon et al. 2008). In infected females, 80% to 90% of the mature oocytes are *Wolbachia*-infected (Genty et al., 2013. submitted). These oocytes will develop a female phenotype independent of their genotype, meaning that ZZ genetic males harboring *Wolbachia* will also develop into functional phenotypic females.

achia–*arthropod interactions*

x uniraptor, are *Wolbachia* symbionts (Rou
ıay also be the case for *Aphytis lingnanes*
ı which *Wolbachia* alter the sexuality of the
.e females of the woodlouse *Armadillidiu*
ɜed broods (Vandel, 1941; Juchault *et al.*, 1
is strong sex ratio deviation; they reverse

Figure 15: Feminization of *A. vulgare* genetic males, characterized by sex- ratio distortion towards females through conversion of genetic ZZ males into phenotypic ZZ females. ZZ/ZW, homo/heterogametic status of the individual. The endosymbiont transmission rate from mother to offspring is assumed to be 100% (Rigaud and Rousset 1996) .

The conservation of real males in the population is maintained by the "imperfect" transmission of the bacteria to 10%-20% of the oocytes. Uninfected, they develop according to their genotype. The prevalence of the ZW females decreases at each generation because of the predominance of feminized ZZ individuals, eventually producing populations without the W female chromosome (Rigaud 1997a; Bouchon et al. 2008; Cordaux et al. 2011) (Figure 15).

In these *Wolbachia*-infected populations, sex determination is completely dependent on *Wolbachia* presence/absence: individuals inheriting *Wolbachia* develop as females and individuals without *Wolbachia* develop into males. This sex determination model corresponds to a cytoplasmic sex determination; all individuals in infected wild populations are chromosomic males (Figure 16). Even if the molecular mechanism of the feminization is still unknown, observations have shown that feminized males never present androgenic gland differentiation. It is hypothesized that bacterial factors might interact with host targets, including the androgenic hormone pathway, no later than the developmental stage 4, just before the beginning of gonad differentiation (Bouchon et al. 2008; Cordaux et al. 2011; Negri et al. 2010).

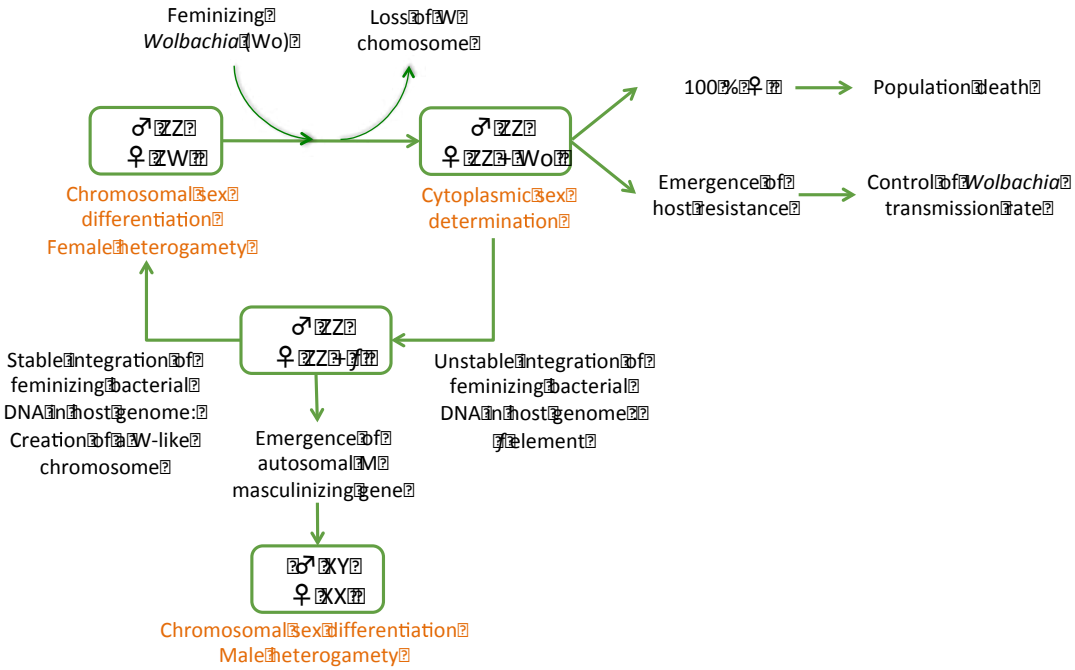


Figure 16: Schematic sex determination mechanisms in the woodlice *A. vulgare*. Adapted from (Juchault and Mocquard 1993; Rigaud and Rousset 1996; Cordaux et al. 2011)

Another feminizing factor, known as the *f* element, has been identified in *A. vulgare*. This element is hypothesized to have integrated pieces of the bacterial genome that carry the feminizing factor(s) into host chromosomes (Legrand and Juchault 1984). Stable integration of these bacterial sequences into a Z chromosome lead to the creation of a W-like chromosome, restoring the chromosomal sex differentiation. However, the presence of this feminizing element has resulted in the selection of host resistance gene(s): dominant autosomal masculinizing *M* gene(s), which restores male development in the presence of the *f* element (Rigaud and Juchault 1993), leading to a chromosomal sex differentiation via a male heterogamety.

IV. *Wolbachia* genome projects

In 1999, to unravel the biology of *Wolbachia*, the "*Wolbachia* Genome Consortium" planned to sequence various genomes representing the diversity of *Wolbachia* (Slatko et al. 1999). During the first meeting of this Consortium held in Beverly, MA (USA) in January 1999, John Werren from the University of Rocherster (NY, USA), presented the biology of *Wolbachia* in association with Scott O'Neill (Yale University, New Haven, CT, USA) and Claudio Bandi (University of Milano, Italy). David Guiliano from the University of Edinburgh (UK) discussed the implications of studying the *Wolbachia* genome from filarial nematodes because the phylogeny of *Wolbachia* endosymbiont is congruent with the phylogeny of the host worms and all the worms studied so far were harboring the bacterium, suggesting a long co-evolutionary history. Since *Wolbachia* is a group of closely related bacteria that can induce a wild range of phenotypic effects on their host, the idea from this consortium was to study and compare the genomes of these bacteria to examine the different *Wolbachia*-induced phenotypes and the basic biology of these different symbioses. The underlying reason for this project was to determine the molecular mechanisms of host reproduction that could be used to control insect pest and disease vectors. For nematodes, *Wolbachia* could provide molecular target for diagnosis and control of filarial disease.

One month later, in February 1999, a second meeting of the Consortium was held to decide which *Wolbachia* genome should be sequenced and the technique to use, aiming to create a collaborative network for these sequencing projects. Kostas Bourtzis (Institute of Molecular Biology and Biotechnology, Heraklion, Greece) presented the European initiatives, which aim at sequencing genomes of three *Wolbachia* strains, each responsible for the induction of cytoplasmic incompatibility (*w*Ri), parthenogenesis

(*w*Uni) and feminization (*w*VulC). During this meeting, Mark Blaxter (University of Edinburgh, UK) exposed the state-of-the-art in genome studies on *Wolbachia* from the filarial nematode *B. malayi*. (Bandi et al. 1999). However, the first challenge in *Wolbachia* genome sequencing was to purify enough high quality *Wolbachia* DNA from host DNA and to estimate the *Wolbachia* genome sizes (Sun et al. 2001).

The first complete *Wolbachia* genome was published five years later, in 2004. It was the CI-inducing *w*Mel strain from the fly *D. melanogaster* yw^{67c23} that naturally contains the *w*Mel infection (Wu et al. 2004). The DNA was extracted from young adult flies and separated by pulsed-field gel electrophoresis. Sequencing was performed using the whole-genome shotgun method (Venter et al. 1996) producing 1.5-2kb and 4-8kb libraries. They pointed out a difficulty (which was later confirmed by all *Wolbachia* genome projects) in assembling *Wolbachia* genome, due to the presence of large amount of repetitive DNA. This genome is a single circular molecule of 1.27 Mb with a GC content of 35.2%. This first sequenced genome highlighted *Wolbachia* genome specificities such as an important number of repetitive DNA as well as mobile elements including insertion sequences (IS), transposable elements (TE) and phages, and the presence of many metabolic differences with the closely related Rickettsia species: *Wolbachia w*Mel presented an intact glycolysis and purine synthesis but an apparent inability to synthetize lipopolysaccharide. Concerning host-symbiont communication, a complete type IV secretion system (T4SS) was identified: in addition to the five *vir* genes previously described from *Wolbachia w*KueYo (Masui et al. 2000), they identified four other in *w*Mel. Of these nine *vir* genes, they showed that eight of them were organized into two separate operons. The T4SS has been considered as a putative route by which bacterial determinants are translocated into host cells and subsequently interact with host molecules. *w*Mel genome also revealed several genes encoding eukaryote motif-containing proteins including an unexpected large amount of ankyrin genes identified as putative bacterial effectors involved in *Wolbachia*-induced phenotypes.

All insect *Wolbachia* genomes also contained a large number of repetitive elements such as prophages (Kent and Bordenstein 2010; Kent et al. 2011a; Metcalf and Bordenstein 2012), insertion elements (Cordaux 2008; Cordaux et al. 2008; Cordaux 2009; Cerveau et al. 2011) and group II introns (Leclercq et al. 2011). Evidences have been found that several recombination events occurred within and between *Wolbachia* genomes (Ling and Cordaux 2010) leading to horizontal DNA transfers within and between divergent *Wolbachia* supergroups, explaining the difficulties to establish a reliable phylogenetic

tree of the *Wolbachia* strains and the finding of the most recent common ancestor of this clade of bacteria (Lo et al. 2002; Baldo et al. 2006; Bordenstein et al. 2009).

At the onset of this current study, only four *Wolbachia* genomes had been completely sequenced and deposited in GENBANK, *w*Mel, *w*Bm of *B. malayi* (Foster et al. 2005), *w*Pip of *C. quinquefasciatus Pel (Klasson et al. 2008)*, *w*Ri of *D. simulans* (Klasson et al. 2009). Three *Drosophila* endosymbiont genomes were also retrieved from previous host sequencing (Salzberg et al 2005). More recently, three additional complete genomes have been published: *w*Oo of *O. ochengi* (Darby et al. 2012), and *w*No and *w*Ha of *D. simulans* (Ellegaard et al. 2013). Additional genomes are currently being sequenced/annotated and partial sequences for several *Wolbachia* strains are available (Table 3).

Table 3: *Wolbachia* genome project summary.

Strain	Host	Supergroup	Phenotype	Genome size (Mb)	Status	Refs or project leaders
wAtab3	*Asobara tabida*	A	Depedence	Unknown	Unfinished	Vavre et al.
*w*MelPop	*Drosophila melanogaster*	A	Cytoplasmic incompatibility	1.3	Unfinished	S. O'Neill
*w*Mel	*Drosophila melanogaster*	A	Cytoplasmic incompatibility	1.27	Complete	Xi, Z. et al
*w*Sim	*Drosophila simulans*	A	Cytoplasmic incompatibility	Unknown	629 scaffolds	Salzberg, S. L. *et al.*
*w*Au	*Drosophila simulans*	A	Not cytoplasmic incompatibility	Unknown	Unfinished	S. O'Neill
*w*Ha	*Drosophila simulans*	A	Cytoplasmic incompatibility	1.3	complete	Ellegaard et al.
*w*Ana	*Drosophila simulans*	A	Cytoplasmic incompatibility	Unknown	464 scaffolds	Salzberg, S. L. et al.
*w*Ri	*Drosophila simulans*	A	Cytoplasmic incompatibility	1.44	Complete	S. Andersson et al
*w*Suzi	*Drosophila suzukii*	A	Cytoplasmic incompatibility	1.4	110 scaffolds	Siozios, S et al
*w*Wil	*Drosophila willistoni*	A	Cytoplasmic incompatibility	0.86	260 scaffolds	Remington, K et al
*w*Gmm	*Glossina morsitans*	A	Cytoplasmic incompatibility	Unknown	Unfinished	Bourtzis et al.
*w*Uni	*Muscidifurax uniraptor*	A	Parthogenesis	Unknown	256 scaffolds	S. Anderson and K. Bourtzis
*w*VitA	*Nasonia vitripennis*	A	Cytoplasmic incompatibility	Unknown	Unfinished	J. Werren and S. Richards
*w*AlbB	*Aedes albopictus*	B	Cytoplasmic incompatibility	1.24	165 scaffolds	Mavingui et al.
*w*VulC	*Armadillidium vulgare*	B	Feminization	1.7	103 scaffolds	R. Garrett et al.
*w*Pip Pel	*Culex pipiens*	B	Cytoplasmic incompatibility	1.48	Complete	J. Parkhill and S. Sinkins
*w*Pip-Mol	*Culex pipiens molestus*	B	Unknown	1.5	888 scaffolds	Aslett M. et al
*w*Pip HB	*Culex quinquefasciatus HB*	B	Cytoplasmic incompatibility	1.54	21 scaffolds	Salzberg, S. L. et al.
*w*Di	*Diaphorina citri*	B	Unknown	1.2	124 scaffolds	W. Hunter et al.
*w*No	*Drosophila simulans*	B	Cytoplasmic incompatibility	1.3	Complete	Ellegaard et al.
*w*Bol1	*Hypolimnas bolina*	B	Male killing	1.6	144 scaffolds	A. Duplouy and S. O'Neill
*w*VitB	*Nasonia vitripennis*	B	Cytoplasmic incompatibility	1.1	523 scaffolds	Kent et al.
*w*Dim	*Dirofilaria immitis*	C	Mutualist	0.9	2 scaffolds	C. Bandi and B. Slatko
*w*Oo	*Onchocerca ochengi*	C	Mutualist	0.96	Complete	Darby, A.C. et al.
*w*Ov	*Onchocerca volvulus*	C	Mutualist	0.44	341 scaffolds	M. Taylor et al.
*w*Bm	*Brugia malayi*	D	Mutualist	1.08	Complete	Foster, J. et al.
*w*Wb	*Wuchereria bancrofti*	D	Mutualist	1.05	763 scaffolds	Nutman, T.B et al.
*w*Cle	*Cimex lectularius*	F	Nutritional mutualist	1.3	Complete, unpublished	Fukatsu et al.

A common goal of many *Wolbachia* genome projects is to provide comparative genomic information for understanding mechanisms of genome evolution and mechanisms of host phenotype manipulation by *Wolbachia* (Werren 1997; Brownlie and O'Neill 2005; Werren et al. 2008; Saridaki and Bourtzis 2010; Cordaux et al. 2011).

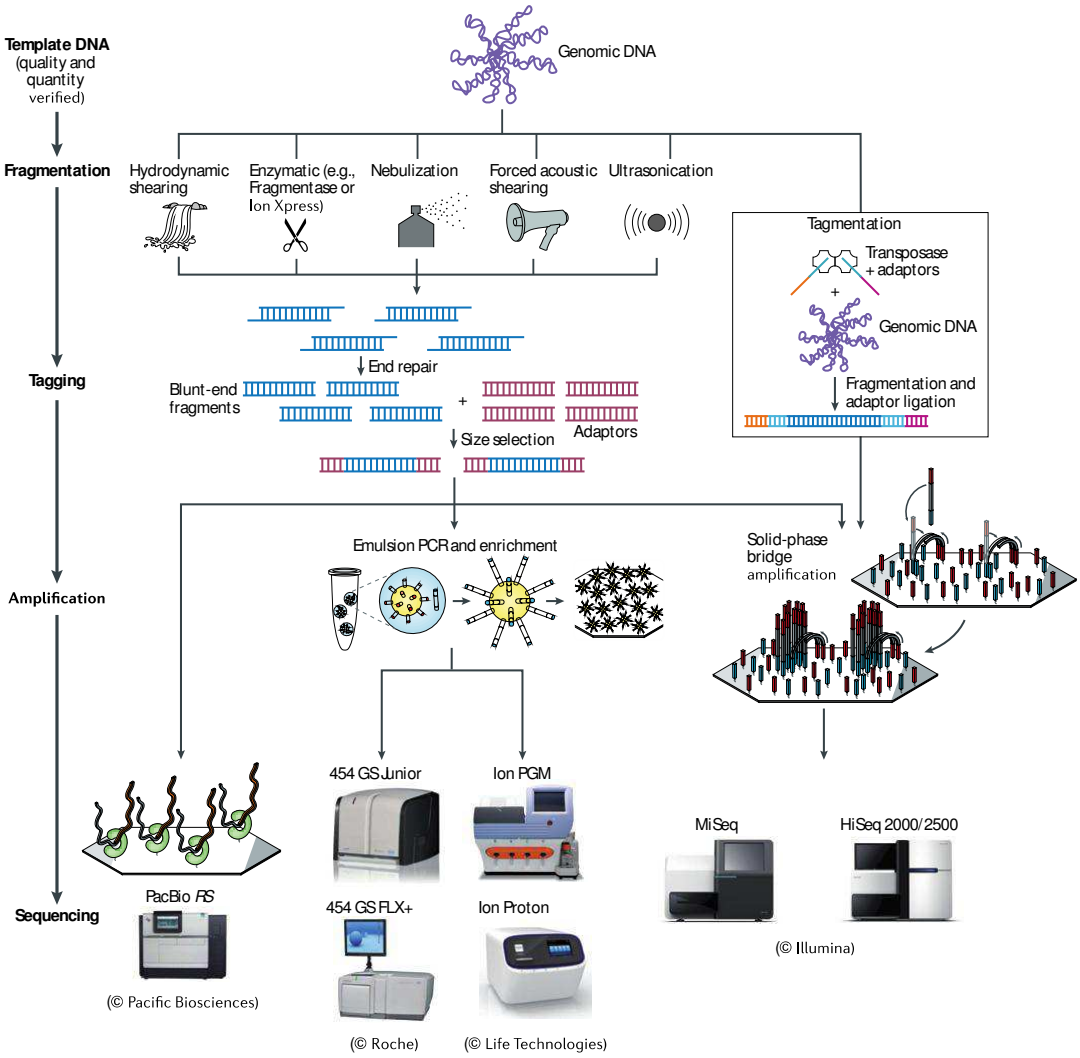


Figure 17: Next-generation sequencing platforms (Loman et al. 2012).

These sequencing projects are currently facilitated by the emergence of the Next-Generation Sequencing (NGS), which makes genome sequencing more accessible (Figure 17). These techniques enable faster, less expensive and deeper sequencing coverage and are now available in small "benchtop" machines. Today, these are four main technologies used with independent specificities that allow researchers to utilize the most adapted machines to their project. While all these technologies need a sequencing library construction step prior to sequencing, the Illumina technologies do not need any additional amplification step that is required for both Roche 454 and Ion Torrent technologies. However, the 454 was initially preferred by scientists as it allowed sequencing of larger fragments, but with the MiSeq machines, Illumina may also produce the same read length for a reduced cost. Indeed, the limit of these high-throughput machines is the short length of the sequences produced (from 50bp to ~400bp). Pacific Biosciences (PacBio) recently developed a new technology that allows

the sequencing of single molecule of a size up to 40kb. The limitation of this new technique is the relatively high percentage of error, but because of the depth of coverage obtained and combined with an Illumina approach, it facilitates assemblies of genomes with repetitive sequences such as *Wolbachia*.
However, the bottleneck of these NGS whole-genome sequencing is the subsequent bioinformatics data analysis that is needed to perform the analysis with servers to handle this huge amount of generated data.

Applying these new technologies to *Wolbachia* genome sequencing will allow whole genome comparison analyses to help understand the biology of these particular symbiosis relationships.

V. Aims of the thesis

A key step in the investigation to develop new anti filarial drugs or to use *Wolbachia* as agricultural pest and disease vector control is the understanding of the relationship between *Wolbachia* and their hosts. As part of that goal, the research developed in this PhD project aims at understanding the molecular mechanisms of the symbiosis between *Wolbachia* bacteria and their hosts. This research was focused on the identification of putative bacterial effectors involved in *Wolbachia*-host interaction in mutualistic (*B. malayi*) and parasitic (*A. vulgare*) models using genomic, transcriptomic and proteomic approaches.

While most of the arthropods are infected by at least a strain of *Wolbachia*, only a few *Wolbachia* strains from arthropods have been sequenced so far and only one strain infecting terrestrial crustaceans is currently assembled in 10 contigs.

In order to generate a larger *Wolbachia* genome dataset to perform comparative genomics, one step of this PhD project was to develop an efficient *Wolbachia* DNA purification procedure, using the *w*Bm (*Wolbachia* from *B. malayi*) genome as the test case, to allow the sequencing of additional *Wolbachia* strains infecting isopods that are inducing either host feminization or CI.

Whole-genome comparative genomics was performed to identify (i) symbiosis related gene patterns; genes only present in strains inducing the same phenotype and (ii) bacterial effectors putatively involved in symbiont-host interaction such as membrane and eukaryote-like proteins and secretion systems. This approach was complemented in parallel by a transcriptomic analysis and the investigation of protein interactome between *Wolbachia* and both the nematode *B. malayi* and the isopod *A. vulgare*.

This PhD work was done in collaboration with the Ecology Evolution Symbiosis lab (UMR CNRS7267) at the University of Poitiers, France and the parasitology lab at New England Biolabs, USA, under the supervision of Pierre Grève and Barton Slatko.

GENOMIC ANALYSIS OF MULTIPLE *WOLBACHIA* STRAINS

Introduction

With the emergence of the new sequencing technologies, whole genome sequencing became accessible for bacterial genomes such as *Wolbachia*. However, obtaining enough pure *Wolbachia* DNA of high quality remained a challenge. Purification strategies such as techniques including chemical gradients (Charles and Ishikawa 1999), pulsed-field gel purification with or without whole genome amplification, library construction followed by gene walking, etc. (Sun et al. 2001; Mavingui et al. 2005; Foster et al. 2005; Iturbe-Ormaetxe et al. 2011) remained unsuccessful or organism specific. More recently, informatics (Kumar and Blaxter 2011) and cell sorting methods (Santos-Garcia 2012) have been used.

To circumvent this general problem for endosymbiotic organisms, we devised a novel strategy based on methods used to sequence only the coding regions of large eukaryotic organisms. The method can be used to specifically "capture" and isolate *Wolbachia* DNA away from host DNA (Mamanova et al. 2010). After our preliminary experiments verified that this procedure worked for the previously completely sequenced mutualistic *w*Bm strain from *B. malayi* (Foster et al. 2005), we applied this method to the feminizing *w*VulC strain from *A. vulgare*, which is distantly related to *w*Bm and almost totally sequenced and assembled in ten contigs. Then, we optimized the procedure by supplementing the bait library and improving the DNA preparation and sequencing. Finally, we extended this approach using 8 other isopod strains chosen among the collection held by the Ecologie Evolution Symbiose laboratory (Table 4).

Table 4: Chart of the 39 species of terrestrial crustacean collection held by the Ecology Evolution Symbiosis laboratory in Poitiers. *Wolbachia*-infected species are in bold.

Acaeroplastes melanurus	*Armadillo officinalis*	***Porcelliodilatatus dilatatus***
Armadillidium badium	***Atlantoscia floridana***	***Porcelliodilatatus petiti***
Armadillidium decorum	***Balloniscus glaber***	***Porcellio dispar***
Armadillidium depressum	*Balloniscus sellowi*	*Porcellio gallicus*
Armadillidium granulatum	*Chaetophiloscia elongata*	***Porcellio laevis***
Armadillidium maculatum	***Cubaris murina***	***Porcellio lamellatus***
Armadillidium nasatum	***Cylisticus convexus***	*Porcellio variabillis*
Armadillidium siculorum	*Eluma purpurescens*	***Porcellio scaber***
Armadillidium tunisiense	*Haplophtalmus danicus*	*Porcellionides tingendus*
Armadillidium versicolor	***Helleria brevicornis***	***Porcellionides pruinosus***
Armadillidium virgo	***Oniscus asellus***	*Porcellionides sexfaciatus*
Armadillidium vulgare	***Oniscus lusitanus***	*Trichoniscus pusillus*
Armadillidium vulgare albinos	*Platyarthrus hiasensis*	*Trichorhina tomentosa*

This large collection consists of 39 different species of crustacean isopods of which 21 are infected (Table 4). Several *Wolbachia* strains have been identified for some isopod

species such as *A. vulgare,* which can be infected by either of the three strains *w*VulC, *w*VulM, *w*VulP. However, there are no multi-infections within the same individual in these laboratory lineages (Bouchon et al. 1998; Cordaux et al. 2001; Cordaux et al. 2004; Verne et al. 2007; Bouchon et al. 2008; Wiwatanaratanabutr et al. 2009; Cordaux et al. 2012).

Hence genomic sequencing was performed on *Wolbachia* DNA isolated from two CI strains, *w*Dil from *P. dilatatus dilatatus* and *w*Pet from *P. dilatatus petiti*, five feminizing strains *w*VulM from *A. vulgare*, *w*Albu from *A. album, w*Nas from *A. nasatum, w*Pru from *P. pruinosus and w*Ase from *O. Asellus* and one 'unknown phenotype' strain *w*Bre from *Helleria brevicornis*. These isopod *Wolbachia* strains were chosen either because of their phenotype (CI vs. feminization) or their position in the *wsp* phylogeny tree (Figure 18). While the virulent *w*Albu strain, which induces a high rate of intersexes in natural populations, and *w*Nas were chosen because of their high proximity to the *w*VulC reference strain, *w*VulM was sequenced because it infests the same host as *w*VulC but does not present the same virulence nor the same position in the phylogenetic tree. The *w*Bre strain is at the deep-branching position of the subgroup of *Wolbachia* from terrestrial crustaceans and *w*Pru is the most distant strain, closer to insect strains. Furthermore, the two strains *w*Pru and *w*Ase also infect males, which is unexpected for feminizing strains. The *w*Dil and *w*Pet *Wolbachia* strains are two strains from *P. dilatatus* species that induce CI.

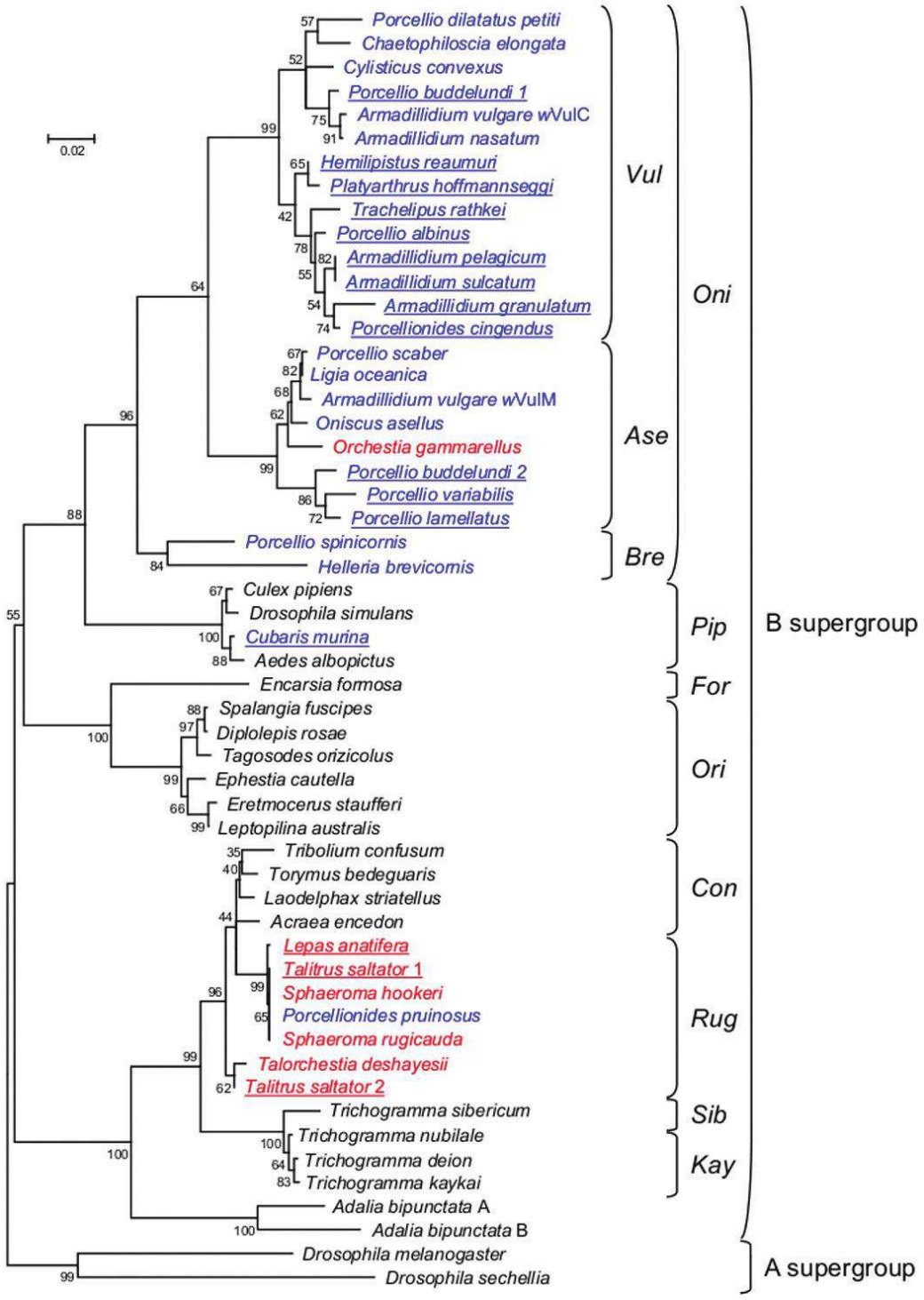


Figure 18: Phylogeny of B-supergroup *Wolbachia* strains based on *wsp* sequences, using Minimum Evolution analysis. (Cordaux et al. 2012)

I. Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA

To develop an efficient purification for endosymbiotic bacteria we adapted the Agilent SureSelect kit, originally designed for exon capture. This kit is a solution-based method conceived to enrich a DNA sample by pulling down the exon sequences using a biotinylated complementary RNA library. Once exons are hybridized with the complementary sequences, regions of interest are isolated using magnetic streptavidin beads. After RNA digestion, a NGS genomic library is prepared (Figure 19).

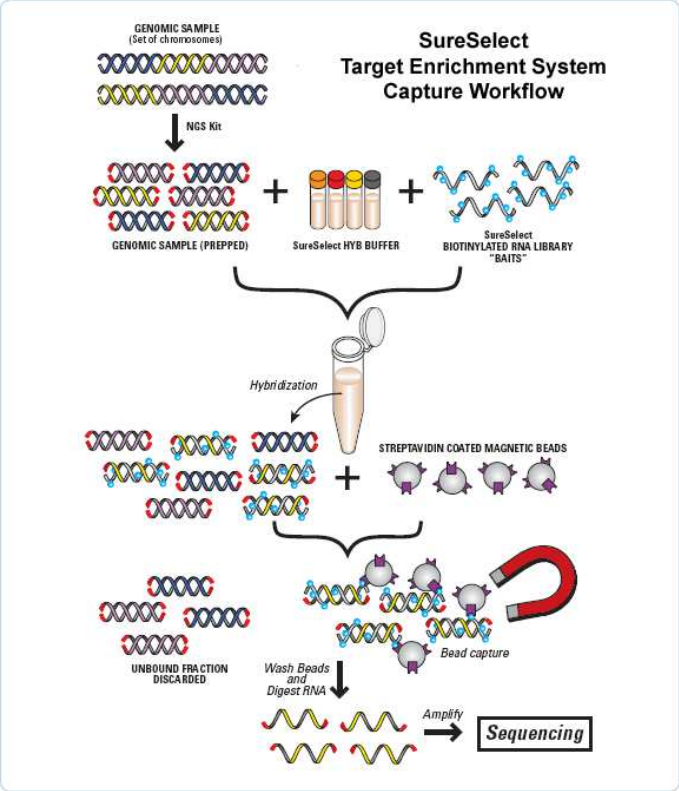


Figure 19: SureSelect Target Enrichment workflow.

We adapted this method by designing a cRNA bait library specific to *Wolbachia* genomes. Instead of targeting the exome, we targeted the endosymbiotic bacteria genome. This procedure was published in Symbiosis.

A. Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA (Geniez et *al*. 2012).

Symbiosis (2012) 58:201–207
DOI 10.1007/s13199-012-0215-x

Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA

Sandrine Geniez · Jeremy M. Foster · Sanjay Kumar · Bouziane Moumen · Emily LeProust · Owen Hardy · Moraima Guadalupe · Stephen J. Thomas · Braden Boone · Cynthia Hendrickson · Didier Bouchon · Pierre Grève · Barton E. Slatko

Received: 28 August 2012 / Accepted: 10 December 2012 / Published online: 6 January 2013

Abstract *Wolbachia* endosymbionts are widespread in arthropods and are generally considered reproductive parasites, inducing various phenotypes including cytoplasmic incompatibility, parthenogenesis, feminization and male killing, which serve to promote their spread through populations. In contrast, *Wolbachia* infecting filarial nematodes that cause human diseases, including elephantiasis and river blindness, are obligate mutualists. DNA purification methods for efficient genomic sequencing of these unculturable bacteria have proven difficult using a variety of techniques. To efficiently capture endosymbiont DNA for studies that examine the biology of symbiosis, we devised a parallel strategy to an earlier array-based method by creating a set of SureSelect™ (Agilent) 120-mer target enrichment RNA oligonucleotides ("baits") for solution hybrid selection. These were designed from *Wolbachia* complete and partial genome sequences in GenBank and were tiled across each genomic sequence with 60 bp overlap. Baits were filtered for homology against host genomes containing *Wolbachia* using BLAT and sequences with significant host homology were removed from the bait pool. Filarial parasite *Brugia malayi* DNA was used as a test case, as the complete sequence of both *Wolbachia* and its host are known. DNA eluted from capture was size selected and sequencing samples were prepared using the NEBNext® Sample Preparation Kit. One-third of a 50 nt paired-end sequencing lane on the HiSeq™ 2000 (Illumina) yielded 53 million reads and the entirety of the *Wolbachia* genome was captured. We then used the baits to isolate more than 97.1 % of the genome of a distantly related *Wolbachia* strain from the crustacean *Armadillidium vulgare*, demonstrating that the method can be used to enrich target DNA from unculturable microbes over large evolutionary distances.

Keywords *Wolbachia* · Obligate endosymbiont · Target enrichment · NextGen sequencing · DNA capture · SureSelect™

S. Geniez · J. M. Foster · S. Kumar · B. E. Slatko (✉)
New England Biolabs, Inc., Ipswich, MA 01938, USA
e-mail: slatko@neb.com

S. Geniez · B. Moumen · D. Bouchon · P. Grève
Écologie et Biologie des Interactions, Équipe Écologie, Évolution, Symbiose, University of Poitiers UMR CNRS 7267, 86022 Poitiers, France

E. LeProust · O. Hardy · M. Guadalupe
Agilent Technologies Inc., Santa Clara, CA 95051, USA

S. J. Thomas · B. Boone · C. Hendrickson
HudsonAlpha Institute of Biotechnology, Huntsville, AL 35806, USA

1 Introduction

Whole genome sequencing (WGS) using Next-Generation sequencing (NGS) technologies is becoming a cost efficient and effective procedure for many research laboratories. While metagenomic projects (purification and DNA sequencing of mixed populations of bacteria *en masse*) are becoming common, purification methods for efficient sequencing of selected unculturable bacteria from amongst other DNA populations, such as symbiont hosts, have proven difficult. Purification strategies have generally used a variety of techniques including chemical gradients (Charles and Ishikawa 1999), pulsed-field gel (PFG) purification with or without whole genome amplification, library construction followed by gene walking, etc. (Foster et al. 2005; Iturbe-Ormaetxe et al. 2011; Mavingui et al. 2005;

Sun et al. 2001). Informatics methods (Kumar and Blaxter 2011) and cell sorting methods (Santos-Garcia et al. 2012) have also been used.

More recently, DNA capture methods using oligonucleotide probes have been used to isolate symbiont or parasitic DNA away from host DNA, providing significant enrichment for more efficient sequencing (Kent et al. 2011; Melnikov et al. 2011). Our goal was to further assess this capture methodology using SureSelect™ (Agilent) technology that could be applied to purify DNA from widely divergent unculturable bacteria (e.g. symbionts) away from other DNA sources as an initiation point for WGS, when some homologous or similar DNA sequence is known. We used the obligate endobacterium *Wolbachia* as the target species for capture, to eliminate host DNA from downstream library construction and subsequent DNA sequencing. *Wolbachia* are obligate alpha-proteobacteria closely related to rickettsial organisms and are present in insects, mites, crustaceans, spiders and parasitic filarial nematodes (Werren et al. 2008; Werren and Windsor 2000; Cordaux et al. 2012; Bouchon et al. 1998). It has been recently estimated that ~40 % of arthropod species are infected with *Wolbachia* (Zug and Hammerstein 2012), making it the most widespread intracellular bacterial species. Phylogenetic analyses using single genes or multilocus sequence typing currently describe seven well-resolved *Wolbachia* groups, designated as supergroup lineages (A to H, no group G) together with a number of additional lineages (Lo et al. 2002; Casiraghi et al. 2005; Casiraghi et al. 2003; Lo and Evans 2007; Baldo and Werren 2007; Bordenstein et al. 2009).

In 1999, the *Wolbachia* Genome Consortium planned to sequence various genomes representing the diversity of *Wolbachia* (Slatko et al. 1999). At the onset of our study, only four *Wolbachia* genomes had been completely sequenced and deposited in GENBANK, *w*Mel of *Drosophila melanogaster* (Wu et al. 2004), *w*Bm of *Brugia malayi* (Foster et al. 2005), *w*Pip of *Culex quinquefasciatus Pel* (Klasson et al. 2008), *w*Ri of *Drosophila simulans* (Klasson et al. 2009). Additional genomes are currently being sequenced/annotated (Werren et al. 2008) and partial sequences for several *Wolbachia* strains are available, e.g., http://www.ncbi.nlm.nih.gov/nuccore/?term=wolbachia. A common goal of many *Wolbachia* genome projects is to provide comparative genomic information for understanding mechanisms of genome evolution and mechanisms of host phenotype manipulation by *Wolbachia* (Cordaux et al. 2011; Saridaki and Bourtzis 2010; Werren 1997). In addition, the evolutionary analysis of *Wolbachia* prophages is of interest and capturing the symbiont DNA from divergent sources not only provides DNA sequence of *Wolbachia* but also can capture their prophages (Kent et al. 2011).

To test the selectivity of the designed baits, we used the *Wolbachia* *w*Bm strain (supergroup D) as a test case for the efficiency of *Wolbachia* DNA capture. The sequence of the 1.1 Mb *Wolbachia* genome is already known (Foster et al. 2005) and was included in the bait pool. Using these baits, we then attempted to selectively capture *Wolbachia* DNA from a phylogenetically distant strain, the feminizing *w*VulC strain from the isopod *Armadillidium vulgare* (supergroup B) (Cordaux et al. 2004) for which no genomic sequences were included in the bait design.

In the case of *w*VulC, DNA sequencing and comparative analysis from related strains with alternate phenotypes (cytoplasmic incompatibility and feminization) (Bouchon et al. 2008) will aid in the identification of the genetic basis of the phenotypic differences induced by *Wolbachia* within this group of isopod crustaceans. In the case of *w*Bm, it has been shown that this endosymbiont is a novel drug target against human filariasis (Slatko et al. 2010), and identification of worldwide variants will be useful. The described DNA capture technique should find application in drug discovery, evolutionary analysis and populational/ecological studies. For example, a similar approach with Roche Nimblegen arrays was utilized by the Bordenstein lab to isolate and analyze the *Wolbachia* genome and WO prophages of the strain *w*VitB from the parasitic wasp *Nasonia vitripennis* (Kent et al. 2011).

2 Materials and methods

2.1 *Wolbachia* strains

Two *Wolbachia* strains were analyzed in this study: the obligate *w*Bm strain from the nematode *B. malayi* (TRS Labs, Georgia, USA) (Foster et al. 2005) and the feminizing *w*VulC strain from the terrestrial isopod *A. vulgare* (maintained in the EES lab) (Cordaux et al. 2004). The 1.1 Mb *w*Bm genome sequence is known and the *w*VulC genome is in final assembly steps. The size of the *w*VulC genome has been estimated by PFG electrophoresis at 1.75 Mb (Bouchon et al. 2008) and the current sequence consists of 10 contigs of 1.66 Mb, which agrees with pulsed-field estimates (Liu et al. manuscript in preparation).

2.2 RNA bait library design

To design the targeted genome enrichment library, we created a SureSelect™ set of enrichment oligonucleotides for solution hybrid selection. The library of biotinylated complementary RNA baits was designed and synthesized by Agilent (Santa Clara, CA). RNA baits were utilized because of the stability of RNA-DNA hybrids in the selection process and the ease of their removal in subsequent steps. The

120-mer RNA bait library was created based upon compiling 11 *Wolbachia* complete and partial genome sequences found in GenBank (*Wolbachia* infections of *Muscidifurax uniraptor* (NZ_ACFP00000000.1), *Wuchereria bancrofti* (PRJNA43539), *Onchocerca volvulus* (PRJNA12625), *Drosophila willistoni* (PRJNA16739), *Drosophila simulans gdsi 540* (AAGC01000629.1), *Drosophila simulans* (AAGC00000000.1), *Drosophila melanogaster* (AE017196.1), *Drosophila ananassae gdan 143* (AAGB00000000), *Culex quinquefasciatus pel* (AM999887.1), *Culex quinquefasciatus jhb* (ABZA00000000), *Brugia malayi* (AE017321.1)). Bait sequences were tiled across each genomic sequence with 60 bp overlap and pooled, resulting in approximately 215,000 baits with about 207,000 unique sequences. Baits were then filtered for homology against select host genomes that contain *Wolbachia* (*Brugia malayi* (GCF_000002995.1), *Onchocerca volvulus* (ADBW00000000.1), *Wuchereria bancrofti* (ADBV00000000.1)) using BLAT, a high speed and more accurate BLAST-Like Alignment Tool with the ability to use an internal set of sequences for assembly and rapidly find high similarity sequences of relatively short length (Kent 2002). 6,000 baits with significant host homology were removed from the bait pool. The final bait count was 201,776 after removing BLAT rejects. Baits were also tested for uniqueness against two nematodes that do not harbor the endosymbiont (*Caenorhabditis elegans* (GCA_000002985.2), *Loa loa* (GCA_000183805.1)). This BLAT search did not produce significant hits and thus no additional baits were removed.

2.3 DNA extraction and preparation

Total DNA from the nematode *B. malayi* and from the isopod *A. vulgare* were extracted as described (Sambrook and Russell 2001; Bouchon et al. 1998). Quantification of the DNA samples was performed using a Nanodrop 1,000 spectrophotometer (Thermo Scientific) and the Qubit 2.0 fluorimeter (Invitrogen). DNA samples were normalized to 3 μg of DNA for each sequence capture protocol. DNA samples were sheared by sonication to an average length of 200 bp using a Covaris S1 then end repaired, followed by 3′dA addition using the NEBNext® Sample Preparation Kit (New England Biolabs). Adaptors were ligated onto the ends and following purification the DNA was PCR amplified (6 cycles) using indexed PCR primers and the Illumina InPE1.0 forward PCR primer. After purification, quality assays were performed using the Caliper GX (Life Science) and Bioanalyzer (Agilent) to determine the average fragment sizes and concentrations.

2.4 Capture and sequencing

Wolbachia DNA was captured from the prepared total DNA by hybridization to the biotinylated cRNA baits for 24 h at 65 °C, following the Agilent SureSelect™ protocol, but supplemented with custom blocking oligos complementary to the barcoded adaptors. Bound DNA was recovered using magnetic streptavidin beads, PCR amplified (12 cycles) using Illumina forward and reverse primers and purified. Sequencing samples were prepared using the NEBNext® Sample Preparation Kit (New England Biolabs). The library was paired-end sequenced on the Illumina HiSeq 2000 at HudsonAlpha, Inc.

2.5 Bioinformatics analysis

Sequence reads provided by HudsonAlpha were quality controlled using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and mapped against the complete *w*Bm genome and the partial *w*VulC genome using Bowtie (version 0.12.7) (Langmead et al. 2009) and Bowtie2 (version 2.0.0-beta5) (Langmead and Salzberg 2012). Mapping results were processed by SAMtools (Li et al. 2009) and visualized using Artemis (Carver et al. 2012). Assemblies were performed using Velvet assembler (Version 1.2.03) (Zerbino and Birney 2008). Optimization of the assembly was performed by using different k-mers (from 19 to 49 by steps of 2 bases); the optimal assembly was chosen considering the N50, the length of the longest contig and the total bases in the contig.

3 Results and discussion

3.1 *w*Bm sequence capture and sequencing

Nearly ninety-two million (91,724,826) reads from duplicated one third lanes on an Illumina HiSeq 2000 were produced from DNA captured by the oligo bait set. High quality metrics were obtained from the run (Q scores above 30 for bases 1–13 and above 40 for bases 14–50). Sensitive parameters of Bowtie2 were used to map these reads onto the genome sequence of the *Wolbachia* endosymbiont from the TRS strain of *Brugia malayi* (NC_006833). The entire *Wolbachia* genome was captured at a depth of over 3000×. These results for *w*Bm were anticipated as the capture oligos we designed were based on all *Wolbachia* genomic sequences in GenBank or RefSeq including those of *w*Bm (Fig. 1).

Depth of coverage was generally uniform but showed spikes over certain regions (Y axis, Fig. 1). These regions could either be repetitive in the genome or could be regions of duplicated lateral gene transfers into the host genome (Dunning Hotopp et al. 2007).

Only 5.20 % of the reads did not map to the *Wolbachia* *w*Bm genome. These unmapped reads were assembled by Velvet and the resulting 1,107 contigs were used to query the NCBI database by BLASTn (Fig. 2). 27.8 % of these

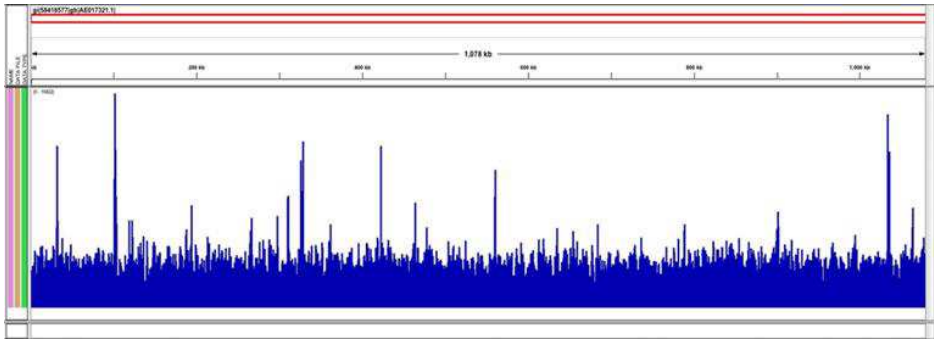


Fig. 1 The average number of mapped reads per genomic location averaged over a 25 nt window. The X axis represents a linear map of the *Wolbachia* genome from *B. malayi* and the Y axis (1–15,000) represents relative sequence coverage. The count file was generated from a BAM file using IGV Tools (Thorvaldsdottir et al. 2012). 100 % coverage was obtained

~5 % unmapped sequences BLAST to host nematode sequences and 51 % appeared to BLAST to *w*Bm *Wolbachia* sequence. These represent contigs assembled from reads that do not map to the reference genome (reads with more than 3 mismatches were not mapped) but have significant similarity to *w*Bm by BLAST. Unknown sequences (15.59 %) were investigated for GC content and they appeared to have an average GC content of 28.1 %, a value similar to *Wolbachia* and the host (Foster et al. 2005; Ghedin et al. 2007). Thus they can not be differentiated as to whether they represent *Wolbachia* or *B. malayi* genes or derive from other organisms with similarly low GC content.

3.2 *w*VulC sequence capture and sequencing

Over thirty-six million reads (one-third of a 50 nt paired-end sequencing lane on the Illumina HiSeq™ 2000) were produced from DNA capture by the oligo bait set. As with the *w*Bm mapping set, sequences were quality assayed. The entire dataset was of high quality (Q scores above 30 for bases 1–9 and above 40 for bases 10–50).

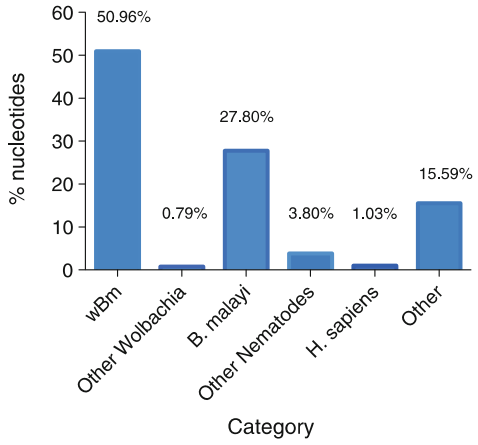


Fig. 2 Distribution of the 5.2 % reads not mapped to *w*Bm. Bar graph denotes % of nucleotides out of the total assembled contigs that show significant BLAST scores to *w*Bm genome, other *Wolbachia* the *B. malayi* host, other nematodes or *H. sapiens*. The "Other Nematodes" category corresponds to sequences with a significant BLAST score to other species of nematodes (mainly filarial nematodes). The "*H. sapiens*" category may represent DNA contamination. The "Other" category corresponds to sequences with no significant BLAST similarity to any sequences in the NCBI database

Different parameters of Bowtie were tested to improve the mapping of the total paired-end read dataset to the incompletely sequenced *w*VulC genome and the final mapping using Bowtie2 provided a total of 34,524,894 millions reads mapping to the reference genome (94.73 %).

This mapping covered 97.1 % of the partially known *w*VulC sequence (Fig. 3). Several regions appear to be over-represented in the data set, as with the *w*Bm mapping. These include the single copy DNA-directed RNA polymerase, beta/beta' subunit gene as well as the phage major capsid protein E gene, which is likely repetitive in the genome.

The unmapped reads (5.27 % of the total reads) were assembled by Velvet and the resulting 28 contigs were used to query the NCBI database by BLASTn. Of these contigs, 24 (94.78 % of the total unmapped reads) corresponded to other *Wolbachia* sequences. These sequences may represent *w*VulC genes not yet identified in the incomplete genome as the average GC% of these sequences is similar to the GC content of *Wolbachia* strain. None of the remaining 4 contigs had any significant match to any other sequences in the database.

As the wVulC genome is still in draft form, it cannot be excluded that there are errors in the preliminary sequence. To correct the sequence, mapping results were used to call variants and to detect errors, which were manually corrected. By this method, more than 300 base substitutions and over 30 indels have been corrected by use of the oligo bait-captured DNA sequence.

4 De novo assembly of the wVulC genome sequence

A primary goal of this study was to sequence an unknown genome and thus we performed a *de novo* assembly of the oligo-captured *w*VulC sequence data. After optimization of the Velvet parameters (k-mer=45, insert length=170, standard deviation for insert length=63 and minimum contig

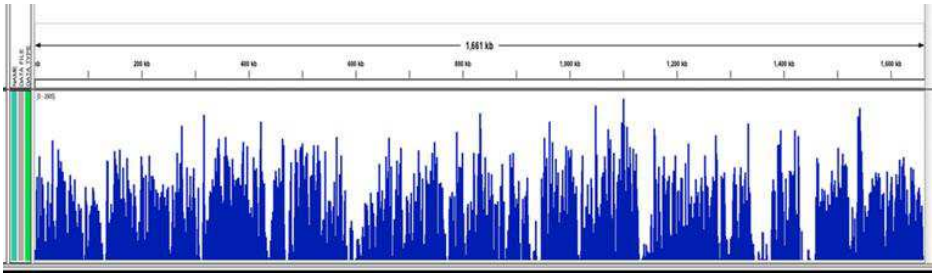


Fig. 3 The average number of mapped reads per genomic location averaged over a 25 nt window. The X axis represents a linear map of the artificially concatenated *Wolbachia* pseudo-contig from *A. vulgare* and the Y axis (1–3,000) represents relative sequence coverage. The count file was generated from a BAM file using IGV Tools (Thorvaldsdottir et al. 2012). 97.1 % coverage was obtained

length=200 bases), a manual assembly produced 523 contigs with N50 of 5826, a maximum length of 26,288 bases. This used over 87 % of the oligo-captured reads (31.81/36.45 million reads) providing a total length of 1.32 Mb, representing 75.5 % of the estimated length of the genome and 79.5 % of the partially known sequence.

To ensure that all the contigs built from the oligo-captured DNA were *w*VulC sequences, we aligned and ordered them to the previously sequenced *w*VulC genome using Mauve (Darling et al. 2010) and r2cat (Husemann and Stoye 2010). Overall, 95.7 % of the bases in contigs built from oligo-captured DNA were *w*VulC, which represents 76 % of the 1.66 Mb concatenated reference genome. 61 contigs did not match the incomplete *w*VulC genome (Fig. 4). Several of these contain sequences that show similarities with other *Wolbachia* strain sequences (0.1 %) while a contig of 16 kb appeared to be a region from the *Armadillidium vulgare* 18S ribosomal gene (1.22 %), certainly host DNA contamination (Fig. 4). Since the *w*VulC

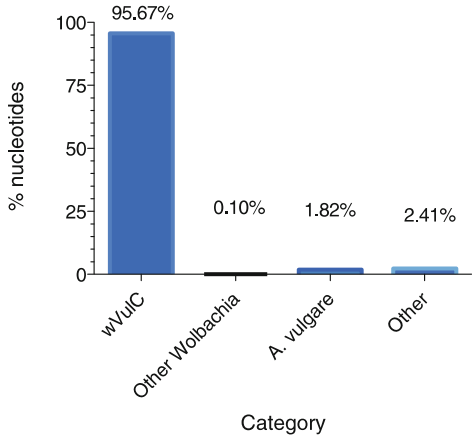


Fig. 4 Capture efficiency of the *w*VulC *Wolbachia* genome. Bar graph shows the distribution of nucleotides out of the total assembled contigs that show significant BLAST scores to the partial *w*VulC genome, other *Wolbachia* or the *A. vulgare* host. The "Other" category corresponds to sequences with no significant BLAST similarity to any sequences in the NCBI database

genome is not finished and gaps remain between the 10 contigs, analysis of the contigs built from oligo-captured DNA that matched other *Wolbachia* strains or didn't match any known sequence in NCBI (2.41 %) are being used to complete the reference genome sequence as they may represent yet unknown *w*VulC sequences.

Our results confirm and extend the results of Kent et al. (2011) and demonstrate that this method can potentially be used with any *Wolbachia* strain and on any endobacterium with either a reference genome or a highly similar sequence, providing an approach for isolating a significant fraction of symbiont DNA from host DNA for sequencing and comparative genomic analysis. In this context, of interest is the observation that both *w*Bm and *w*VulC were successfully purified away from host and mitochondrial DNA with the same bait library even though they are members of phylogenetically distant clades. In filarial nematodes, this approach will enable rapid isolation and analysis of *Wolbachia* strains from worldwide populations to identify polymorphisms related to drug discovery initiatives and evolutionary analysis of *Wolbachia* prevalence and distribution. In isopods, this approach will allow DNA isolation for genomic comparison of *Wolbachia* strains, which induce either various types of feminizing phenotypes or cytoplasmic incompatibility. For example, the isopod *A. vulgare* may harbor another *Wolbachia* strain, *w*VulM, which has a lower feminizing effect (~70 %) than the *w*VulC strain (~80 %) (Cordaux et al. 2004). In the *Porcellionides pruinosus* complex of species, 3 distinct feminizing *Wolbachia* strains have been identified which are present in ~60 % of populations where *Wolbachia* are present only in females or ~90 % in populations where both males and females are infected (Marcadé et al. 1999; Lefebvre and Marcadé 2005), a situation also encountered in populations of *Oniscus asellus* (Rigaud et al. 1999). Further, most *Wolbachia* strains infecting isopods induce a feminizing phenotype (Bouchon et al. 2008), however, 3 of them induce cytoplasmic incompatibility (Legrand et al. 1978; Moret et al. 2001). One of them, *w*Con, which infects *Cylisticus convexus*, is closely related to *w*VulC (Cordaux et al. 2012). Genomic comparisons

will thus be useful in helping decipher the evolution of *Wolbachia* and its various biological manifestations.

Acknowledgments This work was funded by New England Biolabs and by a grant from the Bill and Melinda Gates Foundation to the Liverpool School of Tropical Medicine as part of the A-WOL consortium (BES). We thank Drs William Jack, Tilde Carlow and Don Comb for their interest and support. Research in Poitiers was funded by the Centre National de la Recherche Scientifique (CNRS) and the French Ministère de l'Education Nationale, de l'Enseignement Supérieur et de la Recherche. The authors also thank Seth Bordenstein for helpful discussions in preparation of the manuscript.

References

Baldo L, Werren JH (2007) Revisiting *Wolbachia* supergroup typing based on WSP: spurious lineages and discordance with MLST. Curr Microbiol 55:81–87

Bordenstein SR et al (2009) Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say. Mol Biol Evol 26:231–241

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

Bouchon D, Cordaux R, Grève P (2008) Feminizing *Wolbachia* and the evolution of sex determination in isopods. In: Bourtzis K, Milller TA (eds) Insect symbiosis. CRC Press, Boca Raton, pp 273–294

Carver T et al (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28:464–469

Casiraghi M et al (2003) dnaA gene sequences from *Wolbachia* pipientis support subdivision into supergroups and provide no evidence for recombination in the lineages infecting nematodes. Parassitologia 45:13–18

Casiraghi M et al (2005) 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 151:4015–4022

Charles H, Ishikawa H (1999) Physical and genetic map of the genome of Buchnera, the primary endosymbiont of the pea aphid *Acyrthosiphon pisum*. J Mol Evol 48:142–150

Cordaux R et al (2004) Evidence for a new feminizing *Wolbachia* strain in the isopod *Armadillidium vulgare*: evolutionary implications. Heredity (Edinb) 93:78–84

Cordaux R, Bouchon D, Greve P (2011) The impact of endosymbionts on the evolution of host sex-determination mechanisms. Trends Genet 27:332–341

Cordaux R et al. (2012) Widespread Wolbachia infection in terrestrial isopods and other crustaceans. Zookeys 176:123–131

Darling AE, Mau B, Perna NT (2010) ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147

Dunning Hotopp JC et al (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317:1753–1756

Foster J et al (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3:e121

Ghedin E et al (2007) Draft genome of the filarial nematode parasite *Brugia malayi*. Science 317:1756–1760

Husemann P, Stoye J (2010) r2cat: synteny plots and comparative assembly. Bioinformatics 26:570–571

Iturbe-Ormaetxe I et al (2011) A simple protocol to obtain highly pure *Wolbachia* endosymbiont DNA for genome sequencing. J Microbiol Methods 84:134–136

Kent WJ (2002) BLAT-the BLAST-like alignment tool. Genome Res 12:656–664

Kent BN et al (2011) Complete bacteriophage transfer in a bacterial endosymbiont (*Wolbachia*) determined by targeted genome capture. Genome Biol Evol 3:209–218

Klasson L et al (2008) Genome evolution of *Wolbachia* strain wPip from the Culex pipiens group. Mol Biol Evol 25:1877–1887

Klasson L et al (2009) The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci U S A 106:5725–5730

Kumar S, Blaxter ML (2011) Simultaneous genome sequencing of symbionts and their hosts. Symbiosis 55:119–126

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Langmead B et al (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Lefebvre, F, Marcadé I (2005) New insights in the *Porcellionides pruinosus* complex (Isopoda, Oniscidea): biological, behavioural, and morphological approaches. Crustacceana 78:465–480

Legrand JJ, Martin G, Artault JC (1978) Correlation between the presence of a bacterial symbiont in oocytes of *Parcellio dilatatus* petiti, and the sterility of the cross P.d petiti male x P.d dilatatus female. Arch Inst Pasteur Tunis 55:507–504

Li H et al (2009) The sequence alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

Lo N, Evans TA (2007) Phylogenetic diversity of the intracellular symbiont *Wolbachia* in termites. Mol Phylogenet Evol 44:461–466

Lo N et al (2002) How many *Wolbachia* supergroups exist? Mol Biol Evol 19:341–346

Marcadé II et al (1999) Mitochondrial DNA variability and *Wolbachia* infection in two sibling woodlice species. Heredity (Edinb) 83(Pt 1):71–78

Mavingui P et al (2005) Efficient procedure for purification of obligate intracellular *Wolbachia pipientis* and representative amplification of its genome by multiple-displacement amplification. Appl Environ Microbiol 71:6910–6917

Melnikov A et al (2011) Hybrid selection for sequencing pathogen genomes from clinical samples. Genome Biol 12:R73

Moret Y, Juchault P, Rigaud T (2001) *Wolbachia* endosymbiont responsible for cytoplasmic incompatibility in a terrestrial crustacean: effects in natural and foreign hosts. Heredity 86:325–322

Rigaud T, Moreau J, Juchault P (1999) *Wolbachia* infection in the terrestrial isopod Oniscus asellus: sex ratio distortion and effect on fecundity. Heredity 83:469–475

Sambrook J, Russell DW (2001) Isolation of high-molecular-weight DNA from mammalian cells using proteinase K and phenol. In: Sambrook J, Russell DW (eds) Molecular cloning a laboratory manual, Vol 1, 3rd ed. Cold Spring Harbor Press, Cold Spring Harbor, pp 6.4–6.12

Santos-Garcia D, Peris-Bondia F, D'Auria G, Moya A, Silva F, Latorre A (2012) Enrichment of insect samples with bacterial symbionts using flow cytometry, ABSTRACT 7th International Symbiosis Society Congress

Saridaki A, Bourtzis K (2010) *Wolbachia*: more than just a bug in insects genitals. Curr Opin Microbiol 13:67–72

Slatko BE et al (1999) The *Wolbachia* genome consortium. Microb Comp Genomics 4:161–165

Slatko BE, Taylor MJ, Foster JM (2010) The *Wolbachia* endosymbiont as an anti-filarial nematode target. Symbiosis 51:55–65

Sun LV et al (2001) Determination of *Wolbachia* genome size by pulsed-field gel electrophoresis. J Bacteriol 183:2219–2225

Thorvaldsdottir H, Robinson JT, Mesirov JP (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. doi:10.1093/bib/bbs017

Werren JH (1997) Biology of *Wolbachia*. Annu Rev Entomol 42:587–609

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

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

Wu M et al (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2:E69

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829

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

B. Capture optimization

As demonstrated in the Symbiosis manuscript, this targeted genome enrichment method is efficient in capturing *Wolbachia* DNA from a distantly related strains but can still be improved. First, to perfect the efficiency of the capture: about 3% of the known *w*VulC genome sequence was not captured presumably because the homologous sequences were not present in the probe sets. Further, and more importantly, it is necessary to integrate into the approach any newly available sequencing tools such as long read sequencing by Illumina MiSeq, PacBio and new DNA library preparation kits (e.g., the NEBNext Ultra Library), which is more efficient than previously commercialized kits.

The optimization process included 4 main steps:

1- The capture library was supplemented with the newly published *Wolbachia* genome sequences and the unfinished *w*VulC genome sequences.
2- The sequencing library was prepared with the NEBNext Ultra Library preparation kit with only 1.5 µg of starting DNA
3- The genomic DNA was sheared at 500 bp
4- Sequencing was performed on a MiSeq Illumina sequencer with 250bp paired-end reads.

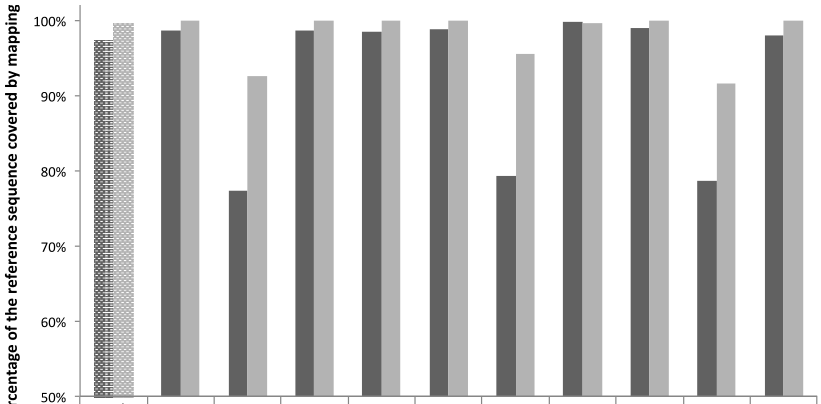
These optimizations allowed us to clearly improve the capture. To measure the efficiency of this optimization, a mapping on the reference *w*VulC contigs was performed with both the HiSeq Illumina reads from the 1st capture (labeled *w*VulC_HiSeq in Table 5) and the MiSeq Illumina reads from the 2nd capture (labeled *w*VulC_MiSeq in Table 5). 38,360 bp (85.5%) of the missing known sequence from the 1st capture was then captured by the optimized procedure leading from 97.3% to 99.6% of captured sequence.

Table 5: Comparison of the read mapping between the initial capture and the optimized capture.

	Number of reads	Mapped Reads	Mean Coverage	Mean insert size	wVulC reference genome (1.66Mb)	contig 1 (581kb)	contig 2 (44kb)	contig 3 (29kb)	contig 4 (38kb)	contig 5 (218kb)	contig 6 (38kb)	contig 8 (10kb)	contig 8 (179kb)	contig 9 (17kb)	contig 10 (510kb)
wVulC_HiSeq	72,894,720	68,914,878 94.54%	1035.41x	170.92	97.30%	98.63%	77.34%	98.60%	98.45%	98.83%	79.24%	99.73%	99.05%	78.68%	98.00%
		number of non covered base			*44,884*	*7,987*	*9,868*	*410*	*580*	*2,543*	*7,938*	*27*	*1,697*	*3,635*	*10,199*
wVulC_MiSeq	6,158,542	5,757,186 93.48%	836.84x	378.49	99.608%	99.991%	92.579%	99.993%	99.891%	99.996%	95.538%	99.578%	100%	91.658%	99.997%
		number of non covered base			*6,524*	*55*	*3,232*	*2*	*41*	*8*	*1,706*	*42*	*0*	*1,422*	*16*

The still missing sequences (0.4%) can be explained by phage regions (e.g. contig 6 & 9) or repeat sequences (Figure 20) which encompass respectively 12.2% (7 complete or

partial prophages) and 15.9% (372 sequences) of the *w*VulC genome. Seven contigs are almost totally covered (contigs 1, 3, 4, 5, 7, 8, 10; 0 to 55 uncovered bases) and 3 contigs (contigs 2, 6, 9; 1,422 to 3,232 uncovered bases) show larger uncovered sequences. The first 1kb of contig 2, which does not present any coding sequence, is not covered as well as a large phage sequence, which is poorly covered. The main uncovered sequence of contig 6 is a 1kb sequence coding for 4 transposases Tn5, proteins involved in recombination. The short 17kb contig 9, only coding for phage proteins, shows a global poor coverage. Phage sequences and repeat region are sequences that evolve quickly with a lot of mutations and recombination, which induce a high diversity of these particular sequences (Kent and Bordenstein 2010). As the reads with more than 3 SNPs are not mapped on the reference sequence, high variant sequences leads to a low coverage.



	wVulC reference genome (1.66Mb)	contig 1 (581kb)	contig 2 (44kb)	contig 3 (29kb)	contig 4 (38kb)	contig 5 (218kb)	contig 6 (38kb)	contig 8 (10kb)	contig 8 (179kb)	contig 9 (17kb)	contig 10 (510kb)
wVulC_HiSeq	97.30%	98.63%	77.34%	98.60%	98.45%	98.83%	79.24%	99.73%	99.05%	78.68%	98.00%
wVulC_MiSeq	99.608%	99.991%	92.579%	99.993%	99.891%	99.996%	95.538%	99.578%	100%	91.658%	99.997%

Figure 20: Mapping distribution across the reference genome *w*VulC and the 10 separated contigs of the reference (Contig lengths are indicated in brackets). The table indicates the percentage of base in the reference sequence covered by at least one read.

Most of the *w*VulC sequences were captured after optimization of the bait library and a good example of the increased capture efficiency is the biotin operon. This operon was not captured with the first bait library as shown by the absence of coverage (Figure 21 # A) but was fully captured with an average of 1,000x coverage with the second round of capture (Figure 21 # B). Because *w*VulC is the first sequenced *Wolbachia* genome to harbor all the genes of the biotin pathway (*BioB*, *BioF*, *BioH*, *BioC*, *BioD*, *BioA*), these 5.5kb were absent from the 1st library of RNA baits, which explains these sequences were not initially captured and sequenced. This example illustrates a potential limit of this method to look at non-related genomes, but with the increasing number of

published *Wolbachia* genomes, reiteration of probe constructions will make this method even more efficient.

Figure 21: Reads mapping before (A) and after (B) capture optimization on the biotin operon. Reads are illustrated in the upper part in blue and green (Artemis screenshot (Carver et al. 2012)).

We have demonstrated that this method is a robust and very efficient method for isolation of endosymbiont DNA from host DNA. Concerning *Wolbachia*, this method is a powerful tool to potentially sequence all *Wolbachia* supergroup strains. We used this method to efficiently capture 8 other *Wolbachia* strains from isopods including the strain *w*Pru from *P. pruinosus*, which is the *Wolbachia* strain from isopod the most distant from *w*VulC according to (Cordaux et al. 2012) (Figure 18).

II. Whole genome sequencing

Using this *Wolbachia* DNA purification method, two round of sequencing were performed, first with the initial capture method (Geniez et al. 2012), then with the optimized procedure (I.B p74) (Table 6).

In addition to the two strains *w*Bm and *w*VulC described previously, the first set of *Wolbachia* genomes sequenced was composed of the strains *w*Albu, *w*Bre, *w*Dil, *w*Pet, *w*Pru, *w*VulM from the isopods *A. album*, *H. brevicornis*, *P. dilatatus dilatatus*, *P. dilatatus petiti*, *P. Pruinosus* and *A. vulgare*, respectively. These DNA were fragmented in 300 bp sequences, captured with the first set of RNA baits and sequenced by HiSeq Illumina using 50bp paired end reads. The extracted DNA samples have been sent to HudsonAlpha (Huntsville, AL, USA) for capture and sequencing.

The second set of *Wolbachia* genomes sequenced was composed of the strains *w*VulC, *w*Ase, *w*Dil, *w*Pru, *w*Gla and *w*Nas from the isopods *A. vugare*, *O. Asellus*, *P. dilatatus dilatatus*, *P. Pruinosus*, *B. Glaber* and *A. nasatum*, respectively. These DNA were fragmented in 500 bp sequences, captured with the optimized second set of RNA baits and sequenced by MiSeq Illumina using 250 bp paired end reads. For this second set, I performed all the steps of these experiments and the NEB sequencing core performed the MiSeq sequencing. Unfortunalty, DNA from *w*Gla and from *w*Nas failed sequencing due to poor quality DNA samples. The DNA of *w*Nas is currently being sequenced by HudsonAlpha.

Table 6: Summary of the DNA isolation and sequencing procedures.

	First set	Second set
Capture library	initial	optimized
DNA fragment size	300 bp	500 bp
Sequencer	HiSeq	MiSeq
Reads	50bp paired end reads	250 bp paired end reads
Strains	*w*Albu_HiSeq	*w*Ase_MiSeq
	*w*Bre_HiSeq	*w*Dil_MiSeq
	wDil_HiSeq	*w*Pru_MiSeq
	*w*Pet_HiSeq	*w*VulC_MiSeq
	wPru_HiSeq	
	*w*VulC_HiSeq	
	*w*VulM_HiSeq	

Because the NGS technologies used in this study produced a large amount of data but in short reads, producing a complete genomic sequence is very difficult due to the shortness of the reads and the high rate of repetitive sequences in *Wolbachia* genomes.

In order to produce a first complete genome of a *Wolbachia* strain from isopod, the DNA of the strain *w*VulC was also prepared for PacBio sequencing.

A. Genome coverage

The first step to perform to ensure that *Wolbachia* DNA sequences were actually obtained is to map the raw sequencing reads on a reference genome, the *w*VulC genome, which is the only *Wolbachia* strain from isopod already sequenced (Liu et al. 2013).
The mapping consists on the alignment of raw or filtered sequencing reads on a reference sequence to ensure the quality of the sequencing by checking two parameters: the mapped read depth or base coverage, which is the total number of bases sequenced and aligned at a given reference base position; and the percentage of genome mapped or genome coverage, which is the percentage of the reference genome that is successfully mapped after alignment.

As a result of the *Wolbachia* DNA capture, the majority of the Illumina reads obtained mapped to the *w*VulC reference genome for all the sequenced strains for both sets of sequencing. For the seven isopod strains captured during the first round and sequenced by HiSeq, about 70 million 50bp paired-end reads were obtained and about 5-6 millions of 250bp paired-end reads were obtained for the four strains secondly captured and sequenced by MiSeq (Table 7).
Quality controls were performed on the raw reads by FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The entire dataset was of high quality for the first round of sequencing by HiSeq and two of the second round sequencings (*w*Dil_MiSeq and *w*Pru_MiSeq) had a few poor quality reads due to a high number of unknown nucleotide sequence calls ('N') in the sequences. These reads were filtered out with fastx-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) to only keep the high quality dataset for further assembly analysis. For both genomes, about half of the forward reads (53% for *w*Dil and 43% for *w*Pru) and most of the reverse reads (91% for *w*Dil and 86% for *w*Pru) were filtered out using a high stringency cut off (100% of bases with a minimum quality score of 20) suitable for efficient *de novo* assembly.
To ensure that the low percentage of raw *w*Pru_HiSeq reads that mapped *w*VulC is not a consequence of a bad quality capture but to a higher sequence difference, a mapping of the *w*Bm reads on the *w*VulC reference genome was performed to confirm the previous hypothesis. 97% of these *w*Bm reads mapped to the *w*Bm genome with 100% coverage

but only 0.39% of these same reads mapped *w*VulC genome with 5% of genome coverage, indicating that a low coverage is not necessarily a hint of bad capture quality.

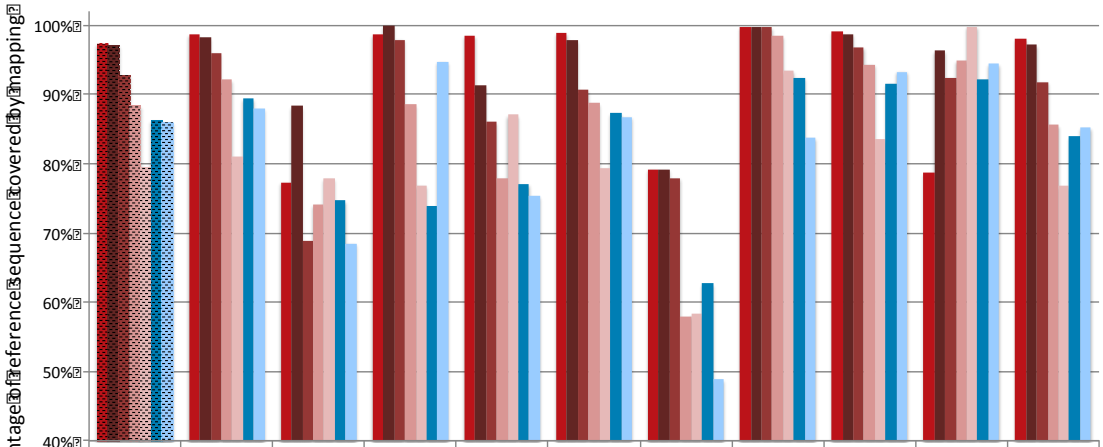
Table 7: Reads mapping on the incomplete *w*VulC and complete *w*Bm reference genomes. The percentages indicate the proportion of bases covered by at least one read. Table A refers to the mapping of the HiSeq reads. Table B refers to the mapping of the MiSeq reads.

A	Number of reads	Mapped Reads	Mean Coverage	Mean insert size	*w*VulC reference genome (1.66Mb)
*w*VulC_HiSeq	72,894,720	68,914,878 / 94.54%	1,035x	171	97.30%
*w*Albu_HiSeq	71,789,634	44,639,996 / 62.18%	1,341x	223	92.74%
*w*Bre_HiSeq	71,322,214	45,614,467 / 63.96%	1,370x	193	88.34%
*w*Dil_HiSeq	74,953,452	45,849,720 / 61.17%	1,377x	182	86.19%
*w*Pet_HiSeq	61,568,590	43,787,206 / 71.12%	1,315x	184	85.99%
*w*Pru_HiSeq	77,645,824	33,651,632 / 43.34%	1,011x	187	79.48%
*w*VulM_HiSeq	64,625,816	45,311,090 / 70.11%	1,361x	184	97.06%
*w*Bm_HiSeq	91,724,826	354,236 / 0.39%	11x	90	5.44%
					*w*Bm reference genome (1.1Mb)
wBm	91,724,826	88,973,082 / 97%	3,000x		100.00%

B	Number of reads	Mapped Reads	Mean Coverage	Mean insert size	*w*VulC reference genome (1.66Mb)
*w*VulC_MiSeq	6,158,542	5,757,186 / 93.48%	837x	378	99.61%
*w*Ase_MiSeq	1,056,048	755,143 / 71.51%	126x	334	84.52%
*w*Dil_MiSeq	5,685,220	291,308 / 5.12%	39x	362	85.82%
*w*Pru_MiSeq	4,542,836	3,183,986 / 70.09%	418x	340	80.38%

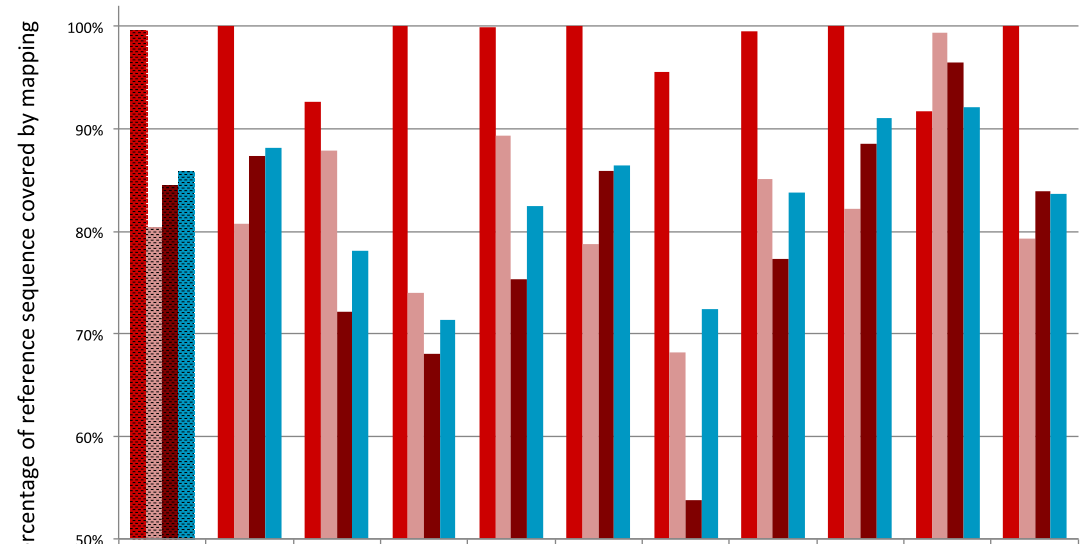
With respect to MiSeq reads, mapping results are less consistent in term of number of reads (between 1 and 6.1 million reads) and coverage (between 39x and 836x). The exceptionally low coverage of *w*Dil reads (39x) can be explained by the low percentage of reads that mapped to the reference genome (only 5% of the reads), likely caused by a lower efficiency of the capture (this decrease in efficiency is well known for this in-solution capture method when evaporation occurs during hybridization). However, 39x coverage is more than enough to create a draft genome; a Sanger genome sequence yielded an average about 8x coverage. Actually, for the all four genomes sequenced by MiSeq, the sequencing reads mapped to the reference *w*VulC genome with a high percentage, reflecting the global high quality of the sequencing (Table 7).

The mapping on the 10 separated *w*VulC reference contigs illustrates the partially uneven distribution of the mapped reads. Non-surprisingly, less conserved regions such as phage regions (e.g. contig 6) were those with the lower coverage (Figure 22 and Figure 23). As expected from the different technologies, MiSeq sequencings yield a lower coverage (40 to 800x) than HiSeq sequencings (~ 1,200x coverage) but the reads are longer (~350bp) which allows more accurate assemblies.



	wVulC reference genome (1.66Mb)	contig 1 (581kb)	contig 2 (44kb)	contig 3 (29kb)	contig 4 (38kb)	contig 5 (218kb)	contig 6 (38kb)	contig 8 (10kb)	contig 8 (179kb)	contig 9 (17kb)	contig 10 (510kb)
wVulC	97.30%	98.63%	77.34%	98.60%	98.45%	98.83%	79.24%	99.73%	99.05%	78.68%	98.00%
wVulM	97.06%	98.20%	88.28%	99.95%	91.26%	97.80%	79.21%	99.65%	98.76%	96.31%	97.17%
wAlb	92.74%	95.97%	68.78%	97.93%	86.01%	90.61%	77.97%	99.66%	96.78%	92.43%	91.79%
wBre	88.34%	92.11%	74.14%	88.54%	77.98%	88.70%	57.98%	98.42%	94.32%	94.88%	85.61%
wPru	79.48%	80.97%	77.90%	76.90%	87.04%	79.45%	58.30%	93.45%	83.48%	99.83%	76.75%
wDil	86.19%	89.42%	74.70%	73.83%	77.12%	87.33%	62.74%	92.47%	91.60%	92.24%	83.90%
wPet	85.99%	87.94%	68.52%	94.70%	75.35%	86.76%	48.94%	83.75%	93.31%	94.38%	85.21%

Figure 22: Mapping distribution across the reference genome *w*VulC and the 10 separated contigs of the reference (length of the contigs indicated in brackets). The table indicates the percentage of nucleotides in the reference sequence covered by at least one read. In red are the feminizing strains and in blue are the CI strains.



	wVulC reference genome (1.66Mb)	contig 1 (581kb)	contig 2 (44kb)	contig 3 (29kb)	contig 4 (38kb)	contig 5 (218kb)	contig 6 (38kb)	contig 8 (10kb)	contig 8 (179kb)	contig 9 (17kb)	contig 10 (510kb)
wVulC_MiSeq	99.61%	99.99%	92.58%	99.99%	99.89%	99.996%	95.54%	99.578%	100.00%	91.66%	99.997%
wPru_MiSeq	80.38%	80.71%	87.87%	74.01%	89.38%	78.82%	68.23%	85.09%	82.14%	99.35%	79.29%
wAse_MiSeq	84.52%	87.38%	72.16%	68.06%	75.34%	85.94%	53.81%	77.32%	88.51%	96.49%	83.97%
wDil_MiSeq	85.82%	88.16%	78.07%	71.36%	82.44%	86.43%	72.36%	83.74%	91.07%	92.05%	83.65%

Figure 23: Mapping distribution across the reference genome *w*VulC and the 10 separated contigs of the reference (length of the contigs indicated in brackets). The table indicates the percentage of nucleotides in the reference sequence covered by at least one read. In red are the feminizing strains and in blue are the CI strains.

In conclusion, with the optimized capture procedure, the coverage of the reference genome in term of percentage of bases mapped by at least one read is better, meaning than more sequences of the *Wolbachia* genomes were captured and sequenced (an exception is *w*Dil, for which an obvious experimental capture issue occured).

However, for the three other sequenced genomes, the proportion of reads that mapped the reference is almost the same (Table 7) meaning that some contaminants were still sequenced, but these sequences can easily be filtered out before the computational analyses due to a different CG percentage or highly different sequences from other *Wolbachia* genomes.

Since the mapping was performed on the same reference genome *w*VulC, we were expecting that the phylogenetically closer strains presented the higher base coverage and vice versa, particularly for optimized capture since the *w*VulC sequences were included in the bait library. As illustrated in Figure 23, this statement is correct for the *w*VulC_MiSeq reads, which have the highest base coverage (99.6%). However, the three other strains, which do not have the same phylogenetic distance with *w*VulC (Figure 18) presented about the same percentage (~80-85%). With these whole-genome sequences, a phylogenomic analysis at the genome level is now possible using these strains and will probably generate a slightly different tree that may explain these surprising base coverages.

B. Genome assembly

1 - *De novo* Assembly

Reads from HiSeq sequencing were assembled in scaffolds with Velvet (Zerbino and Birney 2008) after optimization of the parameter with VelvetOptimiser (Appendix 1). The assemblies of the seven genomes sequenced by HiSeq produced between 523 and 1755 scaffolds for an average number of nucleotides of 1.55Mb (Table 8).

Table 8: Velvet *de novo* assembly summary

	number of scaffolds	number of nucleotides in scaffold	number of gaps/number of nucleotides forming gaps
wAlb	1755	1,724,208	418/26,205
wBre	786	1,253,219	860/25,004
wDil	1743	1,854,492	642/34,146
wPet	856	1,466,213	442/23,807
wPru	1282	1,659,900	303/12,422
wVulC	523	1,289,143	179/6,741
wVulM	1325	1,582,488	695/53,076

As quality assay of the assemblies revealed a high number of gaps filled by poly "N" sequences (Table 8), *de novo* assemblies were optimized by gap closing using the software SSPACE and GapFiller. These optimizations resulted in better assemblies for all the genomes, with a lower number of contigs and a higher number of bases in scaffolds (Table 9). Since we also sequenced the *w*Dil and *w*Pru genomes by MiSeq technology during the second round of capture, we did not perform the assembly optimization for these strains nor the annotation. Further analyses were performed on the wDil_MiSeq and wPru_MiSeq data since they are more complete.

Table 9: Optimized assembly summary after gap closing.

	Final assembly		
strains	Post Gapfiller (final assembly)	number of scaffolds	number of nucleotides in scaffold
*w*Albu	169 gaps / 178 N'	1679	1,761,570
*w*Bre	143 gaps / 148 N'	734	1,361,676
*w*Pet	154 gaps / 205 N'	794	1,484,643
*w*VulC	171 gaps / 627 N'	354	1,302,723
*w*VulM	150 gaps / 796 N'	1301	1,597,374

Qualities of the sequences of the optimized assemblies were also assayed by gene finding. As shown in Table 10, the optimized assemblies presented all the tRNA and the tmRNA as opposed to those observed in initial *de novo* assemblies. Only the second copy of the serine tRNA is missing but this second copy has been only identified in *w*VulC. All the optimized assemblies presented the same pattern with 34 tRNA and one tmRNA identified.

Table 10: Summary of the tRNA annotated on the *w*VulC reference genome published (*w*VulC_ref), the Velvet de novo assembly (*w*VulC_assembly_velvet) and the optimized assembly (*w*VulC_optimized assembly).

wVulC_ref		wVulC_assembly_velvet		wVulC_optimized_assembly	
tmRNA*	ANDNFAAEDNVALAA*	tmRNA*	ANDNFAAEDNVALAA*	tmRNA*	ANDNFAAEDNVALAA*
tRNA-Ala	(tgc)	missing		tRNA-Ala	(tgc)
tRNA-Arg	(acg)	tRNA-Arg	(acg)	tRNA-Arg	(acg)
tRNA-Arg	(ccg)	tRNA-Arg	(ccg)	tRNA-Arg	(ccg)
tRNA-Arg	(cct)	missing		tRNA-Arg	(cct)
tRNA-Arg	(tct)	tRNA-Arg	(tct)	tRNA-Arg	(tct)
tRNA-Asn	(gtt)	tRNA-Asn	(gtt)	tRNA-Asn	(gtt)
tRNA-Asp	(gtc)	tRNA-Asp	(gtc)	tRNA-Asp	(gtc)
tRNA-Cys	(gca)	tRNA-Cys	(gca)	tRNA-Cys	(gca)
tRNA-Gln	(ttg)	missing		tRNA-Gln	(ttg)
tRNA-Glu	(ttc)	tRNA-Glu	(ttc)	tRNA-Glu	(ttc)
tRNA-Gly	(gcc)	tRNA-Gly	(gcc)	tRNA-Gly	(gcc)
tRNA-Gly	(tcc)	tRNA-Gly	(tcc)	tRNA-Gly	(tcc)
tRNA-His	(gtg)	missing		tRNA-His	(gtg)
tRNA-Ile	(gat)	missing		tRNA-Ile	(gat)
tRNA-Leu	(caa)	missing		tRNA-Leu	(caa)
tRNA-Leu	(gag)	tRNA-Leu	(gag)	tRNA-Leu	(gag)
tRNA-Leuc	(taa)	tRNA-Leu	(taa)	tRNA-Leu	(taa)
tRNA-Leu	(tag)	tRNA-Leu	(tag)	tRNA-Leu	(tag)
tRNA-Lys	(ttt)	missing		tRNA-Lys	(ttt)
tRNA-Met	(cat)	missing		tRNA-Met	(cat)
tRNA-Met	(cat)	tRNA-Met	(cat)	tRNA-Met	(cat)
tRNA-Met	(cat)	tRNA-Met	(cat)	tRNA-Met	(cat)
tRNA-Phe	(gaa)	tRNA-Phe	(gaa)	tRNA-Phe	(gaa)
tRNA-Pro	(tgg)	tRNA-Pro	(tgg)	tRNA-Pro	(tgg)
tRNA-Ser	(cag)	tRNA-Ser	(cag)	tRNA-Ser	(cag)
tRNA-Ser	(gct)	missing		tRNA-Ser	(gct)
tRNA-Ser	(gct)	missing		missing	
tRNA-Ser	(gga)	tRNA-Ser	(gga)	tRNA-Ser	(gga)
tRNA-Ser	(tga)	tRNA-Ser	(tga)	tRNA-Ser	(tga)
tRNA-Thr	(ggt)	tRNA-Thr	(ggt)	tRNA-Thr	(ggt)
tRNA-Thr	(tgt)	tRNA-Thr	(tgt)	tRNA-Thr	(tgt)
tRNA-Trp	(cca)	tRNA-Trp	(cca)	tRNA-Trp	(cca)
tRNA-Tyr	(gta)	tRNA-Tyr	(gta)	tRNA-Tyr	(gta)
tRNA-Val	(gac	tRNA-Val	(gac)	tRNA-Val	(gac)
tRNA-Val	(tac)	tRNA-Val	(tac)	tRNA-Val	(tac)

Reads from MiSeq sequencing (250 bp paired-end reads) were assembled using Newbler (Roche), which produces assemblies in contigs (Table 11). No optimization was performed on wVulC_MiSeq, wDil_MiSeq and wAse_MiSeq assemblies since contigs, in contrast with scaffolds, do not contain gaps and none of these contigs presented unknown 'N' bases. For wPru_MiSeq, the assembly generated 567 contigs, which contain more than 4.5 kb of 'N', revealing a lower quality of assembly. Except for wPru_MiSeq, all the assemblies of quality filtered MiSeq reads contained the 34 tRNAs. Actually, the wPru_MiSeq assembly presented 91 tRNAs, which confirmed the bad quality of the assembly. Even in combination with the previous HiSeq reads, no generated assemblies for *w*Pru were usefull for further comparative analysis. Therefore this strain was omitted for the following analyses and should be resequenced.

2 - *Wolbachia* sequence isolation and genome annotation

Because the most common contaminant of endosymbiotic genome sequencing is host mitochondrial DNA, a specific clean up was performed by BLAT against a database I established with all known isopod mitochondrial sequences. Using this procedure, host mitochondrial sequences were removed from the assemblies. All assemblies but the one from *H. brevicornis* presented sequences of partial or complete mitochondrial genomes (Table 11). To validate the nature of these sequences, they were aligned on the published *A. vulgare* mitochondrial genome using Mauve (Darling et al. 2010) (Figure 24). These mitochondrial sequences showed a nice but imperfect alignment attesting of close but not identical isopod species. These genomes, which are by-products of the *Wolbachia* genome sequencings, will be used in a side-project in collaboration with Dr. Isabelle Marcadé (EES). After closure of these 6 mitochondrial genomes by PCR sequencing, a phylogenomic analysis will be performed and the genomes investigated for heteroplasmy (Doublet et al. 2008).

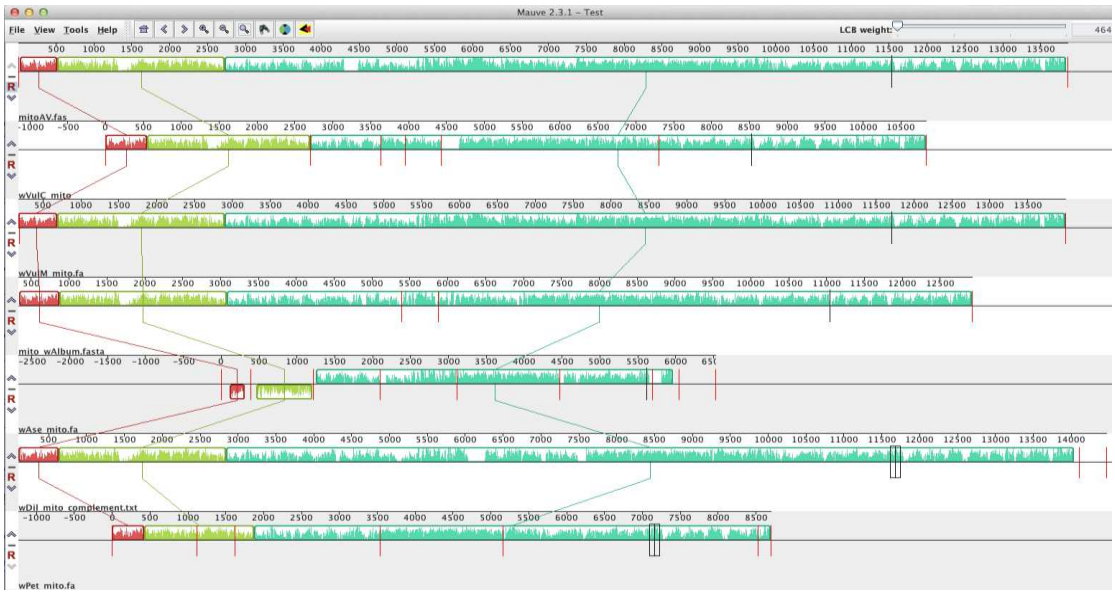


Figure 24: Mitochondrial genome alignment using Mauve.

After removing these mitochondrial sequences from the assemblies, *Wolbachia* sequences were isolated from contaminants by BLAT against an internal database containing all the published sequences of *Wolbachia* (Table 11; Appendix 2).

Scaffolds identified as *Wolbachia* sequences were analyzed with Prodigal (Prokaryotic Dynamic Programming Gene finding Algorithm (Hyatt et al. 2010) to isolate protein-coding DNA sequences (CDS) and an annotation draft was generated using the pipeline Prokka (Prokaryotic Genome Annotation System - http://vicbioinformatics.com/) (Table 11).

Table 11: Scaffold distribution of the Velvet assemblies of the HiSeq reads (top) and the Newbler assemblies of the MiSeq reads (bottom).

	Final assembly			mitochondrial DNA		contaminant DNA			*wolbachia*-like DNA				
strains	Velvet assembly (optimized assembly)	number of scaffolds	number of nucleotides in scaffold	number of scaffolds	number of nucleotides in scaffold	number of scaffolds	number of nucleotides in scaffold	Percentage of total nucleotides in scaffold	number of scaffolds	N50 / number of scaffolds larger than N50	number of nucleotides in scaffold	Percentage of total nucleotides in scaffold	number of predicted proteins (prodigal)
*w*Albu_HiSeq	169 gaps / 178 N'	1679	1,761,570	3	12,923	386	164,107	9.3%	**1290**	**6,511 / 60**	**1,585,022**	**90.0%**	**2,615**
*w*Bre_HiSeq	143 gaps / 148 N'	734	1,361,676	0	0	59	20,956	1.5%	**675**	**5,422 / 66**	**1,340,868**	**98.5%**	**1,834**
*w*Pet_HiSeq	154 gaps / 205 N'	794	1,484,643	6	8,695	285	134,599	9.1%	**503**	**8,567 / 46**	**1,341,349**	**90.3%**	**1,665**
*w*VulC_HiSeq	171 gaps / 627 N'	354	1,302,723	0	0	55	51,796	4.0%	**299**	**8,604 / 45**	**1,250,927**	**96.0%**	**1,450**
*w*VulM_HiSeq	150 gaps / 796 N'	1301	1,597,374	3	13,997	765	268,494	16.8%	**533**	**7,275 / 50**	**1,314,883**	**82.3%**	**1,684**

	Final assembly			mitochondrial DNA		contaminant DNA			*wolbachia*-like DNA				
strains	Nembler assembly	number of contigs	number of nucleotides in contig	number of contigs	number of nucleotides in contig	number of contigs	number of nucleotides in contig	Percentage of total nucleotides in contig	number of contigs	N50 / number of scaffolds larger than N50	number of nucleotides in contig	Percentage of total nucleotides in contig	number of predicted proteins (prodigal)
*w*VulC_MiSeq	Nembler assemblies do not have gaps	6,067	3,837,349	7	12,048	5,351	2,289,960	59.7%	**709**	**9,479 / 44**	**1,537,389**	**40.1%**	**1,880**
*w*Ase_MiSeq		4,018	2,794,237	8	6,532	3,253	1,268,725	45.4%	**757**	**6,731 / 65**	**1,518,980**	**54.4%**	**1,924**
*w*Dil_MiSeq		16,824	6,935,381	2	14,468	16,060	5,379,271	77.6%	**762**	**5,925 / 72**	**1,541,642**	**22.2%**	**1,925**

The Velvet assemblies of the HiSeq reads produced fewer contigs than the Newbler assemblies from MiSeq reads, which totaled more nucleotides in assembled contigs. After filtered out the mitochondrial sequences and the contaminant sequences, it appeared that the second capture round, which optimized the sequencing of the *Wolbachia* sequences, also co-captured more contaminants. The assembled *Wolbachia* sequences totaled about 1.3 Mb for the strains sequenced by HiSeq and about 1.5 Mb for the strains sequenced by MiSeq. The contaminant sequences are largely unknown sequences; they have no BLAST hit in the NCBI database. These contigs likely are host DNA sequences since none isopod crustacean genome has currently been sequenced.

To compare the quality of the Velvet assemblies of the short 50bp Illumina HiSeq reads with the Newbler assemblies of the longer 250bp Illumina MiSeq reads, an alignment of both assemblies *w*VulC_HiSeq and *w*VulC_MiSeq on the reference *w*VulC genome was generated with the software Mauve (Figure 25). Both assemblies produced contigs that covered the entire reference genome (graphs on the right of Figure 25), but the *w*VulC_MiSeq assembly presented fewer miss-assemblies than the *w*VulC_HiSeq assembly.

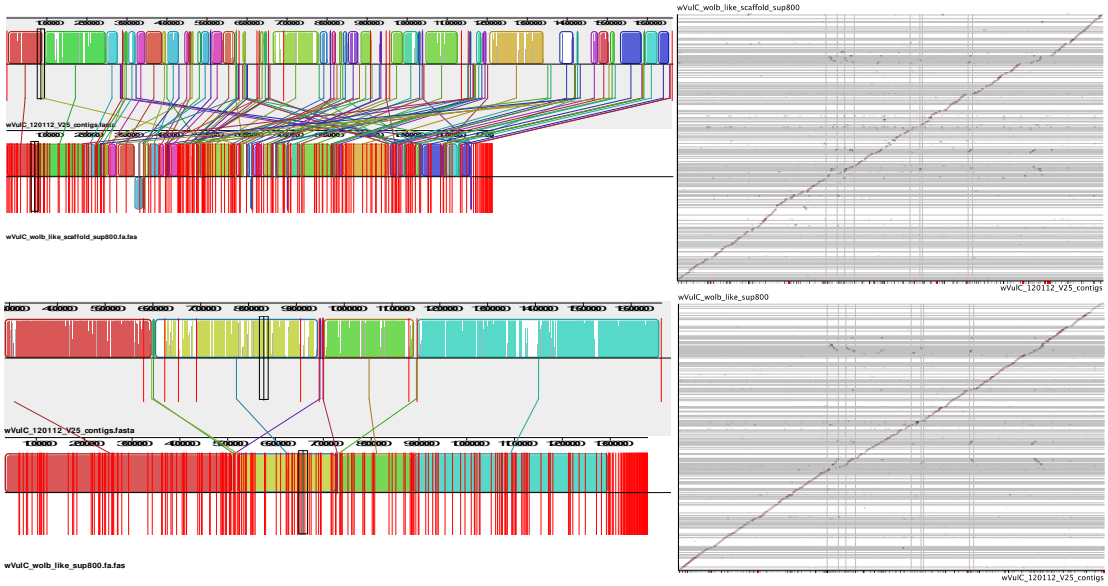


Figure 25: Comparison of HiSeq (top) and MiSeq (bottom) *w*VulC assemblies. Mauve alignments on the reference 10 contigs.

In conclusion, the assemblies of short reads are not sufficient to correctly assemble the genomes because of the highly repeated structure of *Wolbachia* genomes, but they produced sequences good enough to obtain the coding-sequence information, which was the aim of this strategy. Assemblies of longer MiSeq reads already gave better quality contigs and still produced a large number of contigs (~700).

3 - 'Closure' of *w*VulC

To produce a first complete genome of a *Wolbachia* strain from isopod, I prepared a DNA sample of *w*VulC for PacBio sequencing.

A first attempt of sequencing was performed using a short (2kb) and a long (6kb) libraries built from DNA extracted with a regular phenol / chloroform extraction followed by an enrichment in *Wolbachia* DNA. Original crude DNA extract was composed of 93% host DNA (*A. vulgare*). After a round of enrichment by a specific filtration that separates entire *Wolbachia* bacteria from nuclear DNA, a first DNA sample was composed of 88% *Wolbachia* DNA in 1µg and a second DNA sample was composed of 73% *Wolbachia* DNA in 5µg. The first sample was entirely used for the construction of the 2kb library and the second sample was entirely used for the construction of the 6kb library.

After the library constructions performed according to the manufacturer manual (III.A.4 - p164), I obtained 73ng DNA, which represents a DNA yield of 27% for the 2kb library and 1,368 ng of DNA (DNA yield of 31.2%) for the 6kb library. These yields were even higher than estimated by the library construction manual; only few picograms are needed to load the PacBio sequencer. A 120-minute PacBio movie and two successive 55-minute movies were performed for respectively the 6kb and 2kb libraries.

Even if these library constructions produced more DNA than needed for PacBio sequencing, the low qualities of the reads for both libraries did not allow an acceptable analysis. Both sequencing totalized 37,589 reads (~ 27 Mb) of which 17,425 (46.4%) were isolated as previously done by BLAST as *Wolbachia* sequences. An assembly using Newbler produces 509 contigs for a total of 339 kb, which is far away from the 1.76 Mb expected.

The main explanation could be the use of phenol / chloroform DNA extraction procedure, which could have left some traces of phenol with the DNA fragments. These traces are now known to inhibit the PacBio DNA sequencing, favoring polymerase detachment during sequencing. Since this sequencing technology is still quite recent and constantly optimized, this issue was unknown at the time of our sequencing.

Consequently, more DNA was extracted using a more PacBio compatible extraction kit (DNeasy Blood & Tissue Kit, Qiagen) and sent for sequencing at NEB after enrichment. Sequencing is in progress.

C. Genome features

The seven sequenced genomes present similar genome features, which are common to *Wolbachia* genomes. These genomes are about 1,5Mb with 34.5/34.6 % GC content (Table 12). They encoded two ribosomal operons (16S-23S and 5S) and 34 tRNAs.

The numbers of predicted protein-coding sequences are higher than previously sequenced genomes and the average gene size is smaller, which could be explained by fragmented genes.

Table 12: Comparative genome feature. Strains sequenced in this study are in bold.

	*w*VulC	*w*Pip-Pel	*w*Mel	*w*Ri	*w*Bm	*w*Oo	*w*No	*w*Ha	***w*VulC**	***w*Albu**	***w*Ase**	***w*Bre**	***w*Dil**	***w*Pet**	***w*VulM**
Host	*Armadillidium vulgare*	*Culex pipiens*	*Drosophila melanogaster*	*Drosophila simulans*	*Brugia malayi*	*Onchocerca ochengi*	*Drosophila simulans*	*Drosophila simulans*	***Armadillidium vulgare***	***Armadillidium album***	***Oniscus Asellus***	***Helleria brevicornis***	***Porcellio dilatatus dilatatus***	***Porcellio dilatatus petiti***	***Armadillidium vulgare***
Genome statistics															
Genome size (bp)	1,663,852	1,482,455	1,267,782	1,445,873	1,080,084	957,990	1,301,823	1,295,804	**1,537,389**	**1,585,022**	**1,518,980**	**1,340,868**	**1,541,642**	**1,341,349**	**1,314,883**
GC content (%)	34.4	34.2	35.2	35.2	34.2	32.1	34.5	35.34	**34.6**	**34.6**	**34.6**	**34.6**	**34.6**	**34.5**	**34.6**
Gene predictions															
Predicted protein-coding genes	1,568	1,385	1,269	1,264	903	842	1,040	1,010	**1,880**	**2,615**	**1,924**	**1,834**	**1,925**	**1,665**	**1,684**
Coding density (protein-coding nucleotides, %)	82.52	81.20	80.10	78.10	67.00	66.50	80.00	78.00	**82.65**	**84.05**	**82.45**	**83.82**	**83.98**	**82.75**	**82.98**
Average gene size (bp)	728	944	851	951	899		1013	1000	**676**	**509**	**651**	**613**	**673**	**667**	**648**
Number of rRNA	3	3	3	3	3	3	3	3	**3**	**3**	**3**	**3**	**3**	**3**	**3**
Number of tRNA	35	34	34	34	34	34	34	34	**34**	**34**	**34**	**34**	**34**	**34**	**34**

Discussion

Using the targeted genome enrichment procedure we developed to specifically purify endosymbiotic DNA from host DNA, we successfully partially sequenced 6 new *Wolbachia* genomes of strains infecting isopod crustaceans that induce either CI or feminization. While these DNA genome sequences are not complete (they are assembled in 300-700 contigs) most of the genic information is present. Actually, for all these sequenced strains, we obtained about 1.5Mb of sequences and 83% of coding density, which is a rate close to those obtained for the already published genome. Since our objective was to produce a genic dataset complete enough to perform comparative genomics between the strains, no complementary approach by PCR sequencing was performed even if it could easily take the number of contigs down to dozens having the *w*VulC genome as reference. This NGS approach producing partial genomes was previously validated for the sequencing of the *Wolbachia* strains *w*AlbB and *w*Bol1-b (Mavingui et al. 2012; Duplouy et al. 2013).

However, some assembly optimizations still can be performed for *w*VulC and *w*Dil, for which we generated reads from HiSeq and MiSeq sequencings, by combining both assemblies. Moreover, PacBio sequencing of *w*VulC may lead to the first complete *Wolbachia* genome from isopod crustacean. This approach was already successfully used to sequence whole genomes rich in repetitive sequences, in combination with Illumina reads (English et al. 2012).

Additionally, since no biotin (H vitamin or B_7) operon has been published in *Wolbachia* genomes yet, a characterization of the *w*VulC biotin genes by RT-PCR should determine if the biotin biosynthetic pathway is functional in this strain of *Wolbachia*. Actually previous published *Wolbachia* genomes only had the BioY importer, suggesting that these *Wolbachia* strains import host biotin. This is likely the case for the strain *w*Cle, which infects the bedbug *C. lectularius* since antibiotic treatment induced a reduction of the growth and an alteration of the reproduction that are reversed by B vitamins (including biotin) supplementation (Hosokawa et al. 2010). The strain *w*VulC possesses the 6 genes involved in the biotin biosynthesis and the synteny of these genes (BioB/BioF/BioH/BioC/BioD/BioA) is highly conserved with *Neorickettsia sennetsu*, *N. risticii* and in the Rickettsial strain (REIS) of *Ixodes scapularis* (Dunning Hotopp et al. 2006; Gillespie et al. 2012; Penz et al. 2012). Biotin is important for carboxylation reactions and cannot be synthetized by many multicellular eukaryotes, including insects. This result may suggest that the *w*VulC strain, by synthetizing its own vitamin B_7, may play a role in host nutrition and favors host fitness.

As previously mentioned, three *Wolbachia* strains (*w*VulC, *w*VulM, *w*VulP) have been identified in different *A. vulgare* populations (Rigaud 1991; Cordaux et al. 2004; Verne et al. 2007, 2012). Keeping in mind that the strain *w*VulP has been characterized as a recombinant strain of *w*VulC and *w*VulM, bi-infection of a same individual might have occurs at least transiently (Ironside et al. 2011) since multiple infections with feminizing *Wolbachia* strains are unstable (Caubet et al. 2000). A recent re-investigation of the same natural populations (Verne et al. 2012) by a new genotyping method allowing the detection of multi-infection of some individuals, suggested that bi-infection could be a transitory phase towards the fixation of the most virulent feminizing strain (Valette et al. 2013).

The simultaneous localization of both strains in host tissues could help understanding host-symbiont interactions. Therefore, the list of strain-specific genes, particularly those presenting a high transcriptional rate, should be investigated for the identification of strain-specific markers usable in FISH imaging.

More generally, this strain-specific genes list represents a repertoire of potential strain marker that could be used in routine PCR for strain screening and identification.

Comparative genomic approaches are powerful tools to study organism similarities and differences. It allows for example to give significant biological insights of genome structure such as global genome statistic, repeats, genome rearrangements, gene recombination, and synteny. Indeed, characterization of IS and mobile introns dynamics (Cerveau et al. 2011; Leclercq et al. 2011) revealed insights of *Wolbachia* genome structure evolution. In filarial nematodes, sequencing of *Wolbachia* strains from worldwide populations can lead to the identification of polymorphisms. These polymorphisms can be of interest for the identification of drug targets and also for the evolutionary analysis of *Wolbachia* prevalence and distribution under natural conditions. Most of the actual researches of drug target are based on the *w*Bm laboratory strain from the TRS lab, which have been kept in lineage for over 30 years. Recent studies on insects (Weeks et al. 2007; Carrington et al. 2011) showed genetic divergences between natural populations and laboratory lineages. Thus, sequencing of natural population of *w*Bm could highlight these eventual polymorphisms in order to adapt more efficient drug discovery to wild-type conditions.

Comparative genomics can also leads to the comparison of non-coding regions as prediction of regulatory elements. In order to answer our question of the molecular mechanism of the symbiosis, we focused our comparative genomic approach to the coding regions and protein orthologs.

WHEN COMPARATIVE GENOMICS LEADS TO SYMBIOSIS-RELATED PAN-GENOMES

Introduction

Using whole-genome sequencing, genome comparisons between different phenotypic inducing strains from a same species offer a new perspective for investigation of virulence factors, in particular for unculturable bacteria (Hansen et al. 2012; Ghignone et al. 2012). Whole-genome sequences also provide a novel inroad for phylogenic studies. Where lateral gene transfer, convergence and gene-dependent evolution rate limit traditional single-gene phylogeny, phylogenomics draws reliable information based on large number of genes (Dagan 2011).

Cross-strain genome studies are based on orthologs, genes present in different species that have only evolved through speciation events. More rarely, paralogs, genes arisen by duplication events (Fitch 2000), are used. Orthologs are known to have the same biological functions (Tatusov et al. 1997) whereas duplication allows development of new functions (Ohno et al. 1968). Identification of orthologs is a powerful tool of understanding the genealogy of genes to investigate the mechanism of evolutionary process. These groups of genes with the same biological function are also an inestimable pool of information for functional analyses such as pathway comparison across species.

On the basis of orthologous databases, sets of bacterial genomes can be divided into (1) a “core-genome”, containing genes shared by all strains, (2) a “dispensable-genome” composed by genes present in two or more but not all strains and finally (3) genes unique to single strains (Figure 26) (Medini et al. 2005; Muzzi et al. 2007; Tettelin et al. 2008; Bentley 2009; Mira et al. 2010). Core-genome, dispensable-genome and unique genes composed the “pan-genome” (pan, from the Greek word π αν, meaning whole), also known as “species-genome” or “supra-genome” (Bentley 2009). As a direct consequence, a species pan-genome has an infinite size; each new sequenced genome would add new genes.

The core-genome is assigned to be composed of important or essential genes for the studied species. While core-genome analysis leads to phylogenomics and analysis of species evolution, dispensable-genome and strain specific genes would be great tools for studying host-symbiont symbiosis.

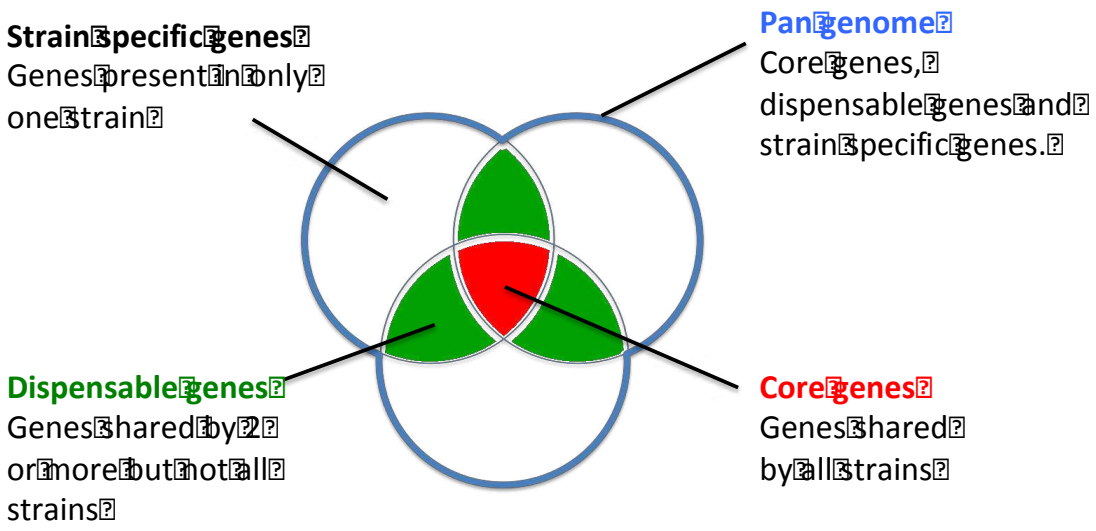


Figure 26: Schematic representation of species pan-genome. Core, dispensable, strain specific genes are respectively represented in red, green and white. Circles represent 3 schematic genomes. Pan-genome is in bold blue, composed by core, dispensable and strain specific genes.

Wolbachia endosymbiosis is an excellent model for studies of phenotypic inducing symbiosis; very closed phylogenetic strains induce a wild variety of phenotypes through the symbiosis continuum going from the parasitism to the mutualism. To date, 10 cytoplasmic incompatibility strains, 4 mutualistic strains, one parthenogenetic strain and one male killing strain have been sequenced (Table 3). In this study, we complement this dataset by adding 2 cytoplasmic strains (*w*Dil, *w*Pet) and 4 feminizing strains (*w*Albu, *w*Ase, *w*VulC, *w*VulM) and one unknown phenotype strain (*w*Bre).

Therefore, we generated the database of the *Wolbachia* orthologous clusters and established the *Wolbachia* pan-genome of all the available genomes (published and previously sequenced) as well as the *Wolbachia* pan-genome of the strains infecting isopod crustaceans. Based on all the single-copy core-genes, a phylogenomy of these *Wolbachia* strains was inferred.

We then investigated phenotype-specific gene patterns considering the 5 main known symbiotic interactions, allowing the identification of genes that could be possibly related to symbiosis and host-symbiont interactions. *Wolbachia* genomes were clustered by phenotypes to investigate phenotypic differences at the gene level, such as shared and phenotype-specific genes for the purpose of finding a genomic interpretation to the different phenotype-inducing *Wolbachia* symbiosis and define a set of genetic factors related to the *Wolbachia* symbiosis.

I. Comparative genomics

Pan-genome analyses are based on the OrthoMCL output file, which is produced from an all-against-all protein BLAST of all the proteins of the 23 studied *Wolbachia* strains (Figure 27). Clusters of orthologs with recent paralogs are generated based on similarity. The pan-genome is composed by all these clusters and by the strain specific proteins; it represents the global gene repertoire of the *Wolbachia* species.

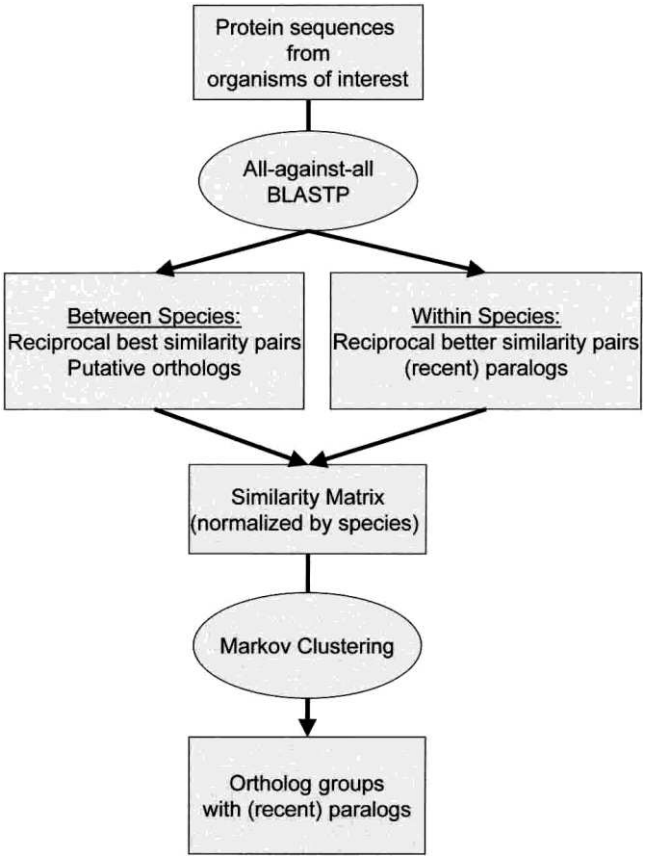


Figure 27: Flow chart of the OrthoMCL algorithm for clustering orthologous proteins. (Li et al. 2003)

A. Pan genome analysis

Whole-genome comparison of the different *Wolbachia* strains leads to the generation of a first *Wolbachia* pan-genome (all-*Wolbachia* pan-genome) that includes the sequences of the 23 *Wolbachia* strains for which a complete or partial genome sequence is known and a partial isopods' *Wolbachia* pan-genome only composed by the genomic sequences of the *Wolbachia* infecting isopod crustaceans.

1 - Global all-*Wolbachia* pan-genome

Out of the 31,123 predicted protein-coding sequences belonging to the 23 *Wolbachia* strains, 6,403 homologous clusters, which represent the pan-genome, (see supplemental_file_1) were generated including 4,270 clusters unique to one strain (see supplemental_file_2). The majority of the 31,123 predicted proteins are part of the core-genome or the dispensable-genome, meaning that they have homologies in other *Wolbachia* proteomes (26,686 proteins; 85.7%). Since most of the genomes used for this study are draft genomes in contigs, we can consider that proteins present in 22 of the 23 proteomes are more likely conserved proteins, and also part of the core-genome. The distribution of the 6,403 clusters of proteins across these strains reflects a bimodal model: most of the clusters (53.6%) are either components of the core-genome or strain-specific proteins (Figure 28).

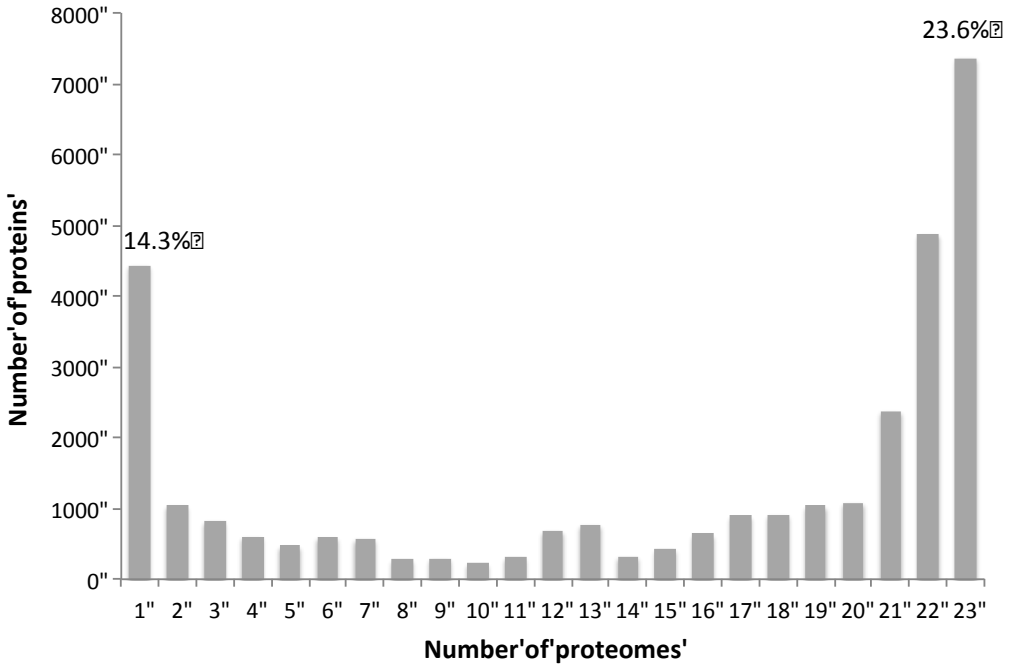
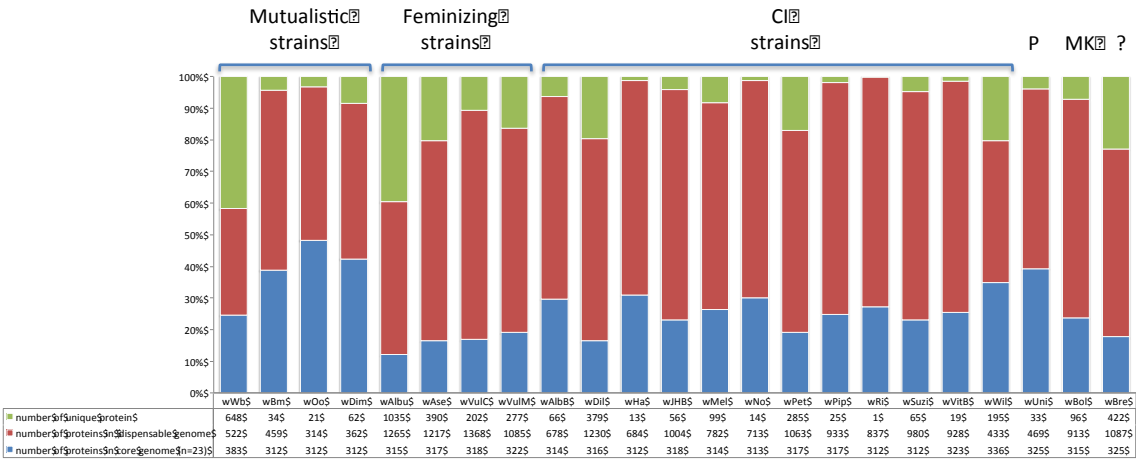


Figure 28: Frequency of proteins within the 23 studied *Wolbachia* strains. Each bar represents the group of proteins shared by the indicated number of proteomes. Proteins present in a single genome represent species-specific proteins, while proteins present in the 23 proteomes represent the *Wolbachia* core-genome. Proteins present in 2 to 22 proteomes compose the dispensable-genome.

Most of the proteins (96%) occurring only in a single proteome are unique single-gene proteins, with no paralogs. Few proteins are strain-specific and present in more than one copy; this is more likely due to the draft-state of the genomes. In contrast, 7,360 proteins belong to 311 core clusters; clusters composed of at least one protein of every studied *Wolbachia* strain. 210 of these core clusters are "single-gene" core clusters,

composed by orthologous proteins that do not have paralogs. The percentage of each genome that comprised the different components of the *Wolbachia* pan genome (core-genome, dispensable genome and strain-specific genes) is represented in Figure 29.
The proportion of each genome assigned to the core-genome (in blue in Figure 29) ranges from 12% for *w*Albu to 48% for *w*Oo, whereas species-specific genes (in green in Figure 29) represent from 0.1% (a single protein) for *w*Ri to 41.7% (648 proteins) for *w*Wb. The average of strain-specific proteins on the 23 *Wolbachia* strains is 11.5%.



	wWb	wBm	wOo	wDim	wAlbu	wAse	wVulC	wVulM	wAlbB	wDil	wHa	wJHB	wMel	wNo	wPet	wPip	wRi	wSuzi	wVitB	wWil	wUni	wBol	wBre
number of unique protein	648	34	21	62	1035	390	202	277	66	379	13	56	99	14	285	25	1	65	19	195	33	96	422
number of proteins in dispensable genome	522	459	314	362	1265	1217	1368	1085	678	1230	684	1004	782	713	1063	933	837	980	928	433	469	913	1087
number of proteins in core genome (n=23)	383	312	312	312	315	317	318	322	314	316	312	318	314	313	317	317	312	312	323	336	325	315	325

Figure 29: *Wolbachia* pan genome distribution. Percentage of genome included in core-genome (blue), dispensable-genome (red) and strain-specific gene (green). Numbers in table indicate the number of proteins for the *Wolbachia* pan-genome and for each strain that belong to these categories.

The exceptionally low percentage of unique genes for *w*Ri is due to the presence of a phylogenetically very close strain (*w*Suzi). In contrast, the presence of a relatively high percentage of unique genes could be a sign of a more distant strain such as *w*Wb and *w*Albu. However, both strains are not completely sequenced, (they have respectively 763 and 1290 scaffolds) and as a result, these proteins could be truncated proteins due to their position at the border of a contigs. Each strain contains a relatively large number of dispensable genes, genes shared by at least two but not all genomes. This high proportion shows a weak positive selection for these genes that can carry important genes such as genes for the biotin biosynthesis pathway. Even if these genes are not necessary for survival or maintenance of the species as they are absent from some strains, they may play important roles in virulence or attested of a host-adaptation. It is known that mutualistic strains have smaller genomes as they use some essential host biosynthesis pathways for their own survival (Moran 2001; Moran 2002; Foster et al. 2005; Murfin et al. 2012).

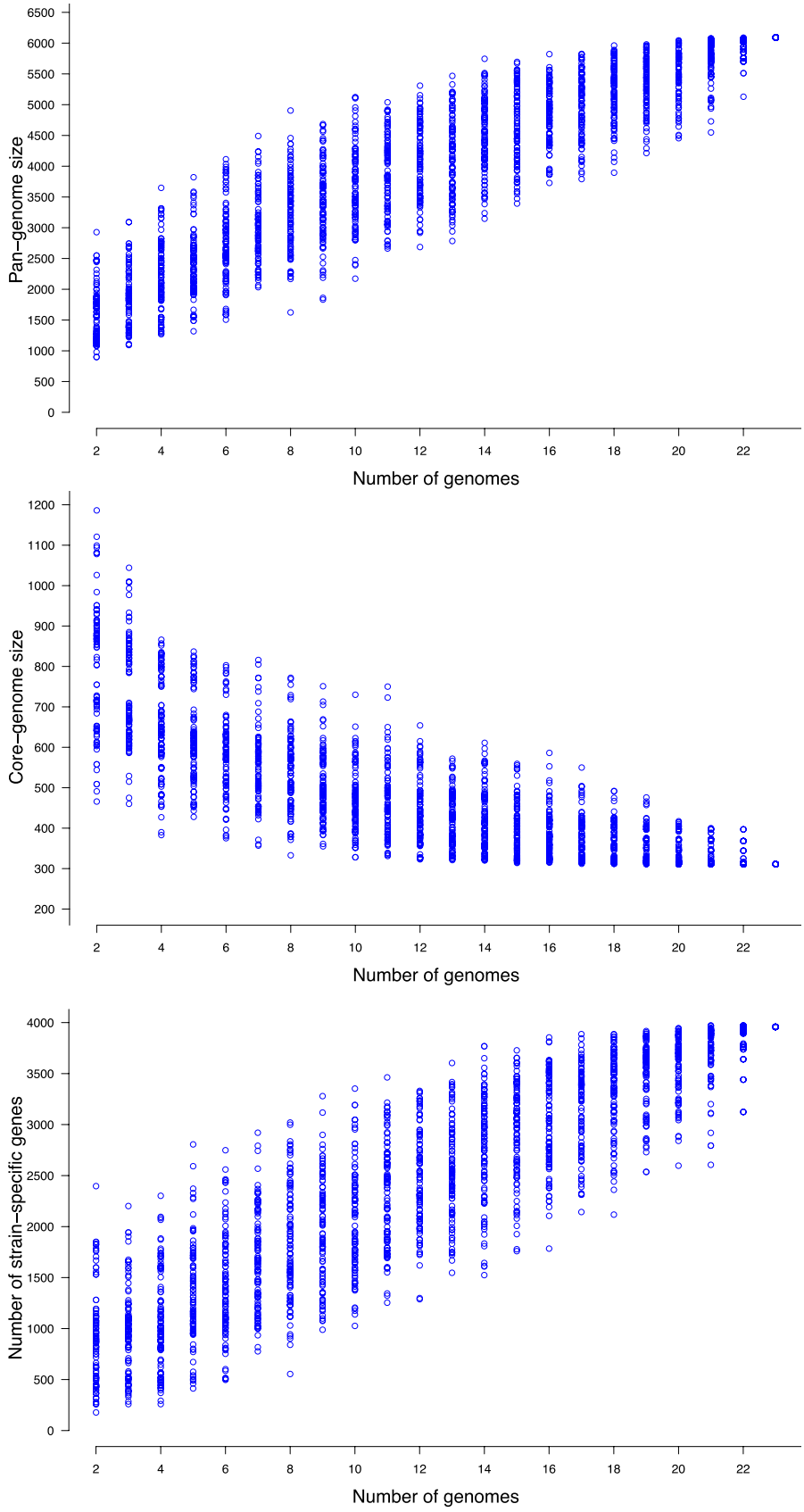


Figure 30: Accumulation curves for *Wolbachia* pan-genome, core-genome and strain-specific genes, given a number of genomes analyzed for the different strains of *Wolbachia.* The plots correspond to the distribution after repeating one hundred random input orders of the genomes.

Analysis of the accumulation curves of the pan-genome, core-genome and strain specific genes gave an overview of the high diversity of the *Wolbachia* genomes (Figure 30). The size of the pan-genome increases with every new additional strain while the number of conserved genes composing the core-genome is decreasing; increasing by consequence, the size of the dispensable genome. Similarly, the number of strain-specific genes increases for each new additional genome.

A power law regression analysis of the pan-genomic distribution ($y=961.21x^{0.5806}$) describes an open pan-genome, which is estimated to contain 13,932 different genes after 100 *Wolbachia* strains have been sequenced. This estimate pan-genome size is 10-fold larger than the average number of gene per *Wolbachia* genome (1,375 genes) indicating a high genomic diversity in *Wolbachia* genomes. Similarly, the power law regression analysis of the core-genomic distribution ($y=1125.7x^{-0.391}$) estimates the core-genome size at 186 genes after 100 *Wolbachia* strains have been sequenced.

The power law regression analysis of the species-specific genes distribution ($y=394.48x^{0.7289}$) estimates that there will be 11,320 species-specific genes after 100 *Wolbachia* strains have been sequenced. This will represents ~81% of the estimated pan-genome. These results illustrate an open pan-genome; each new sequenced genome adds new genes to the *Wolbachia* pan-genome, which virtually has an unlimited size.

2 - Isopod *Wolbachia* pan-genome

The isopod pan-genome represents a subset of the all-*Wolbachia* pan-genome based on the 7 *Wolbachia* strains from isopods. It contains a total of 13,535 predicted protein-coding sequences; however, 309 orthologous clusters from the previous all-*Wolbachia* pan-genome are absent from isopod *Wolbachia* pan-genome.

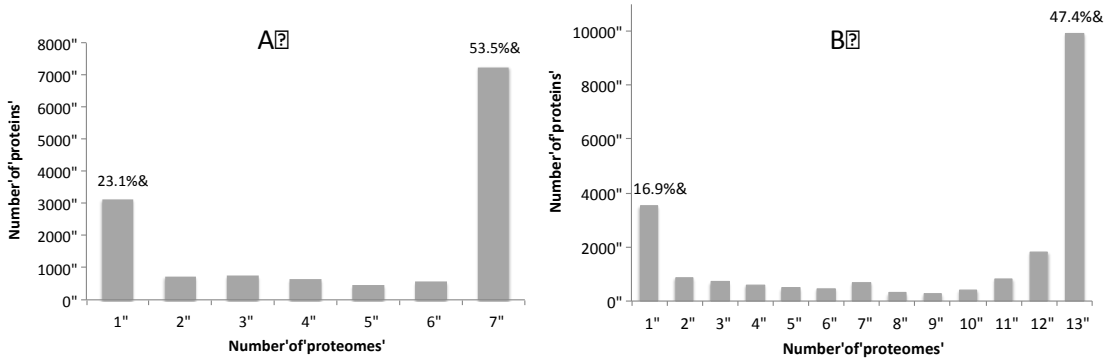
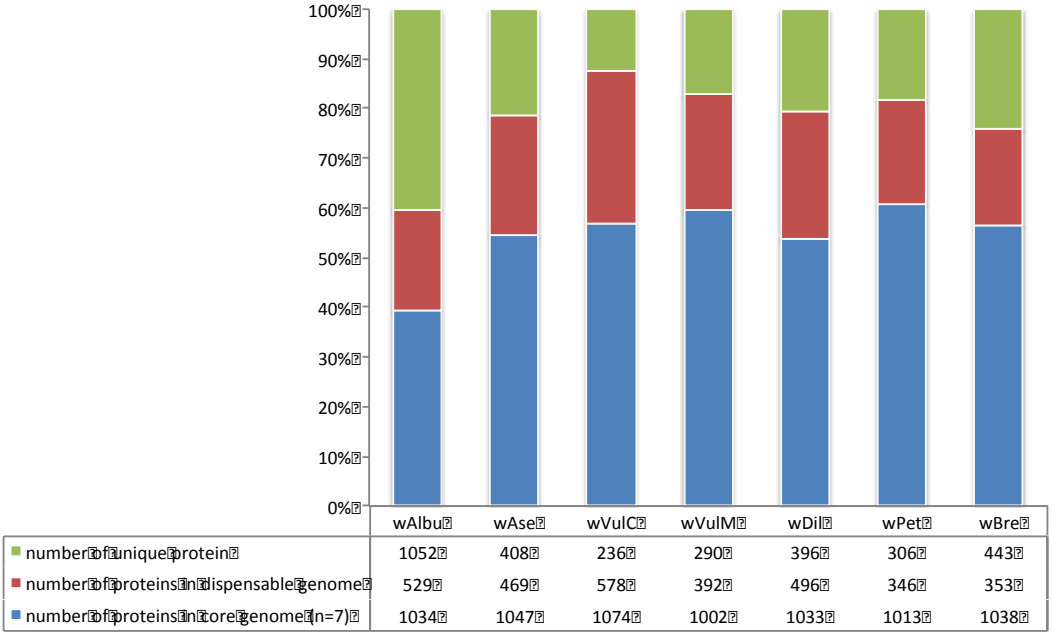


Figure 31: Frequency of proteins within A: the 7 isopod *Wolbachia* strains; B: the 13 supergroup B *Wolbachia* strains. Proteins present in a single genome represent species-specific proteins in opposition to proteins present in the 7/13 proteomes, which are the proteins coded by the isopod/Supergroup B *Wolbachia* core-genome.

As with the all-*Wolbachia* pan-genome, the protein frequency within the 7 isopod *Wolbachia* strains is bimodal (Figure 31). 23% of the proteins are strain-specific proteins, whereas 53.5% (913 orthologous clusters) compose the core-genome of isopod *Wolbachia*. This is more than twice the proportion of proteins in the all-*Wolbachia* core genome. This profile is representative of phylogenetically close strains, but according to the *wsp* phylogeny (Figure 18), some isopod strains such as *w*Pru from *P. pruinosus* are closer to *Wolbachia* strains from butterflies (e.g. *Acraea encodon*) than to other *Wolbachia* strains from isopods. Therefore, these pan-genome distributions should be complemented by the *w*Pru genome after its resequencing.

As it is well-known that single-gene phylogeny is dependent on the evolutionary rate of the marker gene, which may not be representative of the evolutionary rate of the whole-genome, this newly-generated core-genome will be used to infer the *Wolbachia* phylogenomy, a phylogeny based on the entire set of conserved protein-coding-genes.

The comparison of the protein distributions across the strains from isopod crustaceans and the B-supergroup *Wolbachia* confirms that isopod symbiont genomes are more similar to each other than to other B-supergroup *Wolbachia* (Figure 31).



	wAlbu	wAse	wVulC	wVulM	wDil	wPet	wBre
number of unique protein	1052	408	236	290	396	306	443
number of proteins in dispensable genome	529	469	578	392	496	346	353
number of proteins in core genome (n=7)	1034	1047	1074	1002	1033	1013	1038

Figure 32: Isopod-*Wolbachia* pan-genome distribution. Percentage of genome included in core-genome (blue), dispensable-genome (red) and species-specific gene (green). Numbers in table indicate the number of protein-coding sequences for each strain that belong to these categories.

This statement is also illustrated by the pan-genome distribution of the *Wolbachia* from crustacean isopods (Figure 32), for which half of the proteins of each genome constitute the core-genome. Actually, the core-genome of the 7 *Wolbachia* strains from isopods is composed of 913 orthologous clusters whereas the formerly established power law

regression predicts a core-genome of 531 clusters for 7 genomes. This core-genome size reinforces the finding that isopod *Wolbachia* genomes are more conserved than the *Wolbachia* found in other host and confirms that the difference in the core-genome size is not only due to the strain-number factor. Sharing more conserved genes is coherent with the evolution of these strains within the same ecological niche; they are exposed to the same environmental conditions, which differ from those of insect symbionts.

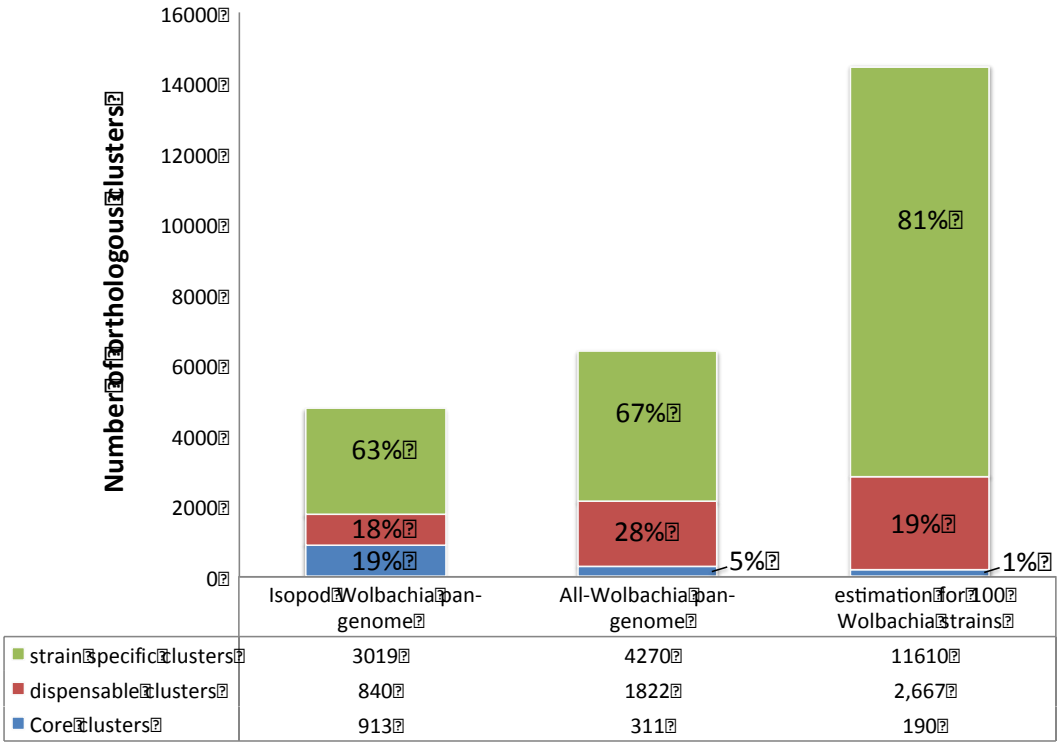


Figure 33: Pan-genome distributions comparison.

A comparison of the isopod *Wolbachia* pan-genome, the all-*Wolbachia* pan-genome and the estimated 100-*Wolbachia* pan-genome compositions is illustrated in Figure 33. While the core-genome represents 19% of the isopod *Wolbachia* pan-genome, it represents 5% of the all-*Wolbachia* pan-genome and 1.3% of the estimate pan-genome at 100 *Wolbachia*. These are more conserved-sequences between the strains infecting isopods than other hosts, for which the dispensable genome is more important.

B. Phylogenomics

In order to create a phylogenomy that is more representative of the whole-genome evolutionary relationship than a single-gene phylogeny, phylogenomic trees were inferred based on the entire set of conserved protein-coding-genes, formerly identified in the *Wolbachia*-core genomes. Phylogenomics is based on the alignments of single-gene orthologous clusters (one copy per studied strain) identified by OrthoMCL and phylogenomic trees were created using the partitioned maximum likelihood method (Figure 34).

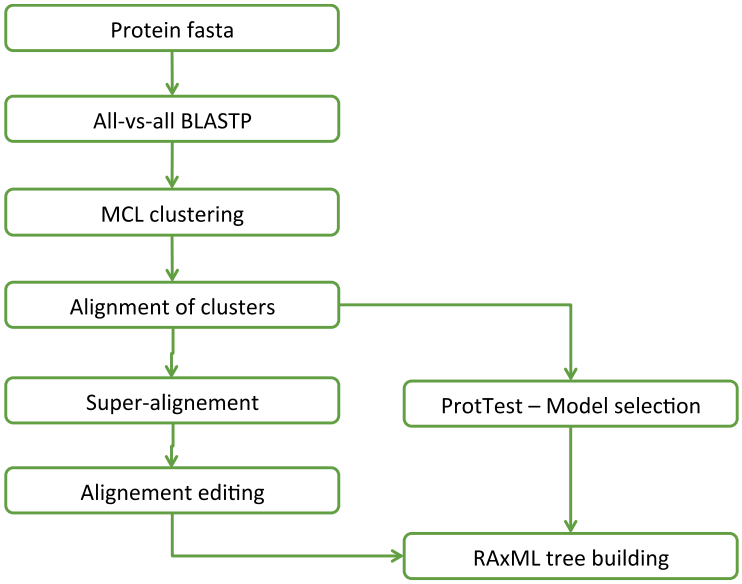


Figure 34: Phylogenomics flow diagram.

Alignments were generated for each gene at the protein level and concatenated into a super alignment, which was edited to remove all gap positions. Single-gene alignments were used to estimate the best amino acid substitution model for each gene using ProtTest (Darriba et al. 2011). Based these ProtTest output files, a shell parsing script was developed to generate a single alignment partition that lists for each gene its position in the concatenated alignment and its best substitution model. This file was used in RAxML (Stamatakis 2006) for tree generation. A series of 1000 bootstrap replicates was performed by randomly sampling positions. The statistical support for each branch was calculated as the percentage of trees where the corresponding group was recovered.

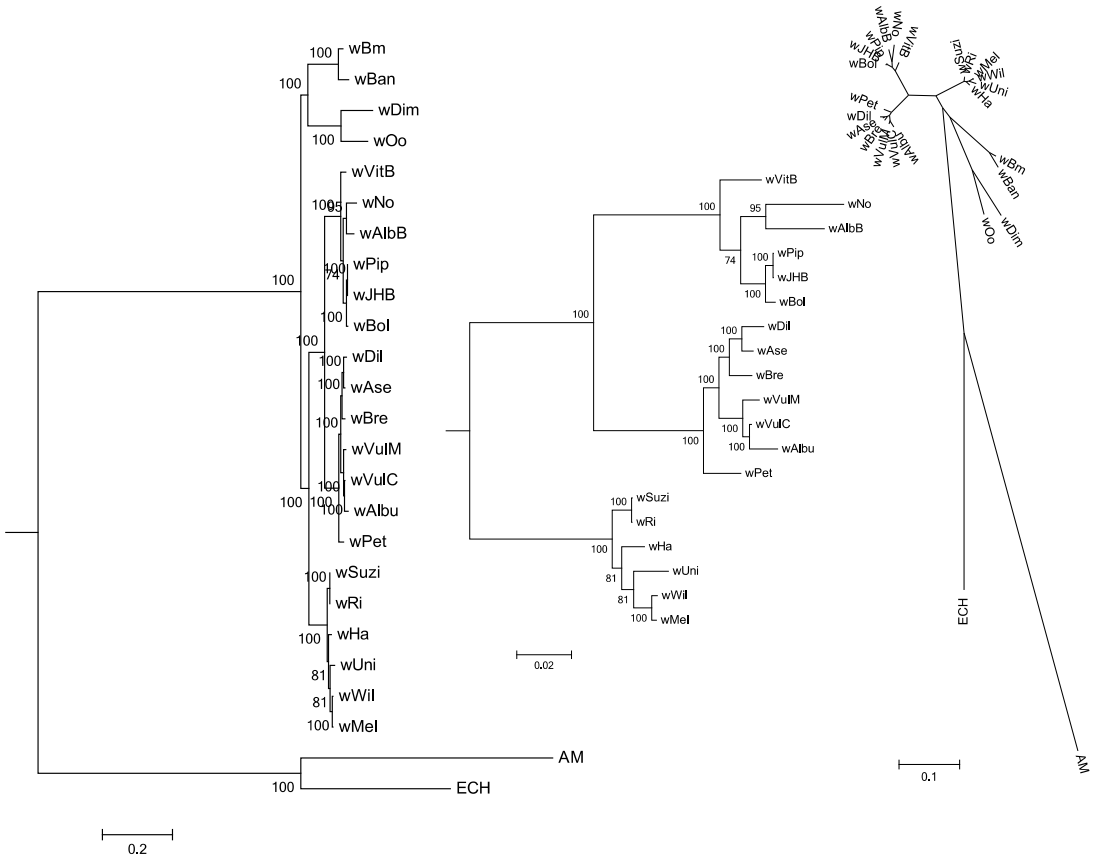


Figure 35: Phylogenomics of the *Wolbachia* strains. The phylogenetic tree was inferred at the protein level from the concatenated alignment of 175 single gene orthologs (48,384 amino acids) in RaxML using the maximum likelihood method. Numbers on the branches represent the support from 1,000 bootstraps. The tree is rooted by the two closest rickettsial strains: *Erlichia (ECH)* and *Anaplasma (AM)*.

The first unrooted phylogenomic tree was built for the 7 sequenced *Wolbachia* strains from isopods together with the 7 completely sequenced and published *Wolbachia* strains. This analysis was performed on 529 single-gene orthologs present in all 14 *Wolbachia* strains (Appendix 3, p212). In addition, a rooted phylogenomy was also inferred on these strains with the two closest rickettsial bacteria *Ehrlichia* and *Anaplasma* as outgroups, based on 175 single-gene core-orthologs (Figure 35). These two phylogenomic trees gave the same topology as the tree created from 209 single-gene orthologs between the all-23 *Wolbachia* strains completely or partially sequenced (Figure 36). They illustrated the well-known supergroup subdivision of the *Wolbachia* strains with two supergroups (C and D) composed by mutualistic strains from nematodes and two supergroups (A and B) composed by arthropod *Wolbachia* strains.

Additionally, these phylogenomic trees reinforced the idea established with the pan-genomic approach that the supergroup B may be composed of two subgroups: (i) the *Wolbachia* strains from isopods and (ii) the *Wolbachia* strains from insects. This

hypothesis should be confirmed by the resequencing of the *w*Pru genome and the sequencing of new non-isopod *Wolbachia* strains to ensure that this topology is not only induced by the relatively large number of isopod *Wolbachia* strains present in this study.

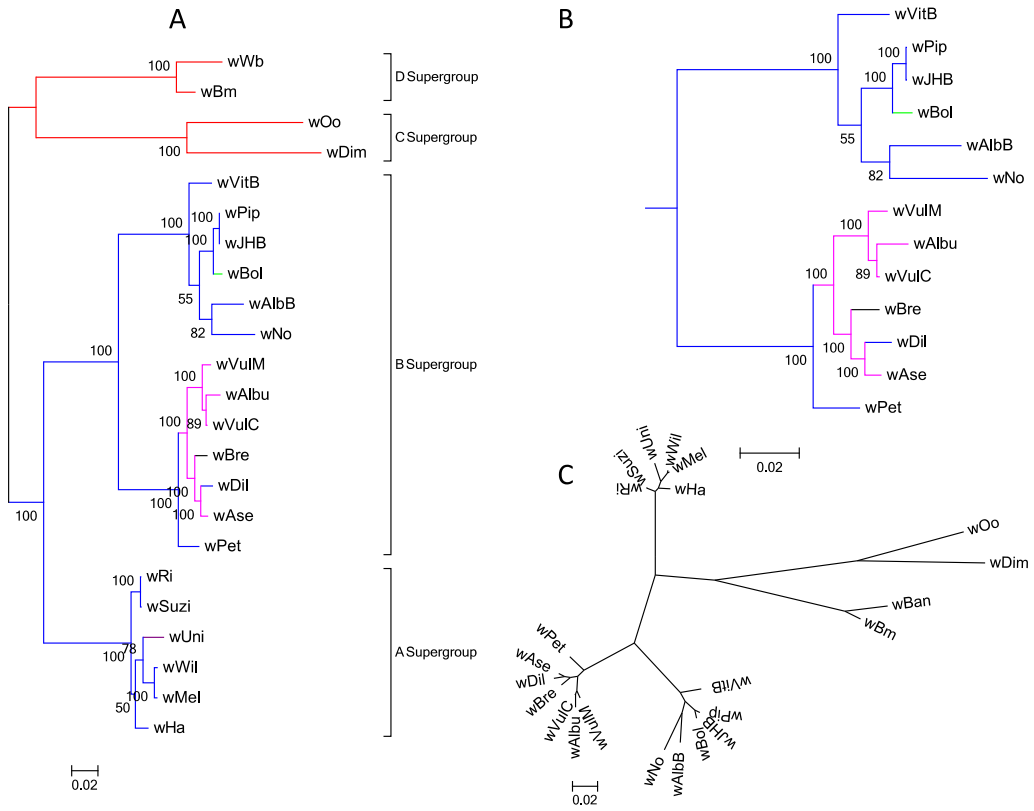


Figure 36: Unrooted phylogenomic trees of the *Wolbachia* strains (23). The phylogenetic tree was inferred from the concatenated alignment of the 209 single-gene orthologs in RaxML using the maximum likelihood method. Numbers on the branches represent the support from 1,000 bootstraps. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. A: general tree of the supergroups A, B, C and D. B: Zoom on the B supergroup. C: Radial representation of the general tree. The horizontal bar indicates number of substitutions per site. Red: mutualistic strains; blue: CI strains; pink: feminizing strains; green: male-killing strain; black: unknown phenotype.

From these three *Wolbachia* trees, it appears that all the *Wolbachia* strains from isopods are closer to each other than to any *Wolbachia* from other hosts and that the strain at the basal position of this isopod cluster is *w*Pet from *P. dilatatus petiti*. This phylogenomy of isopod *Wolbachia* strain is particularly surprising relative to the phylogeny based on the *wsp* gene (Figure 18, Figure 37), which positioned the two CI strains together.

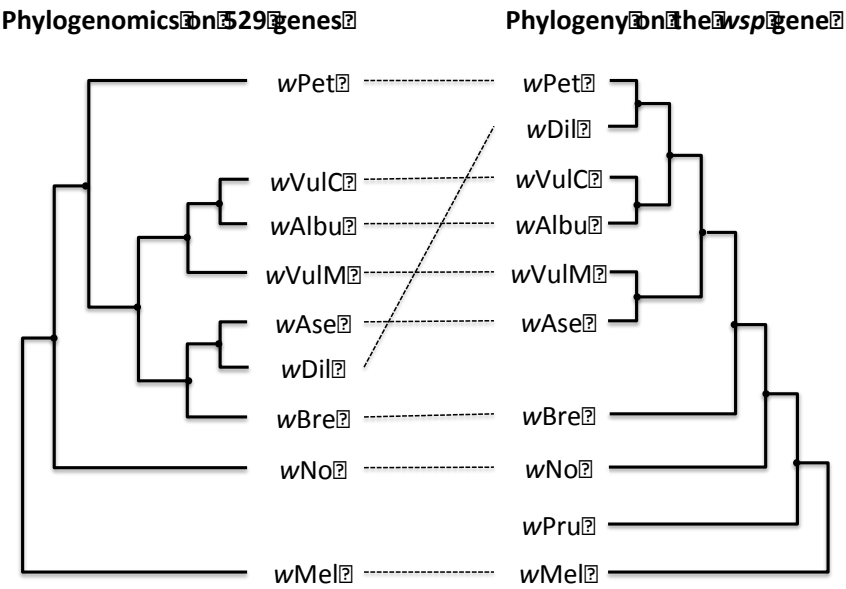


Figure 37: Comparison of the phylogenies inferred on the 209 single-gene orthologs and the *wsp* gene. Manual comparison of the trees from Figure 18 (right) and Figure 37 (left) of isopods strains and the *Drosophila simulans w*No strains and the *Drosophila melanogaster w*Mel strain as outgroups.

Within the isopod strains, two subgroups stand out with the CI strain *w*Pet at the base instead of *w*Bre described by the *wsp* phylogeny. The first subgroup contains all the strains that induce a strong feminization (*w*VulC, *w*Albu, *w*VulM) and a second subgroup is composed of three strains, a CI strain (*w*Dil), a strain whose effect remains unknown (*w*Bre) and an atypical feminizing strain (*w*Ase from *O. asellus*). Indeed, in infected *O. Asellus* populations, both males and females are infected by *w*Ase, which is unusual for feminizing strain infections since males should normally be feminized under *Wolbachia* influence. Moreover, *w*Ase infection induces a negative effect on host fecundity, slightly decreasing the number of offspring (Rigaud et al. 1999) and an CI effect induced by *w*Ase has also been suspected (W. LeClec'h, PhD thesis). Thus, even though there is no monophyletic clade for the CI strains, there is a clade regrouping the strong feminization strains and a clade regrouping the "atypical" *Wolbachia* strains. The resequencing of *w*Pru will allow its positioning in the phylogenomic tree and check if this strain is not phylogenetically close to other isopod *Wolbachia* strains.

In previous single-gene phylogeny (Bouchon et al. 1998; Pichon et al. 2009; Cordaux et al. 2012), no clear correlation between the *Wolbachia*-induced phenotype and the phylogenetic position of the strains was observed.

II. Investigation of phenotype-specific gene patterns: towards a symbiosis-related pan-genome?

In order to identify phenotype-specific gene patterns, the investigated *Wolbachia* strains were divided according to their 5 different symbiotic relationships: mutualism, cytoplasmic incompatibility, male-killing, feminization and parthenogenesis (Table 13). The strain *w*Bre from the isopod *H. brevicornis* was not included in this symbiosis-related pan-genome analysis since the induced phenotype is still unknown. Thus, this symbiosis study was done based on 22 *Wolbachia* genomes.

Table 13: Clusters of studied *Wolbachia* strains according to their symbiotic relationship. The strains sequenced in this study are in bold.

Mutualism	Cytoplasmic incompatibility		Male killing	Feminization	Parthenogenesis
*w*Bm	*w*AlbB	*w*Ri	*w*Bol	***w*Albu**	*w*Uni
*w*Dim	*w*Ha	*w*Suzi		***w*Ase**	
*w*Oo	*w*Mel	*w*VitB		***w*VulC**	
*w*Wb	*w*No	*w*Wil		***w*VulM**	
	*w*Pip-JHB	***w*Dil**			
	*w*Pip-Pel	***w*Pet**			

By comparing the distribution of the all-*Wolbachia* pan-genome orthologs between the 5 defined phenotype-specific clusters, certain genes were found to be exclusively present in strains inducing a particular symbiotic phenotype (Figure 38).

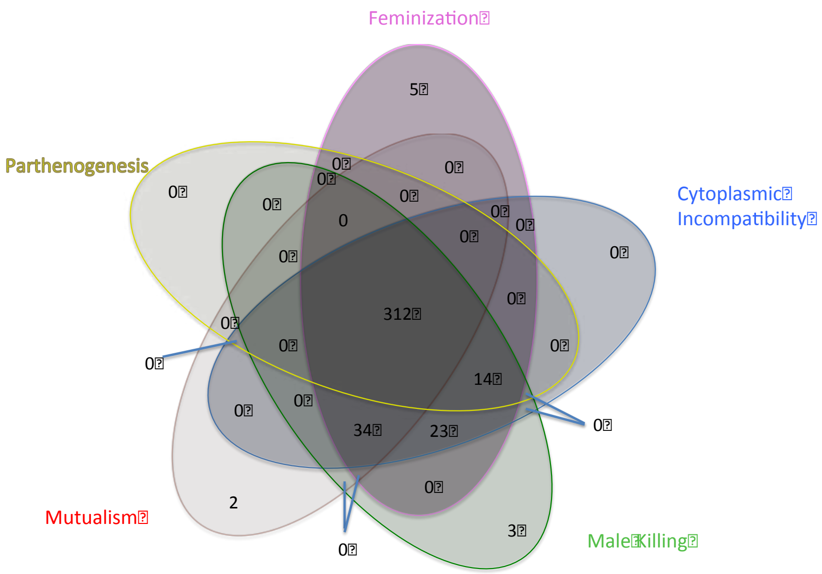


Figure 38: *Wolbachia* genomic diversity between different phenotype-inducing strains. Numbers represent the number of homologous clusters shared by the different phenotype-inducing strains.

The core-genome consisted of the 312 genes found in all strains independent of a particular phenotype. This core-genome is slightly different (an additional protein) than the all-*Wolbachia* core-genome previously established due to the removal of the *w*Bre genome for this phenotype-related analysis. This genomic investigation allowed the identification of 2, 3 and 5 phenotype-specific genes for mutualism, male-killing and feminization, respectively, but no genes were identified as phenotype-specific for CI or parthenogenesis. Interestingly, no genes were found for 2 or more different phenotypes, except for 3 junctions: (1) 14 genes were present in all but the mutualistic strains, (2) 23 genes were common to CI, feminizing and male-killing strains and (3) 34 genes were common in CI, feminizing, male-killing and mutualistic strains. However, these numbers may be over/under estimated as the parthenogenesis phenotype is only represented by an incomplete genome.

All identified genes were investigated for conserved-domain motifs using SMART (Simple Modular Architecture research tool, http://smart.embl.de) an online tool for the identification of protein domain architecture (Schultz et al. 1998; Letunic et al. 2012).

A. Mutualism-related gene pattern

Two genes are only present in the four mutualistic *Wolbachia* strains, the WAS family protein (WASP) and a hypothetical protein corresponding to wBm0076 and wBm0047 in the *Wolbachia* genome of *B. malayi*, respectively (Figure 39).

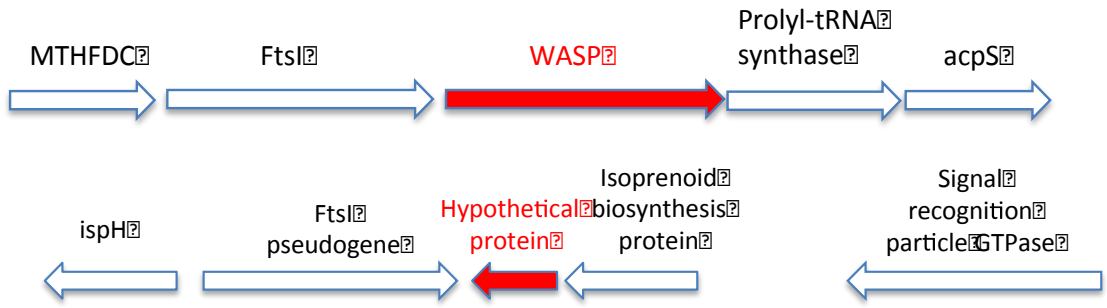


Figure 39: Position of the WASP gene and the wBm0047 gene in the *w*Bm genome.

The WASP protein or Wiskott-Aldrich syndrome protein is a eukaryote-like protein, probably acquired by a common ancestor of *Wolbachia* and *Rickettsia* from an animal host. This protein is disrupted in all studied non-mutualistic strains; when present, the remaining WH2 motif is pseudogenized. This protein, hypothesized to regulate the formation of actin filaments, has already been pointed out as a potential effector in *w*Bm

(Foster et al. 2005; Slatko et al. 2010) because of its WH2 conserved domain, which is known for binding actin monomers and facilitate the assembly of monomers into filaments (Machesky et al. 2001; Paunola et al. 2002; Chereau et al. 2005; Pernier et al. 2013). Since in nematodes, *Wolbachia* are found in the lateral chords, an actin-rich tissue, this eukaryote-like protein may be implicated in the host-symbiont interaction (Foster et al. 2005)

The other mutualistic-specific protein corresponds to the hypothetical protein wBm0047 that does not contain any conserved domain.

In contrast, 14 genes are revealed to be present in all but mutualistic strains (Table 14). These proteins included 5 hypothetical proteins with no significant motif and 1 hypothetical protein with a DUF1284 motif, a motif thought to be present in iron-sulphur binding proteins; 2 ankyrins proteins; 1 protein containing a SPFH domain, which is a membrane protein involved in the regulation of cation exchange; 1 magnesium chelatase-related protein, 1 methyltransferase, 1 16S ribosomal RNA methyltransferase RsmE and 1 sodium/alanine symporter family protein. Further investigations are needed to decipher the role of these non-mutualist-specific proteins in other *Wolbachia* strains and the consequence of their absence in mutualistic *Wolbachia* strains.

Table 14: List of the 14 proteins present in all but the mutualistic strains. The first column corresponds to the number of the orthologous clusters, the second column corresponds to the gene annotation, the third one is the accession number for the *w*Mel orthologs and the last column corresponds to the conserved motif.

		wMel	Motif
OGrps_58.fa	virB2	NP_966999.1	TrbC
OGrps_720.fa	hypothetical	NP_966041.1	NA
OGrps_751.fa	Ankyrin	NP_966003.1	Ank_5 motif
OGrps_816.fa	SPFH domain/Band 7 family protein	NP_966265.1	Band_7_2/ HflC
OGrps_819.fa	Ankyrin	NP_966281.1	Ank2 Ank4 motifs
OGrps_826.fa	Mg chelatase-related protein	NP_966479.1	ChlI/ AAA/ Mg-chelatase_2
OGrps_829.fa	16S ribosomal RNA methyltransferase RsmE	NP_966924.1	Methyltrans_RNA
OGrps_830.fa	hypothetical (histindine kinase ?)	NP_966982.1	NA
OGrps_833.fa	methyltransferase	NP_966812.1	trmB/ Methyltransf_4
OGrps_838.fa	hypothetical	NP_966917.1	DUF1284
OGrps_843.fa	hypothetical	NP_966979.1	NA
OGrps_844.fa	hypothetical	NP_966845.1	NA
OGrps_848.fa	sodium/alanine symporter family protein	NP_966770.1	Na_Ala_aymp superfamily
OGrps_87.fa	hypothetical	NP_966248.1	NA

B. Male killing related gene pattern

Three proteins were identified as specific to the male-killing inducing *w*Bol1-b *Wolbachia* strain: the DNA mismatch repair protein MutL, which is present in two copies in the *w*Bol1-b genome, (wBol1243 and wBol332), the baseplate assembly protein GpV (wBol1296), which is a phage related protein and a hypothetical protein, which is in two 100% identical copies (wBol1318 and wBol1319).

The DNA mismatch repair protein MutL is truncated and thus, probably not functional. The MutL protein is part of the a complex with MutS and MutH that is essential for DNA mismatch reparation of erroneous base incorporation that can occur during DNA replication and recombination. The MutL subunit of this complex is responsible of the binding of the MutH subunit to DNA ("footprinting"). However these proteins are known to be highly recombinant in *Wolbachia* strains (Ellegaard et al. 2013) by their association with prophage elements particularly in *w*Pip and *w*Ha. Thus, the presence of a high level of recombination of these proteins in all *Wolbachia* strains may suggest that this protein is rather a strain specific protein than a male-killing specific protein since only a single genome from male-killing strain is currently sequenced.

The 90AA hypothetical protein does not contain any conserved domain but corresponds (97% of identities) to a fragment of the 412bp hypothetical protein gene wHa_03660. This *D. simulans* gene contains two conserved domains, the "AAA_25" motif and the "RecA-like_NTPases" motif, which is present in proteins involved in DNA recombination. In *Wolbachia* genome, this protein is commonly encoded by prophage sequences.

The baseplate assembly protein GpV, which is also a phage related protein, is likewise truncated.

These three identified genes are prophage genes that were identified as strain specific because they are truncated and therefore too short to be identified as orthologs of the corresponding proteins in the other *Wolbachia* genomes. Thus, they are more likely not good candidates for male-killing inducer genes.

C. Feminization related gene pattern

Five genes are found only in the genomes of feminizing strains, the putative permease gene wVul1344, the putative DNA repair protein radC gene wVul0645 and three genes annotated as hypothetical proteins: wVul0174, wVul1677 and a third one that is present in two identical copies (wVul0281 and wVul560). These three hypothetical proteins were arbitrarily named gene X, gene Y and genes Z, respectively.

1 - Analysis of the putative permease gene

The gene wVul1344 annotated as putative permease encodes a 119 AA protein, which does not contain any conserved domain. This gene is positioned in-between two other CDS, which are present in all the *Wolbachia* genomes, but the start codon of wVul1344 is only present in feminizing strain genomes (Figure 40).

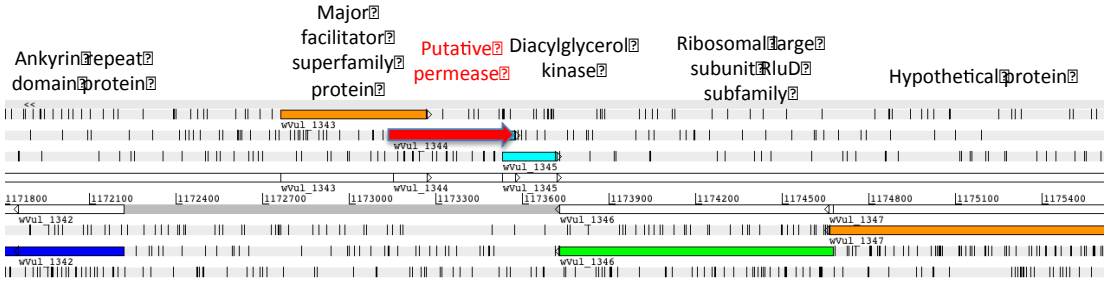


Figure 40: Position of the putative permease gene in *w*VulC genome.

A preliminary expression experiment by RT-qPCR confirmed that this gene is expressed at a low level in adult ovaries and gut tissues. To ensure that primers specifically amplify this gene, the reverse primer was designed at a central position of the gene that does not overlap with the two other adjacent genes.

2 - Analysis of the putative DNA repair protein RadC gene

The locus wVul0645 represents a fragment of the putative DNA repair protein RadC gene found in non-feminizing strains (Figure 40).

This protein was named after an *Escherichia coli* mutant (radC102) that lost the ability of DNA lesion repair, but its function remains unknown since the responsible mutation was finally identified in the recG gene (Lombardo and Rosenberg 2000). Additional experiments on *Streptococcus pneumoniae* suggested that this protein is not involved in DNA lesions reparation, in gene conversion nor in mismatch repair (Attaiech et al. 2008).

A mutation appeared in feminizing strains, which sub-divided the original protein in two CDS (wVul0645 & wVul0646) that have the same annotation. As an example, in the CI *w*Mel strains, the putative DNA repair protein RadC is 224 AA long, whereas both *w*VulC-fragmented proteins are 70 AA and 93 AA long, respectively. The stop codon dividing this gene in 2 separated CDS in *w*VulC was confirmed by mapping of both the HiSeq and MiSeq reads against the reference sequence.

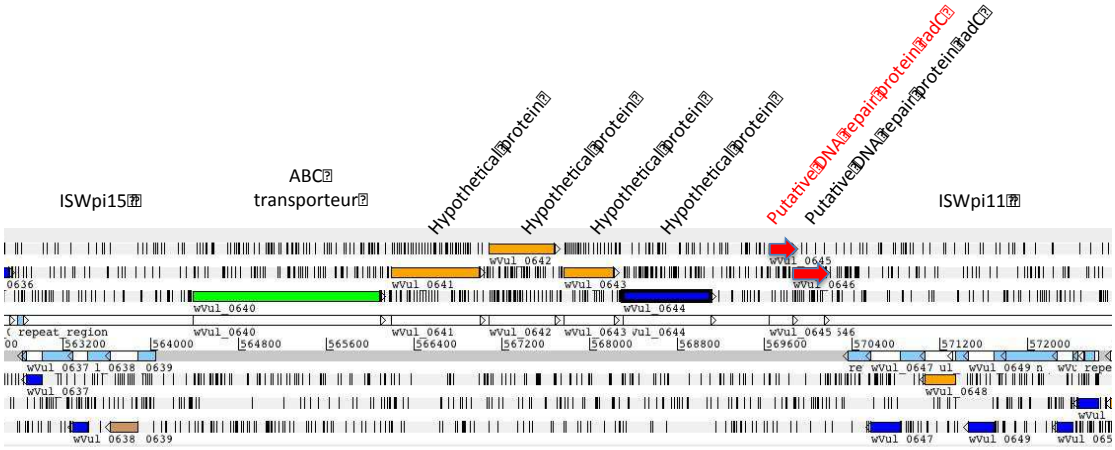


Figure 41: Position of the putative DNA repair protein radC gene in the *w*Vulc genome.

Expression assays on adult ovaries and gut tissues confirmed that both genes are equally expressed even if they do not form an operon: the RT-PCR across both genes is negative. The role of this protein in feminizing *Wolbachia* strains remains unknown.

3 - Analysis of the hypothetical protein gene “gene X”

The hypothetical protein gene wVul0174, arbitrarily named “gene X” encodes an 11.6 kDa protein, which is 106 AA long. This protein does not harbor any known conserved regions but produces a BLASTp hit with weak scores with a membrane protein containing SNARE domain protein from *Rhodopirellula baltica SH28* (27 % of identity over 66 AA).

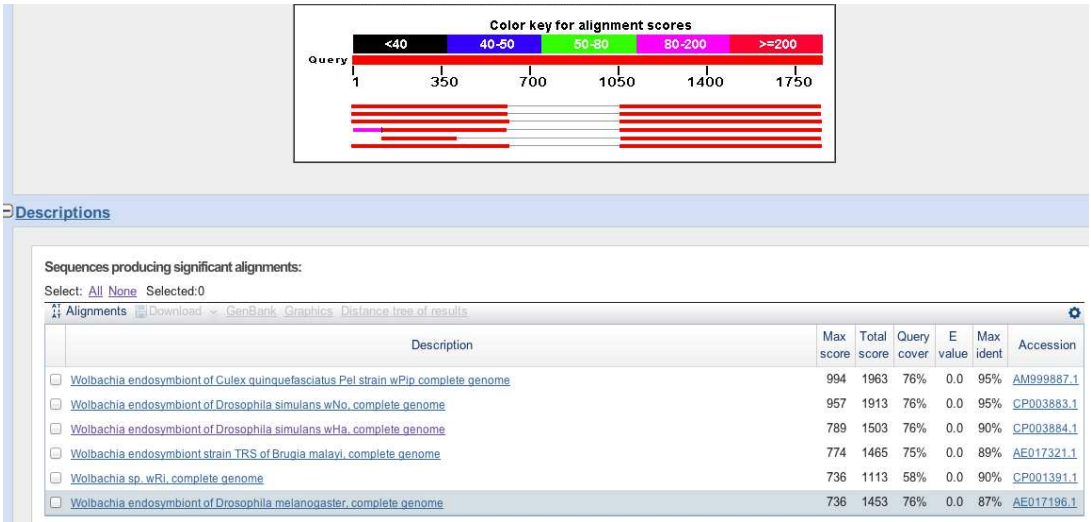


Figure 42: BLASTn of the gene X region on NCBI nt database.

A BLASTn of the gene X environment against the NCBI nucleotide (nr/nt) database revealed a gap in already published *Wolbachia* genomes at the positions corresponding to gene X (Figure 42). None of these genomes belongs to a feminizing strain.

To confirm the gap in the *Wolbachia* genomes sequenced from isopods, a BLASTn was run against the contigs of each genome (Figure 43). Among the published and sequenced genomes, gene X is present only in the feminizing strains (*w*VulC, *w*Albu, *w*Ase, *w*VulM) and the uncharacterized strain *w*Bre.

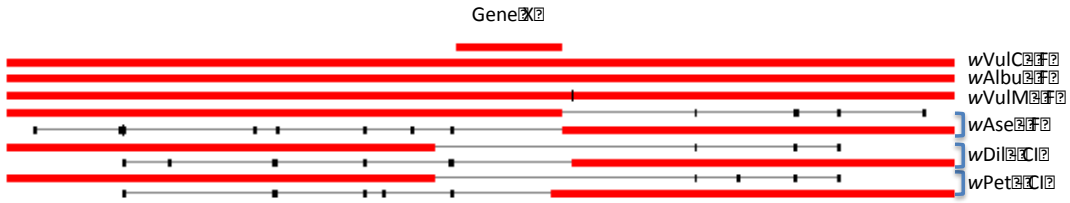


Figure 43: BLASTn of the gene X region on the sequenced strains. F: Feminizing. CI: Cytoplasmic incompatibility.

These results were then confirmed by PCR amplification. Two primer sets were designed to specifically amplify a ~ 200bp sequence inside the gene X (Figure 43; Appendix5 p214). For the sequenced strains, gene X is present in the genome of the feminizing *w*VulC and *w*Ase strains, and absent in the genome of the CI-inducing *w*Dil, *w*Pet strains. PCR results also confirm the absence of gene X in *w*ConV from *C. convexus*, the third known CI-inducing strain of isopods and its presence in the genomes of *w*Nas (from *A. nasatum*) and *w*Sca (from *P. scaber*), two of the most studied feminizing strains. However, this gene seems to be absent in the genome of the feminizing strain *w*Pru, which is phylogenetically closer to insect CI-inducing strains.

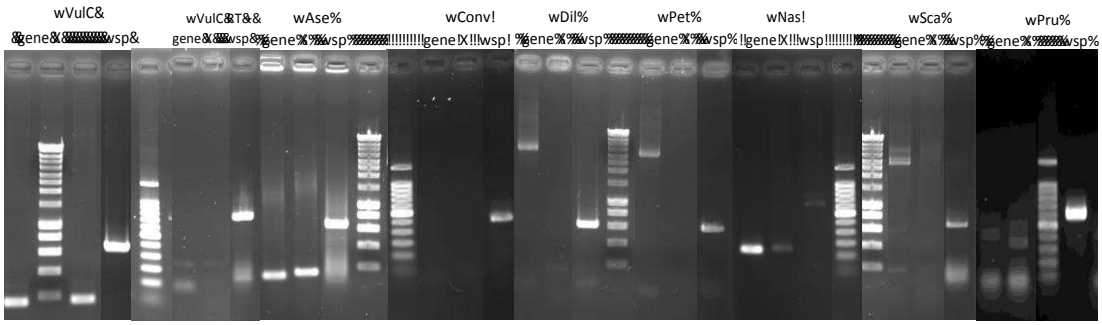


Figure 44: Screening of gene X by PCR on total DNA extract from the isopod *Wolbachia* strains. Amplification with two sets of primers, which produce a 166bp and 172bp fragment of *geneX*. *wsp*: Positive control of *Wolbachia* infection.

The genomic environment was investigated and gene X is positioned between two genes encoding proteins constitutive of a same secretion system: *secF* and *yidc*. Looking for

these two proteins in published genomes from non-feminizing strains, it appeared that the genomic synteny is well conserved around the gene X. Figure 45 illustrates this gene synteny for the feminizing strain *w*VulC, the CI strain *w*Pip and the mutualistic strain *w*Bm.

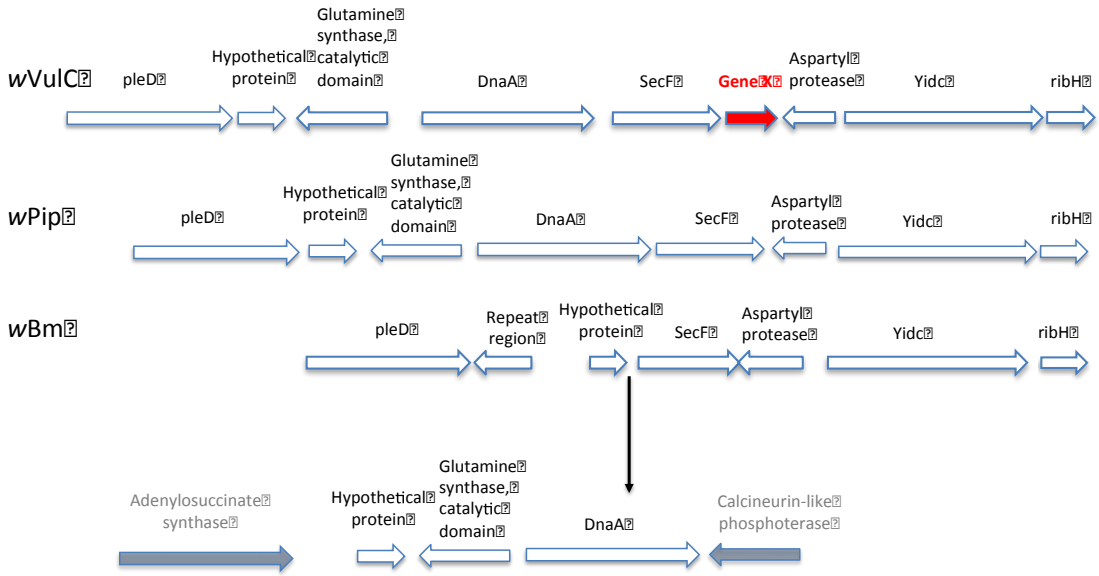


Figure 45: Position of gene X in the *w*VulC genome and gene environment in *w*Pip and *w*Bm.

Extragenic primers were designed around the gene X, in genes *secF* and *aspartyl protease* in order to investigate the potential co-transcription of the gene X with the secretion system. Interestingly, RT-PCR on adult ovaries of *A. vulgare* revealed that gene X is expressed (Figure 43) and co-transcribed with the *secF* gene but not with the *aspartyl protease* gene.

As all the feminizing *Wolbachia* strains of the phylogenetic isopod subgroup possess the gene X in their genome, these sets of primers could be used to screen other *Wolbachia* strains with unknown phenotype for the presence of the gene X and adjacent genes. This approach will allow to speculate whether these strains could be feminizing, as for *w*Bre, which possesses the gene X in its genome.

4 - Analysis of the hypothetical protein gene "gene Y"

The hypothetical protein gene wVul1677, arbitrarily named "gene Y" encodes an 11.2 kDa protein, which is 100 AA long. This protein does not harbor any known conserved domain and is located between two IS in *w*VulC (Figure 46).

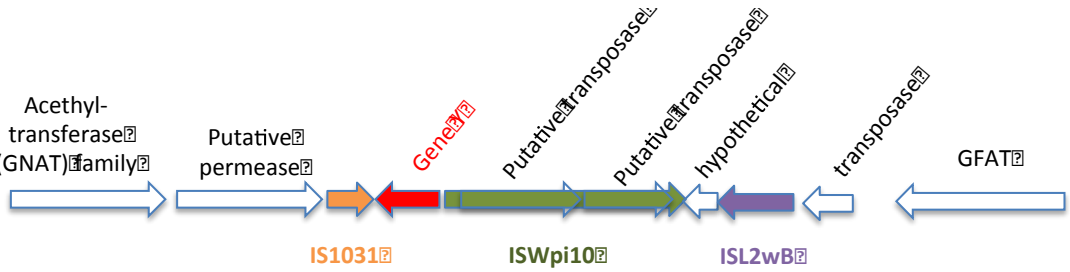


Figure 46: Gene environment of gene Y in the strain *w*VulC. CDS are in blue. Gene Y in red. IS sequences are in orange, green and purple. GFAT: glucosamine--fructose-6-phosphate aminotransferase (isomerizing).

Preliminary expression analyses in adult ovaries and gut tissues showed an absence of expression of this gene in adult tissues. Expression will be checked throughout post-embryonic development to verify whether this gene might only be expressed in the course of sexual differentiation.

The location of the gene Y between IS sequences suggests that this gene has been acquired by horizontal transfer. This acquisition might not be recent since all three IS are disrupted (Liu et al. 2013). A deeper analysis of these IS sequences would allow to decipher the acquisition history of this gene.

5 - Analysis of the hypothetical protein gene "gene Z"

The hypothetical protein gene, arbitrary named "gene Z", is present in 2 copies (wVul0281 and wVul0560) with 100% identity in the *w*VulC genome and encodes a 4.5kDa peptide, which is 39 AA long (Figure 46). This hypothetical protein does not harbor either any known conserved domain.

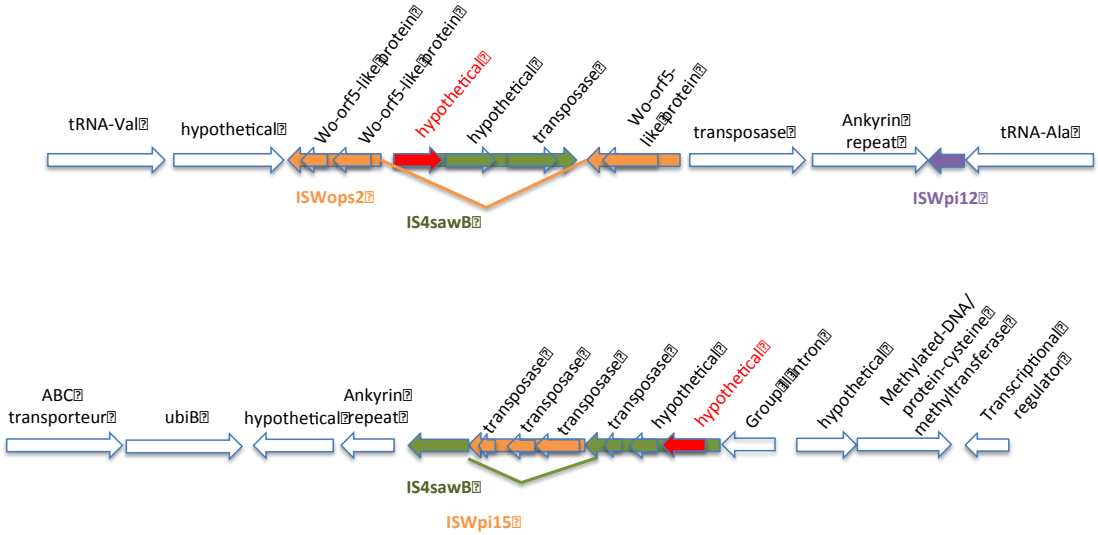


Figure 47: position of the 2 copies of the gene Z in *w*VulC genome.

Both copies are located in *w*VulC in the same IS family: IS4sa-wB, which is a family identified in 8 copies in *w*VulC and *w*Ri – *w*Mel are shearing 6 copies (N. Cerveau PhD thesis, (De Palmenaer et al. 2008).
The gene Z is one of the 3 ORFs of the IS4sa-wB with another hypothetical protein and a transposase gene. The location of the gene Z within an environment of disrupted IS (Liu et al. 2013) also suggests an ancient horizontal acquisition.

Discussion

Wolbachia induce a large repertoire of phenotypic effects in a wide range of arthropod species (Werren et al. 2008). This implies a particular ability to adapt to different cellular environments, which can be attributed to the genetic diversity of these strains. Molecular mechanisms involved in these phenotypes are still unknown. Comparative genomics as well as phylogenomics are powerful approaches that allow deciphering the processes, identifying responsible genes of this adaptation and giving insights of the genetic evolution of *Wolbachia* strains (Boussau et al. 2004; Touchon et al. 2009).
The comparative genomics of the published *Wolbachia* genomes and the genomes sequenced in this study, which represent 1/3 of the total sequenced strains, allowed us to establish the first pan-genome of the *Wolbachia* taxa based on a large number of strains. Created on 23 different strains, this *Wolbachia* pan-genome describes the full repertoire of *Wolbachia* genes composed by more than 6,400 orthologous genes describing an open pan-genome meaning that each new strain sequenced reveals new genes. Mathematical extrapolation of the pan-genomic accumulation curves confirmed the theoretical infinite size of the sets of genes and predicts 13,932 different genes after sequencing 100 *Wolbachia* strains. This high number attests of the extensive genomic diversity of *Wolbachia* strains suggesting that *Wolbachia* genomes are still evolving though gene acquisition and/or duplication that allow the emergence of new functions.
Three main factors are known to induce genomic variability: gene gain, gene loss and genome rearrangement (Figure 48) (Pallen and Wren 2007).

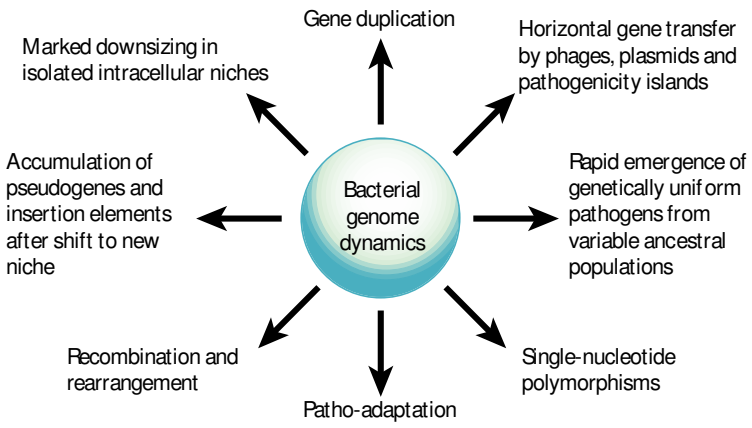


Figure 48: Bacterial genome dynamics (Pallen and Wren 2007).

Horizontal gene transfers or lateral gene transfers are the most potent source of innovation and variation for bacteria. If prokaryote-eukaryote gene transfers have been recently described for insertion of *Wolbachia* sequences inside host genome (Dunning Hotopp et al. 2007; Kondo et al. 2002b; Nikoh et al. 2008; Klasson et al. 2009; McNulty et al. 2012; Duplouy et al. 2013; Wu et al. 2013), horizontal gene transfer more frequently occurs between different strains and species of bacteria, due to the ecological connections of endosymbionts within arthropod intracellular environment (Rousset and Solignac 1995; Kondo et al. 2002a; Reuter and Keller 2003; Hiroki et al. 2004). Recent studies highlighted lateral transfers of insertion sequences between *Wolbachia*, *Cardinium* and *Rickettsia* bacterial endosymbionts (Duron 2013). However, genomic variations the most commonly found are duplications, which result in an increased number of key gene clusters, inviting scientist to pay attention to copy number variations allowed by whole-genome comparative genomic analyses.

The pan-genome analysis also reveals a high number of strain-specific genes that are phage-associated (about 15%) or IS-related (about 3%), highlighting the role of phages and horizontal transfers in the generation of genomic diversity in *Wolbachia* bacteria. The importance of mobile elements and bacteriophage fluxes in genome dynamics in endosymbionts have been extensively studied. (Wu et al. 2004; Klasson et al. 2009; Kent and Bordenstein 2010; Kent et al. 2011b; Leclercq et al. 2011; Gillespie et al. 2012; Penz et al. 2012; Duplouy et al. 2013). These newly closely related sequenced genomes should contribute to the understanding of the phage and mobile element implications on the *Wolbachia* genome evolution.

A deeper analysis of the functional categories of orthologous proteins such as biosynthesis pathways, transporters, regulators, drug resistance, antibiotic

biosynthesis/resistance, contained in the dispensable-genome should also reveal insights of the adaptation processes to their different cellular environments.

Comparison with a pan-genome reduced to selected host *Wolbachia* strains, as performed for the isopod *Wolbachia* strains, highlights the adaptability of the bacteria to its host environment. With the emergence of the NGS technologies multiple-strains sequencing projects will arise, allowing the comparison of host-species-specific pan-genome (e.g. insect vs. isopod; mosquito vs. drosophila, etc.) that aid our understanding of bacterial host-adaptation. Classification of orthologous protein clusters depending on their presence or absence should show strain-specific gene gains or losses and depending of their biological significance highlights genes potentially important for symbiosis. Indeed, attenuation/increase of the virulence for example may be due to the loss/gain of specific genes. The study of gene losses occurred in all mutualistic genomes should also help the understanding of key processes supplied by host genomes.

The *Wolbachia* genome dynamics and evolution was also illustrated by a reliable phylogenomy. More than just illustrate the well-known supergroup division of the *Wolbachia* strains, this phylogenomic approach reveals that the *Wolbachia* B-supergroup may be composed of two subgroups, one clustering isopod *Wolbachia* strains and the other one the *Wolbachia* strains from insects. The resequencing of the *w*Pru genome is necessary to validate this hypothesis. Even inside the subgroup of isopod *Wolbachia* strains, positioning of CI and feminizing strains has been modified from single-gene phylogenies (Bouchon et al. 1998; Pichon et al. 2009; Cordaux et al. 2012), illustrating another example of the strength of phylogeny built with hundreds of genes (Dagan 2011). Keeping in mind the subgroup of feminizing strains, it would be interesting to integrate in this analysis other feminizing strains from non-isopod hosts such as the *Wolbachia* strains from the butterfly *Eurema hecabe* (Hiroki et al. 2002) and the grass-dwelling leafhopper *Zyginidia pullula* (Negri et al. 2006). Actually, based on single-gene *wsp* phylogeny, both of them are B-supergroup *Wolbachia* and close to the *Wolbachia* strain of *Acraea encodon*, which is also close to *w*Pru letting speculate that these 3 strains might be phylogenetically close and a phylogenomy as performed in this study should establish the veracity of another feminizing strains subgroup. In another hand, the cluster of “atypical” phenotype strains should be more closely investigate since the basis of the phenotype evolution might be illustrated by these really closely related strains inducing different phenotypes.

To identify phenotype-specific genes, comparative genomics was performed on *Wolbachia* genome clusters established depending on the phenotype induced by *Wolbachia*. This approach mainly revealed interesting proteins regarding the mutualism

and feminization phenotypes. Indeed, one of the mutualistic strain-specific proteins identified is a WASP protein, which is known to interact with actin filaments (Dean 2011) and may be involved in the mobility of the bacteria, particularly in *B. malayi* in which *Wolbachia* ovaries reinfection is delayed during development. Additionally, the genes present in all but mutualistic strain genomes have to be studied with more attention since their absence may have a link with the obligate life style of these strains (Moran 2001; Moran 2002; Foster et al. 2005; Murfin et al. 2012). Concerning the 5 genes identified as only present in feminizing strain genomes, the gene X drew our attention since it is localized downstream the secF gene and upstream to the yidC gene which are part of the Sec export system (du Plessis et al. 2011) suggesting that it might be translocated by this secretion system.

Since this gene is absent in the genome of the feminizing strain *w*Pru, it will be interesting to check its absence in the genome of the two other closely related feminizing *Wolbachia* strains of *E. hecabe* and *Z. pullula*, that will suggest another feminization mechanism for this second feminizing strain subgroup.

Surprisingly no genes specific to male-killing, CI nor parthenogenesis have been identified even though all these strains interfere with the first cell division cycle of the embryo (Werren et al. 2008; Cordaux et al. 2011). In addition, two sets of genes shared by 3 or 4 phenotypes have not been investigated in this study but might highlight particular processes necessary for molecular mechanisms of several phenotypes. Regarding the male killing-specific genes, none of the 3 proteins identified in the *w*Bol1-b strain might be involved in this phenotype. Genes involved in the male-killing phenotype were also investigated when studying the genome of the *w*Bol1-b strain (Duplouy et al. 2013) by looking for the genes present in the *w*Bol1-b strain but absent from the genome of the closely-related CI strain *w*Pip. They identified with a low stringency E. value threshold of 10, 44 *w*Bol1-b strain-specific proteins of which nine were functionally annotated. A comparison of these results with our higher stringency E. value threshold of 10^{-5} will determine if these genes are specific to male-killing strain or specific to the *w*Bol-1 strain.

Experimental confirmations of these *in silico* results, such as gene expression and protein interactome studies, particularly when the bacteria induce a specific phenotype (i.e. during fecundation or development), will decipher the potential role of these candidate proteins.

WOLBACHIA-HOST INTERACTOME INVESTIGATION

Introduction

The investigation of interacting proteins between hosts (*B. malayi* or *A. vulgare)* and their *Wolbachia* symbionts can aid in the identification of symbiosis effectors. These proteins may be part of the underlying biochemical basis of the obligate endosymbiotic relationships of *Wolbachia* and their hosts. Further, inhibition of these protein-protein interactions may provide information towards a potential drug target set for filariasis disease. The analysis of *Wolbachia* genome sequences revealed the presence of several secretion/export systems that are likely involved in the secretion of numerous identified putative translocated effectors, including proteins containing eukaryote-like motifs and surface membrane proteins (Wu et al. 2004; Foster et al. 2005; Mavingui et al. 2012; Darby et al. 2012; Duplouy et al. 2013; Liu et al. 2013). Indeed, secretion systems such as the T4SS are key factors in bacterial virulence (Baron and Coombes 2007; Durand et al. 2009; Hicks and Galan 2013; Li and Carlow 2012) and are well-conserved in *Wolbachia* genomes (Pichon et al. 2009). Secretion systems of endogenous bacteria have been shown to secrete bacterial effectors into the cytoplasm of host cells including proteins with eukaryote-like motifs (Lin et al. 2007; Rikihisa et al. 2009; Al-Khedery et al. 2012), which interact with host proteins. Eukaryote-like motif containing proteins identified in *Wolbachia* genomes are known to be involved in protein-protein interactions as well as in enzyme activities (Table 15) (Liu et al. 2013).

Table 15: Chart of Eukaryote-like motifs found in *Wolbachia* genomes.

Motif/domain	
Ankyrin repeat	ANK
Tetratricopeptide repeat	TPR
Leucine-rich repeat	LRR
WD-40	WD
SNARE-associated Golgi protein	
Ras-like GTPase	
PKCI-related	
Protein kinase	PKc
Phospholipase	PL
GDP/GTP exchange factor	GEF
Acetyltransferase domain	GNAT

We first established the repertoire of selected putative bacterial effector proteins by the search of specific conserved motifs: ankyrin repeat and tetratricopeptide repeat (TPR) domains, complemented by reciprocal BLAST. In parallel, we also investigated the putative role of selected *Wolbachia* proteins in the interaction with their hosts by both proteomic and transcriptomic approaches.

Two classes of proteins have been targeted: (1) proteins involved in secretion systems and (2) putative bacterial effectors: proteins containing ankyrin repeat domains (ANK) and outer membrane proteins, which are likely to interact with host proteins and modify physiological processes.

I. Screening of putative symbiosis-related bacterial effectors

This investigation was focused on ankyrin repeat domain-containing proteins (ANK) and tetratricopeptide repeat proteins (TPR), two main domains found in eukaryote proteins that are involved in protein-protein interaction, and also found in bacterial genomes (Al-Khodor et al. 2010b; Cerveny et al. 2013).

ANK proteins are mainly involved in protein-protein interactions in eukaryotes inducing a large variety of functions such as cell cycle regulation, transcriptional regulation, mitochondrial enzymes, cytoskeleton interactions, signal transduction and toxins (Sedgwick and Smerdon 1999; Mosavi et al. 2004; Voronin and Kiseleva 2007).
Although originally described as eukaryote-specific proteins, ANK proteins were found in bacteria, especially endosymbiotic bacteria, such as *Legionella pneumophila*, *Anaplasma phagocytophilum*, *Coxiella burnetii* and *Wolbachia* (Luhrmann et al. 2010; Gomez-Valero et al. 2011; Ramabu et al. 2011; Wakeel et al. 2011; Siozios et al. 2013).
These eukaryote-like proteins are believed be acquired through horizontal gene transfer from eukaryotic host DNA sequences. Some of these proteins play key functions in bacterial pathogenesis by mimicking or interfering with the host functions (Walker et al. 2007; Habyarimana et al. 2008; Pan et al. 2008; Al-Khodor et al. 2010a). ANK proteins contain several copies of 33-residue repeating segments that adapt helix-turn-helix confirmations and comprise antiparallel α-helices (Figure 49 A).
ANK proteins are remarkably abundant in *Wolbachia* genomes and have been hypothesized to be involved in symbiont-host interactions (Pan et al. 2008; Slatko et al. 2010; Papafotiou et al. 2011; Siozios et al. 2013). To date, the isopod *Wolbachia* strain *w*VulC genome is the published genome that contains the largest number of ANK proteins (86) including a phage-related ank gene, the *pk2b2* gene, which expression was recently demonstrated as specific to feminizing *Wolbachia* (Pichon et al. 2012; Liu et al. 2013).

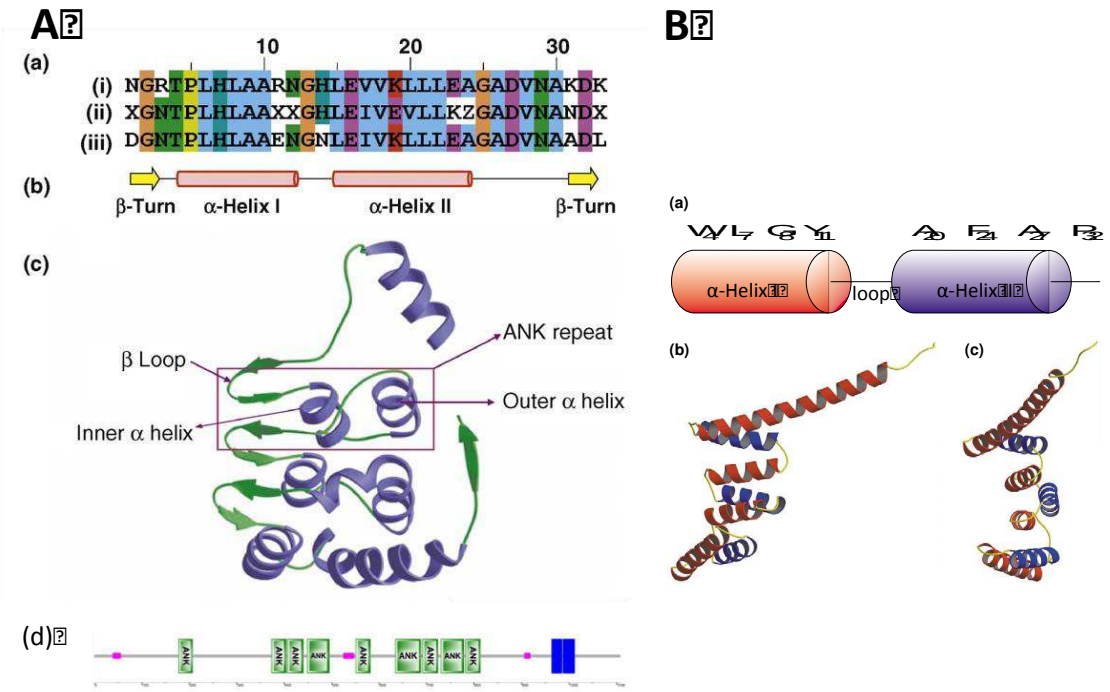


Figure 49: Conserved structural feature of A: the ANK motif, B: the TPR motif. A: (a) Consensus sequences of ANK motifs. (b,c) structural composition of ANK repeats (Al-Khodor et al. 2010b). (d) Structural features of the ANK wBm0447 from the *Wolbachia* *w*Bm strain. B: (a) Schematic representation of the secondary structure of TPR motif. (b,c) Front and perpendicular views of three TPR motifs of a phosphatase protein (D'Andrea and Regan 2003).

The tetratricopeptide repeat motif, originally identified in yeast (Hirano et al. 1990; Sikorski et al. 1990), is a protein-protein interaction module found in multiple copies in a large number of functionally unrelated proteins from all genera, in bacteria as well as in humans, and in a variety of cell localizations (Cerveny et al. 2013). Mutations of particular human TPR-containing proteins have been hypothesized to be responsible of human diseases such as the Down syndrome (Tsukahara et al. 1998). However, this motif is minimally conserved, presenting variability. The consensus motif is a degenerate 34-amino acid repeated sequence (W_4-L_7-G_8-Y_{11}-A_{20}-F_{24}-A_{27}-P_{32}). The TPR motif contains two anti-parallel α-helices, which generate a helical structure with both hydrophobic and hydrophilic regions that might accommodate the complementary region of a target protein (Figure 49 B). TPR proteins are associated with multi-protein complexes and facilitate specific protein-interactions playing an important role in the functioning of chaperone, cell cycle, transcription, and protein transport complexes such as secretion systems (Blatch and Lassle 1999). TPR-containing proteins of bacterial pathogens have been reported to be directly related to virulence-associated functions such as translocation of virulence factors into host cells and adhesion to host cells

(Broms et al. 2006; Edqvist et al. 2006; Cerveny et al. 2013). Class II chaperones of the type III secretion system are the most investigated TPR containing proteins (Pallen et al. 2003). In *Legionella pneumophila*, the strain JR32 contains two virulence-associated T4SSs, the Dot/Icm and the Lvh T4SSs. Functional defect of the Dot/Icm T4SS induces an overexpression of the TPR-containing protein LpnE and EnhC that take over the translocation of Dot/Icm T4SS protein substrates (Bandyopadhyay et al. 2012).

To identify the *Wolbachia* genes repertoire of these two protein families, we first screened the *Wolbachia* genomes for protein family motif recognition and then we supplemented this first analysis with a reciprocal BLAST approach for ANK proteins. Hence, this putative bacterial effector proteins screening was performed using two complementary methods; while the first method targets specific conserved motifs with a non-*a priori* approach, allowing the "blind" detection of new genes, the second method is based on a reciprocal BLAST approach to identify orthologs of already annotated proteins. This homology-finding method is particularly useful for the identification of proteins for which the conserved motif may be disrupted or fragmented, a common phenomenon in *Wolbachia* genomes due to mobile elements (Wu et al. 2004; Liu et al. 2013; Cerveau et al. 2011).

A. Hidden Markov Model (HMM) profiles search

In order to perform non-*a priori* approaches, Hidden Markov Model profiles were used locally with the software HMMER3 (Eddy 2009, 2011; Johnson et al. 2010).

An HMM profile is a probabilistic model of a protein family used for searching a database for other members of this family. This profile is created from a group of sequences aligned by structural or sequence similarity and is composed of the statistical profile and a weight matrix of the possibilities (Gribskov et al. 1987).

Databases of HMM profiles for each eukaryote-like protein family were generated from models downloaded from the Pfam website (http://pfam.sanger.ac.uk). Table 16 lists the models used for the establishment of the ANK and the TPR model databases.

Table 16: Chart of the HMM motifs used for the creation of the eukaryote-like motif databases.

Ankyrins	TPR					
Ank	Adaptin_N	CRM1_C	HEAT_EZ	Neurochondrin	Sel1	TPR_2
Ank_2	Apc3	Cse1	HEAT	Nro1	SHNi-TPR	TPR_3
Ank_3	Apc5	DNA_alkylation	HEAT_PBS	NSF	SPO22	TPR_4
Ank_4	API5	Drf_FH3	HemY_N	Paf67	ST7	TPR_5
Ank_5	Arm_2	Drf_GBD	IBB	ParcG	Suf	TPR_6
DUF3420	Arm	DUF1822	IBN_N	PC_rep	SusD	TPR_7
DUF3447	Avirulence	DUF2225	IFRD	PHAT	SusD-like_2	TPR_8
PRANC	BTAD	DUF2435	KAP	PI3Ka	SusD-like_3	TPR_9
SPX	CAS_CSE1	DUF3385	Leuk-A4-hydro_C	PPP5	SusD-like	TPR_MLP1_2
	ChAPs	DUF3458	LRV_FeS	PPR_1	Tcf25	Upf2
	CHAT	DUF3574	LRV	PPR_2	TOM20_plant	Vac14_Fab1_bd
	CLASP_N	DUF3808	MA3	PPR_3	TPR_10	V-ATPase_H_C
	Clathrin	DUF3856	Methyltransf_31	PPR	TPR_11	V-ATPase_H_N
	Clathrin-link	DUF3928	MIF4G	Proteasom_PSMB	TPR_12	Vitellogenin_N
	Clathrin_propel	EST1_DNA_bind	MIF4G_like_2	PTPS_related	TPR_14	Vps39_1
	Cnd1	FAT	MIF4G_like	PUF	TPR_15	W2
	Cnd3	Foie-gras_1	Mo25	Rab5-bind	TPR_16	Xpo1
	Coatomer_E	GUN4	MRP-S27	Rapsyn_N	TPR_17	YfiO
	Cohesin_HEAT	HAT	NARP1	RPN7	TPR_18	
	Cohesin_load	HEAT_2	NB-ARC	Sdh5	TPR_1	

By searching the published *Wolbachia* genomes for ANK domains, all the annotated proteins were identified with some additional proteins, which correspond to pseudo-genes (Figure 50). The correct identification of ANK proteins of already annotated *Wolbachia* genomes validates the methodology then used for the identification of ANK proteins in the newly-sequenced isopod *Wolbachia* strains. As already noticed for the *w*VulC genome (Liu et al. 2013), some isopod-infecting *Wolbachia* harbor a larger number of ANK proteins than other *Wolbachia* genomes.

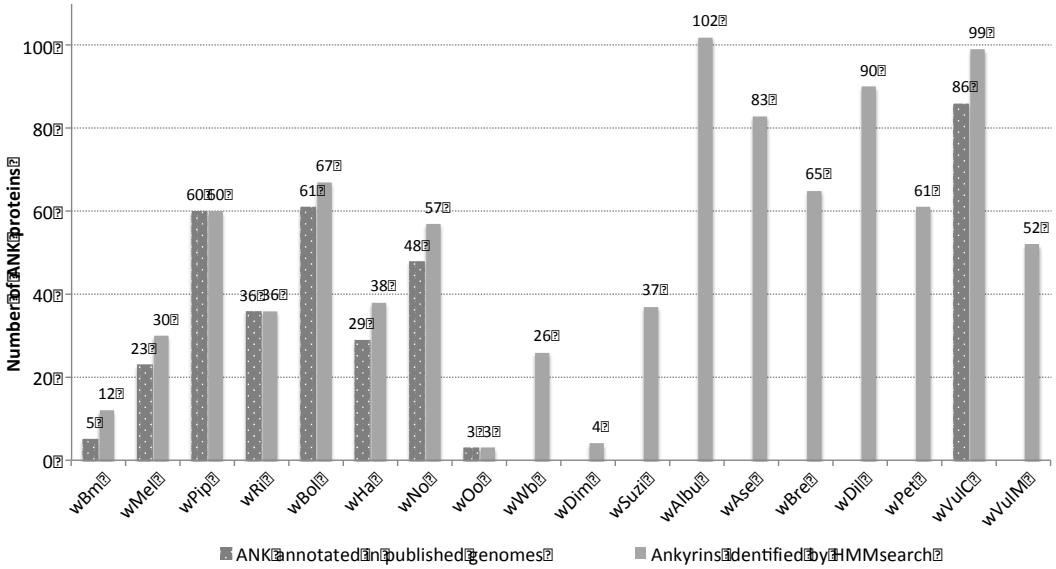


Figure 50: Comparison of the number of Ankyrin proteins found by HMM search with the genome annotation when available.

Similar results were obtained during the second round of eukaryote-like protein screen targeting the TPR proteins: a large number of proteins were identified in all the genomes (Figure 51). However, since this motif is found in a large diversity of proteins involved in different functions, their annotation is rarely linked to the presence of the TPR motif but rather to their function. Nonetheless, it is interesting to note that the number of TPR proteins found in all the genomes is less variable than the ANK proteins. All *Wolbachia* genomes harbor about 30 TPR proteins except for the two closely related *Wolbachia* strains *w*Ri and *w*Suzi that harbor ~10 additional TPR proteins.

To date, this eukaryote-like motif-containing protein family has not been investigated in *Wolbachia* genomes, nor in the two closely related intracellular bacteria *Anaplasma phagocytophilum* and *Ehrlichia chaffeensis* and remains of high interest since they are known to be involved in protein-protein interactions in other bacteria such as *Legionella* (Bandyopadhyay et al. 2012).

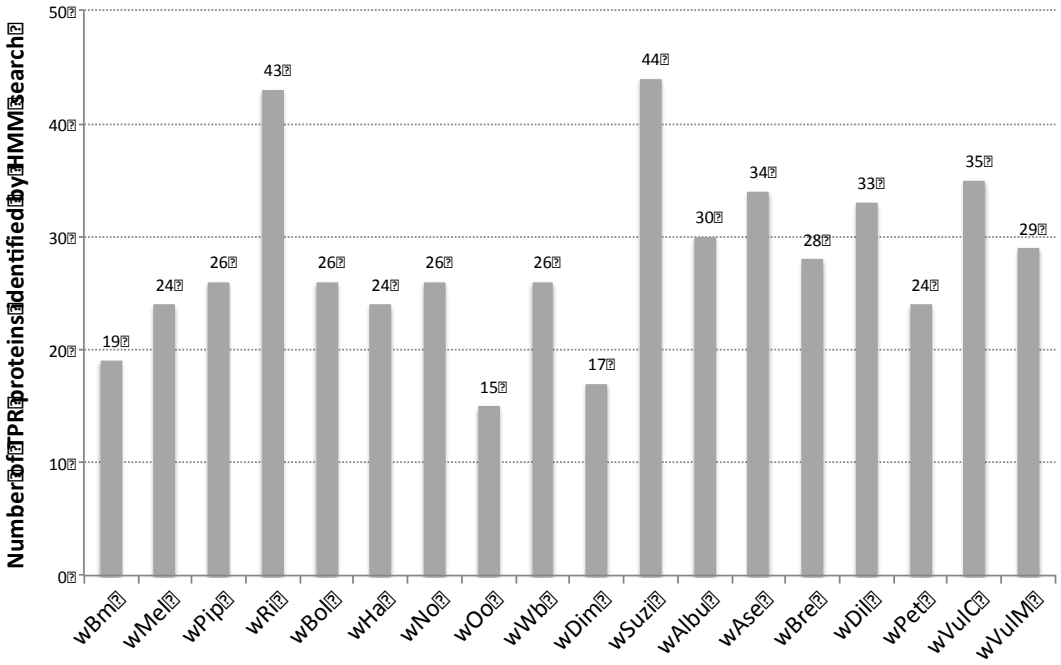


Figure 51: Number of proteins containing a TPR motif identified by HMM search.

These results strengthen that *Wolbachia* genomes present an exceptionally large number of eukaryote-like proteins including ankyrin and TPR-containing proteins, which can be considered as a specific characteristic of this bacterial taxon and might be at the basis of its particular symbiosis relationships.

B. Reciprocal BLASTp

In order to complement this first analysis, ankyrin proteins identified by HMM approach were searched on the reciprocal BLAST output previously performed for phylogenomic analyses (Table 17).

Out of the 1,061 ankyrin proteins previously identified by HMM search, 789 were found in 140 orthologous clusters that contain 1524 proteins (APPENDIX: FILE OGroups_ANK). This result means that 262 ANK identified by HMM search are strain-specific whereas 651 additional putative ANK (homologs to query sequences) were identified by reciprocal BLAST (Figure 52).

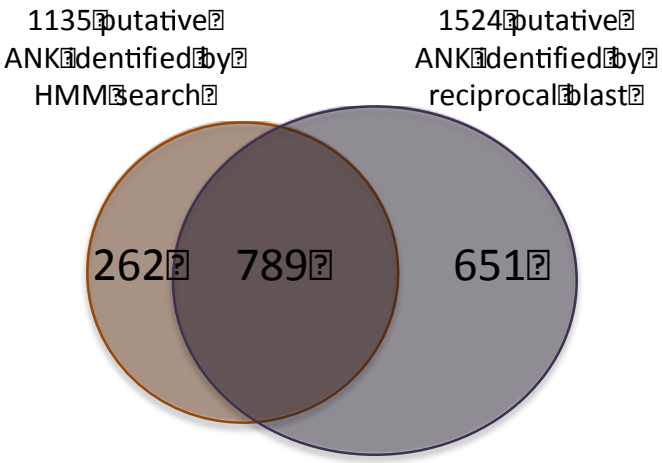


Figure 52: Comparison of ANK screenings by motif search and by reciprocal blast.

A preliminary search for these 651 new putative ANK proteins in the NCBI databases (BLASTn and BLASTp) reveals that some of them are annotated with other functions such as transposase or NADH dehydrogenase, or as hypothetical protein that do not harbor any conserved domain. These proteins will be manually investigated to determine the reliability of the reciprocal BLAST investigation. If the identified proteins that do not harbor the ANK repeat motif are adjacent to truncated proteins containing ANK repeat motifs, they might be real ANK fragments. But these fragmented proteins are likely pseudo-genes that do not harbor any active function anymore. Thus, these proteins will not be retained as candidate for bacterial effectors.

Nonetheless, these results are very encouraging inviting to more detailed investigations on all eukaryote-like motifs to establish the whole *Wolbachia* repertoire of these particular proteins.

Table 17: Chart of Ankyrin protein orthologs. Example of 50 of the 140 orthologous clusters identified as ankyrin proteins.

	*w*VulC	*w*Albu	*w*Ase	*w*Pet	*w*VulM	*w*Bre	*w*Dil	*w*Pip	*w*JHB	*w*No	*w*AlbB
OGrps_5:	wVul1644	wAlb1343 wAlb1731	wAse1340 wAse1679 wAse1621	wPet669 wPet809 wPet914 wPet877	wVulM543 wVulM154	wBre652 wBre946	wDil1247 wDil1618	YP_001975093.1 YP_001975243.1 YP_001975190.1 YP_001975094.1	ZP_03335206.1 ZP_03335447.1 ZP_03335477.1 ZP_03335567.1 ZP_03335003.1 ZP_03335207.1	YP_007885427.1 YP_007886307.1	ZP_09542016.1
OGrps_34:	wVul0779 wVul0908 wVul1312 wVul1601 wVul1667	wAlb1090 wAlb805	wAse960 wAse1404	wPet664 wPet1626 wPet521	wVulM123 wVulM299	wBre307	wDil1570	YP_001975062.1 YP_001975116.1 YP_001976039.1	ZP_03335514.1	YP_007885445.1 YP_007886171.1	ZP_09541999.1
OGrps_39:	wVul1696	wAlb45 wAlb1750	wAse158	wPet1631 wPet1467	wVulM82 wVulM28 wVulM975	wBre981 wBre1335	wDil415	YP_001975534.1 YP_001975533.1	ZP_03334387.1	YP_007885736.1 YP_007885737.1	ZP_09542608.1
OGrps_44:	wVul1049 wVul0887 wVul1581	wAlb1998 wAlb1346	wAse874 wAse1085 wAse1089	wPet1318 wPet1289	wVulM214	wBre1530 wBre90	wDil1469 wDil1255	YP_001975100.1 YP_001975196.1 YP_001975140.1	ZP_03335442.1 ZP_03335561.1 ZP_03335495.1	YP_007885431.1 YP_007886010.1	
OGrps_50:	wVul0888 wVul1582 wVul1048	wAlb1999	wAse1088 wAse1297	wPet518 wPet1316	wVulM1525 wVulM1022	wBre170	wDil1750	YP_001975101.1 YP_001975197.1 YP_001975139.1	ZP_03335441.1 ZP_03335560.1 ZP_03335496.1	YP_007886009.1 YP_007885432.1	
OGrps_59:	wVul0684	wAlb1092 wAlb1257 wAlb364	wAse685 wAse1192	wPet857 wPet1085 wPet277	wVulM164	wBre994	wDil1351 wDil904	YP_001975234.1 YP_001976067.1	ZP_03335196.1 ZP_03335643.1 ZP_03335588.1 ZP_03335123.1	YP_007886274.1	ZP_09542901.1
OGrps_62:	wVul0091	wAlb986 wAlb890	wAse817	wPet545 wPet301	wVulM508 wVulM545	wBre1100 wBre1032 wBre699	wDil1562 wDil454	YP_001974900.1	ZP_03335018.1	YP_007885386.1	ZP_09542380.1
OGrps_71:	wVul0815	wAlb1161	wAse742 wAse839	wPet1530 wPet1102	wVulM553	wBre1041 wBre1078	wDil264 wDil959	YP_001975443.1	ZP_03335411.1	YP_007885540.1	ZP_09542325.1
OGrps_83:	wVul0884	wAlb591 wAlb877	wAse697		wVulM243			YP_001975097.1 YP_001976025.1	ZP_03335564.1 ZP_03335687.1		
OGrps_94:	wVul0802	wAlb80 wAlb1206	wAse821 wAse822	wPet1422 wPet54	wVulM18 wVulM742 wVulM17	wBre973 wBre1763	wDil716	YP_001975179.1	ZP_03335458.1	YP_007886302.1 YP_007886368.1	ZP_09542814.1
OGrps_99:	wVul0064	wAlb231	wAse542	wPet218	wVulM266	wBre1070 wBre1152	wDil407	YP_001975028.1	ZP_03334888.1	YP_007886355.1	ZP_09542097.1
OGrps_136:	wVul1374	wAlb2014	wAse756	wPet1331	wVulM1307	wBre1242	wDil863	YP_001975629.1	ZP_03334475.1	YP_007885839.1	ZP_09542462.1
OGrps_936:	wVul0031	wAlb1808	wAse125	wPet1488	wVulM1136	wBre411	wDil340	YP_001976012.1	ZP_03334863.1	YP_007886334.1	ZP_09542072.1
OGrps_1079:	wVul0046	wAlb1145	wAse82	wPet571	wVulM676	wBre725	wDil159				
OGrps_923:	wVul0112	wAlb1768		wPet557		wBre656	wDil1242	YP_001975955.1	ZP_03334807.1	YP_007886139.1	
OGrps_1080:	wVul0163	wAlb2521		wPet341	wVulM402			YP_001975452.1	ZP_03335420.1		
OGrps_819:	wVul0182	wAlb1550	wAse474	wPet1020	wVulM989	wBre870	wDil976	YP_001975451.1	ZP_03335419.1	YP_007885533.1	ZP_09542332.1
OGrps_1024:	wVul0250	wAlb1883	wAse1283 wAse1759 wAse608	wPet1190	wVulM1178		wDil527 wDil1240			YP_007886212.1	
OGrps_1116:	wVul0274				wVulM1415				wAlb|wAlb121		
OGrps_1595:	wVul0286	wAlb2227									
OGrps_1081:	wVul0303	wAlb1433			wVulM1298	wBre343		YP_001974934.1	ZP_03334983.1		
OGrps_1901:	wVul0544										
OGrps_1425:	wVul0545						wDil1328				
OGrps_1426:	wVul0554		wAse669			wBre1232					
OGrps_932:	wVul0361 wVul0504	wAlb8 wAlb1643	wAse1584		wVulM2 wVulM744	wBre1137	wDil1310 wDil1442 wDil1814				
OGrps_1606:	wVul0502						wDil1713				
OGrps_1630:	wVul1342		wAse1610								
OGrps_254:	wVul0378	wAlb2055 wAlb1037	wAse250	wPet1509 wPet795	wVulM737	wBre1180 wBre1030 wBre649	wDil164	YP_001976008.1	ZP_03334859.1	YP_007886331.1	ZP_09542068.1
OGrps_1395:	wVul1060 wVul1573						wDil751				
OGrps_902:	wVul0533	wAlb2530	wAse452	wPet1546	wVulM1640	wBre1225	wDil822		ZP_03334563.1		ZP_09542598.1
OGrps_1103:	wVul0546 wVul0671 wVul0672		wAse1637							YP_007886287.1	
OGrps_884:	wVul0553	wAlb2543	wAse668	wPet887	wVulM1576	wBre233	wDil563		ZP_03334944.1		
OGrps_1122:	wVul0667 wVul1083	wAlb1510		wPet558 wPet556	wVulM726						
OGrps_784:	wVul0623 wVul0624	wAlb160 wAlb1757	wAse387	wPet76 wPet1059	wVulM72 wVulM1563	wBre326 wBre1001	wDil966	YP_001975547.1 YP_001975548.1	ZP_03334399.1		ZP_09542193.1
OGrps_1227:	wVul0715 wVul0778 wVul1602	wAlb1680			wVulM300						
OGrps_1441:	wVul1065	wAlb1902			wVulM542						
OGrps_1052:	wVul1064	wAlb1437			wVulM862			YP_001976101.1 YP_001976102.1	ZP_03335035.1 ZP_03335036.1	YP_007886176.1	
OGrps_1335:	wVul1121	wAlb1182						wPru791		wPet596	
OGrps_303:	wVul0812	wAlb1164	wAse842	wPet1533	wVulM550	wBre1038	wDil267	YP_001975440.1	ZP_03335408.1	YP_007885543.1	ZP_09542322.1
OGrps_1264:	wVul1268	wAlb1243		wPet1169	wVulM866	wBre603					
OGrps_1458:	wVul1633	wAlb990					wDil813				
OGrps_1460:	wVul1645									YP_007886371.1	
OGrps_1462: OGrps_1934:	wVul1651 wVul1652				wVulM679						
OGrps_1463:	wVul1658		wAse1311				wDil594				
OGrps_941:	wVul0849	wAlb555	wAse284	wPet761	wVulM483	wBre1618	wDil59			YP_007885775.1	ZP_09542539.1
OGrps_588:	wVul0972	wAlb2274	wAse601	wPet1504	wVulM1540	wBre1717	wDil862	YP_001976099.1	ZP_03335033.1	YP_007886255.1	ZP_09542138.1
OGrps_965:	wVul0974	wAlb2276	wAse465	wPet1506	wPru538	wBre1715	wDil684 wDil860		ZP_03335465.1	YP_007886253.1	ZP_09542140.1
OGrps_1028:	wVul0978	wAlb450	wAse461	wPet202	wVulM180		wDil688	YP_001975177.1	ZP_03335460.1	YP_007886249.1	
OGrps_967:	wVul1266	wAlb1339	wAse656		wVulM769	wBre401	wDil202	YP_001975417.1	ZP_03335386.1	YP_007885656.1	ZP_09542809.1

*w*Ha	*w*Bol	*w*VitB	*w*Bol	*w*Mel	*w*Ri	*w*Suzi	*w*Wil	*w*Uni	*w*Dim	*w*Oo	*w*Ban	*w*Bm
YP_007888737.1	wBol1253	ZP_08970164.1	wBol366	NP_965965.1	YP_002726751.1	wSuzi98	ZP_01314584.1	ZP_03787926.1	wDim118		wBan60	
YP_007889074.1		ZP_08970119.1	wBol367		YP_002727224.1	wSuzi424	ZP_01314212.1	ZP_03788397.1			wBan61	
		ZP_08970155.1	wBol365			wSuzi428	ZP_01314200.1				wBan759	
		ZP_08970156.1	wBol69			wSuzi654						
		ZP_08970154.1										
YP_007889003.1	wBol241			NP_966365.1	YP_002727130.1	wSuzi540	ZP_01314677.1					
	wBol266				YP_002727524.1	wSuzi664						
	wBol3					wSuzi976						
		ZP_08970438.1	wBol775	NP_966988.1		wSuzi1237	ZP_01314834.1	ZP_03788112.1			wBan339	YP_197970.1
		ZP_08970432.1									wBan665	YP_197969.1
		ZP_08970439.1									wBan337	
											wBan338	
											wBan336	
YP_007888989.1	wBol343			NP_966401.1	YP_002727112.1	wSuzi522	ZP_01314887.1					
	wBol353				YP_002727506.1	wSuzi958						
YP_007888990.1	wBol344		wBol249	NP_966402.1	YP_002727113.1	wSuzi523	ZP_01314886.1					
	wBol224				YP_002727507.1	wSuzi959						
	wBol352											
YP_007889027.1	wBol294	ZP_08970967.1	wBol323	NP_966338.1	YP_002727237.1							
	wBol372											
YP_007889209.1	wBol85	ZP_08969987.1		NP_966452.1	YP_002727099.1	wSuzi509	ZP_01314298.1					YP_197989.1
											wBan1503	
YP_007888986.1		ZP_08970369.1		NP_966227.1	YP_002726916.1	wSuzi375		ZP_03788091.1			wBan1369	
			wBol690						wDim684	YP_006555990.1	wBan858	YP_198277.1
YP_007888917.1	wBol329	ZP_08970124.1	wBol358	NP_966094.1	YP_002727109.1	wSuzi518	ZP_01314890.1					
	wBol470	ZP_08970129.1		NP_966398.1	YP_002727221.1	wSuzi648	ZP_01314496.1					
					YP_002727503.1	wSuzi954	ZP_01314454.1					
		ZP_08970144.1	wBol399									
YP_007888882.1	wBol398										wBan369	
YP_007889169.1												
YP_007889378.1	wBol215	ZP_08970099.1		NP_966653.1	YP_002727392.1	wSuzi832	ZP_01314743.1	ZP_03788122.1	wDim269	YP_006556130.1	wBan1496	YP_198102.1
							ZP_01315028.1					
YP_007889457.1		ZP_08970536.1	wBol976	NP_966751.1	YP_002727480.1	wSuzi931	ZP_01314809.1	ZP_03788410.1	wDim4	YP_006555948.1	wBan522	YP_198057.1
											wBan525	
	wBol1229	ZP_08970895.1										YP_198127.1
YP_007888920.1												
YP_007889100.1	wBol1176	ZP_08970840.1		NP_966177.1		wSuzi470	ZP_01314706.1	ZP_03787956.1				
	wBol698										wBan1233	
YP_007888933.1	wBol697	ZP_08970376.1		NP_966281.1	YP_002726864.1	wSuzi318	ZP_01314872.1	ZP_03788384.1				
					YP_002726899.1	wSuzi358						
					YP_002726898.1	wSuzi357						
		ZP_08970043.1										YP_198395.1
YP_007889165.1												
	wBol389											
		ZP_08970890.1	wBol1225	NP_965900.1	YP_002726716.1	wSuzi59	ZP_01314372.1					YP_198495.1
		ZP_08970889.1										
YP_007889602.1	wBol894	ZP_08970614.1		NP_966925.1	YP_002727670.1	wSuzi1127	ZP_01314343.1					
		ZP_08971118.1										
YP_007889258.1	wBol156	ZP_08970042.1		NP_966510.1	YP_002727267.1	wSuzi695	ZP_01314612.1				wBan1063	YP_198393.1
	wBol786	ZP_08970455.1										
		ZP_08970454.1										
	wBol25											
				wVulM689								
YP_007888983.1	wBol687	ZP_08970366.1		NP_966230.1	YP_002726913.1	wSuzi372	ZP_01314241.1	ZP_03787883.1	wDim561	YP_006555716.1	wBan1175	YP_198412.1
YP_007889395.1	wBol814	ZP_08970488.1			YP_002727409.1	wSuzi850						
YP_007889110.1	wBol23	ZP_08970914.1			YP_002727003.1	wSuzi481	ZP_01314835.1	ZP_03787603.1	wDim385	YP_006556035.1	wBan352	YP_198126.1
	wBol391	wVulM1538										
	wBol396											
	wBol666	ZP_08970343.1										
		ZP_08970659.1										

II. Protein-protein interactome investigation

In parallel with the wide *in silico* screen of putative symbiosis-related bacterial effectors, experimental protein interactome analyses were performed on two systems, the nematode *B. malayi* and the isopod *A. vulgare*. This work is based on the hypothesis that the bacterium biochemically interacts with its host to maintain the symbiosis. Therefore protein-protein interactions were investigated in order to identify host partners of this interaction.

Potential *Wolbachia* genes of interest were selected based upon the criteria that they might be proteins that could interact with host proteins either because of their location (membrane proteins) or because these proteins are known to contain domains that potentially interact with other proteins. We selected the *Wolbachia* surface proteins (WSP) as well as proteins containing ankyrin domains (ANK), and the Wiskott-Aldrich Syndrome Protein, all potentially interesting protein-protein interacting partners (Dean 2011).

Considering that the *w*Bm strain has only five ANK whereas the *w*VulC strain has 86 ANK, we selected the 5 ANK from *w*Bm as well as their homologs in the isopod strain when present. We complemented this list of putative *Wolbachia* interactors with the WSP proteins of both strains and with the *w*Bm protein WASP as well as the *w*VulC ANK pk2b2 that has been shown to be expressed only in feminizing strains (Pichon et al. 2012).

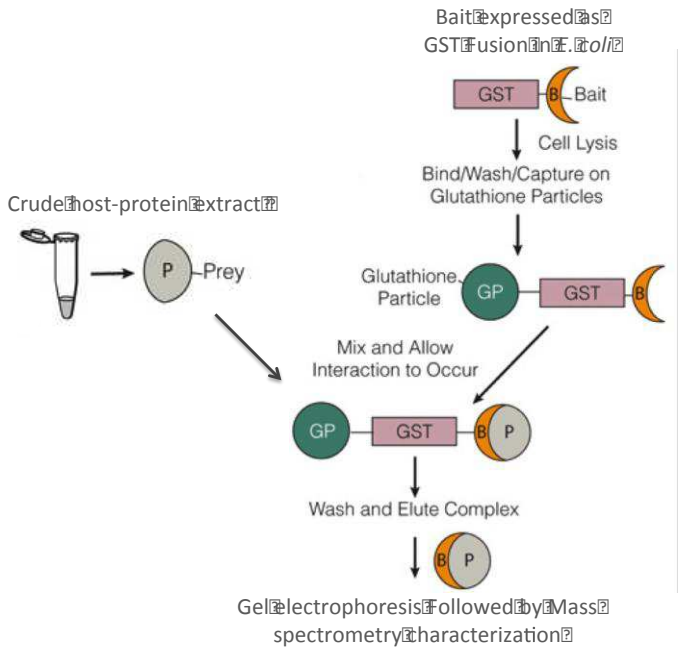


Figure 53: GST-pull down procedure. Modified from MagneGST™ Pull-Down system manual (Promega)

Protein-protein interactions were investigated using two main procedures. First, we used a GST-pull down method where recombinant *Wolbachia* bait proteins were expressed with a GST-tag, bound to glutathione beads to pull down host interactors from protein crude extracts. Bound proteins were then characterized by mass spectrometry (Figure 53). This first analysis was then complemented by a phage display procedure to identify peptides that specifically bound to the selected *Wolbachia* baits.

A. Cloning and expression of selected *Wolbachia* bait genes

In order to clone the selected *Wolbachia* genes, primers were designed according to protein hydrophilicity plots to optimize the expression yield. *Wolbachia* proteins and particularly membrane proteins are known to be hard to express (Bill Sullivan, personal communication). Expressed proteins tended to precipitate if expressed too quickly. Thus, PCR oligonucleotide primers for ANK were specifically designed around the ANK domains and one WSP protein that containing an important hydrophobic sequence in the middle was expressed as two separated proteins before and after this sequence (Figure 54).

A

	*w*Bm gene	*w*VulC gene
wsp	wBm0100	wVul_0072
	wBm0284	wVul_0359
	wBm0432	wVul_1434
WASP	wBm0076	NA
Ankyrins	wBm0287	NA
	wBm0296	wVul_1467
	wBm0394	NA
	wBm0447	wVul_0815
	wBm0582	wVul_0812
	NA	wVul_0887 (pk2b2)

B

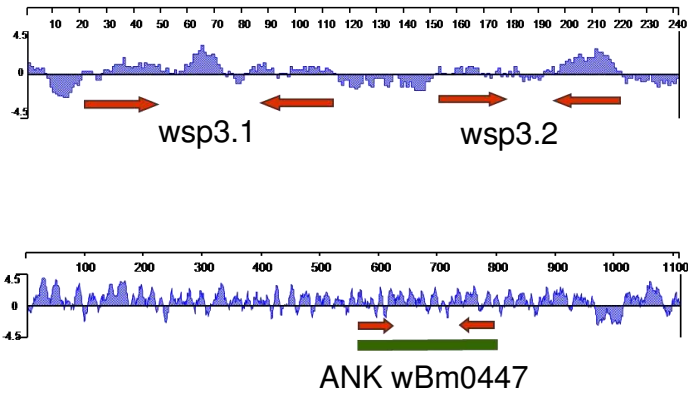


Figure 54: *Wolbachia* proteins used bait for GST-pull down and phage display experiments. A: List of *Wolbachia* proteins homologs in *w*Bm and *w*VulC. B: Example of primer design (in red) according to the hydrophylicity plot for the wsp wBm0432 (wsp3) and the ANK wBm0447.

Expression of recombinant material created by insertion of selected genomic DNA fragments into the pGEX-5X-1 vector, resulted in production of proteins fused at the N-terminal to GST.

The expression of recombinant proteins was only observed in bacteria after induction with IPTG and the solubility of the recombinant proteins was evaluated from the comparison of SDS-PAGE profiles of the bacterial crude extract (total protein fraction) and the supernatant, which only contains the soluble proteins (Figure 55).

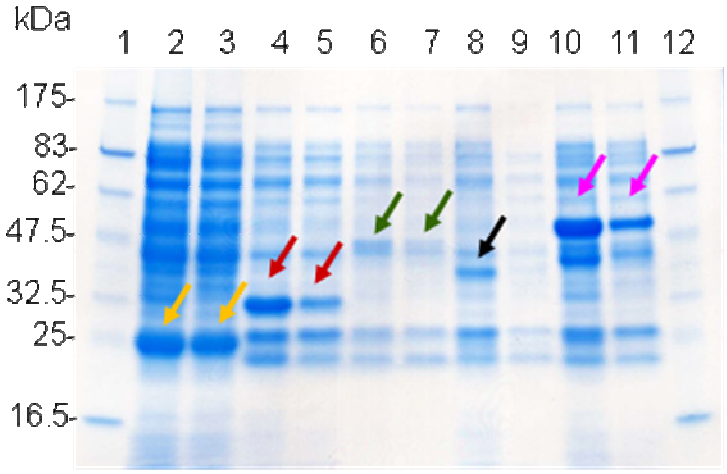


Figure 55: SimplyBlue stained SDS-PAGE gel. Lane 1: Prestained Protein Marker, Broad Range 6-175kDa (NEB). Lanes 2,4,6,8,10: lysates (total protein fractions). Lanes 3,5,7,9: supernatant (soluble fractions). Lanes 2 and 3: GST; 4 and 5: GST-wBm0284; 6 and 7: GST-wBm0296; 8 and 9: GST-wBm0394; 10 and 11: GST-wBm0447. Arrows indicate the expression of recombinant proteins.

GST-fused protein solubility was generally improved by trying several IPTG concentrations (1mM, 100µM, 50µM, 25 µM), temperatures (16°C, 30°C, 37°C) and length of expression (4hrs, 8hrs, 24hrs, 48hrs). The optimum expression condition found was 16°C for 48 hours after an induction with 25 µM IPTG.

Even with all these expression optimization steps, the ANK wBm0394, wBm0582 and wVulC0812, the WSP wVulC0359 and the WASP wBm0076 remained insoluble. Expression of these proteins were then assayed in an *E. coli* strain specifically designed for membrane protein expression (Lemo21(DE3), NEB, C2528H). This strain offers a tunable expression of difficult clones with a gradient of L-rhamnose, but even with these optimization steps, these proteins remained insoluble and were eliminated from the protein-protein interactome studies.

B. GST-pull down experiments

Bead binding buffer had to be optimized since the preliminary experiments we performed using PBST or binding buffer from commercial kit (MagneGST™ Protein Purification System, Promega) were not particularly specific and efficient; the binding yields were relatively low and analysis of bound proteins by SDS-PAGE electrophoresis attested of non-specific binding on beads. First, we attempted to improve the specificity of the binding to the glutathione beads by changing experimental conditions such as the length and temperature of the incubation with the *E. coli* extract containing the recombinant proteins (1 hr. at room temperature and overnight at 4°C). But the results were not acceptable: while the binding efficiency increased with an overnight incubation, nonspecific binding and protein degradation were also higher. In collaboration with Catharina Lindley from Bill Sullivan's lab at the University of Santa Cruz (CA), we developed more efficient buffers for bacteria grinding (50 mM Tris, pH 8, 100 mM NaCl, 2 mM EDTA, 1 mM chymostatin, 1 mM pepstatin, 1 mM leupeptin, 10 mM DTT, 7 mM lysozyme) and bead binding (20 mM Hepes (pH7), 150 mM KOAc, 2 mM MgOAc, 2 mM DTT (freshly made), 0.1% TetraFluoro Acid (TFA), protease inhibitor cocktail). Even if most of the proteomic protocols recommend the use of Tween 20 in binding buffer, we had to find a mass spectrometry-friendly detergent, the TFA. Experimental conditions for bait binding were then optimized to a 15-minute end-over-end rotation incubation with cold beads at 4°C to limit protein degradation. Using the ANK wBm0447, several concentrations of recombinant proteins were assayed and the optimal binding yield was obtained for 100 µL of beads, 200 µL of binding buffer and 50 µL of crude *E. coli* extract.

Incubations with host crude extracts were performed according to these optimization and elution samples were submitted to the NEB mass spectrometry lab for LC/MS/MS characterization.

Analysis of the mass spectrometry results with the Mascot program gave a large list of *B. malayi* proteins as shown in Table 18 for all the proteins but the WSP3.1, for which a single major protein was detected, the Disorganized Muscle Protein 1, putative. The function of this protein is still unknown but the homologous protein in the free-living nematode *Caenorhabditis elegans* is the DIM-1, an immunoglobulin superfamily protein necessary for maintaining body wall muscle integrity. This protein was also found in the list of the putative interactors of the second half of the WSP protein but not in any other bait interactor list, adding a particular interest to this protein. A validation of this result was attempted with a "reverse pull-down" expressing the *B. malayi* protein with the GST-tag, but even after the design of multiple primer sets, amplification of this gene

remained unsuccessful, maybe because of the draft status of the *B. malayi* genome, meaning that the sequences in the database might not have been 100% correct. Thus, the PCR primers may not have been correct for amplifying this gene. A full genome sequence is underway and this control will certainly be possible soon.

Table 18: Tables of the host proteins identified by GST-pull-down. GST is the negative control. Mass spectrometry data analyzed with Mascot (MatrixScience).

WSP wBm0100	WSP wBm0100
	Disorganized muscle protein 1, putative
Calponin homolog OV9M	Calponin homolog OV9M
Actin 1, putative	Actin, putative
	Calcyclin binding protein-like
	Calreticulin, putative
Galectin, putative	Galectin, putative
	Intermediate filament protein, putative
Myosin heavy chain, putative	Myosin heavy chain, putative
	Myosin regulatory light chain 1, putative
Myosin tail family protein	Myosin tail family protein
Sheath protein 5 (Sheath protein 5, identical	Sheath protein 5 (Sheath protein 5, identical
Tropomyosin family protein	Tropomyosin family protein
Tropomyosin, putative (Fragment)	Tropomyosin, putative (Fragment)
Troponin T, putative	Troponin T, putative
Troponin family protein	
Cytoplasmic intermediate filament protein, putative	Cytoplasmic intermediate filament protein, putative
	Tubulin beta-1 chain, putative
	40S ribosomal protein S13, putative
	40S ribosomal protein S6, putative
	60S ribosomal protein L12, putative
60S ribosomal protein L14, putative	60S ribosomal protein L14, putative
60S ribosomal protein L5, putative	60S ribosomal protein L7, putative
	60S ribosomal protein L7a, putative
Heat shock 70 kDa protein C, putative	Heat shock 70 kDa protein C, putative
Heat shock protein 90, putative	Heat shock protein 90, putative
Heterochromatin protein 1, putative	Heterochromatin protein 1, putative
Histone H2B	Histone H2B
Histone H4	Histone H4
	Hypothetical 86.9 kDa protein C30C11.4 in chromosome III, putative
	Ribosomal protein S7
Ribosomal protein S8	
Small heat shock protein	Small heat shock protein
Small heat shock protein OV25-1, putative	Small heat shock protein OV25-1, putative
Structure-specific recognition protein 1, putative (DNA binding)	Structure-specific recognition protein 1, putative (DNA binding)
Putative uncharacterized protein	
Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12)	Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12)
	14-3-3-like protein 2, putative
	Chain A, Structure Of A Brca2-Dss1 Complex., putative
	Alpha amylase, catalytic domain containing protein
	ATP synthase subunit alpha
	Mitochondria bc1 complex core subunit 1
	Endoplasmin, putative
	Major sperm protein
	Phosphoglycerate kinase (EC 2.7.2.3)
	Phosphorylase (EC 2.4.1.1)
	Spectrin alpha chain, putative
	KH domain containing protein

GST	WSP wBm0284	ANK wBm0287	ANK wBm0447
isopentenyl-diphosphate delta-isomerase, type 1 family protein	isopentenyl-diphosphate delta-isomerase, type 1 family protein	isopentenyl-diphosphate delta-isomerase, type 1 family protein	isopentenyl-diphosphate delta-isomerase, type 1 family protein
Beta-ketoacyl synthase, N-terminal domain containing protein	actin-1	Beta-ketoacyl synthase, N-terminal domain containing protein	Fibronectin type III domain containing protein
hypothetical protein Bm1_01365	EGF-like domain containing protein		
hypothetical protein Bm1_16950			

For all the other baits, mass spectrometry characterization of the eluates revealed a large number of host proteins suggesting non-specific binding to *Wolbachia* protein baits. However, it is notable that many of these host proteins are muscle or gene expression-related proteins.

While this GST-pull down procedure may lead to the identification of interacting proteins, it is a method that requires optimization at every step of the protocol and probably for every potentially studied bait. Considering that these optimizations are very time-consuming for uncertain results, we decided to explore a new procedure to filter out the real hit from contaminants and to validate these results.

C. Phage display experiments

Phage display experiments were chosen to continue with the interactome experiments as this method had already given concrete results during a project targeting ligands of *Wolbachia* independent phosphoglycerate mutase (S. Raverdy, PhD thesis).

In the case of *A. vulgare,* the interactome studies were limited to phage display analyses since the host genome sequence is not yet available.

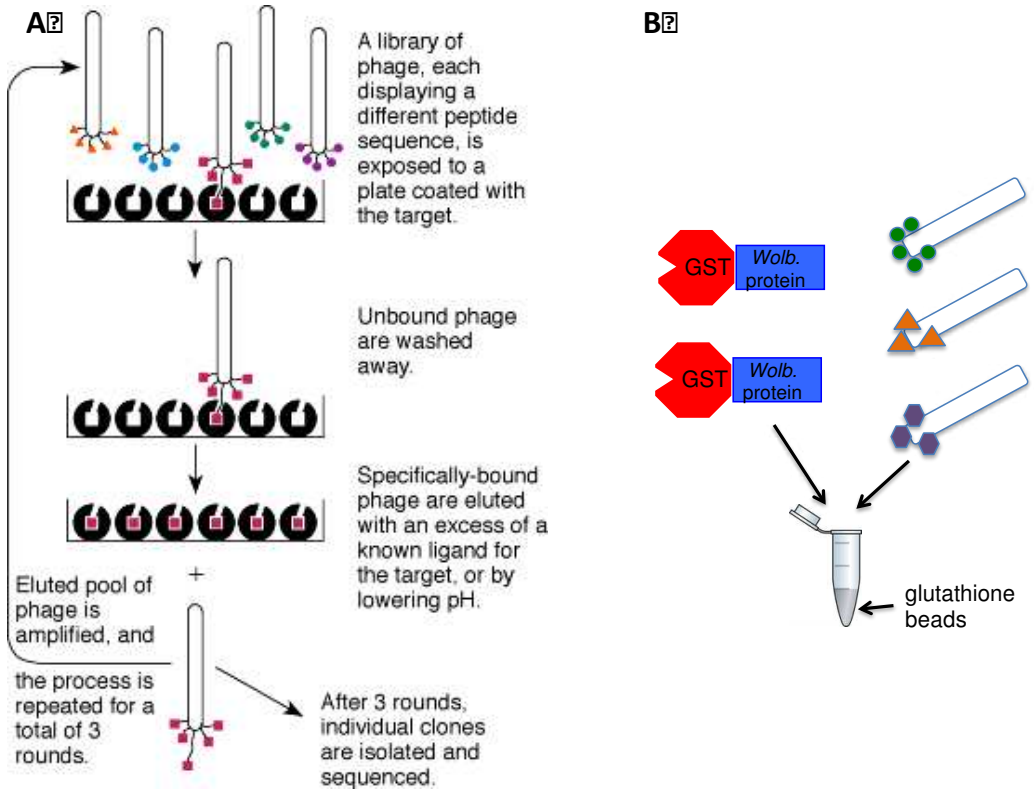


Figure 56: Phage display procedure. A: Theoretical immobilized phage display procedure. B: Used in-solution procedure.

We used two commercial M13 phage display libraries (NEB), displaying randomized 7-mer or 12-mer peptides, to search for peptide sequences which specifically bound to the selected *Wolbachia* baits (Figure 56). Recombinant GST-tagged proteins were used for

these phage display experiments. In order to eliminate non-specific bindings, three rounds of phage selection were performed and a negative selection step was added to the last two rounds by first applying phage pools to the GST protein alone in order to eliminate all peptides that bind to the GST-tag.

Eighteen randomly picked plaques of phage clones from each third elution were sequenced (Table 19). Some of the sequenced phages had no insert encoding a peptide. These are wild-type phages that have an amplification yield higher than phages presenting a peptide because the display of foreign peptides as N-terminal fusions to the infectivity protein pIII slightly attenuates infectivity of the library phage relative to wild-type M13. This is a known phenomenon using this technique, caused by either an environmental contamination of the phage library or a slight remainder of original wild-type phages used for the library construction.

Table 19: Selection of peptides that specifically bind to various *Wolbachia* baits. Eighteen clones were sequenced for each bait, less sequences indicates that wild type clones (with no integration of a peptide) were selected. In Yellow are the peptides found in the control with GST alone, showing non-specific binding. In color are peptides found with more than one bait.

wsp1		ANK447		wsp3		ANK	
wBm	wVulC	wBm	wVulC	wBm	wVulC	wVulC	
wBm0100	ORF72	wBm0447	ANK54	wBm0432	ORF29	pk2b2	GST
GTSFARA	LPGRAHDPWKVP	DLTFTVNPLSKA	STTKLAL	CPHNNNY	QKYELVL	DLMSYMG	AAQTSTP
GXXPSXA	VPRSMAATHSTF	EPLQLKM	STTKLAL	YPPLSTY	YETLWETIPTSG	FTARCQACWGPP	GTSFARA
HNTEVSRYRADH		EYSPKLFPPHRL	GNTPSRA	WTITKHP	YETLWETIPTSG	FTVFPPN	IPPFMRDHIALG
HNTEVSRYRADH		FGADYWENNTML	GNXPSRA	VQPLHKT	YETLWETIPTSG	GKPMPPM	KLPGWSG
HNTEVSRYRADH		GLLFPPESSYKG	TCAKATSTPPLS	SPHLHGA	YETLWETIPTSG	GNTPSRA	LNPHLYA
HNTEVSRYRADH		GWLRSPSLLFSN	XCAKATSTPPLS	SHTLSAK	YETLWETIPTSG	GNTPSRA	NGSLNTHLAPIL
HNTEVSRYRADH		IDRVTSRDPAMN	TCAKATSTPPLS	SHSLLHH	YETLWETIPTSG	GPSALYPNMTQN	NGSLNTHLAPIL
HNTEXSRXRXDH		IPPFMRDHIALG		NGSLNTHLAPIL	YETLWETIPTSG	KPVQLDH	QISFMAN
HXTXXXXXRXXH		LPLTPLP		LPLLPKE	YETLWETIPTSG	LPLTPLP	QLXPGTPAAERL
NGSLNTHLAPIL		MQDHSQIPLLTH		IPTLPSS	YETLWETIPTSG	LPSQTAP	QLYPGTPAAERL
NGSLNTHLAPIL		NGSLNTHLAPIL		IPPFMRDHIALG	YETLWETIPTSG	QAHTIST	SFVLPYY
NGSLNTHLAPIL		QIQNHMILPHIR		ILATPNV	YETLWETIPTSG	SSLHSSHHEMNK	SILPYPY
NGSLNTHLAPIL		VQHNTKYSVVIR		HSSLQTP	YETLWETIPTSG	STPAPGM	SILPYPY
NGSLNTHLAPIL		SDMWENHMIXXL		FLSXAPS	YETLWETIPTSG	TSDLKLIAALRS	SLNHSVF
NQDVPLF				DVTHISRLHHHL	YETLWETIPTSG	WTITKHP	STGSPPP
TNDTWAVTQNS				DLTFTVNPLSKA	YETLWETIPTSG	WTITKHP	STPAPGM
				DGTMVRS	YETLWETIPTSG	YEEASRL	TPPSFLH
					YETLWETIPTSG	YRAPWPP	

An additional control was performed on the GST protein alone to sequence the peptides that specifically bind to the tag instead of the bait. Out of the 18 sequenced clones, 14 different peptides were identified and 4 of them were also present in some bait peptide interacting lists. These 4 peptides were identified as non-specific interactors and

removed from the putative interactors list. Some peptides are also found in different bait lists, these are either non-specific interactors or host peptide motifs that bind to *Wolbachia* proteins. It is also known that peptides containing serine are more likely to be false positives whereas peptides containing cysteine should be strong positives as these peptides appear to be hard to select (C. Noren, personal communication).

Interactions between these peptides and the bait proteins were assayed by ELISA assay. An example of an ELISA measurement is shown in Figure 57. While a negative control of the ELISA assay was performed in presence of the bait protein but without a phage, the specificity of the phage-binding was assayed with BSA. None of the peptides tested showed a non-specific profile where the fluorescence is similar between the well with BSA and the well with the bait protein.

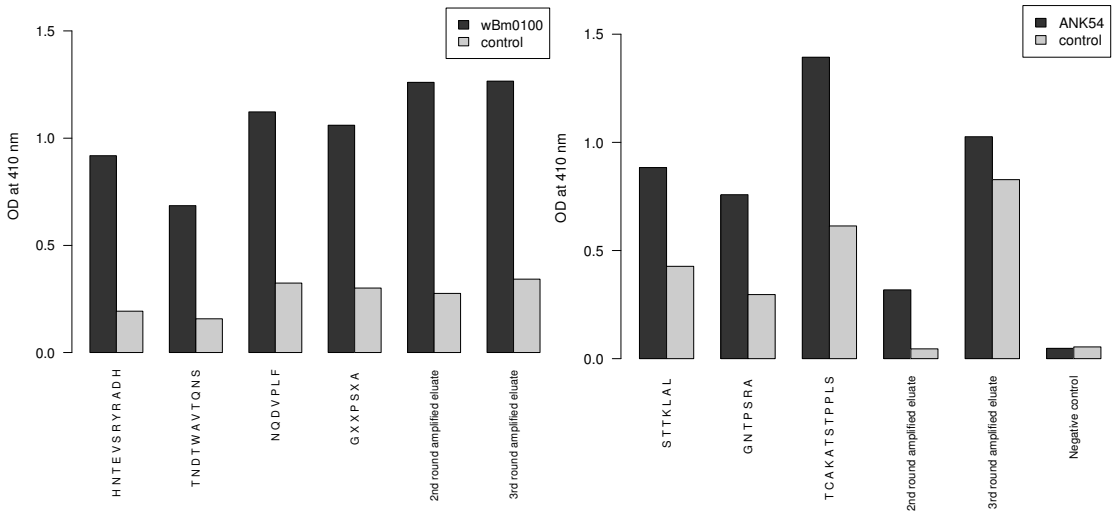


Figure 57: Profile of ELISA measurements on eluted phage clones which bind specific fusion protein after 3 rounds of phage display using PhD7 and PhD12. Blocking buffer with BSA was used as a control to detect unspecific binding. Negative control was performed without phage. Bound phages were detected by incubating with an anti-M13 phage antibody conjugated to HRP and measurements carried out at 410 nm.

Phage display experiments allowed the identification of peptides that specifically bound to the tested recombinant proteins tested. These interactions have to be confirmed by additional methods to remove experimental bias in the protein-protein interactions.

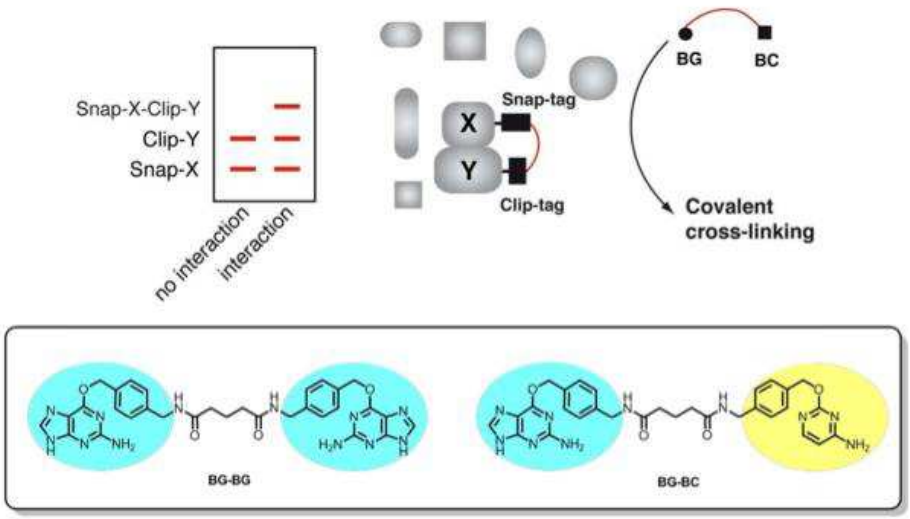


Figure 58: Cross-linking procedure (I. Correa, personal communication).

Protein-protein interactors will be assayed using cross-linking experiments (Figure 58). Each partner will be expressed with either a SNAP-tag or a CLIP-tag that can be added at both ends of the proteins to avoid miss-conformation of the protein due to the tag. This method allows covalent cross-linking between two interacting partners, which can be visualized by SDS-PAGE electrophoresis. .

While still in the preliminary stages, it is interesting to note that both techniques used allow the potential identification of host proteins that interact with selected *Wolbachia* proteins. We will take advantage of these first experiments to test mutualistic and feminizing-specific proteins previously identified in order to identify host proteins that may interact with these factors potentially involved in *Wolbachia*-induced phenotypes.

III. Investigation of putative bacterial virulence factors

Virulence factors are defined as bacterial particles, mainly gene products, which enable the bacteria to establish themselves inside (or on) a host of particular interest and enhance the pathogenesis of the bacteria. These factors are known to be bacterial toxins, cell surface proteins that mediate bacterial attachment or enzymes such as metalloproteases, phosphatases that contribute to the pathogenicity of the bacterium. Many bacterial pathogens that infect eukaryotic cells present a highly conserved set of protein secretion systems that are involved in the transport of bacterial effector proteins into host cell cytosol. These bacterial effectors are able to modulate some host cell functions to serve the bacteria for survival or proliferation (Hicks and Galan 2013).

As many bacteria use their secretion systems to deliver virulence factors to host cells, they are key bacterial weapons and attractive targets for the investigation of virulence mechanisms.
Targeting these secretion systems is of particular interest since they may play a fundamental role in the molecular mechanisms of the host-*Wolbachia* symbiosis e.g. to transport *Wolbachia* feminizing effector(s) to the isopod *A. vulgare.* Furthermore, a better understanding of these systems could lead to the identification of new anti-*B. malayi* drug targets since they have been identified as potential targets for anti-bacterial treatments (Baron and Coombes 2007).

Therefore, we investigated the conservation of secretion systems in all the studied *Wolbachia* genomes and performed transcriptional analyses of both secretion system proteins and selected putative virulence factors.

A. Identification of putative virulence factors pathways

Gram-negative bacteria have evolved up to seven different secretion systems that are classified depending on two secretion mechanisms. In one-step secretion systems (Types I, II, III, and IV secretion systems), substrates are directly translocated from the bacterial cytoplasm to the extracellular medium or into the eukaryotic target cell while in the two-steps secretion systems (Type V and VI secretion systems, Sec an Tat system) substrates are first translocated across the bacterial inner membrane to the periplasm and then translocated through the outer membrane to the exterior (Durand et al. 2009).

The most studied bacterial secretion systems are the Type III secretion system and the Type IV secretion system (T4SS) that are present in a wide range of bacterial symbiont of mammals, plants or insects (Cascales and Christie 2003). However, in *Wolbachia*, the Type III secretion system is absent and the secretion system the most studied so far is the T4SS, which presents functional and highly conserved genes (Pichon et al. 2009).

In all the *Wolbachia* strains sequenced in this study, all the secretion system genes known from published *Wolbachia* genomes were retrieved by reciprocal BLAST and an additional putative virulence pathway was identified: two components of the tol/pal system were annotated (Figure 59). We identified 18 T4SS genes in the sequenced genomes including the well-conserved operons *virB3-virB8* and *virD4-virB8* as well as 10 genes scattered across the genomes. In contrast to other *Wolbachia* genomes, all strains from isopods have 4 copies of *virB2*, which corresponds to a pilus-associated

subunit involved in host-cell attachment (Cascales and Christie 2003; Backert and Meyer 2006). In other *Wolbachia* genomes, this subunit is present in a variable copy number: from 0 for *w*Bm to 6 in *w*Mel. A complete type I secretion system was also present in all strains and only two of the twelve essential genes of the type II secretion system were identified. Furthermore, two proteins of the tol/pal system were identified for the first time; both TolB and PAL are periplasmic proteins. The lipoprotein PAL is involved in outer membrane integrity (Cascales et al. 2002) and uptake of nutrients across the membrane (Llamas et al. 2003). Translocation across the inner membrane occurs through the secretion (Sec) or twin-arginine translocation (Tat) pathways may constitute a route for bacterial effectors since both systems are conserved and complete.

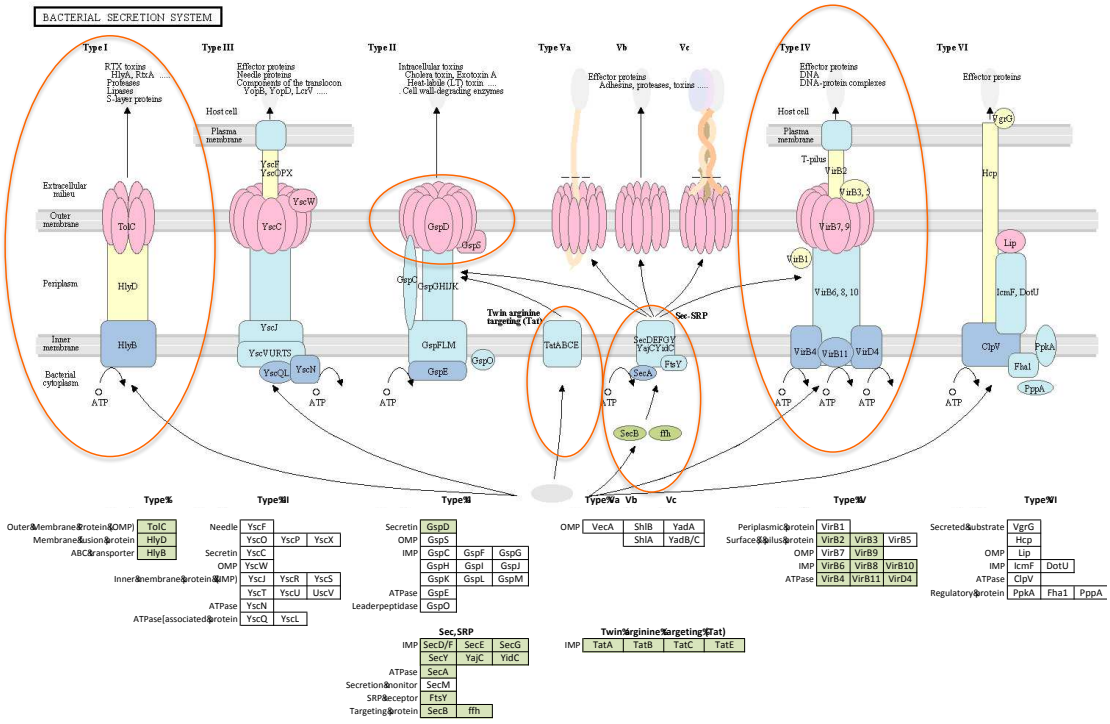


Figure 59: Bacterial secretion systems in *Wolbachia* from isopods. Modified from the KEEG Pathway. Genes present in the sequenced genomes are in green. Red circles represent secretion system identified in sequenced *Wolbachia* strains.

We further decided to follow the expression of two of these secretion systems, the T1SS and the T4SS in addition to putative bacterial effectors throughout *A. vulgare* and *B. malayi* development to investigate the role of these systems in *Wolbachia* virulence.

B. Transcriptomic approach of *Wolbachia* symbiosis in the isopod *A. vulgare*

In *A. vulgare*, the *w*VulC *Wolbachia* strain plays a particular role during post-embryonic development; the presence of the bacteria induces the feminization of genetic males into functional females. Since the relative abundance of the bacteria throughout host development was still unknown, we first quantified the bacterial load for each developmental stage in order to highlight a potential change in bacterial load for a specific stage. We then investigated secretion system and virulence factor gene expression during post-embryonic development of *A. vulgare.*

1 - Bacterial load during host post-embryonic development

Before studying the relative expression of putative effectors, we quantified the number of *Wolbachia* genomes per host cell using the ratio of two single copy genes: the *wsp* gene for the number of bacterial genomes and the androgenic hormone (*AH*) gene to quantify the number of host cells. The quantification was performed by quantitative PCR on DNA samples extracted from a pool a 20 individuals for stages 0, 1 and 2, a pool of 2 individuals for stage 3 and individual females embryos for stages 4 to 8. In this laboratory lineage, all individuals are homogametic ZZ males, only the presence of *Wolbachia* is responsible for female development. The experiment was performed with 2 technical replicates on 15 biological replicates per individual embryo samples and 4 biological replicates for pooled samples. Copy numbers were calculated using a standard curve performed on serial dilutions of regular PCR products of the *wsp* and *AH* genes from adult ovaries.

The bacterial load increased during the post-embryonic development (Figure 60), before reaching a constant load from stage 4 onwards. Bacterial load was significantly different between stage 0 and all stages after stage 5 (p. value <0.05). Other significant differences in bacterial load were found between stages 4 and 6, between stages 2 and 6 and stages 2 and 7 (Kruskal test, p. value < 0.05). A slight decrease in the number of bacteria per host cell was noticeable for stage 8. As these analyses were performed on DNA extracts from entire newborns, this decrease is either due to a diminution in bacterial numbers per infected cells or to a multiplication of host cells that are not or only weakly infected such as muscle tissues.

Microscopic observations such as fluorescence *in situ* hybridization will determine if the tissue distribution of the *Wolbachia* is identical with that found in adults. Actually, it has

been shown in other *Wolbachia* hosts, such as the filarial nematode, that the *Wolbachia* localization during host development is not consistent with the distribution in adults .

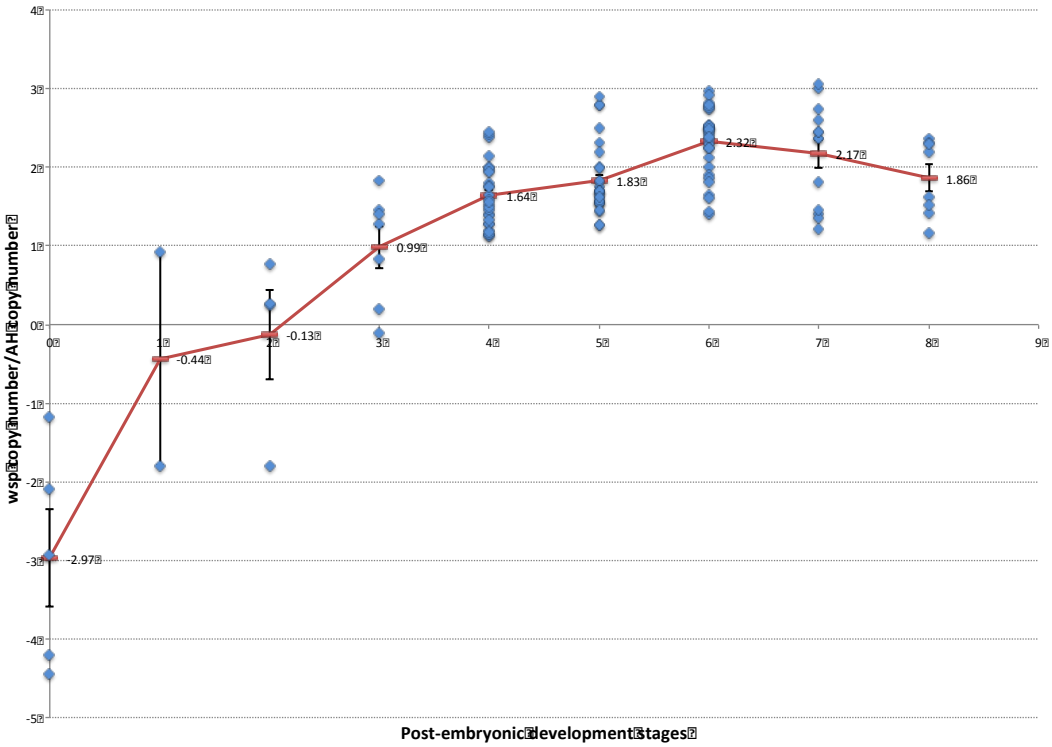


Figure 60: Quantification of *Wolbachia* load in *A. vulgare* during post-embryonic stages, from birth (st0) to fully developed isopod (st8). Data are shown in a 2-log base. Error bars represent standard error.

2 - Bacterial gene expression during host post-embryonic development

The expression of five *Wolbachia* genes was quantified using RT-qPCR on newborns throughout their post-embryonic development. The two operons of the T4SS expression were assayed using genes present in each operon; the expression of the gene *virB3* revealed the expression of the operon *virB3-virB6* and the *virB8* gene to the expression of the operon *virB8-virD4*. Similarly, the expression profile of the gene *TolC* corresponded to the expression profile of the T1SS. Additionally, we studied the expression profiles of two genes potentially involved in feminization: the ankyrin *pk2b2*, which is only expressed in feminizing strains (Pichon et al. 2012) and the gene X previously identified as only present in feminizing strains (II.C.3 - p113).

Expression levels were normalized against the *wsp* gene (wVulC_1434), which has been assayed for presenting a basal expression level. Indeed, gene expression quantifications

need to be normalized by a reference gene, whose expression is not dependent on the studied conditions, to correct for eventual technical differences between samples such as different RT reaction efficiencies or biological differences such as infection rates. To determine an efficient reference gene for these transcriptomic analyses, three genes were assayed: the *wsp* gene, which has already been used in RT-qPCR experiments for the quantification of *Wolbachia* ankyrin expression in adult *Drosophila* adult gonads and carcasses (Papafotiou et al. 2011), the *16S* rRNA gene, which is commonly used as reference gene in RT-qPCR experiments, and the beta-lactamase *hcpA* gene, which is a known housekeeping gene used in MLST studies (Baldo et al. 2006).

The *16S rRNA* gene was rapidly discarded from the analysis because of non-specific amplifications. In *A. vulgare*, amplification of the *16S rRNA* gene is often troubled by co-amplification of the host's *18S rRNA* gene (J. Dittmer, personal communication) and since we had to use a combination of one *Wolbachia*-specific primer with one universal bacterial primer in order to obtain an amplicon short enough to be compatible with qPCR, non-specific amplification could also be due to the amplification of other bacteria *16S rRNA* gene.

The comparison of expression profiles of the *wsp* and *hcpA* genes were performed on RNA samples from each *A. vulgare* post-embryonic stage (Figure 61). With a linear regression determination coefficient of 98% for the two genes, we showed that both genes could be used as reference genes. Considering that *hcpA* presents a lower expression level, we decided to use the *wsp* gene as reference to obtain a more efficient reference.

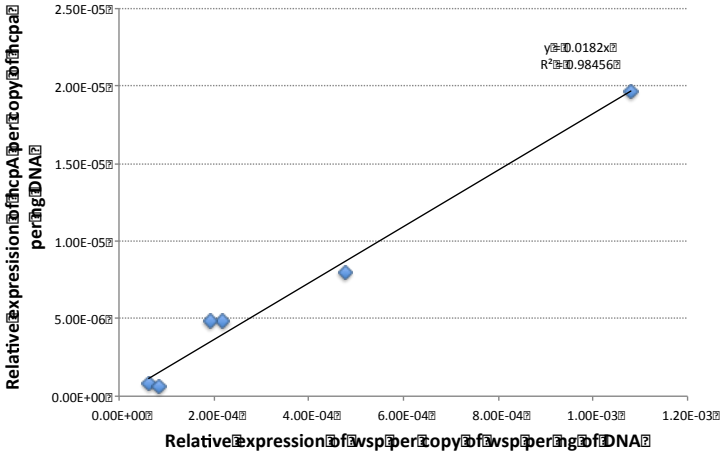


Figure 61: Comparison of expression profiles between two potential reference genes for RT-qPCR: *hcpA* and *wsp*.

Using the *wsp* gene as reference, the experiment was performed with 2 technical replicates on 4 biological replicates for stages 0 and 2, on 2 biological replicates for stages 1 and 3 and on 10 biological replicates for stages 4 to 8. No amplification was recorded when reverse transcriptase was omitted, indicating the signal obtained from cDNA samples was only derived from RNA.

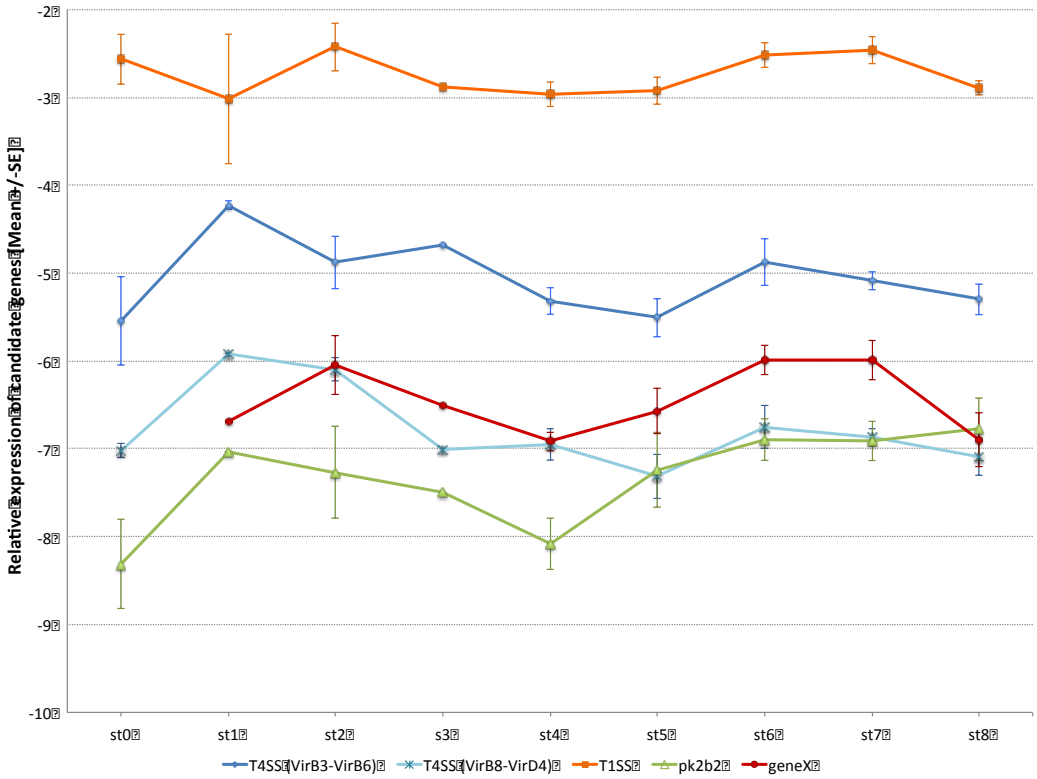


Figure 62: Relative gene expression during *A. vulgare* post-embryonic development. Candidate gene expressions are normalized by the expression of the *wsp* gene taken as reference.

For all tested genes, the relative expression is negative, meaning that these genes are under-expressed compared to the *wsp* gene (Figure 62). Since the gene *wsp* encodes a constitutive membrane protein, a higher *wsp* expression level has been expected beforehand. On the other hand, since the two operons *virB3-virB6* and *virB8-virD4* code for proteins constituting the same secretion system (T4SS), it is surprising to note that both operons have the same expression profile but do not have the same expression level throughout the post-embryonic development, the *virB3-virB6* operon is 30% more expressed than the *virB8-virD4* operon. This first quantitative estimation of the different constituents of the T4SS apparatus has to be extended on all other T4SS genes that are scattered across the genome since they may not all have the same transcriptional rate.

From stage 2 to 7, the two putative feminizing genes, *pk2b2* and gene X present the same expression profile with a decrease of expression up to stage 4 and an increase of

expression from stage 5 to stage 6 followed by the stabilization of the expression level until stage 7. This expression profile is also observed for the *T1SS* gene, which is in average 2.3-fold up-regulated compared to gene X.

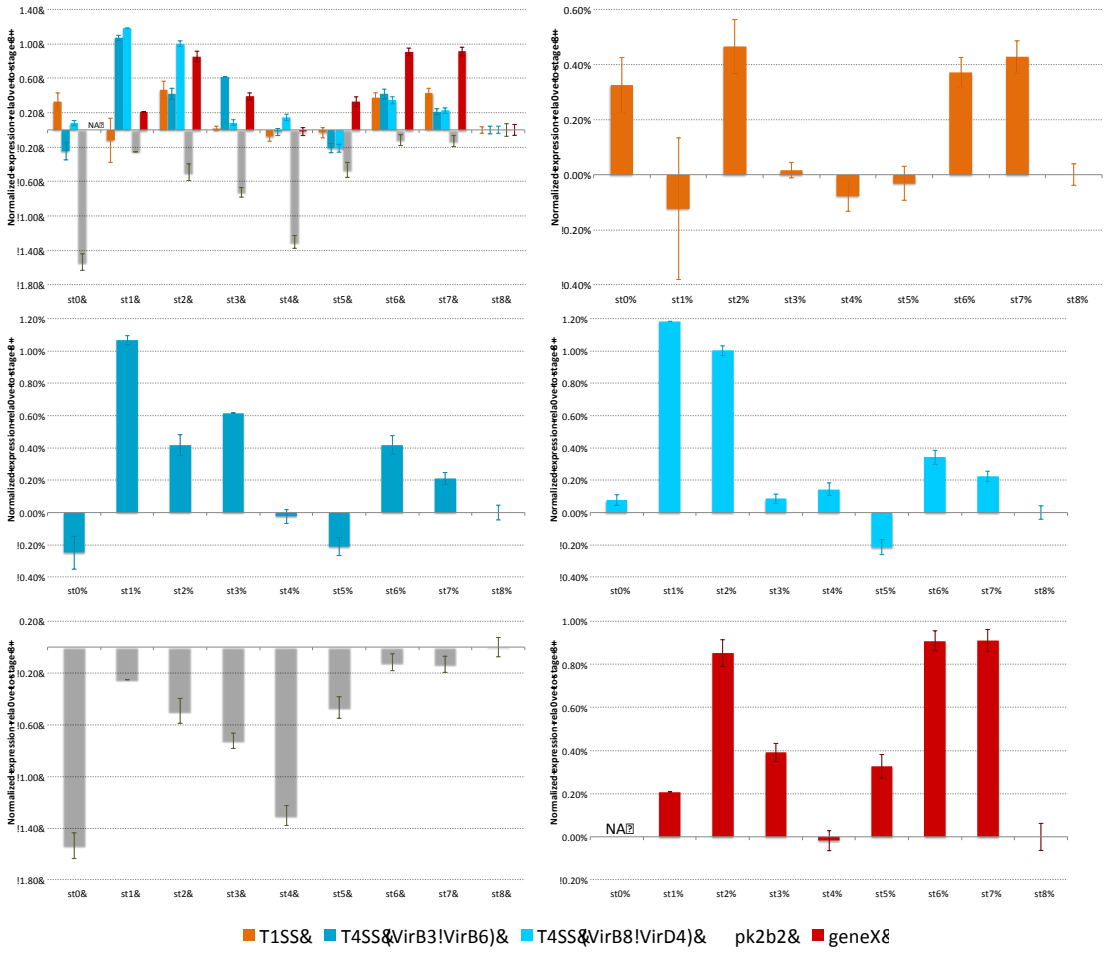


Figure 63: Normalized expression of candidate genes relative to stage 8. Data are shown in a 2-log base: a ratio of 0 means the targeted gene shows the same relative expression as stage 8; a ratio of 1 means a factor 2. Error bars represent standard errors.

In order to compare gene expression levels between post-embryonic developmental stages, relative expressions were normalized to stage 8 (Figure 63).

As previously observed, both putative feminizing genes *pk2b2* and gene X show the same global expression profile throughout post-embryonic development with a lower relative expression at stage 4, but *pk2b2* is the most expressed at stage 8 whereas gene X has the lowest expression rate at this stage.

Statistical analysis will confirm the significance of the observed variations.

C. Transcriptomic approach of *Wolbachia* symbiosis in the nematode *B. malayi*

As the difference with the isopod *A. vulgare*, the quantification of the bacterial load in the nematode *B. malayi* is well documented (Fenn and Blaxter 2004b; McGarry et al. 2004). The lowest number of *Wolbachia* were quantified during MF, L2 and L3 stages whereas a 600-fold increase was observed 7 days post mammalian infection till the complete development of L4 (adult), meaning that the major bacterial population growth occurs during the four weeks after definite host infection. Both studies came to the conclusion that *Wolbachia* is essential for the nematode development and survival in the mammalian host.

The transcriptomic analysis of *Wolbachia* symbionts of the nematode *B. malayi* was performed on 10 genes, the 5 ankyrin genes present in the *w*Bm genome (*ank287*, *ank296*, *ank394*, *ank447* and *ank582*), the three wsp genes (*wsp100*, *wsp284*, *wsp432*) and the two genes previously identified as only present in mutualistic strains (*wBm0047* and *WASP*) (II.A p109). As explained previously, two genes were assayed as reference gene, the *16S rRNA* gene with the primers previously used in other studies (Li et al. 2011; Li and Carlow 2012), and the homologous gene of the *wsp* gene used for *A. vulgare* studies, the *wsp432* gene. We decided to use the *wsp* gene as reference because negative controls using the 16S rRNA genes performed on non-symbiotic *A. vulgare* DNA showed an amplification product, suggesting that these primers may not be as specific as the *wsp* primers.

Using *wsp* as reference gene, the experiment was performed with 2 technical replicates on total *B. malayi* RNA supplied by the Filariasis Research Resource Center (FR3) throughout the host life cycle (II.A.1 - p48; 5 stages: microfilariae – MF, third- and fourth-stage larvae – L3 and L4; adult female – AF and adult male – AM).

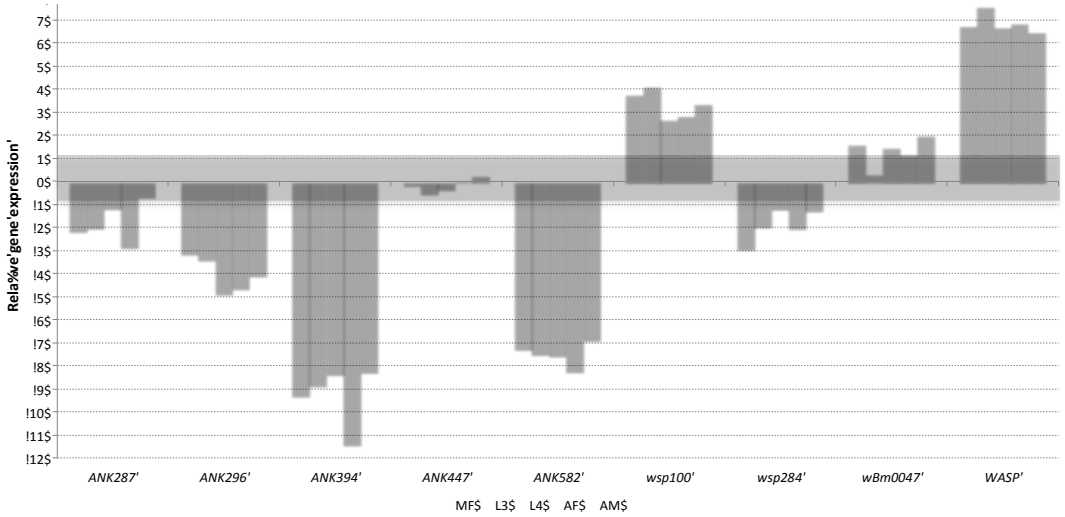


Figure 64: Relative expression of *w*Bm genes to the *w*Bm *wsp432* gene. No error bar is indicated since the experiment was performed on pooled RNA with 2 technical replicates. Data are shown in a 2-log base: a ratio of 0 means the targeted gene shows the same expression level as the reference gene, a ratio of 1 means a factor 2. The grey shape indicates non-significant variation.

All 10 studied genes are transcribed in all examined stages suggesting that all of them are functional (Figure 64). No amplification was recorded when reverse transcriptase was omitted, indicating the signal obtained from cDNA samples was only derived from RNA.

Except for the *ank447* gene, which shows the same expression level as the *wsp432* gene, all ankyrin proteins have a lower expression level. *ank394* is the gene with the lowest expression rate with a 2700-fold down expression in adult females and a 300-fold down expression in adult male and L4s. However, the *wsp100* gene is on average 10-fold more expressed than *wsp432,* while the *wsp284* is 4-fold less expressed. Concerning the two genes only found in mutualistic strains, the wBm0047 is on average 2.7-fold more expressed than the reference gene while the *WASP* gene has the highest expression rate of the studied genes, on average its expression is 120 times higher than the expression of *wsp432*.

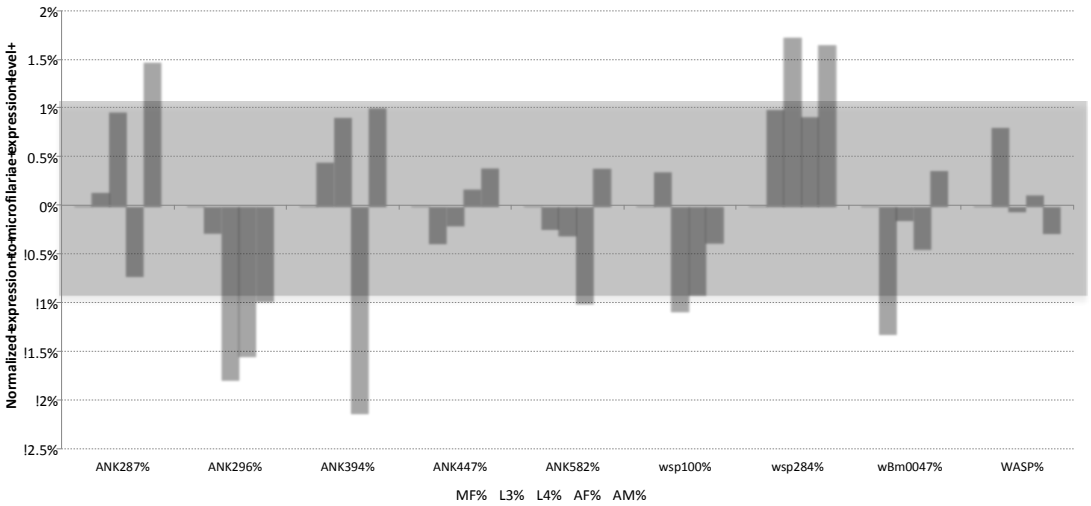


Figure 65: Normalized expression of candidate genes relative to the microfilariae stage. Data are shown in a 2-log base: a ratio of 0 means the targeted gene shows the same expression level as the reference gene, a ratio of 1 means a factor 2. The grey shape indicates non-significant variation.

With the expression normalized by the microfilariae expression rate (Figure 65), it appears that there is no significant variation of expression throughout the host life cycle for the genes *ank447*, *ank582*, *wsp100*, *WASP* and *16S rRNA*. The adult males show a significantly increased expression of the *ank287* and the *wsp284* genes whereas the adult females show a significantly decreased expression of the *ank296* and the *ank394* genes. During the parasite development, there is a down-regulation of the gene *wBm0047* at stage L3 but at stage L4, the *ank296* expression is down-regulated whereas the *wsp284* expression is up-regulated. Finally, no clear expression pattern appears clearly between several genes.

Discussion

The investigation of the molecular mechanism of the *Wolbachia*-host symbiont led us to three complementary approaches performed in parallel to identify putative bacterial effectors: an *in silico* screening, a proteomic approach and a transcriptomic analysis.

The *in silico* screening of bacterial genomes for identification of putative bacterial effectors was focused on the eukaryote-like motif proteins, which are known to be involved in host-symbiont interactions. Even though there are dozens of different eukaryote-like proteins in *Wolbachia* genomes, recent efforts were concentrated on ankyrins. Our study confirmed the exceptionally large number of ANK proteins in *Wolbachia* genomes and particularly in isopod *Wolbachia* genomes, which contain even more ANK than the insect strains, suggesting an adaptation to the host. A similar approach was performed on the tetratricopeptide-containing proteins and a different pattern was observed; most of the *Wolbachia* genomes have a similar number of TPR proteins but the two very closely related *w*Ri and *w*Suzi strains. Since both ANK and TPR proteins are involved in a wide variety of functions, the comparison of these proteins between different strains need an even more detailed analysis considering these diversities of functions. However, this preliminary result illustrated the importance to not underestimate the implication of other eukaryote-like proteins such as TPR containing proteins.

Therefore, we selected some candidates for protein-protein investigations, based on previous finding of their implication in host-symbiont interactions in other bacteria. However, *Wolbachia* proteins are difficult to express *in vitro* and multiple interactome studies performed using different proteomic techniques came to the same conclusion that *Wolbachia* interactome investigation is a long and bumpy road. Actually, both techniques used are based on time-consuming experiments, which need a lot of optimizations to finally give un-correlated results. However, these experiments should be soon aided by the quick development of next-generation proteomics that are targeting whole protein-protein interaction network identification (Altelaar et al, 2013, nature review). Even though high throughput interactome identification is still an actual challenge, the incredibly fast development of genomic NGS lets hope that this technology will be soon available.

If the molecular mechanism of the symbiosis might involve membrane protein for bacterial docking, secretion systems play an important role in the translocation of bacterial effectors through the bacterial membrane to host cytoplasm. Therefore, we screened the newly sequenced *Wolbachia* genomes to identify the genes of the different

secretion systems known in *Wolbachia* and additionally, identify two new component of the tol/pal system.

Considering that feminizing *Wolbachia* strains should interfere with host development before sexual differentiation, which occurs at the developmental stage 4, functional analyses were performed throughout the host *A. vulgare* post-embryonic development. After quantification of the evolution of the *Wolbachia* load during host development, functional analyses by transcriptomic approach were performed on the *w*VulC strain, on both secretion systems (T1SS and T4SS) and putative effectors: the pk2b2, which is only expressed in feminizing strains (Pichon et al. 2009) and the gene X, previously identified by *in silico* analysis as only present in feminizing strain genomes.

The bacterial load quantification reveals an increase of the *Wolbachia* density during infant development until stage 4-5 after what the bacterial density remains constant. It will be interesting to compare this result with the adult bacterial load to see if the adult bacterial load is reached. However, this comparison is technically challenging since DNA extraction on whole adult individuals is difficult due to the strong cuticle. Additionally, it has been recently observed that the bacterial load in adult is also variable dependently to the host reproductive cycle; *Wolbachia* density increases during the maturation of the oocytes (Genty et al. 2013). That being said, we demonstrated that there is not bacterial over-multiplication at a specific developmental stage, leading to the hypothesis of a differential expression of the bacterial effectors inducing the feminization of males into functional females.

In order to perform relative quantification of bacterial genes, several housekeeping genes were assayed to find an acceptable reference gene for both *w*Bm and *w*VulC genes expression. We demonstrated that the *wsp* gene should be used as reference, since the 16S rRNA gene amplification is not specific.

Demonstrating that RT-qPCR experiment is technically doable on a single *A. vulgare* infant from stage 4, we observed that none of the tested genes showed a significantly differential expression in a particular stage. This result is correlated with the observations of a continual action of the feminizing *Wolbachia* during the butterfly *E. hecabe* embryonic development (Narita et al. 2007).

However, global expression profiles were highlighted. Both putative feminization effector genes (pk2b2 and geneX) showed the same expression profile throughout the post-embryonic development than the T1SS gene allowing us to speculate on the implication of this secretion system in the translocation of feminization effectors.

Similarly, expression analysis of the T4SS apparatus through the two operons reveals an intriguing result: if they are both expressed similarly throughout all the post-embryonic

development stages, their level of expression differs. Before this study, co-transcription of both operons has been demonstrated by RT-PCR (Felix et al. 2008; Rances et al. 2008) but never quantified. However, a recent analysis of the regulation of the T4SS on *w*Bm identified two transcription factors (wBmxR1 and wBmxR2) that only regulate the *virB8*-*VirD4* operon suggesting the presence of additional transcription factors involved in the regulation of the *virB3*-*virB6* operon and the scattered *virB8* gene (wBm0641) (Li and Carlow 2012). Thus, if both regulatory systems are synchronal (both operons have the same expression profile), they may not have the same efficiency.

Identically, for all the *w*Bm gene candidates assayed, no clear difference of expression level was noticed between the different stages of the *B. malayi* life cycle. Interestingly, the two genes previously identified as only present in the mutualistic strains (wBm0047 and the WASP wBm0076), showed a higher expression than the reference gene wsp (wBm0432). The incredible high expression level of the WASP gene confirms that this gene may play an important role in the mutualistic symbiosis mechanism. Unfortunately, proteomic interactome analysis was not possible due to the difficulty to obtain a workable amount of soluble protein. Another alternative approach might be to perform an immunostainning of these proteins to see if they are secreted in the host cytoplasm.

By combining transcriptomic and proteomic results, an interesting concordance was observed. Genes presenting the lower expression level such as the ANK wBm0394 and wBm582 correspond to the recombinant proteins for which the production of soluble material was difficult whereas the proteins with the highest expression rate such as the ANK wBm0447 and the WSP wBm0100 were the easiest to express as soluble proteins. Thus, the expression rate of a gene may not be only dependent on the importance of the encoded proteins but also to the hydrophobicity of the protein and thus, the difficulty the bacteria could have to synthetize its own proteins.

An interesting but not surprising outlier of this theory is the WASP protein suggesting future studies would be worthwhile.

GENERAL DISCUSSION AND PERSPECTIVES

Wolbachia are generally maternally inherited endosymbiotic bacteria commonly found in arthropods and nematodes. *Wolbachia* have evolved at least four ways to manipulate arthropod reproduction to maximize their transmission to offspring whereas they have evolved to a mutualistic relationship with nematodes (Werren et al. 2008).
So far, the molecular mechanism of neither *Wolbachia* manipulation of host reproduction nor the basis of the mutualistic symbiosis with nematodes is known.
Herein, we described a first step to the characterization of these molecular interactions using genomic, transcriptomic and proteomic approaches.

Investigation of the genetic factors involved in phenotypic expression was initiated with the project of whole-genome sequencing of several *Wolbachia* strains. However, *Wolbachia* is an obligate endosymbiont that cannot be grown in pure culture. Therefore, the first step was to develop an efficient procedure to specifically isolate *Wolbachia* DNA from host DNA. Based on the targeted genome enrichment method developed and optimized, seven *Wolbachia* strains from isopods were sequenced and partially assembled. These strains were chosen either because of their phenotype (cytoplasmic incompatibly vs. feminization) or their position in the *wsp* based phylogenetic tree. Due to the NGS technologies used, which produced a large number of short reads and due to the large number of repetitive sequences in *Wolbachia* genomes, genomes were partially assembled. We demonstrated that all the necessary data needed for comparative genomics were obtained.

In parallel, a project was initiated to produce a first complete feminizing genome. The strain *w*VulC from the isopod *A. vulgare* was previously partially sequenced using shotgun libraries (Liu et al, 2013), but due to the highly repetitive sequence and the presence of prophages, the assembly was stopped at 10 contigs. In this study, we plan to close this genome using long read NGS such as PacBio sequencing to produce the first complete genome of a feminizing bacteria. Since *Wolbachia* genomes are relatively small genomes (1Mb to 1,7Mb), a nice perspective would be to adapt the direct PacBio sequencing procedure recently published to *Wolbachia* genomes. This procedure proposed to use random primers to allow direct sequencing without library preparation allowing a DNA amount input as low as 1ng (Coupland et al. 2012). Another alternative of the whole-genome sequencing might be to try closing the gaps by PacBio sequencing of long-range PCR product as long as 2.5kb (Zhang et al. 2012).
Even though we have not tried it, we believe that our targeted genome enrichment procedure could also be used to successfully sequence *Wolbachia* strains from multi-

infected host. Actually, since it is improbable that all the strains are present at the same density, the different strain reads could be separated *in silico* considering their abundance. This methodology is commonly use in NGS based metagenomic projects (Lindner and Renard 2013).

Comparative genomics led us to generate the first *Wolbachia* pan-genome, which corresponds to the diversity of genes present in all the sequenced *Wolbachia* strains. This first large pan-genomic analysis highlighted the wide genomic diversity and dynamics of *Wolbachia* genomes and reveals the unlikely possibility to obtain one day the full repertoire of *Wolbachia* genes; each new genome sequenced will add new genes to the pan-genome. Phages and mobile elements are actually inducing novelty in the genomes mainly through duplication, recombination and lateral gene transfers (Siguier et al. 2006). Thus, it will be interesting to sequence strains infecting the same host particularly when recombination was observed. In this instance, sequencing the isopod strain *w*VulP should reveal insights of recombination since this strain is hypothesized to have emerged from a recombination between the two strains *w*VulM and *w*VulC, presently sequenced. Deeper comparative genomics between closely related strains should also highlight evolutionary events that will argument the *Wolbachia* species concept recently discussed (Ellegaard et al. 2013) as well as defining gene loss and gain between strains.

Moreover, having 7 *Wolbachia* strains from isopods currently sequenced will allow comparison of insertion element dynamics in isopods vs. insects to speculate on the role of prokaryote transposable elements in genome dynamics (Cerveau et al. 2011). Several studies already highlighted the horizontal gene transfers between multiple *Wolbachia* strains such as symbionts of the ant *Formica exsecta* (Reuter and Keller 2003), and inter-generic DNA transfers between *Wolbachia*, *Cardinium* and *Richettsia* opening the door to the understanding of ecological connections between different endosymbiont taxa (Duron 2013). These ecological connections can be even deeper in some cases, as illustrated by the bacterial endosymbiont *Tremblaya princeps* of the *Planococcus citri* mealybugs, which possesses its own bacterial endosymbiont, *Moranella endobia*. A recent study highlighted this functional three-way-symbiosis acquired from ancient horizontal gene transfers (von Dohlen et al. 2001; Husnik et al. 2013).

Whole-genome comparisons can also be used to design a “universal vaccine”. This approach was used to design an efficient protective vaccine against group B streptococcus using the dispensable-genome (Maione et al. 2005). To develop a

protective vaccine against this multi-serotype bacterial pathogen life-threatening infection of newborns, eight genome sequences were compared to identify 396 and 193 candidate proteins from the core-genome and dispensable genome, respectively. Out of these candidates, four proteins showed an interesting pattern as a vaccine target but were all part of the dispensable genome. Each of these antigens elicited a protection against most of the strains carrying the antigen-coding gene, but in few cases, protection was not conferred. However, the combination of all four antigens showed an efficient protection against all strains assayed demonstrating the practical importance of multi-strain genome sequencing and the not so accessory dispensable-genome.

In our study, comparative genomics was focused on the identification of putative symbiosis effectors. To date, it is the first study in *Wolbachia* genomics comparing the 5 main *Wolbachia*-related symbioses. This approach revealed interesting gene-patterns. First, numerous genes were only absent from mutualistic strains. These genes should be investigated since the key of the mutualism relationship may reside in their functions.

Second, two genes, the WASP and a hypothetical protein-coding gene were identified as mutualistic-specific. Expression quantification of these two *w*Bm genes revealed a relatively high expression rate throughout the nematode life cycle particularly for the WASP gene. Unfortunately, interactome studies herein performed to identify the host interactome to the WASP proteins remained unsuccessful due to technical difficulties. However, since this protein is known to be involved in actin polymerization, further molecular characterizations should be planned to establish the insight of this protein interactome. Thus, these two proteins should be further investigated as they represent two potential drug target candidates. Actually, targeting molecular mechanisms of *Wolbachia* symbiosis in *B. malayi* and other filarial nematodes, which are human parasites is of high interest for drug discoveries against filariasis. This research orientation is one of the main goals of the anti-*Wolbachia* consortium (A-WOL) that is targeting *Wolbachia* biology to identify drugs that reduce the period of treatment and safe even for children and pregnant women (Taylor et al. 2012; Taylor et al. 2013). Therefore, many research axis have been or are currently investigated such as targeting *Wolbachia* key enzymatic and metabolic pathways (e.g. haem biosynthetic pathways (Wu et al. 2009; Wu et al. 2013) or pyruvate phosphatase dikinase (Raverdy et al. 2008)), investigating the essential set of genes in the biology of *Wolbachia* (Holman et al. 2009) or understanding of antibiotic treatment on *Wolbachia* (Schiefer et al. 2013).

Additionally, 5 genes were identified as only present in feminizing strains and 3 of them are actively expressed in adult tissues. Because of its particular position in the genome, between two genes of the SEC secretion system, the gene encoding a hypothetical protein here named gene X, was of interest.

A quick verification showed that the gene X is co-transcribed with the secF gene leading to hypothesize its translocation through the Sec secretion system where of it's taken by another secretion system to be translocated though the outer membrane. Gene expression quantification throughout *A. vulgare* post-embryonic developmental stages indicated a similar expression profile with the other putative feminization factor, the Ankyrin pk2b2 (Pichon et al. 2009) and the Type-1 secretion system. This result might mean that we have identified a feminizing gene, although its role in the feminization process has to be more clearly demonstrated. There is a laboratory *A. vulgare* lineage called Wx*f* that have integrate in its genome the fragment of *Wolbachia* genome (*f* fragment) that harbors the feminization (Figure 16; p54) (Bouchon et al. 2008; Cordaux et al. 2011). Preliminary PCR and RT-PCR experiments indicated that gene X is integrated in this Wx*f A. vulgare* genome and expressed, allowing us to plan gene-silencing experiments, which are only doable on eukaryote genes. If RNAi experiment confirms that gene X silencing inhibits host feminization, confirmation of gene X as feminizing gene will be done. Microscopic observations should also confirm its actual secretion into host cell cytoplasm. Since injection is also easily realizable in isopods, it will be interesting to inject fluorescent SNAP-tagged gene X protein in the hemolymph, or in the marsupium, to follow the route of this protein.

This discovery will thus lead to deeper characterization of the molecular basis of the *Wolbachia* symbiotic relationship with the isopod *A. vulgare* through the identification of gene X secretion path and host interactors.

Finally, applied approaches of crustacean feminization processes, such as on shrimp, crayfish or lobster, is of high industrial interest since females generally mature faster and are more sociable than males (Chavali Gopal and Gopalapillay Gopikrishna 2010; Aktas and Genc 2011).

MATERIALS AND METHODS

A. Genome sequencing

1 - Biological material

Ten *Wolbachia* strains were isolated from different hosts: one mutualistic strain *w*Bm from *Brugia malayi*, two cytoplasmic incompatibility strains *w*Dil from *Porcellio dilatatus dilatatus (*lineage PddA) and *w*Pet from *P. dilatatus petiti (*lineage Pdp), six feminizing strains *w*VulC and *w*VulM from *A. vulgare* (lineage Zn and BI respectively), *w*Albu from *A. album (lineage XX), w*Nas from *A. nasatum (lineage XX), w*Pru from *Porcellionides pruinosus* (lineage *PruIII*) *and w*Ase from *Oniscus Asellus* (lineage *XX*) and one 'unknown phenotype' strain *w*Bre from *Helleria brevisconis*.

2 - DNA isolation

Total DNA from the nematode *B. malayi* and from isopods were extracted as described (Sambrook and Russell 2001; Bouchon et al. 1998). Quantification of the DNA samples was performed using a Nanodrop 1,000 spectrophotometer (Thermo Scientific) and the Qubit 2.0 fluorimeter (Invitrogen). DNA samples of the strains *w*Bm, *w*Albu, *w*Bre, *w*Dil, *w*Pet, *w*Pru, *w*VulC and *w*VulM were normalized to 3 µg of DNA prior to sonication to an average length of 200 bp using a Covaris S1 ultrasonicator. Illumina sequencing libraries were prepared using the NEBNext® Sample Preparation Kit (New England Biolabs) as previously described (paper capture). DNA samples of the strains *w*Ase, *w*VulC, *w*Dil and *w*Pru were normalized to 1.5 ug of DNA prior to sonication to an average length of 500 bp using a Covaris S1. Illumina sequencing libraries were prepared using the NEBNext® Ultra Sample Preparation Kit (New England Biolabs). All samples were quality assayed by Bioanalyzer (Agilent) and Qubit 2.0 fluorimeter (Invitrogen).

3 - Capture and sequencing

Wolbachia DNA was captured from the prepared total DNA by hybridization to the biotinylated cRNA baits for 24 h at 65 °C, following the Agilent SureSelect™ protocol, but supplemented with custom blocking oligos complementary to the barcoded adaptors. Bound DNA was recovered using magnetic streptavidin beads, PCR amplified (12 cycles) using Illumina forward and reverse primers and purified. The 200 bp libraries were 50 bp paired-end sequenced on the Illumina HiSeq 2000 at HudsonAlpha, Inc and the 500 bp libraries were 250 bp paired-end sequenced on the Illumina MiSeq at New England Biolabs.

4 - DNA preparation for PacBio sequencing

Wolbachia wVulC DNA was extracted from *A. vulgare* (lineage Zn) ovaries using the DNeasy Blood & Tissue Kit (Qiagen). Ovaries in 2 mL ATL buffer were gently crushed using a glass Dounce homogenizer. A 50 µl aliquot was collected for further enrichment calculation. The homogenized sample was filtered through a 5 µm syringe filter and DNA was extracted following the manufacturer's procedure adding the recommended RNAse treatment step. Total DNA was quantified using a nanodrop spectrophotometer and purity was assayed by running an aliquot on a 1% agarose gel.

The enrichment efficiency was estimated by quantitative PCR on single copy genes by comparison of enriched sample with the sample collected before filtration. Host nuclear DNA quantity was estimated using the androgenic hormone gene; host mitochondrial DNA quantity was estimated using the gene COI gene and the *Wolbachia* DNA quantity was estimated using the *wsp* gene. (Appendix 5). qPCR were performed on a LightCycler480 (Roche) with the program 10 min at 95°C followed by 45 cycles of 95°C for 10 s, 60°C for 10 s and 72°C for 20 s. A melting curve (65°C to 97°C) was recorded at the end of each reaction to check the specificity of the PCR reaction. Efficiency of the PCR reaction was calculated. For each gene, absolute quantification was performed using a standard curve plotted with 6 dilutions of PCR product.

B. Genome assembly

1 - Assembly and gap closure

Sequence reads were quality controlled using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Assemblies were performed using Velvet assembler (Version 1.2.03) (Zerbino and Birney 2008). Optimization of the assembly was performed using different k-mers (from 19 to 49 by steps of 2 bases); the optimal assembly was chosen considering the N50, the length of the longest contig and the total bases in the contig. The Velvet assembly was then optimized by gap filling using GapFiller (Nadalin et al. 2012) followed by optimization of the scaffolding to reduce the total number of scaffold. After removing of the Illumina reads containing 'N', SSpace (Boetzer et al. 2011) software was used to scaffold theses pre-assembled contigs using the original paired-end reads. Finally, gaps were closed using again GapFiller.

2 - *Wolbachia* sequences isolation and genome annotation

Potential host mitochondrial DNA contamination was removed by BLAT against a local database of the ten currently published isopod mitochondrial genomes: Asellus aquaticus mitochondrion (GU130252), *Eophreatoicus sp. 14 FK-2009* mitochondrion (FJ790313), *Sphaeroma serratum* mitochondrion (GU130256), *A. vulgare* mitochondrion (GU130251), *A. vulgare* mitochondrion (EF643519), *Ligia oceanica* mitochondrion (NC_008412), *Janira maculosa* mitochondrion (GU130255), *Idotea baltica* mitochondrion (DQ442915), *Glyptonotus cf. antarcticus* FK-2009 mitochondrion (GU130254), *Eurydice pulchra* mitochondrion (GU130253). Lastly, *Wolbachia* DNA contigs were identified from host DNA and potential contaminants by BLAT against a database of the 20 currently published complete or partial *Wolbachia* genomes from *Muscidifurax uniraptor* (*w*Uni, NZ_ACFP00000000.1), *Wuchereria bancrofti* (*w*Wb, ADHD00000000), *Onchocerca volvulus* (*w*Ov, ADHE00000000), *Drosophila willistoni* (*w*Wil, AAQP00000000), *D. simulans* (*w*Sim, AAGC00000000.1), *D. melanogaster* (*w*Mel, AE017196.1), *D. ananassae gdan 143* (*w*Ana, AAGB00000000), *Culex quinquefasciatus pel* (*w*Pip-Pel, AM999887.1), *C. quinquefasciatus jhb* (*w*Pip-JHB, ABZA00000000), *B. malayi* (*w*Bm, AE017321.1), *Onchocerca ochengi* (*w*Oo, NC_018267), *Nasonia vitripennis* (*w*VitB, NZ_AERW00000000), *Aedes albopictus* (*w*AlB, NZ_CAGB00000000), *D. simulans* (*w*Ri, NC_012416), *Hypolimnas bolina* (*w*Bol1-b, CAOH00000000.1), *Diaphorina citri* (*w*Di, AMZJ01000001), *Dirofilaria immitis str. Pavia*, (*w*Dim, PRJNA74405), *C. pipiens molestus* (*w*Pip-Mol, CACK0100001), *D. suzukii* (wSuzi, CAOU02000001), *A. vulgare* (*w*VulC, ALWU00000000).

On final assemblies, the software Prodigal (Hyatt et al. 2010) was used for gene prediction and the pipeline prokka (Prokaryotic Genome Annotation System - http://vicbioinformatics.com/) was used to generate an automatic first draft annotation.

C. Ortholog retrieval

The ortholog retrieval was performed on the 7 studied *Wolbachia* strains and the 6 published complete *Wolbachia* genomes (*w*Pip, *w*Oo, *w*Ri, *w*Mel, *w*No, *w*Ha). Homologous genes between these genomes were determined using an all-against-all protein blast between all the proteins sequences from each genome. Orthologous proteins were then clustered with the OrthoMCL algorithm (Li et al. 2003). Based on the reciprocal BLAST output, this program defines putative pairs of orthologs and recent paralogs. Recent paralogs are identified as genes from the same genome that are more

similar to each other than any sequence from another sequence. Finally, OrthoMCL returns a set of clusters containing orthologs and/or recent paralogs. OrthoMCL was run with an E-value cut-off of 1e-5. The OrthoMCL output was used to generate a table describing genome gene content supplemented by the list of taxon specific genes established with the genes that were not included in any cluster and had no significant hit with any other protein (E-value <1e-10).

This table was used to plot a Venn diagram of the *Wolbachia* genome gene content and to construct a first core-genome data-set of the genome of *Wolbachia* strains from isopods and a second core-genome supplemented by the published complete *Wolbachia* genomes.

D. Phylogenomics

Maximum likelihood phylogenetic trees were computed at the protein level using RAxML (Stamatakis 2006) based on orthologous clusters from the OrthoMCL output of exactly one sequence of each genome. Every cluster of proteins were aligned using Mafft (Katoh et al. 2002) and best-fit models of amino acid replacement were selected with ProtTest (Darriba et al. 2011). The concatenated alignments were used to generate strain phylogenies with specific gene models. A 100 bootstraps run was performed to identify the best scoring tree with maximum likelihood and bootstrap supports were calculated from 1000 rapid bootstraps.

Unrooted *Wolbachia* strains phylogenies were generated as well as a phylogeny rooted with the closest non-*Wolbachia* strains: *Erlichia* and *Analplasma*.

E. DNA and RNA library constructions

1 - Animal collection

A. vulgare (Crustacea Isopoda) individuals used in this study come from two laboratory lineages: the BF lineage originating from Nice (France) and the ZN lineage where females are infected by the feminizing *Wolbachia* strain (*w*VulC) originating from Celles-sur-Belle (France) (Bouchon *et al.*, 1998). These lineages have been stably maintained in the lab since 1967 and 1991, respectively. All animals were reared under laboratory conditions in boxes containing wet compost and food *ad libitum* (fresh slices of carrots and dried limetree leaves), at 20°C under the natural photoperiod of Poitiers (46°40'N).

Several crosses of male BF with female ZN were performed providing *Wolbachia*-infected offspring. These crosses were followed up in order to harvest animals at each post-embryonic developmental stage (Table 2) (Suzuki et al, 1995). 20 individuals were immediately stored separately in liquid nitrogen for all stages but early stages, which were pooled by 20 infants in order to get a workable amount of RNA and DNA.

2 - DNA/RNA isolation and reverse transcription (RT)

DNA and RNA were extracted from each *A. vulgare* individual using the Qiagen AllPrep DNA/RNA/Protein Mini Kit according to the manufacturer's recommendations after sonication. The RNA sample were normalized at 562.5 ng in 15.7µl total and treated with one unit of DNAseI (RNAse free, NEB) in 1X DNAseI buffer in order to minimize genomic DNA contaminations. DNA and RNA integrity, purity and quantity were assessed by nanodrop spectrophotometry.

Each sample was divided in two tubes and reverse transcription was performed using the SuperScript III First Strand Synthesis System (Invitrogen) with random primers accordingly to the manufacturer's instructions with (RT+) or without (RT-) the reverse transcriptase.

On ice, all complementary DNA (cDNA) samples were diluted 1:5, and stored at -80C in 11ul aliquots.

For each gene of interest, primer pairs were designed to have a Tm at 60°C and a length of 18-22bp. Primers used for quantitative PCR are summarized in Appendix 3. Reaction mixture consisted of 1X of Fast SYBR-Green Master Mix (Roche), 500 nM of each primer and the appropriate amount of DNA or cDNA.

All PCR reactions were performed with a technical replicate on a LightCycler LC480 system (Roche) under the following conditions: initial denaturation at 95°C 10 min with 45 cycles of 95°C 10s; 60°C 10s followed by a final extension of 72°C 20 s. A melting curve (65°C to 97°C) was recorded at the end of each reaction to ensure the uniqueness of the PCR product and the efficiency of each primer pair was calculated using the standard curves.

F. Absolute quantification of the *Wolbachia* load by quantitative PCR (qPCR)

In order to generate DNA standard for qPCR calibration, for each primer pairs, a regular PCR was performed on DNA extracted from adult female ovaries using Phusion polymerase. Amplicons were purified with the Qiagen PCR product purification kit. DNA

quality and concentration were assessed by agarose gel and nanodrop quantification. Copy number of gene of interest was calculated using the following formula:

$$\frac{Avogadro's\ number\ (\text{copies/mol}) \times amount\ of\ DNA\ (\text{g})}{amplicon\ size\ (\text{bp}) \times molecular\ weight\ of\ dsDNA\ (\text{g/mol of bp})} = \text{copy number}$$

10 fold serial dilutions (10^8 copies to 10^2 copies) of PCR products were performed to generate the standard curves.

Wolbachia density was determined in each DNA sample previously extracted from young *A. vulgare* at different post-embryonic stages by qPCR amplification of the *Wolbachia surface protein* (*wsp)* gene and the *androgenic hormone (AH)* gene. A standard was run in parallel with the samples to determine the absolute number of *wsp* and *AH* copies and the bacterial density was calculated in copy number of bacterial genome per copy number of host genome. Statistically significant differences between groups were analyzed with a Kruskal test using a p value=0.05.

G. Relative quantification of *Wolbachia* gene expression by reverse transcriptase quantitative PCR (RT-qPCR)

RT-qPCR were performed as previously described using 2.5 µl of 5 fold diluted cDNA as template. Standard curves were plotted for each primer pair using 4 dilutions of cDNA from ovaries. Efficiency of the PCR reaction was calculated for each gene. Expression data were calculated using the efficiency of the primer pair and the crossing point (39fred). Starting quantities of target mRNA were calculated relatively to the starting quantities of the reference mRNA in the same sample following the formula:

$$Relative\ mRNA\ concentration\ = \frac{Efficiency_{target}^{-Cp_{target}}}{Efficiency_{ref}^{-Cp_{ref}}}$$

This calculation allows a correction for both technical differences (such as different efficiency of the RT reaction) and bacterial infection level differences among samples. Normalization to the stage 0 of the post-embryonic development was then applied to facilitate relative expression of target gene through post-embryonic development stages:

$$Ratio\ = \frac{(Efficiency_{target})^{\Delta Cp_{target}\ (stage\ 0 - stage\ of\ interest)}}{(Efficiency_{ref})^{\Delta Cp_{ref}\ (stage\ 0 - stage\ of\ interest)}}$$

H. Analysis of putative symbiosis related bacterial factors

Selected HMM motifs (Table 16) were downloaded from Pfam website (http://pfam.sanger.ac.uk) and compiled in an internal database using the HMMER3 cammand lines hmmbuild (Johnson et al. 2010). This database was used to screen genomic sequences using the command line hmmsearch.

I. Interactors screening

1 - GST-pull down

a) Molecular cloning

(i) Cloning strategies

Cloning was performed utilizing the GST Gene Fusion System (pGEX-5X-1 vector, GE Healthcare), which allows directional cloning and protein expression in *E. coli* of fusion proteins with the GST moiety at the amino terminus and the protein of interest at the carboxyl terminus.

To amplify *Wolbachia* DNA fragments required for obtaining the selected *Wolbachia* genes, polymerase chain reaction (PCR) was performed on *B. malayi* genomic DNA which contains three different genomes, the chromosomal (~100mB) and the mitochondrial (13.657 kb) *Brugia malayi* genomes but also the obligate intracellular symbiont, *Wolbachia* genome (1.08mB).

From Bill Sullivan's research team (UCSC, personal communication) and several NEB teams, we knew that it has been generally difficult to clone and express *Wolbachia* proteins as soluble, active proteins. Thus, primers were designed using the hydrophobicity plot on the non-hydrophobic areas (Appendix 1). Primers used are listed below (Appendix 3). These primers were designed to also contain restriction endonuclease sites for digestion for cloning into cognate sites in appropriate vectors to create the correct protein expression constructs. All primers were produced by Integrated DNA Technologies (IDT) and used at a final concentration of 0.5 pmol/µL each. Underlined regions show the restriction endonuclease recognition sites used for subsequent cloning (GGATCC: BamHI, CTCGAG: XhoI).

To perform the PCRs, 50 µl reactions were constructed using 0.02 unit/µl of PhusionTM High- Fidelity DNA Polymerase (NEB #F530), forward and reverse primers, 10 ng of template DNA, 200 µM of dNTPs (NEB), 1X buffer (stock solution: PhusionTM HF buffer 5X) and water to complete the 50µl reaction.

The PCR program followed (Table3) was run with a BioRad My CyclerTM thermal cycler.

GelPilot Loading Dye 5X from Qiagen was added to a final 1X concentration with the DNA samples. This final mix of 5 µl (1 µl of dye and 4 µl of sample) was then loaded on a 1.2% agarose gel in 1X Tris borate EDTA (TBE, 5X stock from AccuGENE: 0.45M Tris borate, 0.01M EDTA, pH 8.3, supplemented with 10mg/mL Ethidium Bromide).

Electrophoresis was performed in 1X TBE at 100 V until the sample migration was sufficient. Visualization and photography was performed under ultra violet light with the AlphaImagerR HP (AlphaInnotech).

PCR amplified DNA were column-purified using the Qiagen QIAquick PCR purification system, as described in their manual. Elutions were in 50 µl volumes of MilliQ water, pH 8.0.

BamHI (NEB) and XhoI (NEB) restriction endonucleases were used to generate compatible cohesive ends on the PCR fragments and on the pGEX-5X-1 multicloning site in the vector for cloning. In a basic reaction, 1 µg of DNA or 100ng of pGEX-5X-1 vector was incubated at 37°C for 4 hours with 400 units/mL of enzyme, 100 µg/mL of Bovine Serum Albumin (BSA) in 1X of NEBuffer 2 (50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM Dithiothreitol, pH 7.9 at 25°C) in a final volume of 100µl.

Enzyme digests were column purified as described in part II-3.

Ligations were performed with a 3:1 insert: vector molar ratio using 1µL of 10X T4 DNA Ligase Reaction Buffer (50 mM Tris-HCl (pH 7.5), 10 mM MgCl2, 10 mM dithiothreitol (DTT), 1 mM Adenosine triphosphate (ATP)) and 0.5 µL (200 units) of T4 DNA Ligase (NEB) in a 10 µL final volume. The reaction was incubated at room temperature for 10 minutes and then 5 µl were used for transformation.

Transformations were carried out using the following conditions: 50 µL T7 Express Competent *E. coli* (NEB #C2566; *fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1* Δ(mcrC-mrr)114::IS10) suspension and 5 ng plasmid DNA (5 µL of ligation reaction) were incubated on ice for 30 minutes. Cells were heat-shocked for 30 seconds at 42°C and chilled on ice for 2 minutes. After addition of 400 µL of room temperature Luria Bertani media (LBm: 10g Bacto tryptone, 5g yeast extract, 10g NaCl, 1g dextrose, 1g MgCl2,6H2O, 2mL NaOH complete to 1L with distilled water), cells were incubated at 37°C for 1 hour for cell recovery and for expression of antibiotic resistance. Then, the cells were plated on LBm agar plates (15 g/liter agarose) containing 100 µg/mL of ampicillin and incubated overnight at 37°C.

A first antibiotic resistance screening was made on ampicillin containing plates as ampicillin resistance is carried by the pGEX-5X-1 vector. Resistant colonies were

screened by colony PCR followed by separation in agarose gels. 10 widely separated colonies were marked and picked, with a sterile 20 µl micropipette tip, diluted in 20 µL of water and 1 µl used as template for PCR reaction which was run as described in the part II-1 but in a final volume of 20µ l.

Primers used were designed based on the pGEX-5X-1 vector sequence surrounding the cloning sites (Table 4) and the length of the PCR products was checked on agarose gel.

Where the amplified fragment lengths appeared to be of correct length, the screening was confirmed by sequencing of plasmid DNA purified from a 37°C overnight 10 mL LBm culture begun with a pick of the original colony.

Plasmid DNAs were isolated and column-purified using the Qiagen QIAprep Spin Miniprep Kit and submitted to the DNA sequencing lab at NEB.

DNA thermal cycle sequencing was performed by the NEB sequencing lab using an ABI 3100 sequencer based on the Sanger method (Sanger et al., 1977) using 3.2 pmol/µl primer (5 µl) and 8 ng/µl PCR product (5 µl) in a 20 µl reaction containing 1/4X BigDye sequencing reagent. Reactions were purified using an EDGE Biosystems purification matrix (as described in their literature) before electrophoresis.

For comparison, sequences were aligned with known sequences from the NCBI database using the Lasergene (DNA STAR) SeqMan program.

A 1mL of 5 % glycerol stock solution was made from the *E. coli* cultures and stored at -80°C.

(ii) GST-Protein expression and purification

Correct sequence plasmids from step II-7 were transformed as described in part II-6 into the *E. coli* expression strain NEB 2566-CodonPlus competent cells (E. *coli* 2566 F– *ompT hsdS(rB–* mB–) *dcm+* Tetr *gal endA* Hte *[argU ileY leuW* Camr]) and selected on LB + ampicillin plates supplemented with 34µg/mL chloramphenicol. The LB media used for the protein expression was supplemented with 100µ g/mL ampicillin and 34µ g/mL chloramphenicol (LBexp)

Before growing large scale cultures, the expression of each protein was tested in a small scale to check its ability to give sufficient protein yields upon induction and to determine its solubility (i.e protein in the supernatant vs. insoluble pellet). A single ampicillin and chloramphenicol resistant colony, was picked and used to inoculate a 10 mL LBexp pre-culture. Cultures were inoculated overnight at 37°C. These pre-cultures were diluted in ratio 1:40 in 4 mL LBexp culture and incubated at 37°C until OD600nm reached 0.6-0.8. Then, they were induced with a final concentration of 25µM of

IsoPropyl-beta-D-ThioGalactopyranoside (IPTG) and different temperatures and times of growth were tested.

Crude cell extracts and lysate samples from induced and uninduced cells were examined on SDS-PAGE gels as described in part II-3.

For large scale production, one liter of LBexp was inoculated with 20 mL of an overnight culture. The cultures were grown to OD600nm=0.6 - 0.8 at 37°C, induced with 25mM IPTG and incubated at the optimal condition. Cultures were divided in 4 equal volumes and centrifuged at 9 000 rpm for 20 min at 4°C. The resulting bacterial pellets were assembled and stored at – 20°C.

(iii) Separation of proteins on sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

Samples were prepared by adding 2 volumes of protein sample to 1 volume 3X SDS Sample Buffer (187.5 mM Tris-HCl (pH 6.8 at 25°C), 6% (w/v) SDS, 30% glycerol and 0.03% (w/v) phenol red) and 1X reducing agent (stock solution: 30X :1.25 M DTT) then heated to 95°C for 10 min before loading.

Novex Tris-Glycine gels (10-20 %) used were purchased from Invitogen. Electrophoresis was performed in 1X Tris Glycine SDS buffer (125 mM Tris, 1 M glycine, 20mM SDS) at 30-40 mA until the migration was sufficient. A prestained protein marker (Broad range, 20 µL, NEB # P7702) was loaded to monitor the migration. Gels were stained with SimplyBlueTM SafeStain (Invitrogen).

All steps of protein preparation were performed at 4°C or on ice. After thawing frozen bacterial pellets on ice for 15 minutes from the step described in part III-2, they were resuspended in 13 mL lysis buffer (50mM Tris, pH 8, 100mM NaCl, 2mM EDTA, 1mM chymostatin, 1mM pepstatin, 1 mM leupeptin, 10 mM DTT). A final concentration of 100µg/mL lysozyme was added and the solution was incubated for 30 minutes on ice. Bacterial suspensions were then sonicated 6 times, using 1 minute pulses using a sonicator at 60 % capacity. The cell debris was removed by centrifugation at 9,000g for 20 minutes. The soluble supernatant fractions were stored in 1 mL aliquots at -80°C. 40 µL were taken out before and after the centrifugation to be analyzed by SDS-PAGE .

To optimize the efficiency and the specificity of the glutathione beads binding, GST-protein purifications were performed as a binding test.

The GST-proteins were purified from bacterial lysates by affinity chromatography using Glutathione SepharoseTM 4 Fast Flow from GE Healthcare (Table 5).

10 µl of packed glutathione beads (13 µl of the stock solution) were equilibrated five times with 500 µl of binding buffer (20mM Hepes (pH7), 150mM KOAc, 2mM MgOAc, 2mM

DTT (freshly made), 0.1% TetraFluoro Acid (TFA)) and low-speed centrifuged at 2,000 rpm for 2 minutes to remove the supernatant.

(iv) Immobilization of GST fusion protein on beads

The frozen E. coli extracts were added to the beads to achieve 20 µg of GST fusion protein on 10 µg Glutathione beads. The final volume was adjusted to 1 mL with binding buffer.

GST fusion proteins were bound for 15-20 min at 4°C with continual tumbling end over end.

Glutathione beads containing bound GST fusion proteins were washed twice with binding buffer as described for the beads equilibration.

Elution

GST fusion proteins were eluted with 40 µl of freshly made elution buffer (binding buffer supplemented with 20mM reduced glutathione).

b) Detection of protein-protein interactions using the GST-pull-down method

The first 3 steps were identical to the GST-protein purification procedure (III-5)

Glutathione beads were washed twice with binding buffer with a final concentration of 1mM ATP to remove heat shock proteins, followed by once with binding buffer with 1M NaCl to eliminate unspecific bonding and finally twice with standard binding buffer to re-equilibrate the beads.

Frozen *Brugia malayi* worms were obtained from TRS Labs, Athens Gerogia and were treated on binding buffer (1mL/100 worms) supplemented with protease inhibitors (solution stock : 5000X). They were vortexed in a 1.5mL Eppendorf tube for 10 min and then sonicated at 60% capacity by 1 min pulses as often as necessary to break them completely.

The worm lysate (100 frozen worms) was incubated with the glutathione beads from the step IV-2 for 2 hours tumbling at 4°C.

Regular washes with binding buffer were done 3 times.

Potential worm extract proteins that might be bound were eluted with 40 µl of binding buffer supplemented with 1M NaCl.

10 µl of the supernatant was analyzed by SDS PAGE and Sypro Ruby staining and the protein sample was sequencing submitted.

Protein gels for analysis of bond proteins were stained with a more sensitive method than the SimplyBlueTM SafeStain. The Silver stain kit and the Sypro Ruby kit from Invitrogen were tested and used as described in their respective procedures.

All protein characterizations were performed by the NEB protein sequencing lab through a MS/MS analysis after a trypsin digestion.

To validate the potential protein-protein interactions, pull downs should be performed with two recombinant proteins as bait and prey as well as a reverse pull down, using the *B. malayi* protein as the bait and the *Wolbachia* protein as the prey.

One *B. malayi* protein that was identified by mass spectrometry as a bound and eluted protein has to be cloned, as described previously, into a pET21a vector (Invitrogen) to express the His tag protein and in the pGEX-5X-1 vector to express the GST fusion protein. The DNA sequence of the protein was determined from the *B. malayi* genomic sequence database and from that sequence, primers were designed for cloning, as described in part II.

2 - Phage display

The PhD-7 and PhD-12 libraries (New England Biolbas) were used to identify peptides that bind to selected *Wolbachia* proteins from the *Wolbachia* strains *w*Bm and *w*VulC. Recombinant proteins used as bait, were expressed in *E. coli* with a GST tag at the X terminus. The libraries consist of randomized 7 and 12 mers fused to N-terminus of a minor coat protein (pIII) of M13 phage.

Panning phage display libraries

Target proteins (100nM) were immobilized on magnetic glutathione beads previously washed for 1 hour at 4C and washed four times with TBST (1X TBS + 0.1% (v/v) Tween 20). Panning against the targets was performed for 20 minutes at room temperature in a final volume of 200 ul TBST (1X TBS + 0.1% (v/v) Tween 20) containing 2.10^{11} plaque-forming units (pfu) of each library. The beads were washed 10 times with 1 mL TBST. Bound phages were eluted by incubating the beads with elution buffer (50 mM Glutathion, 50 mM Tris buffer) for 15 minutes at room temperature with gentle agitation. A volume of 10 ul of each eluate was titrated as described below.

To increase the selectivity of the phage binding three rounds of panning were performed. For the second and the third rounds of panning, $2x10^{11}$ pfu of amplified phage were incubated with the baits and the Tween 20 concentration for the washes was increased to 0.5%. A negative selection step was performed before these rounds incubating the amplified phage with beads previously bound with GST proteins. The eluate of the negative selection step was used as the phage input for the phage panning.

Amplification

Eluates were amplified by addition to a 20 ml culture of ER2738, grown to early log, and further incubation at 37°C for 4.5 hours with vigorous shaking. The culture was centrifuged at 13,800 x g for 10 min at 4°C. The supernatant was collected and the phage precipitated at 4°C overnight by addition of 1/6 volume of 20% PEG/2.5M NaCl. The phage were pelleted at 13,800 x g for 15 min at 4°C, and resuspended in 1 ml TBS. Following centrifugation for 5 min at 4°C to remove residual cells, a second precipitation was performed by adding 1/6 volume 20% PEG/2.5 M NaCl and incubating on ice for 15-60 min. The solution was then microcentrifuged at maximum speed for 10 min at 4°C. The final pellet was resuspended in 200 µl TBS, 0.02% NaN_3 and the titer of the resulting phage stock was determined.

Titration

The phage concentration was determined by 5 serial 10-fold dilutions of the phage in LB medium and infection of 200 µl of a mid log ER2738 culture, followed by plating of each dilution on the X-gal/IPTG medium to enable rapid identification of the correct plaques by their blue color. Blue plaques were counted after overnight incubation.

Sequencing

Single-stranded DNA sequencing was performed to determine if the panning was converging on particular sequences. After the third round, un-amplified phages were diluted to yield 10-100 plaques per plate. Individual plaques were then selected and cultured independently. Phage in the supernatant (500 µl) were precipitated at RT for 10 min by adding 20% PEG/2.5 M NaCl (200 µl). Following centrifugation, the pellet was resuspended in 100 µl iodide buffer (10 mM Tris-HCl, pH 8.0; 1 mM EDTA, 4 M NaI) to denature coat proteins. DNA was precipitated at RT for 10 min by adding 250µl 95% ethanol. The sample was then microcentrifuged for 10 min at maximum speed, and the pellet washed in 70% ethanol and dried under vacuum. The DNA was resuspended in 30µl TE (10 mM Tris-HCl (pH 8.0), 1 mM EDTA) and used as a template for dideoxy DNA sequencing using the -96 M13 sequencing primer (5´-CCCTCATAGTTAGCGTAACG- 3´, NEB).

Sequences were analyzed and all non-gapped pairwise alignments with a minimum overlap of 7 residues were generated.

Phage ELISA

Microtiter plate wells were coated with 200 µl of bait protein (10 µg/ml) in 0.1 M $NaHCO_3$ (pH 8.6) overnight at 4°C and blocked with 200 µl blocking buffer (5mg/ml BSA in 0.1 M $NaHCO_3$, pH 8.6, 0.02% NaN_3) at RT for 1h. The phage clones at concentration of 1.8-3.0 x 10^{11} pfu/µl were diluted in 200 µl wash buffer (0.5% Tween-20) and added to the micro-plate and incubated 1h at RT. The plate was then washed 6 times with wash buffer. Horseradish peroxidase-labeled mouse anti-M13 antibody (1:2500) in blocking buffer was added and incubated for 1h at RT. Following washing, 200 µl of substrate solution (0.22 mg/ml, diammonium 2,2'-azino-di-(3-ethylbenzthiazoline sulfonate in 50 mM citric acid and 1.7 µl 30% H_20_2/ml, pH 4.0) was added and the plates were incubated for 60 min at RT. Absorbance at 405 nm was then determined.

REFERENCES

Aktas M, Genc MA (2011) The Effects of 17 β-Estradiol on Growth, Survival and Feminization of Green Tiger Shrimp, P. semisulcatus (Decapoda: Penaeidae). Journal of Animal and Veterinary Advances 10 (5):562-565

Al-Khedery B, Lundgren AM, Stuen S, Granquist EG, Munderloh UG, Nelson CM, Alleman AR, Mahan SM, Barbet AF (2012) Structure of the type IV secretion system in different strains of Anaplasma phagocytophilum. BMC Genomics 13:678. doi:10.1186/1471-2164-13-678

Al-Khodor S, Al-Quadan T, Abu Kwaik Y (2010a) Temporal and differential regulation of expression of the eukaryotic-like ankyrin effectors of Legionella pneumophila. Environ Microbiol Rep 2 (5):677-684. doi:10.1111/j.1758-2229.2010.00159.x

Al-Khodor S, Price CT, Kalia A, Abu Kwaik Y (2010b) Functional diversity of ankyrin repeats in microbial proteins. Trends Microbiol 18 (3):132-139. doi:10.1016/j.tim.2009.11.004

Andersen SB, Boye M, Nash DR, Boomsma JJ (2012) Dynamic Wolbachia prevalence in Acromyrmex leaf-cutting ants: potential for a nutritional symbiosis. J Evol Biol 25 (7):1340-1350. doi:10.1111/j.1420-9101.2012.02521.x

Attaiech L, Granadel C, Claverys JP, Martin B (2008) RadC, a misleading name? J Bacteriol 190 (16):5729-5732. doi:10.1128/JB.00425-08

Backert S, Meyer TF (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. Curr Opin Microbiol 9 (2):207-217. doi:10.1016/j.mib.2006.02.008

Baldo L, Ayoub NA, Hayashi CY, Russell JA, Stahlhut JK, Werren JH (2008) 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 Ecol 17 (2):557-569. doi:10.1111/j.1365-294X.2007.03608.x

Baldo L, Bordenstein S, Wernegreen JJ, Werren JH (2006) Widespread recombination throughout Wolbachia genomes. Mol Biol Evol 23 (2):437-449. doi:10.1093/molbev/msj049

Baldo L, Werren JH (2007) Revisiting Wolbachia supergroup typing based on WSP: spurious lineages and discordance with MLST. Curr Microbiol 55 (1):81-87. doi:10.1007/s00284-007-0055-8

Bandi C, Anderson TJ, Genchi C, Blaxter ML (1998) Phylogeny of Wolbachia in filarial nematodes. Proc Biol Sci 265 (1413):2407-2413

Bandi C, Slatko B, O'Neill SL (1999) Wolbachia genomes and the many faces of symbiosis. Parasitol Today 15 (11):428-429

Bandyopadhyay P, Sumer EU, Jayakumar D, Liu S, Xiao H, Steinman HM (2012) Implication of proteins containing tetratricopeptide repeats in conditional

virulence phenotypes of Legionella pneumophila. J Bacteriol 194 (14):3579-3588. doi:10.1128/JB.00399-12

Baron C, Coombes B (2007) Targeting bacterial secretion systems: benefits of disarmament in the microcosm. Infect Disord Drug Targets 7 (1):19-27

Baton LA, Pacidonio EC, Goncalves Dda S, Moreira LA (2013) wFlu: characterization and evaluation of a native Wolbachia from the mosquito Aedes fluviatilis as a potential vector control agent. PLoS One 8 (3):e59619. doi:10.1371/journal.pone.0059619

Bentley S (2009) Sequencing the species pan-genome. Nat Rev Microbiol 7 (4):258-259. doi:10.1038/nrmicro2123

Bian G, Joshi D, Dong Y, Lu P, Zhou G, Pan X, Xu Y, Dimopoulos G, Xi Z (2013a) Wolbachia invades Anopheles stephensi populations and induces refractoriness to Plasmodium infection. Science 340 (6133):748-751. doi:10.1126/science.1236192

Bian G, Xu Y, Lu P, Xie Y, Xi Z (2010) The endosymbiotic bacterium Wolbachia induces resistance to dengue virus in Aedes aegypti. PLoS Pathog 6 (4):e1000833. doi:10.1371/journal.ppat.1000833

Bian G, Zhou G, Lu P, Xi Z (2013b) Replacing a native Wolbachia with a novel strain results in an increase in endosymbiont load and resistance to dengue virus in a mosquito vector. PLoS Negl Trop Dis 7 (6):e2250. doi:10.1371/journal.pntd.0002250

Blatch GL, Lassle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays 21 (11):932-939. doi:10.1002/(SICI)1521-1878(199911)21:11<932::AID-BIES5>3.0.CO;2-N

Bockarie MJ, Fischer P, Williams SA, Zimmerman PA, Griffin L, Alpers MP, Kazura JW (2000) Application of a polymerase chain reaction-ELISA to detect Wuchereria bancrofti in pools of wild-caught Anopheles punctulatus in a filariasis control area in Papua New Guinea. Am J Trop Med Hyg 62 (3):363-367

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27 (4):578-579. doi:10.1093/bioinformatics/btq683

Bordenstein SR, Fitch DH, Werren JH (2003) Absence of wolbachia in nonfilariid nematodes. J Nematol 35 (3):266-270

Bordenstein SR, Paraskevopoulos C, Dunning Hotopp JC, Sapountzis P, Lo N, Bandi C, Tettelin H, Werren JH, Bourtzis K (2009) Parasitism and mutualism in Wolbachia: what the phylogenomic trees can and cannot say. Mol Biol Evol 26 (1):231-241. doi:10.1093/molbev/msn243

Boreham PF, Atwell RB (1983) Adverse drug reactions in the treatment of filarial parasites: haematological, biochemical, immunological and pharmacological changes in Dirofilaria immitis infected dogs treated with diethylcarbamazine. Int J Parasitol 13 (6):547-556

Bouchon D, Cordaux R, Grève P (2008) Feminizing / Wolbachia/ and the evolution of sex determination in isopods. . Insect Symbiosis, pp 273-294

Bouchon D, Rigaud T, Juchault P (1998) Evidence for widespread Wolbachia infection in isopod crustaceans: molecular identification and host feminization. Proc R Soc Lond B Biol Sci 265 (1401):1081-1090

Bourtzis K, and T. A. Miller. (2003) Insect symbiosis. CRC Press, Boca Raton.,

Bourtzis K, Nirgianaki, A., Markakis, G. and Savakis, C. (1996) Wolbachia Infection and Cytoplasmic Incompatibility in Drosophila Species. Genetics 144:1063-1073

Boussau B, Karlberg EO, Frank AC, Legault BA, Andersson SG (2004) Computational inference of scenarios for alpha-proteobacterial genome evolution. Proc Natl Acad Sci U S A 101 (26):9722-9727. doi:10.1073/pnas.0400975101

Breeuwer JAJ, Stouthamer, R., Barns, S.M., Pelletier, D.A., Weisburg, W.G. and Werren, J.H. (1992) Phylogeny of cytoplasmic incompatibility microorganisms in the parasitoid wasp genus Nasonia (Hymenoptera: Pteromalidae) based on 16S ribosomal DNA sequences. Insect Mol Biol 1 (1):25 36

Brelsfoard CL, St Clair W, Dobson SL (2009) Integration of irradiation with cytoplasmic incompatibility to facilitate a lymphatic filariasis vector elimination approach. Parasit Vectors 2 (1):38. doi:10.1186/1756-3305-2-38

Broms JE, Edqvist PJ, Forsberg A, Francis MS (2006) Tetratricopeptide repeats are essential for PcrH chaperone function in Pseudomonas aeruginosa type III secretion. FEMS Microbiol Lett 256 (1):57-66. doi:10.1111/j.1574-6968.2005.00099.x

Brownlie JC, Cass BN, Riegler M, Witsenburg JJ, Iturbe-Ormaetxe I, McGraw EA, O'Neill SL (2009) Evidence for metabolic provisioning by a common invertebrate endosymbiont, Wolbachia pipientis, during periods of nutritional stress. PLoS Pathog 5 (4):e1000368. doi:10.1371/journal.ppat.1000368

Brownlie JC, O'Neill SL (2005) Wolbachia genomes: insights into an intracellular lifestyle. Curr Biol 15 (13):R507-509. doi:10.1016/j.cub.2005.06.029

Bull JJ, Rice WR (1991) Distinguishing mechanisms for the evolution of co-operation. J Theor Biol 149 (1):63-74

Campbell WC (1991) Ivermectin as an antiparasitic agent for use in humans. Annual Rev Microbiol 45:445–474

Carrington LB, Lipkowitz JR, Hoffmann AA, Turelli M (2011) A re-examination of Wolbachia-induced cytoplasmic incompatibility in California Drosophila simulans. PLoS One 6 (7):e22565. doi:10.1371/journal.pone.0022565

Carroll L (1871) Through the Looking-Glass, and What Alice Found There. In: Co M (ed).

Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28 (4):464-469. doi:10.1093/bioinformatics/btr703

Cascales E, Bernadac A, Gavioli M, Lazzaroni JC, Lloubes R (2002) Pal lipoprotein of Escherichia coli plays a major role in outer membrane integrity. J Bacteriol 184 (3):754-759

Cascales E, Christie PJ (2003) The versatile bacterial type IV secretion systems. Nat Rev Microbiol 1 (2):137-149. doi:10.1038/nrmicro753

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 Pt 1:93-103

Casiraghi M, Bain O, Guerrero R, Martin C, Pocacqua V, Gardner SL, Franceschi A, Bandi C (2004) Mapping the presence of Wolbachia pipientis on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. Int J Parasitol 34 (2):191-203. doi:10.1016/j.ijpara.2003.10.004

Casiraghi M, Bordenstein SR, Baldo L, Lo N, Beninati T, Wernegreen JJ, Werren JH, Bandi C (2005) 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 (Pt 12):4015-4022. doi:10.1099/mic.0.28313-0

Casiraghi M, Werren JH, Bazzocchi C, Biserni A, Bandi C (2003) dnaA gene sequences from Wolbachia pipientis support subdivision into supergroups and provide no evidence for recombination in the lineages infecting nematodes. Parassitologia 45 (1):13-18

Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R (2011) Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from Wolbachia endosymbionts. Genome Biol Evol 3:1175-1186. doi:10.1093/gbe/evr096

Cerveny L, Straskova A, Dankova V, Hartlova A, Ceckova M, Staud F, Stulik J (2013) Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. Infect Immun 81 (3):629-635. doi:10.1128/IAI.01035-12

Chambers EW, Hapairai L, Peel BA, Bossin H, Dobson SL (2011) Male mating competitiveness of a Wolbachia-introgressed Aedes polynesiensis strain under

semi-field conditions. PLoS Negl Trop Dis 5 (8):e1271. doi:10.1371/journal.pntd.0001271

Charlat S, Calmet C, Mercot H (2001) On the mod resc model and the evolution of Wolbachia compatibility types. Genetics 159 (4):1415-1422

Charles H, Ishikawa H (1999) Physical and genetic map of the genome of Buchnera, the primary endosymbiont of the pea aphid Acyrthosiphon pisum. J Mol Evol 48 (2):142-150

Charniaux-Cotton H (1954) [Discovery in, an amphipod crustacean (Orchestia gammarella) of an endocrine gland responsible for the differentiation of primary and secondary male sex characteristics]. C R Hebd Seances Acad Sci 239 (13):780-782

Charniaux-Cotton H (1960) Sex determination. In: T. H. Waterman E (ed) The physiology of Crustacea, vol 1.

Charniaux-Cotton H, G.G P, T G-V (1992) Arthropoda-Crustacea: Sexual Différenciation. In: K. G. Adiyodi and R. G. Adiyodi E (ed) Reproductive Biology of Invertebrates, vol 5. pp 281-323

Charniaux-Cotton H. PGGaG-VT (1992) Arthropoda-Crustacea: Sexual Différenciation. In: K. G. Adiyodi and R. G. Adiyodi E (ed) Reproductive Biology of Invertebrates, vol 5. pp 281-323

Chavali Gopal GG, Gopal Krishna, Shrinivas S. Jahageerdar, Morten Rye, Ben J. Hayes, Sivagnanam Paulpandi, Remanibhaskaran P. Kiran, Subramaniapillai M. Pillai, Pitchaiyappan Ravichandran, Alphis G. Ponniah, Dilip Kumar, Gopalapillay Gopikrishna GK, Shrinivas S. Jahageerdar, Morten Rye, Ben J. Hayes, Sivagnanam Paulpandi, Remanibhaskaran P. Kiran, Subramaniapillai M. Pillai, Pitchaiyappan Ravichandran, Alphis G. Ponniah, Dilip Kumar (2010) Weight and time of onset of female-superior sexual dimorphism in pond reared Penaeus monodon. Aquaculture 300 (1-4):237-239

Chereau D, Kerff F, Graceffa P, Grabarek Z, Langsetmo K, Dominguez R (2005) Actin-bound structures of Wiskott-Aldrich syndrome protein (WASP)-homology domain 2 and the implications for filament assembly. Proc Natl Acad Sci U S A 102 (46):16644-16649. doi:10.1073/pnas.0507021102

Chevalier F, Herbiniere-Gaboreau J, Bertaux J, Raimond M, Morel F, Bouchon D, Greve P, Braquart-Varnier C (2011) The immune cellular effectors of terrestrial isopod Armadillidium vulgare: meeting with their invaders, Wolbachia. PLoS One 6 (4):e18531. doi:10.1371/journal.pone.0018531

Cho KO, Kim GW, Lee OK (2011) Wolbachia bacteria reside in host Golgi-related vesicles whose position is regulated by polarity proteins. PLoS One 6 (7):e22703. doi:10.1371/journal.pone.0022703

Combes C (1995) Interactions Durables: Ecologie et Evolution du Parasitisme. Masson edn.,

Cook PE, McGraw EA (2010) Wolbachia pipientis: an expanding bag of tricks to explore for disease control. Trends Parasitol 26 (8):373-375. doi:10.1016/j.pt.2010.05.006

Cordaux R (2008) ISWpi1 from Wolbachia pipientis defines a novel group of insertion sequences within the IS5 family. Gene 409 (1-2):20-27. doi:10.1016/j.gene.2007.10.035

Cordaux R (2009) Gene conversion maintains nonfunctional transposable elements in an obligate mutualistic endosymbiont. Mol Biol Evol 26 (8):1679-1682. doi:10.1093/molbev/msp093

Cordaux R, Bouchon D, Greve P (2011) The impact of endosymbionts on the evolution of host sex-determination mechanisms. Trends Genet 27 (8):332-341. doi:10.1016/j.tig.2011.05.002

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

Cordaux R, Michel-Salzat A, Frelon-Raimond M, Rigaud T, Bouchon D (2004) Evidence for a new feminizing Wolbachia strain in the isopod Armadillidium vulgare: evolutionary implications. Heredity (Edinb) 93 (1):78-84. doi:10.1038/sj.hdy.6800482

Cordaux R, Pichon S, Hatira HB, Doublet V, Greve P, Marcade I, Braquart-Varnier C, Souty-Grosset C, Charfi-Cheikhrouha F, Bouchon D (2012) Widespread Wolbachia infection in terrestrial isopods and other crustaceans. Zookeys (176):123-131. doi:10.3897/zookeys.176.2284

Cordaux R, Pichon S, Ling A, Perez P, Delaunay C, Vavre F, Bouchon D, Greve P (2008) Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont. Mol Biol Evol 25 (9):1889-1896. doi:10.1093/molbev/msn134

Coupland P, Chandra T, Quail M, Reik W, Swerdlow H (2012) Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation. Biotechniques 53 (6):365-372. doi:10.2144/000113962

Cox FE (2002) History of human parasitology. Clin Microbiol Rev 15 (4):595-612

Cross HF, Haarbrink M, Egerton G, Yazdanbakhsh M, Taylor MJ (2001) Severe reactions to filarial chemotherapy and release of Wolbachia endosymbionts into blood. Lancet 358 (9296):1873-1875. doi:10.1016/S0140-6736(01)06899-4

D'Andrea LD, Regan L (2003) TPR proteins: the versatile helix. Trends Biochem Sci 28 (12):655-662. doi:10.1016/j.tibs.2003.10.007

Dagan T (2011) Phylogenomic networks. Trends Microbiol 19 (10):483-491. doi:10.1016/j.tim.2011.07.001

Darby AC, Armstrong SD, Bah GS, Kaur G, Hughes MA, Kay SM, Koldkjaer P, Rainbow L, Radford AD, Blaxter ML, Tanya VN, Trees AJ, Cordaux R, Wastling JM, Makepeace BL (2012) Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res 22 (12):2467-2477. doi:10.1101/gr.138420.112

Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5 (6):e11147. doi:10.1371/journal.pone.0011147

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27 (8):1164-1165. doi:10.1093/bioinformatics/btr088

de Bary A (1879) Die Erscheinung der Symbiose. Verlag von Karl J. Trubner: Strassburg.,

De Palmenaer D, Siguier P, Mahillon J (2008) IS4 family goes genomic. BMC Evol Biol 8:18. doi:10.1186/1471-2148-8-18

Dean P (2011) Functional domains and motifs of bacterial type III effector proteins and their roles in infection. FEMS Microbiol Rev 35 (6):1100-1125. doi:10.1111/j.1574-6976.2011.00271.x

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 (Edinb) 95 (5):394-400. doi:10.1038/sj.hdy.6800739

Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Bouletreau M (2001) Removing symbiotic Wolbachia bacteria specifically inhibits oogenesis in a parasitic wasp. Proc Natl Acad Sci U S A 98 (11):6247-6252. doi:10.1073/pnas.101304298

Dittmar K, Whiting MF (2004) New Wolbachia endosymbionts from Nearctic and Neotropical fleas (Siphonaptera). J Parasitol 90 (5):953-957. doi:10.1645/GE-186R

Dobson SL, Bourtzis K, Braig HR, Jones BF, Zhou W, Rousset F, O'Neill SL (1999) Wolbachia infections are distributed throughout insect somatic and germ line tissues. Insect Biochem Mol Biol 29 (2):153-160

Dobson SL, Rattanadechakul W, Marsland EJ (2004) Fitness advantage and cytoplasmic incompatibility in Wolbachia single- and superinfected Aedes albopictus. Heredity (Edinb) 93 (2):135-142. doi:10.1038/sj.hdy.6800458

Doublet V, Souty-Grosset C, Bouchon D, Cordaux R, Marcadé I (2008) A Thirty Million Year-Old Inherited Heteroplasmy. PLoS One 3 (8):e2938

du Plessis DJ, Nouwen N, Driessen AJ (2011) The Sec translocase. Biochim Biophys Acta 1808 (3):851-865. doi:10.1016/j.bbamem.2010.08.016

Dunning Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, Munoz Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, Ingram J, Nene RV, Shepard J, Tomkins J, Richards S, Spiro DJ, Ghedin E, Slatko BE, Tettelin H, Werren JH (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317 (5845):1753-1756. doi:10.1126/science.1142490

Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen JA, Seshadri R, Ren Q, Wu M, Utterback TR, Smith S, Lewis M, Khouri H, Zhang C, Niu H, Lin Q, Ohashi N, Zhi N, Nelson W, Brinkac LM, Dodson RJ, Rosovitz MJ, Sundaram J, Daugherty SC, Davidsen T, Durkin AS, Gwinn M, Haft DH, Selengut JD, Sullivan SA, Zafar N, Zhou L, Benahmed F, Forberger H, Halpin R, Mulligan S, Robinson J, White O, Rikihisa Y, Tettelin H (2006) Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet 2 (2):e21. doi:10.1371/journal.pgen.0020021

Duplouy A, Iturbe-Ormaetxe I, Beatson SA, Szubert JM, Brownlie JC, McMeniman CJ, McGraw EA, Hurst GD, Charlat S, O'Neill SL, Woolfit M (2013) Draft genome sequence of the male-killing Wolbachia strain wBol1 reveals recent horizontal gene transfers from diverse sources. BMC Genomics 14:20. doi:10.1186/1471-2164-14-20

Durand E, Verger D, Rego AT, Chandran V, Meng G, Fronzes R, Waksman G (2009) Structural biology of bacterial secretion systems in gram-negative pathogens--potential for new drug targets. Infect Disord Drug Targets 9 (5):518-547

Duron O (2013) Lateral transfers of insertion sequences between Wolbachia, Cardinium and Rickettsia bacterial endosymbionts. Heredity (Edinb). doi:10.1038/hdy.2013.56

Eddy SR (2009) A new generation of homology search tools based on probabilistic inference. Genome Inform 23 (1):205-211

Eddy SR (2011) Accelerated Profile HMM Searches. PLoS Comput Biol 7 (10):e1002195. doi:10.1371/journal.pcbi.1002195

Edqvist PJ, Broms JE, Betts HJ, Forsberg A, Pallen MJ, Francis MS (2006) Tetratricopeptide repeats in the type III secretion chaperone, LcrH: their role in substrate binding and secretion. Mol Microbiol 59 (1):31-44. doi:10.1111/j.1365-2958.2005.04923.x

Ellegaard KM, Klasson L, Naslund K, Bourtzis K, Andersson SG (2013) Comparative genomics of Wolbachia and the bacterial species concept. PLoS Genet 9 (4):e1003381. doi:10.1371/journal.pgen.1003381

Engelstadter J, Telschow A (2009) Cytoplasmic incompatibility and host population structure. Heredity (Edinb) 103 (3):196-207. doi:10.1038/hdy.2009.53

English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One 7 (11):e47768. doi:10.1371/journal.pone.0047768

Felix C, Pichon S, Braquart-Varnier C, Braig H, Chen L, Garrett RA, Martin G, Greve P (2008) Characterization and transcriptional analysis of two gene clusters for type IV secretion machinery in Wolbachia of Armadillidium vulgare. Res Microbiol 159 (6):481-485. doi:10.1016/j.resmic.2008.05.007

Fellous S, Salvaudon L (2009) How can your parasites become your allies? Trends Parasitol 25 (2):62-66. doi:10.1016/j.pt.2008.11.010

Fenn K, Blaxter M (2004a) Are filarial nematode Wolbachia obligate mutualist symbionts? Trends Ecol Evol 19 (4):163-166. doi:10.1016/j.tree.2004.01.002

Fenn K, Blaxter M (2004b) Quantification of Wolbachia bacteria in Brugia malayi through the nematode lifecycle. Mol Biochem Parasitol 137 (2):361-364. doi:10.1016/j.molbiopara.2004.06.012

Fischer K, Beatty WL, Jiang D, Weil GJ, Fischer PU (2011) Tissue and stage-specific distribution of Wolbachia in Brugia malayi. PLoS Negl Trop Dis 5 (5):e1174. doi:10.1371/journal.pntd.0001174

Fischer P, Bonow I, Supali T, Ruckert P, Rahmah N (2005) Detection of filaria-specific IgG4 antibodies and filarial DNA, for the screening of blood spots for Brugia timori. Ann Trop Med Parasitol 99 (1):53-60. doi:10.1179/136485905X13339

Fitch WM (2000) Homology a personal view on some of the problems. Trends Genet 16 (5):227-231

Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B (2005) The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3 (4):e121. doi:10.1371/journal.pbio.0030121

Francis H, Awadzi K, Ottesen EA (1985) The Mazzotti reaction following treatment of onchocerciasis with diethylcarbamazine: clinical severity as a function of infection intensity. Am J Trop Med Hyg 34 (3):529-536

Frank AB (1877) Ueber die biologischen Verhaltnisse des Thallus einiger Krusten-flechten. Beitrage zur Biologie der Pflanzen 2:123–200

Fry AJ, Palmer MR, Rand DM (2004) Variable fitness effects of Wolbachia infection in Drosophila melanogaster. Heredity (Edinb) 93 (4):379-389. doi:10.1038/sj.hdy.6800514

Genchi C, Sacchi L, Bandi C, Venco L (1998) Preliminary results on the effect of tetracycline on the embryogenesis and symbiotic bacteria (Wolbachia) of Dirofilaria immitis. An update and discussion. Parassitologia 40 (3):247-249

Geniez S, Foster JM, Kumar S, Moumen B, Leproust E, Hardy O, Guadalupe M, Thomas SJ, Boone B, Hendrickson C, Bouchon D, Greve P, Slatko BE (2012) Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA. Symbiosis 58 (1-3):201-207. doi:10.1007/s13199-012-0215-x

Genty L, Bouchon D, Raimond M, Bertaux J (2013) Wolbachia infect ovaries in the course their maturation: Last minute passengers or priority travellers?

Ghignone S, Salvioli A, Anca I, Lumini E, Ortu G, Petiti L, Cruveiller S, Bianciotto V, Piffanelli P, Lanfranco L, Bonfante P (2012) The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. Isme J 6 (1):136-145. doi:10.1038/ismej.2011.110

Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, Shukla M, Walenz B, Hill CA, Nene VM, Azad AF, Sobral BW, Caler E (2012) A Rickettsia genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. J Bacteriol 194 (2):376-394. doi:10.1128/JB.06244-11

Giordano R, Jackson JJ, Robertson HM (1997) The role of Wolbachia bacteria in reproductive incompatibilities and hybrid zones of Diabrotica beetles and Gryllus crickets. Proc Natl Acad Sci U S A 94 (21):11439-11444

Gomez-Valero L, Rusniok C, Cazalet C, Buchrieser C (2011) Comparative and functional genomics of legionella identified eukaryotic like proteins as key players in host-pathogen interactions. Front Microbiol 2:208. doi:10.3389/fmicb.2011.00208

Gorham CH, Fang QQ, Durden LA (2003) Wolbachia endosymbionts in fleas (Siphonaptera). J Parasitol 89 (2):283-289. doi:10.1645/0022-3395(2003)089[0283:WEIFS]2.0.CO;2

Gottlieb Y, Ghanim M, Gueguen G, Kontsedalov S, Vavre F, Fleury F, Zchori-Fein E (2008) Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies. Faseb J 22 (7):2591-2599. doi:10.1096/fj.07-101162

Gribskov M, McLachlan AD, Eisenberg D (1987) Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci U S A 84 (13):4355-4358

Gulland A (2013) Donated drugs boost protection against neglected tropical diseases. Bmj 346:f351. doi:10.1136/bmj.f351

Habyarimana F, Al-Khodor S, Kalia A, Graham JE, Price CT, Garcia MT, Kwaik YA (2008) Role for the Ankyrin eukaryotic-like genes of Legionella pneumophila in parasitism of protozoan hosts and human macrophages. Environ Microbiol 10 (6):1460-1474. doi:10.1111/j.1462-2920.2007.01560.x

Haegeman A, Vanholme B, Jacob J, Vandekerckhove TT, Claeys M, Borgonie G, Gheysen G (2009) An endosymbiotic bacterium in a plant-parasitic nematode: member of a new Wolbachia supergroup. Int J Parasitol 39 (9):1045-1054. doi:10.1016/j.ijpara.2009.01.006

Hansen AK, Vorburger C, Moran NA (2012) Genomic basis of endosymbiont-conferred protection against an insect parasitoid. Genome Res 22 (1):106-114. doi:10.1101/gr.125351.111

Hasegawa H, E. H, Y K (1993) Homronal control of sexual differentiation and reproduction in Crustacea. Am Zool 33:403-411

Hasegawa Y, Haino-Fukushima K, Katakura Y (1987) Isolation and properties of androgenic gland hormone from the terrestrial isopod, Armadillidium vulgare. Gen Comp Endocrinol 67 (1):101-110

Heath BD, Butcher RD, Whitfield WG, Hubbard SF (1999) Horizontal transfer of Wolbachia between phylogenetically distant insect species by a naturally occurring mechanism. Curr Biol 9 (6):313-316

Hedges LM, Brownlie JC, O'Neill SL, Johnson KN (2008) Wolbachia and virus protection in insects. Science 322 (5902):702. doi:10.1126/science.1162418

Hertig M (1936) Hollow-Ground Slides for Whole Mounts Made with the Dental Engine. Science 83 (2144):110. doi:10.1126/science.83.2144.110

Hertig M, Wolbach SB (1924) Studies on Rickettsia-Like Micro-Organisms in Insects. J Med Res 44 (3):329-374 327

Hicks SW, Galan JE (2013) Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. Nat Rev Microbiol 11 (5):316-326. doi:10.1038/nrmicro3009

Hirano T, Kinoshita N, Morikawa K, Yanagida M (1990) Snap helix with knob and hole: essential repeats in S. pombe nuclear protein nuc2+. Cell 60 (2):319-328

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 (4):167-170

Hiroki M, Tagami Y, Miura K, Kato Y (2004) Multiple infection with Wolbachia inducing different reproductive manipulations in the butterfly Eurema hecabe. Proc Biol Sci 271 (1549):1751-1755. doi:10.1098/rspb.2004.2769

Hoerauf A (2006) New strategies to combat filariasis. Expert Rev Anti Infect Ther 4 (2):211-222. doi:10.1586/14787210.4.2.211

Hoerauf A, Mand S, Volkmann L, Buttner M, Marfo-Debrekyei Y, Taylor M, Adjei O, Buttner DW (2003) Doxycycline in the treatment of human onchocerciasis: Kinetics of Wolbachia endobacteria reduction and of inhibition of embryogenesis in female Onchocerca worms. Microbes Infect 5 (4):261-273

Hoerauf A, Nissen-Pahle K, Schmetz C, Henkle-Duhrsen K, Blaxter ML, Buttner DW, Gallin MY, Al-Qaoud KM, Lucius R, Fleischer B (1999) Tetracycline therapy targets intracellular bacteria in the filarial nematode Litomosoides sigmodontis and results in filarial infertility. J Clin Invest 103 (1):11-18

Hoerauf A, Volkmann L, Hamelmann C, Adjei O, Autenrieth IB, Fleischer B, Buttner DW (2000a) Endosymbiotic bacteria in worms as targets for a novel chemotherapy in filariasis. Lancet 355 (9211):1242-1243. doi:10.1016/S0140-6736(00)02095-X

Hoerauf A, Volkmann L, Nissen-Paehle K, Schmetz C, Autenrieth I, Buttner DW, Fleischer B (2000b) Targeting of Wolbachia endobacteria in Litomosoides sigmodontis: comparison of tetracyclines with chloramphenicol, macrolides and ciprofloxacin. Trop Med Int Health 5 (4):275-279

Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaetxe I, Johnson PH, Muzzi F, Greenfield M, Durkan M, Leong YS, Dong Y, Cook H, Axford J, Callahan AG, Kenny N, Omodei C, McGraw EA, Ryan PA, Ritchie SA, Turelli M, O'Neill SL (2011) Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. Nature 476 (7361):454-457. doi:10.1038/nature10356

Hoffmann D (1969) The development of androgenic glands of protentric schrimp. Biol Bull 137:286-296

Holman AG, Davis PJ, Foster JM, Carlow CK, Kumar S (2009) Computational prediction of essential genes in an unculturable endosymbiotic bacterium, Wolbachia of Brugia malayi. BMC Microbiol 9:243. doi:10.1186/1471-2180-9-243

Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T (2010) Wolbachia as a bacteriocyte-associated nutritional mutualist. Proc Natl Acad Sci U S A 107 (2):769-774. doi:10.1073/pnas.0911476107

Hoti SL, Vasuki V, Lizotte MW, Patra KP, Ravi G, Vanamail P, Manonmani A, Sabesan S, Krishnamoorthy K, Williams SA (2001) Detection of Brugia malayi in laboratory and wild-caught Mansonioides mosquitoes (Diptera: Culicidae) using Hha I PCR assay. Bull Entomol Res 91 (2):87-92

Hougard JM, Alley ES, Yameogo L, Dadzie KY, Boatin BA (2001) Eliminating onchocerciasis after 14 years of vector control: a proved strategy. J Infect Dis 184 (4):497-503. doi:10.1086/322789

Huigens ME, de Almeida RP, Boons PA, Luck RF, Stouthamer R (2004) Natural interspecific and intraspecific horizontal transfer of parthenogenesis-inducing Wolbachia in Trichogramma wasps. Proc Biol Sci 271 (1538):509-515. doi:10.1098/rspb.2003.2640

Hurst GDD, Jiggins FM (2000) Male-killing bacteria in insects: Mechanisms, incidence, and implications. Emerg Infect Dis 6 (4):329-336

Hurst LD, Randerson JP (2002) Parasitic sex puppeteers. Scientific American 286 (4):56-61

Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson AC, von Dohlen CD, Fukatsu T, McCutcheon JP (2013) Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153 (7):1567-1578. doi:10.1016/j.cell.2013.05.040

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. doi:10.1186/1471-2105-11-119

Ijichi N, Kondo N, Matsumoto R, Shimada M, Ishikawa H, Fukatsu T (2002) Internal spatiotemporal population dynamics of infection with three Wolbachia strains in the adzuki bean beetle, Callosobruchus chinensis (Coleoptera: Bruchidae). Appl Environ Microbiol 68 (8):4074-4080

Ironside JE, Smith JE, Hatcher MJ, Dunn AM (2011) Should sex-ratio distorting parasites abandon horizontal transmission? BMC Evol Biol 11:370. doi:10.1186/1471-2148-11-370

Iturbe-Ormaetxe I, Woolfit M, Rances E, Duplouy A, O'Neill SL (2011) A simple protocol to obtain highly pure Wolbachia endosymbiont DNA for genome sequencing. J Microbiol Methods 84 (1):134-136. doi:10.1016/j.mimet.2010.10.019

Jiggins FM, Hurst GD, Jiggins CD, v d Schulenburg JH, Majerus ME (2000) The butterfly Danaus chrysippus is infected by a male-killing Spiroplasma bacterium. Parasitology 120 (Pt 5):439-446

Johnson LS, Eddy SR, Portugaly E (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics 11:431. doi:10.1186/1471-2105-11-431

Juchault P (1977) Corrélation entre différentiation sexuelle mâle externe et fonction andro-inhibitrice de la région médiane du protocérébron chez le Crustacé Sphaeroma serratum Fabr. (Isopoda, Flabellifera). Comptes Rendus de l'Académie des Sciences Paris 285:179-182

Juchault P, Legrand JJ (1964) Mise en évidence d'un inducteur sexuel mâle distinct de l'hormone adulte et contribution à l'étude de l'auto-différentiation ovarienne chez l'Oniscoïde Helleria brevicornis. CR Acad Sc Paris 258:2416-2419

Juchault P, Legrand JJ (1966) Contribution à l'étude expérimentale de l'inducteur sexuel mâle chez les Crustacés Isopodes terrestres. Annales d'Endocrinologie Paris 27 (3):205-209

Juchault P, Mocquard JP (1993) Transfer of a parasitic sex factor to the nuclear genome of the host: A hypothesis on the evolution of sex-determining mechanisms in the terrestrial isopod Armadillidium vulgare Latr. J Evol Biol 6:511-528

Kambris Z, Cook PE, Phuc HK, Sinkins SP (2009) Immune activation by life-shortening Wolbachia and reduced filarial competence in mosquitoes. Science 326 (5949):134-136. doi:10.1126/science.1177531

Katakura Y (1961) Hormonal control of development of sexual characters in the isopod crustacean, Armadillidium vulgare. Annot Zool Japon 34:60-71

Katakura Y (1989) Endocrine and genetic control of sex differentiation in the malacostacan Crustacea. Invertebr Reprod Dev 16:177-182

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30 (14):3059-3066

Kent BN, Bordenstein SR (2010) Phage WO of Wolbachia: lambda of the endosymbiont world. Trends Microbiol 18 (4):173-181. doi:10.1016/j.tim.2009.12.011

Kent BN, Funkhouser LJ, Setia S, Bordenstein SR (2011a) Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (Wolbachia). PLoS One 6 (9):e24984. doi:10.1371/journal.pone.0024984

Kent BN, Salichos L, Gibbons JG, Rokas A, Newton IL, Clark ME, Bordenstein SR (2011b) Complete bacteriophage transfer in a bacterial endosymbiont (Wolbachia) determined by targeted genome capture. Genome Biol Evol 3:209-218. doi:10.1093/gbe/evr007

Kerketta AS, Babu BV, Rath K, Jangid PK, Nayak AN, Kar SK (2005) A randomized clinical trial to compare the efficacy of three treatment regimens along with footcare in the morbidity management of filarial lymphoedema. Trop Med Int Health 10 (7):698-705. doi:10.1111/j.1365-3156.2005.01442.x

Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, Sinkins SP, Parkhill J (2008) Genome evolution of Wolbachia strain wPip from the Culex pipiens group. Mol Biol Evol 25 (9):1877-1887. doi:10.1093/molbev/msn133

Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SG (2009) The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans. Proc Natl Acad Sci U S A 106 (14):5725-5730. doi:10.1073/pnas.0810753106

Kluber S, Supali T, Williams SA, Liebau E, Fischer P (2001) Rapid PCR-based detection of Brugia malayi DNA from blood spots by DNA Detection Test Strips. Trans R Soc Trop Med Hyg 95 (2):169-170

Kondo N, Ijichi N, Shimada M, Fukatsu T (2002a) Prevailing triple infection with Wolbachia in Callosobruchus chinensis (Coleoptera: Bruchidae). Mol Ecol 11 (2):167-180

Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T (2002b) Genome fragment of Wolbachia endosymbiont transferred to X chromosome of host insect. Proc Natl Acad Sci U S A 99 (22):14280-14285. doi:10.1073/pnas.222228199

Kozek WJ (2005) What is new in the Wolbachia/Dirofilaria interaction? Vet Parasitol 133 (2-3):127-132. doi:10.1016/j.vetpar.2005.02.005

Kremer N, Charif D, Henri H, Bataille M, Prevost G, Kraaijeveld K, Vavre F (2009) A new case of Wolbachia dependence in the genus Asobara: evidence for parthenogenesis induction in Asobara japonica. Heredity (Edinb) 103 (3):248-256. doi:10.1038/hdy.2009.63

Kremer N, Dedeine F, Charif D, Finet C, Allemand R, Vavre F (2010) Do variable compensatory mechanisms explain the polymorphism of the dependence phenotype in the Asobara tabida-wolbachia association? Evolution 64 (10):2969-2979. doi:10.1111/j.1558-5646.2010.01034.x

Kumar S, Blaxter ML (2011) Simultaneous genome sequencing of symbionts and their hosts. Symbiosis 55 (3):119-126. doi:10.1007/s13199-012-0154-6

Landmann F, Foster JM, Slatko B, Sullivan W (2010) Asymmetric Wolbachia segregation during early Brugia malayi embryogenesis determines its distribution in adult host tissues. PLoS Negl Trop Dis 4 (7):e758. doi:10.1371/journal.pntd.0000758

Le Clec'h W, Braquart-Varnier C, Raimond M, Ferdy JB, Bouchon D, Sicard M (2012) High virulence of Wolbachia after host switching: when autophagy hurts. PLoS Pathog 8 (8):e1002844. doi:10.1371/journal.ppat.1002844

Le Clec'h W, Chevalier FD, Genty L, Bertaux J, Bouchon D, Sicard M (2013a) Cannibalism and predation as paths for horizontal passage of Wolbachia between terrestrial isopods. PLoS One 8 (4):e60232. doi:10.1371/journal.pone.0060232

Le Clec'h W, Raimond M, Guillot S, Bouchon D, Sicard M (2013b) Horizontal transfers of feminizing versus non-feminizing Wolbachia strains: from harmless passengers to pathogens. Environ Microbiol. doi:10.1111/1462-2920.12172

Leclercq S, Giraud I, Cordaux R (2011) Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. Mol Biol Evol 28 (1):685-697. doi:10.1093/molbev/msq238

Lee T.H. YMaYF (1994) Sex differentiation in the crab Eriocheir japonicus (Decapoda, Grapsidae). Invertebr Reprod Dev 25:123-138

Leggat PA, Melrose W, Durrheim DN (2004) Could it be lymphatic filariasis? J Travel Med 11 (1):56-60

Legrand JJ, Juchault P (1984) Nouvelles données sur le déterminisme génétique et épigénétique de la monogénie chez le crustacés isopodes terrestres Armadillidium vulgare Latr. Génét Sél Evol 16 (1):57-84

Legrand JJ, Martin G, Artault JC (1978) Corrélation entre la présence d'un symbionte bactérien dans les ovocytes de Porcellio dilatatus petiti, et la stérilité du croisement P.d. petiti mâle x P. d. dilatatus femelle. Archives de l'Institut Pasteur de Tunis 55 (4):507-514

Leigh EG, Jr. (2010) The evolution of mutualism. J Evol Biol 23 (12):2507-2528. doi:10.1111/j.1420-9101.2010.02114.x

Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40 (Database issue):D302-305. doi:10.1093/nar/gkr931

Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13 (9):2178-2189. doi:10.1101/gr.1224503

Li Z, Carlow CK (2012) Characterization of transcription factors that regulate the type IV secretion system and riboflavin biosynthesis in Wolbachia of Brugia malayi. PLoS One 7 (12):e51597. doi:10.1371/journal.pone.0051597

Li Z, Garner AL, Gloeckner C, Janda KD, Carlow CK (2011) Targeting the Wolbachia cell division protein FtsZ as a new approach for antifilarial therapy. PLoS Negl Trop Dis 5 (11):e1411. doi:10.1371/journal.pntd.0001411

Lin M, den Dulk-Ras A, Hooykaas PJ, Rikihisa Y (2007) Anaplasma phagocytophilum AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. Cell Microbiol 9 (11):2644-2657. doi:10.1111/j.1462-5822.2007.00985.x

Lindner MS, Renard BY (2013) Metagenomic abundance estimation and diagnostic testing on species level. Nucleic Acids Res 41 (1):e10. doi:10.1093/nar/gks803

Ling A, Cordaux R (2010) Insertion sequence inversions mediated by ectopic recombination between terminal inverted repeats. PLoS One 5 (12):e15654. doi:10.1371/journal.pone.0015654

Liu C, Chen L, Pichon S, Felix C, Cerveau N, Leclercq S, Lesobre J, Moumen B, Bourtzis K, Cordaux R, Bouchon D, Greve P, Garrett R (2013) Genome sequence of the feminizing Wolbachia wVulC from the isopod crustacean Armadillidium vulgare identifies multiple secretion/export systems and putative effectors in a bacterial endosymbiont. in prep

Lizotte MR, Supali T, Partono F, Williams SA (1994) A polymerase chain reaction assay for the detection of Brugia malayi in blood. Am J Trop Med Hyg 51 (3):314-321

Llamas MA, Rodriguez-Herva JJ, Hancock RE, Bitter W, Tommassen J, Ramos JL (2003) Role of Pseudomonas putida tol-oprL gene products in uptake of solutes through the cytoplasmic membrane. J Bacteriol 185 (16):4707-4716

Lo N, Casiraghi M, Salati E, Bazzocchi C, Bandi C (2002) How many wolbachia supergroups exist? Mol Biol Evol 19 (3):341-346

Lo N, Evans TA (2007) Phylogenetic diversity of the intracellular symbiont Wolbachia in termites. Mol Phylogenet Evol 44 (1):461-466. doi:10.1016/j.ympev.2006.10.028

Login FH, Balmand S, Vallier A, Vincent-Monegat C, Vigneron A, Weiss-Gayet M, Rochat D, Heddi A (2011) Antimicrobial peptides keep insect endosymbionts under control. Science 334 (6054):362-365. doi:10.1126/science.1209728

Login FH, Heddi A (2013) Insect immune system maintains long-term resident bacteria through a local response. J Insect Physiol 59 (2):232-239. doi:10.1016/j.jinsphys.2012.06.015

Loman NJ, Constantinidou C, Chan JZ, Halachev M, Sergeant M, Penn CW, Robinson ER, Pallen MJ (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat Rev Microbiol 10 (9):599-606. doi:10.1038/nrmicro2850

Lombardo MJ, Rosenberg SM (2000) radC102 of Escherichia coli is an allele of recG. J Bacteriol 182 (22):6287-6291

Luhrmann A, Nogueira CV, Carey KL, Roy CR (2010) Inhibition of pathogen-induced apoptosis by a Coxiella burnetii type IV effector protein. Proc Natl Acad Sci U S A 107 (44):18997-19001. doi:10.1073/pnas.1004380107

Lustigman S, Prichard RK, Gazzinelli A, Grant WN, Boatin BA, McCarthy JS, Basanez MG (2012) A research agenda for helminth diseases of humans: the problem of helminthiases. PLoS Negl Trop Dis 6 (4):e1582. doi:10.1371/journal.pntd.0001582

Machesky LM, Insall RH, Volkman LE (2001) WASP homology sequences in baculoviruses. Trends Cell Biol 11 (7):286-287

Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Nardi-Dei V, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G (2005) Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science 309 (5731):148-150. doi:10.1126/science.1109869

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, Turner DJ (2010) Target-enrichment strategies for next-generation sequencing. Nat Methods 7 (2):111-118. doi:10.1038/nmeth.1419

Martin G, Juchault P (1999) Androgenic hormone specificity in terrestrial isopods (Oniscidea): systematic involvements. J Crust Biol 19 (4):684-689

Martin G, Juchault P, Legrand JJ (1973) Mise en évidence d'un micro-organisme intracytoplasmique symbiote de l'Oniscoide Armadillidium vulgare L. dont la présence accompagne l'intersexualité ou la féminisation totale des mâles génétiques de la lignée thélygène. Comptes Rendus de l'Académie des Sciences Paris 276:2313-2316

Martin G, Juchault P, Mocquard JP, Raimond R, Rigaud T, Souty-Grosset C (1990) Feminizing endocellular bacteria in the terrestrial Armadillidium vulgare Latr. (Crustacea, Oniscidea). Endocytobiology IV Ed P NARDON et al, INRA Paris:403-408.

Martin G, Sorokine O, Moniatte M, Bulet C, Hetru C, Van Dorsselaer A (1999) The structure of a glycosylated protein hormone responsible for sex determination in the isopod, Armadillidium vulgare. Eur J Biochem 262:727-736

Masui S, Sasaki T, Ishikawa H (2000) Genes for the type IV secretion system in an intracellular symbiont, Wolbachia, a causative agent of various sexual alterations in arthropods. J Bacteriol 182 (22):6529-6531

Mavingui P, Valiente Moro C, Tran-Van V, Wisniewski-Dye F, Raquin V, Minard G, Tran FH, Voronin D, Rouy Z, Bustos P, Lozano L, Barbe V, Gonzalez V (2012) Whole-genome sequence of Wolbachia strain wAlbB, an endosymbiont of tiger mosquito vector Aedes albopictus. J Bacteriol 194 (7):1840. doi:10.1128/JB.00036-12

Mavingui P, Van VT, Labeyrie E, Rances E, Vavre F, Simonet P (2005) Efficient procedure for purification of obligate intracellular Wolbachia pipientis and representative amplification of its genome by multiple-displacement amplification. Appl Environ Microbiol 71 (11):6910-6917. doi:10.1128/AEM.71.11.6910-6917.2005

McGarry HF, Egerton GL, Taylor MJ (2004) Population dynamics of Wolbachia bacterial endosymbionts in Brugia malayi. Mol Biochem Parasitol 135 (1):57-67

McNulty SN, Abubucker S, Simon GM, Mitreva M, McNulty NP, Fischer K, Curtis KC, Brattig NW, Weil GJ, Fischer PU (2012) Transcriptomic and proteomic analyses of a Wolbachia-free filarial parasite provide evidence of trans-kingdom horizontal gene transfer. PLoS One 7 (9):e45777. doi:10.1371/journal.pone.0045777

Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. Curr Opin Genet Dev 15 (6):589-594. doi:10.1016/j.gde.2005.09.006

Melrose WD (2002) Lymphatic filariasis: new insights into an old disease. Int J Parasitol 32 (8):947-960

Metcalf JA, Bordenstein SR (2012) The complexity of virus systems: the case of endosymbionts. Curr Opin Microbiol 15 (4):546-552. doi:10.1016/j.mib.2012.04.010

Michel-Salzat A, Cordaux R, Bouchon D (2001) Wolbachia diversity in the Porcellionides pruinosus complex of species (Crustacea: Oniscidea): evidence for host-dependent patterns of infection. Heredity (Edinb) 87 (Pt 4):428-434

Min KT, Benzer S (1997) Wolbachia, normally a symbiont of Drosophila, can be virulent, causing degeneration and early death. Proc Natl Acad Sci U S A 94 (20):10792-10796

Mira A, Martin-Cuadrado AB, D'Auria G, Rodriguez-Valera F (2010) The bacterial pan-genome:a new paradigm in microbiology. Int Microbiol 13 (2):45-57

Moran N (2001) The process of genome shrinkage in the obligate symbiont
Buchnera aphidicola. Genome Biol 2 (12):12

Moran NA (2002) Microbial minimalism: genome reduction in bacterial pathogens. Cell 108 (5):583-586

Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, Rocha BC, Hall-Mendelin S, Day A, Riegler M, Hugo LE, Johnson KN, Kay BH, McGraw EA, van den Hurk AF, Ryan PA, O'Neill SL (2009) A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium. Cell 139 (7):1268-1278. doi:10.1016/j.cell.2009.11.042

Moret Y, Juchault P, Rigaud T (2001) Wolbachia endosymbiont responsible for cytoplasmic incompatibility in a terrestrial crustacean: effects in natural and foreign hosts. Heredity 86 (3):325-332

Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004) The ankyrin repeat as molecular architecture for protein recognition. Protein Sci 13 (6):1435-1448. doi:10.1110/ps.03554604

Murfin KE, Dillman AR, Foster JM, Bulgheresi S, Slatko BE, Sternberg PW, Goodrich-Blair H (2012) Nematode-bacterium symbioses--cooperation and conflict revealed in the "omics" age. Biol Bull 223 (1):85-102

Muzzi A, Masignani V, Rappuoli R (2007) The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug Discov Today 12 (11-12):429-439. doi:10.1016/j.drudis.2007.04.008

Nadalin F, Vezzi F, Policriti A (2012) GapFiller: a de novo assembly approach to fill the gap within paired reads. BMC Bioinformatics 13 Suppl 14:S8. doi:10.1186/1471-2105-13-S14-S8

Narita S, Kageyama D, Nomura M, Fukatsu T (2007) Unexpected mechanism of symbiont-induced reversal of insect sex: feminizing Wolbachia continuously acts

on the butterfly Eurema hecabe during larval development. Appl Environ Microbiol 73 (13):4332-4341. doi:10.1128/AEM.00145-07

Negri I, Pellecchia M, Greve P, Daffonchio D, Bandi C, Alma A (2010) Sex and stripping: The key to the intimate relationship between Wolbachia and host? Commun Integr Biol 3 (2):110-115

Negri I, Pellecchia M, Mazzoglio PJ, Patetta A, Alma A (2006) Feminizing Wolbachia in Zyginidia pullula (Insecta, Hemiptera), a leafhopper with an XX/X0 sex-determination system. Proc Biol Sci 273 (1599):2409-2416. doi:10.1098/rspb.2006.3592

Nikoh N, Tanaka K, Shibata F, Kondo N, Hizume M, Shimada M, Fukatsu T (2008) Wolbachia genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. Genome Res 18 (2):272-280. doi:10.1101/gr.7144908

Noroes J, Dreyer G, Santos A, Mendes VG, Medeiros Z, Addiss D (1997) Assessment of the efficacy of diethylcarbamazine on adult Wuchereria bancrofti in vivo. Trans R Soc Trop Med Hyg 91 (1):78-81

O'Connor L, Plichart C, Sang AC, Brelsfoard CL, Bossin HC, Dobson SL (2012) Open release of male mosquitoes infected with a wolbachia biopesticide: field performance and infection containment. PLoS Negl Trop Dis 6 (11):e1797. doi:10.1371/journal.pntd.0001797

O'Neill S, Hoffmann A, Werren J (1997) Influential passengers: inherited microorganisms and arthropod reproduction. New York.

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. Proc Natl Acad Sci U S A 89 (7):2699-2702

Ohno S, Wolf U, Atkin NB (1968) Evolution from fish to mammals by gene duplication. Hereditas 59 (1):169-187

Okuno A, Hasegawa Y, Ohira T, Katakura Y, Nagasawa H (1999) Characterization and cDNA cloning of androgenic gland hormone of the terrestrial isopod Armadillidium vulgare. Biochem Biophys Res Commun 264 (2):419-423. doi:10.1006/bbrc.1999.1522

Osborne SE, Leong YS, O'Neill SL, Johnson KN (2009) Variation in antiviral protection mediated by different Wolbachia strains in Drosophila simulans. PLoS Pathog 5 (11):e1000656. doi:10.1371/journal.ppat.1000656

Pallen MJ, Francis MS, Futterer K (2003) Tetratricopeptide-like repeats in type-III-secretion chaperones and regulators. FEMS Microbiol Lett 223 (1):53-60

Pallen MJ, Wren BW (2007) Bacterial pathogenomics. Nature 449 (7164):835-842. doi:10.1038/nature06248

Pan X, Luhrmann A, Satoh A, Laskowski-Arce MA, Roy CR (2008) Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. Science 320 (5883):1651-1654. doi:10.1126/science.1158160

Pannebakker BA, Loppin B, Elemans CP, Humblot L, Vavre F (2007) Parasitic inhibition of cell death facilitates symbiosis. Proc Natl Acad Sci U S A 104 (1):213-215. doi:10.1073/pnas.0607845104

Papafotiou G, Oehler S, Savakis C, Bourtzis K (2011) Regulation of Wolbachia ankyrin domain encoding genes in Drosophila gonads. Res Microbiol 162 (8):764-772. doi:10.1016/j.resmic.2011.06.012

Paunola E, Mattila PK, Lappalainen P (2002) WH2 domain: a small, versatile adapter for actin monomers. FEBS Lett 513 (1):92-97

Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Muller A, Woyke T, Malfatti SA, Hunter MS, Horn M (2012) Comparative genomics suggests an independent origin of cytoplasmic incompatibility in Cardinium hertigii. PLoS Genet 8 (10):e1003012. doi:10.1371/journal.pgen.1003012

Pernier J, Orban J, Avvaru BS, Jegou A, Romet-Lemonne G, Guichard B, Carlier MF (2013) Dimeric WH2 domains in Vibrio VopF promote actin filament barbed-end uncapping and assisted elongation. Nat Struct Mol Biol. doi:10.1038/nsmb.2639

Pfarr KM, Hoerauf A (2007) A niche for Wolbachia. Trends Parasitol 23 (1):5-7. doi:10.1016/j.pt.2006.11.002

Pfarr KM, Hoerauf AM (2006) Antibiotics which target the Wolbachia endosymbionts of filarial parasites: a new strategy for control of filariasis and amelioration of pathology. Mini Rev Med Chem 6 (2):203-210

Pichon S, Bouchon D, Cordaux R, Chen L, Garrett RA, Greve P (2009) Conservation of the Type IV secretion system throughout Wolbachia evolution. Biochem Biophys Res Commun 385 (4):557-562. doi:10.1016/j.bbrc.2009.05.118

Pichon S, Bouchon D, Liu C, Chen L, Garrett RA, Greve P (2012) The expression of one ankyrin pk2 allele of the WO prophage is correlated with the Wolbachia feminizing effect in isopods. BMC Microbiol 12:55. doi:10.1186/1471-2180-12-55

Poole CB, Tanner NA, Zhang Y, Evans TC, Jr., Carlow CK (2012) Diagnosis of brugian filariasis by loop-mediated isothermal amplification. PLoS Negl Trop Dis 6 (12):e1948. doi:10.1371/journal.pntd.0001948

Popovici J, Moreira LA, Poinsignon A, Iturbe-Ormaetxe I, McNaughton D, O'Neill SL (2010) Assessing key safety concerns of a Wolbachia-based strategy to control

dengue transmission by Aedes mosquitoes. Mem Inst Oswaldo Cruz 105 (8):957-964

Prichard R (1994) Anthelmintic resistance. Vet Parasitol 54 (1-3):259-268

Prichard R (2001) Genetic variability following selection of Haemonchus contortus with anthelmintics. Trends Parasitol 17 (9):445-453

Rajan TV (2004) Relationship of anti-microbial activity of tetracyclines to their ability to block the L3 to L4 molt of the human filarial parasite Brugia malayi. Am J Trop Med Hyg 71 (1):24-28

Ramabu SS, Schneider DA, Brayton KA, Ueti MW, Graca T, Futse JE, Noh SM, Baszler TV, Palmer GH (2011) Expression of Anaplasma marginale ankyrin repeat-containing proteins during infection of the mammalian host and tick vector. Infect Immun 79 (7):2847-2855. doi:10.1128/IAI.05097-11

Ramzy RM, el-Setouhy M, Helmy H, Kandil AM, Ahmed ES, Farid HA, Faris R, Weil GJ (2002) The impact of single-dose diethylcarbamazine treatment of bancroftian filariasis in a low-endemicity setting in Egypt. Am J Trop Med Hyg 67 (2):196-200

Rances E, Voronin D, Tran-Van V, Mavingui P (2008) Genetic and functional characterization of the type IV secretion system in Wolbachia. J Bacteriol 190 (14):5020-5030. doi:10.1128/JB.00377-08

Rances E, Ye YH, Woolfit M, McGraw EA, O'Neill SL (2012) The relative importance of innate immune priming in Wolbachia-mediated dengue interference. PLoS Pathog 8 (2):e1002548. doi:10.1371/journal.ppat.1002548

Rao R, Well GJ (2002) In vitro effects of antibiotics on Brugia malayi worm survival and reproduction. J Parasitol 88 (3):605-611. doi:10.1645/0022-3395(2002)088[0605:IVEOAO]2.0.CO;2

Rao RU, Moussa H, Weil GJ (2002) Brugia malayi: effects of antibacterial agents on larval viability and development in vitro. Exp Parasitol 101 (1):77-81

Rao RU, Weil GJ, Fischer K, Supali T, Fischer P (2006) Detection of Brugia parasite DNA in human blood by real-time PCR. J Clin Microbiol 44 (11):3887-3893. doi:10.1128/JCM.00969-06

Raverdy S, Foster JM, Roopenian E, Carlow CK (2008) The Wolbachia endosymbiont of Brugia malayi has an active pyruvate phosphate dikinase. Mol Biochem Parasitol 160 (2):163-166. doi:10.1016/j.molbiopara.2008.04.014

Reuter M, Keller L (2003) High levels of multiple Wolbachia infection and recombination in the ant Formica exsecta. Mol Biol Evol 20 (5):748-753. doi:10.1093/molbev/msg082

Richards FO, Jr., Boatin B, Sauerbrey M, Seketeli A (2001) Control of onchocerciasis today: status and challenges. Trends Parasitol 17 (12):558-563

Richardson DHS (1999) War in the world of lichens: parasitism and symbiosis as exemplified by lichens and lichenicolous fungi. Mycol Res 103 (6):641–650

Rigaud T (1991) Déterminisme extrachromosomique du sexe chez Armadillidium vulgare Latr. (Crustacé, Isopode) : Modifications du sex-ratio par une bactérie endocytobiotique et conséquences sur l'évolution des génotypes sexuels dans les populations. PhD dissertation, Université François Rabelais de Tours, Tours, France

Rigaud T (1997a) Inherited microorganisms and sex determination of arthropod hosts. Influential passengers: inherited microorganisms and arthropod reproduction, pp 81-101

Rigaud T, Juchault P (1993) Conflict between feminizing sex ratio distorters and an autosomal masculinizing gene in the terrestrial isopod Armadillidium vulgare Latr. Genetics 133 (2):247-252

Rigaud T, Juchault, P. and Mocquard, J.P. (1997b) The evolution of sex determination in isopod crustaceans. 19 (5):409-416

Rigaud T, Moreau J, Juchault P (1999) Wolbachia infection in the terrestrial isopod Oniscus asellus: sex ratio distortion and effect on fecundity. Heredity (Edinb) 83 (4):469-475

Rigaud T, Pennings PS, Juchault P (2001) Wolbachia bacteria effects after experimental interspecific transfers in terrestrial isopods. J Invertebr Pathol 77 (4):251-257

Rigaud T, Rousset F (1996) What generates the diversity of Wolbachia-arthropod interactions? Biodiversity and Conservation 5:999-1013

Rikihisa Y, Lin M, Niu H, Cheng Z (2009) Type IV secretion system of Anaplasma phagocytophilum and Ehrlichia chaffeensis. Ann N Y Acad Sci 1166:106-111. doi:10.1111/j.1749-6632.2009.04527.x

Rousset F, Bouchon, D., Juchault, P., Pintureau, B. and Solignac, M. (1992) Molecular diversity of wolbachia, endosymbiont of arthropods. Endocytobiology:127-131

Rousset F, Solignac M (1995) Evolution of single and double Wolbachia symbioses during speciation in the Drosophila simulans complex. Proc Natl Acad Sci U S A 92 (14):6389-6393

Sagi A, Cohen D, Milner Y (1990) Effect of androgenic gland ablation on morphotypic differentiation and sexual characteristics of male freshwater prawns, Macrobrachium rosenbergii. Gen Comp Endocrinol 77 (1):15-22

Sagi A. SEaHI (1997) Sexual differentiation in decapod Crustaceans: Role of the androgenic gland. . Invertebr Repord Dev 31:55-61

Santos-Garcia D, Frances Peris-Bondia, Giuseppe D'Auria, Andrés Moya, Francisco J. Silva, Amparo Latorre (2012) Enrichment of insect samples with bacterial symbionts using flow cytometry. ABSTRACT 7th International Symbiosis Society Congress

Saridaki A, Bourtzis K (2010) Wolbachia: more than just a bug in insects genitals. Curr Opin Microbiol 13 (1):67-72. doi:10.1016/j.mib.2009.11.005

Sasa M, Shirasaka R, Joesoef A, Abdulwahas R, Yamaura H (1976) A study of the microfilarial periodicity at Bireuen, the type locality of Brugia malayi. Southeast Asian J Trop Med Public Health 7 (3):370-376

Schiefer A, Vollmer J, Lammer C, Specht S, Lentz C, Ruebsamen-Schaeff H, Brotz-Oesterhelt H, Hoerauf A, Pfarr K (2013) The ClpP peptidase of Wolbachia endobacteria is a novel target for drug development against filarial infections. J Antimicrob Chemother 68 (8):1790-1800. doi:10.1093/jac/dkt105

Schilthuizen M, Stouthamer R (1997) Horizontal transmission of parthenogenesis-inducing microbes in Trichogramma wasps. Proc Biol Sci 264 (1380):361-366. doi:10.1098/rspb.1997.0052

Schuler H, Bertheau C, Egan SP, Feder JL, Riegler M, Schlick-Steiner BC, Steiner FM, Johannesen J, Kern P, Tuba K, Lakatos F, Koppler K, Arthofer W, Stauffer C (2013) Evidence for a recent horizontal transmission and spatial spread of Wolbachia from endemic Rhagoletis cerasi (Diptera: Tephritidae) to invasive Rhagoletis cingulata in Europe. Mol Ecol. doi:10.1111/mec.12362

Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci U S A 95 (11):5857-5864

Sedgwick SG, Smerdon SJ (1999) The ankyrin repeat: a diversity of interactions on a common structural framework. Trends Biochem Sci 24 (8):311-316

Siguier P, Filee J, Chandler M (2006) Insertion sequences in prokaryotic genomes. Curr Opin Microbiol 9 (5):526-531. doi:10.1016/j.mib.2006.08.005

Sikorski RS, Boguski MS, Goebl M, Hieter P (1990) A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell 60 (2):307-317

Siozios S, Ioannidis P, Klasson L, Andersson SG, Braig HR, Bourtzis K (2013) The diversity and evolution of Wolbachia ankyrin repeat domain genes. PLoS One 8 (2):e55390. doi:10.1371/journal.pone.0055390

Sironi M, Bandi C, Sacchi L, Di Sacco B, Damiani G, Genchi C (1995) Molecular evidence for a close relative of the arthropod endosymbiont Wolbachia in a filarial worm. Mol Biochem Parasitol 74 (2):223-227

Slatko BE, O'Neill SL, Scott AL, Werren JL, Blaxter ML (1999) The Wolbachia Genome Consortium. Microb Comp Genomics 4 (3):161-165

Slatko BE, Taylor MJ, Foster JM (2010) The Wolbachia endosymbiont as an anti-filarial nematode target. Symbiosis 51 (1):55-65. doi:10.1007/s13199-010-0067-1

Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22 (21):2688-2690. doi:10.1093/bioinformatics/btl446

Stouthamer R, Breeuwer, J.A.J., Luck, R.F., & Werren, J.H. (1993) Molecular identification of microorganisms associated with parthenogenesis. Nature 361:66-68

Sun LV, Foster JM, Tzertzinis G, Ono M, Bandi C, Slatko BE, O'Neill SL (2001) Determination of Wolbachia genome size by pulsed-field gel electrophoresis. J Bacteriol 183 (7):2219-2225. doi:10.1128/JB.183.7.2219-2225.2001

Supali T, Djuardi Y, Pfarr KM, Wibowo H, Taylor MJ, Hoerauf A, Houwing-Duistermaat JJ, Yazdanbakhsh M, Sartono E (2008) Doxycycline treatment of Brugia malayi-infected persons reduces microfilaremia and adverse reactions after diethylcarbamazine and albendazole treatment. Clin Infect Dis 46 (9):1385-1393. doi:10.1086/586753

Suzuki S, Yamasaki K (1991) Sex-reversal of male Armadillidium vulgare (Isopoda, Malacostraca, Crustacea) following andrectomy and partial gonadectomy. Gen Comp Endocrinol 83 (3):375-378

Suzuki S, Yamasaki K (1995) Morphological studies on sexual differentiation in Armadillidium vulgare (Isopoda: Armadillidae): Androgenic gland and male sexual characters. . Crustacean Research 24:93-103

Suzuki S, Yamasaki K (1997) Sexual bipotentiality of developing ovaries in the terrestrial isopod Armadillidium vulgare (Malacostraca, Crustacea). Gen Comp Endocrinol 107 (1):136-146. doi:10.1006/gcen.1997.6914

Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278 (5338):631-637

Tay ST (2013) Wolbachia endosymbionts, Rickettsia felis and Bartonella species, in Ctenocephalides felis fleas in a tropical region. J Vector Ecol 38 (1):200-202. doi:10.1111/j.1948-7134.2013.12030.x

Taylor MJ, Bandi C, Hoerauf A (2005) Wolbachia bacterial endosymbionts of filarial nematodes. Adv Parasitol 60:245-284. doi:10.1016/S0065-308X(05)60004-8

Taylor MJ, Bandi C, Hoerauf AM, Lazdins J (2000) Wolbachia bacteria of filarial nematodes: a target for control? Parasitol Today 16 (5):179-180

Taylor MJ, Cross HF, Ford L, Makunde WH, Prasad GB, Bilo K (2001) Wolbachia bacteria in filarial immunity and disease. Parasite Immunol 23 (7):401-409

Taylor MJ, Hoerauf A, Townson S, Slatko BE, Ward SA (2013) Anti-Wolbachia drug discovery and development: safe macrofilaricides for onchocerciasis and lymphatic filariasis. Parasitology:1-9. doi:10.1017/S0031182013001108

Taylor MJ, Voronin D, Johnston KL, Ford L (2012) Wolbachia filarial interactions. Cell Microbiol. doi:10.1111/cmi.12084

Teixeira L, Ferreira A, Ashburner M (2008) The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. PLoS Biol 6 (12):e2. doi:10.1371/journal.pbio.1000002

Tettelin H, Riley D, Cattuto C, Medini D (2008) Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 11 (5):472-477

Thomas F, JF G, F R (2007) Ecologie et évolution des systemes parasites. DE BOECK,

Thrall PH, Hochberg ME, Burdon JJ, Bever JD (2007) Coevolution of symbiotic mutualists and parasites in a community context. Trends Ecol Evol 22 (3):120-126. doi:10.1016/j.tree.2006.11.007

Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Le Bouguenec C, Lescat M, Mangenot S, Martinez-Jehanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Medigue C, Rocha EP, Denamur E (2009) Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths. PLoS Genet 5 (1):e1000344. doi:10.1371/journal.pgen.1000344

Tram U, Sullivan W (2002) Role of delayed nuclear envelope breakdown and mitosis in Wolbachia-induced cytoplasmic incompatibility. Science 296 (5570):1124-1126

Tsukahara F, Urakawa I, Hattori M, Hirai M, Ohba K, Yoshioka T, Sakaki Y, Muraki T (1998) Molecular characterization of the mouse mtprd gene, a homologue of human TPRD: unique gene expression suggesting its critical role in the pathophysiology of Down syndrome. J Biochem 123 (6):1055-1063

Valette V, Essono P-YB, Clec'h WL, Johnson M, Bech N, Grandjean F (2013) Multi-infections of feminizing Wolbachia strains in natural populations of the terrestrial isopod Armadillidium vulgare.

Vandel A (1941) Recherches sur la génétique et la sexualité des isopodes terrestres. VI: Les phénomènes de monogénie chez les Oniscoides. Bull Biol Er Bel 75:316-363

VanValen L (1973) A New Evolutionary Law. Evolutionary Theory 1:1-30

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

Vavre F, Girin C, Bouletreau M (1999b) Phylogenetic status of a fecundity-enhancing Wolbachia that does not induce thelytoky in Trichogramma. Insect Mol Biol 8 (1):67-72

Venter JC, Smith HO, Hood L (1996) A new strategy for genome sequencing. Nature 381 (6581):364-366. doi:10.1038/381364a0

Verne S, Johnson M, Bouchon D, Grandjean F (2007) Evidence for recombination between feminizing Wolbachia in the isopod genus Armadillidium. Gene 397 (1-2):58-66. doi:10.1016/j.gene.2007.04.006

Verne S, Johnson M, Bouchon D, Grandjean F (2012) Effects of parasitic sex-ratio distorters on host genetic structure in the Armadillidium vulgare-Wolbachia association. J Evol Biol 25 (2):264-276. doi:10.1111/j.1420-9101.2011.02413.x

von Dohlen CD, Kohler S, Alsop ST, McManus WR (2001) Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. Nature 412 (6845):433-436. doi:10.1038/35086563

Voronin D, Cook DA, Steven A, Taylor MJ (2012) Autophagy regulates Wolbachia populations across diverse symbiotic associations. Proc Natl Acad Sci U S A 109 (25):E1638-1646. doi:10.1073/pnas.1203519109

Voronin DA, Kiseleva EV (2007) [Functional role of proteins containing ankyrin repeats]. Tsitologiia 49 (12):989-999

Wakeel A, den Dulk-Ras A, Hooykaas PJ, McBride JW (2011) Ehrlichia chaffeensis tandem repeat proteins and Ank200 are type 1 secretion system substrates related to the repeats-in-toxin exoprotein family. Front Cell Infect Microbiol 1:22. doi:10.3389/fcimb.2011.00022

Walker T, Klasson L, Sebaihia M, Sanders MJ, Thomson NR, Parkhill J, Sinkins SP (2007) Ankyrin repeat domain-encoding genes in the wPip strain of Wolbachia from the Culex pipiens group. BMC Biol 5:39. doi:10.1186/1741-7007-5-39

Weeks AR, Turelli M, Harcombe WR, Reynolds KT, Hoffmann AA (2007) From parasite to mutualist: rapid evolution of Wolbachia in natural populations of Drosophila. PLoS Biol 5 (5):e114. doi:10.1371/journal.pbio.0050114

Werren JH (1997) Biology of Wolbachia. Annu Rev Entomol 42:587-609. doi:10.1146/annurev.ento.42.1.587

Werren JH, Baldo L, Clark ME (2008) Wolbachia: master manipulators of invertebrate biology. Nat Rev Microbiol 6 (10):741-751. doi:10.1038/nrmicro1969

Werren JH, Jaenike J (1995) Wolbachia and cytoplasmic incompatibility in mycophagous Drosophila and their relatives. Heredity (Edinb) 75 (Pt 3):320-326

Williams KP, Sobral BW, Dickerman AW (2007) A robust species tree for the alphaproteobacteria. J Bacteriol 189 (13):4578-4586. doi:10.1128/JB.00269-07

Wiwatanaratanabutr I, Kittayapong P, Caubet Y, Bouchon D (2009) Molecular phylogeny of Wolbachia strains in arthropod hosts based on groE-homologous gene sequences. Zoolog Sci 26 (2):171-177. doi:10.2108/zsj.26.171

Wright JD (1979) The etiology and biology of cytoplasmic incompatibility in the Aedes scutellaris group. Thesis (PhD) -- University of California, Los Angeles -- Public Health

Wu B, Novelli J, Foster J, Vaisvila R, Conway L, Ingram J, Ganatra M, Rao AU, Hamza I, Slatko B (2009) The heme biosynthetic pathway of the obligate Wolbachia endosymbiont of Brugia malayi as a potential anti-filarial drug target. PLoS Negl Trop Dis 3 (7):e475. doi:10.1371/journal.pntd.0000475

Wu B, Novelli J, Jiang D, Dailey HA, Landmann F, Ford L, Taylor MJ, Carlow CK, Kumar S, Foster JM, Slatko BE (2013) Interdomain lateral gene transfer of an essential ferrochelatase gene in human parasitic nematodes. Proc Natl Acad Sci U S A 110 (19):7748-7753. doi:10.1073/pnas.1304049110

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA (2004) Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2 (3):E69. doi:10.1371/journal.pbio.0020069

Yen JH, Barr AR (1971) New hypothesis of the cause of cytoplasmic incompatibility in Culex pipiens L. Nature 232 (5313):657-658

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18 (5):821-829. doi:10.1101/gr.074492.107

Zhang X, Davenport KW, Gu W, Daligault HE, Munk AC, Tashima H, Reitenga K, Green LD, Han CS (2012) Improving genome assemblies by sequencing PCR products with PacBio. Biotechniques 53 (1):61-62. doi:10.2144/0000113891

Zug R, Hammerstein P (2012) Still a host of hosts for Wolbachia: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. PLoS One 7 (6):e38544. doi:10.1371/journal.pone.0038544

SUPPLEMENTAL DATA

Appendix 1: Table of the optimized parameters for the Velvet Assemblies

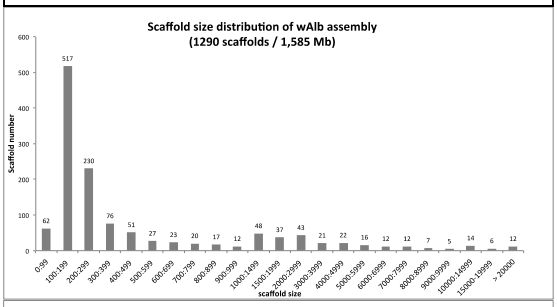
	*w*Albu	*w*Bre	*w*Dil	*w*Pet	*w*Pru	*w*VulM
Optimized Velvet assembly						
velvet hash value	47	49	47	47	47	47
exp_cov	131	94	166	164	177	157
cov_cutoff	1.178	6.171	3.083	1.978	3.999	5.199
Assembly parameters						
total number of contigs	1,755	786	1,743	856	1,282	1,325
N50	5,631	4,790	4,843	6,664	12,566	5,367
length of longest contig	36,966	32,942	40,290	28,707	56,622	40,526
Total bases in contigs	1,724,208	1,253,219	1,854,492	1,466,213	1,659,900	1,582,488
Number of contigs > 1k	290	280	359	273	214	299
Total bases in contigs > 1k	1,389,298	1,107,561	1,519,320	1,300,712	1,351,885	1,339,686

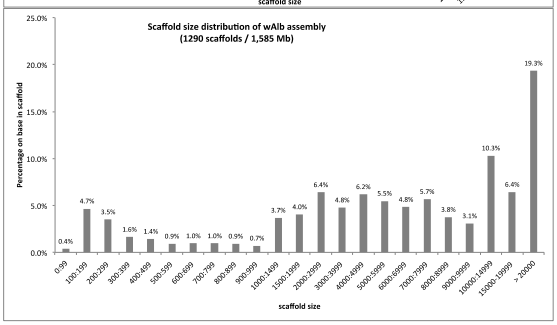
wAlbu assembly

Wolb-like : 1,290 scaffolds / 1,585,022 nt (89.98% total assembly)
(1,156 scaffolds align on wVulC_ref)

FILE: wAlb_wolb_like_scaffold.fa

Total length of sequence:	1,585,022 bp
Total number of sequences:	1,290
N25 stats:	18 sequences >= 15,273 bp
N50 stats:	60 sequences >= 6,511 bp
N75 stats:	161 sequences >= 2,114 bp
GC %:	34.56 %



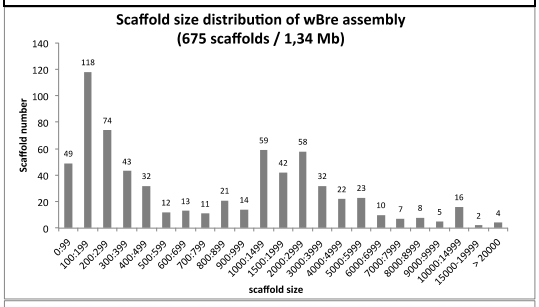


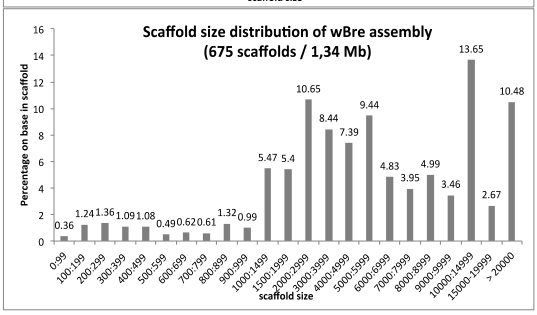
wBre assembly

Wolb-like: 675 scaffolds / 1,340,868 bp (98,5% total assembly)

FILE: wBre_wolb_like_scaffold.fa

Total length of sequence:	1,340,868 bp
Total number of sequences:	675
N25 stats:	20 sequences >= 10265 bp
N50 stats:	66 sequences >= 5422 bp
N75 stats:	158 sequences >= 2441 bp
GC %:	34.57 %



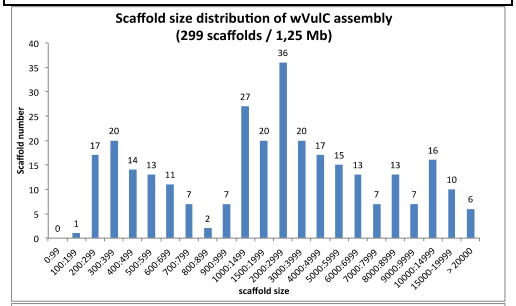


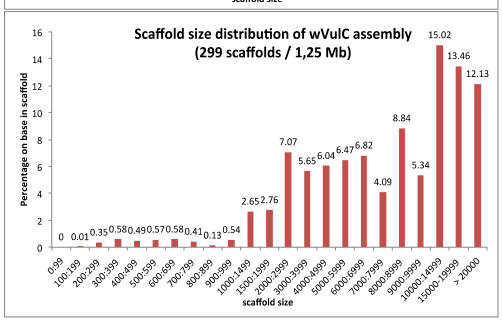
wVulC assembly

Wolb-like: 299 scaffolds / 1,250,927 bp (96.02% total assembly)
(294 scaffolds align on wVulC_ref)

FILE: wVulC_wolb_like_scaffold.fa

Total length of sequence:	1,250,927 bp
Total number of sequences:	299
N25 stats:	16 sequences >= 15638 bp
N50 stats:	45 sequences >= 8604 bp
N75 stats:	95 sequences >= 4463 bp
GC %:	34.55 %



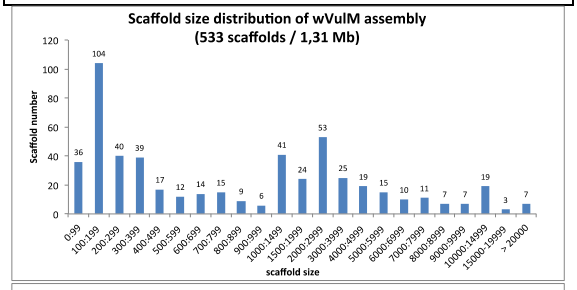


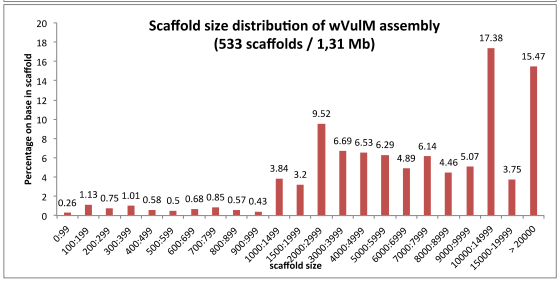
wVulM assembly

Wolb-like : 533 scaffolds / 1,314,883 nt (82.3% total assembly)

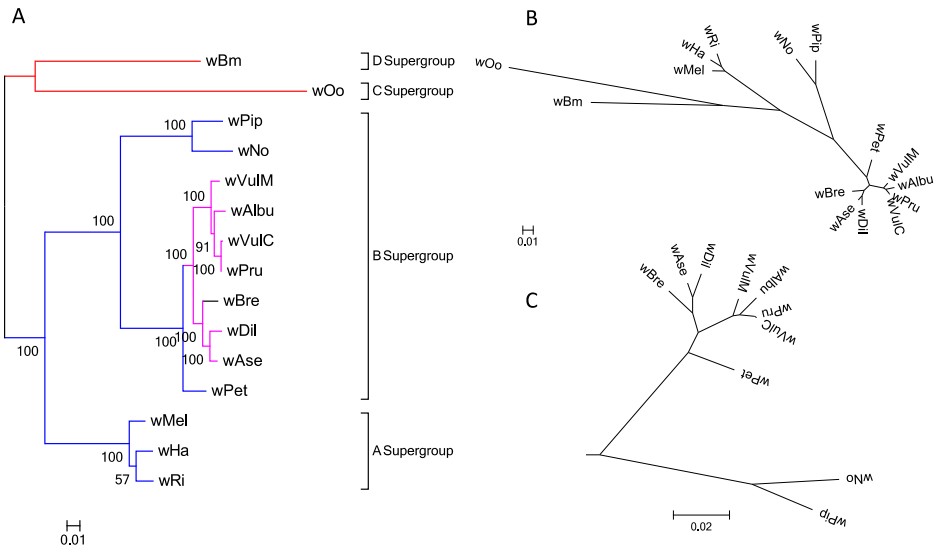
FILE: wVulM_wolb_like_scaffold.fa

Total length of sequence:	1314883 bp
Total number of sequences:	533
N25 stats:	16 sequences >= 12671 bp
N50 stats:	50 sequences >= 7275 bp
N75 stats:	117 sequences >= 3335 bp
GC %:	34.56 %





Appendix 2: Representative examples of the contig distribution of the *Wolbachia* genome assemblies.



Appendix 3: Unrooted phylogenomics of *Wolbachia* strains (15). The phylogeny tree was inferred on the concatenate alignment of the 529 single gene orthologs in RaxML using the maximum likelihood method. Numbers on the branches represent the support from 1,000 bootstraps. Trees are drawn to scale, with branch lengths measured in number of substitutions per site. A: Rectangular representation of the phylogenetic tree. Red: mutualistic strain; blue: CI strains; pink: feminizing strains; black: unknown phenotype. B: Radial representation of the phylogenetic tree. C: Radial representation of B Supergroup *Wolbachia* strains. Horizontal bars indicate number of substitutions per site.

Appendix 4: Table of primers used in cloning PCR

Primers used for GST-cloning

Target		gene	primer name	sequence
wVulC	wsp	wVul_0359	ORF17aF-BamHI	TTTTGGATCCAAGGGTTTTACTTTGGTGGTGGATATT
			ORF17R-XhoI	TTTTCTCGAGTATGGTAGTAGACACTTTATTATCACCTAC
		wVul_0072	ORF72F-BamHI	TTTTGGATCCAAATACCAATAAAATGTCAAGAACTTGAAA
			ORF72R-XhoI	TTTTCTCGAGACTATTTTCAATATTAACTTTAGAATATATG
		wVul_1434	ORF29aF-BamHI	TTTTGGATCCCAGATCCTATTGGTCCAATAAGTG
			ORF29aR-XhoI	TTTTCTCGAGGTCTGCAACAGTTACTGGAGTAAATGCTGCA
			ORF29bF-BamHI	TTTTGGATCCCTTTAGCAACAAAAGTTGCTGGTG
			ORF29bR-XhoI	TTTTCTCGAGATCAGTAACCTTTTTCTCTAGATCATCAG
	Ankyrins	wVul_0815	ANK54F-BamHI	TTTTGGATCCGCAATGATAATTATCAATTACATTTAGC
			ANK54R-XhoI	TTTTCTCGAGTTTCCTTATAGTACTTTTTAGGTATTGTTAT
		wVul_1467	ANK79aF-BamHI	TTTTGGATCCCTGTAGAAACTACATTGTTGCATTTTG
			ANK79bF-BamHI	TTTTGGATCCATGTTGATGCATGGAACAGCG
			ANK79R-XhoI	TTTTCTCGAGCTGCTTACTAACATATGCCCCACTT
		wVul_0812	ANK87F-BamHI	TTTTGGATCCTTTTTTATTGCATTAAGAAGGGTAACTTAAC
			ANK87bR-XhoI	TTTTCTCGAGAATTGCGTTTATTATATCGATCTGCTTTTTCTC
		wVul_0887 (pk2b2)	ANK48_BamHI_F	TTTTGGATCCTGTTAAAGCTTGGCAAATTTAATAAGT
			ANK48_XhoI_R	TTTTCTCGAGTTAGCCTTTCATAAATAACTTTTCAAT
	Vector	pGEX-5X-1	pGEX-5X-1_F	GGGCTGGCAAGCCACGTTTGGTG
			pGEX-5X-1_R	CGGGAGCTGCATGTGTCAGAGG

		gene	primer name	sequence
wBm	wsp	wBm0100	wsp1f-BamHI	TTTTGGATCCATGCTAATGAAAGTTGTTCAAATGATGGAC
			wsp1r-XhoI	TTTTCTCGAGGCCATTTTTAATGTTGACCTTAGAATACATG
		wBm0284	wsp2F-BamHI	TTTTGGATCCAAGGATTCTACTTTGGTGGTGGATACC
			wsp2r-XhoI	TTTTCTCGAGATTTATGCTACCTGTCTGATTATCTGATGC
		wBm0432	wsp3af-BamHI	TTTTGGATCCCAGATCCTGTTGGTCCAATAGCTG
			wsp3ar-XhoI	TTTTCTCGAGATCTACAATATCTGGAGTAGGAGCTC
			wsp3bf-BamHI	TTTTGGATCCCTGCAAAGGCACAAGTTATTGCTG
			wsp3br-XhoI	TTTTCTCGAGCGCAGTAACTTTTGTTTCTTTATCCTCAC
	WASP	wBm0076	076f-BamHI	TTTTGGATCCGTAATAAGAAAGAAGCAGGTC
			076r-XhoI	TTTTCTCGAGCTATTGTTTTATAAGAAAAGCTCTAG
	Ankyrins	wBm0287	287f-BamHI	TTTTGGATCCTGCATGTTCACAAGAATGTTG
			287r-XhoI	TTTTCTCGAGCTTGATTAAGAAAAATTCTACAACG
		wBm0296	296f-BamHI	TTTTGGATCCGGGCAAATGTTAATGCAGAAG
			296r-XhoI	TTTTCTCGAGACGCCGAATTATATAGTGACTTTTG
		wBm0394	394f-BamHI	TTTTGGATCCTGTTAGATCATAGTACTAGCTATG
			394r-XhoI	TTTTCTCGAGTTAGTTATTGACGACAGCAAAATATAAAG
		wBm0447	447f-BamHI	TTTTGGATCCGTAATGATAATTATCTATTACATTCAGC
			447r-XhoI	TTTTCTCGAGTGTAACTTTTTGCTTGCCTTTTG
		wBm0582	ANK582F-BamHI	TTTTGGATCCTTTTTTACTGTATTAAAAAGAATAACTTGGT
			ANK582R-XhoI	TTTTCTCGAGAATTGCATTTATTATATCAACTTGTTCTCTTTC

Appendix 5: Table of primers used in PCR, qPCR

Primers used in qPCR

wVulC	*wsp*	wVul_1434	qwsp-F	TGGTGCAGCATTTACTCCAG
			qwsp-R	TCGCTTGATAAGCAAAACCA
	hcpa		qhcpA-F	CACGTCTTCGCTCTGCTAT
			qhcpA-R	CAGAAGCAGTTCGGTTGCG
	Ankyrin	wVul_0887 (pk2b2)	qANK40F	ACGAATTGGTCGTAGGTTCG
			qANK40R	TCATAGCTTCTTTCCCGCTG
	gene X	wVul174	qVul_43F	GCCGGTTTTATCTGTGGTACA
			qVul_208R	TACCACGACCTGCCAACATA
			qVul_145F	TGTATTGGTTTGATGGCAGTG
			qVul_316R	GCATTCGATGCTGATCCTTT
		external primers	Vul_SecF2_F	TTGGCACTTGCTCTGCAATA
			Vul_0175_R	ACTTGGAATGAGCTTTTTGAGG
			Vul_160448F	TGGCTAGTCGATTGCTTGAA
			Vul_SecF_F	ATTGGCAATGTGGATTTGGT
			Vul_0172_F	GCGAGACCGAGACAAATAGC
			Vul_0176_R	TGGCTTTGGCTTGTGTTAAG
			Vul_1539_F	GCTTGTTTTCTGCCCCATTA
	Biotin	ankyrin	Ank_1546R	CCTTTCCGAGCAGCAAAA
		BioA	BioA_F	ATCAGACCACTCGGCAATTC
			BioA_R	AAGCGTTACTGCAGGCTCAT
		BioB	BioB_F	TCTGGGCAGCTACCTGTTTT
			BioB_R	GTTTATCAGCTGGGCGAGAA
		BioC	BIOC_F	GCAAATCAAGGTAGCCAAGC
			BioC_R	TGTGGATGAAATTCAGGAAGG
		BioD	BioD_F	TGGGGGTGATCTTAAATGGA
			BioD_R	CTCCTCCAGCACCCTCTATG
		BioF	BioF_F	GGGATTTAGTGCCCCATCTT
			BioF_R	AGAGCAACCGAGAACCAGTG
		BioH	BioH_F	TGACACAGCGCTTTCAAATG
			BioH_R	CCCCAACCGTGACAAAATAC
		BioB	qBioB_F	CCATACAGAAACGGGTGCTT
			qBioB_R	AAACAGGTAGCTGCCCAGAA
		BioB	qBioC_F	CCTTCCTGAATTTCATCCACA
			qBioC_R	CGCCTAGGATGAATGCAACT
		wVul0646	qVul645_F	TGGAGACGTTTCTAAGTGCAG
			qVul645_R	TGAGCTTTTTCTCTCAGTACTCTTTC
		wVul0645	qVul646_F	CGCCAATTAATCGGAGAAGA
			qVul646_R	ACATGCTGCTGCCTAAGCTC
		wVul1677	qVul1677_F	TTCAAAGTTTTGAGGAACAAATGA
			qVul1677_R	GCGAAACACACACTGAGGAA
		wVul1344	qVul1344_F	ATTTGATGAATTCGGGTGGA
			qVul1344_R	AGCCTTCTAATGGCCCAATC
			qVul1344_F2	ATGATTGGGCCATTAGAAGG
			qVul1344_R2	AACACCAGGAATAGGCTGCTT

A. vulgare	Androgenic hormone		qAH-F1	GAGGTATGAGATCCGATGTG
			qAH-R1	AAATGCCAATTCATCTTCAGG

organism	Target	gene	primer name	primer sequence
wBm	wsp	wsp 1 (wBm0100)	wBm-wsp1-F	ATAACGGTAGCCGAGGAAGG
			wBm-wsp1-R	ATGGGTCACCTCAAAACTCG
		wsp 2 (wBm0284)	wBm-wsp2-525-F	TGGTGCAACAAGGATGAAGA
			wBm-wsp2-657-R	GCCAATAACGCCAAAATGTC
		wsp3 (Wbm0432)	wBm-wsp3-F	ATCAGCAACCCTGCAAAGGCAC
			wBm-wsp3-R	CGCCATAAGAACCAAAGTAGCGAG
	Ankyrins	wBm0287	wBm0287-92-F	ACGCTGTTCAAAGAGGATGC
			wBm0287-194-R	GCCCTATGCACTGGCTTACT
		wBm0296	wBm0296-510-F	ACACTTTGCTGCTCGATGTG
			wBm0296-687-R	ATCTGCTGCACGAACATCTG
		wBm0394	wBm0394-168-F	CGCCCCCATCTAACAATAAC
			wBm0394-351-R	TTTCGAACAAGTGTTGCCTTA
		wBm0447	wBm0447-2851-F	GCTGTATGGAAGTGCGTTGA
			wBm0447-2964-R	TACAATTTGCCCTTGGAAGC
		wBm0582	wBm0582-640-F	TGCCTGATCTTCTGCCTTTT
			wBm0582-787-R	TTCAAGCGCTGATTTGAATG
	WASP	wBm0076	qWASP_444F	ATCACAAGATGCAGGTGCTG
			qWASP_671R	TTCTCATCGCGTTCTTCCTT
			qWASP_526F	GGTTTATTTGCAGGGGTGAA
			qWASP_734R	CCACCTACGTTATCGCTTGA
			wBm75_F	CTCCAGGATTGTTGCCTGAT
			wBm77_R	ATCCCTGATGATGCCTTCAA
	wBm0047	wBm0047	wbm47_88F	TTCAGGGGCGATGTCTAAAT
			wbm47_280R	TTGCACAGACAGTGGGTGTAG
			wBm46_F	GCATCTTGCGGTTATGGAAA
			wBm48_R	AGAAAAGCTAGGTGGTGAGCA

B. malayi	Tubulin		Bm-tub-1-F	TTTCCCTGGACAGTTGAACGC
			Bm-tub-1-R	CGATAAGCAGCAGCATCACGAG

RESUME EN FRANCAIS

THÈSE

pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITE DE POITIERS

FACULTE DES SCIENCES FONDAMENTALES ET APPLIQUEES

(Diplôme National – Arrêté du 7 août 2006)
École Doctorale **Sciences pour l'environnement Gay Lussac**
Spécialité : **Biologie des organismes – Interactions symbiotiques**
préparée au laboratoire **Écologie et Biologie des Interactions,**
équipe **Écologie Évolution Symbiose** – Poitiers – France
et à **New England Biolabs** – Ipswich – Etats-Unis
présentée le 27 septembre 2013 par

Sandrine GENIEZ

Résumé en français

Etude des relations symbiotiques entre *Wolbachia* et les isopodes et les nématodes par génomique et analyse de l'intéractome

Directeurs de thèse : **Pierre GRÈVE**
Barton SLATKO

JURY

Dr. P. MAVINGUI	Directeur de recherche CNRS, Université de Lyon	*Rapporteur*
Dr. K. PFARR	Research Scientist, Université de Bonn	*Rapporteur*
Dr. E. HERNIOU	Chargée de recherche CNRS, Université de Tours	*Examinatrice*
Pr. T. BERGES	Professeur, Université de Poitiers	*Examinateur*
Pr. P. GREVE	Professeur, Université de Poitiers	*Examinateur*
Dr. B. SLATKO	Senior Scientist, New England Biolabs, Inc., Ipswich	*Examinateur*

Introduction

Le terme de symbiose au sens biologique définit une interaction étroite entre au moins deux partenaires, un hôte et un symbiote. De nombreux types de symbiose ont été définis et de nombreux scientifiques défendent la théorie selon laquelle pratiquement tous les organismes sont impliqués dans au moins une relation symbiotique avec un autre organisme (Combes 1995; Thomas et al. 2007).

Ces relations symbiotiques s'étendent le long d'un continuum composé de trois principaux types de relations symbiotiques : le mutualisme, qui définit une symbiose pour laquelle l'hôte et le symbiote dépendent l'un de l'autre pour leur survie réciproque, le commensalisme, qui définit une symbiose qui bénéficie à l'un des partenaires sans affecter l'autre, et le parasitisme qui bénéficie à l'un des partenaires au détriment de l'autre (Combes 1995). Cependant, la symbiose est une relation dynamique : en fonction de l'environnement et des besoins du partenaire symbiotique, la relation symbiotique peut évoluer le long du continuum allant du mutualisme au parasitisme (Thrall et al. 2007; Fellous and Salvaudon 2009; Leigh 2010). La symbiose constitue ainsi une force majeure de l'évolution qui permet l'adaptation des organismes aux conditions environnementales. Une attention particulière a été accordée aux microorganismes intra-cytoplasmiques parce que leur survie et leur reproduction dépendent directement de la survie et du succès reproducteur de leurs hôtes (Bull and Rice 1991). Ce concept évolutif a été initialement décrit par Van Valen en 1973. En s'appuyant sur la théorie de la Reine Rouge, il décrit une «course aux armements» où deux populations doivent perpétuellement co-évoluer pour maintenir la symbiose (VanValen 1973). Cette théorie est utilisée pour expliquer la résistance des pathogènes aux défenses de l'hôte et l'adaptation des hôtes aux facteurs de virulence des symbiotes via par exemple le système immunitaire. Les facteurs de virulence bactériens sont libérés dans le cytoplasme des cellules hôtes principalement par l'intermédiaire des systèmes de sécrétion bactériens (Durand et al. 2009). Ce sont pour la plupart des protéines contenant des motifs de type eucaryote qui miment ainsi l'action des protéines de l'hôte. Afin d'assurer leur prolifération, les symbiotes profitent de la reproduction de l'hôte pour faciliter leur transmission. Ainsi, toutes les actions qui tendent à favoriser la reproduction de l'hôte seront bénéfiques à la propre prolifération des symbiotes. Cela est particulièrement vrai lorsque les symbiotes sont transmis verticalement : ils infestent les générations suivantes en colonisant les ovocytes. Cette voie de transmission est couramment utilisée par l'alpha-protéobactérie *Wolbachia*. Cet endosymbiote obligatoire est très commun chez les arthropodes et les nématodes

filaires, ce qui en fait l'endosymbiote le plus répandu sur Terre. Les analyses phylogénétiques ont permis de décrire 7 supergroupes différents, avec les groupes A et B qui incluent la plupart des souches parasites d'arthropodes et les groupes C et D la majorité des souches infestant les nématodes. Selon la souche, *Wolbachia* peut induire une grande diversité de phénotype le long du continuum de la symbiose allant du parasitisme au mutualisme. Alors que *Wolbachia* est nécessaire à la survie des nématodes filaires, elle est responsable de différentes modifications de la reproduction chez les arthropodes, en induisant la mort des mâles, l'incompatibilité cytoplasmique, la parthénogénèse ou la féminisation des mâles en femelles fonctionnelles. La bactérie *Wolbachia* est de fait d'un grand intérêt puisqu'elle est l'objet d'études de cible thérapeutique chez les filaires responsable de filarioses humaines, mais aussi fortement étudiée comme agent de contrôle biologique contre les ravageurs des cultures ou les vecteurs de maladies tels que les moustiques. Cependant, les mécanismes moléculaires de ces différentes relations symbiotiques sont encore méconnus.

Au début de cette étude, seuls quatre génomes de *Wolbachia* avaient été entièrement séquencés et déposés dans GenBank, *w*Mel de *D. melanogaster* (Wu et al. 2004), *w*Bm de *B. malayi* (Foster et al. 2005), *w*Pip de *C. quinquefasciatus Pel (Klasson et al. 2008), w*Ri de *D. simulans* (Klasson et al. 2009). Par la suite, les génomes partiels de trois souches de *Wolbachia* infectant des drosophiles ont également été récupérées des séquences obtenues lors des séquençages de leur hôte (Salzberg et al. 2005). Plus récemment, trois génomes complets supplémentaires ont été publiés: *w*Oo *O. ochengi* (Darby et al. 2012), *w*No et *w*Ha de *D. simulans* (Ellegaard et al. 2013). Enfin, de nombreux génomes supplémentaires sont en cours de séquençage et des séquences partielles de plusieurs souches de *Wolbachia* sont également disponibles.

Objectifs de la thèse:

L'étude présentée propose différentes approches visant à mettre en avant les relations entre les souches de *Wolbachia* et leurs hôtes. Dans le cadre de cette recherche, ce projet de thèse vise à comprendre les mécanismes moléculaires de la symbiose entre les bactéries *Wolbachia* et leurs hôtes. Cette recherche a été axée sur l'identification des effecteurs bactériens présumés impliqués dans l'interaction hôte-bactérie dans le cadre d'une souche de *Wolbachia* impliquée dans une relation mutualiste (*B. malayi*) et d'une souche impliquée dans une relation parasitaires (*A. vulgare*) en utilisant des approches de génomiques, de transcriptomiques et de protéomiques.

Alors que la plupart des arthropodes sont infectés par au moins une souche de *Wolbachia*, seuls le génome de quelques souches d'arthropodes sont à ce jour complétement ou partiellement séquencés. Les nématodes filaires responsables de filarioses sont infectés par des souches mutualistes et 4 souches sont à ce jour complétement ou partiellement séquencées, dont *w*Bm la souche de *Wolbachia* infestant le nématode *B. malayi.*

Afin de générer une plus grande banque de données de génomes de *Wolbachia* qui rend accessible les études de génomique comparative, une des étapes de ce projet de thèse était de développer un procédé efficace de séparation de l'ADN de *Wolbachia* de l'ADN de l'hôte. En utilisant le génome connu de *w*Bm comme témoin, une méthode d'isolation d'ADN bactérien a été développée et utilisée pour le séquençage de 8 souches supplémentaires de *Wolbachia* infectant les isopodes terrestres qui induisent soit une féminisation de l'hôte soit une incompatibilité cytoplasmique.

Une étude de génomique comparative a ensuite été réalisée pour identifier les gènes présents uniquement dans les souches induisant un même phénotype ainsi que les effecteurs bactériens potentiellement impliqués dans l'interaction symbiotique tels que les protéines membranaires ou les protéines possédant des domaines eucaryote-like ansi que les protéines des systèmes de sécrétion.

Cette approche a été complétée en parallèle par une analyse transcriptomique et de l'interactome protéique entre *Wolbachia* et deux de ces hôtes : le nématode *B. malayi* et l'isopode *A. vulgare*.

Chapitre I : Conception et développement d'une banque de donnée de génomes de *Wolbachia*.

Avec l'émergence des nouvelles technologies de séquençage, le séquençage de génome entier est devenu accessible pour des bactéries endocellulaires telles que *Wolbachia*. Cependant, l'obtention d'une quantité suffisante d'ADN de *Wolbachia* pur demeure un défi puisque les stratégies de purification telles que les techniques de gradients chimiques (Charles and Ishikawa 1999), de purification en champ pulsé avec ou sans amplification, de construction de banques génomiques suivi de marche sur les gènes, etc … (Mavingui et al. 2005; Sun et al. 2001; Foster et al. 2005; Iturbe-Ormaetxe et al. 2011) sont restées infructueuses ou spécifique d'un organisme.

En adaptant la procédure d'enrichissement génomique par capture de séquence ciblée, nous avons développé une méthode qui permet de spécifiquement isoler l'ADN endosymbiotique de l'ADN de l'hôte et de le séquencer par Illumina HiSeq (50bp paired-

end). Ce procédé, qui a fait l'objet d'une publication dans la revue Symbiosis (Geniez et al. 2012), a été testé sur le génome publié de *w*Bm pour lequel 100% de la séquence a été obtenue et appliqué avec succès (à 97.3%) à *w*VulC, qui est une souche de *Wolbachia* phylogénétiquement éloigné, démontrant la robustesse de la méthode applicable à n'importe quelle souche de *Wolbachia*.

Par la suite, cette méthode a encore optimisée pour parfaire l'efficacité de la capture: environ 3% de la séquence du génome wVulC connu n'avait pas été capturée sans doute parce que les séquences homologues n'étaient pas présentes dans la banque de sondes. De plus, il nous a paru indispensable d'intégrer dans l'approche des outils de séquençage nouvellement disponibles, tels que le séquençage par Illumina MiSeq qui permet d'obtenir des séquences plus longues et d'utiliser les nouveaux kits de préparation de banque d'ADN plus efficace que les kits précédemment commercialisés.

Ainsi 4 étapes principales ont été optimisées:

1. La banque de capture a été complétée avec les séquences des génomes de *Wolbachia* nouvellement publiés et les séquences du génome partiel de *w*VulC.
2. La banque de séquençage a été préparée avec le nouveau kit de préparation d'ADN (NEBNext Ultra Library preparation kit) qui nécessite seulement 1,5 µg d'ADN génomique.
3. L'ADN génomique a été fragmenté à 500 pb.
4. Le séquençage a été réalisé sur un séquenceur Illumina MiSeq qui produit de longues séquences (250bp paired-end)

Ces optimisations ont permis de passer de 97.3 % de *w*VulC capturé à 99.6% et plus important, de capturer également des séquences de *w*VulC telles que les gènes de la voie de biosynthèse de la biotine. Ces gènes n'étant pas présent dans génomes utilisés pour créer la première banque de capture, ces séquences n'avaient alors pas été capturées.

Ces nouvelles approches nous ont permis de capturer l'ADN de 7 nouveaux génomes de *Wolbachia* infectant les crustacés isopodes qui induisent soit une incompatibilité cytoplasmique soit la féminisation des males.

Table 1: Modalités de séquençage des différentes souches de *Wolbachia*.

	First set	Second set
Capture library	initial	optimized
DNA fragment size	300 bp	500 bp
Sequencer	HiSeq	MiSeq
Reads	50bp paired end reads	250 bp paired end reads
Strains	*w*Albu_HiSeq	*w*Ase_MiSeq
	*w*Bre_HiSeq	*w*Dil_MiSeq
	wDil_HiSeq	*w*Pru_MiSeq
	*w*Pet_HiSeq	*w*VulC_MiSeq
	wPru_HiSeq	
	*w*VulC_HiSeq	
	*w*VulM_HiSeq	

Bien que ces génomes soient partiellement assemblés en contigs, nous avons démontré que la majorité de l'information génétique est présente et permet une approche de génomique comparative. Pour toutes ces souches séquencées, nous avons obtenu environ 1,5 Mb de séquences qui ont une densité codante de 83 %, ce qui est un taux proche de ceux obtenus pour les génomes déjà publiés. Seul le séquençage de *w*Pru n'a pas permis d'obtenir des données d'une qualité suffisante pour l'inclure dans les analyses de génomique comparative.

Cette approche utilisant les nouvelles méthodes de séquençage qui conduit à des génomes partiellement assemblés a été précédemment validés pour le séquençage des souches de *Wolbachia* *w*AlbB et *w*Bol1-b (Mavingui et al. 2012; Duplouy et al. 2013). Cependant, certaines optimisations d'assemblage sont encore possibles notamment pour deux des génomes étudiés, en combinant les séquences obtenues par deux techniques NGS différentes (Illumina HiSeq et MiSeq).

Nous avons également initié le séquençage du génome *w*VulC par la nouvelle technique de séquençage de Pacific Bioscience (PacBio) qui permet de séquencer une unique molécule d'ADN (jusqu'à 20kb) sans étape d'amplification. Ce séquençage de longues séquences nous permettrait de compléter le génome et d'obtenir le premier génome complet d'une souche de *Wolbachia* infestant un isopode. Cette approche a déjà été utilisée avec succès pour le séquençage complet de génomes riches en séquences répétitives, en combinaison avec un séquençage Illumina (English et al. 2012) (Table1).

Plus globalement, ces séquençages de génomes bactériens sont à la base des approches de génomique comparative. Ces comparaisons sont des outils puissants pour étudier les similitudes et les différences entre les souches puisqu'ils permettent par exemple de donner d'importantes indications biologiques de la structure du génome comme ses statistiques globales, les répétitions de séquences, les réarrangements du génome, la

recombinaison des gènes, et les études de synténie. En effet, des études sur les éléments transposables et la dynamique des introns mobiles ont permis de donner un premier aperçu de l'évolution de la structure des génomes de *Wolbachia* (Cerveau et al. 2011; Leclercq et al. 2011). Chez les nématodes filaires, le séquençage de souches de *Wolbachia* dans les populations de terrain peut conduire à l'identification de polymorphismes qui pourraient permettre l'identification de cibles thérapeutiques et le suivi de l'évolution de prévalence et de distribution de *Wolbachia* dans des conditions naturelles. En effet, la plupart des recherches en cours de cible thérapeutiques sont basées sur la souche de laboratoire *w*Bm du laboratoire TRS, qui a été conservée en laboratoire depuis plus de 30 ans. Or, des études récentes sur les insectes (Weeks et al. 2007; Carrington et al. 2011) ont montré des divergences génétiques entre populations naturelles et les lignées de laboratoire. Ainsi, le séquençage de la population naturelle de *w*Bm pourrait mettre en lumière ces polymorphismes éventuels afin d'adapter les cibles thérapeutiques identifiées au génotype de terrain.

La génomique comparative peut également conduire à la comparaison des régions non codantes pour prédire les éléments de régulation par exemple. Cependant notre approche de génomique comparative a été axées sur l'étude des régions codantes et des protéines orthologues afin d'identifier les potentiels acteurs moléculaires de la symbiose.

Table 2: Caractéristiques des génomes de *Wolbachia*. Les souches séquencées dans cette étude sont en gras.

	*w*VulC	*w*Pip-Pel	*w*Mel	*w*Ri	*w*Bm	*w*Oo	*w*No	*w*Ha	***w*VulC**	***w*Albu**	***w*Ase**	***w*Bre**	***w*Dil**	***w*Pet**	***w*VulM**
Host	*Armadillidium vulgare*	*Culex pipiens*	*Drosophila melanogaster*	*Drosophila simulans*	*Brugia malayi*	*Onchocerca ochengi*	*Drosophila simulans*	*Drosophila simulans*	***Armadillidium vulgare***	***Armadillidium album***	***Oniscus Asellus***	***Helleria brevicornis***	***Porcellio dilatatus dilatatus***	***Porcellio dilatatus petiti***	***Armadillidium vulgare***
Genome statistics															
Genome size (bp)	1,663,852	1,482,455	1,267,782	1,445,873	1,080,084	957,990	1,301,823	1,295,804	**1,537,389**	**1,585,022**	**1,518,980**	**1,340,868**	**1,541,642**	**1,341,349**	**1,314,883**
GC content (%)	34.4	34.2	35.2	35.2	34.2	32.1	34.5	35.34	**34.6**	**34.6**	**34.6**	**34.6**	**34.6**	**34.5**	**34.6**
Gene predictions															
Predicted protein-coding genes	1,568	1,385	1,269	1,264	903	842	1,040	1,010	**1,880**	**2,615**	**1,924**	**1,834**	**1,925**	**1,665**	**1,684**
Coding density (protein-coding nucleotides, %)	82.52	81.20	80.10	78.10	67.00	66.50	80.00	78.00	**82.65**	**84.05**	**82.45**	**83.82**	**83.98**	**82.75**	**82.98**
Average gene size (bp)	728	944	851	951	899		1013	1000	**676**	**509**	**651**	**613**	**673**	**667**	**648**
Number of rRNA	3	3	3	3	3	3	3	3	**3**	**3**	**3**	**3**	**3**	**3**	**3**
Number of tRNA	35	34	34	34	34	34	34	34	**34**	**34**	**34**	**34**	**34**	**34**	**34**

Chapitre 2: De la génomique comparative à l'identification de pan-génomes représentatifs de certaines relations symbiotiques

Ces 7 nouveaux génomes séquencés dans cette étude ont permis de compléter le set de génomes séquencés en rajoutant 4 souches induisant une féminisation, souches induisant une incompatibilité cytoplasmique et une souche dont le phénotype induit n'est pas connu. L'étude des 23 génomes de *Wolbachia* alors connus, nous a permis d'établir le premier pan-génome de la bactérie *Wolbachia* basé sur un si grand nombre de souches. Ce pan-génome, qui décrit le répertoire complet des gènes de *Wolbachia*, est composées de plus de 6400 gènes orthologues. Cette étude a mis en lumière un pan-génome dit « ouvert » : chaque nouvelle souche séquencée révèle de nouveaux gènes. Par ailleurs, l'extrapolation mathématique de la courbe d'accumulation du pan-génome a confirmé un nombre de gènes théoriquement infini pour la bactérie *Wolbachia* et prédit 13 932 gènes différents après le séquençage de 100 souches de *Wolbachia*. Ce nombre élevé témoigne de la grande diversité génomique des souches de *Wolbachia* suggérant que ces génomes sont encore en évolution via l'acquisition et/ou la duplication de gènes qui permettent l'émergence de nouvelles fonctions.

L'analyse pan-génomique a également révélée un grand nombre de gènes spécifiques à une souche dont un large nombre sont des gènes associés aux phages (environ 15%) ou aux IS (environ 3%), soulignant le rôle des phages et des transferts horizontaux dans la diversité génomique des bactéries *Wolbachia*. L'importance des éléments mobiles et les flux de bactériophages dans la dynamique du génome ont été largement démontrés. (Leclercq et al. 2011; Gillespie et al. 2012; Penz et al. 2012; Duplouy et al. 2013; Wu et al. 2004; Klasson et al. 2009; Kent and Bordenstein 2010; Kent et al. 2011). Les génomes séquencés dans cette étude appartenant à la même niche écologique et devraient contribuer à la compréhension des implications du phage et des éléments mobiles sur l'évolution du génome de *Wolbachia*. D'autre part, une analyse plus approfondie des catégories fonctionnelles des protéines orthologues telles que les voies de biosynthèse, les transporteurs, les organismes de régulation ou de résistance aux antibiotiques contenues dans le dispensable-génome devrait également dévoiler des processus d'adaptation des bactéries à leurs environnements cellulaires souvent différents.

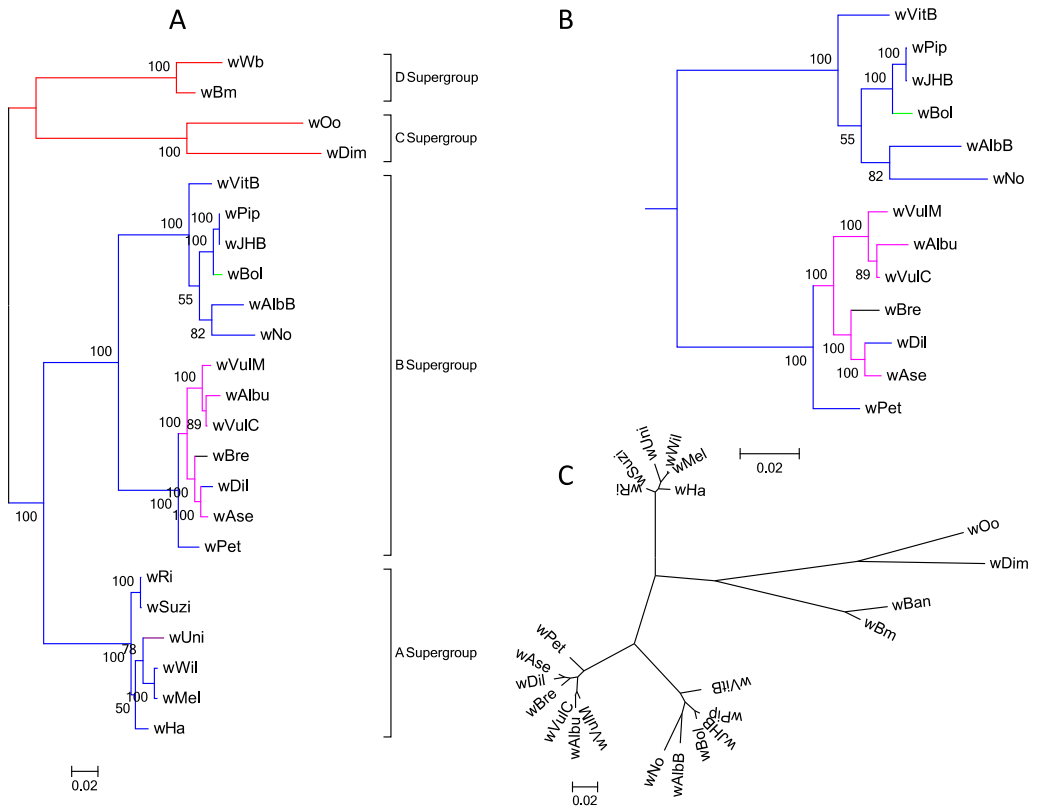


Figure 1: Arbres phylogénomiques non racinées des souches de *Wolbachia* (23). L'arbre phylogénomique a été construit à partir de la concaténation des alignements individuels des 209 orthologues présent en une seule copie dans raxml en utilisant la méthode du maximum de vraisemblance. Les chiffres sur les branches représentent le support de 1 000 bootstraps. L'arbre est dessiné à l'échelle, avec des longueurs de branche représentatives du nombre de substitutions par site. A: arbre général des supergroupes A, B, C et D. B: Zoom sur le supergroupe B. C: représentation radiale de l'arbre général. La barre horizontale indique le nombre de substitutions par site. Rouge: souches mutualistes; bleues: souches IC; rose: souches féminisantes, en vert: la souche qui induit la mort des males; en noir: phénotype inconnu.

La dynamique et l'évolution des génomes de *Wolbachia* ont également été illustrées par une phylogenomie fiable basée sur l'ensemble des gènes du core-génome présent en une seule copie sur l'ensemble des génomes. Au delà de l'illustration de la phylogénie bien connue des souches de *Wolbachia* en super-groupe, cette approche phylogénomique a révélé que les bactérie *Wolbachia* du super-groupe B peut être subdivisées en deux sous-groupes, les *Wolbachia* d'isopodes et les souches de *Wolbachia* des insectes. Le reséquençage de la souche *w*Pru sera nécessaire pour valider cette hypothèse. De plus, un nouveau positionnement des souches de *Wolbachia* d'isopodes a été observé. A l'aide de cette nouvelle analyse basée sur des centaines de gènes orthologues, deux sous-groupes apparaissent au sein du clade des *Wolbachia* d'isopodes avec la souche *w*Pet à la racine. Alors qu'un de ces sous-groupes est composé des souches fortement

féminisantes, l'autre sous-groupe est composé de 3 souches dont le phénotype est plus incertain : une souche IC (*w*Dil), une souche dont l'effet demeure inconnue (*w*Bre) et une souche féminisante atypique (*w*Ase de *O. asellus*). En effet, dans les populations d'isopode *O. asellus* infectés, les mâles et les femelles sont infectées par *w*Ase, ce qui est inhabituel pour une souche féminisante (Rigaud et al. 1999) et un effet IC a également été soupçonné (W. LeClec'h, thèse de doctorat). Ainsi, même si il n'y a pas de clade monophylétique pour les souches IC, il y a un clade regroupant les souches induisant une forte féminisation et un clade regroupant les souches de *Wolbachia* "atypiques".
Ces positionnements des souches de *Wolbachia* d'isopodes diffèrent des phylogénies monogéniques précédemment publiées pour lesquelles aucun lien entre la phylogénie et les phénotypes n'avaient pu être mis en évidence (Bouchon et al. 1998; Pichon et al. 2009; Cordaux et al. 2012). Cette analyse illustre un autre exemple de la force des phylogénies construites à partir de centaines de gènes (Dagan 2011). En effet, les séquences de génomes entiers donnent une nouvelle dimension aux études phylogénétiques. Alors que les phylogénies basées sur un seul gène peuvent être biaisées par le choix même du gène d'étude de par ses potentielles recombinaisons, l'étude phylogénomique donne une information fiable basé sur un grand nombre de gènes (Dagan 2011).

Nous avons ensuite étudié les caractéristiques génétiques communes aux souches induisant un même effet phénotypique en se basant sur les 5 principales interactions symbiotiques connues. Cette étude permet l'identification de gènes qui pourraient être éventuellement liées à la symbiose et à l'interaction hôte-symbiote. Pour ce faire, les génomes de *Wolbachia* ont été regroupés par phénotypes pour décoder les différences phénotypiques au niveau des gènes, tels que les gènes partagés uniquement par les souches induisant un même phénotype dans le but de trouver une interprétation génomique au différent phénotype induits par *Wolbachia* et définir un ensemble de facteurs génétiques liés à la symbiose avec la bactérie *Wolbachia*.

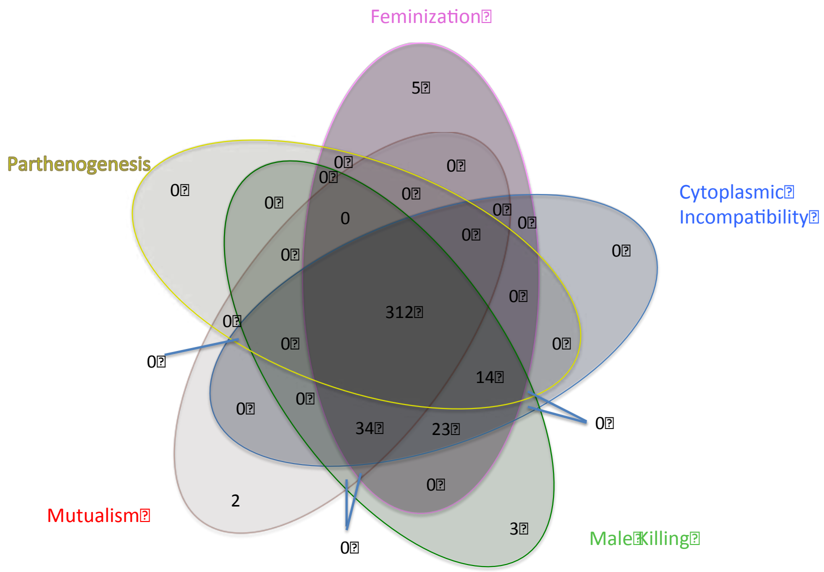


Figure 2: La diversité génomique entre les différentes souches de *Wolbachia* induisant différent phénotypes. Les chiffres représentent le nombre de clusters d'orthologues partagées par les différentes souches induisant les divers phénotypes.

Alors que le core-génome est composé des 312 gènes présents dans toutes les souches indépendamment du phénotype induit, des profils de gènes se sont distingués pour certains phénotypes. En effet, cette étude de génomique comparative a permis l'identification de 2, 3 et 5 gènes spécifiques du mutualisme, du phénotype de la mort des males et de la féminisation. Aucun gène n'a été identifié comme spécifique de l'IC ou de la parthénogenèse. Par ailleurs, très peu de gènes commun à 2 ou plusieurs phénotypes différents ont été identifiés, seule 3 jonctions ont été décrites: (1) 14 gènes étaient présents dans toutes les souches à l'exception des souches mutualistes, (2) 23 gènes étaient communs aux souches induisant l'IC, la féminisation et la mort des males et (3) 34 gènes étaient communs aux souches induisant l'IC, la féminisation, la mort des males et le mutualisme. Cependant, ces chiffres peuvent être plus ou moins surestimés étant donné que le phénotype parthénogenèse n'est représenté que par un génome incomplet.

Cette approche a révélé principalement des protéines intéressantes pour le mutualisme et la féminisation. En effet, une des protéines spécifiques aux souches mutualistes est une protéine WASP, qui est connu pour interagir avec les filaments d'actine (Dean 2011) et être peut être impliquée dans la mobilité des bactéries au sein du cytoplasme des cellules hôtes : chez *B. malayi,* la réinfection des ovaires par *Wolbachia* est retardée au cours du développement. Par ailleurs, les gènes présents dans tous les génomes à l'exception des souches mutualistes devront être plus étroitement étudiés puisque leur

absence peut être liée au style de vie obligatoire de ces souches (Moran 2001; Moran 2002; Foster et al. 2005; Murfin et al. 2012). Concernant les 5 gènes identifiés comme seuls présent dans les génomes des souches induisant une féminisation, le gène X a attiré notre attention car il est localisée en aval du gène secF et en amont du gène yidC qui codent pour des protéines constituant un même système de sécrétion : le système Sec (du Plessis et al. 2011), suggérant que ce gène pourrait être transporté par ce système de sécrétion. Comme le gène X est absent de la souche *w*Pru, qui selon les phylogénies antérieures est plus proche des souches d'insectes, il serait intéressant de vérifier son absence dans le génome des deux autres souches féminisantes d'insectes (E. hecabe et *Z. pullula*), car cela pourrait suggérer un mécanisme de féminisation différent pour les souches féminisantes de ce deuxième sous-groupe.

Étonnamment, aucun gène spécifique au phénotype de la mort des males, de l'IC ou de la parthénogenèse n'a été identifié, même si toutes ces souches interfèrent avec le premier cycle de la division cellulaire de l'embryon (Werren et al. 2008; Cordaux et al. 2011). D'autre part, les ensembles de gènes partagés par les génomes de souches induisant différents phénotypes n'ont pas été étudiés dans cette étude. Cette analyse pourrait mettre en évidence certains processus nécessaires aux mécanismes moléculaires commun à ces phénotypes. En ce qui concerne les 3 gènes initialement identifiés comme spécifiques des génomes des souches induisant la mort des males, aucun de ces gènes codant pour des protéines de prophage n'a été retenu comme gènes potentiellement responsable de ce phénotype.

Chapitre 3 : Etude de l'intéractome entre *Wolbachia* et ses hôtes.

L'analyse de l'interaction entre les protéines hôtes (*B. malayi* ou *A. vulgare*) et de leurs symbiotes *Wolbachia* peut amener à l'identification des effecteurs de la symbiose. Ces protéines peuvent faire partie par exemple de la base biochimique des relations endosymbiotiques obligatoires des *Wolbachia* avec leurs hôtes. De plus, en ce qui concerne les interactions Wolbachia-filaires, l'inhibition de ces interactions protéine-protéine peut révéler une nouvelle cible thérapeutique potentielle contre la maladie de la filariose humaine.

L'analyse des séquences génomiques de *Wolbachia* a révélé la présence de plusieurs systèmes de sécrétion et d'export transmembranaire qui sont probablement impliqués dans la sécrétion de nombreux effecteurs bactériens dans les cellules hôtes, tels que des protéines contenant des motifs de type eucaryote et des protéines membranaires (Wu et al. 2004; Foster et al. 2005; Mavingui et al. 2012; Darby et al. 2012; Duplouy et al. 2013;

Liu et al. 2013). En effet, les systèmes de sécrétion comme le T4SS sont des acteurs clés dans la virulence bactérienne (Baron and Coombes 2007; Durand et al. 2009; Hicks and Galan 2013) et sont très conservés dans les génomes de *Wolbachia* (Pichon et al. 2009). Chez d'autres bactéries endosymbiotiques, ces systèmes de sécrétion sécrètent des effecteurs bactériens dans le cytoplasme des cellules hôtes, dont des protéines avec des motifs "eucaryote-like" (Lin et al. 2007; Rikihisa et al. 2009; Al-Khedery et al. 2012) qui interagissent avec les protéines de l'hôte.

Nous avons d'abord créé le répertoire de certaines protéines connues comme potentiellement impliquées dans la virulence bactérienne par recherche de motifs conservés spécifiques: les motifs répétés ankyrine (ANK) et tetratricopeptide (TPR). Cette étude a été complétée par un BLAST réciproque. En parallèle, nous avons également étudié le rôle potentiel des protéines de *Wolbachia* sélectionnés dans l'interaction avec leurs hôtes par des approches de protéomique et de transcriptomique. Deux classes de protéines ont été ciblées: les protéines impliquées dans les systèmes de sécrétion et des effecteurs bactériens potentiels: les protéines contenant des domaines ANK et des protéines de la membrane externe, qui sont susceptibles d'interagir avec les protéines de l'hôte et de modifier les processus physiologiques cytoplasmiques.

Le filtrage *in silico* des génomes bactériens pour l'identification de potentiels effecteurs bactériens a été axée sur les protéines à motifs "eucaryote-like", qui sont connus pour être impliqués dans les interactions hôte-symbiote. Même si il existe des dizaines de protéines différentes à motif eucaryote, les études récentes se sont concentrées sur les les protéines à domaines ANK. Notre étude a confirmé le nombre exceptionnellement élevé de protéines ANK dans les génomes de *Wolbachia* et en particulier dans les génomes de *Wolbachia* infectant les isopodes, ce qui suggère un rôle particulier de ces protéines, vraisemblablement dans l'adaptation des bactéries à leurs hôtes. Une approche similaire a été effectuée sur les protéines TPR et un profil différent a été observé : la plupart des génomes de *Wolbachia* ont un nombre similaire de protéines TPR à l'exception des deux souches *w*Ri et *w*Suzi qui sont très phylogénétiquement très proches. Etant donné que ces deux types de protéines (ANK et TPR) sont impliqués dans une grande variété de fonctions, la comparaison de ces protéines entre les différentes souches doit être plus détaillée en tenant compte de ces diversités de fonctions. Toutefois, ce résultat préliminaire a illustré l'importance de ne pas sous-estimer l'implication d'autres protéines à motif de type eucaryote tels que les TPR.

En parallèle, nous avons sélectionné certains candidats pour les analyses des interactions protéine-protéine en se basant sur leur implication préalable dans les interactions hôte-symbiote chez d'autres bactéries. Les protéines sélectionnées comme

la plupart des protéines de *Wolbachia* ont été particulièrement difficile à exprimer *in vitro*. Les études d'interactome ont été réalisées en utilisant différentes techniques de protéomique telles que le GST-pull down et le phage display. Les deux techniques utilisées mettent en jeu des expériences qui demandent beaucoup de temps pour leur optimisation et n'ont pas donné de résultats corrélés. Cependant, de nombreux paramètres d'optimisation sont maintenant maîtrisés, ce qui devrait faciliter l'étude des protéines spécifiques de certains phénotypes identifiées dans cette étude.
Par ailleurs, si le mécanisme moléculaire de la symbiose pourrait impliquer des protéines membranaires pour l'amarrage des bactéries, les systèmes de sécrétion jouent un rôle important dans la translocation des effecteurs bactériens à travers la membrane bactérienne vers le cytoplasme de l'hôte. Par conséquent, nous avons recherché et identifié dans les génomes de *Wolbachia* séquencés dans cette étude les gènes des différents systèmes de sécrétion connus et avons identifier deux nouveaux gènes impliqués dans le système de sécrétion tol / pal.

En considérant que la féminisation induite par les souches de *Wolbachia* devrait s'effectuer au cours du développement de l'hôte et de la différenciation sexuelle, qui se produit au stade 4 du développement post-embryonnaire chez *A. vulgare*, des analyses fonctionnelles ont été réalisées au cours de ce développement post-embryonnaire. Après quantification de l'évolution de la charge de *Wolbachia* au cours du développement, des analyses fonctionnelles par approche transcriptomique ont été effectuées sur la souche *w*VulC, sur deux systèmes de sécrétion (T1SS et T4SS) et sur deux effecteurs potentiels de la féminisation: le gène pk2b2 , qui ne s'exprime que dans les souches féminisantes (Pichon et al. 2012) et le gène X précédemment identifiés par analyse *in silico* comme étant seulement présent dans les génomes des souches féminisantes étudiées.
La quantification de la charge bactérienne a révélé une augmentation de la densité de *Wolbachia* au cours du développement jusqu'au stade 4-5 au-delà desquels la densité des bactéries reste constante. Il serait intéressant de comparer ce résultat avec la charge bactérienne des adultes pour voir si la densité de *Wolbachia* a déjà atteint un maximum équivalent à celle observée chez les femelles adultes. Il a cependant déjà été observé que la charge bactérienne chez l'adulte est également variable au cours du cycle de reproduction de l'hôte, la densité de *Wolbachia* augmente avec la maturation des ovocytes (Genty et al. 2013). Nous avons toutefois démontré qu'il n'y a pas de multiplication bactérienne à un stade spécifique du développement, conduisant à

l'hypothèse d'une expression différentielle des effecteurs bactériens induisant la féminisation des mâles en femelles fonctionnelles.

Afin d'effectuer une quantification relative des gènes bactériens ciblés, plusieurs gènes de ménage ont été analysés pour trouver un gène de référence acceptable pour l'étude de l'expression des gènes de *w*Bm et de *w*VulC. Nous avons ainsi démontré que le gène *wsp* pouvait être utilisé comme référence, contairement au gène de l'ARN 16S dont l'amplification n'est pas suffisamment spécifique.

Après avoir démontré que l'expérience de RT-qPCR est techniquement réalisable sur un seul embryon d'*A. vulgare* prélevé le stade 4, nous avons observé qu'aucun des gènes testés n'a montré une expression clairement différente à un stade particulier. Ce résultat est en corrélation avec les observations d'une action continue de la *Wolbachia* féminisante au cours du développement embryonnaire chez le papillon *E. hecabe* (Narita et al. 2007).

Un profil d'expression global a été observé : les deux gènes potentiellement impliqués dans la féminisation (*pk2b2* et gène X) présentent le même profil d'expression que celui du gène du T1SS au cours du développement post-embryonnaire, ce qui nous permet de spéculer sur l'implication de ce système de sécrétion dans la translocation des effecteurs de féminisation. De même, l'analyse de l'expression du T4SS à travers les deux opérons révèle un résultat intéressant: si les opérons sont tous les deux exprimés selon le même profil au cours du développement post-embryonnaire, leur niveau d'expression diffère. Avant cette étude, la co-transcription des deux opérons avait été démontrée par RT-PCR (Felix et al. 2008; Rances et al. 2008) mais jamais quantifiée. Une analyse récente de la régulation du T4SS de la souche *w*Bm a identifié deux facteurs de transcription (wBmxR1 et wBmxR2) qui ne régulent que l'opéron *virB8-virD4*, ce qui suggère la présence de facteur(s) de transcription supplémentaire(s) pour la régulation de l'expression de l'opéron *virB3-virB6* et du gène *virB8* (wBm0641) indépendant de l'opéron (Li and Carlow 2012). Ainsi, si les deux systèmes de régulation semblent synchrones, ils pourraient ne pas avoir la même efficacité.

De la même façon, aucune différence claire du niveau d'expression a été observée pour tous les de gènes du génome *w*Bm testés, entre les différentes étapes du cycle de vie du nématode *B. malayi*. Néanmoins, les deux gènes précédemment identifiés comme étant uniquement présents dans les souches mutualistes (wBm0047 et le WASP wBm0076) présentent une expression plus élevée que le gène de référence *wsp* (wBm0432). Le taux incroyablement élevé de l'expression du gène WASP confirme que ce gène pourrait jouer un rôle important dans le mécanisme de symbiose mutualiste. Malheureusement, l'analyse protéomique de l'interactome n'a pas était possible en raison de la difficulté à

obtenir une quantité exploitable de protéine soluble. Une approche alternative et complémentaire pourrait consister à effectuer un marquage immuofluorescent de ces protéines pour voir si elles sont sécrétées dans le cytoplasme de l'hôte.
En combinant les résultats de transcriptomique et de protéomique, une concordance intéressante a été observée : les gènes présentant le taux d'expression le plus faible, tel que les ANK wBm0394 et wBm582 correspondent aux protéines recombinantes pour lesquelles la production de manière soluble était difficile. Inversement, les gènes avec le taux d'expression les plus élevés comme l'ANK wBm0447 et le *wsp* wBm0100 correspondent aux protéines les plus faciles à exprimer de façon soluble. Ainsi, le taux d'expression d'un gène peut ne pas être uniquement dépendant de l'importance des protéines codées mais aussi dépendant du caractère hydrophobe de la protéine et donc de la difficulté pour les bactéries à synthétiser ses propres protéines. Une exception intéressante à cette observation est la protéine WASP accentuant d'autant plus l'intérêt de cette protéine.

Discussion

Les bactéries *Wolbachia* sont des bactéries endosymbiotiques qui sont héritées principalement maternellement et qui sont communément trouvées chez les arthropodes et les nématodes. Les *Wolbachia* ont développé au moins quatre façons de manipuler la reproduction des arthropodes afin de maximiser leur transmission à la descendance, et ont évolué vers une relation mutualiste avec des nématodes (Werren et al. 2008). Jusqu'à présent, ni le mécanisme moléculaire de manipulation de la reproduction de l'hôte par *Wolbachia*, ni la base de la symbiose mutualiste avec des nématodes n'est connue.
Dans cette étude, nous avons marqué une première étape dans la caractérisation de ces interactions moléculaires en utilisant des approches génomiques, transcriptomiques et protéomiques.

L'étude des facteurs génétiques impliqués dans l'expression phénotypique liés à la présence de *Wolbachia* a été initiée avec le projet de séquençage de plusieurs souches de *Wolbachia*. Cependant, *Wolbachia* étant un endosymbiote obligatoire qui ne peut être cultivé, la première étape consistait à élaborer une procédure efficace capable d'isoler spécifiquement l'ADN de *Wolbachia* de l'ADN de l'hôte. Basé sur la méthode d'enrichissement de séquences ciblées, développée et optimisée dans cette étude, sept souches de *Wolbachia* d'isopodes ont été séquencées et partiellement assemblées. Ces

souches ont été choisies soit en raison de leur phénotype (IC vs. féminisation) ou de leur position dans l'arbre phylogénétique basé sur le gène *wsp*. En raison des technologies NGS utilisées, un grand nombre de séquences courtes a été produits et en raison du grand nombre de séquences répétées dans les génomes de *Wolbachia*, les génomes n'ont pu être que partiellement assemblés. Toutefois, nous avons démontré que la grande majorité de l'information génique était présente, ce qui nous a permis de réaliser des études de génomique comparative fiables.
Même si nous n'avons pas essayé d'appliquer notre méthode de capture à des individus multi-infectés par différentes souches de *Wolbachia*, il y a tout lieu de penser qu'un traitement informatique des données de séquençage permettrait de trier les séquences de chaque souche en se basant notamment sur la profondeur de couverture. Cette méthodologie est bien connue pour les projets de métagénomique qui utilisent le séquençage haut débit (Lindner and Renard 2013).

En parallèle, un projet a été initié pour générer un premier génome complet d'une souche de *Wolbachia* féminisante. Le génome *w*VulC de l'isopode *A. vulgare* a été précédemment partiellement séquencé et publié (Liu et al. 2013), mais en raison de la nature hautement répétitive des séquences et de la présence de prophages dans le génome, l'assemblage n'a pu aller au delà de 10 contigs. Avec cette nouvelle approche de PacBio et de séquençage de longues séquences, nous espérons pouvoir fermer ce génome.

Les études de génomique comparative nous ont permis de générer le premier pan-génome des bactéries du taxon *Wolbachia*, autrement dit d'établir un répertoire de la diversité des gènes présents dans toutes les souches de *Wolbachia*. Cette analyse novatrice de part le nombre de souches prises en compte a souligné la grande diversité génomique et la dynamique des génomes de *Wolbachia* tout en mettant en lumière l'éventuelle impossibilité d'obtenir un jour le répertoire complet des gènes de *Wolbachia.* En effet, une estimation à génomes a montré que chaque nouveau génome séquencé va ajouter de nouveaux gènes au pan-génome. Cette diversité peut s'expliquer par la présence de phages et d'éléments mobiles, qui sont vecteurs de nouveauté génomique, principalement par duplication, recombinaison et transferts horizontaux de gènes (Siguier et al. 2006). Ainsi, il serait intéressant de séquencer les différentes souches infectant un même hôte, en particulier lorsque de la recombinaison a été observée. C'est pourquoi, le séquençage de la souche *w*VulP devrait aider à la compréhension des mécanismes de recombinaison puisque cette souche est supposée

avoir émergé d'une recombinaison entre les deux souches *w*VulM et *w*VulC, dont les génomes ont été séquencés dans cette étude. De même, l'analyse comparative des génomes de souches phylogénétiquement très proche devraient mettre en évidence des événements évolutifs qui illustreraient le concept d'espèce pour *Wolbachia* dont le status a été récemment discuté (Ellegaard et al. 2013) ainsi que les notions de perte et de gain de gènes entre les souches.

Par ailleurs, les génomes des 7 souches de *Wolbachia* d'isopodes actuellement séquencées permettront de comparer la dynamique évolutive des éléments d'insertion chez les isopodes avec celle des insectes, et de spéculer sur le rôle des éléments transposables procaryotes dans la dynamique du génome (Cerveau et al. 2011). Plusieurs études ont déjà mis en évidence des transferts horizontaux de gènes entre différentes souches de *Wolbachia*, comme chez de la fourmi *Formica exsecta* (Reuter and Keller 2003), et des transferts d'ADN inter-spécifiques entre *Wolbachia*, *Cardinium* et *Richettsia*, ouvrant la porte à la compréhension des liens écologiques entre les différents taxons d'endosymbiotes (Duron 2013). Ces connexions écologiques peuvent être encore plus fines dans certains cas, comme illustré par les bactéries *Tremblaya princeps,* endosymbiotes de la cochenille *Planococcus citri*, qui possèdent leur propre endosymbiote, la bactérie *Moranella endobia*. Une étude récente a mis en évidence cette symbiose à trois niveaux basée sur l'acquisition de transferts horizontaux de gènes anciens (von Dohlen et al. 2001; Husnik et al. 2013).

Dans notre étude, nous avons axé l'analyse de génomique comparative sur l'identification d'effecteurs potentiels de la symbiose. À ce jour, il s'agit de la première étude génomique qui compare les 5 principales symbioses à *Wolbachia*. Cette approche a révélé des profils géniques intéressants. Tout d'abord, de nombreux gènes sont communs à l'ensemble des souches de *Wolbachia* à l'exception des souches mutualistes. Ces gènes nécessitent alors d'être étudié plus finement car la clé de la relation de mutualisme peut résider dans leurs fonctions.
Deuxièmement, deux gènes ont été identifiés comme présent uniquement chez les souches mutualistes. L'étude quantitative de l'expression de ces deux gènes chez *w*Bm a révélé un taux d'expression relativement élevé tout au long du cycle de vie des nématodes en particulier pour le gène WASP confirmant la fonctionnalité de ces gènes et leur rôle potentiel dans le mutualisme. Malheureusement, les études d'interaction protéine-protéine effectuées pour identifier les protéines de l'hôte qui interagissent avec la protéine WASP sont restées infructueuses en raison de difficultés techniques.

Toutefois, cette protéine étant connue pour être impliquée dans la polymérisation de l'actine, l'hypothèse selon laquelle la protéine WASP interagirait directement avec les protéines de l'hôte est confortée incitant à poursuivre ces études d'interactome.

Ainsi, ces deux protéines représentent deux candidats potentiels pour définir des cibles thérapeutiques. En effet, cibler les mécanismes moléculaires de la symbiose entre *Wolbachia* et le parasite humain *B. malayi* est d'un intérêt certain pour la recherche de nouvelle cible thérapeutique contre la filariose. Cet axe de recherche est l'un des principaux objectifs du consortium anti- *Wolbachia* (A- WOL) qui vise *Wolbachia* pour identifier de nouveaux traitements plus rapides et non nocifs pour les enfants et les femmes enceintes (Taylor et al. 2012; Taylor et al. 2013). De nombreux axes de recherche ont été ou sont actuellement entrepris, comme l'étude des voies enzymatiques et métaboliques clés chez *Wolbachia* (par exemple, les voies de biosynthèse de l'hème (Wu et al. 2009; Wu et al. 2013) ou de la pyruvate dikinase phosphatase (Raverdy et al. 2008)), l'identification des gènes essentiels à la biologie de *Wolbachia* (Holman et al. 2009) ou la compréhension de l'effet des traitements antibiotiques sur *Wolbachia* (Schiefer et al. 2013).

Par ailleurs, 5 gènes ont été identifiés comme étant uniquement présent chez les souches de *Wolbachia* induisant la féminisation et nous avons démontré que 3 d'entre eux sont activement exprimés dans les tissus hôtes adultes. En raison de sa position particulière dans le génome, entre deux gènes du système de sécrétion Sec, le gène codant pour une protéine hypothétique ici nommé gène X, a particulièrement attiré notre attention. Une première vérification a montré que le gène X semble co-transcrit avec le gène *secF*, ce qui conduit à l'hypothèse de sa translocation par le système de sécrétion Sec à travers la membrane interne suggérant sa prise en charge par un autre système de sécrétion pour sa translocation de la membrane externe. Son expression a été quantifiée au cours du développement post-embryonnaire d'*A. vulgare* et un profil d'expression similaire à celui du deuxième facteur potentiel de féminisation, l'ankyrine pk2b2 (Pichon et al . 2009) et du système de sécrétion de type 1 a été observé. Ce résultat pourrait signifier que nous avons identifié un gène impliqué dans la féminisation, bien que son rôle dans le processus de féminisation doit être plus clairement démontrée. Pour ce faire, nous pouvons utiliser une lignée d'*A. vulgare* du laboratoire EES, appelée Wx*f.* Les individus de cette lignée sont connus pour avoir intégré dans leur génome un fragment du génome de *Wolbachia* (fragment *f*) qui porte la féminisation (Bouchon et al. 2008; Cordaux et al. 2011). Des expériences

préliminaires de PCR et RT-PCR ont indiqué que le gène X est intégré dans le génome des isopodes W*x*f et est exprimé. Ce résultat nous permet de planifier des expériences d'ARN interférents qui inhibent l'expression du gène ciblé, puisque ces expériences ne sont pour l'instant réalisables que pour des gènes eucaryotes. Si l'inhibition de l'expression du gène X induit la suppression de la féminisation de l'hôte alors le gène X pourra être annoté comme facteur de la féminisation. Des observations microscopiques devraient également confirmer sa sécrétion dans le cytoplasme de la cellule hôte. Par ailleurs, les injections étant également facilement réalisable sur les isopodes, il serait intéressant d'injecter dans l'hémolymphe ou dans le marsupium, la protéine du gène X étiquetée de façon fluorescente pour suivre son itinéraire et ses voies d'interactions.

Cette découverte pourrait donc conduire à une caractérisation plus fine de la base moléculaire de la relation symbiotique de la bactérie *Wolbachia* avec l'isopode *A. vulgare* grâce a l'identification du gène X, de ses voies de sécrétion et d'interaction avec l'hôte.

Al-Khedery B, Lundgren AM, Stuen S, Granquist EG, Munderloh UG, Nelson CM, Alleman AR, Mahan SM, Barbet AF (2012) Structure of the type IV secretion system in different strains of Anaplasma phagocytophilum. BMC Genomics 13:678. doi:10.1186/1471-2164-13-678

Baron C, Coombes B (2007) Targeting bacterial secretion systems: benefits of disarmament in the microcosm. Infect Disord Drug Targets 7 (1):19-27

Bouchon D, Cordaux R, Grève P (2008) Feminizing / Wolbachia/ and the evolution of sex determination in isopods. . Insect Symbiosis, pp 273-294

Bouchon D, Rigaud T, Juchault P (1998) Evidence for widespread Wolbachia infection in isopod crustaceans: molecular identification and host feminization. Proc R Soc Lond B Biol Sci 265 (1401):1081-1090

Bull JJ, Rice WR (1991) Distinguishing mechanisms for the evolution of co-operation. J Theor Biol 149 (1):63-74

Carrington LB, Lipkowitz JR, Hoffmann AA, Turelli M (2011) A re-examination of Wolbachia-induced cytoplasmic incompatibility in California Drosophila simulans. PLoS One 6 (7):e22565. doi:10.1371/journal.pone.0022565

Cerveau N, Leclercq S, Leroy E, Bouchon D, Cordaux R (2011) Short- and long-term evolutionary dynamics of bacterial insertion sequences: insights from Wolbachia endosymbionts. Genome Biol Evol 3:1175-1186. doi:10.1093/gbe/evr096

Charles H, Ishikawa H (1999) Physical and genetic map of the genome of Buchnera, the primary endosymbiont of the pea aphid Acyrthosiphon pisum. J Mol Evol 48 (2):142-150

Combes C (1995) Interactions Durables: Ecologie et Evolution du Parasitisme. Masson edn.,

Cordaux R, Bouchon D, Greve P (2011) The impact of endosymbionts on the evolution of host sex-determination mechanisms. Trends Genet 27 (8):332-341. doi:10.1016/j.tig.2011.05.002

Cordaux R, Pichon S, Hatira HB, Doublet V, Greve P, Marcade I, Braquart-Varnier C, Souty-Grosset C, Charfi-Cheikhrouha F, Bouchon D (2012) Widespread Wolbachia infection in terrestrial isopods and other crustaceans. Zookeys (176):123-131. doi:10.3897/zookeys.176.2284

Dagan T (2011) Phylogenomic networks. Trends Microbiol 19 (10):483-491. doi:10.1016/j.tim.2011.07.001

Darby AC, Armstrong SD, Bah GS, Kaur G, Hughes MA, Kay SM, Koldkjaer P, Rainbow L, Radford AD, Blaxter ML, Tanya VN, Trees AJ, Cordaux R, Wastling JM, Makepeace BL (2012) Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis. Genome Res 22 (12):2467-2477. doi:10.1101/gr.138420.112

Dean P (2011) Functional domains and motifs of bacterial type III effector proteins and their roles in infection. FEMS Microbiol Rev 35 (6):1100-1125. doi:10.1111/j.1574-6976.2011.00271.x

du Plessis DJ, Nouwen N, Driessen AJ (2011) The Sec translocase. Biochim Biophys Acta 1808 (3):851-865. doi:10.1016/j.bbamem.2010.08.016

Duplouy A, Iturbe-Ormaetxe I, Beatson SA, Szubert JM, Brownlie JC, McMeniman CJ, McGraw EA, Hurst GD, Charlat S, O'Neill SL, Woolfit M (2013) Draft genome sequence of the male-killing Wolbachia strain wBol1 reveals recent horizontal gene transfers from diverse sources. BMC Genomics 14:20. doi:10.1186/1471-2164-14-20

Durand E, Verger D, Rego AT, Chandran V, Meng G, Fronzes R, Waksman G (2009) Structural biology of bacterial secretion systems in gram-negative pathogens--potential for new drug targets. Infect Disord Drug Targets 9 (5):518-547

Duron O (2013) Lateral transfers of insertion sequences between Wolbachia, Cardinium and Rickettsia bacterial endosymbionts. Heredity (Edinb). doi:10.1038/hdy.2013.56

Ellegaard KM, Klasson L, Naslund K, Bourtzis K, Andersson SG (2013) Comparative genomics of Wolbachia and the bacterial species concept. PLoS Genet 9 (4):e1003381. doi:10.1371/journal.pgen.1003381

English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, Worley KC, Gibbs RA (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLoS One 7 (11):e47768. doi:10.1371/journal.pone.0047768

Felix C, Pichon S, Braquart-Varnier C, Braig H, Chen L, Garrett RA, Martin G, Greve P (2008) Characterization and transcriptional analysis of two gene clusters for type IV secretion

machinery in Wolbachia of Armadillidium vulgare. Res Microbiol 159 (6):481-485. doi:10.1016/j.resmic.2008.05.007

Fellous S, Salvaudon L (2009) How can your parasites become your allies? Trends Parasitol 25 (2):62-66. doi:10.1016/j.pt.2008.11.010

Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B (2005) The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3 (4):e121. doi:10.1371/journal.pbio.0030121

Geniez S, Foster JM, Kumar S, Moumen B, Leproust E, Hardy O, Guadalupe M, Thomas SJ, Boone B, Hendrickson C, Bouchon D, Greve P, Slatko BE (2012) Targeted genome enrichment for efficient purification of endosymbiont DNA from host DNA. Symbiosis 58 (1-3):201-207. doi:10.1007/s13199-012-0215-x

Genty L, Bouchon D, Raimond M, Bertaux J (2013) Wolbachia infect ovaries in the course their maturation: Last minute passengers or priority travellers?

Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, Shukla M, Walenz B, Hill CA, Nene VM, Azad AF, Sobral BW, Caler E (2012) A Rickettsia genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle. J Bacteriol 194 (2):376-394. doi:10.1128/JB.06244-11

Hicks SW, Galan JE (2013) Exploitation of eukaryotic subcellular targeting mechanisms by bacterial effectors. Nat Rev Microbiol 11 (5):316-326. doi:10.1038/nrmicro3009

Holman AG, Davis PJ, Foster JM, Carlow CK, Kumar S (2009) Computational prediction of essential genes in an unculturable endosymbiotic bacterium, Wolbachia of Brugia malayi. BMC Microbiol 9:243. doi:10.1186/1471-2180-9-243

Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson AC, von Dohlen CD, Fukatsu T, McCutcheon JP (2013) Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. Cell 153 (7):1567-1578. doi:10.1016/j.cell.2013.05.040

Iturbe-Ormaetxe I, Woolfit M, Rances E, Duplouy A, O'Neill SL (2011) A simple protocol to obtain highly pure Wolbachia endosymbiont DNA for genome sequencing. J Microbiol Methods 84 (1):134-136. doi:10.1016/j.mimet.2010.10.019

Kent BN, Bordenstein SR (2010) Phage WO of Wolbachia: lambda of the endosymbiont world. Trends Microbiol 18 (4):173-181. doi:10.1016/j.tim.2009.12.011

Kent BN, Funkhouser LJ, Setia S, Bordenstein SR (2011) Evolutionary genomics of a temperate bacteriophage in an obligate intracellular bacteria (Wolbachia). PLoS One 6 (9):e24984. doi:10.1371/journal.pone.0024984

Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, Sinkins SP, Parkhill J (2008) Genome evolution of Wolbachia strain wPip from the Culex pipiens group. Mol Biol Evol 25 (9):1877-1887. doi:10.1093/molbev/msn133

Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SG (2009) The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans. Proc Natl Acad Sci U S A 106 (14):5725-5730. doi:10.1073/pnas.0810753106

Leclercq S, Giraud I, Cordaux R (2011) Remarkable abundance and evolution of mobile group II introns in Wolbachia bacterial endosymbionts. Mol Biol Evol 28 (1):685-697. doi:10.1093/molbev/msq238

Leigh EG, Jr. (2010) The evolution of mutualism. J Evol Biol 23 (12):2507-2528. doi:10.1111/j.1420-9101.2010.02114.x

Li Z, Carlow CK (2012) Characterization of transcription factors that regulate the type IV secretion system and riboflavin biosynthesis in Wolbachia of Brugia malayi. PLoS One 7 (12):e51597. doi:10.1371/journal.pone.0051597

Lin M, den Dulk-Ras A, Hooykaas PJ, Rikihisa Y (2007) Anaplasma phagocytophilum AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. Cell Microbiol 9 (11):2644-2657. doi:10.1111/j.1462-5822.2007.00985.x

Lindner MS, Renard BY (2013) Metagenomic abundance estimation and diagnostic testing on species level. Nucleic Acids Res 41 (1):e10. doi:10.1093/nar/gks803

Liu C, Chen L, Pichon S, Felix C, Cerveau N, Leclercq S, Lesobre J, Moumen B, Bourtzis K, Cordaux R, Bouchon D, Greve P, Garrett R (2013) Genome sequence of the feminizing Wolbachia wVulC from the isopod crustacean Armadillidium vulgare identifies multiple secretion/export systems and putative effectors in a bacterial endosymbiont. in prep

Mavingui P, Valiente Moro C, Tran-Van V, Wisniewski-Dye F, Raquin V, Minard G, Tran FH, Voronin D, Rouy Z, Bustos P, Lozano L, Barbe V, Gonzalez V (2012) Whole-genome sequence of Wolbachia strain wAlbB, an endosymbiont of tiger mosquito vector Aedes albopictus. J Bacteriol 194 (7):1840. doi:10.1128/JB.00036-12

Mavingui P, Van VT, Labeyrie E, Rances E, Vavre F, Simonet P (2005) Efficient procedure for purification of obligate intracellular Wolbachia pipientis and representative amplification of its genome by multiple-displacement amplification. Appl Environ Microbiol 71 (11):6910-6917. doi:10.1128/AEM.71.11.6910-6917.2005

Moran N (2001) The process of genome shrinkage in the obligate symbiont

Buchnera aphidicola. Genome Biol 2 (12):12

Moran NA (2002) Microbial minimalism: genome reduction in bacterial pathogens. Cell 108 (5):583-586

Murfin KE, Dillman AR, Foster JM, Bulgheresi S, Slatko BE, Sternberg PW, Goodrich-Blair H (2012) Nematode-bacterium symbioses--cooperation and conflict revealed in the "omics" age. Biol Bull 223 (1):85-102

Narita S, Kageyama D, Nomura M, Fukatsu T (2007) Unexpected mechanism of symbiont-induced reversal of insect sex: feminizing Wolbachia continuously acts on the butterfly Eurema hecabe during larval development. Appl Environ Microbiol 73 (13):4332-4341. doi:10.1128/AEM.00145-07

Penz T, Schmitz-Esser S, Kelly SE, Cass BN, Muller A, Woyke T, Malfatti SA, Hunter MS, Horn M (2012) Comparative genomics suggests an independent origin of cytoplasmic incompatibility in Cardinium hertigii. PLoS Genet 8 (10):e1003012. doi:10.1371/journal.pgen.1003012

Pichon S, Bouchon D, Cordaux R, Chen L, Garrett RA, Greve P (2009) Conservation of the Type IV secretion system throughout Wolbachia evolution. Biochem Biophys Res Commun 385 (4):557-562. doi:10.1016/j.bbrc.2009.05.118

Pichon S, Bouchon D, Liu C, Chen L, Garrett RA, Greve P (2012) The expression of one ankyrin pk2 allele of the WO prophage is correlated with the Wolbachia feminizing effect in isopods. BMC Microbiol 12:55. doi:10.1186/1471-2180-12-55

Rances E, Voronin D, Tran-Van V, Mavingui P (2008) Genetic and functional characterization of the type IV secretion system in Wolbachia. J Bacteriol 190 (14):5020-5030. doi:10.1128/JB.00377-08

Raverdy S, Foster JM, Roopenian E, Carlow CK (2008) The Wolbachia endosymbiont of Brugia malayi has an active pyruvate phosphate dikinase. Mol Biochem Parasitol 160 (2):163-166. doi:10.1016/j.molbiopara.2008.04.014

Reuter M, Keller L (2003) High levels of multiple Wolbachia infection and recombination in the ant Formica exsecta. Mol Biol Evol 20 (5):748-753. doi:10.1093/molbev/msg082

Rigaud T, Moreau J, Juchault P (1999) Wolbachia infection in the terrestrial isopod Oniscus asellus: sex ratio distortion and effect on fecundity. Heredity (Edinb) 83 (4):469-475

Rikihisa Y, Lin M, Niu H, Cheng Z (2009) Type IV secretion system of Anaplasma phagocytophilum and Ehrlichia chaffeensis. Ann N Y Acad Sci 1166:106-111. doi:10.1111/j.1749-6632.2009.04527.x

Salzberg SL, Dunning Hotopp JC, Delcher AL, Pop M, Smith DR, Eisen MB, Nelson WC (2005) Serendipitous discovery of Wolbachia genomes in multiple Drosophila species. Genome Biol 6 (3):R23. doi:10.1186/gb-2005-6-3-r23

Schiefer A, Vollmer J, Lammer C, Specht S, Lentz C, Ruebsamen-Schaeff H, Brotz-Oesterhelt H, Hoerauf A, Pfarr K (2013) The ClpP peptidase of Wolbachia endobacteria is a novel target for drug development against filarial infections. J Antimicrob Chemother 68 (8):1790-1800. doi:10.1093/jac/dkt105

Siguier P, Filee J, Chandler M (2006) Insertion sequences in prokaryotic genomes. Curr Opin Microbiol 9 (5):526-531. doi:10.1016/j.mib.2006.08.005

Sun LV, Foster JM, Tzertzinis G, Ono M, Bandi C, Slatko BE, O'Neill SL (2001) Determination of Wolbachia genome size by pulsed-field gel electrophoresis. J Bacteriol 183 (7):2219-2225. doi:10.1128/JB.183.7.2219-2225.2001

Taylor MJ, Hoerauf A, Townson S, Slatko BE, Ward SA (2013) Anti-Wolbachia drug discovery and development: safe macrofilaricides for onchocerciasis and lymphatic filariasis. Parasitology:1-9. doi:10.1017/S0031182013001108

Taylor MJ, Voronin D, Johnston KL, Ford L (2012) Wolbachia filarial interactions. Cell Microbiol. doi:10.1111/cmi.12084

Thomas F, JF G, F R (2007) Ecologie et évolution des systemes parasites. DE BOECK,

Thrall PH, Hochberg ME, Burdon JJ, Bever JD (2007) Coevolution of symbiotic mutualists and parasites in a community context. Trends Ecol Evol 22 (3):120-126. doi:10.1016/j.tree.2006.11.007

VanValen L (1973) A New Evolutionary Law. Evolutionary Theory 1:1-30

von Dohlen CD, Kohler S, Alsop ST, McManus WR (2001) Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. Nature 412 (6845):433-436. doi:10.1038/35086563

Weeks AR, Turelli M, Harcombe WR, Reynolds KT, Hoffmann AA (2007) From parasite to mutualist: rapid evolution of Wolbachia in natural populations of Drosophila. PLoS Biol 5 (5):e114. doi:10.1371/journal.pbio.0050114

Werren JH, Baldo L, Clark ME (2008) Wolbachia: master manipulators of invertebrate biology. Nat Rev Microbiol 6 (10):741-751. doi:10.1038/nrmicro1969

Wu B, Novelli J, Foster J, Vaisvila R, Conway L, Ingram J, Ganatra M, Rao AU, Hamza I, Slatko B (2009) The heme biosynthetic pathway of the obligate Wolbachia endosymbiont of Brugia malayi as a potential anti-filarial drug target. PLoS Negl Trop Dis 3 (7):e475. doi:10.1371/journal.pntd.0000475

Wu B, Novelli J, Jiang D, Dailey HA, Landmann F, Ford L, Taylor MJ, Carlow CK, Kumar S, Foster JM, Slatko BE (2013) Interdomain lateral gene transfer of an essential ferrochelatase gene in human parasitic nematodes. Proc Natl Acad Sci U S A 110 (19):7748-7753. doi:10.1073/pnas.1304049110

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA (2004) Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2 (3):E69. doi:10.1371/journal.pbio.0020069

RESUME

Les *Wolbachia* sont des alpha-proteobacteries présentes chez de nombreux arthropodes et nématodes filaires. Ces bactéries héritées maternellement induisent chez leurs hôtes des phénotypes allant du parasitisme au mutualisme, avec le long de ce continuum des phénotypes tels que la féminisation (F), l'incompatibilité cytoplasmique (IC) ou la mort des males. *Wolbachia* est ainsi un modèle particulièrement intéressant pour étudier les différents types de relations symbiotiques. Chez *Brugia malayi*, comme pour les autres nématodes filaires, *Wolbachia* vit en symbiose obligatoire avec son hôte. L'élimination de la bactérie par des traitements antibiotiques entraîne une perte de fertilité voire la mort du nématode. Chez l'isopode terrestre *Armadillidium vulgare, Wolbachia* induit la féminisation des mâles génétiques en femelles fonctionnelles entraînant des biais de sex-ratio vers les femelles dans la descendance. Pour comprendre les mécanismes impliqués dans ces deux symbioses, nous avons mis au point une nouvelle méthode de capture pour isoler l'ADN de *Wolbachia* et séquencer 7 souches de *Wolbachia* d'isopodes (F et IC). Une étude de génomique comparative a permis d'établir un premier pan-génome des bactéries du genre *Wolbachia* et d'identifier 2, 5 et 3 gènes présents seulement chez les souches mutualistes, féminisantes ou induisant la mort des males. L'expression des gènes potentiellement impliqués dans la féminisation ou le mutualisme a été étudiée au cours du développement de l'hôte. L'étude de l'interactome protéique bactérie-hôte a ensuite été initiée en utilisant comme appât des protéines bactériennes à domaines eucaryotes en vue d'identifier les cibles de *Wolbachia* chez l'hôte.

MOTS-CLES

Wolbachia, symbiose, séquençage de génomes bactériens, NGS, pan-genome, génomique comparative, effecteurs bactériens, interactome hôte/symbiote, transcriptomique, *Armadillidium vulgare*, crustacés isopodes, féminisation, *Brugia malayi*, nématodes, mutualisme.

ABSTRACT

Bacteria of the genus *Wolbachia* are gram-negative alpha-proteobacteria present in many arthropods and filarial nematodes. These obligate intracellular bacteria are maternally inherited and induce a large number of phenotypes across the symbiosis continuum from mutualism to parasitism, including feminization (F), cytoplasmic incompatibility (CI) or male killing. Studying *Wolbachia* symbioses is therefore of particular interest in the investigation of symbiotic relationships. In *Brugia malayi* and other filarial nematodes, they are obligate, leading to a loss of worm fertility, and eventual death upon their depletion with certain antibiotics. In arthropods, they are parasitic. In the isopod crustacean *Armadillidium vulgare* they cause feminization when present: genetic males develop as functional females leading to female biased sex-ratio progenies. In order to understand the molecular mechanisms of these two symbioses, we set up a new procedure to capture *Wolbachia* DNA and perform whole-genome sequencing on 7 *Wolbachia* strains, symbionts of isopods (F & CI). Comparative genomics led to the establishment of the *Wolbachia* pan-genome as well as the identification of phenotype related gene patterns. We identified 2, 5 and 3 protein-coding genes that are only found in mutualist, feminizing and male killing strains, respectively. Expression of genes potentially involved in feminization and mutualism were also analyzed throughout host post-embryonic development. A host-symbiont interactome approach was then initiated by protein-protein interaction studies using bacterial proteins with eukaryote like motifs as bait in order to identify *Wolbachia* host targets involved in symbiosis.

KEYWORDS

Wolbachia, symbiosis, whole-genome sequencings, NGS, pan-genome, comparative genomics, bacterial effectors, host-symbiont interactome, transcriptomics, *Armadillidium vulgare*, isopod crustaceans, feminization, *Brugia malayi*, nematodes, mutualism.