EVOLUTION AND GENE EXPRESSION OF *WOLBACHIA* IN *D. MELANOGASTER*

A thesis submitted to the University of Manchester
for the degree of Doctor of Philosophy
in the Faculty of Life sciences

2016

By
Florence Gutzwiller

Contents

Word Count: 20501

List of Tables

List of Figures

The University of Manchester

Doctor of Philosophy

Florence Gutzwiller

EVOLUTION AND GENE EXPRESSION OF *Wolbachia* IN *D. melanogaster*

28 March, 2016

Wolbachia is a facultative endosymbiont infecting *D. melanogaster*, among many other arthropod species. In *D. melanogaster*, *Wolbachia* is a reproductive manipulator, but also gives some benefits to its host, such as protection against viruses (1; 2; 3; 4). We used open access libraries of *D. melanogaster* resequencing data (3; 5; 6; 7; 8; 9; 10; 11; 12) and yet unpublished sequences from our lab, to study the phylogenomics and geographical diversity of *Wolbachia* in *D. melanogaster* from five continents. We confirmed the clade structure of *Wolbachia* infecting *D. melanogaster* and the vertical transmission of *Wolbachia* from a single ancestral infection in *D. melanogaster* for almost all samples as shown in previous studies (3; 13; 14; 15; 15; 16; 17). However, for the first time, we found some discrepancies between *Wolbachia* and mitochondrial clades, indicating potential horizontal transfers. The geographical distribution of the clades, and the variation in proportion of uninfected flies between clades suggest that the clade composition and infection rate in different populations is the result of trade-off between local adaptations and *Wolbachia* transmission efficiency. After confirming the tight coevolution between *Wolbachia* and *D. melanogaster*, and the existence of diversity in *Wolbachia* strains, we analysed the gene expression of *Wolbachia* in different clades and host backgrounds. We first looked at the variations in gene expression of *Wolbachia w*Mel across the developmental life cycle of the ISO1 line of *D. melanogaster*. We found 93 *Wolbachia* genes differentially expressed across the *D. melanogaster* life cycle and/or between adult males and females. Most of the stage-specific genes follow a similar pattern of expression with up-regulation after embryonic stages and stronger up-regulation in early larvae and late pupae, and most of sex-specific genes were up-regulated in males. Then, we analyzed *Wolbachia w*MelCS gene expression in a nos-gal4, UAS-DCR2 driver mutant line of *D. melanogaster* (DCR2), in embryo, ovaries and ovarectomized carcasses. We found 118 genes differentially expressed between embryo and female carcasses, a lot of them (46) being also differentially expressed between embryo and ovaries, and up-regulated in embryo. A small number of the genes differentially expressed between embryo and female carcasses are up-regulated in female carcasses and are, for most of them, also found differentially expressed between ovaries and carcasses in the same dataset, and between embryo and female adult in *Wolbachia w*Mel infecting the ISO1 line of *D. melanogaster*. Both RNAseq analyses showed that most of the *Wolbachia* genes were expressed and that the expression level of most of the genes was relatively constant across the life cycle stages, sex and tissues tested.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy[1], in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations[2] and in The University's policy on presentation of Theses.

[1] http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487
[2] http://www.manchester.ac.uk/library/aboutus/regulations

Acknowledgements

I would like to thank my supervisor Casey Bergman for his trust, his guidance, and teachings during this Ph.D., and also for the understanding and support he showed to his students through those years, that extends beyond academic life. I would also like to thank my advisor Richard Preziosi for his insightful comments and discussion about my work. Many thanks to Iryna Kozeretska and Danny Miller for producing some of the data that I used in this thesis. I also want to thank people from the lab, Pedro, Manee, Mar, Michael, and Raquel, who helped me to learn, and for the interesting discussions.

I also want to thank all the people who made these years in Manchester a great time, with bubbles and barbecues despite the rain. I want to thank especially Matias for his help and indefectible support. I also want to thank Tamara and Tania for the symbiosis in writing, which made this process much more bearable, and also Thomas, Pedro, Carolina, Felipe, Cesar, Cristobal, Wamz, Alicia, Sotiris, Fudge, and all the people I met during my time here and who helped me in many ways. Many thanks to my parents and my sister, for being always encouraging and supportive, and helping me in many ways throughout my studies, making this journey easier.

Finally I want to thank the N.E.R.C. for funding this project, and the O.M.P. for the follow-up throughout the Ph.D.

Chapter 1

Introduction

In this chapter, I will start by introducing the model organisms used in this thesis: *Drosophila melanogaster* and the endosymbiotic bacteria *Wolbachia*. Then, I will present the basis and history of the symbiotic relationship between these two organisms.

1.1 *Drosophila melanogaster*: a model for studying host-symbiont interactions

1.1.1 *D. melanogaster* life history

D. melanogaster is a model insect in the order *Diptera* and family *Drosophilidae*; it is a human commensal species, and has a worldwide cosmopolitan range (18). Its developmental cycle takes around eight days after hatching, at 25°C (18). Adult, *D. melanogaster* can live around 30 days (18). Females can lay up to 100 eggs a day, that will hatch after 22-24 hours and go through three larval stages (18). At 25°C, the first two larval stages last around 24 hours and the third larval stage lasts 48 hours (18). After four days of pupation, adults emerge and will start mating around 12 hours after eclosion (18). In the wild, larvae usually feed on decaying vegetable matter such as rotten fruits (19).

D. melanogaster is a well known model organism and has been shown to be a useful tool to study host-microbe interaction (20; 21; 22; 23; 24; 25). Indeed, its microbiome is well characterized and is relatively simple, composed of few taxa (21; 24; 25; 26). Moreover, *D. melanogaster* genome, physiology, and behaviour is extensively documented and studied, making it a good model for studying the

effects of symbionts.

1.1.2 *D. melanogaster* biogeography

To interpret genetic data of *D. melanogaster* symbionts, like *Wolbachia*, it is important to have an understanding of the biogeography and demography of this species (27). While it is found worldwide, the site of origin of *D. melanogaster* is thought to be in sub-Saharan Africa (6; 28; 29). This is based on the fact that all species in the *melanogaster* species subgroup have an African origin, and that levels of genetic diversity observed in sub-Saharan African populations of *D. melanogaster* are higher than elsewhere ((6; 27; 30; 31) reviewed in (29; 32)). In Zimbabwe, *D. melanogaster* populations are very diverse at a molecular level (31; 32) and show an excess of derived mutations (33; 34). A study based on populations from Europe and Africa suggests that expansion of *D. melanogaster* in Africa started from Zimbabwe, 60,000 years ago (35). However a more recent study by Duchen *et al.* (30), proposed that a bottleneck model explains the diversity of the Zimbabwe population better than a simple expansion model. Moreover, a study of sequence variation in 25 *D. melanogaster* populations (21 from Sub-Saharan Africa) suggests that the Uganda population is closest to the original range, from which flies spread to western and eastern Africa and later Eurasia (33).

It is currently thought that *D. melanogaster* spread from Africa first to Eurasia and that secondary migrations from Europe allowed the colonisation of Australia, Japan, America and Pacific islands with the help of human activities in the last 500 years (30; 36). The date of the bottleneck event corresponding to European colonisation was first thought to be between 6,400 and 16,000 years ago, depending on the study (reviewed in (29)). However, most recently, Duchen *et al.* (30) estimated the time of split between European and African population to be 19,000 years ago. The duration of the bottleneck was estimated by Li and Stephan (35) to last about 340 years.

Caracristi and Schlötterer (27) found the phylogenetic position of North American populations to be between European and African populations. These authors interpreted this finding to be consistent with a scenario of admixture between

African and European populations (27). More recently Duchen *et al.* (30) confirmed these findings by showing that a model including admixture between European and African populations best explains North American population diversity. Duchen *et al.* determined the proportion of European and African ancestry in North American populations to be 85% and 15% respectively (30). Admixture is also seen in African populations with evidence of non-African sequences in many African genomes (6). The proportion of admixture in African populations varies between populations and across the genome, which suggests a non-neutral introgression process (6).

The current scenario for recent biogeographic events in *D. melanogaster* is summarised in Figure 1.1. In red is the African expansion, likely from an origin population close to Uganda (33). In orange is the migration to Eurasia estimated to have occurred around 19,000 years ago (30). In yellow are recent colonisations of North America from both Europe and Africa during the last 500 years (30). In green are the introgression events of non-African populations back into African ones. These migrations likely occurred several times after the expansion of *D. melanogaster* out of Africa (6). Biogeographic events for other regions of the world have not yet been investigated in detail.

1.1.3 *D. melanogaster* genome

The *D. melanogaster* genome is organised as four chromosomes comprised of three autosomes (two large and one small) and a large X chromosome in females and one X and Y in males. The genome size is around 180 mega-bases (Mb) in total with about 120 Mb of euchromatin and 60 to 100 Mb of heterochromatin (the Y chromosome being 40 Mb long and only constituted of heterochromatin) (37; 38). Euchromatin contains most of the genes and corresponds to the major part of the reference genome sequence. Heterochromatin is mainly composed of repetitive sequences that are challenging for the whole genome shotgun assembly and therefore underrepresented in the reference genome sequence (38).

The *D. melanogaster* mitochondrial genome is a circular, double stranded molecule of around 19 kilo-bases (Kb) in length (39; 40). It has a biased A+T composition and contains a region particularly rich in A+T, known as the control region (36). The control region includes a sequence involved in replication

Figure 1.1: **Summary of the recent biogeographic event in *D. melanogaster*.** This figure summarise the main migration events in recent *D. melanogaster* history. In red: African expansion, in orange: migration to Eurasia, in yellow: recent colonisation of North America and in green: introgression events of non-African populations back into African ones. See text for details.

and also sequences having a role in transcription (40). This sequence is also rich in variable repetitive sequences, that makes the mitochondrial DNA size variable within individuals (40; 41). Longer mitochondrial genomes are favourably transmitted to the offspring (40; 41).

While, it is not used to resolve precisely the phylogeography of *D. melanogaster* (32; 42), the mitochondrial genome, due to its maternal inheritance, has been successfully used in several studies to investigate evolution of *D. melanogaster* and *Wolbachia* (14; 15; 16; 43).

The availability of the first *D. melanogaster* complete reference genome sequence (37) and advances in high throughput sequencing enabled several major resequencing projects such as the *Drosophila* Genetic Reference Panel (DGRP) and the *Drosophila* Population Genomics Project (DPGP2 and DPGP3). The DGRP aimed to study variation in one outbred population by comparing genomes from 192 inbred strains coming from North Carolina, USA (5). The DPGP 2 sequenced 161 genomes from 22 sub-Saharan wild population and one European

population in order to study diversity and demographic history of *D. melanogaster* in Africa. The DPGP3 added 197 samples from Zambia to the DPGP2 project (6). Many other resequencing projects were published (7; 9; 10; 11; 12) and further studies used these available projects and added new resequencing data to expand the knowledge of the *D. melanogaster* genome. For instance, Lack *et al.* in 2015 (8) added new sequences from sub-Saharan populations to existing projects (including DPGP3 and DGRP), reaching a sample size of 623 genomes in a project called The Drosophila Genome Nexus (DGN). The aim of this project was to develop a methodology to assemble *D. melanogaster* genome sequence, minimising the methodological differences between data sets (8). Together with the raw data from these projects (DGRP, DPGP and DPGP3), we used data from several other study. Grenier *et al.* (7) sequenced 85 *D. melanogaster* samples from populations in Australia, China, Netherlands, USA and Zimbabwe with the aim of modeling metabolic regulation by using natural genetic variation. The data published by Grenier *et al.* (7) was also used by Early *et al.* (15), who showed monophyly of *Wolbachia* in *D. melanogaster* and geographical variation of infection rate. Another 18 *D. melanogaster* genomes were sequenced by King *et al.* (9) from an advanced generation cross between eight inbred founders in order to improve the understanding of the genetics of complex traits in *D. melanogaster*. A project by Campo *et al.* (10) sequenced 37 *D. melanogaster* genomes from two USA populations to study demographics and adaptations of *D. melanogaster* in north America. Versace *et al.* (11) sequenced the genome of 48 flies in a study using experimental evolution to understand fitness dynamics of *Wolbachia* in *D. melanogaster*. They showed how cold environment could influence relative competitiveness between different *Wolbachia* strain in a same host background (11). 56 more genomes were sequenced by Bergman and Haddrill (12) from Ghana, France, and USA, and finally, 9 lab strains were sequenced by Chrostek *et al.* (3) to determine the genome sequence of the reference genotypes of *Wolbachia* infecting *D. melanogaster*. From those projects and with the addition of unpublished sequences, we used a total of 839 samples to study the phylogenomics of *Wolbachia* in *D. melanogaster*.

D. melanogaster, as a model organism, is also part of the two organisms used by the modENCODE project with the aim of decoding the genomic information and characterizing functional networks (44). As part of the modENCODE

project, the *D. melanogaster* transcriptome was sequenced across the fly life cycle, from embryo to adult (45; 46; 47). As the *D. melanogaster* line used in this RNA-seq project was infected with *Wolbachia*, we were able to use the produced dataset to study the dynamics of *Wolbachia* gene expression across the host life cycle.

1.2 *Wolbachia pipientis*: a model for studying endosymbiosis

Wolbachia is an endosymbiotic α-proteobacteria from the *Rickettsiae* family infecting many species of nematodes and arthropods (48; 49). The transmission of *Wolbachia* in its host is mainly vertical through ovaries but horizontal transfer of the bacteria between two organisms and cases of paternal transmission have been shown (48). The *Wolbachia* clade has been separated into seven main supergroups (A, B, C, D, E, F, H), plus a few unclassified *Wolbachia* that have also been detected (50). The impact of *Wolbachia* interaction with its hosts ranges across the spectrum from mutualist to parasite.

On one end of the spectrum, *Wolbachia* from groups C and D that infect nematodes are obligate (used here as obligate from the standpoint of the host) and mutualist symbionts, providing hosts with essential amino acids (51; 52). This makes *Wolbachia* a potential target for new treatments for nematode infections such as river blindness (53; 54). As an obligate mutualist endosymbiont, *Wolbachia* in nematodes has a reduced genome compared to the other supergroups (50; 55). Gerth *et al.,* (50) showed that those two supergroups were paraphyletic, suggesting that *Wolbachia* switched from arthropod to nematode hosts twice and that the genome reduction in supergroups C and D happened independently.

Wolbachia from supergroup F can infect both arthropods and nematodes, and can act both as mutualist and reproductive parasite (56; 57; 58). Two *Wolbachia* screening studies found new groups among the arthropod-only infecting groups: supergroup E which infects collembola (springtail), and supergroup H which infects termites (59; 60).

The A and B *Wolbachia* supergroups infect arthropods and display more diverse interactions from mutualist to parasite and are even pathogenic in some laboratory strains (4; 56; 61; 62; 63; 64; 65). Those two groups form a monophyletic clade, suggesting that the capacity of infecting and adapting to a wide range of arthropod hosts evolved only once (50). The arthropod infecting *Wolbachia* are mainly facultative for the host and often involve reproductive parasitism in order to increase maternal transmission (65). Several reproductive manipulation strategies are implemented by the bacteria, each optimizing the reproductive success of infected females: male killing, male feminization, induction of parthenogenesis, and cytoplasmic incompatibility (48; 65; 66). Cytoplasmic incompatibility is a type of reproductive manipulation, where infected males are unable to produce a viable offspring with uninfected females, while other crossings are possible. Cytoplasmic incompatibility-inducing *Wolbachia* have been envisaged as a biological control of insects pest to reduce the number of viable offspring by releasing infected males into natural populations, and therefore reducing the reproductive chances of uninfected males by introducing "sterile" competitors (66). This method has been attempted in insect vectors of human diseases such as *Culex pipiens* and in agricultural pests such as *Ceratitis capitata* (67; 68).

Among arthropods, *Wolbachia* infects a very large proportion of arthropod species. It has been estimated that 66% of arthropod species harbour *Wolbachia* (69), however, more recent studies have lowered this estimate to 40% (70). Due to its effects on sexual reproduction, such as cytoplasmic incompatibility, *Wolbachia* is also thought to potentially play a role in speciation by inducing reproductive isolation (48). Besides reproductive manipulation, *Wolbachia* in insects causes a wide range of effects such as behavioural modification, impact on nutrition, protection against viruses (4; 56; 61; 62; 63; 71; 72). The fact that *Wolbachia* can protect some insects against viruses has been seen as a potential tool for fighting dengue transmission through mosquitos (73; 74; 75; 76). Although *Wolbachia* is mainly a facultative symbiont in insects, it has been found in several mutualistic obligate symbioses such as in a parasitic wasp and a bedbug (56; 61).

1.3 *Wolbachia* in *Drosophila melanogaster*

1.3.1 Phenotypic effects of *Wolbachia* in *D. melanogaster*

Wolbachia in *D. melanogaster* has been found to induce low-level cytoplasmic incompatibility (1; 2). When it induces cytoplasmic incompatibility, *Wolbachia* in the male pronucleus causes a delay in *de novo* nucleosome assembly with maternal core histones and a delay in replication in the early embryo (77). These effects leads to a delay in Cdk1 activation and improper paternal chromosome segregation (78; 79). Lethality is rescued in eggs containing *Wolbachia* (48). However, the molecular factors that cause the modification and rescue are unidentified.

Wolbachia cytoplasmic incompatibility induction varies with strains and hosts (80). The *Wolbachia* endosymbiont of *D. melanogaster* (*w*Mel) cytoplasmic incompatibility is moderate in its host, but high in *D. simulans*, while other strains of *Wolbachia* infecting various *Drosophila* species can display stronger or no cytoplasmic incompatibility (80).

Besides sexual parasitism, *Wolbachia* in *Drosophila* can cause other phenotypes likely taking place in different specific tissues (62; 64; 81). Zug and Hammerstein (82) show for instance, that in *D. melanogaster*, *Wolbachia* infection helps reducing oxidative stress caused by external factors (83). *Wolbachia* was also found to increase *D. melanogaster* fecundity in iron-restricted and iron-overloaded diets (62). Garlapow *et al.* also showed that *Wolbachia* infection increases food intake in females of *D. melanogaster* (84).

Distinct *Wolbachia* variants infecting *D. melanogaster* were shown to provide protection against viruses, at different levels (3; 4). The protection against viruses was associated with higher replication rate of the bacteria and reduced longevity of *D. melanogaster* (3; 4). The virus protection is not a side effect of a *Wolbachia*-triggered immune reaction (85). While *Wolbachia* doesn't trigger an immune response in long-term associations, it might induce an immune response in new and/or virulent infections such as *w*MelPop which induce mortality in its host by over-replicating (3; 74; 86; 87).

Wolbachia infection in *D. melanogaster* is also linked with physiological and behavioural modification such as chill coma recovery, sleep time during the day, ethanol sensitivity, locomotor startle response and olfaction response (88). The chill recovery influence of *Wolbachia* is not systematic and observed only in a few lines, between which the effects on recovery time vary (88). The starvation

survival is affected by *Wolbachia* in different ways in different lines as well, some of which show higher longevity before tetracycline treatment and some showing no effect or higher longevity after the antibiotic treatment (88).

Those positive effects, with the addition of the reproductive manipulation, make *Wolbachia* a very efficient symbiont in *D. melanogaster*, able to persist in the natural populations. In an experimental evolution experiment, Versace *et al.* (11) showed that when a *D. melanogaster* population was partially infected with *Wolbachia* in the lab, the bacteria always went to fixation, no matter the initial frequencies and temperature (11).

However, *Wolbachia* can also produce negative effects in *D. melanogaster*. Vale and Jardine observed an increased somnolence in both males and females and an increased lethargy in males of *D. melanogaster* infected with *Wolbachia* (89). Likewise, some variants of *Wolbachia*, *w*MelCS-like have been shown to have a deleterious effect on the fly longevity (3).

1.3.2 *Wolbachia* distribution inside *D. melanogaster*

Being maternally inherited, *Wolbachia* is present in female germline in order to infect the next generation (65). The distribution of *Wolbachia* in embryo is determined during oogenesis stages 8 to 10 and does not change until late embryogenesis (90).

Differences in the distribution of *Wolbachia* can be observed in embryo depending on the *Wolbachia* strain (90). In the *w*Mel *Wolbachia* strain, the bacteria are concentrated in the germ plasm, located in the posterior part of the embryo (90). *Wolbachia w*Mel positioning in the posterior cortex of the oocyte and redistribution into the posterior region is done by cargo transport by the kinesin-1 motor along the microtubules (91; 92; 93). The distribution stay the same when *Wolbachia w*Mel is transinfected to *D. simulans*, suggesting a bacterial control of the distribution pattern (90). However, another study shows that in *Trichogramma*, the normal *Wolbachia* distribution in embryo was lost in naturally uninfected lines transinfected with *Wolbachia*, suggesting that the *Wolbachia* distribution can be context-dependent (94). The density of *Wolbachia* in the posterior part of the embryo, rather than in the whole embryo, has been proposed to affect cytoplasmic incompatibility intensity (90).

At the end of *D. melanogaster* embryogenesis, *Wolbachia* located in the germ

plasm start colonizing other tissues (90). The intracellular distribution of *Wolbachia* during development is dynamic and guaranties infection of the germline. Somatically infecting *Wolbachia* can migrate to the germline to be passed to the next generation (90; 91; 95). Indeed, *Wolbachia* present in the gonads are usually in only a fraction of the total bacterial load (96). *Wolbachia* has been localized to the *Drosophila* brain during larval and adult stages (97). Albertson et al., (88) found that, in *D. melanogaster*, *Wolbachia* *w*Mel is always present in the brain, not in axons but in, or next to, the cell bodies. In brains of *D. melanogaster* infected with the pathogenic variant *w*MelPop, the bacteria density is higher and seems to form extracellular clusters (88). The position of *Wolbachia* in the brain is dependent on the *Wolbachia* strain (88). In *w*Mel, the bacteria are present evenly in the central brain except for the soma surrounding the optic lobe but the titer is affected also by the host (88). The amount of infected cells is lower than 40% in *D. melanogaster* infected with *w*Mel (88). *Wolbachia* is also found in the ventral nerve cord in similar amount than it is in the brain (88). The infection of *D. melanogaster* brain by *Wolbachia* can explain some of the behavioural modifications induced in *D. melanogaster* (88). Besides the brain cells, *Wolbachia* is also always found in *D. melanogaster* fat bodies (88).

In males, meiosis starts in testes of the third instar larvae and *Wolbachia* can be observed in sperm cysts at this stage (98). During spermatogenesis, *Wolbachia* inside the spermatocyte aggregate near the periphery when the spermatids begin elongation (98). The presence of *Wolbachia* in male gonads is probably linked with the modification of sperm cells involved in cytoplasmic incompatibility, as cytoplasmic incompatibility is driven by a delay in the de novo nucleosome assembly with maternal core histones (77). In *D. simulans* infected with *Wolbachia* wRi, the number of *Wolbachia* in testes increases during the pupal development and decreases with age in adult males, while the density of *Wolbachia* in the somatic cells increases (98). It has also been shown by Chrostek *et al.* that overall density of *Wolbachia* increases with age in *D. melanogaster* (3).

1.3.3 *Wolbachia* genomics in *Drosophila*

The genome of *Wolbachia* *w*Mel, endosymbiont of *D. melanogaster*, was fully sequenced by Wu *et al.* (99). It consists of a circular double stranded DNA molecule of 1,267,782 base pairs (bp) which is AT rich with a G+C content of 35%. The coding sequences represent 85.4% of the total length and are distributed

across 1195 genes (99). A characteristic of the *w*Mel genome is that it contains a large proportion of repeated sequences, among the highest for a prokaryotic genome, many of which correspond to mobile elements (99). Other endosymbionts like *Buchnera*, and *Wigglesworthia* show less repetitive DNA and mobile elements (99; 100; 101).

Little is known about the functions of *Wolbachia* genes and how they relate to the phenotype induced in their hosts. Studies based on comparative genomics have brought attention to two main features of the *Wolbachia* genome: a eight-gene region (called octomom) that is highly triplicated or deleted in *w*MelPop and *w*MelPop-PGYP respectively, two pathogenic strain of *Wolbachia* (3; 102; 103; 104), and genes coding for proteins with ankyrin repeat-like domains (99; 102; 105; 106; 107). The number of copies of the octomom region has been shown to be associated not only with pathogenicity but also to higher *Wolbachia* replication and higher protection against viruses (3). The ankyrin repeat is a motif of 33 amino acids that mediates protein-protein interactions in eukaryotes (108). These domains are found in many proteins with a wide variety of functions (108). In some intracellular pathogens, ankyrin repeat proteins are type IV secretion system effectors (109). Ankyrin repeat proteins occur more in eukaryotes but can be seen in some bacteria, mainly in Proteobacteria (110). However, *Wolbachia* *w*Mel has a higher number of ankyrin repeat proteins than is usually found in the Proteobacteria (99; 110). This high number of ankyrin repeat proteins is also observed in other arthropod-infecting *Wolbachia* but is absent in the nematode-infecting *Wolbachia* (55; 107; 111). Various studies have tried to link ankyrin repeat proteins with the phenotype that *Wolbachia* induces in its host, such as cytoplasmic incompatibility (105; 107; 111; 112; 113; 114; 115).

In bacteria, regulation of gene expression relies mainly on operons, regulatory proteins and sigma factors (116). However, in obligate intracellular bacteria, genome reduction lead to a loss of those mechanisms (116; 117). In this case, mechanisms such as post-transcriptional regulation through antisense RNAs are thought to provide regulatory potential (116). The genome of *Wolbachia* *w*Mel does encode a few proteins for regulatory responses (99). Among those, some regulatory systems, such as two-component signaling pairs, are well conserved in *Wolbachia*, but some differences compared to other α-proteobacteriae suggest that there could be differences of regulation in *Wolbachia* (118). In the *w*Mel strain, one type of two-component regulatory system, a sensor histidine kinase

and paired response regulator which are involved in sensing and responding to the environment, is encoded by the genes WD1215-16, WD0732, WD1284, WD0221, WD0728 (118).

1.3.4 History of *Wolbachia* infection in *D. melanogaster*

Knowledge about infection rates and *Wolbachia* diversity in *D. melanogaster* are necessary to have a good understanding of the co-evolution process. The prevalence of *Wolbachia* in *D. melanogaster* is widespread, with the proportion of infected individuals per population ranging from 10 to 90% in general, however in some cases population showing around 1% of infection have been described ((13; 14; 17; 43; 119; 120; 121; 122), reviewed in (16), (15)). In an early screen of *Wolbachia* infection in 1994, 266 lines of *D. melanogaster* from Africa, Europe and North-America showed around 30% of infected strains, with no evidence for a geographical pattern of infection rate (13). The infection rate in this study did not appear to vary with the collection date of the lines (13). The authors also suggested that infection occurred only once early in the history of *D. melanogaster* and that the polymorphism in *Wolbachia* infection results from repeated loss events rather than independent acquisition (13). Later, studies on co-phylogeny between *Wolbachia* and various *D. melanogaster* populations worldwide also found evidence for a single origin with incomplete vertical transmission and no horizontal transfer (14; 15; 16).

Using the *Wolbachia* genome sequence published in 2004 by Wu *et al.* (99), Riegler *et al.* (17) identified polymorphic markers and designed PCR primers to identify distinct *Wolbachia* genotypes in *D. melanogaster*, which did not show polymorphism for the gene sequences, such as 16S, ftsZ, wsp, studied previously. The authors defined five distinct *Wolbachia* genotypes on the basis of the presence or absence of structural variants including IS5 transposon insertion site and several variable number tandem repeats: *w*Mel, *w*Mel2, *w*Mel3, *w*MelCS and *w*MelCS2 (Figure 1.2 from Riegler *et al.* (17)). Another genotype called *w*Mel4 was found later by Ilinsky *et. al* in 2013 from a single population in the Sinai Peninsula (Egypt) (16).

The use of whole genome resequencing enabled the different *Wolbachia* genotypes (and the associated *D. melanogaster* mitochondria genotypes) to be further separated into distinct clades based on a phylogenomics analysis of SNPs. Richardson *et al.* first defined six clades (I to VI) based on the phylogeny of

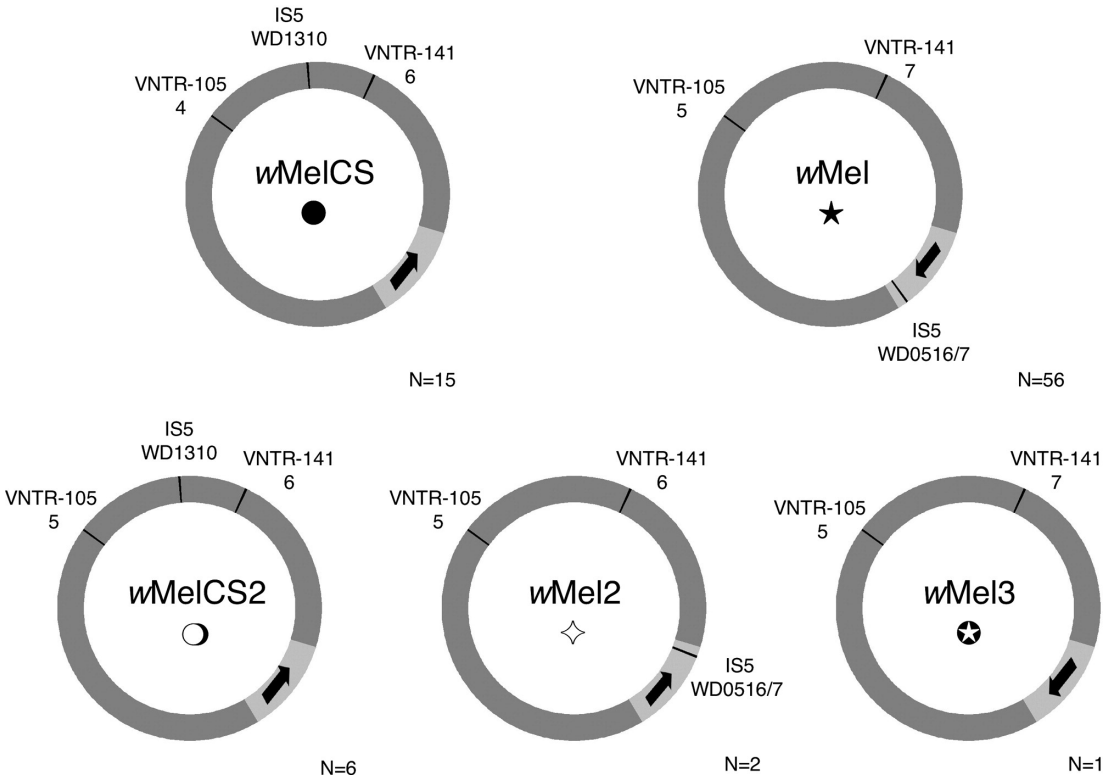

Figure 1.2: **Chromosomal Maps of Five Different Genotypes of *Wolbachia* Isolated from *Drosophila melanogaster* from Riegler *et al.* (17)** . The genotypes are differentiated by two variable number tandem repeat (VNTR) loci, two differential insertion sites of IS5, and a large chromosomal inversion.

D. melanogaster mitochondrial and *Wolbachia* genomes (14) (Figure 1.3 from Richardson *et al.* (14)). Clades I to IV were infected with *w*Mel-like genotype and clade VI was infected by *w*MelCS-like genotypes (14). No infection was detected in flies carrying clade V mtDNA in this study (14). Clades I and III were found in strains from Africa, Europe and North America (the three sampled locations) (14). Clade II was composed of only African strains and clade IV was composed of only five Ethiopian strains (14). Clade V contained only two European strains while clade VI contained strains from North America (14). Clade III was found associated with both *w*Mel and *w*Mel3 genotypes (3).

Because *Wolbachia* in *D. melanogaster* has been shown to be transmitted only vertically with loss events from a single origin, the phylogeny of both *Wolbachia* and *D. melanogaster* mitochondria are congruent. Therefore, the clades defined on the basis of *Wolbachia* phylogeny have corresponding mitochondrial clades,

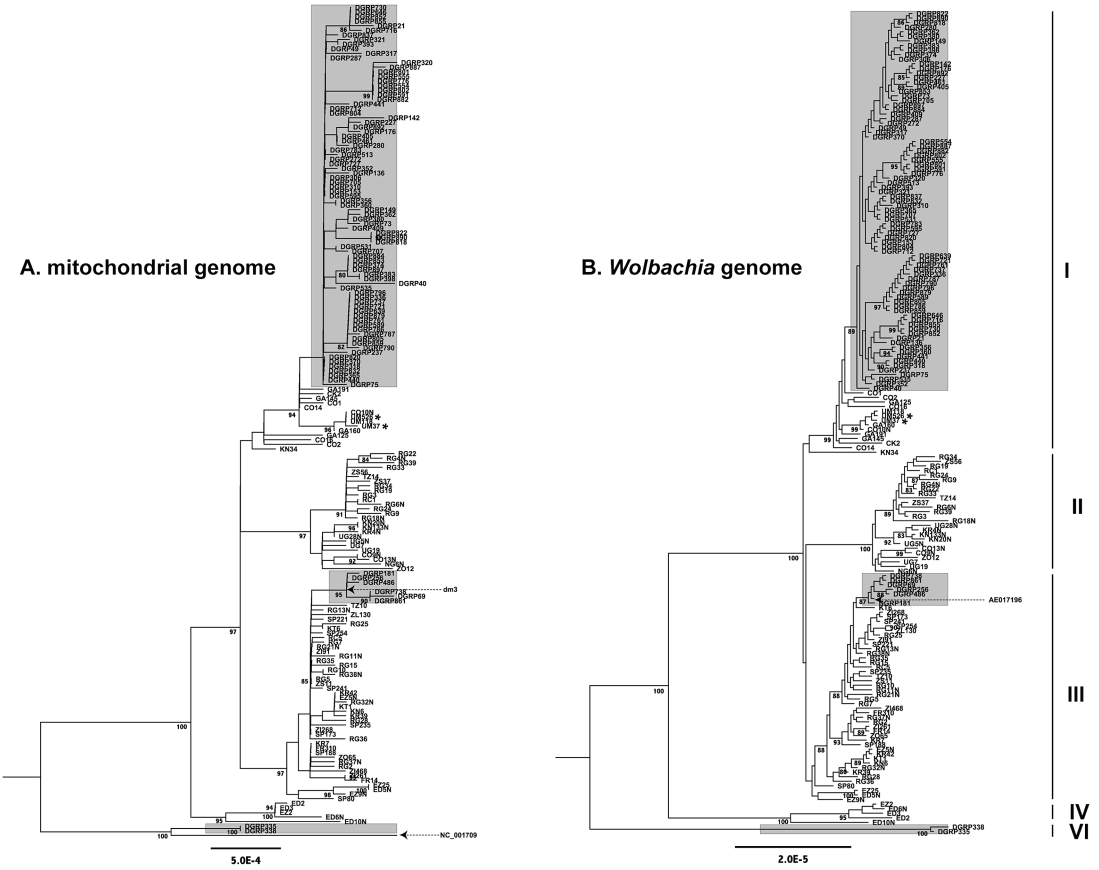

Figure 1.3: **Maximum likelihood genealogies from Richardson *et al.* (14).** Maximum likelihood genealogies of mtDNA (A) and *Wolbachia* (B) from infected DGRP and DPGP strains. Strains highlighted with a grey background are from the North American DGRP sample and the remainder are from the African and European DPGP sample. The major cytoplasmic lineages discussed in the main text are shown as clades I–V. Reference sequences for mtDNA (dm3 and NC_001079) and Wolbachia (AE017196) are labelled and dashed arrows represent their positions on the tree. Asterisks represent two Ugandan strains that are also predicted to be infected with Spiroplasma. The mitochondrial tree is based on an ungapped multiple alignment of 181 sequences of 12,236 bp in length, and the *Wolbachia* tree is based on an ungapped multiple alignment of 180 sequences of 957,546 bp in length. Unrooted ML trees were midpoint rooted for visualization and branches with ¿80% RAxML bootstrap values are shown with coloured boxes. Scale bars for branch lengths are in terms of mutations per site.

and uninfected flies can be associated to one of those clades (14; 15; 16).

A study by Ilinsky *et al.* (16) reported another cytoplasmic lineage, named clade VII, based on the mtDNA and associated with the *w*MelCS2 genotype

in populations from eastern Europe (Ukraine, Moldavia, Georgia, Tajikistan, Belorussia, and Russia). This study also found *Wolbachia* infection in *D. melanogaster* from clade V in populations from Central Asia, Russia and Ukraine (16), extending the European distribution seen by Richardson *et al.* (14) to Eurasia. The *w*Mel genotype was also found to be associated with clade III in populations from Russia, USA, Egypt and Zimbabwe (16), confirming the wide geographical range of clade III found in Richardson *et al.* (14). Ilinsky *et al.* (16) found *w*Mel2 associated with clade IV, in a population from Japan, adding to the Ethiopian strains found by Richardson *et al.* (14). *Wolbachia* from the same *w*Mel2 genotype were found in populations from Japan, China, India and Southeast Asia, (14; 17; 43). However, those *w*Mel2 genotypes might not belong to clade IV, since *Wolbachia* genotypes are defined at low resolution and can be associated with several clades.

Early *et al.* report in 2013 finding a new clade associated with a *w*Mel-like genotype, based on *Wolbachia* SNPs and named clade VIII, in a population from Beijing (15). Based on sequencing type strain defined by Riegler *et al.* (17), the *w*Mel2 genotype was found associated with clade VIII in a population from Japan (3). The cytotypes found by Early *et al.* (15) are geographically structured but not isolated, with the exception of clade VIII found only in Beijing and clade V and VI in the Netherlands.

Riegler *et al.,* found that the frequency of *Wolbachia* genotypes in samples of lab and wild stocks changed over time; with *w*MelCS being predominant in old stocks collected prior to 1970 and *w*Mel being more frequent in recent collections (17). They suggested that *w*Mel replaced *w*MelCS starting from an origin in North America in 1930.

However, the date of the most recent common ancestor (MRCA) of the *Wolbachia* infecting *D. melanogaster* was estimated to be around 8,000 years ago by Richardson *et al.,* based on phylogeny of whole *Wolbachia* and *D. melanogaster* mitochondrial genomes, suggesting that the scenario originally proposed by Riegler *et al.* (17) concerning the *w*Mel replacement event requires revision. Another study by Ilinsky *et al.,* estimated the divergence between the *w*Mel-type and *w*MelCS-type clades to be over 1000 years ago and suggest that it happened before the spread of *D. melanogaster* by humans and before the last glaciation (10 to 12000 years ago) (16). These authors also propose that the replacement of *w*MelCS-type by *w*Mel-type happened at different rates in different regions of the world (16). Early *et al.* also attempted to date the separation between

*w*Mel and *w*MelCS clades by dating of the most recent common ancestor of all *w*Mel-like clade using two different methods: one using the mitochondrial mutation rate like in Richardson *et al.*, and the other using age priors for the nodes of the clades found only in America and Australia (15). The divergence between *w*Mel and *w*MelCS was found to be around 6000 and around 2239 years ago respectively using these two methods, the latest being the favored hypothesis by the authors (15). Although the dates of the separation between the *w*Mel-like and *w*MelCS-like clades diverged from one study to another, they all agree that there was a recent replacement of *w*Mel-like variants but that the emergence of the *w*Mel genotype was much older than hypothesised by Riegler *et al.* (17).

Concerning the separation of the newer subgroups, Richardson *et al.,* suggested that *w*Mel-like clade started to diverge about 5000 years ago, most likely in Africa. Ilinsky *et al.* (16), based on the geographical location of the *w*MelCS-like clades, notably the *w*MelCS2 genotype only found in eastern Europe, propose that the *w*MelCS2 clade arose in Middle Asia and Eastern Europe (16). So far, clades I, III, IV and VI were found widely distributed (found on at least two continents), clade V has been found in Eurasia, clade VII in eastern Europe only and clade VIII in Asia.

The geographical structuring of the *Wolbachia* and mitochondrial clades, and relative proportions of different clades in a given location, have been suggested to be based mainly on selection upon *Wolbachia* (15; 16; 42). This has also been suggested by the experimental evolution experiments from Versace *et al.,* in 2014 (11) who found difference in fitness between the different *Wolbachia* clades but only when flies were infected (11). Versace *et al.,* recorded the relative frequencies of the different *Wolbachia*/mitochondrial clades in a fly population derived from a natural population and maintained during more than 30 generations. In cold environment, *Wolbachia* from clade V and VI had a significant advantage over *Wolbachia* from clades I, II and III (the relative frequencies of clades V and VI were increasing); and clade V *Wolbachia* were providing a significantly higher fitness than clade VI *Wolbachia* (11). The fitness difference was not observed in warmer environment (11).

However, it remains to be seen if the geographical distribution of some clades could be due to sampling effects. Geographic sampling is still to be completed to understand the origin and diversity of *Wolbachia* in *D. melanogaster*.

1.4 Aim

The aim of this research is to investigate further the symbiosis between *Wolbachia* and *D. melanogaster*, both on a functional and evolutive side. In light of the evolutive history and geographical variation of *Wolbachia* in *D. melanogaster*, identifying genes potentially involved in the host manipulation could shed light on the mechanisms and adaptation used by *Wolbachia* for its successful transmission in *D. melanogaster*, which in turn could be used for potential applications such as pest control or human health. Chapter 2 shows the methods for all the analyses used in the following chapters.

In Chapter 3, I use a phylogenomic approach to study the evolution and co-adaptation of *Wolbachia* in *D. melanogaster*, and set the geographical and historical context to the following chapters.

In Chapter 4, I use RNA sequencing (RNAseq) data to study *Wolbachia* gene expression across the *D. melanogaster* developmental life cycle, in an attempt to find genes involved in host interaction, but also identify key moments in the *D. melanogaster*-*Wolbachia* interaction.

The diversity of *Wolbachia* strains infecting *D. melanogaster* comes with variability in the effects that *Wolbachia* has on its hosts. In the Chapter 5, I used RNAseq data in a different host/strain context to try to evaluate the conservation of the previous findings through the variation of *Wolbachia*-*D. melanogaster* couples.

Finally, Chapter 6 brings a conclusion to the results of all the precedent chapters and propose perspectives for this research work.

Chapter 2

Material and methods

2.1 Phylogenomic analysis

2.1.1 Datasets

We gathered *D. melanogaster* resequencing data from numerous projects, to which we added 59 unpublished sequences, sampled in four locations in Ukraine (Chornobyl, Kiev, Uman, and Varva) by Iryna Kozeretska.

The unpublished *Wolbachia* sequences, sampled in Ukraine by Iryna Kozeretska, were sequenced by Casey Bergman using the following protocol. Genomic DNA was prepared from isofemale strains 11-13 generations after collection in 2011 from four field sites (Uman, Varva, Kiev, Chornobyl Nuclear Plant). Fifty females were pooled and snap-frozen in liquid nitrogen, then DNA was extracted using a standard phenol-chloroform extraction protocol with ethanol and ammonium acetate precipitation. Sequencing libraries were generated using Nextera DNA Sample Preparation Kit (Product No. FC-121-1030). Following a cleanup using the Zymo-Spin kit (Cat. No. D4023) the purified, fragmented DNA was then amplified via limited-cycle PCR which also added the indices (i7 and i5) and sequencing primers. AMPure XP beads (Cat. No. A63881) were then used to purify and size select the library DNAs. The libraries were then normalized to 2nM and pooled prior to cluster generation using a cBot instrument. The loaded flow-cell was then paired-end sequenced on an Illumina HiSeq instrument. Demultiplexing of the output data (allowing one mismatch) was performed with CASAVA 1.8.3. In total, three sequencing runs were performed. The first included 10 samples from Uman and were sequenced as 101x85 nt on an Illumina

HiSeq2000. The second included 19 samples from Varva, 15 from Kiev and 4 from Uman and were sequenced as 101x101 nt on an Illumina HiSeq2500. The third included 9 samples from Chornobyl Nuclear Plant and 4 from Uman and were sequenced as 101x101 nt on an Illumina HiSeq2500.

The published projects we used are the following:

1. The DGRP (project ID SRP000694) consisting of 213 samples from populations in North Carolina, USA (5).
2. The DPGP2 (project ID SRP005599) consisting of 161 samples from many populations from Africa and one European population (6).
3. The DPGP3 (project ID SRP006733) which consists of 194 more samples from Zambia www.dpgp.org/dpgp3.
4. A project of 85 samples (project ID SRP050151) from populations in Australia, China, Netherlands, USA and Zimbabwe, described by Grenier *et al.* (7).
5. 53 samples (ID SRP050307) sequenced from five African populations by Lack *et al.* to build the Drosophila Genome Nexus Project (8),
6. The DSRP project (ID SRP011971) consisting of 18 samples from eight lab lines (9).
7. The 37 *D. melanogaster* lines from two USA populations sequenced by Campo *et al.* in 2013 (project ID SRP009033) (10).
8. A project of 48 samples from Portugal sequenced by Versace *et al.* (project ID ERP003956) (11).
9. A project of 56 sample from Ghana, France and USA sequenced by Bergman and Haddrill (12) (project ID ERP009059).
10. Finally we used 9 lab strains from Chrostek *et al.*, used to determine the different reference genotypes of *Wolbachia* (project ID ERP002662) (3).

We discarded 84 samples from these datasets to keep only Illumina paired end reads. For ERP003956, we merged the fastq files of *Wolbachia* and mitochondrial sequencing for each infected samples and used only mtDNA samples for uninfected

lines. For this project, we also discarded 36 experimental evolution samples that were not base population lines. From the DPGP2 project, we discarded 33 samples that had been crossed with a balancer line since the origin of their cytoplasm could not be determined. We removed 8 samples containing pooled data from project ERP009059. Finally we discarded the sample SRS167162 because it did non contain mitochondrial DNA, and the sample SRS172501 because it was a mislabeled *D. simulans*.

In total we used 839 samples from 1220 runs. The flies were sampled from 22 countries from five continents (North America, Africa, Europe, Asia, and Australia).

2.1.2 Generating consensus sequences for *Wolbachia* and mitochondrial genomes

To map the sequencing reads, we used a reference hologenome consisting of the Release 6 version of the *D. melanogaster* genome (BDGP Release 6 + ISO1 MT/dm6 from UCSC) and the *Wolbachia* *w*Mel reference genome (Ensembl Genomes Release 24, Wolbachia_endosymbiont_of_drosophila_melanogaster.GCA_-000008025.1.24) (123; 124).

The raw data from sequencing were downloaded from EBI, using aspera version 2.7.9, for each run. Then, the fastq files for each run were mapped to the hologenome using bwa mem version 0.7.5a (125), in paired end mode. The resulting bam files were then sorted before being merged by sample and sorted again using samtools version 0.1.19 (126). Once all the sample bam files were sorted, they were indexed, and the reads mapping to *Wolbachia* and mtDNA were extracted to produce *Wolbachia* and mtDNA bam files, respectively, using samtools.

Wolbachia and mtDNA bam files were indexed using samtools, and variants were called with samtools mpileup, with a maximum read depth of 100000, and converted in vcf files using vcfutils.pl varFilter and a maximum read depth of 1000. Using the same mpileup parameters, the consensus sequences were produced using vcfutils.pl vcf2fq taking in account only SNPs with coverage between 10 and 100 as described in Richardson *et al.* (14).

The consensus sequence-making process did not take into account insertion or deletions. Therefore, all sequences should have the same length than the

reference sequence. However, if there was no coverage for the end of the genome, the resulting consensus sequence would miss the last nucleotides. To restore the proper length, "N"s were added to the end of the incomplete sequences using a custom perl script.

Bam files were also used to make wiggle track files of the mapped read using bedtools version 2.22.0 (127). We also used samtools flagstat to record statistics about the reads mapping to *D. melanogaster* and *Wolbachia* genomes.

2.1.3 Predicting infection status

We determined if a sample was infected based on the breadth (number of sites with a non-zero coverage) and average depth of coverage of the *Wolbachia* genome. We used bedtools genome coverage to compute per base coverage and calculated the mean depth and breadth using a custom perl script. We considered a sample to be infected when the breadth of coverage was greater than 90% and the mean depth was greater than one, as used previously by Bergman and Haddrill in 2015 (12).

2.1.4 Phylogenetic analysis

For all the infected samples, the *Wolbachia* consensus sequences were gathered in a single fna alignment file. For all samples, the mitochondrial consensus sequences were put together in another fna alignment file.

For both *Wolbachia* and mitochondrial alignments, fastree version 2.1.8 was used to make phylogenies using the generalized time-reversible option (128).

The phylogenies were analysed in R v3.1.1 using the package ape (129). To determine the correspondence between previously described clades and our phylogeny, we initially used a custom function, working with ape package in R, to identify the monophyletic group encompassing two most distant samples in the same clade from the other study. We used this approach to determine clades I, II, III and IV and VI, based on the Richardson *et al.* study (14) Only two samples ZI420-HE and ZI374-HE couldn't be attributed to any clade since they were branched from the base of clades I and II respectively, but seemed to belong to clades III and I respectively on the mitochondrial phylogeny. We left them as intermediate samples. Clades V and VIII, were defined based on which samples were found in these clades in other studies (3; 11).

Clades for the mitochondial tree were defined in the same way as *Wolbachia* clades.

Clade definitions made in R were used to produce data files to plot the phylogenies using iTOL website http://itol.embl.de (130).

The distribution of the aforementioned clades from the various geographical location on a world map was plotted using mapPie function of the rworldmap package (131) in R v3.1.1.

2.2 Gene expression analysis

2.2.1 Datasets

We studied *Wolbachia* gene expression in two different contexts. First, we studied *Wolbachia* transcriptome variation across *D. melanogaster* development in the BDSC ISO1 substrain (later refered as ISO1). To do so, we used the total RNA-seq libraries from the modENCODE developmental time course (45; 46; 47). This dataset consists of 30 time points from embryo to adult, within which males and females are mixed in non-adult samples, in unknown proportions, and adult females samples are a mix of virgin and mated females in unknown ratio. Biological replicates are available for 24 of the 30 time points. The RNA-seq libraries are 100bp reads, rRNA-depleted, paired-end and stranded.

In addition, we studied *Wolbachia* gene expression in a *nos*-gal4, UAS-*dcr*-2 driver transgene line of *D. melanogaster* (later refered as DCR2) in a dataset produced by Czech *et al.,* (132). In this dataset, samples were taken from whole embryos, ovaries, and from the ovariectomized virgin female carcasses. Biological replicates are available for all samples. The reads are 50bp long, single-end and stranded. The females from which ovaries and carcasses samples were taken, at around two days post eclosion, come from heat-shocked embryos (132). Embryo samples were collected between zero to one hour after egg laying, from virgin and mated females, and were not heat-shocked (132)(Czech personal communication).

2.2.2 Genome sequencing and phylogeny

To identify which strain of *Wolbachia* infects the ISO1 and DCR2 *D. melanogaster* lines, the genomes were sequenced by Danny Miller, and Casey Bergman put

together those sequences with other genome sequences from known *Wolbachia* genotypes to build a phylogeny.

Genomic DNA for the BDSC ISO1 and DCR2 strains was prepared from 10 starved, adult males using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, 69504). 1 µg of DNA from each was fragmented using a Covaris S220 sonicator (Covaris Inc.) to 250 base pair (bp) fragments by adjusting the treatment time to 85 seconds. Following manufacturer's directions, short fragment libraries were made using the KAPA Library Preparation Kits (KAPA Biosystems, KK8201) and Bioo Scientific NEXTflex DNA Barcodes (Bioo Scientific, 514104). The resulting libraries were purified using Agencourt AMPure XP system (Beckman Coulter, A63880), then quantified using a Bioanalyzer (Agilent Technologies) and a Qubit Fluorometer (Life Technologies). Libraries were pooled with other strains, re-quantified and run for 100 cycles in paired-end high output mode over multiple lanes on an Illumina HiSeq 2000 instrument using HiSeq Control Software v1.5.15.1 and Real-Time Analysis v1.13.48.0. CASAVA v1.8.2 was run to demultiplex reads and generate fastq files. Raw fastq reads were submitted to ENA as project ERP009035.

DNA-seq fastq sequences from ERP009035 were downloaded and mapped against a "hologenome" consisting of the Release 5 version of the *D. melanogaster* genome (Ensembl Genomes Release 24, Drosophila_melanogaster.BDGP5.24.dna .toplevel.fa) and the *Wolbachia* wMel reference genome (Ensembl Genomes Release 24, Wolbachia_endosymbiont_of_drosophila_melanogaster.GCA_000008025.1.24) (124; 133). Hologenome reference mapping was performed using bwa mem v0.7.5a (125) with default parameters in paired-end mode. Mapped reads for all runs from the same sample were merged, sorted and converted to BAM format using samtools v0.1.19 (126). BAM files were then used to create BCF and fastq consensus sequence files using samtools mpileup v0.1.19 (options -d 100000). Fastq consensus sequence files were converted to fasta using seqtk v1.0-r76-dirty (https://github.com/lh3/seqtk) and concatenated with consensus sequences of *Wolbachia* type strains from Chrostek *et al.* (3). Maximum-likelihood phylogenetic analysis on resulting multiple alignments was performed using raxmlHPC-PTHREADS v8.1.16 (options -T 6 -f a -x 12345 -p 12345 -N 100 -m GTRGAMMA) (134).

2.2.3 Estimating gene expression levels

RNA-seq fastq sequences were downloaded from SRP001696 (ISO1) and SRP021103 (DCR2). Reads were then mapped against a "hologenome" file containing both *D. melanogaster* and *Wolbachia* genomes using bwa mem v0.7.5a (125) with default parameters, in paired-end mode for ISO1 and single-end mode DCR2. The hologenome consists of the Release 5 version of the *D. melanogaster* genome (Ensembl Genomes Release 24, Drosophila_melanogaster.BDGP5.24.dna.toplevel.fa) and the *Wolbachia* *w*Mel reference genome (Ensembl Genomes Release 24, Wolbachia_endosymbiont_of_drosophila_melanogaster.GCA_000008025.1.24) (124; 133).

Once mapped, the reads were sorted and converted to BAM format using samtools v0.1.19. (126). We used the resulting BAM files to estimate the expression level of the different *Wolbachia* genes annotated by Wu *et al.* (99) found in the gtf file from ensembl (ftp://ftp.ensemblgenomes.org/pub/bacteria/release-21/gff3/bacteria_23_collection/wolbachia_sp_wri) after excluding non-coding RNA genes. Because of the operon structure of bacterial genomes, transcripts can correspond to several genes expressed together and therefore, a single read can overlap more than one gene model. However, since it has been shown that inside an operon the gene transcription level can vary (135) and the operon structure of *Wolbachia* is not known, we decided to keep the expression level estimation at a gene level rather than transcript level. As the probability of overlap decrease with the read length, individual read counts are preferable than fragment (pair of reads) counts. Therefore, we decided to count the reads independently for both ends of the sequenced fragment for ISO1 and the single reads for DCR2. We counted reads using BEDtools coverageBed v2.22.0 (127), allowing the read to extend beyond the gene limits, and counting reads independently for each gene, meaning that reads overlapping more than one gene are counted twice. We did so after making sure that any potential bias of counting reads overlapping genes twice was negligible by comparing with a method discarding the overlapping reads.

To compare gene expression between the different time points, we normalized within-sample read counts using units of transcripts per million (TPM) (136). Effective gene length in TPM calculations was set to be $gene\_length+read\_length-1$ to account for reads that extend beyond annotated gene models. TPM normalization was performed on R v3.1.1 (137).

2.2.4 Comparing global expression pattern

To compare the global gene expression between the different life stages in ISO1, we computed a Pearson correlation coefficient across gene expression values for all genes in TPM between all samples. The correlation coefficients between the different stages was visualized on a heatmap using R v3.1.1.

2.2.5 Detecting variation in gene expression

To study variation of *Wolbachia* gene expression across *D. melanogaster* life cycle, we used two approaches. First we used the read counts from ISO1 for probabilistic clustering of genes using all samples, with and without replicates, to look for common expression profiles across the host life cycle. We used MBcluster.seq v.1.0 (138) with the Poisson model with two clusters and the expectation maximization method. Because the method is probabilistic, the results may vary over clustering iteration. In order to take this variability in account, we repeated the clustering a thousand times and recorded the results. To match clusters across iterations, we defined Cluster 1 as the cluster containing most of the genes, and Cluster 2 as the cluster with the smallest number of genes.

To investigate further gene expression variation across the entire host life cycle, we also used read counts from ISO1 to perform differential expression analysis using edgeR v3.6.8 GLM approach (139) on all the 24 time points with biological replicates. We then adjusted the P-values using the Benjamini, Hochberg, and Yekutieli method (140) to correct for multiple testing. We set an adjusted p-value threshold of 0.05 to select differentially expressed genes.

We also performed a series of pairwise differential expression analysis to detect genes showing differences between: (i) adult males and females in ISO1; (ii) embryo and adult females in both ISO1 and DCR2 using most similar samples (embryo 0-2h and female one day old in ISO1, and embryo and female carcasses in DCR2); (iii) ovaries and somatic tissues in DCR2. Raw read counts were used in edgeR v3.6.8 with an exact test approach (141). We also corrected the P-values using the Benjamini, Hochberg, and Yekutieli method (140). We set a selection threshold to a P-value less than 0.01 and fold change more than two.

Chapter 3

Phylogenomics of *Wolbachia* in *Drosophila melanogaster* populations

3.1 Introduction

While the history of *Wolbachia* and *D. melanogaster* has been widely investigated (11; 13; 14; 15; 16; 42), some questions remain unanswered. Many studies agree that *Wolbachia* in *D. melanogaster* is strictly vertically transmitted, with some uninfected lineages due to recent loss of infection (11; 13; 14; 15; 16; 42). However, the dating of the first divergence between the two main groups of *Wolbachia* (*w*Mel and *w*MelCS) and the history of the evolution and interaction between the different clades of *Wolbachia* remains unclear (14; 15; 16; 17).

In an attempt to bring new elements to answer those questions, we used available resequencing data of *D. melanogaster* populations from ten projects (3; 5; 6; 7; 8; 9; 10; 11; 12), to which we added unpublished data of 59 samples from several Ukraine populations, from our own lab. Out of the ten projects, three (DGRP, DPGP3 and DPGP2) account for 536 of the 839 samples used in total. From the DGRP project we used 204 samples, all coming from North Carolina, USA (5). 194 samples were from the DPGP3 project, coming from a Zambia population, and 138 samples from DPGP2 were coming from 13 Sub-Saharan African populations and one European population www.dpgp.org/dpgp3 (6). 85 samples came from a resequencing of *D. melanogaster* from five population across the world (Australia, China, Netherlands, USA and Zimbabwe) by Grenier

et al. (7). 53 samples came from five African populations (Egypt, Ethiopia, Kenya, South Africa, and Uganda) sequenced by Lack *et al.* (8). 50 samples were coming from sequencing of *D. melanogaster* from populations in USA, France and Portugal by Bergman and Haddrill (12). A population in Portugal sequenced by Versace *et al.* gave 12 of the sequences we used (11). 17 sequences came from a resequencing by Campo *et al.* from two USA populations (10). Finally, we used 18 samples from the DSRP project (9) and nine samples resequenced by Chrostek *et al.*, all coming from lab strains (3).

Using these 839 samples, we reconstructed *Wolbachia* and mitochondrial genome sequences and built a *Wolbachia* phylogeny to study the diversity of cytoplasmic clades in *D. melanogaster*. Finally we looked at the geographic distribution of the *Wolbachia* and mitochondrial clades to understand the spacial patterns of diversity in *Wolbachia* infection.

3.2 Results and discussion

3.2.1 Rates of Infection by *Wolbachia*

We predicted *Wolbachia* infection in each sample using the breadth and depth of coverage of the *Wolbachia* genome. A sample was considered infected when breadth of coverage was greater than 90% and the average depth of coverage was greater than one. Figure 3.1 shows a distribution of the samples based on their breadth and depth of coverage. In blue are the uninfected samples and in red the samples we determined as infected. For infected samples, the maximum depth is 4573.2, and the median of depth of coverage is 36.1.

Although most of the samples are neatly split between infected and uninfected in terms of coverage, there are 44 sample which have an intermediate breadth of coverage (between 20% and 90%) and low depth of coverage. Most of those are project-specific and come from the DGRP, DPGP3 and Drosophila Genome Nexus projects (project ID SRP000694, SRP006733, and SRP050307). These projects are among those with the largest sample sizes. Such samples with intermediate breadth of coverage had been reported by Richardson *et al* (14) for the DGRP project. We can't determine if those samples are truly infected, and because our study is based on phylogenomic analysis, we chose to take a more conservative approach and considered those samples as uninfected, and kept only

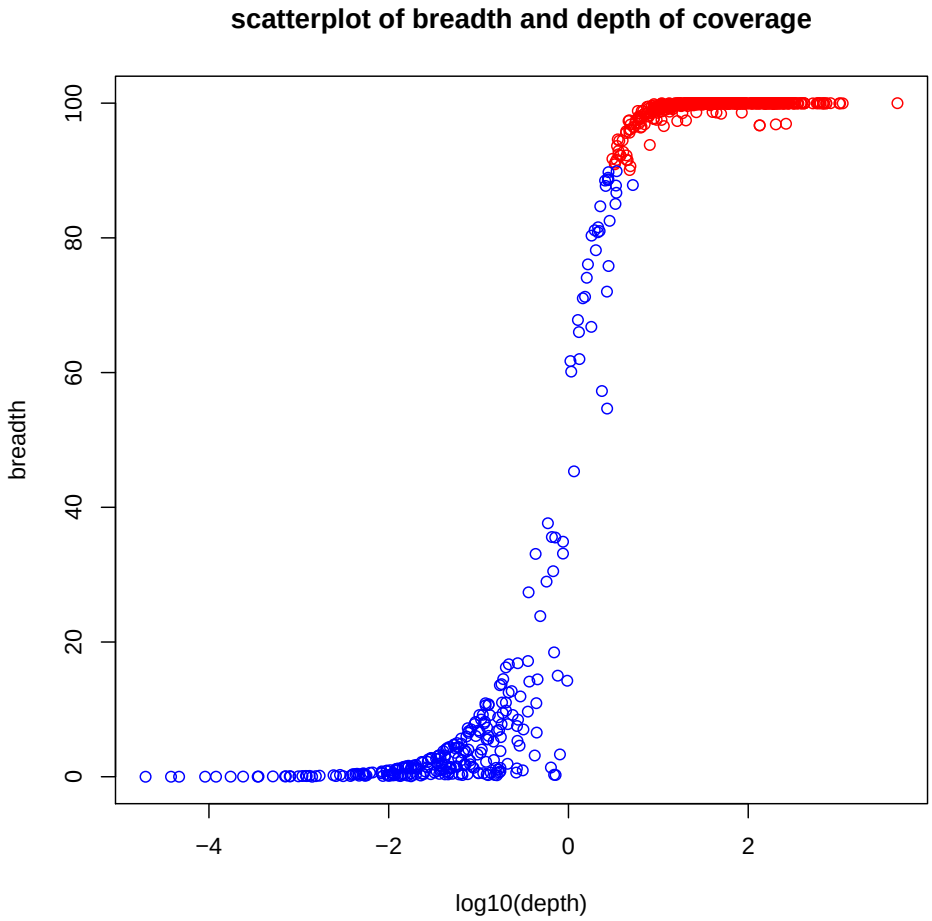

Figure 3.1: **Breadth and depth of coverage of the *Wolbachia* genome for 839 samples.** The 542 infected samples are plotted in red and the 297 uninfected samples are plotted in blue.

good quality genome sequences.

Counting those intermediate coverage samples as uninfected could affect the infection rate estimates. However, most of those intermediate samples come from Raleigh (USA) and Siavonga (Zambia), which are the largest sampled areas, with respectively 207 and 198 samples. Therefore, the impact of considering those samples as uninfected is quite low. These intermediate coverage samples could be due to a polymorphic *Wolbachia* infection in the stock, leading to *Wolbachia* sequence being on the edge of the detection levels.

We also compared predicted infection status determined here with the infection status found previously in other studies determined by PCR or genomic data

(5; 12; 14; 88; 142). For 369 out of 839 samples, we could compare infection status with previous studies. Only 7 samples showed discrepancies between infection status between studies. None of those samples were part of the samples with intermediate breadth of coverage. For four of those samples, the discrepancies lie between other studies and the infection status we find matches most of the other studies (Table 3.1). Only three samples (BCM-DGRP303, BCM-DGRP307, and BCM-DGRP774) are found infected here but were found uninfected by many other studies (Table 3.1). Overall there is a very good between-studies concordance in infection status and we can be confident about the criteria used here to determine the infection status.

sample	infection prediction	2	3	4	5	6	7	8	9
BCM-DGRP303	1	0	0	0	0	0	0		
BCM-DGRP307	1	0	0	0	0	0	0		
BCM-DGRP38	0	1	0		0	0	0		
BCM-DGRP705	1	1	1	0	1	1	1		
BCM-DGRP774	1	0	0	0	0	0	0		
BCM-DGRP911	0	1	0		0	0	0		
GA08	1							1	0

Table 3.1: **Infection status of samples with discrepencies between studies.** "1" represents presence of *Wolbachia* infection and "0" absence of infection. Columns 2 and 3 corresponds to Richardson *et al.* study (14), PCR and predictions respectively. Column 4 corresponds to Albertson *et al.* study (88) predictions. Columns 5 and 6 corresponds to the DGRP website information and Mackay *et al.* study (5). Column 7 corresponds to Huang *et al.* study (142). Columns 8 and 9 corresponds to Bergman and Haddrill study (12), predictions and PCR respectively

The total infection rate across samples is 64.6%, which is within the previously reported range for *Wolbachia* infection in *D. melanogaster* (16). However, the proportion of infected individuals seems quite variable across populations as shown in Figure 3.2 and Table 3.2. Indeed, we can see from the map in Figure 3.2, that few populations have a similar rate of infection. As seen in Table 3.2, most of the locations are represented by a relatively small sample size, but two locations have been sampled extensively: Raleigh (USA) and Siavonga (Zambia), with respectively 207 and 198 samples. Between those two locations, there is a clear difference, with the proportion of infected samples being much higher in Siavonga (82.32%) than in Raleigh (54.59%). Moreover, in the East part of sub-Saharan Africa, although the sample sizes are low, the infection rate is very

high for most of the populations (Figure 3.2 and Table 3.2). Those differences in infection rates could be due to an environmentally dependent advantage of *Wolbachia*, or variation in *Wolbachia* strain transmission efficiency, or a mixture of both. It is worth noting that very few populations are entirely uninfected.

Country	locality	sample_nb	Infected	Infection rate
Congo	Kisangani	2	1	50
Ethiopia	Gambella	5	3	60
Ethiopia	Bonga	5	5	100
Ethiopia	Dodola	5	5	100
Ethiopia	Fiche	5	2	40
Ethiopia	Masha	3	2	67
Ethiopia	Debre Birhan	5	4	80
Ethiopia	Ziway	4	4	100
Kenya	Malindi	4	2	50
Kenya	Nyahururu	5	4	80
Kenya	Marigat	4	4	100
Kenya	Thika	2	2	100
Rwanda	Cyangugu	2	2	100
Rwanda	Gikongoro	27	26	96
South Africa	Barkly East	5	3	60
South Africa	Dullstroom	5	4	80
South Africa	Port Edward	3	2	67
South Africa	Fouriesburg	5	5	100
South Africa	Phalaborwa	7	7	100
Tanzania	Uyole	3	2	67
Uganda	Namulonge	4	4	100
Uganda	Kisoro	5	2	40
Uganda	Masindi	3	3	100
Zambia	Siavonga	198	163	82
Zambia	Livingstone	1	1	100
Zambia	Solwezi	2	2	100
Zimbabwe	Harare	14	12	86
Zimbabwe	Lake Kariba	3	3	100
Zimbabwe	Sengwa	4	4	100
Cameroon	Oku	10	7	70

Gabon	Franceville	10	5	50
Ghana	Accra	15	1	7
Guinea	Dalaba	5	0	0
Nigeria	Maiduguri	6	1	17
Egypt	Cairo	3	2	67
France	Lyon	9	3	33
France	Montpellier	20	8	40
Netherlands	Houten	19	12	63
Portugal	Povoa de Varzim	12	4	33
Ukraine	Varva	29	17	59
Ukraine	Kiev	15	11	73
Ukraine	Chornobyl	9	6	67
Ukraine	Uman	6	5	83
China	Beijing	15	8	53
Australia	Tasmania	18	14	78
USA	Raleigh NC	207	113	55
USA	Athens	15	10	67
USA	Ithaca	19	14	74
USA	Winters	17	0	0

Table 3.2: **Summary of sample size and infection at the different sample sites.** The sample size and number of infected samples are summarized for the samples coming from known location. The location are colored by geographical area. Blue for Sub-Saharan Africa West, green for Sub-Saharan Africa East, light green for north Africa, yellow for Western Europe, light orange for Eastern Europe, dark orange for Asia, Red for North America and purple for Oceania.

3.2.2 *Wolbachia* clades

We used complete genome sequences of infected samples to reconstruct phylogeny and matched our clades to previous results. The phylogeny is aproximative and used at a large scale to group samples at the level of clade. On the phylogeny of *Wolbachia* in Figure 3.3, we do find the same clades found by Richardson *et al.* (14). In both studies the branches supporting the relationship of clades I, II and III are very short, and little supported. In our phylogeny, we also see two *Wolbachia* clades found in previous studies that were not seen in Richardson *et al.* (14): clade VIII containing notably *w*Mel2 samples (3; 15) and clade V,

Figure 3.2: **Map of the proportion of infected samples from the various locations.** The infected samples are represented by the black portion of the circle, and the circle area is proportional to the samples size.

that was found in Richardson *et al.* (14) for the mitochondrial phylogeny but only contained uninfected samples. Ilinsky *et al.* (16) described a clade VII, associated with *w*MelCS2 genotype. As no sample in our phylogeny clustered with the two *w*MelCS2 type samples, we labeled them as clade VI, with the other *w*MelCS samples, for simplicity. We also find two samples for which we assigned no clades because they branch quite deep between the clades I, II and III.

Although we didn't find any new clade, the two samples to which no clade were assigned (ZI374 and ZI420) might be a sign that more clades are present at very low levels. However, it could also be that these flies were infected with two different strains of *Wolbachia* as has been observed in *D. simulans* (143; 144), and that the resulting sequence is a mixture of both genomes.

3.2.3 Geographical distribution of *Wolbachia* clades

Figures 3.4 to 3.9, show the proportion of the *Wolbachia* clades, defined in Figure 3.3, plotted on a world map, split by geographical zones. These data clearly show that *Wolbachia* clades are geographically structured.

In the east side of sub-Saharan Africa, Figure 3.4, the sampled *D. melanogaster* populations are mostly infected by clade III between Kenya and South-Africa. These populations are also infected with the less frequent clade II. In Ethiopia, the dominant clade is clade IV, which is not found elsewhere in the world, with a lower proportion of clade III. In the east part of Sub-Saharan Africa, clade I is only present in Uganda and Congo.

In contrast, on the western part of sub-Saharan Africa (Figure 3.5), clade I is the dominant clade, while clade II is also present at lower frequency. This *Wolbachia* infection pattern extends to some populations in Uganda and Congo (Figure 3.4).

As seen in Figure 3.6, the population from Egypt is infected with *Wolbachia* from clade III, which is also dominant in most populations in western Europe. In Western Europe, only the Portuguese population contains *Wolbachia* from clades I and VI (Figure 3.6). clade V is present only in European populations, in larger proportion in eastern Europe than in western Europe. Indeed, populations from eastern Europe show the presence of clades III and V in similar proportions, a smaller proportion of clade VI, and a few samples from clade I.

The Figure 3.7 shows that populations from USA are dominated by clade I, with a small proportion of clade III and a few samples from clade VI. The

populations from Australia (Figure 3.8) is mainly infected with clade III, with a few samples infected with clade I.

In China, clade VIII is found together with clades I and III (Figure 3.9). clade VIII is associated with *w*Mel2 and not found elsewhere.

We can see that clades I and III, although more prevalent in certain regions, are very widespread. In contrast, clades IV and VIII have only been found in specific regions, Ethiopia and China, respectively. We only found clade II in Sub-Saharan Africa, but with a relatively wide geographical distribution within Africa. However, it is almost always found in lower frequencies compared to other clades (I and III), except for one population from Uganda. Clade V is found only in Europe, with higher frequencies in eastern Europe. Finally, clade VI is found in Eurasia and USA, with significant frequencies only in eastern Europe.

The geographical distribution of the various clades matches the previous studies (14; 15; 16; 17; 43) except for clade IV which is only found in Africa in our study. Ilinsky *et al.* (16) finds *Wolbachia* clustering with *w*Mel2 in Japan, which he reports as associated with clade IV. This is the only time it had been reported outside Africa. However two other studies found *w*Mel2 associated with clade VIII in Asian populations (3; 15). In the study by Ilinsky *et al*, clade IV divides in two main clades (16). Since clade VIII had not been defined yet at that time and clades IV and VIII branch consecutively in the phylogeny (see Figure 3.3), it is possible that the Asian samples associated with *w*Mel2 found by Ilinsky *et al* were in fact samples from clade VIII that were wrongly affiliated to clade IV. Under this hypothesis, clade IV is restricted to Ethiopia, while clade VIII is restricted to Asia.

Comparing Figure 3.2 and Figures 3.4 to 3.9, we can also notice that clade I is mainly present in zones with lower infection rates such as North America and West Africa, and clade III and IV are found in areas with high infection rates such as East Africa. This might be a reflection of transmission efficiency of the different *Wolbachia* clades, with clade III and IV having better transmission rates than clade I. However, clade I is very widespread and therefore potentially more competitive than other clades such as clade II. It is also difficult to separate the environmental effect from a clade effect. It could be that in Eastern Sub-Saharan Africa, selection to keep the *Wolbachia* infection is higher than in the Western Sub-Saharan Africa or North America, independent of the relative competitiveness and transmission rate of the strains. Moreover, the fact that both strains

from clade I and III are widely distributed suggests that the geographical structuring of those clades is not simply due to a lack of migration, but may rather result from local adaptation. Many studies suggest that the diversity of *Wolbachia* in *D. melanogaster* is the result of selection and local adaptation on the *Wolbachia* strains (11; 15; 16).

3.2.4 Coevolution of *Wolbachia* and *D. melanogaster*

To get a better understanding of the efficiency of *Wolbachia* transmission between the various strains, it is important to know to which clade the uninfected samples belong, since uninfected samples have been proposed to be the result of recent loss of infection (11; 13; 14; 15; 16; 42). To do so, we built a phylogeny based on mitochondrial sequences and assigned the samples to the clades corresponding to those found in the *Wolbachia* phylogeny.

Figure 3.10 shows the mitochondrial phylogeny with the different cytoplasmic clades in different colors. On the mitochondrial phylogeny, we find the same clades as found for the *Wolbachia* phylogeny, with some slight variation in the relationships between clades I, II and III. However, the branch length between those clades are again very short, as they are for the *Wolbachia* phylogeny. Figure 3.10 also shows the clade for the associated *Wolbachia* genome (based on Figure 3.3) for infected samples (first circle). For almost all samples, there is a strict concordance between the *Wolbachia* and mitochondrial clades. This is consistent with previous studies that found that *Wolbachia* transmission in *D. melanogaster* was strictly vertical with periodic loss of infection (14; 15; 16). However, three samples show discrepancies between *Wolbachia* and mitochondrial phylogeny: In the mitochondrial clade V, two samples from Ukraine are infected with *Wolbachia* from clade III, and one sample from Ukraine is infected with *Wolbachia* from clade I (See zoomed area in Figure 3.10). This suggest that some horizontal transfer of *Wolbachia* from another clade happened in those samples. This is the first time that evidence for potential horizontal transfer has been found in natural populations of *D. melanogaster* (14; 15; 16). The two clade V samples infected with *Wolbachia* from clade III are closely related to each other (zoomed area in Figure 3.10). These two *Wolbachia* from clade III infecting samples from clade V are also closely related to a third clade III bacteria, infecting a clade III sample from the same population (see zoomed area in Figure 3.3, the *Wolbachia* associated with a different mitochondrial clade are labeled with an asterisk.).

This pattern could be the result of a single transfer of a clade III *Wolbachia* into a clade V population in Ukraine. In contrast, the clade I *Wolbachia* infecting the clade V mitochondria is not closely related to the other *Wolbachia* in clade I from the same region (see Figure 3.3). These putative horizontal transfer events come from a newly sampled population in Ukraine. European populations are infected with several *Wolbachia* clades and don't present a dominant clade 3.6. This might increase the likelihood of detecting a horizontal transfer of *Wolbachia* into a different cytoplasmic clade, which would not be detected in populations that have a dominant cytoplasmic type.

3.2.5 Geographical distribution of mitochondrial clades for uninfected samples

We observed variation in infection rates between populations and geographical structure of the *Wolbachia* clades. To understand whether the variation of infection rate between population might be explained by the *Wolbachia* clade structure, we need to look at the mitochondrial clades distribution of uninfected samples, and their relative proportions in the different populations.

When looking at the maps of the geographical distribution of the mitochondrial clades for uninfected samples in Figures 3.11 to 3.16 we can see that it globally matches the *Wolbachia* clade distribution, as is expected for vertical inheritence. The mitochondrial clades of uninfected samples match the *Wolbachia* clades of infected samples from the same area. This is consistent with what had been found in previous studies (14; 15; 16). However, some variation can be observed in the proportion of infected and uninfected samples across clades. The proportion of clade III and IV, or II and III are rather conserved in uninfected versus infected samples from Sub-Saharan Africa (Figures 3.4 and 3.11). However, in the areas where clades I and III are coexisting such as North America and Australia (Figures 3.7 and 3.12; Figures 3.8 and 3.13), the proportion of clade I relative to clade III is higher in uninfected samples. This suggests that the transmission rate of clade I is lower than the transmission rate of clade III. Indeed, when looking at the infection rate by clades, clade I shows a global infection rate of 46.47% compared to a global infection rate of 79.47% for clade III. Therefore, it is possible that the lower transmission rate of *Wolbachia* from clade I partly explains the association between low infection rate and clade I in several regions

of the world. *Wolbachia* from clade I might confer higher selective advantages to be able to be maintained in those populations together with other *Wolbachia* clades despite the lower transmission rate.

Clade II, on the other hand, seems to be less competitive due to its more restricted geographic distribution and lower prevalence in most population. However, it seem to have a higher global rate of transmission, possibly explaining that it can maintain itself in populations infected with more abundant clades such as clade I. Indeed, while clade I has the lowest global infection rate, clade II has one of the highest infection rates with 81.08% of samples infected with *Wolbachia*. Clade IV seem very competitive in its area of distribution in Ethiopia, where it is largely dominant, and has very good transmission rates since this zone show high rates of infection (83.33% of samples from clade IV are infected).

The clades V and VI, mainly found in Europe have similar global infection rate (respectively 51.06% and 53.85%), lower than the global infection rate of clade III (79.47%). This is consistent with the fact that the proportion of clade V and VI is higher in uninfected samples than in infected samples in the European populations (Figures 3.6 and 3.15), suggesting again that infection rates might be driven by *Wolbachia* clades more than by local environmental conditions. Clade VIII has a similar infection rate to clades I, V and VI, with 50% of infected samples.

Differences in fitness have been measured between *Wolbachia* clades when exposed to cold temperatures (11). Versace *et al.,* showed that in a cold environment, clade V and VI had a significant advantage over clades I, II and III. Such environmental effects on *Wolbachia* clade competition could be driving the geographical structure and area of distribution of the *Wolbachia* clades. It is also worth noting that while *Wolbachia* in clade VI are only found in USA and Europe, mainly in eastern Europe, two uninfected samples from clade VI are found in Sub-Saharan Africa, in Gabon and Rwanda, suggesting that *Wolbachia* from this clade might be present in African populations at a low frequency, perhaps due to migration back into Africa from Europe or North America (6).

3.3 Conclusion

Here we show that *Wolbachia* clades are geographically structured and that infection rates of *Wolbachia* in *D. melanogaster* vary across populations. We also

compared the diversity of mitochondrial and *Wolbachia* clades to study infection rates across clades and populations.

For almost all samples, we found a perfect concordance between *Wolbachia* and mitochondrial clades, confirming the prevalence of vertical transmission of *Wolbachia* in *D. melanogaster* with subsequent loss of the infection in some populations as shown in other studies (13; 14; 15; 16).

However, for the first time, we find three samples which show discrepancies between mitochondrial and *Wolbachia* clades, in a population from Ukraine. We propose that two horizontal transfer events occurred, one from a *Wolbachia* from clade III in a clade V mitochondrial background, and that we witnessed in two samples; and one from a *Wolbachia* from clade I into a clade V mitochondrial background. Further work is needed to fully confirm these as horizontal transfer and discard the possibility of contamination or mislabeling.

Based on the geographical concordance between clades of infected and uninfected samples and the variations in infection rates among clades, we suggest that the presence and proportion of the difference clades at a given location are the result of local adaptation and selection of the cytoplasmic clades, while the proportion of infected individuals is mainly a result of clade-specific transmission rates. A trade-off between those different factors could explain the coexistence of several *Wolbachia* clades at a single location.

Clade I and III seem to be competitive in a variety of environments while clades IV, VIII and V are more competitive in smaller geographic areas, perhaps due to more specialized adaptation to a particular environment. Clade IV seems to be better at competing against generalist clades like III, maybe partly due to high transmission rate.

Although the *w*MelCS-type and *w*Mel-type genotype diverged between 8000 and 2239 years ago (14; 15), the global replacement of the *w*MelCS-type by *w*Mel-type bacteria proposed by Riegler *et al.* (17) is a much more recent event (after 1930). It could be the result of a recent spreading of generalist *w*Mel clades such as clades I and III. We saw that different clades seem to have different transmission rates, and Versace *et al.* showed that *w*MelCS strains performed better than many *w*Mel clades in cold environments (11). It has also been shown that the *w*MelCS strains generate stronger CI and better protection against viruses (3; 4). Based on these fitness differences between *w*Mel and *w*MelCS, we propose, as an alternative hypothesis, that the bias toward *w*MelCS infection in lines collected before 1930

shown by Riegler *et al.* (17), could be an artefact of a better persistence of *Wolbachia* *w*MelCS compared to *w*Mel in those old *D. melanogaster* lines.

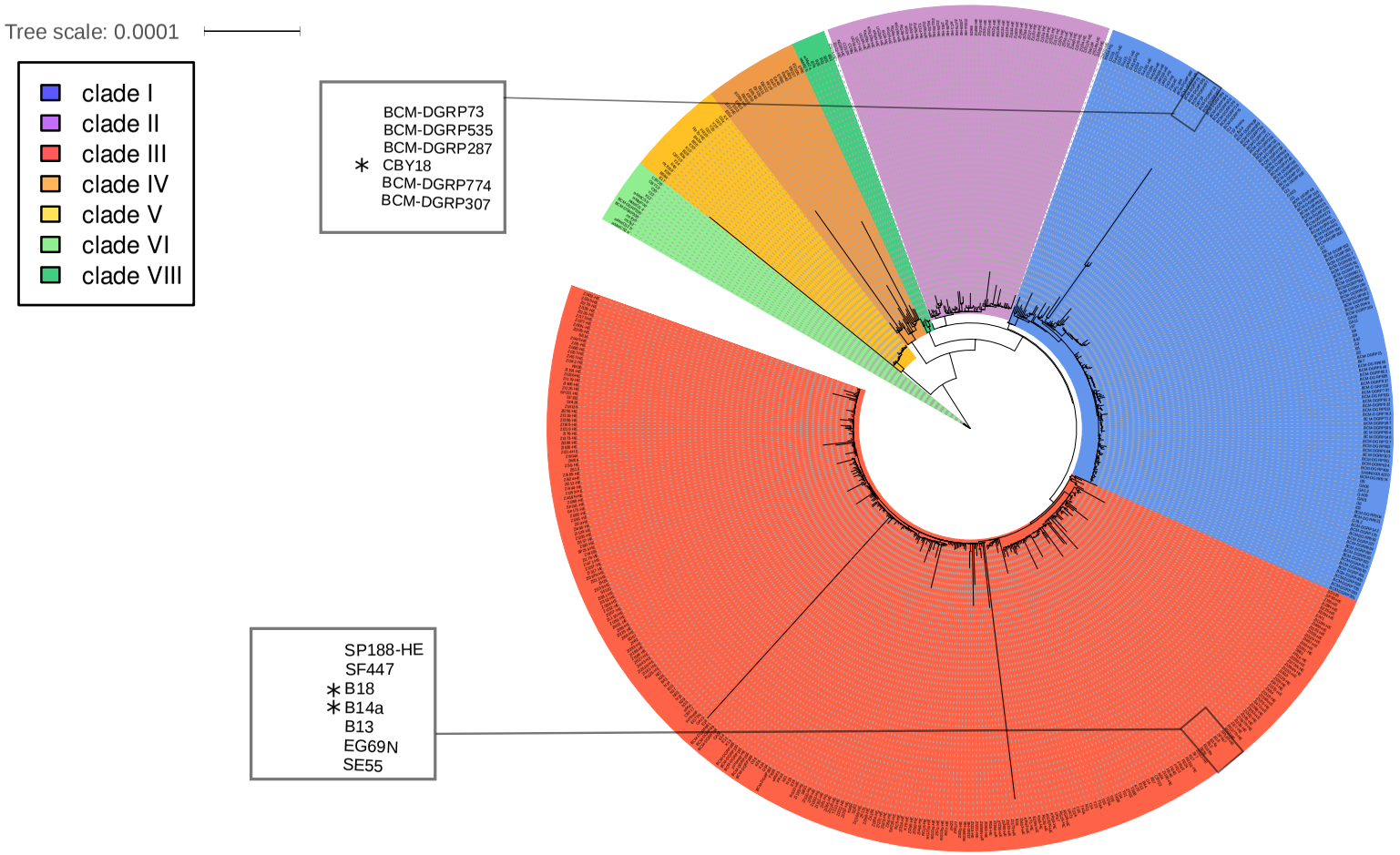

Figure 3.3: **Phylogeny based on *Wolbachia* genome.** The colors represent the different clades. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green. The stars indicate the samples where the *Wolbachia* clade is different than the mitochondrial clade.

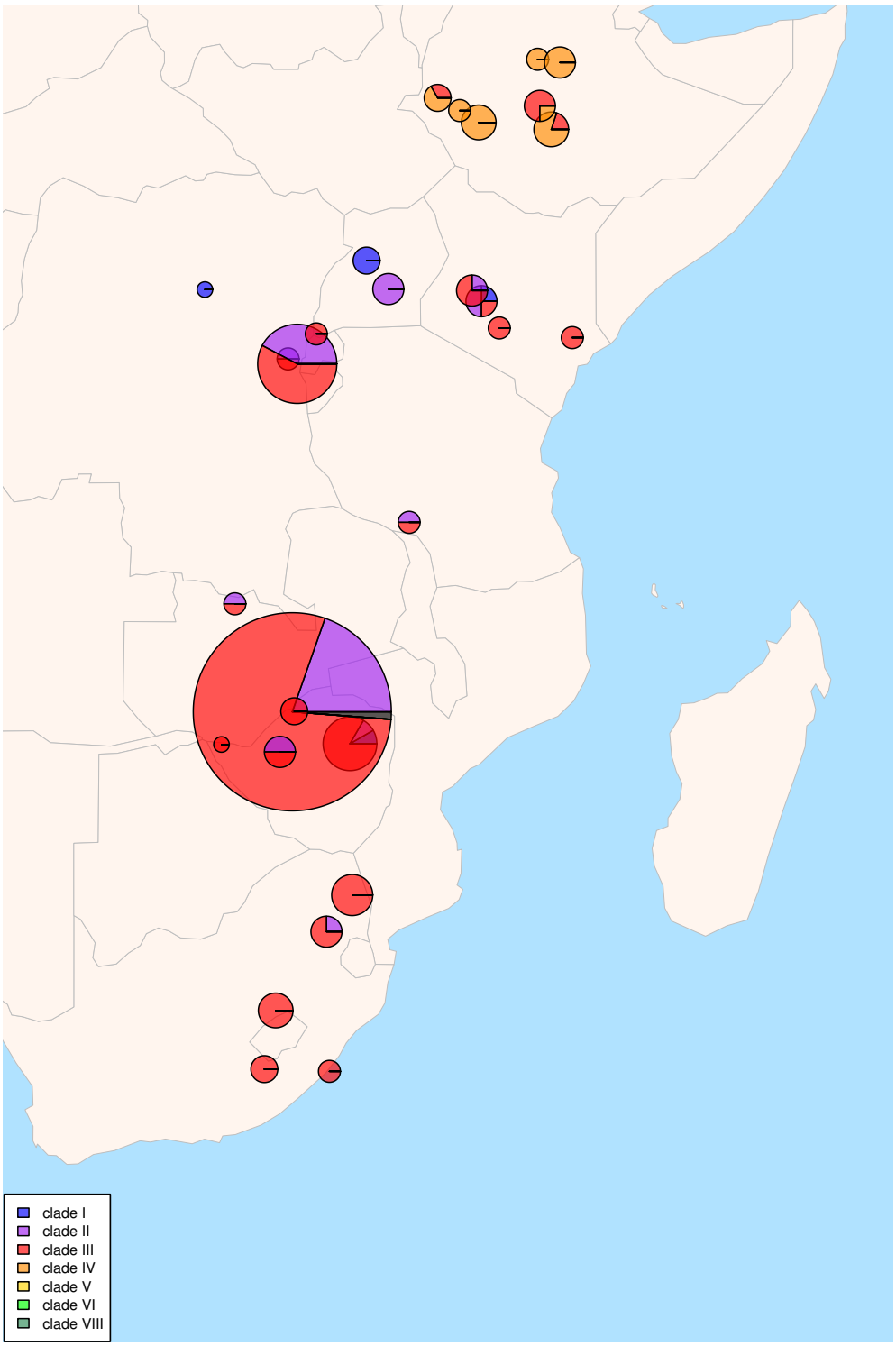

Figure 3.4: **Map of the proportion of the different *Wolbachia* clades for infected samples from East Sub-Saharan Africa.** The circle area is proportional to the samples size and the colors represent the *Wolbachia* clades for the infected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

Figure 3.5: **Map of the proportion of the different *Wolbachia* clades for infected samples from West Sub-Saharan Africa.** The circle area is proportional to the samples size and the colors represent the *Wolbachia* clades for the infected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

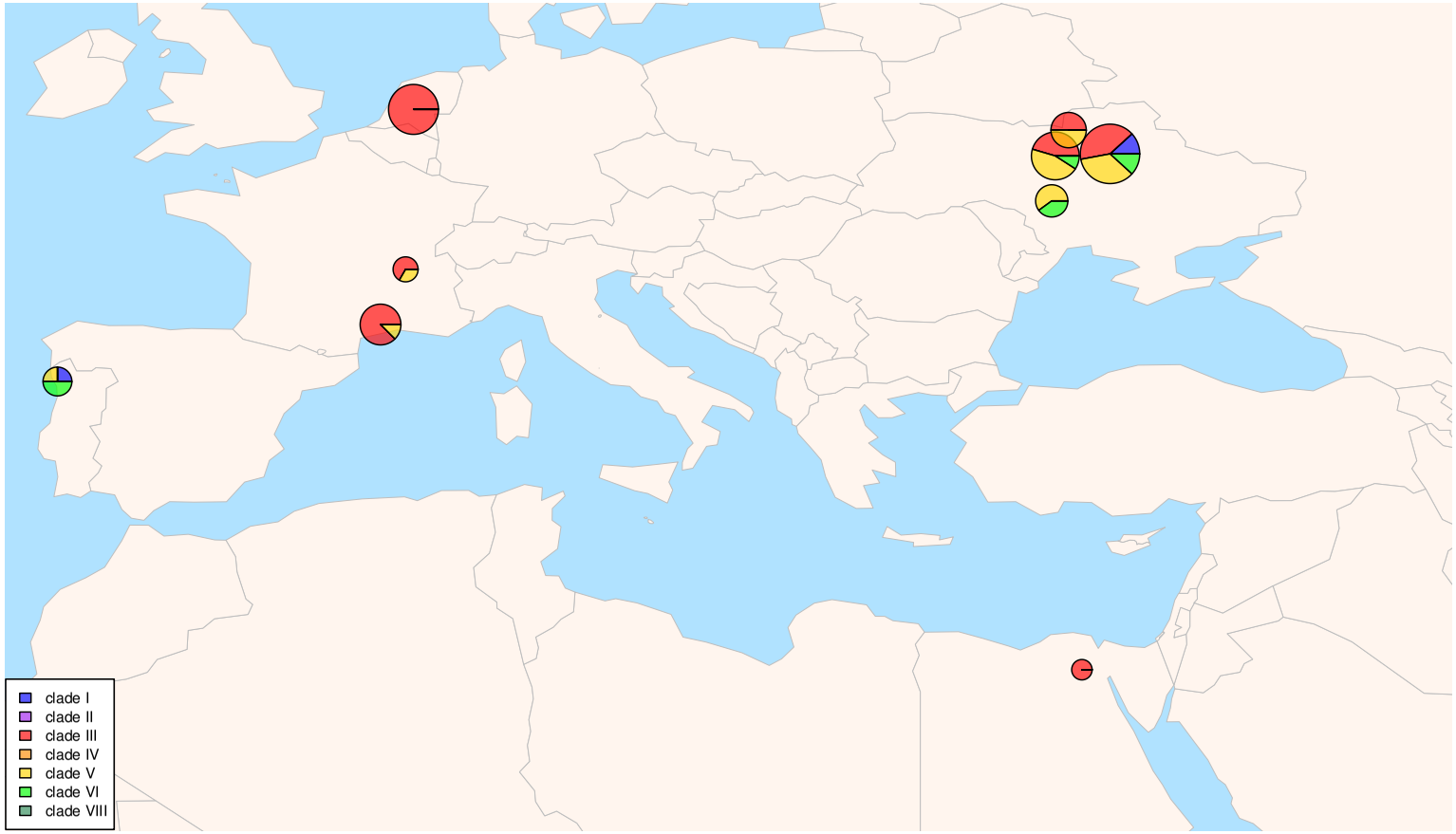

Figure 3.6: **Map of the proportion of the different *Wolbachia* clades for infected samples from North Africa and Europe.** The circle area is proportional to the samples size and the colors represent the *Wolbachia* clades for the infected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

Figure 3.7: **Map of the proportion of the different *Wolbachia* clades for infected samples from North America.** The circle area is proportional to the samples size and the colors represent the *Wolbachia* clades for the infected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

Figure 3.8: **Map of the proportion of the different *Wolbachia* clades for infected samples from Australia.** The circle area is proportional to the samples size and the colors represent the *Wolbachia* clades for the infected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

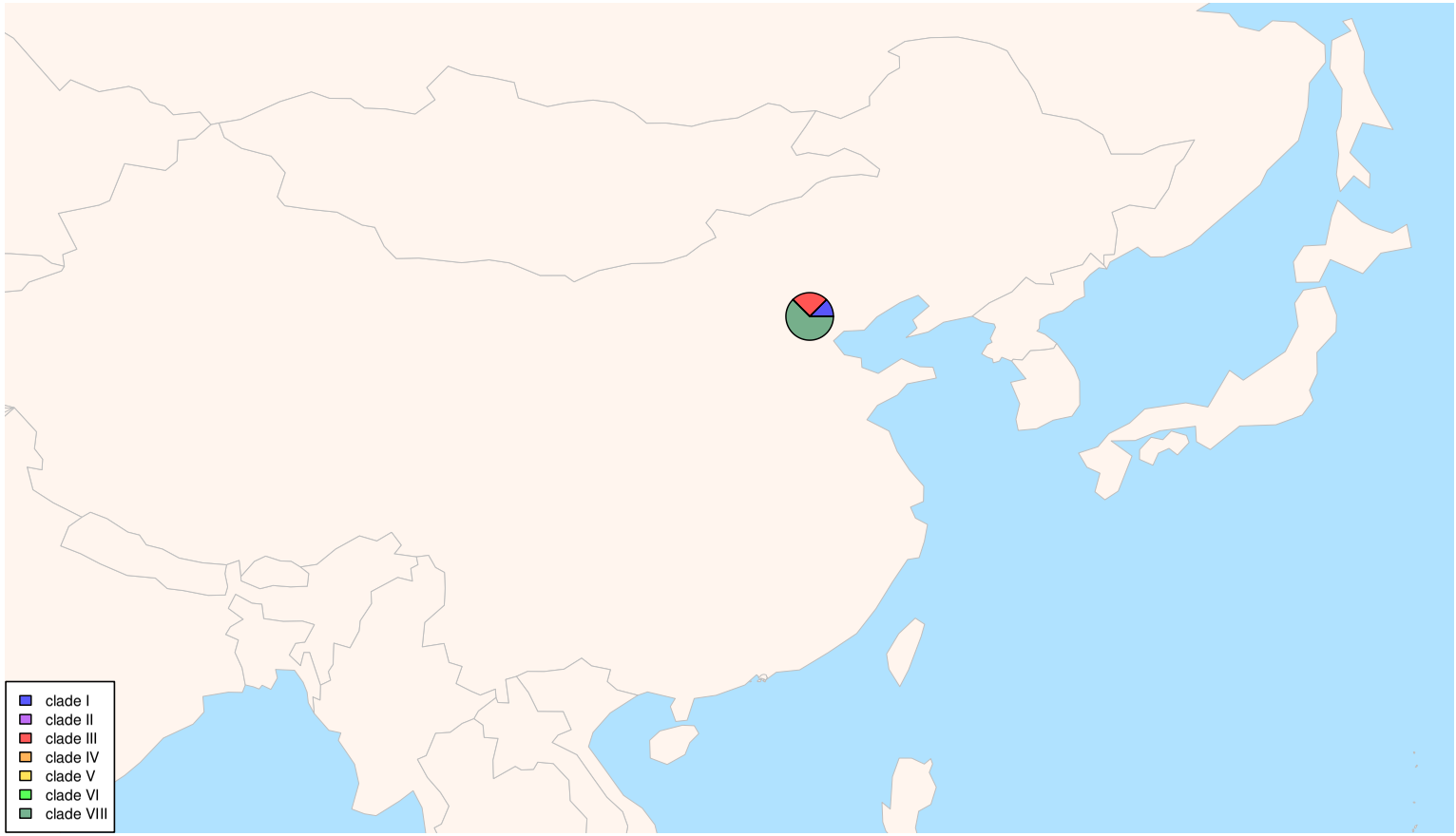

Figure 3.9: **Map of the proportion of the different *Wolbachia* clades for infected samples from Asia.** The circle area is proportional to the samples size and the colors represent the *Wolbachia* clades for the infected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

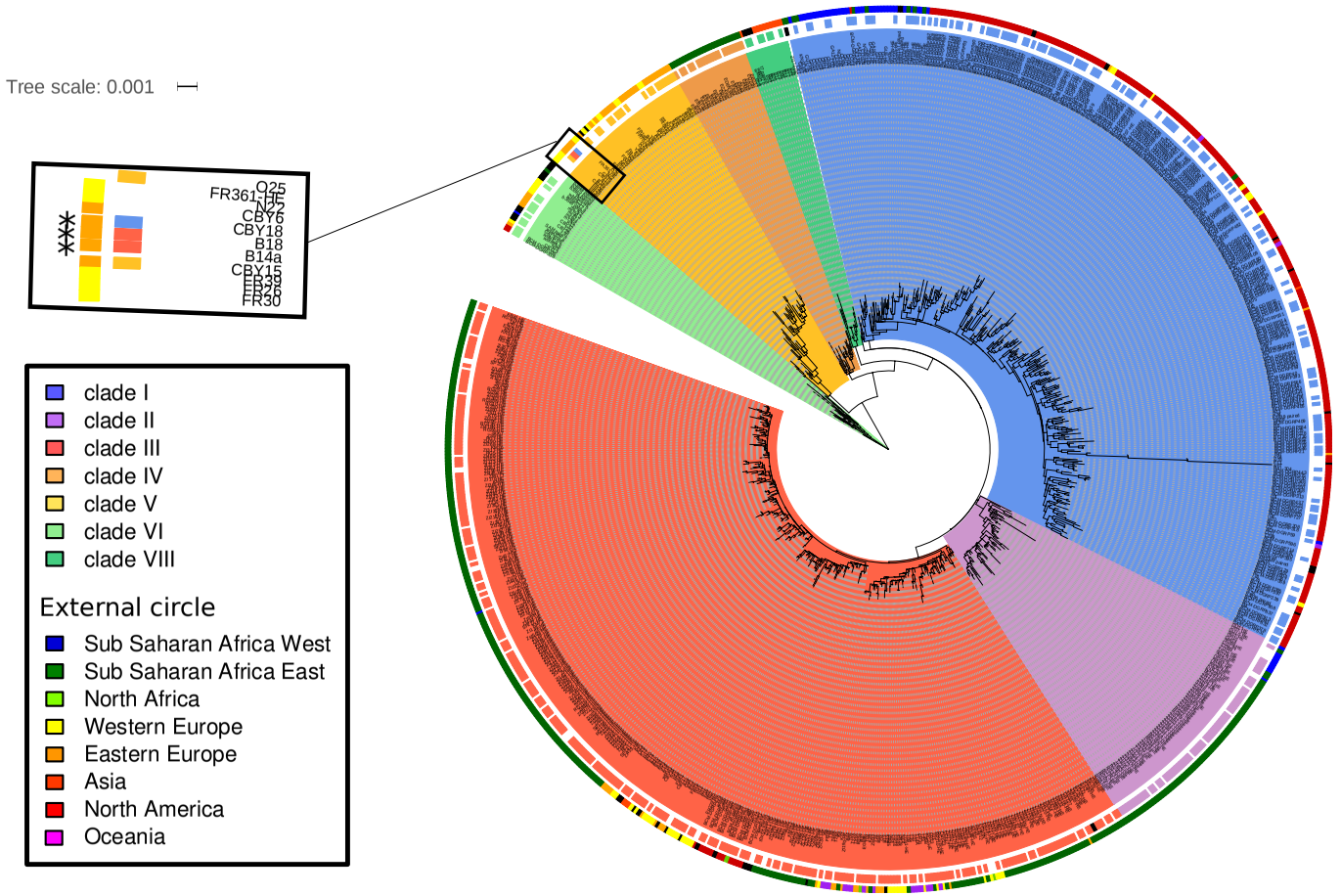

Figure 3.10: **Phylogeny based on *D. melanogaster* mitochondrial genome.** The colors represent the different clades for the mitochondrial genome clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green. The first circle shows corresponding clades for the *Wolbachia* genome from the same samples (with the same color code). The outside circle shows the geographical location of the samples. In blue, samples from Sub-Saharan Africa West, in green samples from Sub-Saharan Africa East, in light green samples from north Africa, in yellow samples from Western Europe, in light orange samples from Eastern Europe, in dark orange samples from Asia, in Red samples from North America and in purple samples from Oceania

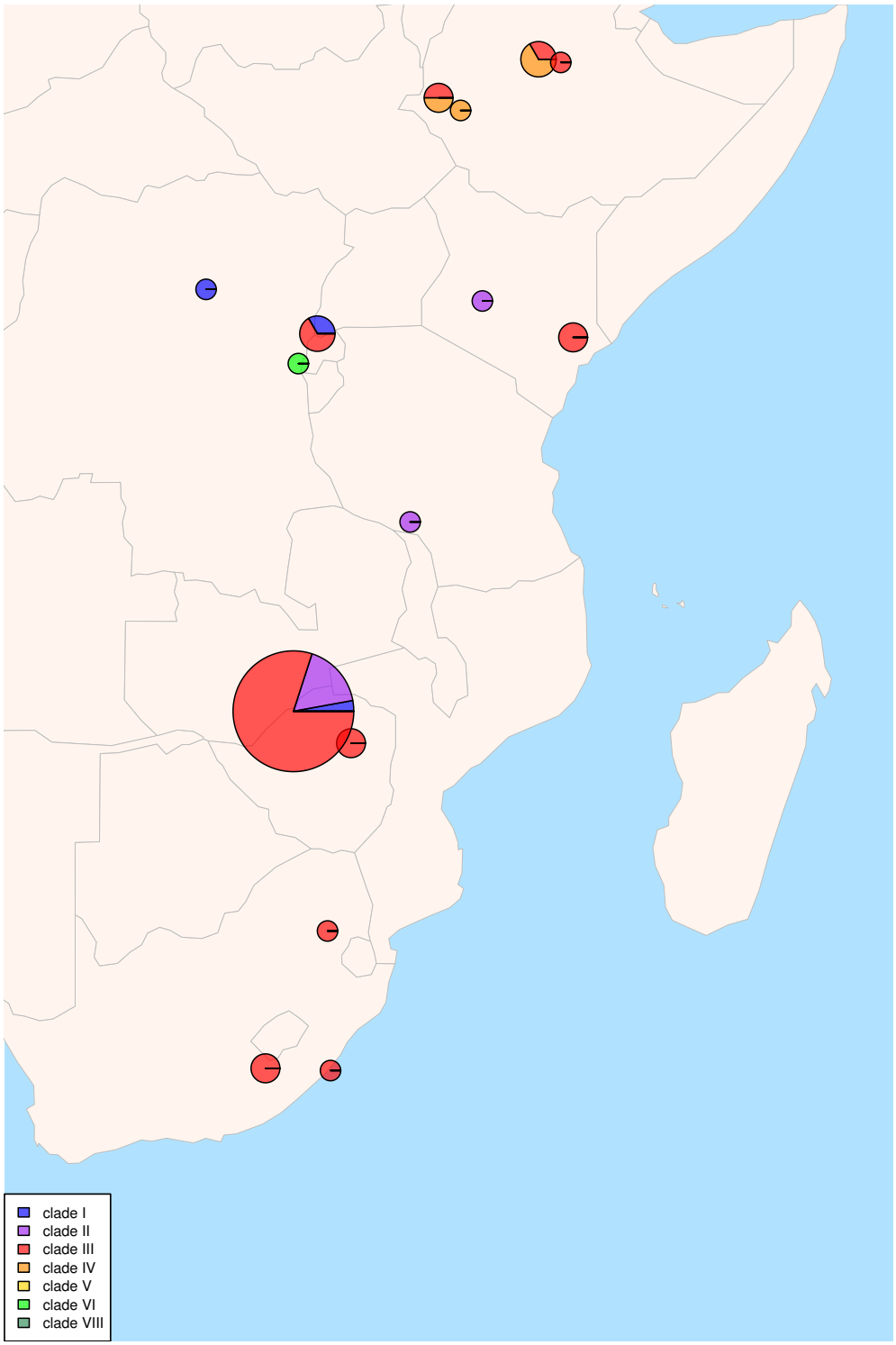

Figure 3.11: **Map of the proportion of the different mitochondrial clades for uninfected samples from East Sub-Saharan Africa.** The circle area is proportional to the samples size and the colors represent the mitochondrial clades for the uninfected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

Figure 3.12: **Map of the proportion of the different mitochondrial clades for uninfected samples from North America.** The circle area is proportional to the samples size and the colors represent the mitochondrial clades for the uninfected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

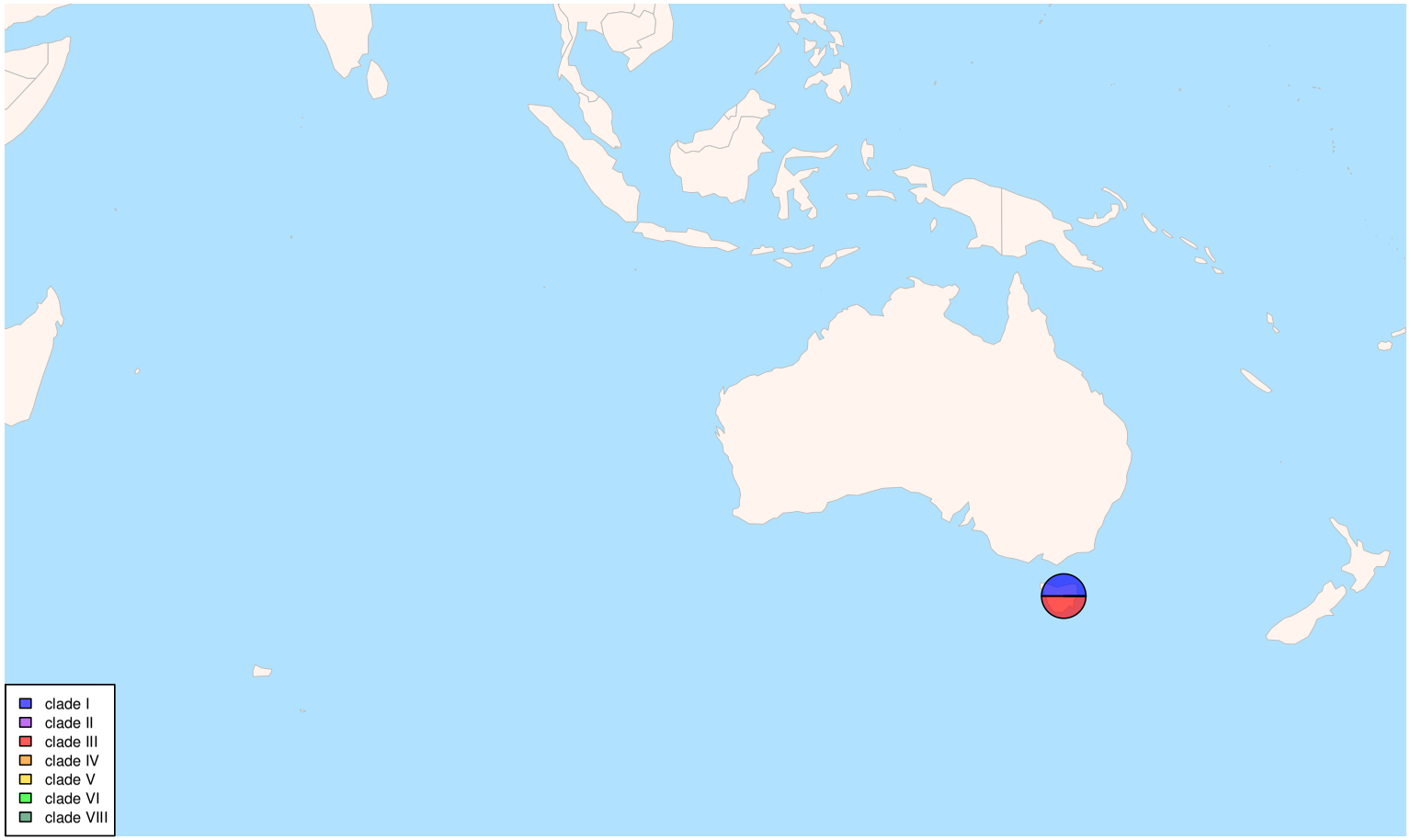

Figure 3.13: **Map of the proportion of the different mitochondrial clades for uninfected samples from Australia.** The circle area is proportional to the samples size and the colors represent the mitochondrial clades for the uninfected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

Figure 3.14: **Map of the proportion of the different mitochondrial clades for uninfected samples from West Sub-Saharan Africa.** The circle area is proportional to the samples size and the colors represent the mitochondrial clades for the uninfected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

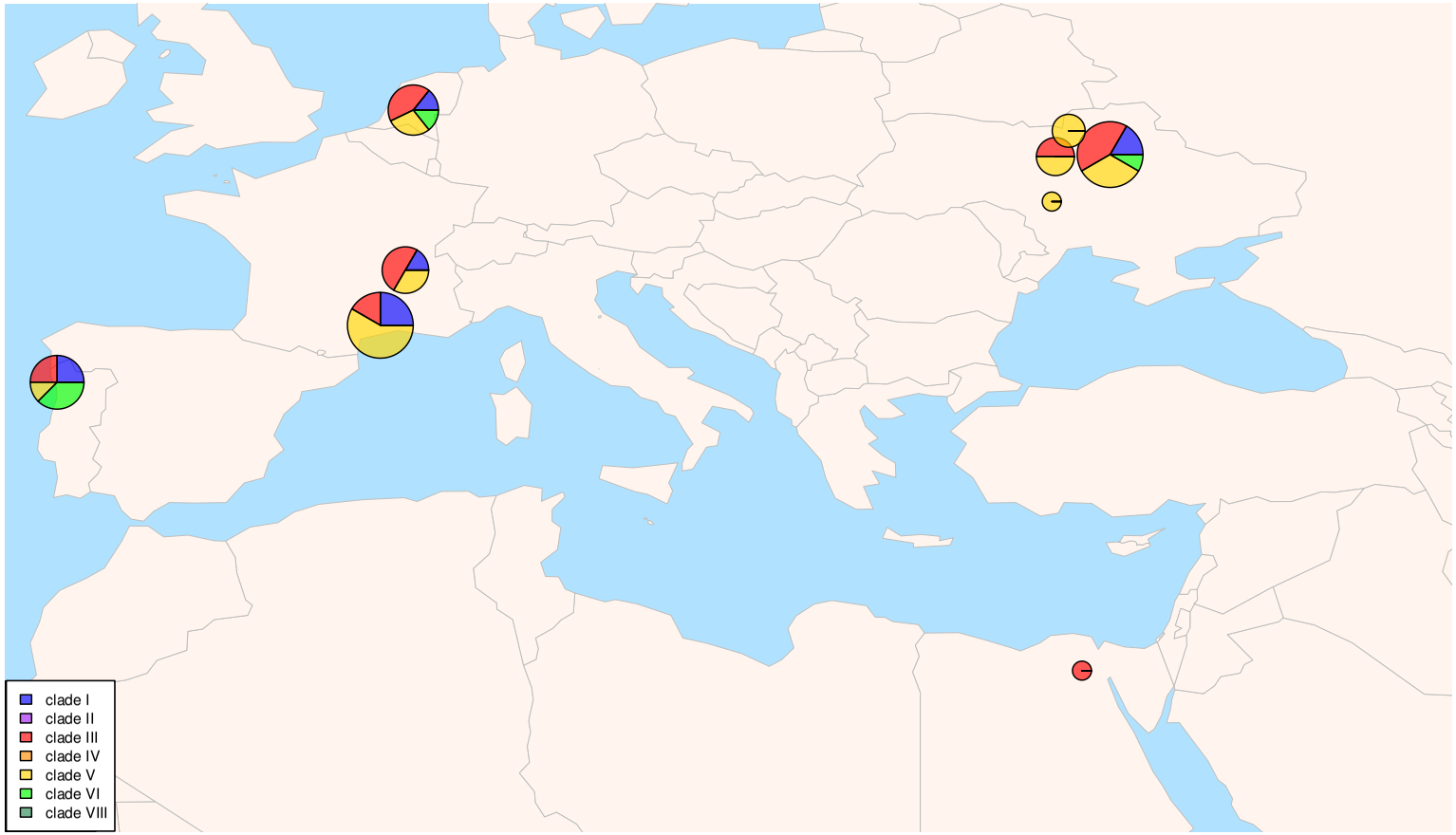

Figure 3.15: **Map of the proportion of the different mitochondrial clades for uninfected samples from North Africa and Europe.** The circle area is proportional to the samples size and the colors represent the mitochondrial clades for the uninfected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

Figure 3.16: **Map of the proportion of the different mitochondrial clades for uninfected samples from Asia.** The circle area is proportional to the samples size and the colors represent the mitochondrial clades for the uninfected samples. Clade I is in blue, clade II in purple, clade III in red, clade IV in orange, clade V in yellow, clade VI in green, and clade VIII in dark green.

Chapter 4

Wolbachia gene expression over *D. melanogaster* life cycle

4.1 Introduction

We saw, in Chapter 3, that several *Wolbachia* clades infect *D. melanogaster* and that the infection rate and relative efficiency seem subject to variation between the different clades. Moreover, the different *Wolbachia* variants have been shown to produce distinct effects on their host, with more or less cytoplasmic incompatibility, protection against viruses, and deleterious effect on longevity (3; 64). The wide range of effects induced by *Wolbachia* on its hosts makes it a bacteria of interest to understand coevolution of insect-bacteria symbiosis, but also for human health and agriculture as it can be used as a tool to control insect vectors such as mosquitoes (73; 74; 75; 76).

However, finding which *Wolbachia* genes are involved in host/bacteria interactions is made difficult by the inability to culture and manipulate this species in a free-living state. Therefore, candidate genes that might mediate interaction with their hosts have been primarily identified using comparative genomic approaches. For example, an unusually large number of ankyrin repeat domain encoding genes has been found in the genome of *Wolbachia* *w*Mel, and other strains in facultative associations with arthropod hosts, relative to other bacteria (99; 105; 113; 145), while few ankyrin repeat domain encoding genes are found in the obligate *Wolbachia* endosymbionts of nematodes (52; 55). Comparative genomic analysis of more closely-related strains of *Wolbachia* has also been used to identify candidate genes involved in host-symbiont interaction (3; 104; 105; 113; 114). For example,

a cluster of eight genes (called the Octomom region), identified as being specifically amplified in the pathogenic "Popcorn" (*w*MelPop) strain of *Wolbachia* from *D. melanogaster* (3; 104), was recently shown to cause the high bacterial titres and virulence associated with this strain (103).

Other studies have successfully used RNAseq data to understand better functional interactions between host and symbiont. Humphrys *et al.* (146), 2013, show an *in vitro* expression profile of the pathogenic bacteria *Chlamydia trachomatis* and its host cells through the first stages of infection, shedding light on potential mechanisms underlying this disease. Darby *et al.* (52) found tissue-specific differential expression revealing different roles of *Wolbachia* in the soma and germline of the nematode host *Onchocerca volvulus*. Darby *et al.* (147) also have shown that *Wolbachia* expression is regulated *in vitro* when its host cells were exposed to a potent stressor.

In this chapter, we study the gene expression of *Wolbachia* inside *D. melanogaster* throughout fly development. We used a time series of gene expression data from *D. melanogaster* at different life stages generated by the modENCODE project, which also contains transcriptomic data for *Wolbachia* (45; 46; 47).

4.2 Results and discussion

4.2.1 The modENCODE life cycle expression dataset contains a complete *Wolbachia* transcriptomic.

We discovered that the ISO1 line of *D. melanogaster* was infected with *Wolbachia* (148). The phylogeny build from genome sequences of *Wolbachia* in ISO1 among other characterized *Wolbachia* genomes (Figure 4.1) shows that the ISO1 line of *D. melanogaster* is infected with a *Wolbachia w*Mel strain, close to the *Wolbachia* reference genome. This line, notably used as the *D. melanogaster* reference genome (37), was also used to generate total RNAseq libraries across the *D. melanogaster* life cycle by modENCODE (45; 46; 47). This project contains RNAseq data from 30 time points from embryo to adult stage, 24 of them with biological replicates. The samples from embryo to pupae contain RNA from a mix of males and females, while adult stages are separated into males and females.

We mapped the raw RNAseq reads to a file containing both *D. melanogaster* and *Wolbachia w*Mel reference genomes (37; 99). All the samples contained reads

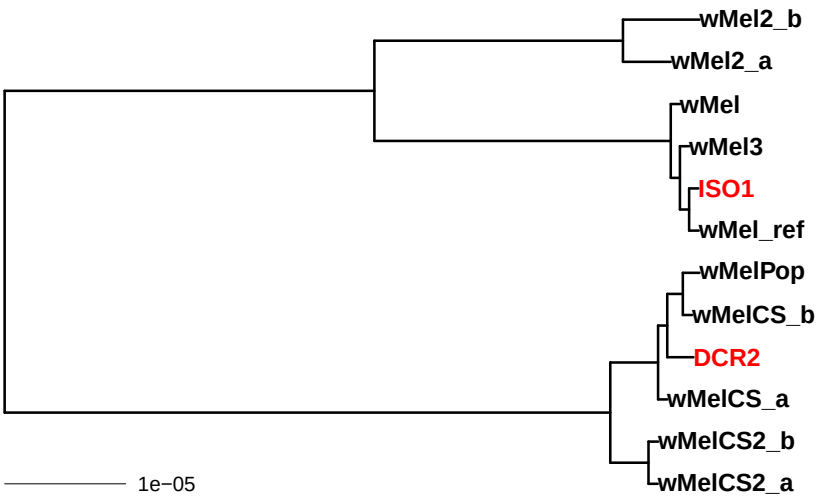

Figure 4.1: **Genome-based phylogeny of *Wolbachia* strains used in RNAseq analysis.** The other *Wolbachia* genome sequences used were produced and characterized by Chrostek *et al.* (3). The samples in red are the *Wolbachia* from Iso1 and *Dcr-2* samples used for the RNA-seq analysis.

mapping to the *Wolbachia* genome, ranging from 0.1 to 7 million reads per sample (0.3%-8.5% of total reads) with a median of 1.7 million reads per sample (1.6% of the total reads).

Figure 4.2 shows the reads mapping to *Wolbachia* in a strand-specific manner for a few time points. The coverage and strand-specificity of the modENCODE RNAseq dataset is high enough to show clear correspondence with the boundaries of the majority of annotated *Wolbachia* gene models, given their presumed operonic structure and lack of annotated untranslated regions. The correspondence between the genes and read orientation also shows the absence of DNA contamination into RNAseq data.

We next counted reads mapping to each gene and estimated expression levels in transcripts per million (TPM) for each of the 1195 characterized protein-coding genes in the *Wolbachia* genome in each of the modENCODE RNAseq samples. On average, 1104.4 genes were expressed (defined as having a non-zero TPM) across all samples, ranging from 948 genes to 1170 genes. 1193 genes out of 1195 were found to be expressed in at least one of the samples. This indicates that the majority of the *Wolbachia* genome is transcriptionally active throughout the *D.*

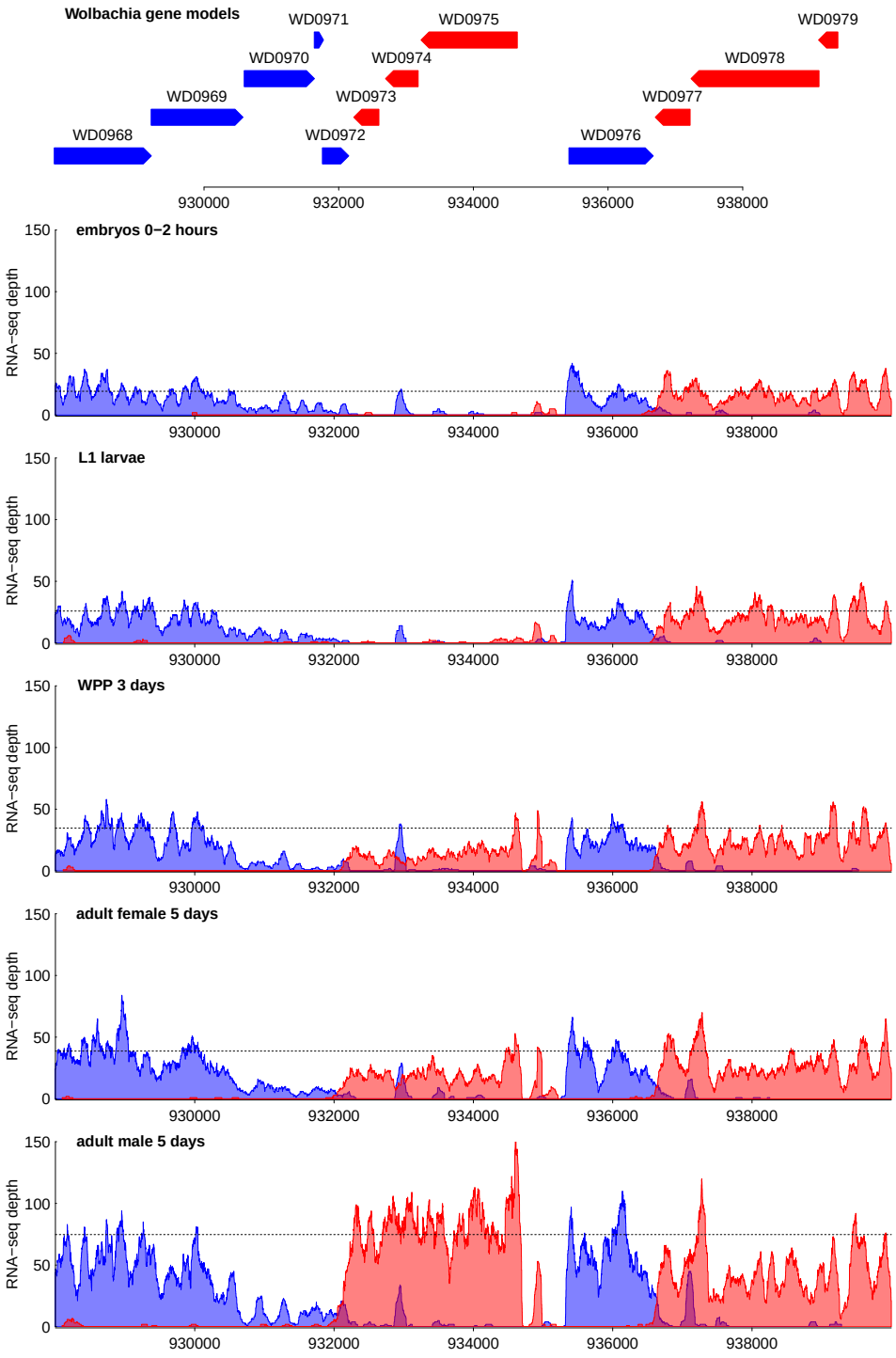

Figure 4.2: **Expression landscape of *Wolbachia* in *D. melanogaster*.** Gene models and RNAseq coverage plots for a 12-gene window of the *Wolbachia* genome in representative stages of the *D. melanogaster* life-cycle. Gene models and RNAseq coverage are shown on the forward and reverse strands in blue and red, respectively. RNAseq plots are shown on the same absolute y-axis scale. To provide an internal normalization factor for comparison across samples, mean coverage of the stably-expressed Wsp/WD1063 gene (not shown in this interval) divided by twenty is depicted by the dashed line in each panel.

melanogaster life cycle.

4.2.2 Most *Wolbachia* genes are transcriptionally stable across the host's life cycle.

To evaluate how *Wolbachia* global gene expression varies across the host life cycle, we visualized the correlation of *Wolbachia* gene expression levels between every pair of samples across the host life stages (Figure 4.3). Expression levels are in units of TPM, so they are normalized by the sample total read count and can be compared between samples.

Correlation among biological replicates of the same stage is very high ($r > 0.94$) and highly significant ($p < 2.2e-16$). The lowest correlation among biological replicates is observed for three stages from the third larval instar (L3 dark blue gut, L3 light blue gut and L3 clear gut), which show higher similarity between stages from the same replicate series than they do between biological replicates from the same stage.

Correlation of expression levels is also reasonably high and highly significant among all life-cycle stages ($r > 0.61$, $p < 2.2e-16$). Despite this low global variability, two weakly-differentiated, partially-overlapping clusters can be observed that include embryonic to white prepupal (WPP) stages, and late larval to adult stages, respectively (square blocks of yellow spanning multiple stages in Figure 4.3). Overall, these results suggest that differences in *Wolbachia* gene expression do exist among *D. melanogaster* life-cycle stages, but that the global pattern of *Wolbachia* gene expression does not change dramatically across the *D. melanogaster* life cycle when assayed at the level of the whole organism.

To investigate further how *Wolbachia* gene expression varies across the host life cycle, we looked for genes which would display similar behaviour across time by clustering the genes based on their expression levels in all samples (138). This clustering approach enables us to use information from samples with no biological replicates. After testing several parameters, we found that the genes were dividing best in two clusters. However, the probabilistic nature of this method leads to slightly different outcomes over iterations so we recorded the cluster affiliation for each gene over 1000 clustering iterations.

We identified two main clusters that could be matched across independent clustering runs (Figure 4.4 A). The first cluster contains 1033 genes out of 1195

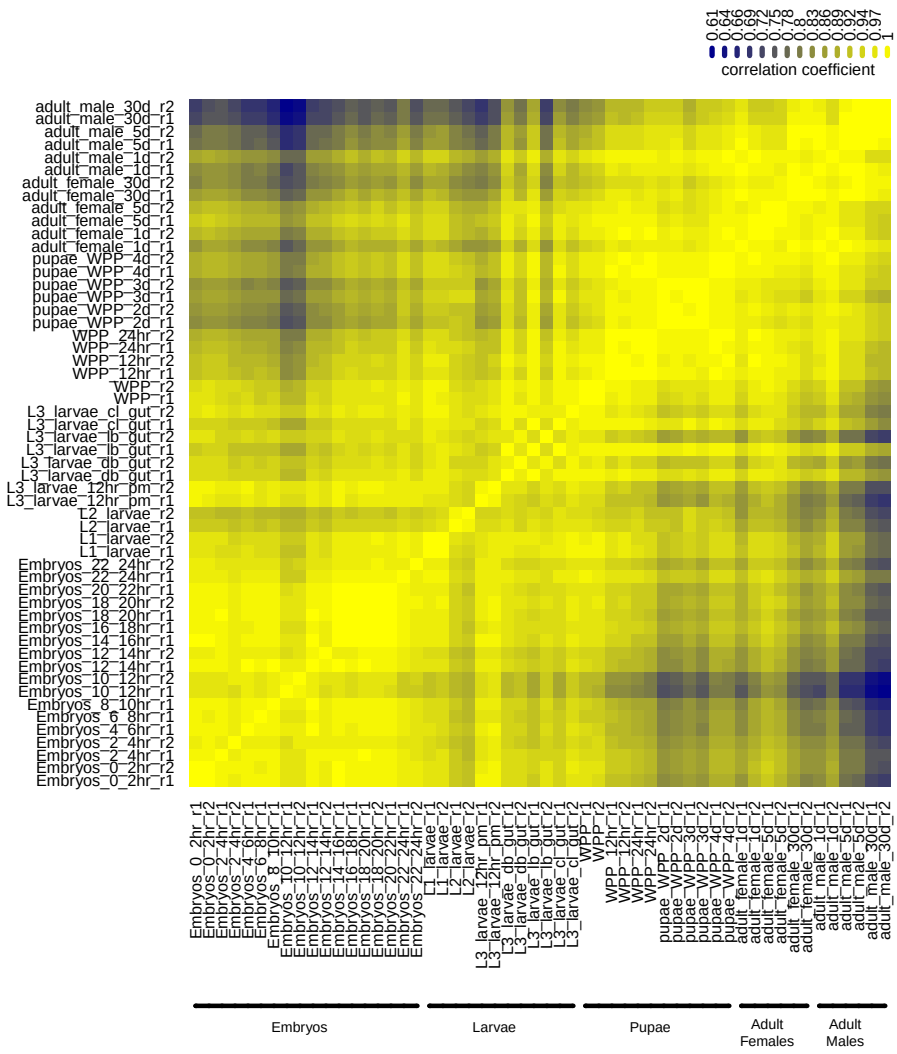

Figure 4.3: ***Wolbachia* gene expression levels are highly-correlated across biological replicates and life cycle stages.** Each cell in the heatmap represents a Pearson correlation coefficient of expression levels across all genes (in units of TPM) for a pair of samples in the ISO1 total RNAseq dataset. Higher similarity among pairs of samples is represented by bright yellow and lower similarity by dark blue. All but six stages in the modENCODE total RNAseq time course have biological replicates (Embryos 4-6 hrs, Embryos 6-8 hrs, Embryos 8-10 hrs, Embryos 14-16 hrs, Embryos 16-18 hrs, and Embryos 20-22 hrs). Replicate samples from the same stage were collected in two independent series, denoted by _r1 and _r2 suffixes.

(86.4%) whose expression pattern remains relatively stable across the host life cycle (Figure 4.4 B). We still can observe variation in expression for individual genes in this cluster (Figure 4.4 B) but these peaks are mainly single-sample effects and not a trend across the developmental life cycle. It is worth noting that the general variability is higher in embryo and larvae samples. The lower variation in expression observed in pupae and adult samples could explain the pupae and adult cluster in the correlation heatmap (Figure 4.3). The second cluster contains the remaining 162 genes (13.6%) (Figure 4.4 A) which show variation in expression pattern across the host life cycle (Figure 4.4 C to E).

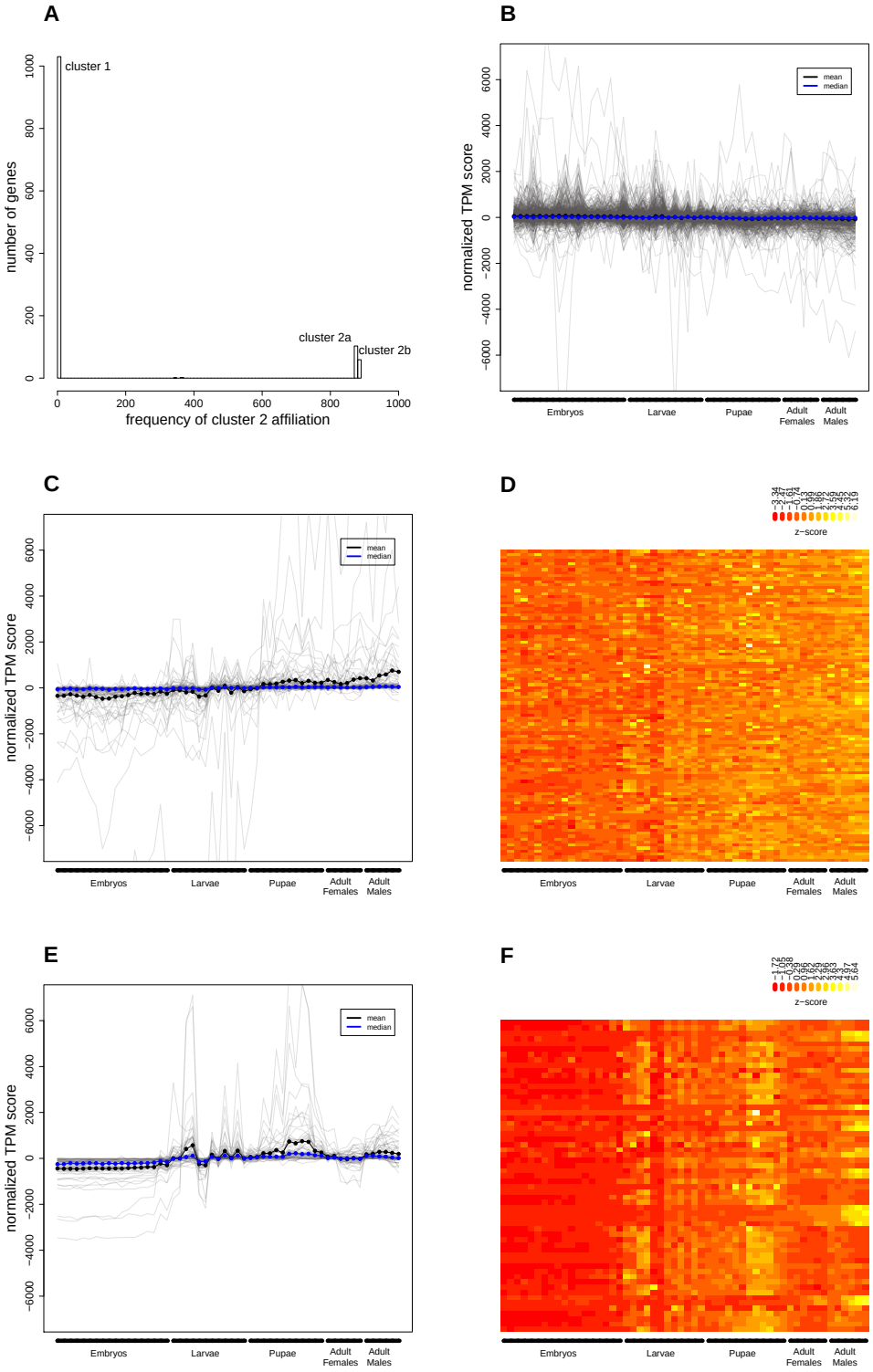
A
cluster 1
cluster 2a
cluster 2b
number of genes
frequency of cluster 2 affiliation
B
mean
median
normalized TPM score
Embryos
Larvae
Pupae
Adult Females
Adult Males
C
D
z−score
E
F

Figure 4.4: **Clustering analysis of *Wolbachia* gene expression in the modENCODE life cycle time course.** (A) Histogram showing the number of independent clustering runs (out of a 1000 runs) that a gene was affiliated to Cluster 2. Normalized expression profiles for genes in Cluster 1 are shown in panel (B), for genes in Cluster 2a in panel (C) and for genes in Cluster 2b in panel (E). Heat map of row-normalized expression levels for genes in Cluster 2a is shown in panel (D) and for genes in Cluster 2b in panel (F). For panels (B), (C) and (E), TPMs for each gene at each stage are normalized (shown in grey) by subtracting the mean TPM for that gene across different life cycle stages. The mean and median of normalized expression levels for all genes at each stage are show for each cluster in black and blue, respectively. For panels (D) and (F), row-normalized expression levels are visualized as a heatmap where each row represents a gene; expression levels is expressed in terms of Z-scores (observed TPM minus row mean TPM, divided by the standard deviation of TPMs for that row). Note that the heatmap color scale differs in panels (D) and (F).

4.2.3 A small subset of *Wolbachia* genes show dynamic expression across the host life cycle

Both correlation and clustering analysis reveal stable gene expression profiles for most *Wolbachia* genes across the life cycle of its host, but also that some variation exist for a subset of genes. The clustering analysis showed that most of the genes formed a first cluster with relatively constant expression across *D. melanogaster* life cycle, while a subset of 162 genes formed another cluster (Figure 4.4 A). Closer inspection of the number of runs in which a gene was assigned to Cluster 2 (Figure 4.4 A) shows that a group of 103 genes is affiliated to the second cluster 87.1% of the iterations and a group of 59 genes is affiliated to the second cluster 88.9% of the iterations. We will refer to these two groups as respectively Cluster 2a and Cluster 2b.

Both Cluster 2a and 2b show up-regulation after embryogenesis. The first, Cluster 2a, shows modest up-regulation after embryogenesis with the characteristic dip in expression at larval L3 (12 hrs) for the set of 103 genes (Figure 4.4C and D). The second, Cluster 2b, shows clearer up-regulation for 59 genes and the dip in expression at larval L3 (12 hrs) but stronger up-regulation in early larvae and late pupae (Figure 4.4E and F). For a few genes in Cluster 2b, the pattern is less strong but they show an important increase in expression in adult males after five days.

Genes found in Cluster 2 are a principal factor in the high correlation found between embryos and larvae in Figure 4.3. Indeed, when redoing the correlation heatmap after removing expression levels for those genes, we find that only the pupae and adult high correlation remains (data not shown). We propose that the correlation heatmap (Figure 4.3) reveals the stability of gene expression of *Wolbachia* across the host life cycle, and two clusters of more similar expression: one coming mainly from the expression of genes in Cluster 2, and the other coming mainly for a lower variability of gene expression in adult samples compared to the other samples. This could come from the fact that adult samples contain a higher number of bacteria, making the signal clearer.
The clustering experiment shows that despite the relative constancy of expression for *Wolbachia* genes at a whole organism level, some gene show not only variation across the host life cycle, but also a common expression pattern, revealing potential co-regulation and/or implication in a same function.

We also investigated if variation in expression pattern could be detected using a differential expression analysis on the samples with biological replicates (24 of the 30 stages). Because of the large number of life-cycle stages and their complex developmental dependencies, an analysis involving all-by-all pairwise comparisons was deemed unfeasible. Instead, we used an omnibus test of changes in expression across all stages simultaneously using an ANOVA-like GLM approach (149). This analysis allowed us to identify 80 genes (6.7%) whose expression changes in a reproducible manner across the *D. melanogaster* life cycle (in one or more stages) independently, using a threshold of adjusted p-value ≤ 0.05 (Figure 4.5). All 80 genes have a greater than two-fold change between at least one pair of stages. We note that both up- and down-regulated genes exhibit a wide range of absolute expression levels, and many show quantitative shifts rather than dramatic qualitative changes in expression level.
The majority of the *Wolbachia* genes found differentially expressed (75/80, 93.8%) show a common pattern of expression with higher expression in either larval, pupal and/or adult stages, and a transient decrease in expression at larval L3 (12 hrs) as described for the genes in Cluster 2. Indeed, 74 out of the 80 (92.5%) differentially expressed genes are found in Cluster 2 (a or b). Only six genes (GroES/WD0308, ABC transporter/WD0455, succinate dehydrogenase

subunit/WD0727, WD0804, DnaK/WD0928, Hsp90/WD1277) identified as differentially expressed in the life-cycle GLM were not found in Cluster 2 and are instead merged together with the stably-expressed genes in Cluster 1, most of which (with the exception of succinate dehydrogenase subunit/WD0727) are up-regulated in embryo, and down regulated in the later stages. Indeed, five genes show the pattern of having higher relative expression in embryos with down-regulation later in the life cycle (Figure 4.5). The dynamics of up-regulated and down-regulated genes show nearly complementary transitions at the end of embryogenesis, suggesting response to common signals or possible cross-talk between these gene sets.

Overall, both the differential expression and clustering results support the same conclusions that only a small proportion of *Wolbachia* genes show robust differences in expression across the modENCODE life-cycle time course at the level of the whole organism, and that the majority of dynamically-expressed genes show a common pattern of up-regulation after embryogenesis. However, the greater number of genes identified in Cluster 2 relative to those identified by the life-cycle GLM suggests that our differential expression analysis may have only detected a conservative subset of *Wolbachia* genes that show the strongest expression differences across *D. melanogaster* development.

4.2.4 Dynamically-expressed *Wolbachia* genes are predicted to be involved in stress response and host-microbe interactions

The 80 *Wolbachia* genes that exhibit dynamic expression across the *D. melanogaster* life-cycle fall into three broad classes (Figure 4.5). The first is a small class of five genes that show high relative expression in embryos with down-regulation later in the life cycle. Three of these genes are involved in chaperone function (GroES/WD0308, DnaK/WD0928, and Hsp90/WD1277). The chaperone GroEL/WD0307, which putatively forms a complex with GroES/WD0308, is co-transcribed with GroES/WD0308 and shows similar down-regulation at later stages of the life cycle, but does not pass the significance threshold in the life-cycle GLM (adjusted p-value=0.15). Both GroES/WD0308 and GroEL/WD0307 are in the top 15 most abundant transcripts based on average TPM across all stages,

Figure 4.5: **A small subset of *Wolbachia* genes show differential expression across the *D. melanogaster* life-cycle.** Row-normalized expression levels are visualized as a heatmap where each row represents a gene (ordered top-to-bottom by its position in the genome) and each cell represents the relative expression level for a particular sample in terms of Z-scores (observed TPM minus row mean TPM, divided by the standard deviation of TPMs for that row). Gene names and identifiers are shown on the left. Membership in dynamically-expressed gene classes is shown by dots on the right. Stages that lack biological replicates in the modENCODE total RNAseq time course were not used in this analysis and are not shown here.

confirming that chaperones are among the most highly expressed genes in *Wolbachia* (52; 147; 150). High basal expression of GroEL or other chaperone proteins was suggested to be a compensatory mechanism for the accumulation of slightly deleterious non-synonymous mutations in endosymbionts that arise because of their small population size and lack of recombination (151; 152). The differential expression of *Wolbachia* chaperones during the *D. melanogaster* life cycle that we observe may indicate higher stress level for *Wolbachia* in the embryo compared to larvae, pupae and adults.

The second class, comprising the majority of up-regulated genes detected (57/80), shows an increase in relative expression starting at the larval L1 or L2 stages and persisting until adult stages, with a decrease in expression level at the larval L3 (12 hr) stage and stronger up-regulation at the white pre-pupal 2 and 3 day stages. Genes in this class are mostly unannotated, but include eight genes that code for proteins with membrane or secretion system function (WspB/WD0009, TerC/WD0194, SPFH domain/WD0482, type II secretion/WD0500, HlyD/WD0649, type I secretion/WD0770, VirB3/WD0859, Rhoptry surface protein related/WD1041) and four ANK-containing genes (WD0191, WD0385, WD0438, WD1213). The ANK-containing genes from several bacterial species have been shown to be type IV secretion system effector molecules that have diverse effects on eukaryotic cells (109; 153; 154; 155; 156). Thus, these results suggest secretion of ANK-containing genes into the host cell may be enriched during early larval and mid-to-late pupal stages of *D. melanogaster* development. Up-regulation of components for secretion systems (type III) has been observed in pupal stages for other arthropod endosymbionts (157), suggesting that metamorphosis may be a general period that is enriched for up-regulation of secreted symbiont effector proteins involved in host interaction. These genes could also be involved in cytoplasmic incompatibility and have some early effects on male sperm. Indeed, the density of *Wolbachia* increases in testes of male pupae, where those genes show high up-regulation (98). Other effects such as tissue colonization or protection against viruses could be connected to these up-regulated genes. Indeed, it has been shown that *Wolbachia* in embryos stay in the same location and start colonizing tissues only at the end of embryogenesis (90). This could also explain the higher expression in late pupal stages where the adult tissues start differentiating and where *Wolbachia* might need to colonize the newly formed tissues.

A third class of 22 *Wolbachia* genes show up-regulation primarily in *D. melanogaster* adults, with higher expression in adult males relative to adult females at the same age (see more below). Most of the genes in this "sex-biased" class are also unannotated, however three are ANK-containing genes (WD0291, WD0292, WD0438). Our observation of sex-specific expression of ANK-containing genes based on global gene expression profiles of *Wolbachia* in *D. melanogaster* extends results from targeted RT-PCR analysis in *Wolbachia* strains from other insects (107; 113; 114; 115; 158), and provides further evidence for their possible involvement in cytoplasmic incompatibility.
Finally, we note that our qualitative classification of up-regulated genes in classes 2 and 3 is not mutually exclusive, and the existence of four genes (WD0438, WD1288, WD1289 and WD1290) with sex-biased expression that also show differential expression at larval or pupal stages suggests possible shared regulation of these classes.

4.2.5 Some *Wolbachia* genes have a sex-biased expression with an age-dependent effects

The global clustering and differential expression analysis showed that some *Wolbachia* genes are preferentially expressed in adult males. To further explore gene expression variation in response to the host sex, we did a pairwise differential expression analysis between males and females at matched developmental stages. For this analysis, we used an exact testing framework (141) because the GLM-based approach used for the complete life cycle is not the optimal method to use in a pairwise context.
We identified a total of 41 genes that exhibited greater than 1.5-fold difference at an adjusted p-value cutoff of 0.01 in pairwise tests between male and female samples at either 1, 5 or 30 days post-eclosion (Figure 4.6). Most of the genes differentially expressed between males and females have a higher expression in males than in females. The expression of those genes seems to increase with age for both males and females, but do so earlier and more drastically in males than in females. This might explain why more genes are detected as differentially expressed between 5 days old (35 genes) males and females than at one and thirty days old (1 and 15 genes respectively). This demonstrates that sex-biased gene expression in *Wolbachia* is also age-dependant.

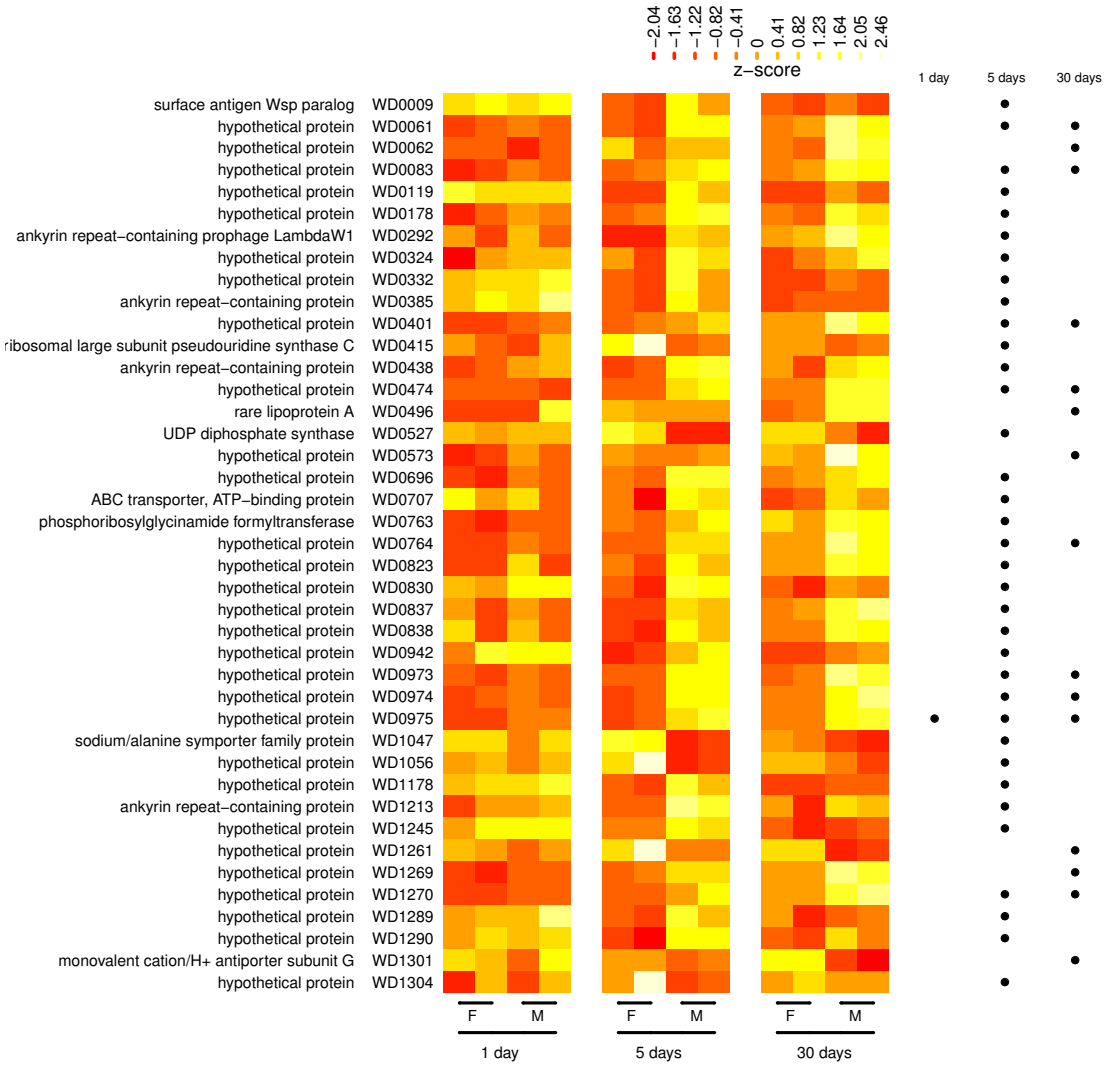

Figure 4.6: ***Wolbachia* genes show age-dependent sex-biased expression.** Row-normalized expression levels are visualized as a heatmap where each row represents a gene (ordered top-to-bottom by its position in the genome), and each cell represents the relative expression level for a particular sample in terms of Z-scores (observed TPM minus row mean TPM, divided by the standard deviation of TPMs for that row). For each stage, two biological replicates are shown for each female (F) and male (M) sample as distinct columns. Values higher than row means are represented by yellow, and values lower than row means are represented by red. Gene names and identifiers are shown on the left. Dots on the right indicate if a gene is differentially expressed between males and females at 1-day, 5-days or 30-days post eclosion, respectively.

Our whole-organism RNAseq analysis at 5 days post-eclosion correctly predicted the presence (3/3) or lack (13/14) of sex-biased expression differences for 16/17 ANK-containing genes in a *w*Mel strain previously classified by RT-qPCR to have

over 1.5-fold difference in expression level between testes and ovaries of 2-day old flies (115) (the only exception being that WD0292 shows sex-biased expression in the RNAseq data at 5-days that is not observed in the RT-qPCR at 2-days). The capacity to detect genes differentially expressed between adult gonads using RNAseq on whole organism suggests that gene expression patterns observed in adults are largely dominated by *Wolbachia* in ovaries and testes and/or that the quality of the RNAseq data from modENCODE is good enough to detect small yet significant variation in expression.

The majority of sex-biased genes in the pairwise male-vs-female analyses showed higher expression in males relative to females at matched stages, with only seven genes (rluC/WD0415, uppS/WD0527, sodium/alanine symporter/WD1047, WD1056, WD1261, cation antiporter subunit G/WD1301, WD1304) showing relatively higher expression in females at one or more time point. This could be due to the fact that modifications leading to cytoplasmic incompatibility mainly happen in the sperm cells (77).

Most sex-biased genes in this analysis were identified in one or both of the up-regulated classes in the life-cycle GLM above (28/41, 68.3%), indicating these complementary approaches identify a similar set of *Wolbachia* genes with detectable sex-biased expression in the modENCODE data set. Likewise, sex-biased genes comprise over one-third of differentially-expressed genes identified in the life-cycle GLM (28/80, 35%), suggesting that sex-biased expression is a dominant component of major differences in *Wolbachia* gene expression that can be observed across the *D. melanogaster* life cycle. This could also mean that genes implicated in sex-biased function are also implicated in other non-sex-related functions, or that the sex-biased function have a basis in earlier life stages.

Our finding that *Wolbachia* genes with sex-biased expression also show age-dependent effects is consistent with a decline in the strength of cytoplasmic incompatibility reported in males at 1-day versus 5-day post eclosion in *D. melanogaster* (1; 159; 160). The observed pattern of sex-biased genes being up-regulated in older males is compatible with these *Wolbachia* genes playing a role in attenuating the modification of *D. melanogaster* sperm that leads to embryonic lethality in incompatible crosses (161). Alternatively, if the host is responsible for reducing the effects of *Wolbachia* on the sperm of older males, up-regulation of *Wolbachia* genes in older males could represent a compensatory effect and hence indicate these genes play a role in promoting cytoplasmic incompatibility.

4.3 Conclusion

The work presented here represents the most comprehensive gene expression profiling to date of an endosymbiotic bacteria in its native host context. We establish that most *Wolbachia* genes are expressed in all *D. melanogaster* life-cycle stages, but that major changes in expression levels of *Wolbachia* genes are rare when studied simultaneously across all *D. melanogaster* tissues. Additional work is needed to rule out lack of power to detect real differences due to the low number of replicates and averaging of tissue-specific expression in our study, but global stability in *Wolbachia* expression is mechanistically consistent with the limited number of regulatory genes encoded in the *w*Mel genome (99). Global stability of *Wolbachia* gene expression across the *D. melanogaster* life-cycle may be an adaptive response that simply reflects the stable environment of an intracellular endosymbiont, or that suggests *Wolbachia* co-exists in non-obligatory symbioses largely in a stable "stealth mode" that reduces the chances of being detected by the host as a pathogen.

Nevertheless, a set of 93 *Wolbachia w*Mel genes that show robust stage- and/or sex-specific differential expression can be identified at the whole-fly level, many of which share common expression dynamics and therefore may be co-regulated. These genes provide many new candidate genes for understanding, and possibly manipulating, the genetic basis of how *Wolbachia* interacts with arthropod hosts. Furthermore, differences in expression levels among *Wolbachia* genes previously implicated to mediate host-symbiont interaction can provide insight into the likelihood and mechanistic basis of existing candidates.

Importantly, we also provide the first detailed insight into the developmental dynamics of *Wolbachia* gene expression in an insect host, which suggest larval and pupal stages (where *Wolbachia* have been detected cytologically (162)) merit further study to understand how *Wolbachia* manipulates host biology to maintain persistent infections and affect transmission.

The fact that the same genes are found showing stage-specific and sex-specific expression could be a sign of overlapping functions or regulation, but could also be a sign of generally reduced regulation of gene expression, with only a few genes remaining susceptible to regulation. Investigating gene expression of *Wolbachia* in different context and or tissue might help understanding if more genes can show variation in expression when the bacteria is under different conditions.

Chapter 5

Comparison of *Wolbachia* gene expression in different contexts

5.1 Introduction

As described in Chapter 1, *Wolbachia* has a large diversity of effects, positive or negative, on its hosts (4; 56; 61; 62; 63; 64; 65), even inside a single species (3; 88). We confirmed in Chapter 3, that in *D. melanogaster*, *Wolbachia* is separated in two main groups: *w*Mel and *w*MelCS (17). The effect on *D. melanogaster* varies between the two groups of strains: for instance, *w*MelCS induces higher levels of CI and protection against viruses than *w*Mel (3; 4). *w*MelCS also has deleterious effects on *D. melanogaster* longevity that are not observed in *w*Mel infections (3). At a finer scale, different *Wolbachia* clades found in different region of the world, demonstrate variation in efficiency when in competition at cold temperatures (11).

In Chapter 4, we showed that a group of 80 genes from *Wolbachia* *w*Mel vary across the life cycle of *D. melanogaster*. Most of those genes are up-regulated after the embryonic stages, with higher expression in pupae and adults. Only a few genes were up-regulated in the embryo, and these were mainly involved in stress response. A group of 41 genes were found differentially expressed between males and females, mostly at 5 days post eclosion, and most of which also show variation during host development.

Given the variable effects that *Wolbachia* can have on its hosts, we wanted to test whether these findings apply to different conditions and strains. Indeed, in order to leverage those findings for further studies, it is important to know

how specific they are to one strain and host context. Studying *Wolbachia* gene expression in other strains also can give us insights into the gene expression regulation capacities of *Wolbachia* in *D. melanogaster*.
Here, we investigated *Wolbachia* gene expression variation in a different host-microbe context using data from Czech *et al.,* (132). This dataset consists of total RNA from embryos, ovaries from one day old females and the ovarectomized one day old female carcasses from a *nos*-gal4, UAS-*Dcr*-2 transgene line of *D. melanogaster*, referred to as DCR2. We built a phylogeny using the genome from the *Wolbachia* strain infecting the DCR2 line and other genomes from known clades, and determined that the DCR2 line of *D. melanogaster* is infected with a *w*MelCS-like genotype of *Wolbachia* (Figure 4.1). The nos-gal4, UAS-*Dcr*-2 transgene induce an over-expression of *Dcr*-2 specifically in ovaries (132). The over-expression of *Dcr*-2 in ovaries enhance the RNA interference effects (163). This is a limitation of this dataset as we can't predict how this might affect *Wolbachia* biology in ovaries, and its gene expression.
In this chapter, we first show that a few genes are differentially expressed between embryo, ovaries and female carcasses in the DCR2 line of *D. melanogaster*. Then we compare those results to genes found differentially expressed between *Wolbachia* infecting embryo and adult females of ISO1 line of *D. melanogaster*, using the stages the closest to the samples from the DCR2 line.

5.2 Results and discussion

5.2.1 *Wolbachia* gene expression can be detected in the DCR2 line of *D. melanogaster*

For the six samples in the DCR2 dataset, a median of 0.4 out of 13.8 million reads (2.9 percent) mapped to *Wolbachia* genome. This is lower than the median of 1.7 million reads for the data from ISO1, but gene expression is still detected for 1123 of the 1195 genes (TPM > 0). The Figure 5.1 shows the amount of read mapping to *Wolbachia* genome in a window containing the genes WD0968 to WD0978. Compared to the same plot for ISO1 in Figure 4.2, the coverage is much lower for the DCR2 dataset. Indeed, the scale goes only up to 30 in DCR2 but up to 150 in ISO1. The orientation of most of the mapped reads in the DCR2 dataset is consistent with the orientation of the genes, but a small number of antisense

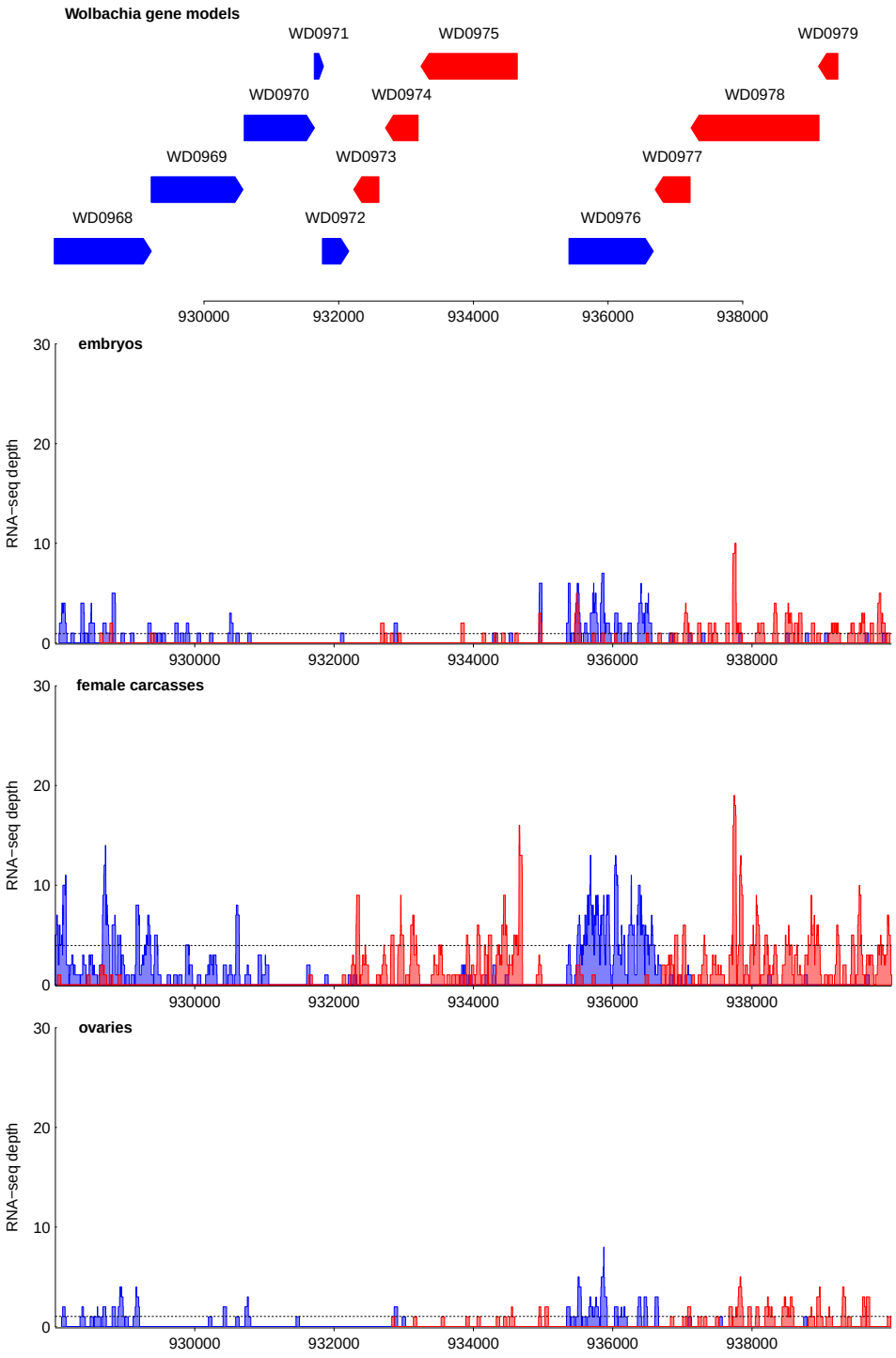

Figure 5.1: **Expression landscape of *Wolbachia* in Dcr2 transgene nos-gal4 driver *D. melanogaster*.** Gene models and RNA-seq coverage plots for a 12-gene window of the *Wolbachia* genome showing gene expression levels in representative stages of the *D. melanogaster* DCR2 dataset. Gene models and RNA-seq coverage are shown on the forward and reverse strands in blue and red, respectively. RNA-seq plots are shown on the same absolute y-axis scale. To provide an internal normalization factor for comparison across samples, mean coverage of the stably-expressed Wsp/WD1063 gene (not shown in this interval) divided by twenty is depicted by the dashed line in each panel.

reads, possibly coming from a low amount of genomic contamination, are visible. Overall the DCR2 data, although of lower quality than the ISO1 dataset, provides expression data for almost all *Wolbachia* genes enabling qualitative evaluation of gene expression between the different time points and tissues in a second strain of *D. melanogaster*.

5.2.2 A small group of genes show changes in expression between tissues and stages in DCR2.

We used a pairwise differential expression analysis to compare gene expression between the embryo, carcasses and ovaries. We found that 118 genes are differentially expressed between embryo and carcasses, 18 between ovaries and carcasses and 59 between embryo and ovaries. As in ISO1, the majority of genes don't show detectable variation between the different life stages. It is important to note that due to the lesser quality of the data compared to ISO1 samples, the sensitivity to detect smaller differences in expression level between stages is also lower.
All genes differentially expressed between at least two of the DCR2 samples (138 genes in total) are shown in Figure 5.2, 5.3 and 5.4. The columns on the right show in which pairs of samples the genes are found to be differentially expressed. A venn diagram in Figure 5.5 shows how genes differentially expressed overlap between the different pairwise comparisons.
Comparing *Wolbachia* gene expression between *D. melanogaster* tissues, in ovaries and female carcasses, we can see that only 18 genes are differentially expressed. This is consistent with the findings from the ISO1 dataset where most of *Wolbachia* genes keep a relatively similar gene expression between the host time stages and seems to expand this rule to the variations between female tissues as well. For the genes that do show variation in expression between ovaries and female carcasses, most of them are up-regulated in female carcasses compared to ovaries, suggesting a role in female soma rather than in ovaries. Most of the genes showing differential expression between ovaries and female carcasses, also show differential expression between embryo and female carcasses (Figure 5.5), in both case, being up-regulated in female carcasses. In the DCR2 line of *D. melanogaster*, *Wolbachia* gene expression does not show much differences between ovaries and carcasses, and when it does, the expression in ovaries is similar to embryos.

While some of the 118 genes differentially expressed between embryo and female carcasses are up-regulated in female carcasses and show similar differences between ovaries and female carcasses, the majority of them are up-regulated in embryos. This differs compared to the findings for ISO1 in Chapter 4, where most of the genes showing differential expression across the life cycle were up-regulated after embryonic stages. In the heatmaps in Figure 5.2, 5.3 and 5.4 there is a visually notable difference in color between embryo and other samples. This shows that *Wolbachia* in embryonic samples of the DCR2 dataset has a quite distinct gene expression profile compared to female carcasses and ovaries. Ideed, the differential expression between the embryo and other stages counts for most of the differentially expressed genes between all samples (118 out of 138) and many genes differentially expressed between embryo and carcasses, are also differentially expressed between embryo and ovaries (46 out of 138) (Figure 5.5). Moreover, a lot of the genes, up-regulated in embryo compared to female carcasses, not significantly differentially expressed between embryo and ovaries still show a similar trend, with a lower expression in ovary samples. This also shows that the number of differentially expressed genes is probably an underestimation, likely due to the lowed data quality and therefore a higher variation between replicates.
The effects of the *Dcr*-2 over-expression in ovaries could be disrupting the expression pattern of *Wolbachia* in ovaries. However, unlike in embryo, *Wolbachia* in ovaries does not show a peculiar expression compared to the two other samples: most of the genes showing difference in expression between ovaries and female carcasses also show differential expression between embryo and female carcasses; and most of the genes showing differential expression between ovaries and embryo also show differential expression between female carcasses and embryo (Figure 5.5). However, it is a possibility that the low amount of differentially expressed genes between ovaries and carcasses is the result of a *Dcr*-2 disruption of *Wolbachia* gene expression.

5.2.3 Differentially expressed genes between stages and tissues belong to distinct functional groups

Genes differentially expressed between embryo and female carcasses or ovaries, up-regulated in embryo, are mainly involved in stress response, transcription,

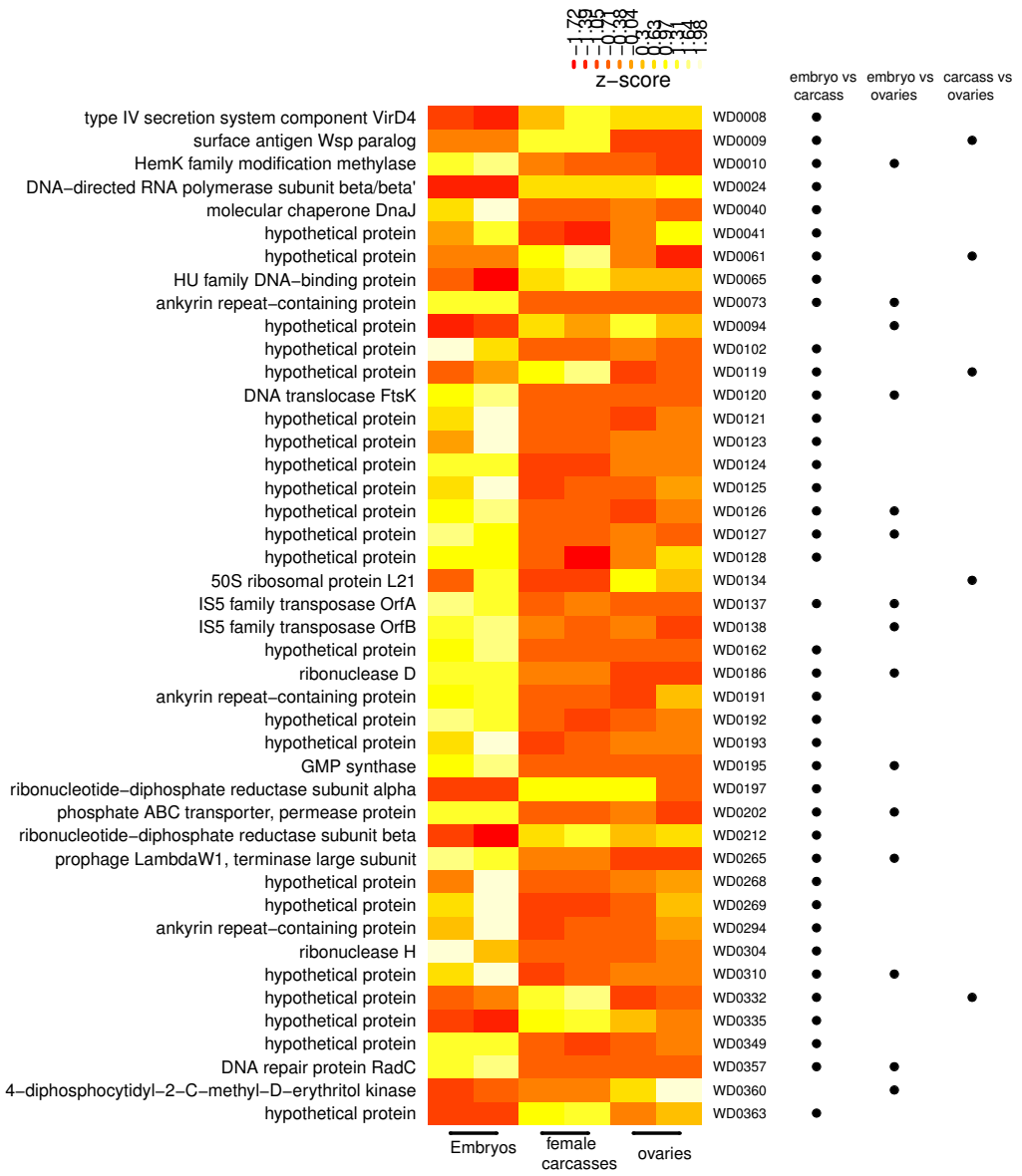

Figure 5.2: ***Wolbachia*** **genes differentially expressed between** ***Dcr*****-2 samples. Part 1/3.**

Genes differentially expressed between embryos and adults, adults and ovaries, or ovaries and embryos in DCR2. Each row represents a gene in order top-to-bottom of its position in the genome, and each cell represents the relative expression level for a particular sample in terms of Z-scores (observed TPM minus row mean TPM, divided by the standard deviation of TPMs for that row). Values higher than row means are represented by yellow, and values lower than row means are represented by red. The column on the right of the heatmap show in which pairwise comparison a gene is found to be differentially expressed.

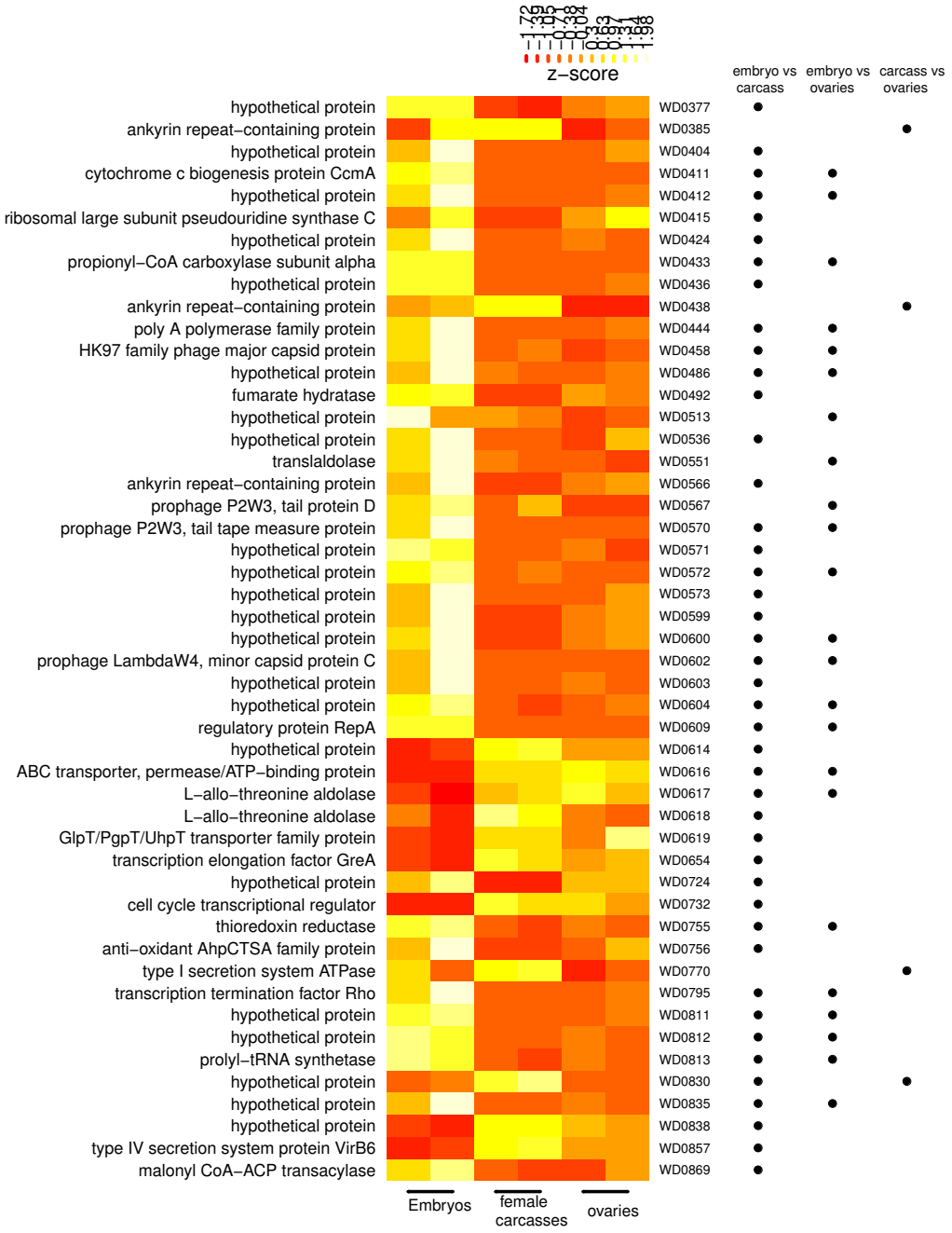

Figure 5.3: ***Wolbachia* genes differentially expressed between dcr2 samples. Part 2/3.**

translation or DNA replication (Figure 5.2, 5.3 and 5.4). Besides those main functions, a number of differentially expressed genes between embryo and female carcasses and/or ovaries are phage genes and parts of the IS5 transposable element (Figure 5.2, 5.3 and 5.4).

The large number of metabolic genes with a higher expression in embryo than

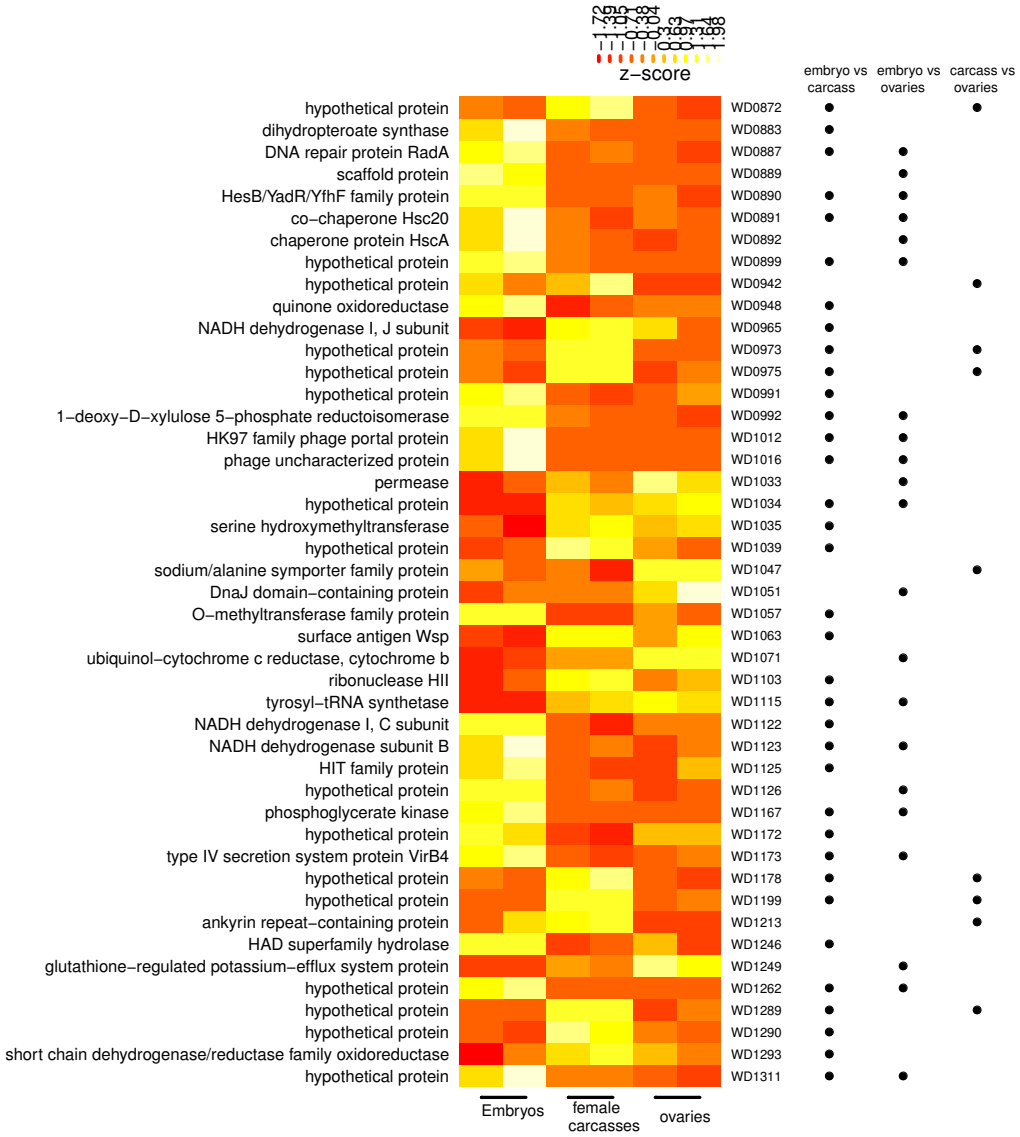

Figure 5.4: ***Wolbachia* genes differentially expressed between dcr2 samples. Part 3/3.**

in the adult tissues could be due to a stress response (164). Moreover, it was shown that phage titres in *Wolbachia* increase in response to heat-shock in *Nasonia vitripennis* (165). While the embryo from the DCR2 dataset were not heat-shocked, increase in phage titers might be part of a more general stress response, potentially activated by other stimuli. Results from ISO1 already suggested that up-regulated genes in embryo were linked with stress response. However, the number of genes up-regulated in the DCR2 embryo is much higher than the few genes observed in ISO1 dataset. The difference in the intensity of the

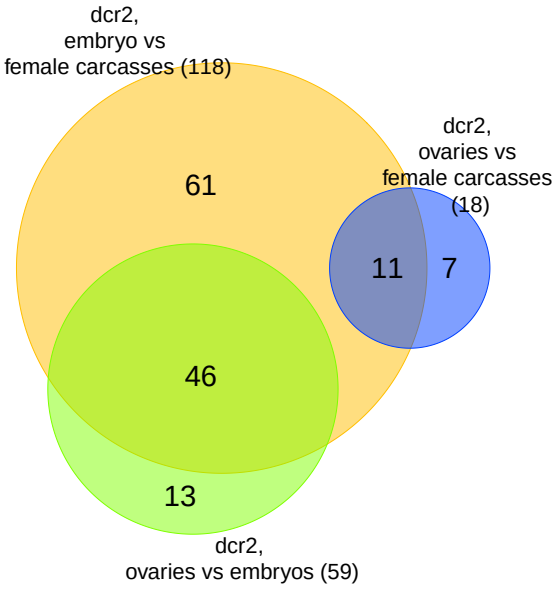

Figure 5.5: **Overlap between groups of differentially expressed genes in *Dcr-2*** Venn diagram of the genes differentially expressed between embryos and adults, adults and ovaries, or ovaries and embryos in DCR2. The circle areas are proportional to the size of the gene groups.

embryo response could be due to the *Wolbachia* strain, the host background or environmental conditions. It has been shown by Chrostek *et al.* (3), that the *w*MelCS variant, while providing better virus protection to its *D. melanogaster* host, showed higher titers than the *w*Mel variant. The high number of metabolic genes expressed in *Wolbachia* *w*MelCS infecting *D. melanogaster* embryo might be partly due to the high replication of this *Wolbachia* strain, making the embryo a potential key stage to determine *Wolbachia* titers in *D. melanogaster*. However, another study suggested that the *Wolbachia* titers were mostly controlled by the *D. melanogaster* rather than the *Wolbachia* strain (90).

The small proportion of genes differentially expressed between embryo and female carcasses, down-regulated in embryo compared to female carcasses, are

mainly genes coding for membrane protein or involved in secretion systems; and the genes differentially expressed between ovaries and carcasses, mostly up-regulated in female carcasses, are mainly secretion system proteins of membrane proteins, similarly to the functions of differentially expressed genes in ISO1 that are mostly up-regulated in adults (Figure 5.2, 5.3 and 5.4).

5.2.4 Comparison of differentially expressed genes in DRC2 with genes differentially expressed in ISO1

After identifying the genes differentially expressed between samples of the DCR2 dataset, we compared the expression pattern of *Wolbachia* *w*MelCS in DCR2 to the expression pattern of *Wolbachia* *w*Mel in ISO1. Since the host background, *Wolbachia* genotype, and experimental procedures are all batch effects that differ for the DCR2 and ISO1 datasets, we cannot perform direct quantitative analyses between these datasets, nor can we fully control for all factors in determining the causal basis of any potential differences that we may observe in their *Wolbachia* gene expression. However, comparison of DCR2 life-cycle stages *inter se* does offer the ability to test the generality of our findings about stage-specific *Wolbachia* expression based on comparisons within ISO1. Using the ovarectomized carcasses in DCR2 to compare with adult females from ISO1 data can be an issue if most of the expression pattern observed in the adult females in ISO1 comes from *Wolbachia* in the ovaries. This is a possibility since *Wolbachia* density is high in females ovaries (96). However, comparing expression pattern of *Wolbachia* in adult females with its expression in females carcasses and ovaries, can give insights as to whether the expression pattern observed in adult females samples is driven by *Wolbachia* in somatic tissues or ovaries.

To compare the differentially expressed genes in ISO1 and DCR2, we decided to compare the set of genes showing differential expression between embryo, ovaries and female carcasses in DCR2 with the set of genes differentially expressed between embryo and adult females in ISO. To do so, we used a pairwise differential expression analysis between the ISO1 stages closest to the DCR2 samples: embryo 0-2 hours and adult females one day old. We found 65 genes differentially expressed between those two time points in ISO1 dataset. 54 of these genes were found in the set of genes found differentially expressed across the life cycle in ISO1.

Figure 5.6 shows a heatmap visualization of the genes differentially expressed between samples from the DCR2 dataset and embryo and adult females in the ISO1 dataset. The block A represents 8 genes differentially expressed between embryo and carcasses in DCR2 and embryo and adult females in ISO1 but not between ovaries and carcasses in DCR2. The block B are 11 genes differentially expressed between embryo and carcasses in DCR2, ovaries and carcasses in DCR2 and embryo and female adult in ISO1. The block C contains 7 genes differentially expressed between ovaries and carcasses in DCR2 and embryo and adult female in ISO1, but not between embryo and carcasses in DCR2. The genes in blocs D to F are genes not found to be differentially expressed between embryo and adult females in ISO1. The block D represents 54 genes differentially expressed between embryo and carcasses in DCR2, not found to be differentially expressed between embryo and ovaries in DCR2. The block E represents 45 genes differentially expressed between both embryo versus carcasses in DCR2 and embryo versus ovaries in DCR2, and the bock F shows 13 genes differentially expressed between embryos and ovaries in DCR2 but not between embryo and carcasses in DCR2. Finally the block G shows the 41 genes differentially expressed between embryo and adult females in the ISO1 dataset, but not differentially expressed between samples of the DCR2 dataset.

As mentioned earlier, only 18 genes were found differentially expressed between ovaries and female carcasses in the *D. melanogaster* DCR2 line. We started by comparing the genes differentially expressed between ovaries and carcasses in the DCR2 dataset, to the genes differentially expressed between embryo and whole females in the ISO1 dataset. As shown in the Venn diagram Figure 5.7, 16 of the 18 genes differentially expressed between ovaries and carcasses are also found differentially expressed between embryo and adult females in the ISO1 dataset. Figure 5.6 B and C shows the 18 genes differentially expressed between ovaries and carcasses in DCR2. We can see that for all of them, including the 2 genes not differentially expressed between embryo and adult females in ISO1, the gene expression pattern is similar in ISO1 with embryo being similar to ovaries. This confirms that, where differences between ovaries and female carcasses are observed, *Wolbachia* in ovaries seem to behave in a more similar way to *Wolbachia* in embryo than in female carcasses. Moreover, this indicates that the females carcasses in DCR2 can be used for the comparison with adult whole females in ISO1 since they show similar expression for genes showing variation between

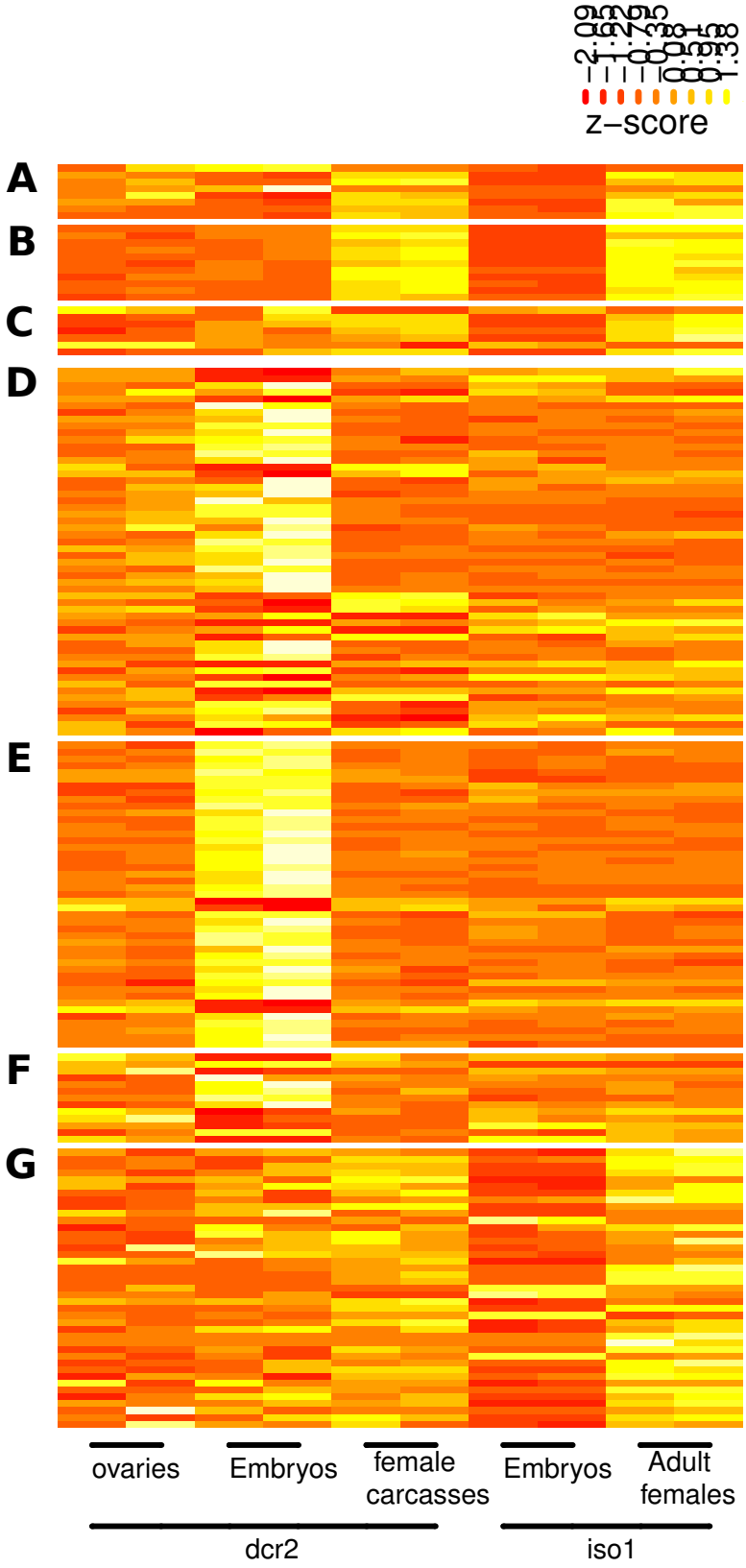

Figure 5.6: Summary of the comparison of *Wolbachia* gene expression in ISO1 and DCR2 background. Row-normalized expression levels are visualized as heatmaps where each row represents a gene, clustered by presence in different pairwise comparions. Relative expression is represented by Z-scores (observed TPM minus row mean TPM, divided by the standard deviation of TPMs for that row).

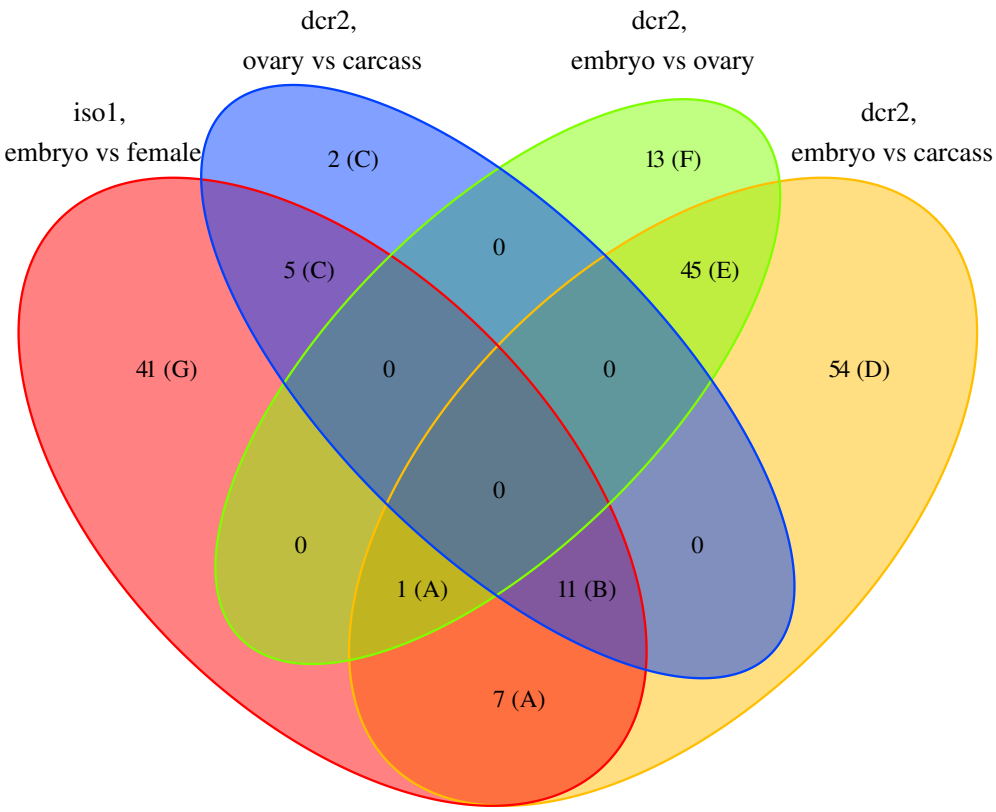

Figure 5.7: Venn diagram of the comparison of *Wolbachia* gene expression in ISO1 and DCR2 background. Diagram of genes found to be differentially expressed in the three pairwise analysis between DCR2 samples and embryo vs adult female in ISO1. Next to the number in each section is displayed a letter corresponding to the bloc in which those genes are displayed in the Figure 5.6.

ovaries and female carcasses. This also suggest that *Wolbachia* expression pattern in female might not be driven by the expression in ovaries.

Of the 118 genes differentially expressed between embryo and females carcasses in DCR2, 19 are also found differentially expressed between embryo and adult females in the ISO1 dataset (Figure 5.7). Figure 5.6 A and B show the heatmap of the 19 genes differentially expressed between embryo and female adult/carcasses in both ISO1 and DCR2. For most of them, as seen in Figure 5.7, the expression in ovaries is similar to the expression in embryo with a significative difference to female carcasses (Figure 5.6 B). For the rest of the 19 genes, the *Wolbachia* gene expression pattern in ovaries tends to be similar to the expression in embryo although not in a significant way (Figure 5.6 A). There are therefore a few genes where expression varies between embryo and adult females or carcasses regardless of changes in context such as *Wolbachia* and *D. melanogaster* strain, and for which

the expression in ovaries is similar to the expression in embryos.
However, for most of the genes differentially expressed between embryo and female carcasses in DCR2, no differences are found between embryo and adult females in ISO1 (99 genes out of 118). For this set of genes, displayed in Figure 5.6 D and E, the *Wolbachia* gene expression in ISO1 is more similar to the expression of *Wolbachia* in ovaries and carcasses. This is also the case for the 59 genes differentially expressed between ovaries and embryo in DCR2, as shown in Figure 5.6 E and F. All but one of the genes differentially expressed between ovaries and embryo do not show differential expression between embryo and adult females in ISO1 (Figure 5.7). This confirm that the strong up-regulation of those genes is specific to the embryo samples of the DCR2 dataset.
Finally, the block G of the Figure 5.6 show the 41 genes found differentially expressed between embryo and adult females in ISO1 but not found differentially expressed between samples of the DCR2 dataset. For this set of genes, while no significant differences can be observed between samples from the DCR2 dataset, we can see that while the female carcasses sample is more similar to the adult female samples from ISO1, the DCR2 sample which is more similar to ISO1 embryo, is the DCR2 ovary samples. The greater similarity between ovaries from DCR2 and embryo from ISO1 than between embryos from DCR2 and ISO1 confirms the particularity of *Wolbachia* gene expression in embryo samples from DCR2, which could be due to host context, *Wolbachia* strain or experimental conditions.
The absence of significant differential expression between ovaries and female carcasses for the genes differentially expressed between embryo and adult female in ISO1 where a similar trend can be seen might be due to the lower quality of the data in the DCR2 dataset, or to the strong up-regulation in DCR2 embryo masking more subtle differences. However, the trend observed for many of those 41 genes between ovaries and female carcasses, together with the embryo-like expression of ovaries for the 18 genes differentially expressed between ovaries and female carcasses, suggest that the expression patter in females for those genes is not driven by *Wolbachia* in ovaries. As 26 of the 65 genes differentially expressed between embryo and adult females in the ISO1 dataset are also found differentially expressed between males and females in the ISO1 dataset, it is possible that some of the sex-biased genes in ISO1 are detected as down-regulated in females because of a significant proportion of *Wolbachia* in ovaries, displaying an

embryo-like expression. A study by Papafotiou *et al.* (115) shows that for two genes (WD0438 and WD1213) differentially expressed between adult testes and ovaries (which we find differentially expressed between ovaries and female carcasses), the expression in adult soma is similar between males and females, while the expression is lower in ovaries and higher in testes. This is consistent with the difference we find between ovaries and female carcasses, and suggest that the embryo-like expression in ovaries would not account for all the difference between males and females for genes up-regulated in males, since the expression in testes is higher than in males and females soma.

The comparison of *Wolbachia* gene expression pattern between ISO1 and DCR2 datasets confirmed that embryo samples in DCR2 have a very specific up-regulated set of genes, and that ovaries from the samples in DCR2 have an expression pattern more similar to embryo than to female carcasses for the genes that show variation in expression between embryo and adults in ISO1.

5.2.5 Summary and conclusion

Here we show that *Wolbachia* gene expression in the DCR2 dataset shows detectable variation between samples only for a small number of genes (138 out of 1195 genes detected to be differentially expressed). This stability in expression, while partly due to the low sensitivity of the differential expression analysis on this dataset, is broadly consistent with what was seen in ISO1 in Chapter 4. When variation is detected, genes fall into two broad categories: one set of genes shows variations between embryo and the two other samples (ovaries and carcasses) and are involved in stress response, cell cycle and phage/IS5 genes; a second set of genes, show variation between embryo and/or ovaries and carcasses in DCR2 in the same way they do between embryo and adult females in ISO1 dataset.
The first three blocks in Figure 5.6 contains the genes for which variation in expression is relatively conserved between ISO1 and DCR2. The differences between embryos and ovaries when compared to carcasses, in the first and third blocks, could be biological differences between *Wolbachia* expression between embryo and ovaries. But, as we can see in the venn diagram (Figure 5.7), only one of those genes (in block one) is also found differentially expressed between embryo and

ovaries. It is possible that the noise in the data or the peculiarity of the *Wolbachia* expression in the embryonic samples of DCR2 where some genes show a very strong up-regulation masks more subtle differences in expression.
The similarity in gene expression patterns between embryos and ovaries for those genes might give some indications regarding the role of those genes. Indeed, it has been shown that the distributions of *Wolbachia* inside the oocyte takes place in the ovaries and does not change in embryo. Genes with similar expression in ovaries and embryo are unlikely to be involved in this process. However, genes up-regulated in somatic tissues of adults could be involved in other function such as behavioural changes, protection against viruses, metabolic variations etc...

The main part of the expression variation between the DCR2 samples is found between the embryo and the other samples and is mostly with genes for which no changes are found between *Wolbachia* gene expression in embryos and adult females in ISO1. The genes differentially expressed between embryos and carcasses in DCR2 split in two groups: one that clusters with the genes differentially expressed between ovaries and carcasses in DCR2 and embryos and adult females in ISO1, and one that clusters with genes differentially expressed between embryos and ovaries in DCR2. Genes in the first set are mainly up-regulated in carcasses and genes in the second set are mainly up-regulated in the embryo. The second set of genes is represented in the heatmap Figure 5.6, blocks D to F. In those blocks, the expression of *Wolbachia* in embryonic samples from DCR2 is different to all the other samples, despite a lack of statistical detections for one of the pairwise comparison in blocks D and F. The genes from the second cluster notably contain IS5 transposable elements and prophage genes. It is not clear whether the differences in expression observed between the two datasets are due to the *Wolbachia* strain difference or to the host background. However it has been shown that the transposable elements found differentially expressed are present in both strain (104) and therefore the difference observed is on the expression rather than presence/absence of the sequence.
The main differences we found between *Wolbachia* *w*Mel gene expression in ISO1 line and *Wolbachia* *w*MelCS gene expression in DCR2 line are: (i) There are fewer genes up-regulated in female carcasses compared to embryo and/or ovaries in the DCR2 dataset than genes up-regulated in females compare to embryos in the ISO1 dataset. This difference, while probably not due to the absence of

ovaries in the female carcasses from DCR2, could be due to biological differences coming from the *Wolbachia* strain or *D. melanogaster* line. However, there is a possibility that it is due to a noisier dataset in DCR2. (ii) There is a large number of genes up-regulated in embryos from DCR2 dataset compared the ovaries and females carcasses. The genes up-regulated in embryo could correspond to a stress response, or high replication in the embryo. It is unclear if the strong pattern in DCR2 embryos is due to an abiotic environment, host background, or *Wolbachia* strain specificities.

Although this analysis remain mainly inconclusive regarding the causes of the observed variation due to many changing parameters and lower dataset quality, it gives insights on possible generalisation of the findings on the ISO1 dataset, and potential leads to investigate the variability of *Wolbachia* gene expression regulation. Studying gene expression of *Wolbachia* *w*MelCS in the ISO1 host background would help identify gene expression variation related to *Wolbachia* strain and avoid effects coming from host. Focusing on embryo might be interesting since it displays such a peculiar expression pattern in the samples from the DCR2 dataset. In another experiment, the gene expression of *Wolbachia* *w*Mel could be tested in the DCR2 host background to study whether the nos-gal4, UAS-*Dcr*-2 transgene could affect *Wolbachia* biology.

Chapter 6

Conclusion and perspectives

We started investigating the *Wolbachia-D. melanogaster* symbiosis by looking at the diversity of *Wolbachia* clades infecting *D. melanogaster* and their geographical distribution. In Chapter 3, we show a phylogenomic study of *Wolbachia* in *D. melanogaster* from various populations across North America, Africa, Europe, Asia and Australia. We find that for almost all samples, there is a correspondence between *Wolbachia* and mitochondrial clades, confirming a large prevalence of vertical transmission of *Wolbachia* in *D. melanogaster* and subsequent loss of the infection in some populations, as shown by previous studies (13; 14; 15; 16). However, for three samples, we find discrepancies between *Wolbachia* and mitochondrial clades, indicating possible horizontal transfer. It is the first time that horizontal transfer of *Wolbachia* in *D. melanogaster* has been reported in the wild. However, the limited number of clades in most populations might have made other horizontal transfer events more difficult to detect, and they might be more frequent than thought previously.

The geographical distribution of the clades suggests different adaptation capacities between them. The *Wolbachia* and mitochondrial clade structure, and the variation in infection rates of the different populations, suggest differences in competitiveness between the various *Wolbachia* clades, as well as differences in transmission efficiency.

Our findings confirm that *Wolbachia* transmission is almost exclusively vertical, from a single ancestral infection, making the bacteria and its host's evolution tightly tied. Moreover, the geographical distribution and relative abundance of the different *Wolbachia* clades suggest variations in the adaptations and efficiency

of the different clades. The result of the conjoint evolution and various adaptations of the *Wolbachia* probably left its mark on the bacterial genome. However, some of these adaptations might rely on expression regulation, rather than on gene mutations and duplication or deletions.

We continued the study of the *Wolbachia-D. melanogaster* symbiosis by looking at *Wolbachia* gene expression in *D. melanogaster* at different time points and tissues, in two different *Wolbachia* strains and host backgrounds. In Chapter 4, we looked for *Wolbachia* *w*Mel genes showing differential expression between the life stages of the ISO1 line of *D. melanogaster*, using the developmental transcriptomic data from modENCODE (45; 46; 47). Across the *D. melanogaster* life cycle, 80 genes were found to be differentially expressed, most of them following a similar pattern of expression with up-regulation after embryonic stages and stronger up-regulation in early larvae and late pupae. We also found 41 genes differentially expressed between adult males and females, most of them being identified as differentially expressed at five days old and most of them being up-regulated in males compared to females. 28 genes were found to be differentially expressed in both sets of genes, showing age and sex-biased expression. In total, 93 genes were found to have a stage- and/or sex-specific differential expression, and are candidate genes for understanding the basis of *Wolbachia* host manipulation.
In Chapter 5, we investigated *Wolbachia* gene expression in *D. melanogaster* in a second RNAseq dataset. We used an experiment by Czech *et al.* (132) containing expression data from *Wolbachia* *w*MelCS in the DCR2 line of *D. melanogaster*. We compared *Wolbachia* gene expression between the embryo, ovaries and female carcasses samples. We found 118 genes differentially expressed between embryo and female carcasses, a lot of them (46) being also differentially expressed between embryo and ovaries, and up-regulated in embryo. Another part of the genes differentially expressed between embryo and female carcasses is up-regulated in female carcasses and are, for most of them, also found differentially expressed between ovaries and carcasses in the same dataset, and between embryo and female adult in *Wolbachia* *w*Mel infecting the ISO1 line of *D. melanogaster*.
The RNAseq analysis done on two different strains of *Wolbachia*, *w*Mel and *w*MelCS, in different host backgrounds showed that most of the genes were expressed and that the expression levels of most of the expressed genes was relatively constant across the life cycle stages, sex and tissues tested. This is consistent with

a scenario of intimate coevolution, where the bacteria does not express strong virulence (3) and it does not trigger an immune reaction (82; 85).

The common gene expression pattern for genes differentially expressed over the life cycle shows a possible involvement in same function and co-regulation. The fact that some of those genes are also differentially expressed between embryo and adult of both *w*Mel and *w*MelCS in *D. melanogaster* ISO1 and DCR2 lines respectively indicates that some of the age-specific functions are conserved over strains and host contexts. Most of the genes showing a similar expression patterns in the two datasets also show differential expression between ovaries and female carcasses, with expression in ovaries similar to the embryo. This suggests that some of the adult-specific functions in females might not take place in ovaries, but in the somatic tissues, despite the high concentration of *Wolbachia* in ovaries. Most of the genes found to be differentially expressed in between males and females are up-regulated in males, and also differentially expressed across the host life cycle. This suggests that sex-biased *Wolbachia* functions are mainly taking place in males, and that similar mechanisms could be involved in stage and sex-biased functions.

The main difference we found between the two strains and backgrounds was the number of genes up-regulated in the embryo. Indeed, we found that a lot of genes were up-regulated in DCR2 embryonic samples compared to adult stages. Those genes were involved in translation, transcription, and chaperone functions, indicating a potential stress response. However, this might be a sign of higher replication observed in *w*MelCS, compared to *w*Mel. The variation in gene expression patterns between the two datasets indicates that *Wolbachia* in *D. melanogaster* kept the capacity to regulate its gene expression, and that some of the adaptations leading to the variety of fitness and transmission efficiency between *Wolbachia* clades might be based on gene expression regulation.

This work provides a set of candidate genes potentially involved in host-symbiont interactions. These findings can be leveraged to investigate further the mechanisms behind *Wolbachia* host-manipulations, and potentially be used in manipulating or selecting *Wolbachia* strains that are more adapted for purposes such as pest control or blocking insect-mediated virus transmission. Moreover, we found that besides adult stages, different time points across the development of *D. melanogaster* might be linked with *Wolbachia* activity, such as early larvae and late pupae for *Wolbachia* *w*Mel in ISO1, and embryo for *Wolbachia*

*w*MelCS in DCR2. Studying *Wolbachia* distribution and *Wolbachia* titers in *D. melanogaster* in those stages could shed light on the potential functions of those genes. To continue looking for the mechanisms of interaction between *Wolbachia* and *D. melanogaster*, the RNAseq dataset from modENCODE can still be used in many ways: after the study of *Wolbachia* and *D. melanogaster* gene expression, independently, conjoint gene expression from *D. melanogaster* and *Wolbachia* could be studied using the same dataset to find host and bacterial genes with a similar pattern of expression across fly development, possibly indicating some interaction. Moreover, as we saw in the Figure 4.2, we are able to detect yet unannotated small non-coding RNAs in *Wolbachia* transcriptome. The dataset could be therefor used to identify those small non coding RNAs and measure their expression variation, as they might be involved in gene expression regulation and are still not fully characterized to this day. The gene expression from the modENCODE dataset, thanks to the good sequencing depth of *Wolbachia* transcriptome, could also be used to characterize the operon structure in *Wolbachia*. This would help understanding co-regulation and functional links between *Wolbachia* genes. Since we couldn't distinguish between *Wolbachia* strain and host background effects when comparing the two RNAseq datasets, it would be interesting to investigate further the effects of both host background and *Wolbachia* strain, in a more controlled experiment. *Wolbachia* *w*MelCS gene expression could be studied after transinfection in an iso1 line of *D. melanogaster*. It would also be interesting maybe to study the gene expression of *Wolbachia* *w*Mel in the DCR2 host background. This would help to link more precisely variations in gene expression with phenotypic differences.

Previous studies showed that gene copy number variations were associated with phenotypic differences generated by *Wolbachia* (3; 103). The *D. melanogaster* resequencing data that we used in the phylogenomic analysis could be also used to investigate copy number variation in the different strains, adding complementary information to the RNA sequencing to identify genes implicated in host manipulation. Indeed, *Wolbachia* being an endosymbiont, it is difficult to manipulate and using natural diversity of *Wolbachia* strain is a good way to investigate the effect of *Wolbachia* genes. The large amount of available *Wolbachia* sequences make it possible to detect these structural variation and use this to link gene and phenotype in the *Wolbachia* and *D. melanogaster* symbiosis.

The many *D. melanogaster* resequencing experiments allowed us to study the variability of *Wolbachia* strains in wild populations of *D. melanogaster*, and we found for the first time a potential horizontal transfer of *Wolbachia* in *D. melanogaster* populations. In order to be able to confirm whether horizontal transfer happened in Ukraine, the populations could be sampled more heavily in order to see if the discrepancies we observe between *Wolbachia* and mitochondrial clades are confirmed. Although the resequencing data of *D. melanogaster* is very large, some areas remain poorly sampled. Sampling *D. melanogaster* populations from Central and South America, and Asia would help complete our understanding of *Wolbachia-D. melanogaster* coevolution. It would also be interesting to sample more deeply several populations, to confirm or correct the trends that we observed concerning the infection rate and clade proportions. Indeed, deeper sampling would allow to quantitatively verify the findings of our qualitative approach. Moreover, as some clades might subsist in low prevalence, such as *w*MelCS in Sub-Saharan African populations, a deeper sampling in those populations could help characterize better the clade diversity.

Bibliography

[1] K. T. Reynolds and A. A. Hoffmann, "Male age, host effects and the weak expression or non-expression of cytoplasmic incompatibility in Drosophila strains infected by maternally transmitted Wolbachia," *Genet. Res.*, vol. 80, pp. 79–87, Oct. 2002.

[2] U. Friberg, P. M. Miller, A. D. Stewart, and W. R. Rice, "Mechanisms Promoting the Long-Term Persistence of a Wolbachia Infection in a Laboratory-Adapted Population of Drosophila melanogaster," *PLoS ONE*, vol. 6, p. e16448, Jan. 2011.

[3] E. Chrostek, M. S. P. Marialva, S. S. Esteves, L. A. Weinert, J. Martinez, F. M. Jiggins, and L. Teixeira, "Wolbachia variants induce differential protection to viruses in Drosophila melanogaster: a phenotypic and phylogenomic analysis," *PLoS Genet.*, vol. 9, no. 12, p. e1003896, 2013.

[4] L. Teixeira, A. Ferreira, and M. Ashburner, "The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster," *PLoS Biol*, vol. 6, p. e1000002, Dec. 2008.

[5] T. F. C. Mackay, S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, M. F. Richardson, R. R. H. Anholt, M. Barrón, C. Bess, K. P. Blankenburg, M. A. Carbone, D. Castellano, L. Chaboub, L. Duncan, Z. Harris, M. Javaid, J. C. Jayaseelan, S. N. Jhangiani, K. W. Jordan, F. Lara, F. Lawrence, S. L. Lee, P. Librado, R. S. Linheiro, R. F. Lyman, A. J. Mackey, M. Munidasa, D. M. Muzny, L. Nazareth, I. Newsham, L. Perales, L.-L. Pu, C. Qu, M. Ràmia, J. G. Reid, S. M. Rollmann, J. Rozas, N. Saada, L. Turlapati, K. C. Worley,

Y.-Q. Wu, A. Yamamoto, Y. Zhu, C. M. Bergman, K. R. Thornton, D. Mittelman, and R. A. Gibbs, "The Drosophila melanogaster genetic reference panel," *Nature*, vol. 482, pp. 173–178, Feb. 2012.

[6] J. E. Pool, R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno, M. W. Crepeau, P. Duchen, J. J. Emerson, P. Saelao, D. J. Begun, and C. H. Langley, "Population Genomics of Sub-Saharan Drosophila melanogaster: African Diversity and Non-African Admixture," *PLoS Genet*, vol. 8, p. e1003080, Dec. 2012.

[7] J. K. Grenier, J. R. Arguello, M. C. Moreira, S. Gottipati, J. Mohammed, S. R. Hackett, R. Boughton, A. J. Greenberg, and A. G. Clark, "Global Diversity Lines–A Five-Continent Reference Panel of Sequenced Drosophila melanogaster Strains," *G3 (Bethesda)*, vol. 5, pp. 593–603, Feb. 2015.

[8] J. B. Lack, C. M. Cardeno, M. W. Crepeau, W. Taylor, R. B. Corbett-Detig, K. A. Stevens, C. H. Langley, and J. E. Pool, "The Drosophila Genome Nexus: A Population Genomic Resource of 623 Drosophila melanogaster Genomes, Including 197 from a Single Ancestral Range Population," *Genetics*, p. genetics.115.174664, Jan. 2015.

[9] E. G. King, C. M. Merkes, C. L. McNeil, S. R. Hoofer, S. Sen, K. W. Broman, A. D. Long, and S. J. Macdonald, "Genetic dissection of a model complex trait using the Drosophila Synthetic Population Resource," *Genome Res*, vol. 22, pp. 1558–1566, Aug. 2012.

[10] D. Campo, K. Lehmann, C. Fjeldsted, T. Souaiaia, J. Kao, and S. V. Nuzhdin, "Whole-genome sequencing of two North American Drosophila melanogaster populations reveals genetic differentiation and positive selection," *Mol Ecol*, vol. 22, pp. 5084–5097, Oct. 2013.

[11] E. Versace, V. Nolte, R. V. Pandey, R. Tobler, and C. Schlötterer, "Experimental evolution reveals habitat-specific fitness dynamics among Wolbachia clades in Drosophila melanogaster," *Mol Ecol*, vol. 23, pp. 802–814, Feb. 2014.

[12] C. M. Bergman and P. R. Haddrill, "Strain-specific and pooled genome sequences for populations of Drosophila melanogaster from three continents.," *F1000Research*, Jan. 2015.

[13] M. Solignac, D. Vautrin, and F. Rousset, "Widespread occurrence of the proteobacteria Wolbachia and partial cytoplasmic incompatibility in Drosophila melanogaster," *Comptes rendus des séances de l'Académie des sciences. Série III, Sciences de la vie*, vol. 317, pp. 461–470, Jan. 1994.

[14] M. F. Richardson, L. A. Weinert, J. J. Welch, R. S. Linheiro, M. M. Magwire, F. M. Jiggins, and C. M. Bergman, "Population genomics of the Wolbachia endosymbiont in Drosophila melanogaster," *PLoS Genet.*, vol. 8, p. e1003129, Dec. 2012.

[15] A. M. Early and A. G. Clark, "Monophyly of Wolbachia pipientis genomes within Drosophila melanogaster: geographic structuring, titre variation and host effects across five populations," *Mol Ecol*, vol. 22, pp. 5765–5778, Dec. 2013.

[16] Y. Ilinsky, "Coevolution of Drosophila melanogaster mtDNA and Wolbachia Genotypes," *PLoS ONE*, vol. 8, p. e54373, Jan. 2013.

[17] M. Riegler, M. Sidhu, W. J. Miller, and S. L. O'Neill, "Evidence for a global Wolbachia replacement in Drosophila melanogaster," *Curr Biol*, vol. 15, pp. 1428–33, Aug. 2005.

[18] M. Ashburner, K. G. Golic, and R. S. Hawley, *Drosophila: a laboratory handbook*. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press, 2005.

[19] T. N. Kristensen, A. K. Henningsen, C. Aastrup, M. Bech-Hansen, L. B. H. Bjerre, B. Carlsen, M. Hagstrup, S. G. Jensen, P. Karlsen, L. Kristensen, C. Lundsgaard, T. Møller, L. D. Nielsen, C. Starcke, C. R. Sørensen, and M. F. Schou, "Fitness components of Drosophila melanogaster developed on a standard laboratory diet or a typical natural food source," *Insect Science*, pp. n/a–n/a, July 2015.

[20] D. Hoang, A. Kopp, and J. A. Chandler, "Interactions between Drosophila and its natural yeast symbionts-Is Saccharomyces cerevisiae a good model for studying the fly-yeast relationship?," *PeerJ*, vol. 3, p. e1116, 2015.

[21] N. A. Broderick and B. Lemaitre, "Gut-associated microbes of Drosophila melanogaster," *Gut Microbes*, vol. 3, pp. 307–321, Aug. 2012.

[22] B. Erkosar and F. Leulier, "Transient adult microbiota, gut homeostasis and longevity: Novel insights from the Drosophila model," *FEBS Letters*, vol. 588, pp. 4250–4257, Nov. 2014.

[23] N. Buchon, N. A. Broderick, and B. Lemaitre, "Gut homeostasis in a microbial world: insights from Drosophila melanogaster," *Nat Rev Micro*, vol. 11, pp. 615–626, Sept. 2013.

[24] B. Erkosar, G. Storelli, A. Defaye, and F. Leulier, "Host-Intestinal Microbiota Mutualism: "Learning on the Fly"," *Cell Host & Microbe*, vol. 13, pp. 8–14, Jan. 2013.

[25] W.-J. Lee and P. T. Brey, "How Microbiomes Influence Metazoan Development: Insights from History and Drosophila Modeling of Gut-Microbe Interactions," *Annual Review of Cell and Developmental Biology*, vol. 29, no. 1, p. null, 2013.

[26] C. N. A. Wong, P. Ng, and A. E. Douglas, "Low-diversity bacterial community in the gut of the fruitfly Drosophila melanogaster," *Environ. Microbiol.*, vol. 13, pp. 1889–1900, July 2011.

[27] G. Caracristi and C. Schlötterer, "Genetic differentiation between American and European Drosophila melanogaster populations could be attributed to admixture of African alleles.," *Molecular biology and evolution*, vol. 20, pp. 792–799, May 2003.

[28] D. Lachaise, M. Cariou, J. David, F. Lemeunier, L. Tsacas, and M. Ashburner, "Historical biogeography of the Drosophila melanogaster species subgroup.," *Evolutionary Biology*, vol. 22, pp. 159–225, 1988.

[29] W. Stephan and H. Li, "The recent demographic and adaptive history of Drosophila melanogaster," *Heredity*, vol. 98, pp. 65–68, Sept. 2006.

[30] P. Duchen, D. Živković, S. Hutter, W. Stephan, and S. Laurent, "Demographic Inference Reveals African and European Admixture in the North American Drosophila melanogaster Population," *Genetics*, vol. 193, pp. 291–301, Jan. 2013.

[31] D. J. Begun and C. F. Aquadro, "African and North American populations of Drosophila melanogaster are very different at the DNA level.," *Nature*, vol. 365, pp. 548–550, Oct. 1993.

[32] M. D. S. Nunes, H. Neumeier, and C. Schlotterer, "Contrasting patterns of natural variation in global Drosophila melanogaster populations," *Mol Ecol*, vol. 17, pp. 4470–9, Oct. 2008.

[33] J. E. Pool and C. F. Aquadro, "History and structure of sub-Saharan populations of Drosophila melanogaster," *Genetics*, vol. 174, pp. 915–29, Oct. 2006.

[34] S. Glinka, L. Ometto, S. Mousset, W. Stephan, and D. De Lorenzo, "Demography and Natural Selection Have Shaped Genetic Variation in Drosophila melanogaster: A Multi-locus Approach," *Genetics*, vol. 165, pp. 1269–1278, Nov. 2003.

[35] H. Li and W. Stephan, "Inferring the demographic history and rate of adaptive substitution in Drosophila," *PLoS Genet*, vol. 2, p. e166, Oct. 2006.

[36] M. Ashburner and C. M. Bergman, "Drosophila melanogaster: A case study of a model genomic sequence and its consequences," *Genome Research*, vol. 15, pp. 1661–1667, Dec. 2005.

[37] M. D. Adams, S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne, P. G. Amanatides, S. E. Scherer, P. W. Li, R. A. Hoskins, R. F. Galle, R. A. George, S. E. Lewis, S. Richards, M. Ashburner, S. N. Henderson, G. G. Sutton, J. R. Wortman, M. D. Yandell, Q. Zhang, L. X. Chen, R. C. Brandon, Y.-H. C. Rogers, R. G. Blazej, M. Champe, B. D. Pfeiffer, K. H. Wan, C. Doyle, E. G. Baxter, G. Helt, C. R. Nelson, G. L. Gabor, Miklos, J. F. Abril, A. Agbayani, H.-J. An, C. Andrews-Pfannkoch, D. Baldwin, R. M. Ballew, A. Basu, J. Baxendale, L. Bayraktaroglu, E. M. Beasley, K. Y. Beeson, P. V. Benos, B. P. Berman, D. Bhandari, S. Bolshakov, D. Borkova, M. R. Botchan, J. Bouck, P. Brokstein, P. Brottier, K. C. Burtis, D. A. Busam, H. Butler, E. Cadieu, A. Center, I. Chandra, J. M. Cherry, S. Cawley, C. Dahlke, L. B. Davenport, P. Davies, B. d. Pablos, A. Delcher, Z. Deng, A. D. Mays, I. Dew, S. M. Dietz, K. Dodson,

L. E. Doup, M. Downes, S. Dugan-Rocha, B. C. Dunkov, P. Dunn, K. J. Durbin, C. C. Evangelista, C. Ferraz, S. Ferriera, W. Fleischmann, C. Fosler, A. E. Gabrielian, N. S. Garg, W. M. Gelbart, K. Glasser, A. Glodek, F. Gong, J. H. Gorrell, Z. Gu, P. Guan, M. Harris, N. L. Harris, D. Harvey, T. J. Heiman, J. R. Hernandez, J. Houck, D. Hostin, K. A. Houston, T. J. Howland, M.-H. Wei, C. Ibegwam, M. Jalali, F. Kalush, G. H. Karpen, Z. Ke, J. A. Kennison, K. A. Ketchum, B. E. Kimmel, C. D. Kodira, C. Kraft, S. Kravitz, D. Kulp, Z. Lai, P. Lasko, Y. Lei, A. A. Levitsky, J. Li, Z. Li, Y. Liang, X. Lin, X. Liu, B. Mattei, T. C. McIntosh, M. P. McLeod, D. McPherson, G. Merkulov, N. V. Milshina, C. Mobarry, J. Morris, A. Moshrefi, S. M. Mount, M. Moy, B. Murphy, L. Murphy, D. M. Muzny, D. L. Nelson, D. R. Nelson, K. A. Nelson, K. Nixon, D. R. Nusskern, J. M. Pacleb, M. Palazzolo, G. S. Pittman, S. Pan, J. Pollard, V. Puri, M. G. Reese, K. Reinert, K. Remington, R. D. C. Saunders, F. Scheeler, H. Shen, B. C. Shue, I. Sidén-Kiamos, M. Simpson, M. P. Skupski, T. Smith, E. Spier, A. C. Spradling, M. Stapleton, R. Strong, E. Sun, R. Svirskas, C. Tector, R. Turner, E. Venter, A. H. Wang, X. Wang, Z.-Y. Wang, D. A. Wassarman, G. M. Weinstock, J. Weissenbach, S. M. Williams, T. Woodage, K. C. Worley, D. Wu, S. Yang, Q. A. Yao, J. Ye, R.-F. Yeh, J. S. Zaveri, M. Zhan, G. Zhang, Q. Zhao, L. Zheng, X. H. Zheng, F. N. Zhong, W. Zhong, X. Zhou, S. Zhu, X. Zhu, H. O. Smith, R. A. Gibbs, E. W. Myers, G. M. Rubin, and J. C. Venter, "The Genome Sequence of Drosophila melanogaster," *Science*, vol. 287, pp. 2185–2195, Mar. 2000.

[38] A. G. Clark, M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, T. A. Markow, T. C. Kaufman, M. Kellis, W. Gelbart, V. N. Iyer, D. A. Pollard, T. B. Sackton, A. M. Larracuente, N. D. Singh, J. P. Abad, D. N. Abt, B. Adryan, M. Aguade, H. Akashi, W. W. Anderson, C. F. Aquadro, D. H. Ardell, R. Arguello, C. G. Artieri, D. A. Barbash, D. Barker, P. Barsanti, P. Batterham, S. Batzoglou, D. Begun, A. Bhutkar, E. Blanco, S. A. Bosak, R. K. Bradley, A. D. Brand, M. R. Brent, A. N. Brooks, R. H. Brown, R. K. Butlin, C. Caggese, B. R. Calvi, A. Bernardo de Carvalho, A. Caspi, S. Castrezana, S. E. Celniker, J. L. Chang, C. Chapple, S. Chatterji, A. Chinwalla, A. Civetta, S. W. Clifton, J. M. Comeron, J. C. Costello, J. A. Coyne, J. Daub, R. G. David, A. L. Delcher, K. Delehaunty, C. B. Do, H. Ebling,

K. Edwards, T. Eickbush, J. D. Evans, A. Filipski, S. Findeiss, E. Freyhult, L. Fulton, R. Fulton, A. C. Garcia, A. Gardiner, D. A. Garfield, B. E. Garvin, G. Gibson, D. Gilbert, S. Gnerre, J. Godfrey, R. Good, V. Gotea, B. Gravely, A. J. Greenberg, S. Griffiths-Jones, S. Gross, R. Guigo, E. A. Gustafson, W. Haerty, M. W. Hahn, D. L. Halligan, A. L. Halpern, G. M. Halter, M. V. Han, A. Heger, L. Hillier, A. S. Hinrichs, I. Holmes, R. A. Hoskins, M. J. Hubisz, D. Hultmark, M. A. Huntley, D. B. Jaffe, S. Jagadeeshan, *et al.*, "Evolution of genes and genomes on the Drosophila phylogeny," *Nature*, vol. 450, no. 7167, pp. 203–218, 2007. 1476-4687 (Electronic) Journal article.

[39] R. Garesse, "Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations," *Genetics*, vol. 118, pp. 649–63, Apr. 1988.

[40] D. L. Lewis, C. L. Farr, A. L. Farquhar, and L. S. Kaguni, "Sequence, organization, and evolution of the A+T region of Drosophila melanogaster mitochondrial DNA.," *Molecular biology and evolution*, vol. 11, pp. 523–538, May 1994.

[41] D. M. Rand, "Population genetics of the cytoplasm and the units of selection on mitochondrial DNA in Drosophila melanogaster.," *Genetica*, May 2011.

[42] G. D. D. Hurst and F. M. Jiggins, "Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts," *Proc Biol Sci*, vol. 272, no. 1572, pp. 1525–34, 2005. Hurst, Gregory D D Jiggins, Francis M 070535/Wellcome Trust/United Kingdom Wellcome Trust/United Kingdom Research Support, Non-U.S. Gov't Review England Proceedings. Biological sciences / The Royal Society Proc Biol Sci. 2005 Aug 7;272(1572):1525-34.

[43] M. D. S. Nunes, V. Nolte, and C. Schlötterer, "Nonrandom Wolbachia infection status of Drosophila melanogaster strains with different mtDNA haplotypes," *Molecular Biology and Evolution*, vol. 25, pp. 2493–8, Nov. 2008.

[44] S. E. Celniker, L. A. L. Dillon, M. B. Gerstein, K. C. Gunsalus, S. Henikoff,

G. H. Karpen, M. Kellis, E. C. Lai, J. D. Lieb, D. M. MacAlpine, G. Micklem, F. Piano, M. P. Snyder, L. Stein, K. P. White, and R. H. Waterston, "Unlocking the secrets of the genome.," *Nature*, vol. 459, pp. 927–30, June 2009.

[45] B. R. Graveley, A. N. Brooks, J. W. Carlson, M. O. Duff, J. M. Landolin, L. Yang, C. G. Artieri, M. J. van Baren, N. Boley, B. W. Booth, J. B. Brown, L. Cherbas, C. A. Davis, A. Dobin, R. Li, W. Lin, J. H. Malone, N. R. Mattiuzzo, D. Miller, D. Sturgill, B. B. Tuch, C. Zaleski, D. Zhang, M. Blanchette, S. Dudoit, B. Eads, R. E. Green, A. Hammonds, L. Jiang, P. Kapranov, L. Langton, N. Perrimon, J. E. Sandler, K. H. Wan, A. Willingham, Y. Zhang, Y. Zou, J. Andrews, P. J. Bickel, S. E. Brenner, M. R. Brent, P. Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, B. Oliver, and S. E. Celniker, "The developmental transcriptome of Drosophila melanogaster," *Nature*, vol. 471, pp. 473–479, Mar. 2011.

[46] J. B. Brown, N. Boley, R. Eisman, G. E. May, M. H. Stoiber, M. O. Duff, B. W. Booth, J. Wen, S. Park, A. M. Suzuki, K. H. Wan, C. Yu, D. Zhang, J. W. Carlson, L. Cherbas, B. D. Eads, D. Miller, K. Mockaitis, J. Roberts, C. A. Davis, E. Frise, A. S. Hammonds, S. Olson, S. Shenker, D. Sturgill, A. A. Samsonova, R. Weiszmann, G. Robinson, J. Hernandez, J. Andrews, P. J. Bickel, P. Carninci, P. Cherbas, T. R. Gingeras, R. A. Hoskins, T. C. Kaufman, E. C. Lai, B. Oliver, N. Perrimon, B. R. Graveley, and S. E. Celniker, "Diversity and dynamics of the Drosophila transcriptome," *Nature*, vol. 512, pp. 393–399, Aug. 2014.

[47] M. O. Duff, S. Olson, X. Wei, S. C. Garrett, A. Osman, M. Bolisetty, A. Plocik, S. E. Celniker, and B. R. Graveley, "Genome-wide identification of zero nucleotide recursive splicing in Drosophila," *Nature*, vol. 521, pp. 376–379, May 2015.

[48] J. H. Werren, "Biology of Wolbachia," *Annual Review of Entomology*, vol. 42, no. 1, pp. 587–609, 1997.

[49] M. Hertig, "The Rickettsia, Wolbachia pipientis (gen. et sp.n.) and Associated Inclusions of the Mosquito, Culex pipiens," *Parasitology*, vol. 28, pp. 453–486, Oct. 1936.

[50] M. Gerth, M.-T. Gansauge, A. Weigert, and C. Bleidorn, "Phylogenomic analyses uncover origin and spread of the Wolbachia pandemic," *Nat Commun*, vol. 5, p. 5117, Oct. 2014.

[51] M. J. Taylor, D. Voronin, K. L. Johnston, and L. Ford, "Wolbachia filarial interactions," *Cell Microbiol*, vol. 15, pp. 520–526, Apr. 2013.

[52] A. C. Darby, S. D. Armstrong, G. S. Bah, G. Kaur, M. A. Hughes, S. M. Kay, P. Koldkjær, L. Rainbow, A. D. Radford, M. L. Blaxter, V. N. Tanya, A. J. Trees, R. Cordaux, J. M. Wastling, and B. L. Makepeace, "Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis," *Genome Res.*, vol. 22, pp. 2467–2477, Dec. 2012.

[53] F. Tamarozzi, A. Halliday, K. Gentil, A. Hoerauf, E. Pearlman, and M. J. Taylor, "Onchocerciasis: the Role of Wolbachia Bacterial Endosymbionts in Parasite Biology, Disease Pathogenesis, and Treatment," *Clin Microbiol Rev*, vol. 24, pp. 459–468, July 2011.

[54] M. Walker, S. Specht, T. S. Churcher, A. Hoerauf, M. J. Taylor, and M.-G. Basáñez, "Therapeutic Efficacy and Macrofilaricidal Activity of Doxycycline for the Treatment of River Blindness," *Clin Infect Dis*, vol. 60, pp. 1199–1207, Apr. 2015.

[55] J. Foster, M. Ganatra, I. Kamal, J. Ware, K. Makarova, N. Ivanova, A. Bhattacharyya, V. Kapatral, S. Kumar, J. Posfai, T. Vincze, J. Ingram, L. Moran, A. Lapidus, M. Omelchenko, N. Kyrpides, E. Ghedin, S. Wang, E. Goltsman, V. Joukov, O. Ostrovskaya, K. Tsukerman, M. Mazur, D. Comb, E. Koonin, and B. Slatko, "The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode," *PLoS Biol.*, vol. 3, p. e121, Apr. 2005.

[56] T. Hosokawa, R. Koga, Y. Kikuchi, X.-Y. Meng, and T. Fukatsu, "Wolbachia as a bacteriocyte-associated nutritional mutualist," *PNAS*, vol. 107, pp. 769–774, Jan. 2010.

[57] Y. I. Coulibaly, B. Dembele, A. A. Diallo, E. M. Lipner, S. S. Doumbia, S. Y. Coulibaly, S. Konate, D. A. Diallo, D. Yalcouye, J. Kubofcik, O. K.

Doumbo, A. K. Traore, A. D. Keita, M. P. Fay, S. F. Traore, T. B. Nutman, and A. D. Klion, "A Randomized Trial of Doxycycline for Mansonella perstans Infection," *New England Journal of Medicine*, vol. 361, pp. 1448–1458, Oct. 2009.

[58] M. Zabal-Aguirre, F. Arroyo, J. García-Hurtado, J. de la Torre, G. M. Hewitt, and J. L. Bella, "Wolbachia effects in natural populations of Chorthippus parallelus from the Pyrenean hybrid zone," *J. Evol. Biol.*, vol. 27, pp. 1136–1148, June 2014.

[59] T. T. Vandekerckhove, S. Watteyne, A. Willems, J. G. Swings, J. Mertens, and M. Gillis, "Phylogenetic analysis of the 16s rDNA of the cytoplasmic bacterium Wolbachia from the novel host Folsomia candida (Hexapoda, Collembola) and its implications for wolbachial taxonomy1," *FEMS Microbiology Letters*, vol. 180, pp. 279–286, Nov. 1999.

[60] S. Bordenstein and R. B. Rosengaus, "Discovery of a Novel Wolbachia Supergroup in Isoptera," *Curr Microbiol*, vol. 51, pp. 393–398, Oct. 2005.

[61] F. Dedeine, F. Vavre, F. Fleury, B. Loppin, M. E. Hochberg, and M. Boulétreau, "Removing symbiotic Wolbachia bacteria specifically inhibits oogenesis in a parasitic wasp," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 98, pp. 6247–6252, May 2001.

[62] J. C. Brownlie, B. N. Cass, M. Riegler, J. J. Witsenburg, I. Iturbe-Ormaetxe, E. A. McGraw, and S. L. O'Neill, "Evidence for metabolic provisioning by a common invertebrate endosymbiont, Wolbachia pipientis, during periods of nutritional stress," *PLoS Pathog*, vol. 5, p. e1000368, Apr. 2009.

[63] L. M. Hedges, J. C. Brownlie, S. L. O'Neill, and K. N. Johnson, "Wolbachia and virus protection in insects," *Science*, vol. 322, no. 5902, p. 702, 2008.

[64] K. T. Min and S. Benzer, "Wolbachia, normally a symbiont of Drosophila, can be virulent, causing degeneration and early death," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 94, pp. 10792–10796, Sept. 1997.

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

[66] J. H. Yen and A. R. Barr, "New hypothesis of the cause of cytoplasmic incompatibility in Culex pipiens L," *Nature*, vol. 232, no. 5313, pp. 657–8, 1971. Yen, J H Barr, A R England Nature Nature. 1971 Aug 27;232(5313):657-8.

[67] H. Laven, "Eradication of Culex pipiens fatigans through Cytoplasmic Incompatibility," *Nature*, vol. 216, pp. 383–384, Oct. 1967.

[68] S. Zabalou, M. Riegler, M. Theodorakopoulou, C. Stauffer, C. Savakis, and K. Bourtzis, "Wolbachia-induced cytoplasmic incompatibility as a means for insect pest population control," *PNAS*, vol. 101, pp. 15042–15045, Oct. 2004.

[69] K. Hilgenboecker, P. Hammerstein, P. Schlattmann, A. Telschow, and J. H. Werren, "How many species are infected with Wolbachia?–A statistical analysis of current data," *FEMS Microbiol Lett*, vol. 281, no. 2, pp. 215–20, 2008. Hilgenboecker, Kirsten Hammerstein, Peter Schlattmann, Peter Telschow, Arndt Werren, John H Meta-Analysis Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. England FEMS microbiology letters FEMS Microbiol Lett. 2008 Apr;281(2):215-20. Epub 2008 Feb 28.

[70] R. Zug and P. Hammerstein, "Still a Host of Hosts for Wolbachia: Analysis of Recent Data Suggests That 40% of Terrestrial Arthropod Species Are Infected," *PLoS ONE*, vol. 7, p. e38544, June 2012.

[71] F. Gutzwiller, F. Dedeine, W. Kaiser, D. Giron, and C. Lopez-Vaamonde, "Correlation between the green-island phenotype and Wolbachia infections during the evolutionary diversification of Gracillariidae leaf-mining moths," *Ecol Evol*, vol. 5, pp. 4049–4062, Sept. 2015.

[72] C. E. Rohrscheib, E. Bondy, P. Josh, M. Riegler, D. Eyles, B. v. Swinderen, M. W. Weible, and J. C. Brownlie, "Wolbachia influences the production of octopamine and affects Drosophila male aggression," *Appl. Environ. Microbiol.*, pp. AEM.00573–15, May 2015.

[73] L. Lambrechts, N. M. Ferguson, E. Harris, E. C. Holmes, E. A. McGraw, S. L. O'Neill, E. E. Ooi, S. A. Ritchie, P. A. Ryan, T. W. Scott, C. P. Simmons, and S. C. Weaver, "Assessing the epidemiological effect of wolbachia

for dengue control," *The Lancet Infectious Diseases*, vol. 15, pp. 862–866, July 2015.

[74] L. A. Moreira, I. Iturbe-Ormaetxe, J. A. Jeffery, G. Lu, A. T. Pyke, L. M. Hedges, B. C. Rocha, S. Hall-Mendelin, A. Day, M. Riegler, L. E. Hugo, K. N. Johnson, B. H. Kay, E. A. McGraw, A. F. van den Hurk, P. A. Ryan, and S. L. O'Neill, "A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium," *Cell*, vol. 139, pp. 1268–1278, Dec. 2009.

[75] G. Bian, Y. Xu, P. Lu, Y. Xie, and Z. Xi, "The endosymbiotic bacterium Wolbachia induces resistance to dengue virus in Aedes aegypti," *PLoS Pathog.*, vol. 6, p. e1000833, Apr. 2010.

[76] A. A. Hoffmann, B. L. Montgomery, J. Popovici, I. Iturbe-Ormaetxe, P. H. Johnson, F. Muzzi, M. Greenfield, M. Durkan, Y. S. Leong, Y. Dong, H. Cook, J. Axford, A. G. Callahan, N. Kenny, C. Omodei, E. A. McGraw, P. A. Ryan, S. A. Ritchie, M. Turelli, and S. L. O'Neill, "Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission," *Nature*, vol. 476, no. 7361, pp. 454–7, 2011. Hoffmann, A A Montgomery, B L Popovici, J Iturbe-Ormaetxe, I Johnson, P H Muzzi, F Greenfield, M Durkan, M Leong, Y S Dong, Y Cook, H Axford, J Callahan, A G Kenny, N Omodei, C McGraw, E A Ryan, P A Ritchie, S A Turelli, M O'Neill, S L Evaluation Studies Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. England Nature Nature. 2011 Aug 24;476(7361):454-7. doi: 10.1038/nature10356.

[77] F. Landmann, G. A. Orsi, B. Loppin, and W. Sullivan, "Wolbachia-Mediated Cytoplasmic Incompatibility Is Associated with Impaired Histone Deposition in the Male Pronucleus," *PLoS Pathog*, vol. 5, p. e1000343, Mar. 2009.

[78] U. Tram and W. Sullivan, "Role of Delayed Nuclear Envelope Breakdown and Mitosis in Wolbachia-Induced Cytoplasmic Incompatibility," *Science*, vol. 296, pp. 1124–1126, May 2002.

[79] U. Tram, K. Fredrick, J. H. Werren, and W. Sullivan, "Paternal chromosome segregation during the first mitotic division determines Wolbachia-induced cytoplasmic incompatibility phenotype," *J Cell Sci*, vol. 119, pp. 3655–3663, Sept. 2006.

[80] S. Zabalou, A. Apostolaki, S. Pattas, Z. Veneti, C. Paraskevopoulos, I. Livadaras, G. Markakis, T. Brissac, H. Merçot, and K. Bourtzis, "Multiple Rescue Factors Within a Wolbachia Strain," *Genetics*, vol. 178, pp. 2145–2160, Apr. 2008.

[81] K. Koukou, H. Pavlikaki, G. Kilias, J. H. Werren, K. Bourtzis, and S. N. Alahiotis, "Influence of Antibiotic Treatment and Wolbachia Curing on Sexual Isolation Among Drosophila Melanogaster Cage Populations," *Evolution*, vol. 60, pp. 87–96, Jan. 2006.

[82] R. Zug and P. Hammerstein, "Wolbachia and the insect immune system: what reactive oxygen species can tell us about the mechanisms of Wolbachia–host interactions," *Front Microbiol*, vol. 6, Oct. 2015.

[83] L. Wang, C. Zhou, Z. He, Z.-G. G. Wang, J.-L. L. Wang, and Y.-F. F. Wang, "Wolbachia infection decreased the resistance of Drosophila to lead.," *PloS one*, vol. 7, no. 3, 2012.

[84] M. E. Garlapow, W. Huang, M. T. Yarboro, K. R. Peterson, and T. F. C. Mackay, "Quantitative Genetics of Food Intake in Drosophila melanogaster," *PLOS ONE*, vol. 10, p. e0138129, Sept. 2015.

[85] E. Chrostek, M. S. P. Marialva, R. Yamada, S. L. O'Neill, and L. Teixeira, "High Anti-Viral Protection without Immune Upregulation after Interspecies Wolbachia Transfer," *PLOS ONE*, vol. 9, p. e99025, June 2014.

[86] K. Bourtzis, M. M. Pettigrew, and S. L. O'Neill, "Wolbachia neither induces nor suppresses transcripts encoding antimicrobial peptides," *Insect Mol. Biol.*, vol. 9, pp. 635–639, Dec. 2000.

[87] P. E. Cook and E. A. McGraw, "Wolbachia pipientis: an expanding bag of tricks to explore for disease control," *Trends Parasitol.*, vol. 26, pp. 373–375, Aug. 2010.

[88] R. J. Albertson, V. Tan, R. R. Leads, M. Reyes, W. Sullivan, and C. Casper-Lindley, "Mapping Wolbachia distributions in the adult Drosophila brain," *Cell. Microbiol.*, Mar. 2013.

[89] P. F. Vale and M. D. Jardine, "Sex-specific behavioural symptoms of viral gut infection and Wolbachia in Drosophila melanogaster," *J. Insect Physiol.*, Aug. 2015.

[90] Z. Veneti, M. E. Clark, T. L. Karr, C. Savakis, and K. Bourtzis, "Heads or Tails: Host-Parasite Interactions in the Drosophila-Wolbachia System," *Appl. Environ. Microbiol.*, vol. 70, pp. 5366–5372, Sept. 2004.

[91] L. R. Serbus and W. Sullivan, "A Cellular Basis for Wolbachia Recruitment to the Host Germline," *PLoS Pathog*, vol. 3, p. e190, Dec. 2007.

[92] G. Callaini, M. G. Riparbelli, and R. Dallai, "The distribution of cytoplasmic bacteria in the early Drosophila embryo is mediated by astral microtubules," *J Cell Sci*, vol. 107, pp. 673–682, Mar. 1994.

[93] H. Kose and T. L. Karr, "Organization of Wolbachia pipientis in the Drosophila fertilized egg and embryo revealed by an anti-Wolbachia monoclonal antibody," *Mech. Dev.*, vol. 51, pp. 275–288, June 1995.

[94] B. Pintureau, S. Grenier, B. Boléat, F. Lassablière, A. Heddi, and C. Khatchadourian, "Dynamics of Wolbachia Populations in Transfected Lines of Trichogramma," *Journal of Invertebrate Pathology*, vol. 76, pp. 20–25, July 2000.

[95] H. M. Frydman, J. M. Li, D. N. Robson, and E. Wieschaus, "Somatic stem cell niche tropism in Wolbachia," *Nature*, vol. 441, pp. 509–512, May 2006.

[96] S. L. Dobson, K. Bourtzis, H. R. Braig, B. F. Jones, W. Zhou, F. Rousset, and S. L. O'Neill, "Wolbachia infections are distributed throughout insect somatic and germ line tissues," *Insect Biochemistry and Molecular Biology*, vol. 29, pp. 153–160, Feb. 1999.

[97] R. Albertson, C. Casper-Lindley, J. Cao, U. Tram, and W. Sullivan, "Symmetric and asymmetric mitotic segregation patterns influence Wolbachia distribution in host somatic tissue," *J Cell Sci*, vol. 122, pp. 4570–4583, Dec. 2009.

[98] M. E. Clark, Z. Veneti, K. Bourtzis, and T. L. Karr, "The distribution and proliferation of the intracellular bacteria Wolbachia during spermatogenesis in Drosophila," *Mech. Dev.*, vol. 111, pp. 3–15, Feb. 2002.

[99] M. Wu, L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, J. F. Kolonay, W. C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M. B. Young, T. Utterback, J. Weidman, W. C. Nierman, I. T. Paulsen, K. E. Nelson, H. Tettelin, S. L. O'Neill, and J. A. Eisen, "Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: a streamlined genome overrun by mobile genetic elements," *PLoS Biol.*, vol. 2, p. E69, Mar. 2004.

[100] S. Shigenobu, H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa, "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS," *Nature*, vol. 407, pp. 81–86, Sept. 2000.

[101] L. Akman, A. Yamashita, H. Watanabe, K. Oshima, T. Shiba, M. Hattori, and S. Aksoy, "Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia," *Nat Genet*, vol. 32, pp. 402–407, Nov. 2002.

[102] I. I. Goryacheva and B. V. Andrianov, "Biological effects of Wolbachia pipientis: Elucidation of genetic mechanisms," *Biol Bull Rev*, vol. 5, pp. 109–118, Apr. 2015.

[103] E. Chrostek and L. Teixeira, "Mutualism breakdown by amplification of Wolbachia genes," *PLoS Biol.*, vol. 13, p. e1002065, Feb. 2015.

[104] M. Woolfit, I. Iturbe-Ormaetxe, J. C. Brownlie, T. Walker, M. Riegler, A. Seleznev, J. Popovici, E. Rancès, B. A. Wee, J. Pavlides, M. J. Sullivan, S. A. Beatson, A. Lane, M. Sidhu, C. J. McMeniman, E. A. McGraw, and S. L. O'Neill, "Genomic evolution of the pathogenic Wolbachia strain, wMelPop," *Genome Biol. Evol.*, vol. 5, no. 11, pp. 2189–2204, 2013.

[105] I. Iturbe-Ormaetxe, G. R. Burke, M. Riegler, and S. L. O'Neill, "Distribution, expression, and motif variability of ankyrin domain genes in Wolbachia pipientis," *J. Bacteriol.*, vol. 187, pp. 5136–5145, Aug. 2005.

[106] L. Klasson, T. Walker, M. Sebaihia, M. J. Sanders, M. A. Quail, A. Lord, S. Sanders, J. Earl, S. L. O'Neill, N. Thomson, S. P. Sinkins, and J. Parkhill, "Genome evolution of Wolbachia strain wPip from the Culex pipiens group," *Mol. Biol. Evol.*, vol. 25, pp. 1877–87, Sept. 2008.

[107] L. Klasson, J. Westberg, P. Sapountzis, K. Näslund, Y. Lutnaes, A. C. Darby, Z. Veneti, L. Chen, H. R. Braig, R. Garrett, K. Bourtzis, and S. G. Andersson, "The mosaic genome structure of the Wolbachia wRi strain infecting Drosophila simulans.," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, pp. 5725–5730, Apr. 2009.

[108] S. G. Sedgwick and S. J. Smerdon, "The ankyrin repeat: a diversity of interactions on a common structural framework," *Trends in Biochemical Sciences*, vol. 24, pp. 311–316, Aug. 1999.

[109] X. Pan, A. Lührmann, A. Satoh, M. A. Laskowski-Arce, and C. R. Roy, "Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors," *Science*, vol. 320, pp. 1651–1654, June 2008.

[110] S. Al-Khodor, C. T. Price, A. Kalia, and Y. A. Kwaik, "Ankyrin-repeat containing proteins of microbes: a conserved structure with functional diversity," *Trends Microbiol*, vol. 18, pp. 132–139, Mar. 2010.

[111] T. Walker, L. Klasson, M. Sebaihia, M. J. Sanders, N. R. Thomson, J. Parkhill, and S. P. Sinkins, "Ankyrin repeat domain-encoding genes in the wPip strain of Wolbachia from the Culex pipiens group," *BMC Biology*, vol. 5, p. 39, Sept. 2007.

[112] R. Yamada, I. Iturbe-Ormaetxe, J. C. Brownlie, and S. L. O'Neill, "Functional test of the influence of Wolbachia genes on cytoplasmic incompatibility expression in Drosophila melanogaster," *Insect Mol. Biol.*, vol. 20, pp. 75–85, Feb. 2011.

[113] O. Duron, A. Boureux, P. Echaubard, A. Berthomieu, C. Berticat, P. Fort, and M. Weill, "Variability and expression of ankyrin domain genes in Wolbachia variants infecting the mosquito Culex pipiens," *J. Bacteriol.*, vol. 189, pp. 4442–4448, June 2007.

[114] S. P. Sinkins, T. Walker, A. R. Lynd, A. R. Steven, B. L. Makepeace, H. C. J. Godfray, and J. Parkhill, "Wolbachia variability and host effects on crossing type in Culex mosquitoes," *Nature*, vol. 436, pp. 257–260, July 2005.

[115] G. Papafotiou, S. Oehler, C. Savakis, and K. Bourtzis, "Regulation of Wolbachia ankyrin domain encoding genes in Drosophila gonads," *Res. Microbiol.*, vol. 162, pp. 764–772, Oct. 2011.

[116] A. K. Hansen and P. H. Degnan, "Widespread expression of conserved small RNAs in small symbiont genomes," *ISME J*, vol. 8, pp. 2490–2502, Dec. 2014.

[117] J. L. Wilcox, H. E. Dunbar, R. D. Wolfinger, and N. A. Moran, "Consequences of reductive evolution for gene expression in an obligate endosymbiont," *Molecular Microbiology*, vol. 48, pp. 1491–1500, June 2003.

[118] S. Christensen and L. R. Serbus, "Comparative Analysis of Wolbachia Genomes Reveals Streamlining and Divergence of Minimalist Two-Component Systems," *G3 (Bethesda)*, Mar. 2015.

[119] A. A. Hoffmann, M. Hercus, and H. Dagher, "Population dynamics of the Wolbachia infection causing cytoplasmic incompatibility in Drosophila melanogaster.," *Genetics*, vol. 148, pp. 221–231, Jan. 1998.

[120] Y. Y. Ilinsky and I. K. Zakharov, "Infection of the Uman' population of Drosophila melanogaster with the cytoplasmic endosymbiont Wolbachia," *Dokl Biol Sci*, vol. 413, pp. 166–168, Apr. 2007.

[121] Y. Ilinsky and I. Zakharov, "The endosymbiont Wolbachia in Eurasian populations of Drosophila melanogaster," *Genetika*, vol. 43, no. 7, pp. 905–15, 2007. Ilinskii, Iu Iu Zakharov, I K English Abstract Research Support, Non-U.S. Gov't Russia (Federation) Genetika Genetika. 2007 Jul;43(7):905-15.

[122] R. L. Verspoor and P. R. Haddrill, "Genetic Diversity, Population Structure and Wolbachia Infection Status in a Worldwide Sample of Drosophila melanogaster and D. simulans Populations," *PLoS ONE*, vol. 6, p. e26318, Oct. 2011.

[123] R. A. Hoskins, J. W. Carlson, K. H. Wan, S. Park, I. Mendez, S. E. Galle, B. W. Booth, B. D. Pfeiffer, R. A. George, R. Svirskas, M. Krzywinski, J. Schein, M. C. Accardo, E. Damia, G. Messina, M. Méndez-Lago, B. de Pablos, O. V. Demakova, E. N. Andreyeva, L. V. Boldyreva, M. Marra, A. B. Carvalho, P. Dimitri, A. Villasante, I. F. Zhimulev, G. M. Rubin, G. H. Karpen, and S. E. Celniker, "The Release 6 reference sequence of the Drosophila melanogaster genome," *Genome Res.*, vol. 25, pp. 445–458, Mar. 2015.

[124] P. J. Kersey, J. E. Allen, M. Christensen, P. Davis, L. J. Falin, C. Grabmueller, D. S. T. Hughes, J. Humphrey, A. Kerhornou, J. Khobova, N. Langridge, M. D. McDowall, U. Maheswari, G. Maslen, M. Nuhn, C. K. Ong, M. Paulini, H. Pedro, I. Toneva, M. A. Tuli, B. Walts, G. Williams, D. Wilson, K. Youens-Clark, M. K. Monaco, J. Stein, X. Wei, D. Ware, D. M. Bolser, K. L. Howe, E. Kulesha, D. Lawson, and D. M. Staines, "Ensembl Genomes 2013: scaling up access to genome-wide data," *Nucleic Acids Res.*, vol. 42, pp. D546–552, Jan. 2014.

[125] H. Li, "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM," *arXiv*, p. 1303.3997, Mar. 2013.

[126] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, pp. 2078–2079, Aug. 2009.

[127] A. R. Quinlan and I. M. Hall, "BEDTools: a flexible suite of utilities for comparing genomic features," *Bioinformatics*, vol. 26, pp. 841–842, Mar. 2010.

[128] M. N. Price, P. S. Dehal, and A. P. Arkin, "FastTree 2–approximately maximum-likelihood trees for large alignments," *PLoS ONE*, vol. 5, no. 3, p. e9490, 2010.

[129] E. Paradis, J. Claude, and K. Strimmer, "APE: Analyses of Phylogenetics and Evolution in R language," *Bioinformatics*, vol. 20, pp. 289–290, Jan. 2004.

[130] I. Letunic and P. Bork, "Interactive Tree Of Life v2: online annotation

and display of phylogenetic trees made easy," *Nucleic Acids Res.*, vol. 39, pp. W475–478, July 2011.

[131] A. South, "rworldmap: A new r package for mapping global data," *The R Journal*, vol. 3, pp. 35–43, June 2011.

[132] B. Czech, J. Preall, J. McGinn, and G. Hannon, "A Transcriptome-wide RNAi Screen in the Drosophila Ovary Reveals Factors of the Germline piRNA Pathway," *Molecular Cell*, vol. 50, pp. 749–761, June 2013.

[133] F. Cunningham, M. R. Amode, D. Barrell, K. Beal, K. Billis, S. Brent, D. Carvalho-Silva, P. Clapham, G. Coates, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, A. K. Kähäri, S. Keenan, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, B. Overduin, A. Parker, M. Patricio, E. Perry, M. Pignatelli, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, B. L. Aken, E. Birney, J. Harrow, R. Kinsella, M. Muffato, M. Ruffier, S. M. J. Searle, G. Spudich, S. J. Trevanion, A. Yates, D. R. Zerbino, and P. Flicek, "Ensembl 2015," *Nucleic Acids Res.*, vol. 43, pp. D662–669, Jan. 2015.

[134] A. Stamatakis, "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies," *Bioinformatics*, vol. 30, pp. 1312–1313, May 2014.

[135] M. Güell, V. van Noort, E. Yus, W.-H. Chen, J. Leigh-Bell, K. Michalodimitrakis, T. Yamada, M. Arumugam, T. Doerks, S. Kühner, M. Rode, M. Suyama, S. Schmidt, A.-C. Gavin, P. Bork, and L. Serrano, "Transcriptome complexity in a genome-reduced bacterium," *Science*, vol. 326, pp. 1268–1271, Nov. 2009.

[136] B. Li and C. N. Dewey, "RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome," *BMC Bioinformatics*, vol. 12, p. 323, Aug. 2011.

[137] R. D. C. Team, *R: a language and environment for statistical computing*. 2012. ISBN 3-900051-07-0.

[138] Y. Si, P. Liu, P. Li, and T. P. Brutnell, "Model-based clustering for RNA-seq data," *Bioinformatics*, vol. 30, pp. 197–205, Jan. 2014.

[139] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, pp. 139–140, Jan. 2010.

[140] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. R. Stat. Soc. Series. B Stat. Methodol.*, vol. 57, pp. 289–300, Jan. 1995.

[141] M. D. Robinson and G. K. Smyth, "Small-sample estimation of negative binomial dispersion, with applications to SAGE data," *Biostatistics*, vol. 9, pp. 321–332, Apr. 2008.

[142] W. Huang, A. Massouras, Y. Inoue, J. Peiffer, M. Ràmia, A. M. Tarone, L. Turlapati, T. Zichner, D. Zhu, R. F. Lyman, M. M. Magwire, K. Blankenburg, M. A. Carbone, K. Chang, L. L. Ellis, S. Fernandez, Y. Han, G. Highnam, C. E. Hjelmen, J. R. Jack, M. Javaid, J. Jayaseelan, D. Kalra, S. Lee, L. Lewis, M. Munidasa, F. Ongeri, S. Patel, L. Perales, A. Perez, L. Pu, S. M. Rollmann, R. Ruth, N. Saada, C. Warner, A. Williams, Y.-Q. Wu, A. Yamamoto, Y. Zhang, Y. Zhu, R. R. H. Anholt, J. O. Korbel, D. Mittelman, D. M. Muzny, R. A. Gibbs, A. Barbadilla, J. S. Johnston, E. A. Stone, S. Richards, B. Deplancke, and T. F. C. Mackay, "Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines," *Genome Res.*, vol. 24, pp. 1193–1208, July 2014.

[143] H. Mercot, B. Llorente, M. Jacques, A. Atlan, and C. Montchamp-Moreau, "Variability within the Seychelles Cytoplasmic Incompatibility System in Drosophila Simulans," *Genetics*, vol. 141, pp. 1015–1023, Nov. 1995.

[144] F. Rousset and M. Solignac, "Evolution of single and double Wolbachia symbioses during speciation in the Drosophila simulans complex.," *Proc Natl Acad Sci U S A*, vol. 92, pp. 6389–6393, July 1995.

[145] S. Siozios, P. Ioannidis, L. Klasson, S. G. E. Andersson, H. R. Braig, and K. Bourtzis, "The diversity and evolution of Wolbachia ankyrin repeat domain genes," *PLoS ONE*, vol. 8, p. e55390, Feb. 2013.

[146] M. S. Humphrys, T. Creasy, Y. Sun, A. C. Shetty, M. C. Chibucos, E. F. Drabek, C. M. Fraser, U. Farooq, N. Sengamalay, S. Ott, H. Shou, P. M. Bavoil, A. Mahurkar, and G. S. A. Myers, "Simultaneous Transcriptional

Profiling of Bacteria and Their Host Cells," *PLoS ONE*, vol. 8, p. e80597, Dec. 2013.

[147] A. C. Darby, A. Christina Gill, S. D. Armstrong, C. S. Hartley, D. Xia, J. M. Wastling, and B. L. Makepeace, "Integrated transcriptomic and proteomic analysis of the global response of Wolbachia to doxycycline-induced stress," *ISME J*, vol. 8, pp. 925–937, Apr. 2014.

[148] F. Gutzwiller, C. R. Carmo, D. E. Miller, D. W. Rice, I. L. G. Newton, R. S. Hawley, L. Teixeira, and C. M. Bergman, "Dynamics of Wolbachia pipientis Gene Expression Across the Drosophila melanogaster Life Cycle," *G3 (Bethesda)*, vol. 5, no. 12, pp. 2843–2856, 2015.

[149] D. J. McCarthy, Y. Chen, and G. K. Smyth, "Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation," *Nucleic Acids Res.*, vol. 40, pp. 4288–4297, May 2012.

[150] S. Bennuru, Z. Meng, J. M. C. Ribeiro, R. T. Semnani, E. Ghedin, K. Chan, D. A. Lucas, T. D. Veenstra, and T. B. Nutman, "Stage-specific proteomic expression patterns of the human filarial parasite Brugia malayi and its endosymbiont Wolbachia," *Proc Natl Acad Sci U S A*, vol. 108, pp. 9649–9654, June 2011.

[151] N. A. Moran, "Accelerated evolution and Muller's rachet in endosymbiotic bacteria," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 93, pp. 2873–2878, Apr. 1996.

[152] M. A. Fares, E. Barrio, B. Sabater-Muñoz, and A. Moya, "The Evolution of the Heat-Shock Protein GroEL from Buchnera, the Primary Endosymbiont of Aphids, Is Governed by Positive Selection," *Mol Biol Evol*, vol. 19, pp. 1162–1170, July 2002.

[153] P. Caturegli, K. M. Asanovich, J. J. Walls, J. S. Bakken, J. E. Madigan, V. L. Popov, and J. S. Dumler, "ankA: an Ehrlichia phagocytophila group gene encoding a cytoplasmic protein antigen with ankyrin repeats," *Infect. Immun.*, vol. 68, pp. 5277–5283, Sept. 2000.

[154] J. L. Sisko, K. Spaeth, Y. Kumar, and R. H. Valdivia, "Multifunctional analysis of Chlamydia-specific genes in a yeast expression system," *Mol. Microbiol.*, vol. 60, pp. 51–66, Apr. 2006.

[155] M. Lin, A. den Dulk-Ras, P. J. J. Hooykaas, and Y. Rikihisa, "Anaplasma phagocytophilum AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection," *Cell. Microbiol.*, vol. 9, pp. 2644–2657, Nov. 2007.

[156] K. M. O'Brien, E. L. Lindsay, and V. J. Starai, "The Legionella pneumophila Effector Protein, LegC7, Alters Yeast Endosomal Trafficking," *PLoS ONE*, vol. 10, Feb. 2015.

[157] C. Dale, G. R. Plague, B. Wang, H. Ochman, and N. A. Moran, "Type III secretion systems and the evolution of mutualistic endosymbiosis," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, pp. 12397–12402, Sept. 2002.

[158] G.-H. Wang, L.-M. Niu, G.-C. Ma, J.-H. Xiao, and D.-W. Huang, "Large proportion of genes in one cryptic WO prophage genome are actively and sex-specifically transcribed in a fig wasp species," *BMC Genomics*, vol. 15, p. 893, Oct. 2014.

[159] K. T. Reynolds, L. J. Thomson, and A. A. Hoffmann, "The effects of host age, host nuclear background and temperature on phenotypic effects of the virulent Wolbachia strain popcorn in Drosophila melanogaster," *Genetics*, vol. 164, pp. 1027–1034, July 2003.

[160] R. Yamada, K. D. Floate, M. Riegler, and S. L. O'Neill, "Male development time influences the strength of Wolbachia-induced cytoplasmic incompatibility expression in Drosophila melanogaster," *Genetics*, vol. 177, pp. 801–808, Oct. 2007.

[161] D. Poinsot, S. Charlat, and H. Merçot, "On the mechanism of Wolbachia-induced cytoplasmic incompatibility: confronting the models with the facts," *Bioessays*, vol. 25, pp. 259–265, Mar. 2003.

[162] M. E. Clark, C. L. Anderson, J. Cande, and T. L. Karr, "Widespread prevalence of Wolbachia in laboratory stocks and the implications for Drosophila research," *Genetics*, vol. 170, pp. 1667–1675, Aug. 2005.

[163] G. Dietzl, D. Chen, F. Schnorrer, K.-C. Su, Y. Barinova, M. Fellner, B. Gasser, K. Kinsey, S. Oppel, S. Scheiblauer, A. Couto, V. Marra, K. Keleman, and B. J. Dickson, "A genome-wide transgenic RNAi library

for conditional gene inactivation in Drosophila," *Nature*, vol. 448, pp. 151–156, July 2007.

[164] S. Dhanjal, A. K. Singh, and S. S. Cameotra, "Global Gene Expression Analysis of Bacterial Stress Response to Elevated Concentrations of Toxic Metalloids—Selenium and Arsenic," *Geomicrobiology Journal*, vol. 31, pp. 480–492, July 2014.

[165] S. R. Bordenstein and S. R. Bordenstein, "Temperature Affects the Tripartite Interactions between Bacteriophage WO, Wolbachia, and Cytoplasmic Incompatibility," *PLoS ONE*, vol. 6, p. e29106, Dec. 2011.