Wolbachia and the transcriptional landscape of stress:

A comparison between strains *w*MelPop-CLA and *w*AlbB at single-nucleotide resolution

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

(Dec 2022)

By

Youseuf Suliman

Acknowledgements

To my supervisors and mentors for their patience, guidance, and advice, even when my own stress response was not expressed. True mutualists.

List of Abbreviations

CI: Cytoplasmic incompatibility
EAM: Eukaryotic association module
MGE: Mobile genetic element
PBP: Penicillin binding protein
GTase: Glycosyltransferase
TPase: Transpeptidase
PG: Peptidoglycan
RNAP: RNA polymerase
σ: Sigma factor
HSP: Heat shock protein
ssDNA: Single-stranded DNA
TCS: Twp-component system
HK: Histidine kinase
RR: Response regulator
5'PPP: 5' triphosphate
TSS: Transcriptional start site
ncRNA: Non-coding RNA
RT-qPCR: Reverse transcriptase quantitative polymerase chain reaction
DTB-GTP: Desthiobiotin-guanosine triphosphate
pTSS: Primary TSS
gTSS: Secondary TSS
iTSS: Internal TSS
asTSS: Antisense TSS
oTSS: Orphan TSS
TSO: Template switching oligonucleotide
F6PA: Fructose-6-phospho aldolase
CPM: Counts per million
TEX: Terminator exonuclease
TAP: Tobacco acid pyrophosphatase
UIP: Uncultivable intracellular pathogens
dRNA-seq: Differential RNA-seq
OM: Outer membrane

Eσ: RNAP Holoenzyme
ANK: Ankyrin
ACC: Acetyl-CoA carboxylase
T4SS: Type foure secretion system
T1SS: Type one secretion system
KO: Kegg orthology
KEGG: Kyoto Encyclopedia of Genes and Genomes
FDR: False discovery rate
DE: Differential expression
KJE: DnaK-DnaJ-GrpEb system
GroELS: GroEL/GroES system
OMP: Outer membrane protein
IMD: Immune deficiency
PGRP: peptidoglycan recognition protein
EBP: exopolysaccharide biosynthesis protein
PA: phosphatidic acid
ABC: ATP-binding cassette
MFP: Membrane fusion protein
RNAi: RNA interference
MFS: Major facilitator superfamily
IM: Inner membrane
IS: Insertion sequence
ATU: Alternative transcriptional unit
MFE: Minimum free energy

Abstract

Symbiont-mediated control of arbovirus transmission entails the transfection of the alpha-proteobacterium *Wolbachia* into mosquito vectors, which greatly reduces vector competence. The choice of optimal *Wolbachia* strain for mosquito population replacement programmes needs to consider several factors, such as potential negative fitness effects on the mosquito host, the speed of replacement induced by cytoplasmic incompatibility, and the stability of the transinfection in the face of stressors such as heat. Mounting evidence indicates that different *Wolbachia* strains exhibit profound differences in resilience to the high temperatures that are regularly experienced in dengue-endemic countries. For instance, *w*MelPop-CLA (originally isolated from *Drosophila melanogaster*) is more easily lost from transinfected *Aedes aegypti* at elevated temperatures than is *w*AlbB (native to *Aedes albopictus*). However, very little is understood about the regulatory response to stress in *Wolbachia* at the transcriptional level. Here, we apply Cappable-Seq to unveil the primary transcriptome of these *Wolbachia* strains under low, optimal and high temperatures during mosquito cell culture. Stark differences in temperature tolerance between *w*MelPop-CLA and *w*AlbB were confirmed *in vitro.* Transcriptional start sites (TSS) were mapped at single-nucleotide resolution for the first time in an endosymbiotic bacterium responding to stress, demonstrating that both *w*MelPop-CLA and *w*AlbB have the capacity to utilise all four known TSS types, including a high proportion of antisense TSS. Global TSS discovery also enabled the identification of TSS promoter motifs associated with the global regulator sigma-70. Interestingly, wMelPop-CLA exhibited very high levels of expression for prophage-related TSS, whereas insertion sequence-associated TSS were more dominant in *w*AlbB. Evidence is also provided for the ability of *wolbachia* to express alternative transcriptional units/sub-operons that can independently regulate *wspB.* This study represents the first comparative transcriptomic analysis between *Wolbachia* strains in use for transinfected mosquito strategies and will facilitate experimentally guided strain choice in future.

Declaration of contributions to the present study

Youseuf Suliman

- Wrote the thesis
- Data analysis
- Experimental design
- Construction of Cappable-seq libraries
- Cell culture/maintenance
- Aa23 stress induction (Temperature)
- 5'RACE

Li Zhiru

- Experimental design
- Aa23 stress induction (Antibiotic)

Philip Dyer

- RML-12 stress induction (Temperature)

Catherine Hartley

- RML-12 Cell culture/maintenance

Amit Sinha

- Bioinformatic suggestions

Sarah Clarke

- Preliminary *w*MelPop-CLA antibiotic and temperature investigations

Chapter 1: Introduction

Symbiosis

"No man is an island" was a phrase by John Donne in his poem "*Devotions upon Emergent Occasions*" (John Donne, 1624), its meaning implies that no-one is self-sufficient and the same holds true for all living organisms. The term symbiosis was first coined in a biological context by Heinrich Anton de Bary in his 1878 lecture Die Erscheinung der Symbiose "The phenomenon of symbiosis" to be defined as "a phenomenon in which dissimilar organisms live together" (Oulhen, Schulz and Carrier, 2016). Symbiotic relationships have drastically changed the understanding of life on the planet, from the origin of the eukaryotic cell (Sagan, 1967) to the concept of the holobiont as an integrated unit of selection (Zilber-Rosenberg and Rosenberg, 2008). Symbiosis can generally be categorized within a spectrum of fitness costs to each member of the relationship: commensalism where one member increases its fitness with no effect on the partner, mutualism where both members increase fitness, and parasitism when the fitness gain of one member is at the cost of the other. With microbes being the earliest known cellular forms of life (Cavalazzi *et al.*, 2021) and arthropods the first terrestrial animals (Garwood and Edgecombe, 2011), the long span of evolutionary time has made symbiotic associations between the two a common occurrence. Vertically transmitted insect-symbiont relationships, predicted as far back as 270 ma, are older than the dawn of mammals at 178 ma (Moran, Tran and Gerardo, 2005).

Insect microbe symbiosis

The plethora of services provided by symbiont to their hosts has allowed insects to radiate into otherwise inhabitable niches, and these associations have led to the extraordinary diversity of insects (Moran, Tran and Gerardo, 2005). Bacteria are metabolically diverse relative to animals and can synthesis nutrients, such as amino acids and vitamins, which may be lacking in the animal diet. Nutrient provisioning is especially prevalent in sap-feeding insects, as plant sap is high in carbohydrates but poor in nitrogenous compounds, which can be supplemented by endosymbionts. Classic examples of nutritional supplementation include provisioning of essential amino acids to aphids (*Acyrthosiphon pisum*) by the symbiont *Buchnera aphidicola* (Hansen and Moran, 2011) and riboflavin to the blood-sucking Tsetse fly by its endosymbiont

Wiggleworthia glossinidia (Akman *et al.*, 2002). Of the bacterial symbionts which have caught the attention of scientific research, the vertically transmitted alpha-proteobacterium *Wolbachia* is of particular interest.

Discovery of *Wolbachia*

Wolbachia, formally named *Wolbachia pipientis*, is an vertically transmitted alpha-proteobacterium that was first described as a *Rickettsia-like* organism residing within the gonads of the *Culex pipiens* mosquito (Hertig and Wolbach, 1924). Since its initial discovery, *Wolbachia* has been found to inhabit ~50% of all arthropod species (Weinert *et al.*, 2015) and various filarial nematodes (Lefoulon *et al.*, 2016). *Wolbachia* are currently phylogenetically categorized into 17 supergroups; the majority of strains are placed within supergroups A and B, with facultative associations in arthropods typically classified by whole-genome sequencing and multi-locus sequence typing (Bleidorn and Gerth, 2017). *Wolbachia* have also been found to have obligate associations within filarial nematodes of supergroups C and D (Taylor, Bandi and Hoerauf, 2005). The last common ancestor of the arthropod (supergroups A and B) and filarial *Wolbachia* (Supergroups C and D) appears to have diverged around ~100 Ma (Bandi *et al.*, 1998). Interestingly the *Wolbachia* of supergroup F appear to be the only supergroup capable of infecting both arthropods and nematodes (Lefoulon *et al.*, 2012).The closest relatives of *Wolbachia* include *Rickettsia, Ehrlichia,* and *Anaplasma.*

Reproductive manipulation

Wolbachia are famously able to manipulate their host using a wide range of reproductive manipulations earning the status of reproductive parasite. The end goal of *Wolbachia* is to be transmitted through the female germline and spread to subsequent generations. The distinct reproductive strategies of *Wolbachia* include feminization of developing males, parthenogenetic development of unfertilized eggs, termination of male embryos, and cytoplasmic incompatibility. All the manipulative reproductive strategies effectively enhance the reproductive success of infected females and promote the vertical transmission of the symbiont.

The most studied reproductive phenotype of arthropod *Wolbachia* is cytoplasmic incompatibility (CI). CI is a form of sterility which permits infected females reproductive compatibility with all males regardless of *Wolbachia* infection status, although when infected males mate with an uninfected female, or a female with an incompatible *Wolbachia* strain, most developing embryos are terminated (Yen and Barr, 1971). Embryonic lethality is a result of the failure of paternal chromatin to condense, causing lethal mis-segregation and bridging of paternal DNA during the first zygotic mitosis (Reed and Werren, 1995; Callaini, Dallai and Riparbelli, 1997). The result is that infected females have a reproductive advantage over uninfected females; *Wolbachia* is then able to increase its transmission to subsequent progeny and spread throughout a population.

Only recently have the genetic roots of CI been uncovered. The *cifA*-*cifB* (homologous to *cidA*-*cidB*) gene pair has been proposed as an toxin-antitoxin system of prophage WO origin (Beckmann, Ronau and Hochstrasser, 2017; LePage *et al.*, 2017), products of which were found in in the spermathecae of infected females (Beckmann and Fallon, 2014). The *cifA*-*cifB* gene pairs are almost ubiquitous within supergroups A and B, although these contain most genomes that have been sequenced. The CidB protein has deubiquitylase activity, which when mutated, results in loss of CI (Beckmann, Ronau and Hochstrasser, 2017). Although CidA binds to CidB, indicative of a toxin-antitoxin system, expression of both genes is required for proper CI induction, opposing the toxin-antitoxin model (LePage *et al.*, 2017). Another proposed model is called the Two-by-One model, in which expression of both *cifA* and *cifB* is required for CI induction, whilst *cifA* expression within the female germline results in rescue (Shropshire and Bordenstein, 2019). Althought the CI genes have been narrowed down, the exact mechanism of CI is still poorly understood. Interestingly, the ability of *Wolbachia* to manipulate its host for its continued survival has itself been manipulated by humans for theirs.

Disease-vector control

Mosquitoes are the predominant vectors of arthropod borne diseases of humans, able to spread various life-threatening pathogens responsible for malaria, dengue, and Zika. Dengue is caused by one of four serotypes (DENV 1-4) of the single-stranded positive-sense RNA virus within the Flaviviridae family. Primarily spread by the *Aedes aegypti* mosquito, DENV infects up to 390 million people globally each year and is now endemic on over 100 countries (Bhatt *et al.*, 2013). With no effective antiviral therapy and only one FDA approved vaccine, control

strategies are mainly preventative involving the removal of mosquito breeding grounds and application of insecticides that are increasingly becoming less effective (WHO, 2009). The requirement for an effective and reliable therapy or preventative strategy is urgently needed.

Teixeira, Ferreira and Ashburner (2008) were the first to report that *Wolbachia* infection within *Drosophila melanogaster* provided protection again RNA viruses. Shortly after, a similar observation for the *Wolbachia* induced reduction on DENV was reported in *A. aegypti* mosquitos (Bian *et al.*, 2010). The ability of *Wolbachia* to vertically spread through a population and inhibit RNA viruses eventually led to the development of control strategies involving release of *Wolbachia*-infected mosquitos to prevent DENV. The concept was first experimentally validated within a field-simulated caged population of DENV-infected *A. aegypti,* in which *w*Mel was shown to reduce the viral load by up to 2,600-fold, reaffirming its application in reducing vector-competency (Walker *et al.*, 2011). The application of symbiotic bacteria in transmission prevention of arthropod-borne pathogens is not a novel strategy. The reduviid bug (*Rhodnius prolixus*) is an arthropod vector of *Trypanosoma cruzi* responsible for Chagas disease; transformation of its symbiont *Rhodococcus rhodnii* to express the peptide cecropin-A was used to lethal effect against *T. cruzi* as early as 1997 (Durvasula *et al.*, 1997). However, the release of a genetically modified organism has high regulatory and environmental safety rules which must be satisfied before any attempt to release into the field is permitted. Since *Wolbachia* naturally infects many arthropods and its antiviral properties do not need genetic intervention, it is an ideal candidate for field-release.

The first field-release trial of *Wolbachia* infected mosquitos was conducted in January 2011 in Cairns, Australia. The released *w*Mel strain successfully invaded the native *A. aegypti* population to near fixation in just over three months (Hoffmann *et al.*, 2011). The *w*Mel population of infected *A. aegypti* has since been stable for over a decade with local dengue transmission almost eliminated (Ross *et al.*, 2022). Since its initial conception, the *w*Mel strain has now been released into mosquito populations in several countries including Australia, Brazil, Indonesia, and Vietnam (Ross, Turelli and Hoffmann, 2019). Alternative strains of *Wolbachia* are also proving to be highly effective; the *w*AlbB strain originating from *A. albopictus* is also showing promising results in Malaysia (Ahmad *et al.*, 2021).

Wolbachia strains and phenotypic variation

The choice of strain can have a significant impact on the success *Wolbachia*-mosquito release strategies. Strain *w*MelPop, a variant of *w*Mel, was found during an irrelevant screening for *D. melanogaster* gene mutations (Min and Benzer, 1997). The name *popcorn* was given due to its high systemic proliferation within the *D. melanogaster* body; the high cell density of *w*MelPop eventually led to cell rupture and early death (50% lifespan) of the adult fly (Min and Benzer, 1997). The fitness cost of *w*MelPop on *A. aegypti* in the field proved far too great resulting in eventual loss of infection in the field (Nguyen *et al.*, 2015).

Wolbachia density varies between strains and highly correlates with pathogen blocking (Osborne *et al.*, 2009; Chrostek *et al.*, 2013). Current models in which *Wolbachia* causes viral inhibition are generally centred around host immune priming (Kambris *et al.*, 2010) and competition over intracellular resources. *Wolbachia* infection is capable of upregulating components of the innate immune system in mosquitos; for instance, *w*MelPop-CLA infection of *Ae. aegypti* resulted in 39% of genes involved in immune-related functions such as peptidoglycan recognition proteins and Toll pathway genes being upregulated (Kambris *et al.*, 2009). *Wolbachia* lack a typical cell wall but appear to derive their membranes from the hosts endoplasmic reticulum (Fattouh, Cazevieille and Landmann, 2018). Proteome and lipidome analysis of *Wolbachia*-infected mosquito cells have shown decreases in components related to metabolism, including lipids such as cholesterol (Molloy *et al.*, 2016; Geoghegan *et al.*, 2017). Cholesterol is a key lipid in enveloped RNA viruses involved in virion assembly, replication, and host cell entry. A decrease in the concentration of intracellular free cholesterol is accompanied by an increase in esterified cholesterol; *w*MelPop infected *Ae. aegypti* cells supplemented with a solubilizing compound targeting esterified cholesterol led to an increase in DENV by 100-fold (Geoghegan *et al.*, 2017). *Wolbachia* nutritional competition over lipids such as cholesterol appears to be one of the mechanisms behind the pathogen blocking phenotype.

Understanding the causes of phenotypic variations is crucial to ensuring the long-term success of field-releases. Martinez. *et al*, (2017) conducted a comparative study to determine whether the strain of *Wolbachia* had a larger impact on viral protection over the host genetic background. The authors reported that the effects of *Wolbachia* strain explained 90% of the variation in antiviral protection. It is becoming apparent that strain choice can have a major

impact on the effectiveness of *Wolbachia* as a biocontrol agent, and understanding the effects of genetic variation amongst strains is imperative in understanding the phenotypic affects *Wolbachia* will have when confronted with conditions in the field. Understanding the mechanistic differences between *Wolbachia* strains is hindered by the inability to grow *Wolbachia* axenically to allow genetic manipulation. Although genetic transformation of *Wolbachia* is currently out of reach, genomic and transcriptomic analysis is beginning to home-in on these key differences.

Genomic features of *Wolbachia*

Sequencing of the first complete *Wolbachia* genome was conducted on the *Drosophila melanogaster* arthropod strain *w*Mel in 2004 by Wu. *et al.* (2004) followed by the first filarial *Wolbachia,* (*w*Bm) of *Brugia malayi* (Foster *et al.*, 2005). *Wolbachia* genome sizes typically range from 1.2-1.6 Mb and 0.9-1.1 Mb for arthropod and filarial *Wolbachia* respectively. The differences in genome sizes arthropod and filarial *Wolbachia* is assumed to be due to the nature of the relationship with their host, obligate associations are assumed to evolved greater co-dependence and thus greater loss of redundant genes (Dale and Moran, 2006). However the discovery of the largest *Wolbachia* genome to date comes from the obligate *Wolbachia of the* springtail *Folsomia candida* at 1.8 Mb which challenges the reduction model (Kampfraath *et al.*, 2019). The size of the *w*Mel genome is 1.27 Mb (Mb) encoding 1,227 protein coding genes, the *w*AlbB strain originating from *Aedes albopictus* has a relatively larger genome of 1.48 Mb encoding 1,284 protein coding genes. Currently there are a total of 37 *Wolbachia* genomes in NCBI Refseq (O'Leary *et al.*, 2016) of which 11 are complete.

One of the noticeably striking features of arthropod *Wolbachia* are the high abundance of mobile elements including WO phage, insertion sequences, and a putative plasmid. The temperate WO phage has been reported in at least five supergroups, excluding the filarial mutualists, with the majority in supergroups A and B. Electron microscopic studies have observed the presence of extracellular tail-like particles indicative of their capacity for lytic rupturing of cell membranes (Masui *et al.*, 2001). Interestingly some of the associated protein-domains with WO phage appear to be of eukaryotic origin encompassing up to 70% of phage genes, these eukaryotic-like regions are termed the eukaryotic association module (EAM) and have functional implications for host-microbe interactions (Bordenstein and Bordenstein, 2016). Fascinatingly, the presence of a EAM suggests multiple occasions in which WO phage

have exchanged genes between their arthropod hosts (Bordenstein and Bordenstein, 2016). Transposons also appear to be abundant in *Wolbachia* genomes occupying up to 10% of the genome (Cordaux *et al.*, 2008).

A hallmark of intracellular pathogens and endosymbionts are their relatively small genome size in relation to free-living bacteria. Small genomes are common amongst endosymbiotic bacteria; genes which are no longer required or are complemented by their hosts are lost as endosymbionts adapt within the relatively stable environment of their host. A major mechanisms of genome reduction within endosymbionts involves the accumulation of mildly deleterious mutations and genetic drift through small population bottlenecks. *Wolbachia,* and other cytoplasmic inherited symbionts, must transmit to the next generation of their host and in doing so only a small fraction of the bacterial population are transmitted. These effectively small populations are bottlenecks which permit the sub-population with accumulated mutations to spread in subsequent hosts via genetic drift (Moran, 1996).

Opportunities for genetic recombination in obligate mutualistic endosymbionts are low. A comparative analysis between two *B. aphidicola* strains, symbionts of *Schizaphis graminum* and *Acyrthosiphon pisum,* revealed no recombination between the strains despite a divergence of 50 million years (Tamas *et al.*, 2002). Minimal recombination and chromosomal stability within obligate symbionts is reflected by their high phylogenetic congruence with their hosts as seen *B. aphidicola* (Clark *et al.*, 2000) and *Wolbachia* of filarial (Bandi *et al.*, 1998; Lefoulon *et al.*, 2016) and bedbug hosts (Balvín *et al.*, 2018). Despite also undergoing genome reduction, *Wolbachia* of arthropod hosts do not share phylogenetic congruence indicative of frequent host shifts and (Van Meer, Witteveldt and Stouthamer, 1999). Although both are vertically transmitted, one of the key genomic features of arthropod *Wolbachia* that distinguishes them from their obligate filarial counterparts is their high rate of recombination and high volume of mobile genetic elements (MGEs). MGEs, such as phage, transposases and ISs, together with DNA recombination proteins, such as RecA, are present in arthropod *Wolbachia* but noticeable absent in those residing in filarial hosts (Fallon, 2021). High recombination rates could theoretically mitigate the accumulation of mildly deleterious mutations and promote gene preservation by promoting DNA repair before reaching population bottlenecks. An example of gene preservation amongst *Wolbachia* symbionts is that of *w*Fol whose genome has undergone expansion involving genes related to phage and DNA repair genes absent in *B. aphidicola* and *w*Bm (Kampfraath *et al.*, 2019).

Reduced endosymbiotic genomes generally lose genes involved in DNA repair, cell envelope biosynthesis, and gene regulation (McCutcheon and Moran, 2012). Contrary to most reduced symbiont genome, arthropod *Wolbachia* appear to have retained genes involved in DNA repair and have even duplicated some genes such as the DNA mismatch repair protein MutL, the only known bacteria to do so. As a result of gene loss via genome reduction, the *Wolbachia* are predicted to have limited metabolic capability only encoding those required for essential survival. Contrary to *B. aphidicola,* the retention of pathways involved in amino acid transport and metabolism suggest amino acids are the main source of energy due to a lack of ATP-uptake systems (Wu *et al.*, 2004). Carbohydrate metabolism is also limited; the tricarboxylic acid cycle, nonoxidative pentose phosphate pathway, and glycolysis are complete but rely on fructose-1,6-biphosphate as the starting material. Regarding cell envelope biogenesis, the peptidoglycan biogenesis pathways of obligate intracellular bacteria lack the class-A penicillin binding proteins (PBP) for glycosyltransferase (GTase) and transpeptidase (TPase) activity required for the β1-4-glycosidic bond and cross-linking formation respectively. Loss of peptidoglycan (PG) pathway components have been suggested to reflect an adaption to the isotonic intracellular environment of the host and to evade the hosts antibacterial mechanisms. Many members of the Rickettsiales, such as *Ehrlichia* and *Anaplasma,* along with various strains of *B. aphidicola* have lost components of the PG precursor lipid II biosynthesis pathway and thus cannot a classical PG. interestingly, *Wolbachia* encode a class-B PBP which has been demonstrated to perform both GTase and TPase activity suggesting a PG-like cell wall (Atwal *et al.*, 2021).

Bacterial gene regulation in response to stress

The ability for bacteria to respond to changing condition in the environment is essential to their survival. The primary mechanism in which bacteria interact and adapt to changing conditions is at the level of the transcription regulation. The DNA-dependent RNA polymerase (RNAP) is the principal enzyme involved in the synthesis of RNA. The bacterial RNAP is a multi-subunit enzyme composed of a core enzyme of five-subunits ($\alpha_2\beta'\beta\omega$) along with an additional subunit termed the sigma (σ) factor together forming the RNAP holoenzyme. The σ-factor acts as a transcriptional activator for directing the RNAP holoenzyme towards specific sequences called promoters upstream of genes to be expressed. Multiple σ-factors exist and competitively bind to the RNAP holoenzyme to guide it towards specific promoters and transcribe condition

dependent genes. A total of seven σ-factors are known in *E. coli,* the primary sigma factor σ^{70} known as the housekeeping sigma factor and responsible for most of the growth-related gene expression. During periods of stress, alternative σ-factors guide the RNAP to promoters of genes whose functions alleviate and counter the effects of the specific stressor. In gram-negative bacteria, the σ^{32}-factor directs the expression of genes involved in the heat shock response upon exposure to increasing temperature.

The heat shock response is typified by the expression of heat shock proteins (HSPs) such as GroEL, GroES, DnaK, and DnaJ which act as molecular chaperones that restore the state of unfolded proteins induced by increasing temperatures (Richter, Haslbeck and Buchner, 2010). Besides chaperones, the HSPs also include ATP-dependant proteases such as ClpP, ClpX, and Lon which degrade and recycle abnormal proteins. As the level of unfolded proteins stabilises, DnaK and DnaJ are free to form a ternary complex with σ^{32} repressing further expression of HSPs (Gamer *et al.*, 1996). The *E. coli* alternate sigma factor σ^{S} is involved in directed the transcription of genes responding to carbon starvation. Many α-proteobacteria harbour an σ^{S} ortholog σ^{EcfG}, σ^{EcfG} is tightly controlled by PhyR and NepR which act as anti-sigma factors to sequester σ^{EcfG} under non stress inducing conditions (Francez-Charlot *et al.*, 2009). Although implicated in carbon starvation in *E. coli,* genes regulated by the general stress response have been shown to be involved in protection against oxidative stress and osmoprotection providing a level of cross-protection (Francez-Charlot *et al.*, 2015).

Besides alternative sigma factors, various proteins can regulate gene expression by masking the sequence motifs upstream of stress related genes. The SOS stress response pathway is activated by the accumulation of single stranded DNA (ssDNA) as a by-product of replicating damaged DNA (Pagès and Fuchs, 2003). The LexA repressor and RecA co-protease are two proteins which are key in the regulation of SOS genes. LexA is a dimer acting as a transcriptional repressor which binds to operator sequences with specific SOS sequence motifs thus preventing expression of downstream genes. However, the RecA co-protease functions by promoting the self-cleavage of the LexA dimer and allowing expression of downstream genes. For RecA to promote self-cleavage of LexA, RecA converts to its active form only once bound to ssDNA. Genes upregulated during SOS induction include; *uvr* for nucleotide excision repair, *sulA* which inhibits cell division to enable adequate DNA repair, and the error prone DNA

polymerase IV which enabled continued replication and survival at the expense of increased mutagenesis (Courcelle *et al.*, 2001; Henrikus, van Oijen and Robinson, 2018). Interestingly both *lexA* and *recA* are also upregulated in the SOS regulon, expression of these proteins enables continuous upregulation of the SOS pathway via RecA provided the ssDNA signal is provided, upon repair of DNA damage the ssDNA signal resides and LexA repression governs (Walker, 1984). Unlike *B. aphidicola, Wolbachia* from both arthropod and are capable of DNA recombination and repair (Wu *et al.*, 2004; Foster *et al.*, 2005). However, the LexA repressor is absent from all *Wolbachia* genomes suggesting that they are incapable of mediated the RecA induced SOS response (Fallon, 2021).

Bacteria also have the capacity to response to external signals via two-component systems (TCS). As the name suggests bacterial TCSs are composed of two units involving a membrane bound sensor histidine kinase (HK) and a cytoplasmic response regulator (RR). Upon binding to a specific signal, the HK autophosphorylates which in turn phosphorylates the RR enhancing its affinity to bind to DNA sequences (Gao, Mack and Stock, 2007).

Transcription regulation in streamlined endosymbionts

Due to reductive gene loss, *Wolbachia* encode limited potential for transcriptional regulators. TCSs are limited but the *cckA/ctrA* and *pleC/pleD* systems appear to be ubiquitous amongst *Wolbachia* strains (Christensen and Serbus, 2010). In terms of sigma factors *Wolbachia* retain only two sigma factors, the housekeeping σ^{70} and the heat shock σ^{32} (Amelia R.I. Lindsey, 2020). Highly reduced small genomes such as those of *Mycoplasma genitalium* have only retained one sigma factor whereas bacteria with larger genomes such as *E. coli* contain seven (Mittenhuber, 2002). The preservation of the heat-shock sigma-factor σ^{32} reflects the instability of proteins within streamlined genomes due to the accumulation of mutations.

The heat-shock response is one of the most ancestral mechanisms in cellular organisms, which functions to re-stabilize unfolded proteins when exposed to increasing temperatures. Most obligate pathogens and symbionts constitutively express heat shock proteins to counter these globally destabilised proteins; this is reflected in the aphid symbiont *Buchnera aphidicola*, where 10% of the protein content is dedicated to heat shock proteins (Baumann, Baumann and Clark, 1996). Regardless, *B. apidicola* was still shown to further upregulate heat shock genes,

such as *dnaK, grpE,* and *dnaJ,* in response to 37°C temperature upshifts (Wilcox *et al.*, 2003). The lack of encoded regulatory genes within *Wolbachia* has led to the notion that their streamlined genomes have limited capacity to respond to environmental stimuli.

Symbionts and stress

Temperature represents an abiotic stressor in which all living organisms must effectively respond to survive; the reduced genomes of endosymbionts along with their accumulation of deleterious mutations may hamper their ability to respond to such stressors. Heat exposure of the *Blochmannia* endosymbiont of the carpenter ant tribe Camponotini to up to 37°C was shown to significantly reduce symbiont densities by >99% (Fan and Wernegreen, 2013). The *Buchnera* symbiont of aphids is similarly impacted by thermal stress; field populations of the pea aphid *Acyrthosiphon pisum* exposed to 30°C also exhibited reduced *Buchnera* densities (Montllor, Maxmen and Purcell, 2002). Symbiont susceptibility to thermal stress can also have effects on the ecological range of their hosts. Simulated climate change from 25°C to 30°C could eliminate symbionts within a pentatomid stinkbug, reducing the fitness of its host. Indeed, thermally-induced symbiont loss has been suggested to be a factor in the ecological shift of the stinkbug from tropical to more temperate areas in Japan (Prado and Almeida, 2009; Musolin, Tougou and Fujisaki, 2010).

The effects of thermal stress on *Wolbachia* are increasingly becoming a topic of interest with implications on the long-term success of *Wolbachia* field-releases. The detrimental effects of heat on *Wolbachia* and its capacity of CI was first demonstrated in the spider mite *Tetranychus urticae,* in which heat treatment at 32°C resulted in gradual elimination of *Wolbachia* and complete loss of the CI phenotype (Van Opijnen and Breeuwer, 1999). The loss of CI via heat-induced elimination of *Wolbachia* was also demonstrated in *Drosophila* (Feder *et al.*, 1999). Thermal susceptibility to *Wolbachia* strains native to mosquitos was demonstrated with *w*AlbB in *Ae. albopictus;* thermal shifts maintained above 30°C during larval development significantly reduced *w*AlbB densities in both male and female *Ae. albopictus* (Dobson and Rattanadechakul, 2001; Wiwatanaratanabutr and Kittayapong, 2006). An electron microscopic study observed the morphological changes of *Wolbachia* at 36°C within the ovarian cells of *D. melanogaster,* and the authors reported that although no structural changes of the host cells were observed, *Wolbachia* exhibited invaginations and protrusions of its cell wall (Zhukova,

Voronin and Kiseleva, 2008). This study was the first to demonstrate that thermal stress directly affected the morphology of the symbiont by means of temperature-induced membrane instability.

Some *Wolbachia* strains harbour the bacteriophage WO (Wu *et al.*, 2004), which has been reported to be capable of forming free phage particles (Fujii *et al.*, 2004). Bordenstein and Bordenstein (2011) thermally shocked *Wolbachia* within the *Nasonia vitripennis* parasitoid wasp, as the authors wanted to determine whether the WO phage WOVitA1 was capable of lytic activation under temperature stress. The authors reported that phage expression was inversely correlated with *Wolbachia* density. Heat shock at 30°C reduced *Wolbachia* density by 3-fold whilst phage expression increased by 500%, with particular susceptibility during the egg stages (Bordenstein and Bordenstein, 2011). This was the first report that induction of lytic phage may be the cause of *Wolbachia* density decline during thermal stress and *Wolbachia* thermal susceptibility between developmental stages. Further investigating thermal sensitivity of *Wolbachia* between host developmental stages, Ulrich. *et al*, (2016) subjected the field-released *w*Mel within *Ae. aegypti* hosts to heatwave simulating conditions (30°C -40°C). The study reported that *w*Mel is most susceptible to thermal stress during early developmental stages in eggs, with simulated heatwave conditions reducing *w*Mel densities to <0.1% compared to controls.

Whilst the *w*Mel strain has quickly become a common strain being deployed in field-release strategies, research was suggesting its susceptibility to heat may hinder the success of the strategy. A seminal study conducted by Ross. *et al.* (2017) compared the thermal susceptibility of *w*Mel and *w*AlbB, two ideal candidate *Wolbachia* strains suitable for field-release. In contrast to the continuous heat stress applied by Wiwatanaratanabutr and Kittayapong (2006), the study by Ross. *et al.* (2017) reported that under cyclical heat stress *w*Mel density decreased during heat exposure resulting in partial loss of CI whilst *w*AlbB density did not change during identical thermal conditions and maintained complete CI (Ross *et al.*, 2017). These finding appear to suggest that whilst both *w*Mel and *w*AlbB are susceptible to heat stress, *w*AlbB is capable of thermal recovery within periods of thermal stability. Additionally, when the offspring of heat-treated *w*Mel and *w*AlbB infected adult flies were compared for *Wolbachia* transmission, infections were completely lost in *w*Mel whilst 88.5% of offspring in *w*AlbB heat treated adults preserved infection status. The culmination of these studies suggests that the success of *Wolbachia* field-release strategies may be vulnerable to extreme climates; they also

suggest that alternative strains such as *w*AlbB may be better suited to the thermal conditions encountered in tropical climates where dengue is most prevalent.

Few studies have investigated the genetic causes of endosymbiont susceptibility to thermal stress. One such study conducted by Dunbar. *et al*, (2007) reported a single point mutation in the aphid symbiont *B. aphidicola* located in the promotor region of the heat shock gene *ibpA,* which significantly reduced symbiont density during heat-shock exposure and resulted in reproductive failure of the host (Dunbar *et al.*, 2007). This study emphasizes the informative application of bioinformatic analyses on endosymbiont genomes to investigate phenotypic variations in response to environmental conditions. Due to the genetic intractability of endosymbionts and high ratio of host-symbiont transcripts, functional genomic studies are difficult to perform on endosymbionts. However various methods are being implemented and developed to overcome such technical barriers.

Overview of *Wolbachia* gene expression studies

Few global transcriptional studies have been conducted on endosymbionts relative to their free-living counterparts. The first *Wolbachia* transcriptional analysis was conducted by Darby, *et al.* (2012) comparing the *w*Oo transcriptome within male soma and female germline of the filarial nematode *Onchocerca ochengi* . Using RNA-seq, the authors were able to report that *w*Oo constitutively expressed heat-shock genes such as chaperones and proteases (*groEL, groES, clpB, dnaK*, and *lon*), consistent with the previously mentioned protein instability of degraded genomes. The authors also reported upregulation of transcripts for membrane transport and respiration in the male soma whilst the female germline upregulated DNA replication and transcription. This pioneering transcriptional study was the first to report that *Wolbachia* are capable of regulating genes relevant to its lifestyle. Luck. *et al.* (2014) further investigated *Wolbachia* transcriptional potential by investigating the filarial *Wolbachia* (*w*Di) throughout the lifecycle of its host, *Dirofilaria immitis*. The study observed differential expression of stage-specific *Wolbachia* expression ranging from heme, nucleotide, and riboflavin synthesis, suggesting supplementation to the host during relevant developmental stages, as well as lipid synthesis during early microfilarial stages. This study was able to confirm that *Wolbachia* can regulate genes involved in host-symbiont interactions in the highly dynamic environment of a developing host.

The first global transcriptional profiling of an arthropod *Wolbachia* was conducted by Darby. *et al*, (2014) investigating the transcriptional response of *w*MelPop-CLA to the antibiotic doxycycline. This study was the first to apply an external stressor on the host-symbiont system. The authors were able to report that *w*MelPop-CLA upregulated ribosomal genes in response to the known ribosome inhibiting action of doxycycline; genes also involved in nucleotide and phospholipid synthesis were also upregulated, suggesting that *Wolbachia* increases expression of genes for membrane remodelling in response to antibiotic stress. Interestingly, the study was the first to show evidence for antisense transcripts associated to prophage regions. The study demonstrated that *Wolbachia* of arthropods can respond to specific stressors by upregulating genes besides those related to the heat-shock response, this provided the first clues that *Wolbachia* can express antisense transcripts indicative of non-coding RNA. In terms of arthropod *Wolbachia,* the only remaining global transcriptional study conducted was by Gutzwiller. *et al.* (2015). The authors investigated *w*Mel gene expression within 24 developmental life cycle stages of the *D. melanogaster* host, including male and female sex specificity. The study reported differential expression specific to embryonic, larval, and adult stages. During embryonic stages, regulated genes mainly consisted of those involved in chaperones such as *groES, dnaK, hsp90,* and *groEL.* Larval stages displayed upregulation for genes involved in protein and membrane secretion and interestingly, ankyrin genes were highly regulated, suggesting a possible role in host interactions during metamorphosis. Lastly, genes also displayed male-specific upregulation in the adult stages, which mainly consisted of ankyrins. Additionally, the study analysed differentially expressed genes that were specific to arthropod *Wolbachia,* including genes involved in ankyrin and membrane secretion, which highlighted the importance of intracellular host communication in arthropods compared to filarial *Wolbachia.* Interestingly, the study highlighted the presence of antisense transcripts associated to phage components that were previously reported by Darby. *et al*, (2014).

ncRNA in Wolbachia

Studies investigating ncRNA in *Wolbachia* are limited but illuminating. Currently, only two studies have focused on ncRNA within arthropod *Wolbachia* (Mayoral, *et al.*, 2014 and Woolfit, *et al.*, 2015). The study conducted by Mayoral. *et al*, (2014) reported the presence of two snRNA (*WsnRNA-49* and *WsnRNA-59*) in *w*MelPop-CLA within *Ae. aegypti,* providing evidence that *WsnRNA-49* was capable of both downregulating a *Wolbachia murD* peptidoglycan-related gene and being exported into the host. The authors also supported the

repeated presence of an antisense transcript (*WsnRNA-59*) associated with the phage major capsid protein. The study conducted by Woolfit. *et al,* (2015) used conserved intergenic regions between *w*Mel and *w*Pip to discover two candidate snRNA, *ncrwmel01* and *ncwmel02*. The *ncrwmel02* transcript was shown to be urregulated in the male *D. melanogaster* testes, suggesting a role in host reproduction. Overall, it is becoming clear that ncRNA is an understudied feature in *Wolbachia* biology; furthur research into ncRNA within *Wolbachia* should improve our understanding of the regulatory capacity of *Wolbachia.*

Wolbachia enrichment methods

Although next-generation sequencing has allowed the transcriptomes of endosymbionts to be studied, they are still compromised by the abundance of host transcripts and bacterial rRNA resulting in little coverage of symbiont mRNA. Methods involving the use of terminator-5'-phosphate-dependent exonuclease, which enzymatically degrades 5' phosphate associated rRNA, aim to deplete the majority of ribosomal transcripts which dominate the transcriptional pool (Sharma *et al.,* 2010). However, these methods still lack efficiency as the enzymatic degradation is incomplete. Other methods apply the use of bait libraries, such as the Agilent SureSelect system, with predesigned probes in order to detect specific transcripts within a mixed RNA pool (Chung *et al.,* 2018). Although the Agilent SureSelect system is highly effective at enriching mappable reads, probe design can omit unique genes which may be absent in genome annotations. It is also possible that probes may capture degraded RNA, captured transcripts may not accurately represent nascent transcriptional activity of the cell.

Recently a new method has been developed which involves the selective capping of prokaryotic 5' triphosphate (5'PPP) RNA with biotinylated GTP, effectively separating primary prokaryotic mRNA from host mRNA and bacterial rRNA in the RNA pool. This method is termed Cappable-seq (Ettwiller *et al.,* 2016) and has been successfully applied to the *w*Bm *Wolbachia* within the *B. malayi* filarial host (Luck, Slatko and Foster, 2017). The study was able to show up to 5-fold enrichment of non-ribosomal mRNA transcripts within *Wolbachia,* allowing a more comprehensive transcriptional profiling of *Wolbachia* over conventional methods. A benefit of the selective enrichment of 5'PPP is that it can define the transcriptional start site (TSS) of the captured nascent primary mRNA at single-nucleotide resolution, highlighting an early transcriptional snapshot of gene expression within the bacterium without the need for predesigning probes.

Research questions and aims

This study aims to apply the effective enrichment and TSS identification capabilities of Cappable-seq to map out the global TSS location and diversity within *w*Mel and *w*AlbB. The enrichment capabilities of Cappable-seq will also permit this study to investigate differential expression of *w*Mel and *w*AlbB under stress-inducing conditions. This study will then attempt to compare the transcriptional response of *w*Mel and *w*AlbB under temperature stress to investigate the transcriptional differences in heat susceptibility between these commonly used *Wolbachia* strains that are being used in field-release vector control strategies. Along with the abiotic stress of temperature, this study will also aim to investigate the transcriptional response of *w*AlbB to various antibiotics with varying mechanisms of action including doxycycline, rifampicin, and moxifloxacin. Lastly, with the identification of global TSS, this study will aim to uncover operon structure within both *w*Mel and *w*AlbB, along with the discovery of potentially novel non-coding RNA (ncRNA). This study aims to further the understanding of *Wolbachia* and its transcriptional responses to stress to aid strain selection in vector-release strategies.

Chapter 2: Global mapping of TSS within wMelPop-CLA and wAlbB

Introduction

The TSS represents the immediate moment of transcription in an organism transcriptome. By identifying the location of a TSS the upstream sequences can be investigated for evidence of promoter motifs which represents regions capable of regulation for regulatory elements such as sigma factor binding sites. The classic view of transcription assumes that for a gene to be functionally expressed its TSS is generally found upstream to a protein coding gene. With the emergence of cost-effective high throughput sequencing, research is beginning to demonstrate that this is not the case for majority of TSS within reduced bacterial genomes (Sharma *et al.*, 2010). However, investigating global TSS within endosymbionts is hindered by the overwhelming majority of transcripts belonging to both the host and rRNA (Giannoukos *et al.*, 2012). In an attempt to improve the understanding of *Wolbachia* gene expression, this study

has applied Cappable-seq (Ettwiller *et al.*, 2016) to reveal a level of transcriptional complexity as yet unseen between arthropod associated *Wolbachia*.

The following analysis provides the first comparative global TSS maps between two *Wolbachia* strains significant to biocontrol strategies of vector-born diseases. This following chapter aims to subject wMelPop-CLA and *w*AlbB to abiotic (temperature) and biotic (antibiotics) stressors in order detect a diverse set of TSS in response to environmentally relevant stressors. The study begins by recapitulating the known effects of environmental stressors within an *Ae. albopictus* cell line to support a model in which *Wolbachia* stress responses can be investigated. The subsequent analyses also demonstrate the efficacy of Cappable-seq for the enrichment of *Wolbachia* transcripts. The single-nucleotide resolution of Cappable-seq has uncovers the true complexity of *Wolbachia* gene expression, this study reports that arthropod associated genomes are replete with intergenic and antisense transcripts which may give insight into regulatory mechanism with reduced genomes. lastly, the chapter ends with a comparative view of gene expression between *w*MelPop-CLA and *w*AlbB at baseline conditions hinting at transcriptional differences that may define how *Wolbachia* strains differ.

Methods

Preliminary 34°C and doxycycline exposure (Conducted at University of Liverpool, Liverpool, UK)

Cell culture

To investigate the transcriptional response of *w*MelPop-CLA and *w*AlbB exposed to stress, RML-12 and Aa23 cell lines were used as a model to study gene expression within the *Ae. albopictus* host. Cell lines were chosen due to their relatively ease of use and logistical maintenance compared to the whole insect body. RML-12 *Wolbachia* positive (w^+), RML-12 *Wolbachia* negative (w^-), and Aa23 *Aedes albopictus* cell lines harbouring *w*MelPop-CLA (McMeniman *et al.*, 2008) and *w*AlbB (O'Neill *et al.*, 1997) respectively were maintained at 28°C in T75 flasks. Cell lines were maintained in 15ml insect media composed of 45% Mitsuhashi & Maramorosch (HiMedia), 45% Scheiders (Sigma), 10% fetal calf serum, and

0.18% L-glutamine. Cell lines were passaged every 4-6 days until confluency in a 1:5 ratio (3ml suspended cells to 12ml fresh media).

Preliminary 34°C and doxycycline exposure

Doxycycline was diluted from doxycycline hyclate in ultrapure water to 0.25µg/ml at minimum bactericidal concentrations as previously reported (Hermans, Hart and Trees, 2001). Total concentration added to RML-12 (w^+) was 0.25µg/ml, an equivalent volume of ultrapure water was added to control cells. Cells were maintained in 6-well plates with each well containing 3 ml of RML-12 (w^+) either with or without doxycycline in triplicate. Cell plates were maintained at their respective temperatures of 28°C or 34°C with and without doxycycline. Cell plates were passaged every 7 days (1:5 as above) into new 6-well plates and returned to their respective temperatures. A 20% sample of resuspended cells was taken every 7 days and pelleted by centrifuged at 300*g* for 10 minutes. Cell pellets underwent RNA extraction using the Qiagen RNeasy as manufacturers instructions. RNA was quantified with ribogreen Quant-it RNA assay (ThermoFisher Scientific) and measured on a Tecan reader.

Table 1. Summary of primer sequences for *Wolbachia* 16S and *Ae. albopictus* 18S

Target	Orientation	Sequence (5'-3')	Amplicon size	Genbank Accession	Reference
Aedes albopictus 18S	Sense	CCGTGATGCCCTTAGATGTT	100	X51572	Baldridge and Fallon, (1991)
	Antisense	ATGCGCATTTAAGCGATTTC			
Wolbachia pipientis 16S	Sense	TTGCTATTAGATGAGCCTAT ATTAG	99	X61767	O'Neill et al., (1992)
	Antisense	GTGTGGCTGATCATCCTCT			

cDNA synthesis and RT-qPCR

To measure the relative infection density of *Wolbachia,* reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) was performed on the *Wolbachia* 16S and *Ae. albopictus* 18S ribosomal RNA (Table 1). RT-qPCR was chosen over qPCR as it would better represent metabolically active *Wolbachia* as qPCR may still be able to detect dead or dying *Wolbachia* genomic DNA. The Turbo DNA-free kit (Invitrogen) was applied to RNA samples prior to RT-qPCR. DNA-free RNA then underwent cDNA synthesis using Tetro cDNA synthesis kit (Bioline) as per manufacturer instructions, cDNA samples were then applied to RT-qPCR SensiMix SYBR (Bioline) and analysed on a CFX96 Real-Time PCR Detection System and Biorad CFX Manager software. The 16S:18S ratio was analysed and plotted to investigate *Wolbachia* density relative to the host.

Immunostaining

To visually confirm the infection status of *Wolbachia,* an immunofluorescent antibody was applied targeting the outer membrane protein WSPA to stress induced cells. A 30µl samples was taken from the initial starting point of week 0 and each subsequent week exposed to stress, samples were then concentrated onto frosted glass slides via cytospin at 1000 rpm. Samples were then fixed using IC fixation buffer (Invitrogen) for 30 minutes, permeabilised with 1% tween for 15 minutes, incubated with blocking buffer for 30 minutes, and incubated in PBS twice for 10 minutes per incubation. Samples were then incubated with 8µg/ml of primary antibody-WSPA for 1hr. After incubation samples were incubated in PBS twice for 10 minutes per incubation. After PBS incubation, samples were incubated with Alexaflour488 conjugate and donkey anti-rabbit IgG secondary antibody for 2hrs in darkness. After incubation with the secondary antibody, samples were again incubated in PBS twice for 10 minutes per incubation. After incubation 1 drop of vectashield hard set mounting medium was applied to the glass slide sample with DAPI, covered with a slip, and set overnight in darkness at 4°C. Slides were viewed under 63x/1.25 oil immersion fluorescent microscope.

Cappable-seq enrichment (Conducted at New England Biolabs, Massachusetts, USA)

Cell culture and stress induction

RML-12 and Aa23 cell lines were maintained as above. For temperature stress, RML-12 cells were maintained in T25 flasks (5ml culture, triplicate) at 28°C for 5 days and then transferred to incubators at their respective temperature stress (21°C, 28°C, 34°C) for either 4, 24, or 48hrs for a total of 7 days incubation before harvest (Figure 1A). Temperatures were chosen that reflect values experienced in tropical climates such as the wet season in Cairns, Australia (Richardson *et al.*, 2013) whilst avoiding temperature limits that may damage the host cell line during continuous exposure (Dobson and Rattanadechakul, 2001). Minimizing damage to the host maximised the likelihood that gene expression from *Wolbachia* is in response to the temperature stress and not the result of host cell death. Exposure times of 4, 24, and 48hrs were chosen as to examine the effects of short and long-term exposures that may be experienced

during regular and heat-wave periods. Aa23 cells were maintained in T225 (45ml culture, triplicate) flasks at 28°C then transferred to incubators (21°C, 28°C, 34°C) for 4, 24, or 48hrs before harvesting (Figure 1B). Different incubation periods were chosen as RML-12 and Aa23 displayed different *Wolbachia* growth rates (Discussed with colleague), Aa23 cells reached highest *w*AlbB titres after 5 days incubation whereas *w*MelPop-CLA titres in RML12 were relatively constant throughout incubation. Due to these differences, Aa23 cell were maintained at optimal temperature conditions for as long as possible before being processed whilst maintaining a comparable total incubation time to RML12. Once the 7-day incubation period was complete, cells were prepared for RNA extraction.

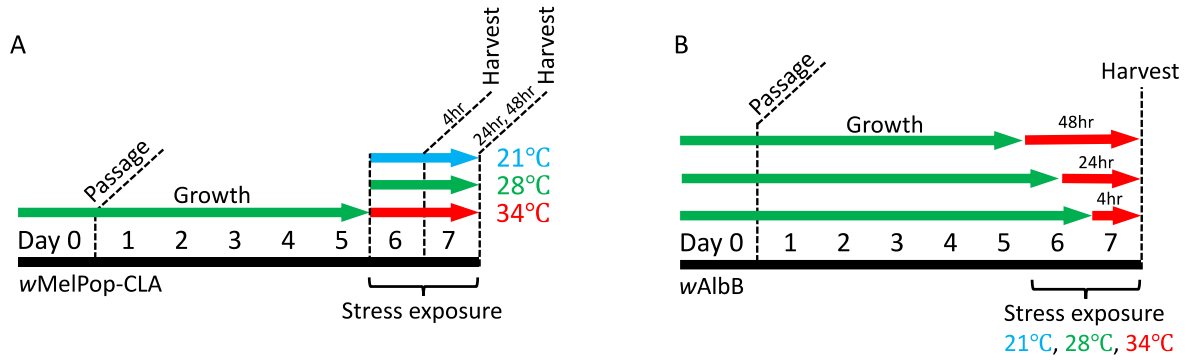


Figure 1. Schematic representation of temperature stress induction procedure. A) *w*MelPop-CLA (RML-12), B) *w*AlbB (Aa23).

Due to the presence of high density extracellular wAlbB, Aa23 sample preparation differed to RML12 as follows. Cells were processed by resuspension and transfer into 50ml falcon tubes placed directly onto ice to stall metabolic and transcriptional activity. Samples were then centrifuged at 300*g* for 10 minutes at 4°C to form a concentrated cell pellet. After removal of the supernatant, the cell pellet was then resuspended in 10ml fresh media and 5ml of 3mm glass beads were then added to each sample to aid in host cell lysis. Samples were disrupted by vortexing for 15 seconds after which they were centrifuged at 2500*g* at 4°C for 15 minutes to pellet lysed cells after which supernatant and glass beads were removed. The remaining pellet was resuspended in 1ml fresh media and filtered through a 2.7µm syringe filter to separate *Wolbachia* from host cell debris. Filtered samples were then further centrifuged at 18,000*g* at 4°C for 15 minutes after which supernatant was discarded and the remaining pellet resuspended in 1ml QIAzol (Qiagen) stored in -80°C.

After their respective incubation periods, RML-12 cells were resuspended and transferred to 15 ml falcon tubes and centrifuged at 300*g* at 4°C for 10 minutes. Cell bead lysis and filtration was not conducted for RML12 cell to prevent loss of *w*MelPop-CLA due to its low infect titre relative to *w*AlbB. Supernatant was discarded after which the cell pelleted samples were placed onto ice. Samples were then resuspended in 1ml QIAzol and stored in -80°C.

The *w*AlbB antibiotic stress induction was performed by growing Aa23 cells in T25 flasks (5ml culture, triplicate) for 6 days. Antibiotics were then added at 0.25μg/ml (Doxycycline), 0.25 μg/ml (Rifampicin), and 1 μg/ml (Moxifloxacin) for 24 hrs before harvesting cells. Cells were then harvested by resuspending cells and transferring them into 15 ml flacon tubes. Samples were centrifuged at 1000 rpm for 5 minutes, supernatant removed, and 2ml of QIAzol added to each cell pellet. QIAzol suspended pellets were then stored at -80°C. Only the *w*AlbB Aa23 cell line was treated with antibiotic exposure due to time restrains and that previous research had already been conducted on *w*MelPop-CLA gene expression to doxycycline (Darby *et al.*, 2014). Antibiotic dosages were based on physiologically relevant concentrations and those chosen by previous studies (Fenollar, Maurin and Raoult, 2003; Darby *et al.*, 2014; Johnston *et al.*, 2014)

RNA extraction

RNA was extracted via QIAzol chloroform extraction. Briefly, 200μl of 100% chloroform was added to each QIAzol suspended sample and shaken by hand for 15 seconds. Samples were then incubated at room temperature for 3 minutes after which they were centrifuged at 12,000*g* at 4°C for 15 minutes. After incubation, 500μl of the aqueous phase was transferred to DNA LoBind tubes (Eppendorf) to which 750μl of 100% alcohol was added. RNA was extracted from each samples using the Monarch total RNA miniprep kit (New England Biolabs product number T2010S) TRIzol extraction protocol with on-column DNase treatment as per manufacturer's instructions. RNA concentration was examined using NanoDrop spectrophotometer (ThermoFisher Scientific).

Cappable-seq enrichment

For the capping reaction, 30µl of 2-10µg of DNase treated RNA was added to 5µl of 10X vaccinia capping enzyme buffer (New England Biolabs product number M2080) with 5µl of 5 mM 3' -desthiobiotin-guanosine triphosphate (DTB-GTP) (New England Biolabs product number N0761), 5µl vaccinia capping enzyme (New England Biolabs product number M2080), 5µl inorganic pyrophosphate (New England Biolabs product number M2403), and ddH_2O (as needed) to a total volume of 50µl and incubated at 37°C for 1hr. After incubation samples underwent RNA clean-up using the Monarch RNA clean-up kit (New England Biolabs product number T2030S) modified to increase the washing step to a total of 5-times to ensure removal of free DTB-GTP, RNA was the transferred to a PCR tube. DTB-GTP capped RNA then underwent fragmentation by adding 1.25µl of 10X T4 polynucleotide kinase buffer and incubated at 94°C for 5 minutes, samples were then placed directly onto ice. The removal of the 3'phosphates from fragmented RNA was accomplished by adding 7µl of 10X T4 polynucleotide buffer and 3.2µl of ATP-free T4 polynucleotide kinase to the samples and incubated at 37°C for 15 minutes. Sampled then underwent RNA clean-up using the Monarch RNA clean-up kit (New England Biolabs product number T2030S) and eluted with 40µl ultrapure H_2O. Samples then underwent two rounds of enrichment with streptavidin beads by adding 30µl of DTB-GTP capped RNA to 30µl of preprepared streptavidin beads and placed on a rotation mixer at room temperature for 30 minutes. Samples were then placed onto a magnetic stand for 5 minutes or until solution became clear. Whilst on the magnetic stand, samples were washed 4 times with 200µl washing buffer (10mM Tris-HCL pH 7.5, 120mM NaCl, 1mM EDTA). Beads were resuspended in 11µl low TE elution and incubated in a thermomixer at 80°C for 10 minutes, samples were then placed back onto a magnetic stand where RNA was collected. Bead resuspension was performed twice before a second round of streptavidin enrichment was performed with 20µl of DTB-TP capped RNA and 20µl of streptavidin beads, enrichment was repeated as above. Decapping of the 5' DTB-GTP cap was performed to leave the 5' monophosphate for ligation by adding 20µl of capped RNA to 2.2µl of 10X thermopol buffer (New England Biolabs product number B9004), 2µl RppH (New England Biolabs product number 0356S) and then incubated at 37°C for 1 hr. After incubation,

1µl of 1:10 diluted proteinase K was added to solution and incubated at 37°C for 10 minutes. RNA solution then underwent a final round of RNA clean-up using (New England Biolabs product number T2030S) before library preparation.

Library preparation and sequencing

Enriched RNA underwent library preparation by using the NEBNext multiplex small RNA library prep kit (New England Biolabs product number E7300S) as per manufacturers instructions. After library completion, samples underwent PCR-clean up by adding 1x vol ampure beads followed by incubation at room temperature for 10 minutes. Samples were placed onto a magnetic stand until solution became clear. Supernatant was removed and beads were washed with 80% ethanol for a total of four time. Once washed, beads were allowed to air dry. Before beads showed signs of cracking, beads were resuspended in 40µl low TE followed by incubation at room temperature for 10 minutes. After incubation, beads were again placed onto a magnetic stand until solution turned clear at which point the supernatant was collected. Resulting library was then diluted 1:10 and analysed on the high sensitivity DNA bioanalyzer using a DNA 1000 chip according to manufacturers instructions. Samples were then sent for NEXTSeq high output single-end 150bp sequencing.

Bioinformatics: Mapping

Adapter trimming and contaminant filtering was performed using BBduk from the BBtools software package (version 38.86). Unpaired single-end reads were mapped using Bowtie2 (version 2.4.5) with the local alignment option. Reads were mapped simultaneously onto a concatenated fasta file containing the genomes of *Ae. albopictus* mitochondria (Accession NC_006817), *Ae. albopictus* (Assembly GCA_001876365.2), and *w*Mel (Assembly GCA_000008025.1) or *w*AlbB (Assembly GCA_004171285.1). A minimal sequencing depth was chosen based on the recommended minimal sequencing depth of 2 million for differential gene expression in *E. coli* (Haas *et al.*, 2012). The 2 million *E. coli* minimal depth translates to 0.43 million reads per Mb of genome length, this translated to a minimal desired library depth of 0.54M and 0.63M reads for *w*MelPop-CLA and *w*AlbB respectively. The *w*MelPop-CLA genome is highly similar to *w*Mel, besides the known differences (Woolfit *et al.*, 2013), allowing a *w*MelPop-CLA to be a suitable model proxy for *w*Mel gene expression. Aligned reads that mapped to *Wolbachia* were further filtered for those with a unique mapping score of

mapq20 using samtools (version 1.11) to ensure reads that associated to repetitive regions were confidently assigned. Conservatively removing non-uniquely mapped reads may limit additional information from repeat elements, such as transposons, however their removal will give greater confidence in expression of TSS that remain associated to repeat elements. Reads were then converted into count values using the software HTSeq-count (version 0.11.1).

Bioinformatics: TSS Calling

TSS calling of mapped reads were conducted using the TSS calling software pipleine (Accessible https://github.com/Ettwiller/TSS) using a clustering filter of a 10-nucleotide window. Designating a TSS type was performed using a custom python script implementing pandas and Bedtools closest (v2.29.2). TSS types were assigned based on location of the TSS relative to closest neighbouring genes by the following: pTSS with the highest read count within 300 nucleotides upstream to the start codon, gTSS within 300 nucleotides upstream to the start codon, iTSS internal to coding sequence, asTSS antisense to coding sequence and/or 100 nucleotides downstream to the antisense stop codon, and oTSS if none of the previously mentioned rules applied. High confidence TSS were determined by only selecting those TSS which were present in all replicates within a specific condition. TSS that were absent in a condition but present in others were given a prior count of 10. To apply the same minimal expression value to all samples, a minimum count threshold for a TSS to be called was based on a value of 10 counts per million (CPM) based on the smallest sample library for temperature and antibiotic experiments separately. TSS statistics and graphs were generated via the python library packages matplotlib and seaborn. Heatmaps were generated using heatmap.2 on the R package ggplot2 (version 3.3.2).

Bioinformatics: Orthologous gene comparisons

To find orthologous genes between *w*Mel and *w*AlbB, the amino acids sequences of all protein coding genes were compared via the OrthoFinder package (Version 2.5.4). As a conservative approach to ensure the functional similarity between protein coding genes, only genes whose

protein products shared 1-to-1 homology between *w*Mel and *w*AlbB were considered for further analysis.

Bioinformatics: Motif discovery

To find sequence motifs indicative of promoter regions, the upstream 50-nucleotide regions of all TSS were used as input into the web browser-based MEME suite (Version 5.4.1) with up to 10 motifs patterns selected in the settings. Motifs were visualised using software package WebLogo (Version 2.8.2).

5'RACE

To confirm the single-nucleotide accuracy of Cappable-seq, 5'RACE analysis was conducted on the highly expressed 30S Ribosomal gene of *w*AlbB. The study used the NEB 5'RACE protocol using the template switching RT enzyme mix (New England Biolabs product number M0466) and cloned into a single-use stable competent *E. coli* using the NEB PCR cloning kit (New England Biolabs product number E1203S) as per manufactures instructions.

Table 2. list of 5'RACE inner and outer primers for *w*AlbB *rpsM* (30S Ribosome subunit)

Gene ID	Gene	Strain	Primer sequence
DEJ70_RS04995	*rpsM*	*w* AlbB	Inner: TTTCTGAGCTCACCCTCTATGACA
DEJ70_RS04995	*rpsM*	*w* AlbB	Outer: AGCATTAGTATGGGTTCTTTGTCCCC

Results

Recapitulating the effects of heat and doxycycline in a cell culture model

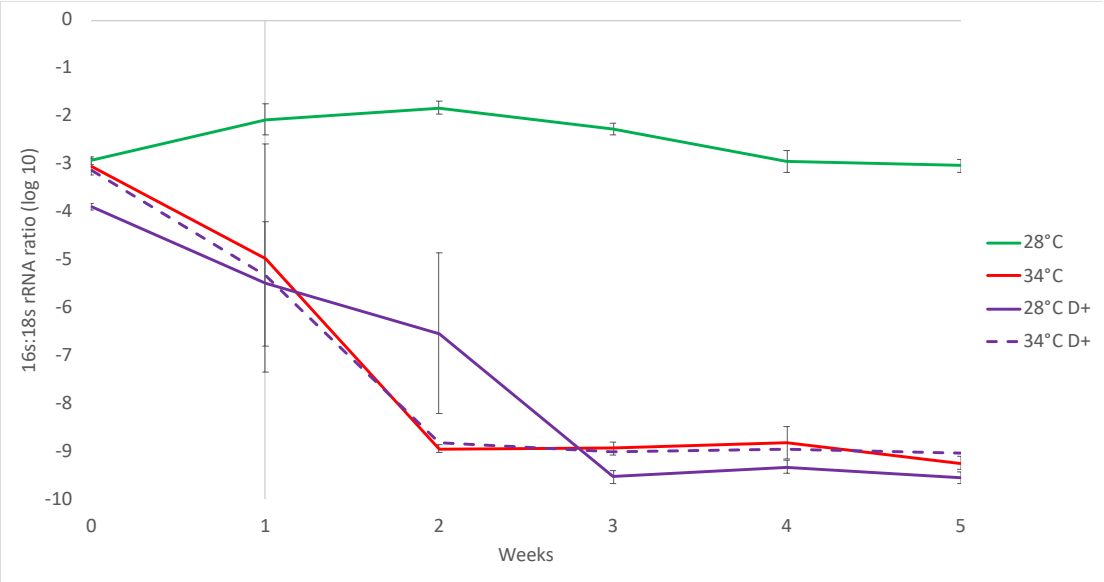


Figure 2. Preliminary effects of temperature and doxycycline on *w*MelPop-CLA density. RT-qPCR 16S:18S rRNA ratios (log10) over a 5-week exposure period. Each sample was repeated in triplicate. Each datapoint represents the average of the replicates. Error bars represent standard deviation of the mean. Doxycycline positive (D+).

Preliminary experiments on *w*MelPop-CLA subjected to 34°C alone and in combination with doxycycline was conducted prior to capable-seq experiments to determine if the *w*MelPop-CLA cell line can recapitulate thermal susceptibility seen in previous studies. Exposure to 34°C alone resulted in reduced infection density with *w*MelPop-CLA becoming undetectable after 2 weeks continual exposure. Heat stress at 34°C in combination with doxycycline treatment did not differ significantly from 34°C with *w*MelPop-CLA becoming undetectable after 2 weeks continual exposure. Doxycycline exposure resulted in gradual reduction in infection density with *w*MelPop-CLA becoming undetectable after 3 weeks continual exposure (Figure 2). Immunostaining of the WSPA protein on preliminary temperature stress *w*MelPop-CLA visually confirmed the presence of *w*MelPop-CLA at 28°C and its absence after 2 weeks of continual 34°C stress induction (Figure 3).

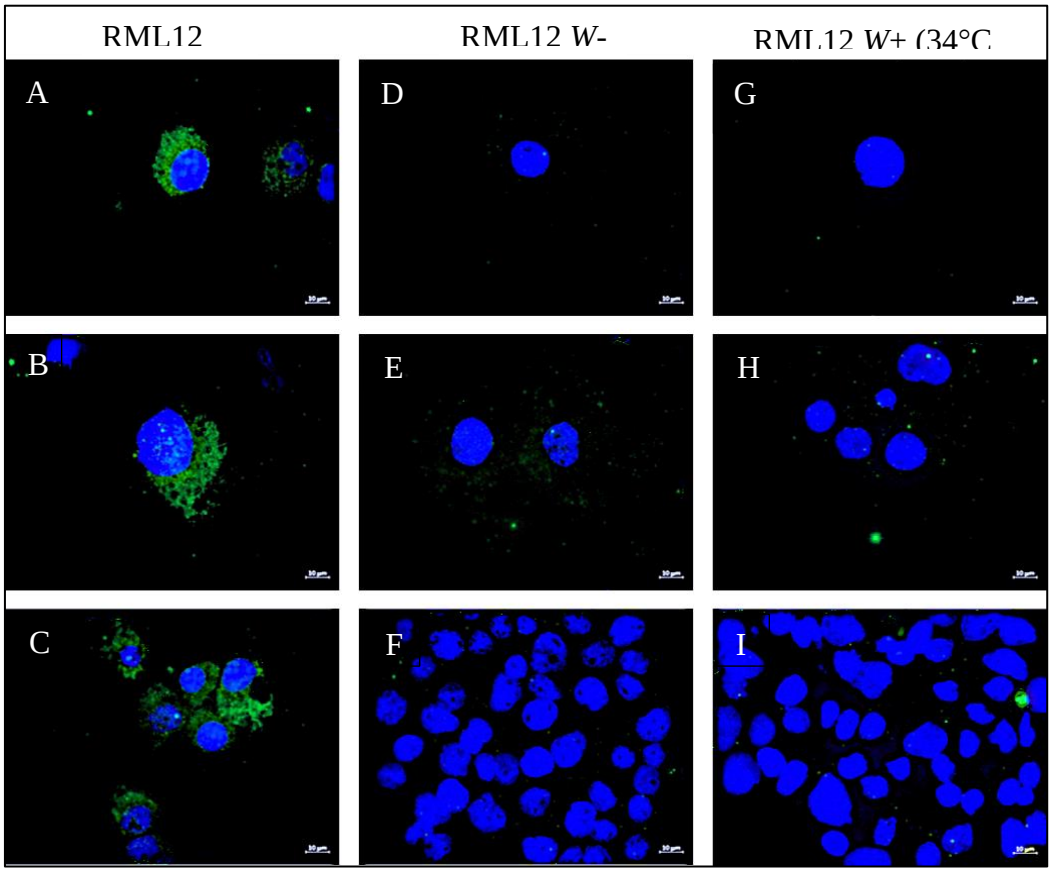


Figure 3. Visualization of thermally depleted *w*MelPop-CLA. Immunostaining of RML-12 cells. A-C: RML-12 (w^+) 28°C at week 0, D-F: RML-12 (w^+) negative control 28°C, G-I: RML-12 (w^+) 34°C heat treated after 3 weeks. *Wolbachia* targeted WSPA antibody (Green) detected by Alexaflour488 conjugate, DAPI stained host nucleus (Blue). Taken at 63x magnification in oil immersion. Scale bar represents 10 μm.

Enrichment of *Wolbachia* primary transcripts

To confirm the targeted enrichment of *Wolbachia* primary transcripts, a comparative analysis of reads mapping to primary transcripts relative to processed tRNA and rRNA were compared between unenriched and enriched libraries (Figure 4). Processed tRNA and rRNA constituted most transcripts under unenriched samples (88.45% *w*MelPop-CLA, 89.81% *w*AlbB). With the implementation of cappable-seq both strains showed significant increases in the proportions of *Wolbachia* primary transcripts compared to unenriched control. Total percentage composition of *Wolbachia* primary transcripts amongst overabundant host RNA increased over 94-fold (0.02% to 2.52%) and over 14-fold (0.94% to 13.53%) for *w*MelPop-CLA and *w*AlbB respectively (Figure 4). Simultaneously, reads mapping to host mitochondria were also enriched (0.18% to 10.24% for *w*MelPop-CLA and 0.05% to 0.12% for *w*AlbB). Amongst the *Wolbachia* specific reads, the percentage of *Wolbachia* specific primary-RNA amongst total *Wolbachia* RNA increased 8.3-fold (11.55% to 96.28%) for *w*MelPop-CLA and 9.0-fold (10.19% to 92.40%) for *w*AlbB.

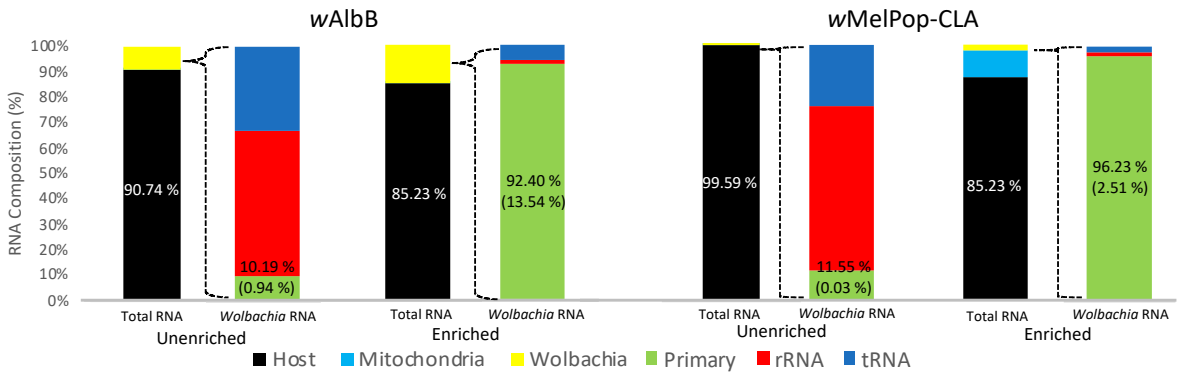


Figure 4. Comparison of Cappable-seq enrichment efficacy of *Wolbachia* primary transcripts from total RNA. Total RNA extracted from *w*MelPop-CLA and *w*AlbB infecting *A. albopictus* RML12 and Aa23 cell lines respectively at 28°C were divided and subjected to Cappable-seq without (Unenriched) and with (Enriched) capable-seq. Bar plot shows the percentage of transcripts that mapped to host (black), mitochondria (light blue), total *Wolbachia* (yellow). Expanded regions show percentage of *Wolbachia* specific transcripts that map to tRNA (dark blue), rRNA (red) and primary transcripts (green), bracketed values represent percentage of *Wolbachia* primary transcripts relative to total RNA.

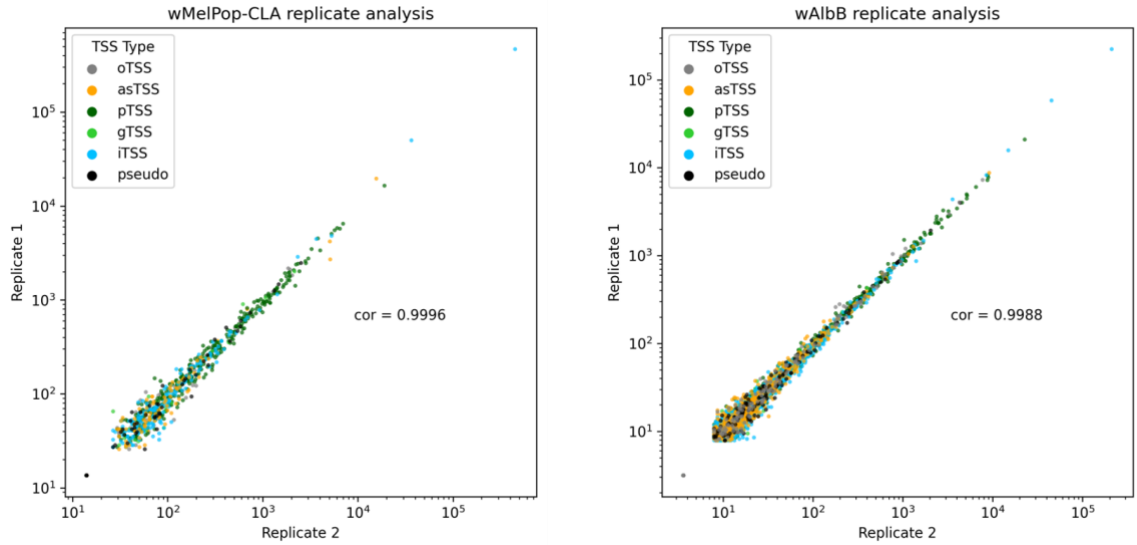


Figure 5. Cappable-seq expression between replicates. Scatter graphs of TSS expression between 28°C replicates of A) *w*MelPop-CLA and B) *w*AlbB. Each dot represents the CPM of each TSS coloured by its designated TSS type. cor = correlation coefficient.

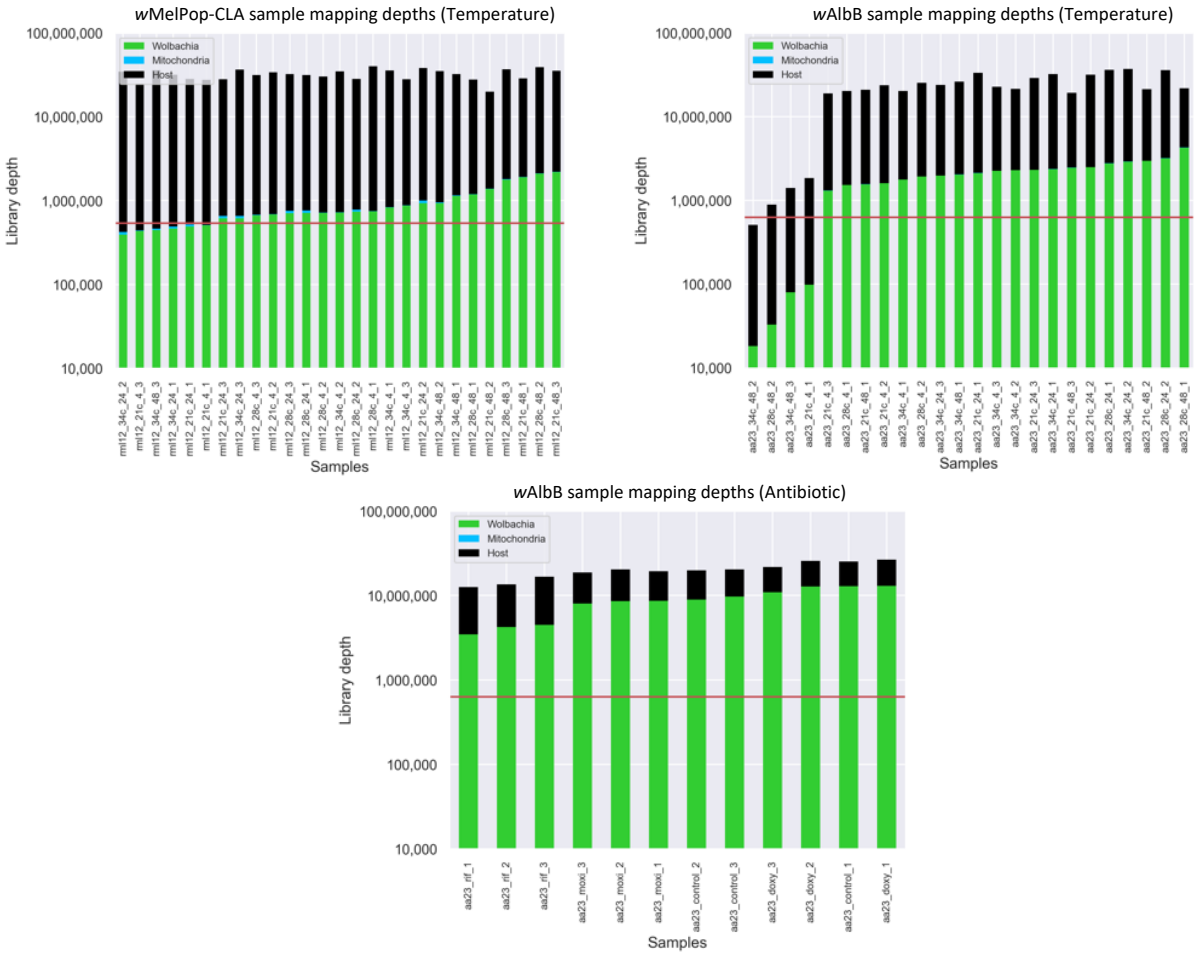


Figure 6. Summary of mapped reads across conditions and their mapped target. Bar chart representing reads depths of mapped reads for each sample. Red horizontal line represents the suggested minimum read depth required for confident differential expression analysis.

Table 3. Summary of mapped *Wolbachia* reads. Sequencing summary of *Wolbachia* mapped reads used throughout the study from *w*MelPop-CLA and *w*AlbB genomes.

Strain	Total mapped reads (mapq20)	Range	Average	Total Samples	Total TSS
***w* MelPop-CLA**	24,872,374	392,459 - 2,189,964	921,199	27	966
***w* AlbB**	151,846,942	18,210 - 13,011,712	4,217,971	36	4,318

Identification and categorization of *Wolbachia* transcriptional start sites

In total, 24,872,374 and 151,846,942 single-end reads for primary *Wolbachia* transcripts were uniquely mapped to *w*MelPop-CLA and *w*AlbB respectively (Table 3). A total of 966 and 4,318 TSS for *w*MelPop-CLA and *w*AlbB respectively were found within all conditions tested (Figure 8). The most abundant TSS type for *w*MelPop-CLA were pTSS (42.1%) whereas for *w*AlbB iTSS were predominant (45.6%) while the least abundant TSS type for *w*MelPop-CLA and *w*AlbB were gTSS (3.2% and 5.4% respectively). For *w*MelPop-CLA and *w*AlbB annotated genes (1,286 and 1,418 respectively), approximately 29.63% (407) and 40.62%

(682) harboured a canonical pTSS. *w*AlbB contained a greater proportion intragenic TSS (iTSS and asTSS) with over half of all CDS's containing an iTSS (53.74%) and over a third containing an asTSS (39.56%) compared to *w*MelPop-CLA (iTSS 16.10%, asTSS 10.19%). Template switching 5'RACE was performed to validate TSS via an alternate method which confirmed the pTSS of the 30S ribosomal protein to single-nucleotide resolution (Figure 7, clones 2-3), minor pTSS present before clustering were also confirmed (Figure 7, clone 1).

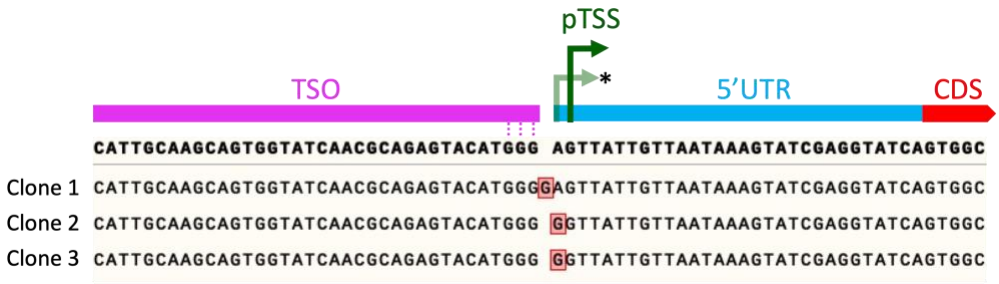


Figure 7. TSS validation with 5'RACE. The pTSS of the *w*AlbB 30S ribosomal gene was validated using a template switching 5'RACE protocol. Segments shown are the template switching oligonucleotide (TSO) sequence with 3' poly-G tail (Pink) followed by the pTSS discovered by Cappable-seq (Dark green), 5'UTR (Blue) and the initial 5 nucleotides of the 30S ribosomal coding sequence. * represents an adjacent minor pTSS one nucleotide away. Three 5'RACE sequences derived from 3 separate clones are displayed, highlighted guanosine (G) nucleotide represents known non-template nucleotide additions.

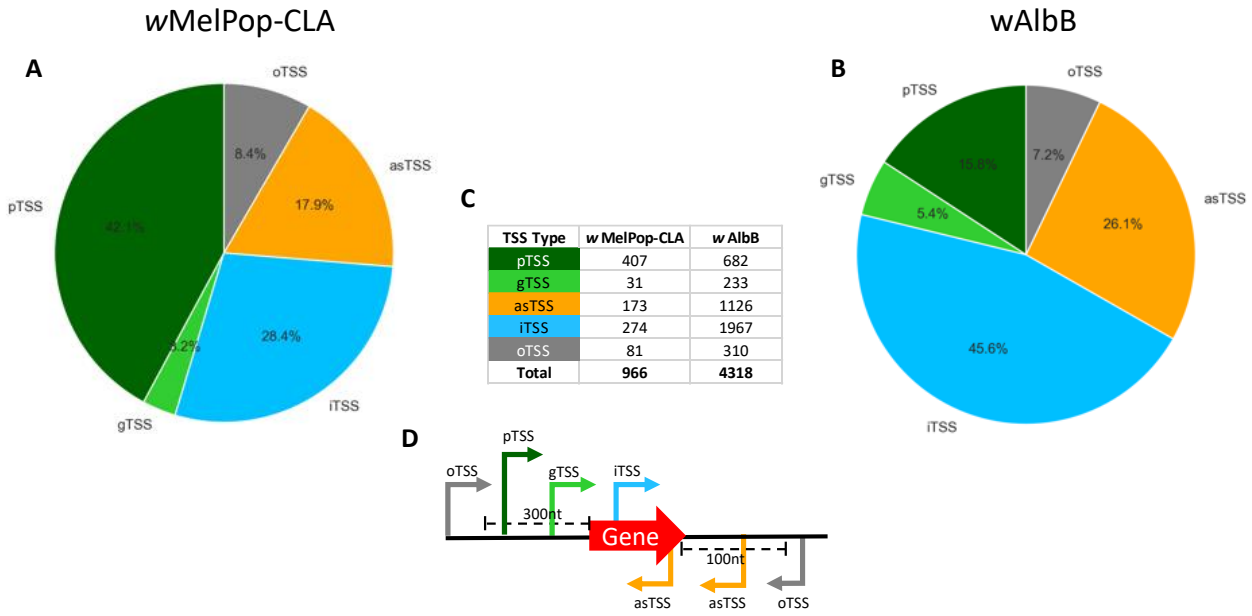


Figure 8. Comparative summary of total TSS type diversity found in *w*MelPop-CLA and *w*AlbB overall tested conditions. A-B) Pie charts representing percentage abundance of each TSS type in wMelPop-CLA and wAlbB genomes respectively, C) Table of total TSS types discovered, and D) visual schematic of each TSS type. pTSS (dark green), gTSS (light green), asTSS (orange), iTSS (light blue) and oTSS (grey).

Table 4. Percentage of coding sequences containing each TSS type. Summary of CDSs that contained each TSS type, oTSS were excluded due to the low probability of association with a CDS.

	CDS	pTSS	gTSS	asTSS	iTSS
w **MelPop-CLA**	1,286	381	30	131	207
		29.63%	2.20%	10.19%	16.10%
w **AlbB**	1,418	576	149	561	762
		40.62%	10.51%	39.56%	53.74%

The number of TSS expressed per condition ranged from 244-764 and 1,174-2,702 under temperature stress for *w*MelPop-CLA and *w*AlbB respectively, and 1,497-3,180 for *w*AlbB under antibiotic stress. The majority of expressed TSS were shared between all conditions for both *w*MelPop-CLA (62%, n=598) and wAlbB (44%, n=1913) between temperatures (Figure 9), this was also apparent for *w*AlbB exposed to antibiotic stress (59%, n=1913). For both *w*MelPop-CLA and *w*AlbB respectively, TSS expression under specific conditions were observed (Figure 9. A-B) for temperatures 21°C (13%, n=124 and 5%, n=151), 28°C (5%, n=45 and 9%, n=297) and 34°C (4%, n=34 and 7%, n=233). Condition specific TSS were also observed for *w*AlbB between antibiotics doxycycline (21%, n=722), rifampicin (2%, n=70) and moxifloxacin (5%, n=184). In *w*AlbB TSS were also observed specific to temperature stress (12%, n=501) and antibiotic stress (12%, n=503) (Figure 9. C-D).

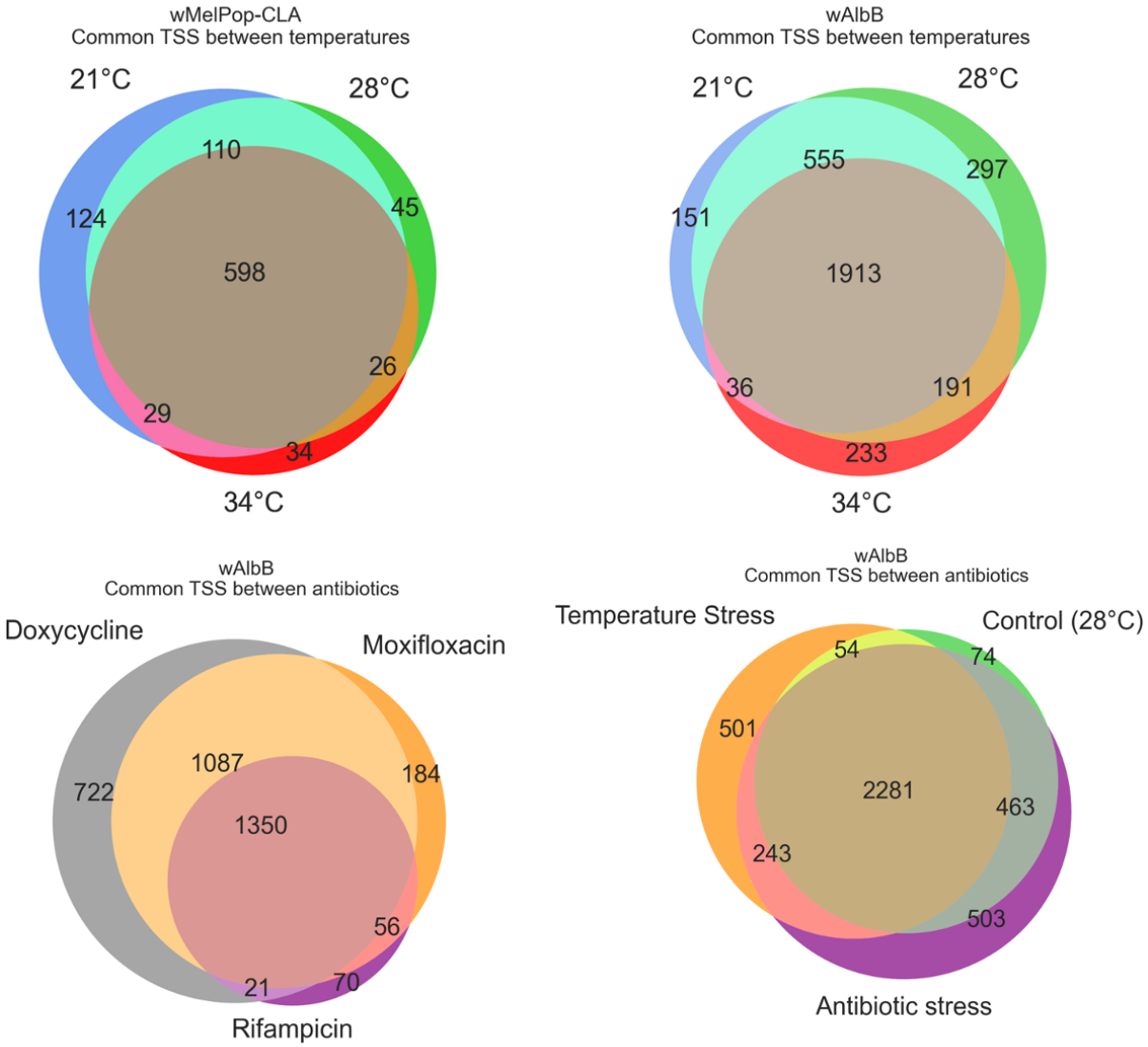


Figure 9. Commonality of TSS between conditions. Venn diagrams show common TSS between stress conditions. A) *w*Mel common TSS between temperatures, B) *w*AlbB common TSS between temperatures, C) *w*AlbB common TSS between antibiotics, D) *w*AlbB common TSS between temperature and antibiotic stressors.

At baseline conditions (28°C) pTSS were the most highly expressed TSS type in both wMelPop-CLA and wAlbB with a median expression of 111.0 and 134.5 CPM respectively (Figure 10. A-B). While intragenic TSS were observed as highly abundant, the median expression level for both iTSS and asTSS were ranked amongst the lowest for both *w*MelPop-CLA and *w*AlbB (Figure 10. A-B). The MEME motif discovery suite was used to detect the presence of promotor motifs, two distinct motifs were recognized displaying the common AT rich -10 TATAAT box and to a lesser extent the -35 TTGACA core promoter consensus sequence of housekeeping sigma-70 observed in *E. coli* (Figure 10. C-D).

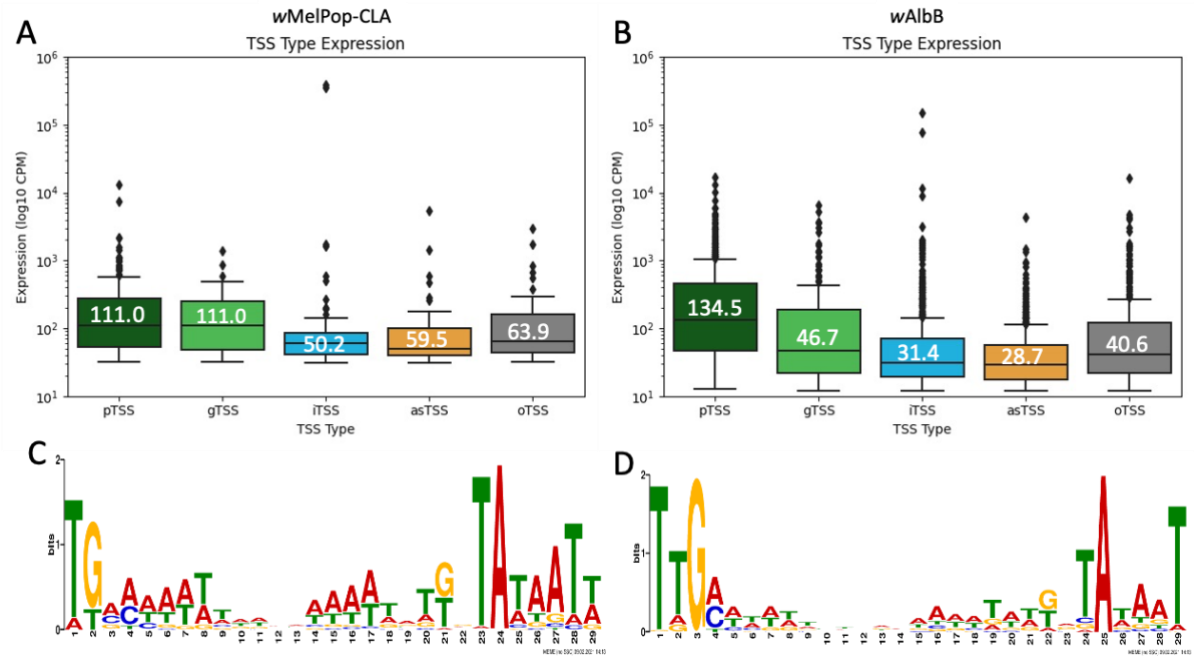


Figure 10. Summary of TSS characteristics for both *w*MelPop-cla and *w*AlbB at baseline (28°C). A-B) Expression (CPM) of each TSS type for wMelPop-CLA and wAlbB respectively, shown values represent median CPM. C-D) Sequence logo motifs of the 50-nucleotide upstream region all TSS identified via MEME discovery suite.

Intragenic TSS (asTSS and iTSS) represented a substantial portion of all detected TSS for both *w*MelPop-CLA (46.3%) and *w*AlbB (71.7%). The location distribution of both iTSS and asTSS were analysed to detect positional bias along associated genes (Figure 11. A-D). Both *w*MelPop-CLA and *w*AlbB iTSS displayed a slight bias towards the 3' and 5' end of genes, asTSS also showed a slight 3' and 5' end although to a lesser extent. With regards to leaderless transcripts, 2.5% (25) and 0.16% (7) of TSS from *w*MelPop-CLA and *w*AlbB respectively were located precisely on the first nucleotide of the associated gene. When extending the leaderless window by 5 nucleotides upstream and downstream to the first nucleotide of the CDS the total leaderless TSS increased to 3% (31) and 1.5% (66) for *w*MelPop-CLA and *w*AlbB respectively. The length of the 5' untranslated region (5'UTR) between the TSS and the first nucleotide of the CDS was determined for pTSS and gTSS for both *Wolbachia* strains (Figure 11. E-F). The median lengths for the 5'UTR for both pTSS and gTSS were 24/21.5 nucleotides for *w*MelPop-CLA and 23/60 nucleotides for *w*AlbB under all conditions tested.

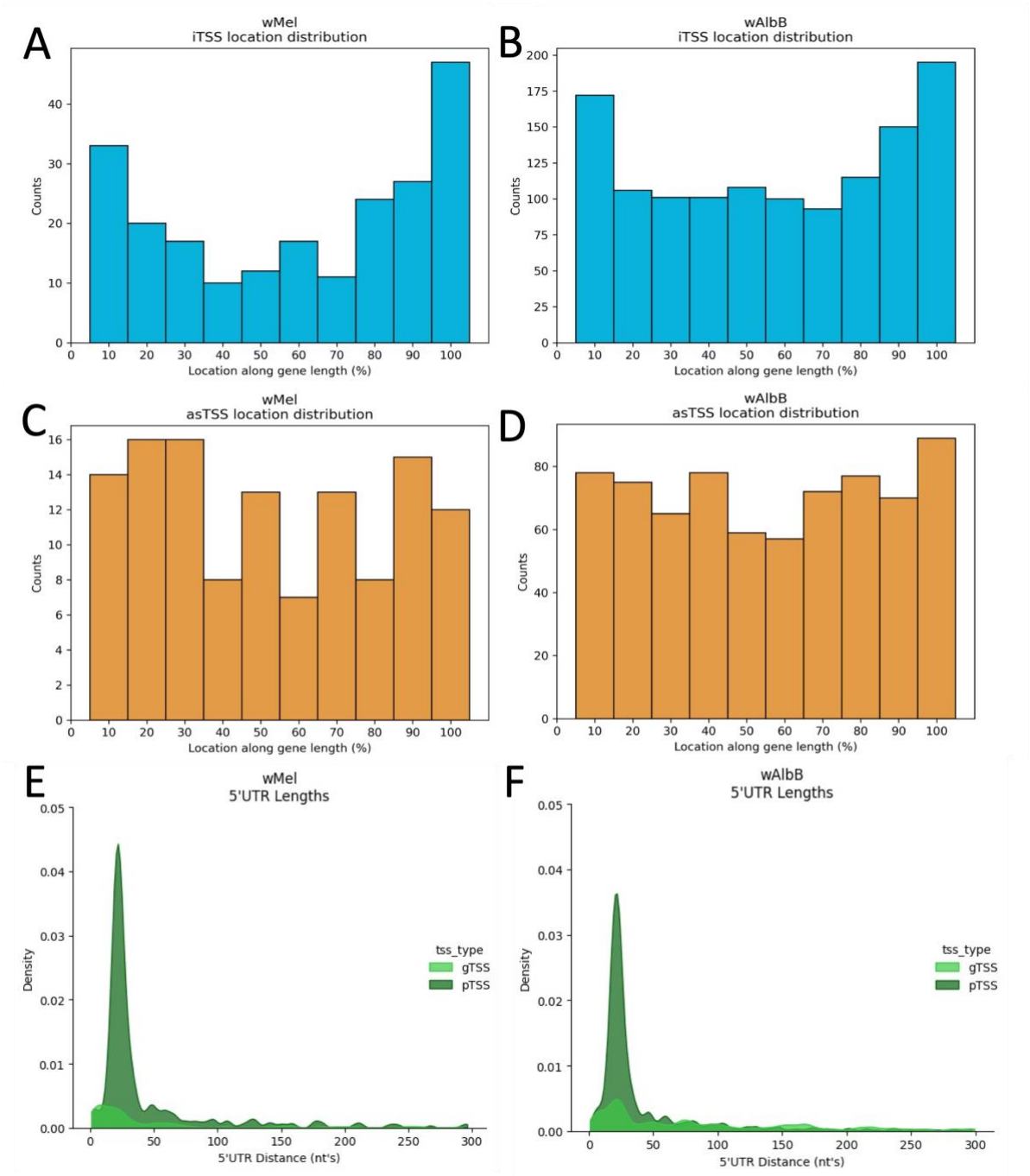


Figure 11. Summary of TSS locations relative to associated gene. A-B) Distribution represented as a percentage of the gene length of iTSS (blue) locations within associated genes, C-D) Distribution of asTSS (orange) within associated genes, E-F) distribution of 5'UTR lengths amongst pTSS (dark green) and gTSS (light green).

w MelPop-CLA				wAlbB			
CPM	Product	Locus	TSS Type	CPM	Product	Locus	TSS Type
4.20E+05	6S RNA	WD_RS06775	iTSS	2.63E+05	6S RNA	DEJ70_RS02130	iTSS
2.93E+05	fructose-6-phosphate aldolase	WD_RS02455	iTSS	1.15E+05	fructose-6-phosphate aldolase	DEJ70_RS05135	iTSS
8.41E+03	hypothetical protein	WD_RS02845	pTSS	1.87E+04	23S ribosomal RNA	DEJ70_RS05220	iTSS
8.12E+03	phage tail protein	WD_RS01240	pTSS	1.81E+04	signal recognition particle sRNA small type	DEJ70_RS02165	iTSS
5.39E+03	DNA mismatch repair endonuclease MutL	WD_RS05920	asTSS	1.75E+04	16S ribosomal RNA	DEJ70_RS05685	pTSS
3.58E+03	signal recognition particle sRNA small type	WD_RS06780	iTSS	1.52E+04	hypothetical protein	DEJ70_RS01070	pTSS
2.45E+03	succinate dehydrogenase assembly factor 2	WD_RS04920	oTSS	1.23E+04	preprotein translocase subunit YajC	DEJ70_RS04650	pTSS
2.25E+03	23S ribosomal RNA	WD_RS00885	pTSS	1.14E+04	hypothetical protein	DEJ70_RS01345	pTSS
2.08E+03	16S ribosomal RNA	WD_RS05540	pTSS	8.79E+03	DNA mismatch repair endonuclease MutL	DEJ70_RS05750	asTSS
2.04E+03	thioredoxin family protein	WD_RS05680	pTSS	6.92E+03	single-stranded-DNA-specific exonuclease RecJ	DEJ70_RS01320	oTSS
1.95E+03	virulence RhuM family protein	WD_RS01225	asTSS	6.32E+03	tRNA-Phe	DEJ70_RS03590	pTSS
1.79E+03	adenylosuccinate lyase	WD_RS03555	iTSS	6.00E+03	tRNA-Leu	DEJ70_RS01720	pTSS
1.76E+03	Bax inhibitor-1/YccA family protein	WD_RS04290	pTSS	5.71E+03	hypothetical protein	DEJ70_RS04380	pseudo
1.64E+03	hypothetical protein	WD_RS04725	pseudo	5.52E+03	Rne/Rng family ribonuclease	DEJ70_RS03135	pTSS
1.57E+03	hypothetical protein	WD_RS00565	pTSS	5.26E+03	ankyrin repeat domain-containing protein	DEJ70_RS06440	pTSS
1.56E+03	ankyrin repeat domain-containing protein	WD_RS01925	pTSS	4.91E+03	tRNA-Arg	DEJ70_RS04270	pTSS
1.23E+03	aspartate-semialdehyde dehydrogenase	WD_RS04305	oTSS	4.91E+03	hypothetical protein	DEJ70_RS01670	pTSS
1.17E+03	hypothetical protein	WD_RS02135	pTSS	3.88E+03	IS66 family transposase	DEJ70_RS06135	oTSS
1.10E+03	TrbC/VirB2 family protein	WD_RS02935	pTSS	3.27E+03	hypothetical protein	DEJ70_RS01180	pTSS
1.02E+03	hypothetical protein	WD_RS00560	pTSS	3.08E+03	50S ribosomal protein L28	DEJ70_RS04330	pTSS

Figure 12. Top 20 expressed TSS at baseline 28°C. CPM represents the average expression of 28°C replicates

Of the top 20 most highly expressed TSS the majority of TSS types were observed (excluding gTSS), 60% consisted of pTSS for *w*MelPop-CLA and 65% *w*AlbB (Figure 12). The two most highly expressed TSS for both *w*MelPop-CLA and *w*AlbB were iTSS associated with 6S RNA and fructose-6-phospho aldolase (F6PA). Using a single baseline sample (28°C), the percentage of total reads these two most highly expressed TSS represented for *w*MelPop-CLA and *w*AlbB was 38.3% and 8.7% for 6S RNA, and 33.9% and 8.7% for F6PA, respectively. Hypothetical proteins were also highly expressed in both strains observed in 25% of the top 20 expressed TSS for both *w*MelPop-CLA and *w*AlbB. Within *w*MelPop-CLA a pTSS was present in the top 20 list associated to a phage tail (WD_RS01240). An asTSS associated to the mismatch repair protein MutL was present within the top 10 most highly expressed TSS for both *w*MelPop-CLA and *w*AlbB.

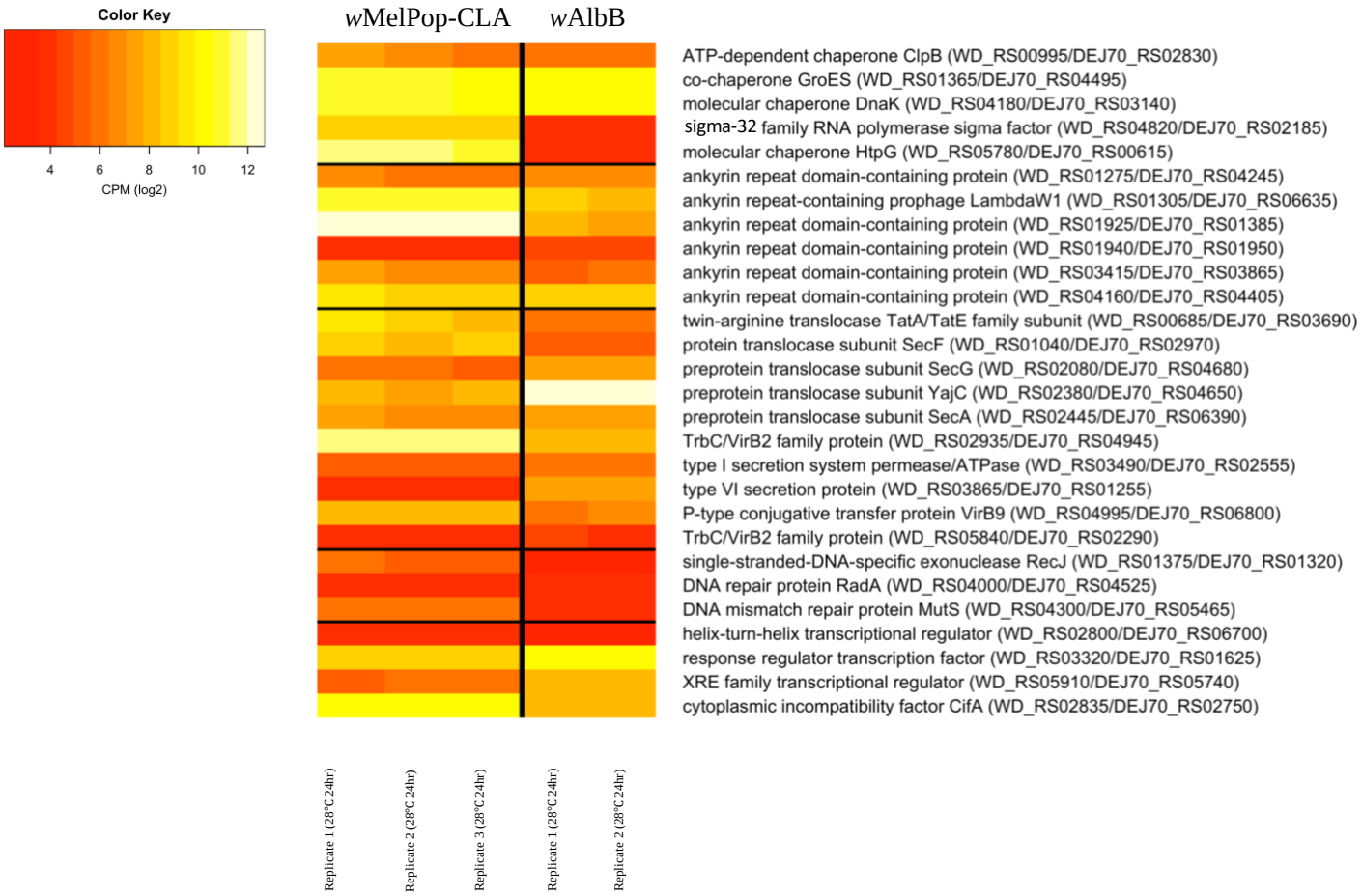


Figure 13. Heatmap of select orthologous pTSS at baseline 28°C. Expression (Log2 CPM) of orthologous genes with robustly assigned pTSS. Locus tags are derived from *w*MelPop-CLA and *w*AlbB respectively.

Table 5. Comparison of TSS studies. Summary of select TSS studies comparing methodology and TSS quantification in relation to genome size. Total mapped reads only include libraries treated with either TEX (terminator exonuclease) or TAP (tobacco acid pyrophosphatase), mapped reads do not include control libraries.

Organism	Genome Size (Mb)	Protein Coding Genes	Total TSS	TSS per Kb	Mapped Reads (M)	Method	Study
Bacteroides thetaiotaomicron	6.36	4,660	4,507	7.1	30.97	dRNA-seq	(Ryan *et al.*, 2020)
Clostridium phytofermentans	4.8	3,926	9,457	19.7	23.37	Cap-seq	(Boutard *et al.*, 2016)
Salmonella enterica	4.8	4,446	1,873	3.9	3.62	dRNA-seq	(Ramachandran, Shearer and Thompson, 2014)
Escherichia coli	4.64	4,228	14,868	32.0	43.09	dRNA-seq	(Förstner *et al.*, 2014)
Escherichia coli	4.64	4,242	16,539	35.6	20.00	cap-seq	(Ettwiller *et al.*, 2016)
Leptospira interrogans	4.61	3,341	2,866	6.2	2.02	dRNA-seq	(Zhukova *et al.*, 2017)
Campylobacter jejuni	1.69	1,572	2,167	12.8	2.58	dRNA-seq	(Dugar *et al.*, 2013)
Helicobacter pylori	1.65	1,468	1,907	11.6	1.38	dRNA-seq	(Sharma *et al.*, 2010)
***Wolbachia* (*w* AlbB)**	1.48	1,202	4,318	29.2	24.90	Cap-seq	Present
Borrelia burgdorferi	1.29	1,212	6,042	46.8	6.18	5'RNA-Seq	(Adams *et al.*, 2017)
***Wolbachia* (*w* MelPop-CLA)**	1.27	1,144	966	7.6	151.90	Cap-seq	Present
Candidatus Phytoplasma asteris	0.85	752	231	2.7	0.14	dRNA-seq	(Nijo *et al.*, 2017)
Chlamydia trachomatis	1.04	887	363	3.5	0.31	dRNA-seq	(Albrecht et al., 2009)
Chlamydia pneumonia	1.23	1,053	565	4.6	0.85	dRNA-seq	(Albrecht *et al.*, 2011)
Mycobacterium marinum	6.32	5,136	6,808	10.8	1.00	Cap-seq	(Huang et al., 2021)

Discussion

Wolbachia displays full TSS diversity and an unprecedented number of globally mapped TSS of any uncultivable bacterium

One of the goals of this study was to detect the global positioning of transcriptional start sites over a range of stress related conditions. A major obstacle in studying transcriptional profiles in uncultivable intracellular pathogens (UIP) is the ability to separate bacterial RNA from the host within which it resides. The most widely used method to increase the proportion of primary RNA from a host-bacterium RNA mixture is differential RNA-seq (dRNA-seq). The dRNA-seq method relies on the statistical comparison of a divided RNA mixture; one sub-sample is treated with terminator exonuclease (TEX+) that degrades processed RNA (5'-monophosphate), thus enriching primary RNA (5'-triphosphate), whilst the other sub-sample remains untreated. Studies which have attempted to map global TSS within UIPs have achieved modest success with *Chlamydia pneumonia*, *Chlamydia trachomatis* and *Candidatus* Phytoplasma asteris, revealing 565, 363 and 231 TSS respectively (Albrecht *et al.*, 2011, Albrecht *et al.*, 2009 and Nijo *et al.*, 2017). The present study succesfully detected 966 and 4,318 TSS for *w*MelPop-CLA and *w*AlbB respectfully; this represents the most TSS detected in any UIP to date.

The apparent scarcity of TSS in previously studied UIPs may partly be attributed to the high percentage of reads mapping to tRNA and rRNA (over 96% for *Ca.* P. asteris and over 50% for *C. trachomatis*) in enriched libraries. In contrast, the proportion of tRNA and rRNA in *w*MelPop-CLA and *w*AlbB enriched libraries were 3.72% and 7.59% respectively. The explanation for such a discrepancy in tRNA and rRNA proportions in Cappable-seq relative to dRNA-seq may be attributable to both the physical removal of 5'PPP transcripts from the RNA mixture and the known incomplete enzymatic degradation of processed transcripts by TEX (Schlüter *et al.*, 2013), these factors may result in greater isolation of primary transcripts when Cappable-seq is applied. Sequencing depths may also contribute greatly to the number of TSS detected; the number of mapped reads in the *Candidatus* Phytoplasma asteris, *Chlamydia trachomatis* and *Chlamydia pneumonia* studies were 0.14, 0.31 and 0.85 million, respectively, whereas the present study obtained 24.87 and 151.85 million for *w*MelPop-CLA and *w*AlbB, respectively, enabling greater detection of low expression TSS.

Of the limited number of bacteria whose transcriptional landscape has been mapped, *Borrelia burgdorferi* shares traits comparable with *Wolbachia* such as biological development within arthropods, a small genome (<1.5 Mb) and a gram negative-like outer membrane (OM) lacking lipopolysaccharides. In the study conducted by Adams, et al. (2017), a total of 6,042 TSS were detected across the *B. burgdorferi* genome, which when normalized to genome size, represents a TSS density of 46.8 TSS per Kb. This is not only higher than *w*MelPop-CLA and *w*AlbB (7.6 and 29.2 TSS per Kb respectively), but higher than reported in any other TSS study to date (Table 5). When comparing TSS studies, the highest correlation was found between total mapped reads compared TSS density. Unsurprisingly, an increase in total TSS relative to genome size and number of protein coding genes can be expected given the nature of TSS; however, an increase in TSS density appear to be related to read depth as more TSS pass the minimum expression threshold as sequencing depths increase particularly so for lowly expressed iTSS. The total TSS discovered for *w*MelPop-CLA and *w*AlbB do not lie outside the range of TSS discovered for other bacteria whose TSS have been globally mapped. The use of triplicate sampling, RRS thresholds and a conservative TSS consensus approach gives the TSS uncovered in this study a high degree of confidence.

As the present study acquired up to 25-fold greater sequencing depth in comparison to the *Borrelia* study, the discrepancy in detected TSS between *Wolbachia* and *Borrelia* cannot be explained by sequencing depth alone. The total number of protein coding genes for *w*MelPop-CLA, *w*AlbB and *Borrelia* are highly similar; thus, the explanation for such a dense TSS allocation in the latter is not due to a greater number of protein coding genes. The discrepancy may be due in part to the dynamic lifestyle of *Borrelia* compared to *Wolbachia*, as the *Borrelia* lifecycle involves the colonization of both the arthropod vector and the mammalian reservoir, whereas arthropod associated *Wolbachia* reside strictly within the arthropod host. Thus, in *Borrelia*, complex regulation of genes with functional roles involving adaptation for both host environments would be expected to harbour a greater number of TSS.

The number of TSS discovered may not reflect the real total; factors such as sequencing depth, the variety of conditions tested, TSS definitions and methodological approach can modify this number. As seen for *E. coli*, an approach that used the well-established dRNA-seq (Förstner *et al.*, 2014) may detect fewer total TSS than an alternative method such as Cappable-seq

(Ettwiller *et al.*, 2016), regardless of greater sequencing depth. Unlike Cappable-seq, the reason that dRNA-seq cannot rely on the TEX+ treated sample alone is that the treated mixture still possesses processed 5'P transcripts and needs to be assessed relative to an untreated control (TEX-). Cappable-seq functions not by analysing the enzymatic remains of an RNA pool but by selectively capturing the unprocessed 5'PPP RNA using an enzymatic cap and magnetic pull-down strategy to separate the primary transcripts from the undesired RNA mixture; this strategy negates the critical need for an untreated control library as only 1.4% of non-ribosomal TSS are deemed as false-positive (Ettwiller *et al.*, 2016).

Within this study a considerable discrepancy was found regarding the total TSS detected between *w*MelPop-CLA (966) and *w*AlbB (4,318). However, a greater similarity between the strains was observed when matching conditions, sequencing depths and TSS thresholds were considered which reduced the total number of *w*AlbB TSS to 1,388. Upon downsampling the *w*AlbB sequencing depth to match that of *w*MelPop-CLA, the TSS most affected were intragenic TSS showing reductions of 45.6% to 32.6% and 26.1% to 14.7% for iTSS and asTSS, respectively. Taken in conjunction with TSS expression, intragenic TSS were consistently the lowest expressed TSS type within both strains (Figure 7), which suggests that deeper sequencing depths may increasingly detect poorly expressed intragenic TSS relative to all other TSS types. This may place limits on the extent to which deeper sequencing can uncover biologically meaningful TSS. Nonetheless, it is not entirely surprising that *w*AlbB should have a greater number of TSS relative to *w*MelPop-CLA given the larger genome size (+17%) and number of protein coding genes (+5%), as this would inherently support more TSS. The number of total TSS within *w*MelPop-CLA could be expected to be larger if greater sequencing depth and a greater variety of stress-inducing conditions were applied to uncover a broader range of condition-dependent TSS.

Intragenic TSS are major features of *Wolbachia* biology

The presence of intragenic TSS (iTSS and asTSS) has been observed in all TSS studies to date. The canonical view of transcription assumes a single origin of transcription initiation upstream to the coding sequence of a gene. This study found that only a small proportion of TSS (pTSS) exhibited this orientation (42.1% wMelPop-CLA, 15.8% wAlbB), indicating that a considerable proportion of transcripts have non-canonical placement and/or function. The high abundance of pervasive intragenic TSS in both *w*MelPop-CLA (46.3%) and *w*AlbB (71.7%)

has also been observed in other global TSS studies such as *E. coli* (63%), *Mycobacterium marinum* (38%), *Salmonella enterica* (10.57%) and *Bacteroides thetaiotaomicron* (45%), to name a select few (Table 5).

For both *w*MelPop-CLA and *w*AlbB, the 6S ncRNA and FSPA iTSS were consistently the two most highly expressed TSS. The 6S ncRNA, encoded by the *ssrS* gene, is known to form a hairpin secondary structure in which a premelted bubble mimics a DNA promoter capable of binding to the RNAP holoenzyme (Eσ). In *E. coli*, the bound 6S-Eσ complex results in downregulation of σ^{70} regulated genes by outcompeting σ^{70} binding to RNAP (Wassarman and Storz, 2000). The expression of 6S RNA has previously been detected in both filarial and arthropod *Wolbachia* strains. Darby, *et al.* (2012) observed upregulation of 6S RNA from the *Wolbachia* (*w*Oo) in the gonads of the filarial parasite *Onchocerca ochengi* and Darby, *et al.* (2014) reported downregulation of 6S RNA of the arthropod-associated *w*MelPop-CLA strain subjected to doxycycline. These observations suggests that *Wolbachia* 6S RNA is involved in regulating the intracellular replication rate in sync with host replication or when exposed to the bacteriostatic effects of antibiotics. Supporting 6S RNA involvement with intracellular replication, deletion of the *ssrS* in *Legionella pneumophila* resulted in a 10-fold reduction of intracellular growth along with downregulation of genes involved in amino acid metabolism and stress adaptation such as *groES* (Faucher *et al.*, 2010). Other studies have observed the increased expression of 6S RNA under stress-induced conditions such *as Burkholderia coenocepacia* under oxidative stress (Ghosh *et al.*, 2017), *E. coli* in the nutrient-poor stationary phase (Trotochaud and Wassarman, 2004), *B. burgdorferi* persistence during tick infection (Wassarman, 2018) and *Rickettsia* intracellular growth (Schroeder *et al.*, 2016). Integrating the current perspective of the role of 6S RNA, the high transcriptional output observed in this study may suggest that *Wolbachia* utilizes 6S RNA to not only control intracellular replication but also to prioritize the expression of stress-related genes, such as *groES*, without the use of sigma factors as an adaptation to mitigate the instability of proteins within streamlined symbionts.

Hypothetical proteins of unknown function are particularly abundant in *Wolbachia* genomes constituting 18.1% and 15.2% of all genes for *w*Mel and *w*AlbB, respectively. Understanding the transcriptional behaviour and genomic context of these hypotheticals may provide clues to the unique biology of *Wolbachia*. Within the top 20 expressed TSS, all TSSs that were associated with a hypothetical protein were identified as primary TSS types (pTSS). A total of

five hypothetical proteins were amongst the top 20 expressed TSS in both *w*MelPop-CLA and *w*AlbB emphasizing a need to understand their role in *Wolbachia* biology.

Ankyrin effectors are highly expressed, display strain specific behaviour and are key candidates for *Wolbachia*-host interactions

Ankyrin (ANK) repeats are tandem repeat motifs of around 33-amino acids more commonly found in eukaryotes, where they mediate protein-protein interactions (Caturegli *et al.*, 2000). The *Wolbachia* genomes harbour the highest number of ANK proteins of any prokaryote to date with *w*Mel and *w*AlbB containing 19 and 34 ANK proteins, respectively (Sinha *et al.*, 2019). The functional role of many ANK proteins in relation to *Wolbachia* biology has yet to be elucidated. Evidence highlighting the ability of ANK proteins from the closely related *Anaplasma phagocytophilum* to bind to condensed host cell chromatin (Caturegli *et al.*, 2000), and that ANK proteins of intracellular pathogens are secretedtype four secretion system (T4SS) effectors (Pan *et al.*, 2008), suggest that *Wolbachia* ANK proteins are key candidates for host-*Wolbachia* interactions.

One of most highly expressed pTSS in *w*MelPop-CLA at basal conditions (28°C) was associated with an ANK repeat containing protein (WD_RS02845 annotated as hypothetical), of which an orthologous pTSS was detected in *w*AlbB (DEJ70_RS06635) but not expressed in the top 20 TSS. The *w*MelPop-CLA ANK protein WD_RS02845 is located within the WO-B prophage region along with CI genes *cifA* and *cifB.* Expression of orthologous ANK genes have also been confirmed in *Wolbachia* strain *w*Pip of *Culex quinquefasciatus* (Sinkins *et al.*, 2005). Contrary to this study, early genomic analysis predicted that the ANK WD_RS02845 of *w*Mel would not be highly expressed (Wu *et al.*, 2004). In a study conducted by Iturbe-Ormaetxe, *et al.* (2005), ANK WD_RS02845 (previously annotated WD0633) expression was observed only in *Wolbachia* strains belonging to supergroup A but not supergroup B. In the present study, although expression was detected in *w*AlbB it was considerably lower than *w*MelPop-CLA, supporting the discrepancy between ANK expression of supergroups A and B previously mentioned. Downstream of the *w*MelPop-CLA ANK WD_RS02845, a recombinase family protein (WD_RS02850) is encoded followed by phage components such as baseplate assembly proteins (WD_RS02875, WD_RS02890, WD_RS02880) and tail proteins (WD_RS02900). Similarly, downstream of the homologous *w*AlbB ANK DEJ70_RS06635, an IS982 family transposase (DEJ70_RS06630) is encoded along with a pseudogenized major

capsid protein (DEJ70_RS06620). The location of the *w*AlbB IS982 transposase appears to have truncated the 3'end of the *w*AlbB ANK CDS (1.94Kb) compared to the *w*MelPop-CLA homolog (2.90kB) by a third (-0.96Kb). Transposase-induced truncation of the *w*AlbB ANK may negatively impact the translation of the protein and result in a defective protein relative to the *w*MelPop-CLA ANK, hence the low expression of the ANK associated pTSS may reflect functional redundancy within *w*AlbB. The high expression of ANK WD_RS02845 in *w*MelPop-CLA compared to *w*AlbB in addition to the truncation of the *w*AlbB homolog suggests that the ANK is of unique functional importance to *w*MelPop-CLA compared to *w*AlbB.

Another ANK associated pTSS (WD_RS01925) was also amongst the top 20 highly expressed TSS in *w*MelPop-CLA; an ortholog of which was detected in *w*AlbB (DEJ70_RS01385) but again expressed to a lesser degree. Based on intergenic spacing, both ANK genes are predicted to exist in a two-gene operon containing a downstream biotin dependant acetyl-CoA carboxylase (ACC), the function of which is to convert acetyl-CoA to malonyl-CoA for fatty acid biosynthesis. If the ACC is simultaneously transcribed with the ANK within an operon, then its high expression may represent a prioritization on cell membrane remodelling. The amino acid sequence of the *w*Mel ANK WD_RS02845 was analysed (SignalP-5.0 Server) to predict the presence of signal peptides that may indicate the ability for the protein to be translocated via the Sec or Tat secretion system. No signal relating to Sec or Tat recognition could be detected; however, this does not exclude the possibility that the proteins are translocated via the T1SS or T4SS.

Stress-related genes are constitutively expressed yet *w*MelPop-CLA exhibits higher stress-related gene expression than *w*AlbB

Until now, a comparative analysis of global gene expression between two arthropod *Wolbachia* has not been investigated. Here, we observed that even under optimal growth conditions, *Wolbachia* strains do not exhibit identical gene expression of stress-related orthologs. At baseline conditions (28°C) *w*MelPop-CLA and *w*AlbB exhibit similar gene expression patterns across temperature conditions with noticeable differences. Both strains show constitutive expression of heat-shock genes such as g*roES*, *dnaK* and *dnaJ*; however, *w*MelPop-CLA expressed higher transcriptional levels of the heat shock gene *htpG* (Figure 12). The exact function of HtpG is poorly understood but it has been linked to the protection of proteins under

environmental stress. Heat-shock proteins often function alongside cellular proteases such as ClpB, which degrade misfolded proteins (Dougan, Mogk and Bukau, 2002). The higher expression of the HtpG heat-shock proteins suggest that *w*MelPop-CLA is under a higher state of stress than *w*AlbB at baseline conditions. Unlike *w*AlbB, *w*MelPop-CLA is residing within cells of a non-native mosquito host instead of its native *Drosophila* host. It has been well documented that host genetic background significantly contributes to *Wolbachia* phenotypes such as delayed egg laying (Reynolds, Thomson and Hoffmann, 2003), CI (Bordenstein and Werren, 1998) and infection density (Voronin *et al.*, 2010). The co-evolutionary relationship between specific *Wolbachia* strains and their native host suggests that both host and symbiont have evolved mechanisms to synchronise growth cycles and co-adapt to each other's presence via signals secreted to or by the symbiont. It is possible that *w*MelPop-CLA outside of its native host is ill-equipped to communicate effectively within the mosquito host and thus enters a state of destabilization and stress. The present study also observed differences of expression between *w*MelPop-CLA and *w*AlbB in both ankyrin-containing proteins and components of the T4SS. It is of particular interest that *w*MelPop-CLA displays high expression of the T4SS component VirB2 alongside an ANK (WD_RS01305/DEJ70_RS06635) that is predicted to be an effector protein secreted by the T4SS and induces significant growth defects when expressed in yeast (Rice, Sheehan and Newton, 2017). It is possible that such ANK effectors are maladaptive when expressed outside of their native host because of a response to stress signals induced by the host, perhaps explaining why expression of orthologous T4SS components and ankyrins are relatively poorly expressed in *w*AlbB.

Limited promotor motifs suggest less reliance on alt-sigma factors for gene regulation

Promotor motifs are specific DNA sequences upstream of coding sequences that recruit RNA polymerase to initiate transcription. Various sigma factors can bind to the RNA holoenzyme to initiate transcription at alternative promoter regions to transcribe genes in response to environmental triggers. In the present study, we analysed the 50-nucleotide sequence upstream of all TSS to detect the presence of sequence patterns indicative of promoter motifs for σ-factor binding sites. The MEME motif suite was used to detect the overrepresentation of sequence patterns characteristic of promoter regions. The study successfully detected the presence of a sequence motif that resembles the -35 (TTGACA) and -10 (TATAAT) boxes of the *E. coli* σ^{70} housekeeping sigma factor in both *w*MelPop-CLA and *w*AlbB. It is interesting to note that

although *w*MelPop-CLA and *w*AlbB exhibited expression of the σ^{32} heat-shock sigma factor, only the σ^{70} motif was detected even when taking into account TSS type specific motifs. This suggests that either the *Wolbachia* σ^{32} motif is highly similar to the σ^{70} motif so that the sequences are difficult to distinguish apart from one another, or *Wolbachia* is less dependent on the σ^{32} protein for the initiation of stress-related genes. In support of this latter hypothesis, both *w*MelPop-CLA and *w*AlbB display high expression of the global regulator 6S RNA, described previously, and various heat-shock proteins are already highly expressed under optimum growth conditions. It is possible that due to the highly recombinant nature of arthropod *Wolbachia* genomes, the sigma σ^{32} promotor motifs have undergone sufficient sequence divergence that they are no longer able to efficiently recruit σ^{32} which may contribute to the constituent expression of HSPs. In accordance with the structure of promotor motifs detected in other TSS studies, the present study also observed an extended periodic AT-rich region upstream of the -10 motif, similar to that found in *Helicobacter pylori* (Sharma *et al.*, 2010) and *Ca*. P. asteris (Nijo et al., 2017).

Condition-specific TSS implies differential TSS regulation in response to stress

Comparing the total TSS detected amongst all conditions tested, this study found condition-specific TSS. As expected, the majority of TSS were commonly expressed at baseline conditions of 28°C. In addition to temperature stress, exposing *w*AlbB to various antibiotic stressors uncovered TSS specific to these treatments. However, due to the majority of TSS being low expression intragenic TSS, these condition specific TSS may be indicative of spurious transcription as instead of and functional relevance to the experienced condition. Regardless, these observations lay the foundations for *Wolbachia'*s ability to express TSS specific to environmental conditions as well as the ability for a streamlined endosymbiont to differentially express TSS when faced with environmental perturbations. The precise nature of *Wolbachia'*s ability to regulate TSS will be discussed in the following chapter.

Chapter 3: TSS regulation of *Wolbachia* in response to environmental and chemical stressors.

Introduction

Chapter 2 introduced a new viewpoint that the transcriptomes of arthropod associated *Wolbachia* are highly complex with only a minority of TSS associated with canonical gene expression. The previous chapter also began to touch upon key differences and similarities in gene expression between *w*MelPop-CLA and *w*AlbB, expression of phage components in *w*MelPop-CLA appear to be highly expressed. Previous research has shown that *w*Mel and *w*AlbB have significantly different susceptibilities to thermal temperatures, these differences in response to thermal stress have major implications for their suitability as biocontrol agents for *Wolbachia*-mosquito field-releases in tropical climates (Ross *et al.*, 2017). However, investigating the genetic components responsible for such differences are still hindered by the inability to culture and modify *Wolbachia* in pure culture. Therefore, differences in the ability between *Wolbachia* strains have mainly been demonstrated via phenotypic disparities in infection density, CI strength, and pathogen blocking (Ant *et al.*, 2018; Flores *et al.*, 2020). The following chapter aims to investigate the transcriptional differences between *w*MelPop-CLA and *w*AlbB to regulate TSS in response to environmental stressors.

This chapter demonstrates that differential gene expression in *Wolbachia* extends beyond regulation of chaperones. This study reports that genes involved in transmission and host interactions are regulated in response thermal stress specifically, these observations may be central to the thermal susceptibility reported in *w*MelPop-CLA (Ross *et al.*, 2017). Regulation of TSS associated to MGEs, such as phage, also appear to response to thermal stress however their relevance may not be associated to their original function. Observations reported in this chapter also include evidence for regulation of nucleotide and membrane metabolism which may be key pathways involved in survival to environmental stress. This chapter also present the first transcriptional response of *w*AlbB to antibiotic stress, these results expands upon previous research on antibiotic associated gene regulation in *Wolbachia* (Darby *et al.*, 2014).

The observations reported in this chapter aims to improve the understanding of *Wolbachia* gene expression in a manner which may assist strain choice in vector-release strategies.

Methods

Bioinformatics: Differential expression analysis

To detect the differential expression of TSS under the stressors chosen differential expression (DE) analysis was conducted using the empirical analysis of digital gene expression package edger (version 3.30.3) with the settings: prior count at 2, minimal fold change threshold of 2-fold, and an adjusted p-value false discovery rate (Benjamini-Hochberg Procedure) of 5%. Multidimensional scaling plots were analysed and produced via the plotMDS function within edgeR and heatmaps were created using the previously mentioned Heatmap.2 in the ggplot2 R package.

Bioinformatics: Functional annotation

To aid in the biological interpretation of differentially expressed TSS under stress, genes were given the Kegg Ortholog (KO) annotation so that they could be analysed into the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis. Functional gene annotations were retrieved with the use of EggNOG (version 5.0.0), KO annotations of differentially expressed genes were then used as input into the KEGG pathway analysis under the organism prefixes of 'wol' and 'wpp' for *w*Mel and *w*AlbB respectively. Pathways which resulted in the most hits or were associated with highly differentially expressed TSS were the focus of biological interpretation.

Results

All TSS samples were analysed via an MDS plot to visualise the level of similarity between samples of varying stressors by their log-fold change (Figure 14). Both *w*MelPop-CLA and *w*AlbB replicates clustered tightly amongst their conditions indicating strong similarity between replicates. Clear distances were observed between temperature and time for both *w*MelPop-CLA and *w*AlbB. Temperature for wAlbB was most separated on the first dimension followed by time in the second dimension (Figure 13. B) whilst temperature and time were more evenly separated for *w*MelPop-CLA amongst both dimensions of the MDS pot (Figure

13. A). *w*AlbB also displayed clear separation within the first dimension between doxycycline and rifampicin whilst moxifloxacin and control conditions were tightly clustered indicating minimal differential expression between moxifloxacin and 28°C controls and high differential expression between doxycycline and rifampicin relative to 28°C controls (Figure 13. C).

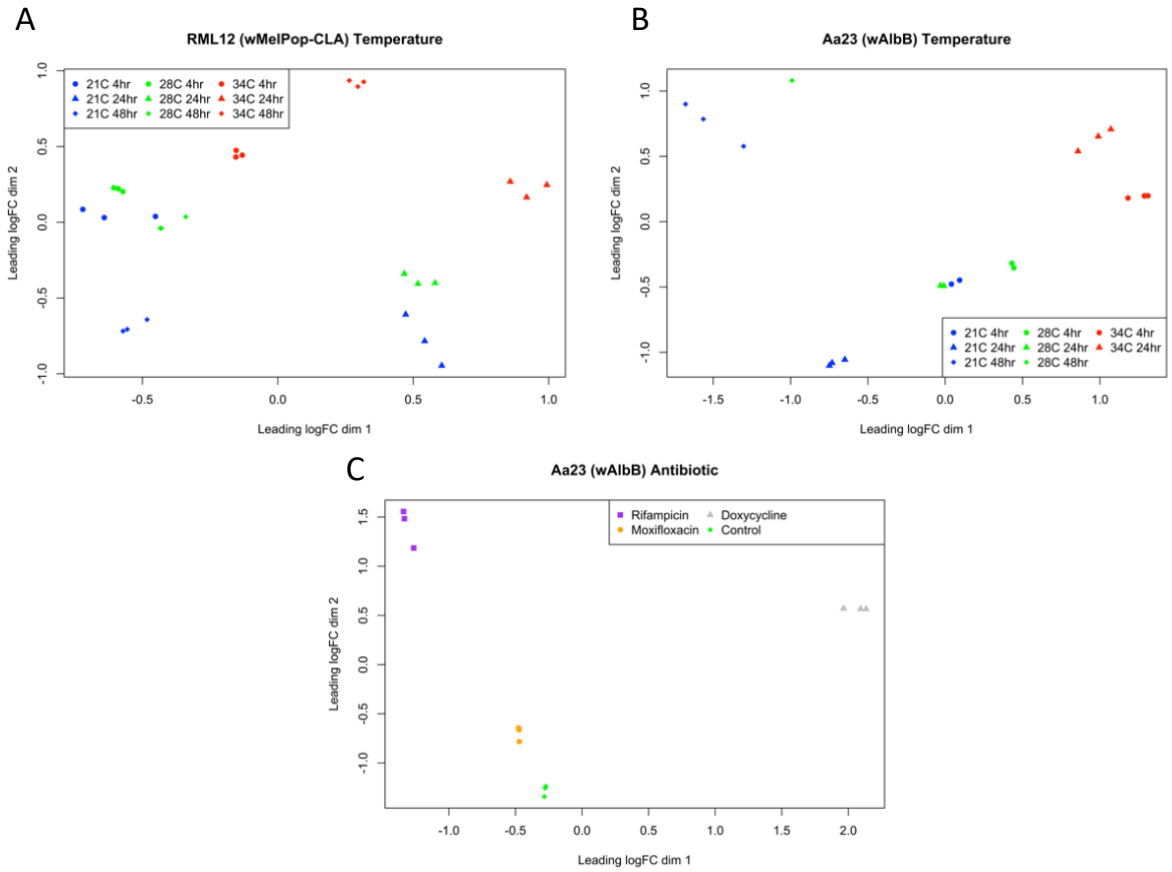


Figure 14. MDS plots of stress induced *w*MelPop-CLA and *w*AlbB samples. A) *w*MelPop-CLA exposed to temperature stress, B) *w*AlbB exposed to temperature stress (Green 28°C control, Blue 21°C, Red 34°C, square 4hrs exposure, circle 24hrs and triangle 48hrs, C) *w*AlbB 24hr exposure to antibiotic stress (Green 28°C control (square), grey doxycycline (circle), orange moxifloxacin (triangle), rifampicin purple (diamond).

Total TSS from each strain underwent differential expression analysis via edger. Of the 966 and 4,318 TSS detected, 311 and 688 TSS were found to be differentially expressed (2-fold change, FDR 5%) under at least 1 temperature stress condition for *w*MelPop-CLA and *w*AlbB respectively (Figure 15). The proportion of TSS types that exhibited differential expression were similar to the proportion of TSS types discovered. For both *w*MelPop-CLA and *w*AlbB, gTSS represented the smallest group of DE TSS whereas the largest represented group of DE TSS were pTSS and iTSS for *w*MelPop and *w*AlbB respectively (Figure 16).

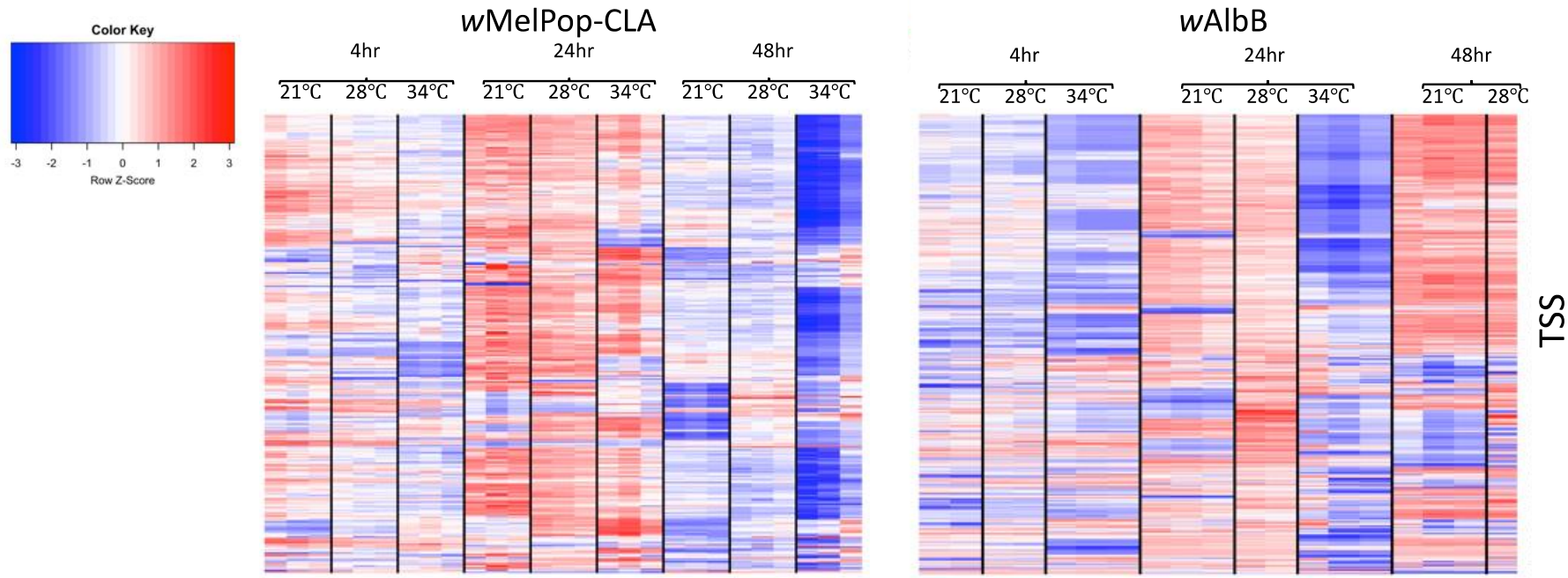
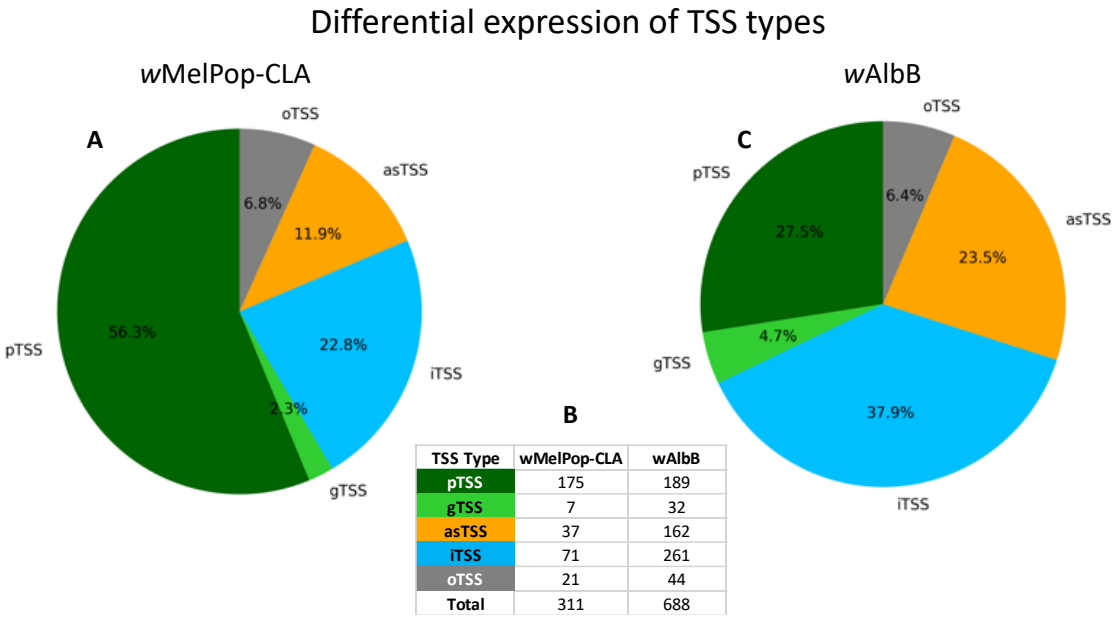


Figure 15. Heatmap of wMelPop-CLA and wAlbB differentially expressed total TSS types under temperature stress. Heatmap of all TSS types that were differentially expressed (2-fold change, 5% FDR) within *w*MelPop-CLA (311) and *w*AlbB (688). Expression levels displayed in Z-scores (CPM minus mean CPM of associated row, divided by standard deviation of CPMs of that row per strain).



B

TSS Type	wMelPop-CLA	wAlbB
pTSS	175	189
gTSS	7	32
asTSS	37	162
iTSS	71	261
oTSS	21	44
Total	311	688

Figure 16. Proportion of differentially expressed TSS types under temperature stress. A & C) Proportion of TSS types that exhibited differential expression under at least 1 temperature condition stress (2-fold change, FDR 5%) for *w*MelPop-CLA and *w*AlbB respectively, B) Summary of differentially expressed TSS types.

For both *w*MelPop-CLA and *w*AlbB the majority the of DE TSS were regulated at 34°C with 201 and 335 DE TSS respectively. At 21°C temperature stress, *w*MelPop-CLA and *w*AlbB displayed 110 and 353 DE TSS respectively. The majority of DE TSS were observed to occur at the 48hr timepoint for both 21°C and 34°C in both strains (Figure 17), conversely the least number of DE TSS occurred under the earliest time exposure of 4hr at 21°C and 34°C in both strains. The largest overlap for temperature induced DE TSS occurred between 24hr and 48hr timepoints for both strains (Figure 17).

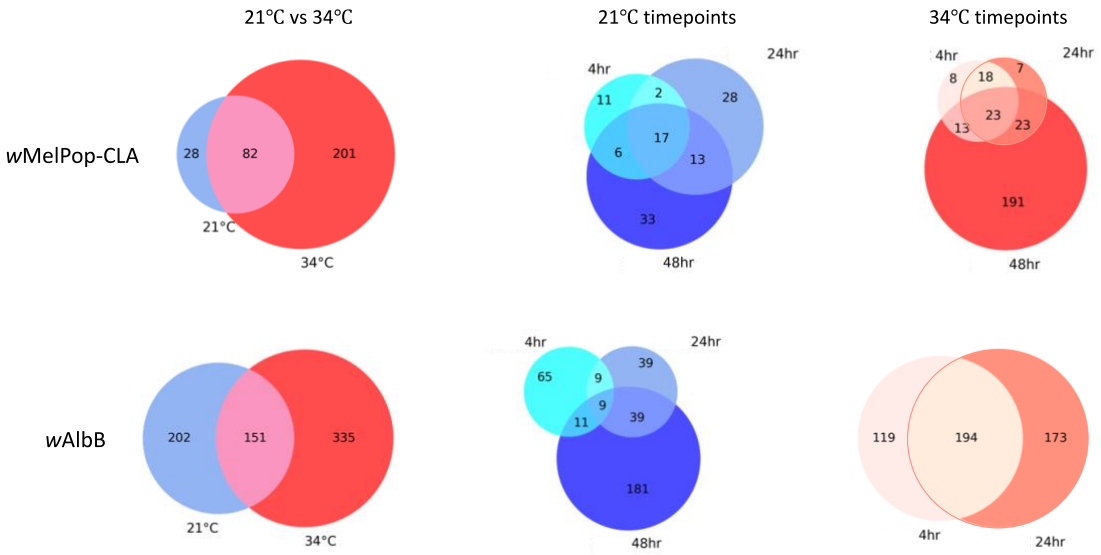


Figure 17. Overlapping differentially expressed TSS between temperatures and time points. Due to two out of three *w*AlbB 34°C 48hr not reaching the recommended sequencing depths, only 4hr vs 24hr is shown for *w*AlbB 34°C.

Table 6. Summary of differentially expressed TSS types between all stressors for both *w*MelPop-CLA and *w*AlbB.

	w MelPop-CLA				*w* AlbB									
	21°C		34°C		21°C		34°C		Doxy		Moxi		Rif	
TSS Type	DOWN	UP	DOWN	UP	DOWN	UP	DOWN	UP	DOWN	UP	DOWN	UP	DOWN	UP
pTSS	51	16	135	27	83	16	122	12	26	40	2	0	221	8
gTSS	0	0	6	1	14	0	22	2	8	2	2	0	22	0
iTSS	18	6	65	5	125	9	166	20	37	26	19	0	239	15
asTSS	7	3	35	3	73	7	99	16	24	15	7	0	89	27
oTSS	4	5	14	6	18	13	22	8	9	3	1	0	30	8
Total	80	30	255	42	313	45	431	58	104	86	31	0	601	58

Of the 175 *w*MelPop-CLA pTSS that were differentially expressed under temperature stress, 16 and 27 were upregulated under 21°C and 34°C respectively. Of the 189 *w*AlbB pTSS that were differentially regulated under temperature stress, 16 and 12 were upregulated under 21°C and 34°C respectively. For both *w*MelPop-CLA and *w*AlbB, the majority of differentially expressed pTSS were downregulated in response to temperature stress. Of the 760 *w*AlbB pTSS that were differentially expressed under antibiotic stress, 40, 0, and 8 were upregulated whilst 122, 26, and 221 were downregulated under doxycycline, moxifloxacin, and rifampicin stress respectively (Table 6). Like temperature stress, moxifloxacin and rifampicin exposure resulted in the majority of pTSS being downregulated whilst doxycycline resulted in the majority of pTSS being upregulated. Using the KEGG mapper search tool, both wMelPop-CLA and wAlbB differentially regulated pTSS were associated to functions and pathways related to chaperone and HSPs, translations, repair and recombination, nucleotide metabolism,

lipid metabolism, MGEs, ankyrins, secretion and transport, and carbon metabolism (Table 7). The largest group of upregulated pTSS during temperature stress were the chaperones and HSPs for both *w*MelPop-CLA and *w*AlbB. For *w*AlbB antibiotic stress, doxycycline displayed the largest category of upregulated pTSS involved in secretion and transport.

Table 7. Selection of differentially expressed pTSS and their assigned Kegg functional/pathway category. Upregulated pTSS are in red whilst downregulated pTSS are blue.

	*w*MelPop-CLA		*w*AlbB				
Function/Pathway	**21°C**	**34°C**	**21°C**	**34°C**	**Doxycycline**	**Moxifloxacin**	**Rifampicin**
Chaperone & HSP	clpB (WD_0224) dnaK (WD_0928) htpG (WD_1277) dnaJ (WD_0040) grpE (WD_0800) groES (WD_0308)	dnaK (WD_0928) htpG (WD_1277) dnaJ (WD_0040) grpE (WD_0800) groES (WD_0308) rpoH (WD_1064) hslV (WD_1189)	Hsp20 (DEJ70_01400) GroES (DEJ70_04505)	hsp20 (DEJ70_01400) GroES (DEJ70_04505) grpE (DEJ70_05955) dnaJ (DEJ70_04285) DnaK (DEJ70_03145) clpB (DEJ70_02835)	dnaJ (DEJ70_04285) DnaK (DEJ70_03145) hsp20 (DEJ70_01400) GroES (DEJ70_04505) HtpG (DEJ70_00615)		
Translation	rpmI (WD_0864) ribosomal large subunit (WD_0415) rpsT; ribosomal protein S20 (WD_1150) rpsL; ribosomal protein S12 (WD_0014)	rpsL; ribosomal protein S12 (WD_0014) rpsP (WD_0798) ribosomal large subunit (WD_0415) ribosomal protein L36 (WD_0342)	30S ribosomal protein S12 (DEJ70_01765) 50S ribosomal protein L33 (DEJ70_03135)	50S ribosomal protein L13 (DEJ70_01425) 50S ribosomal protein L25 (DEJ70_02230) rpmG (DEJ70_03135)	50S ribosomal protein L25 (DEJ70_02230) 50S ribosomal protein L13 (DEJ70_01425) 50S ribosomal protein L34 (DEJ70_05270) 30S ribosomal protein S16 (DEJ70_05945)		50S ribosomal protein L13 (DEJ70_01425)
Repair & Recombination	uvrD (WD_0963) recA (WD_1050)	recA (WD_1050)	nrdF (DEJ70_03295)				
Nucleotide metabolism	guaB (WD_0089) endonuclease III (WD_0789)	RNase P RNA component class A (WD_RS06150) tmk (WD_1251)	purE (DEJ70_05765) ribonucleotide-diphosphate reductase subunit beta (DEJ70_03295)	dgt (DEJ70_00095)	nucleoside-diphosphate kinase (DEJ70_00725) dCTP deaminase (DEJ70_04615)		RNase P RNA component class A (DEJ70_01630)
Lipid metabolism	dgkA (WD_1163) ispE (WD_0360) coaD (WD_0524)	ispE (WD_0360)	phospho-N-acetylmuramoyl-pentapeptide-transferase (DEJ70_06815)		1-acyl-sn-glycerol-3-phosphate acyltransferase (DEJ70_02635) pyruvate dehydrogenase complex E1 (DEJ70_04440) ElbB (DEJ70_00165)		
Transposases & Phage		Transposase (WD_RS03395)		Phage portal protein (DEJ70_00855)	Transposase (DEJ70_06920) Transposase (DEJ70_04650)		
Ankyrins	WD_0691	WD_0691 WD_0754 WD_0434 WD_0191		DEJ70_02765	DEJ70_01550		DEJ70_03875 DEJ70_00610 DEJ70_02915
secretion & Transport	MFS transporter (WD_1033)	multidrug effflux MFS transporter (WD_0384) MFS transporter (WD1064)	type IV secretion system protein (DEJ70_04590) virB9 (DEJ70_06825) cation:proton antiporter (DEJ70_00405) ftsA (DEJ70_02265)	efflux RND transporter permease subunit (DEJ70_02955) type I secretion system permease/ATPase (DEJ70_02565) phosphate ABC transporter substrate binding protein (DEJ70_04630) VirB3 (DEJ70_01260) SecA (DEJ70_06415) OmpA (DEJ70_02485)	VirB3 (DEJ70_01260) TatA/TatE (DEJ70_03700) type IV secretion system protein (DEJ70_04590) TolC (DEJ70_01430) OmpA (DEJ70_02485) MFS transporter (DEJ70_00025) HlyD (DEJ70_03750) YidC (DEJ70_02990) Msp2 family outer membrane protein (DEJ70_04560) efflux RND transporter permease subunit (DEJ70_02950)	cation:proton antiporter (DEJ70_00405) 50S ribosomal protein L13 (DEJ70_01425)	YidC (DEJ70_02990) efflux RND transporter permease (DEJ70_02950) VirB3 (DEJ70_01260) virB9 (DEJ70_06825)
Carbon metabolism	pccA (WD_0433) fumC (WD_0492)		sucA (DEJ70_05990)		pyruvate dehydrogenase complex E1 (DEJ70_04440)		

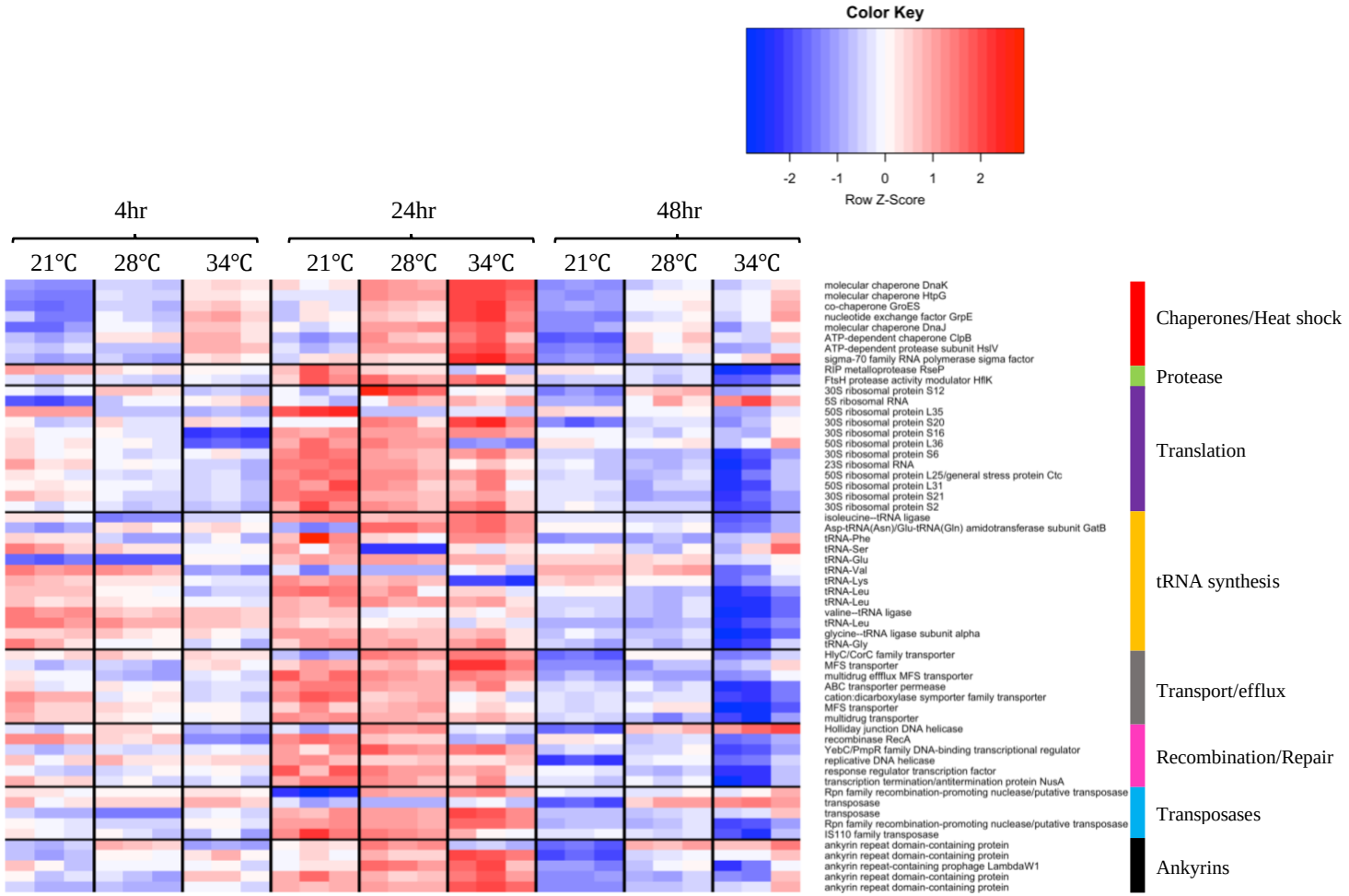


Figure 18. Heatmap of *w*MelPop-CLA DE pTSS under temperature stress. A selection of various DE pTSS based on their functional categories. TSS are normalized to CPM within each replicate (n = 3, minimum of 2). Expression levels displayed in Z-scores (CPM minus mean CPM of associated row, divided by standard deviation of CPMs of that row).

Of the orthologous pTSS shared between *w*MelPop-CLA and *w*AlbB, 40 pTSS were differentially regulated under at least 1 temperature stress condition in both strains (Figure 19). Orthologous DE pTSS contained genes annotated for heat shock proteins (*dnaJ, groES, dnaK, dnaJ, htpG*), proteases (*clpB*), T4SS secretion components (*virB9*), translation (Ribosomes), cell division (*ftsZ*), outermembrane proteins (*ompA*), and hypotheticals. Both strains also exhibited regulation of heat shock related protease components such as *hslV* and *clpB*. HSPs displayed upregulation at 34°C and polar expression at 21°C. Contrary to *w*MelPop-CLA, *w*AlbB strain did not display DE of sigma-32.

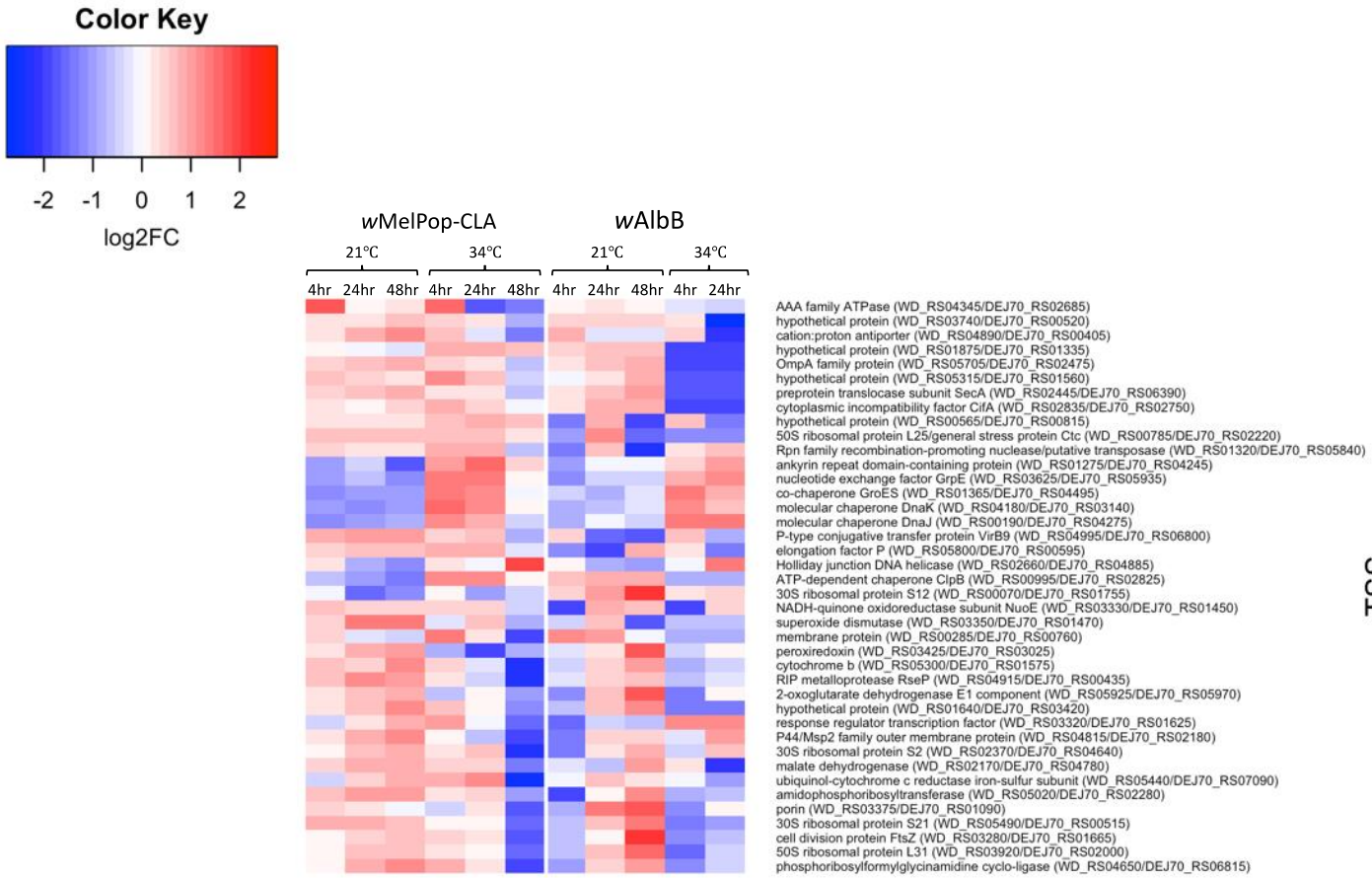


Figure 19. Heatmap of differentially expressed orthologous pTSS exposed to temperature stress. Orthologous pTSS containing genes related to heat shock proteins, ankyrins, secretion systems, and hypothetical proteins. The colour key is shared between strains however the normalized colour value (blue to red) is exclusive to the individual strain. Expression levels displayed in Z-scores (CPM minus mean CPM of associated row, divided by standard deviation of CPMs of that row specific to strain)

For the 760 wAlbB DE TSS exposed to antibiotic stress the pattern of expression was markedly different between each specific antibiotic (Figure 20). Under rifampicin exposure, the expression of the large majority of TSS were significantly lower relative to moxifloxacin, doxycycline, and 28°C conditions. Contrary to rifampicin, doxycycline exposure resulted in high expression patterns relative to moxifloxacin, rifampicin, and 28°C. Both moxifloxacin and 28°C conditions resulted in similar TSS expression patterns (Figure 20) as reflected in the MDS plots (Figure 14). Expression patterns exclusive to DE *w*AlbB pTSS across all tested conditions were similar to the expression pattern of all TSS types (Figure 21).

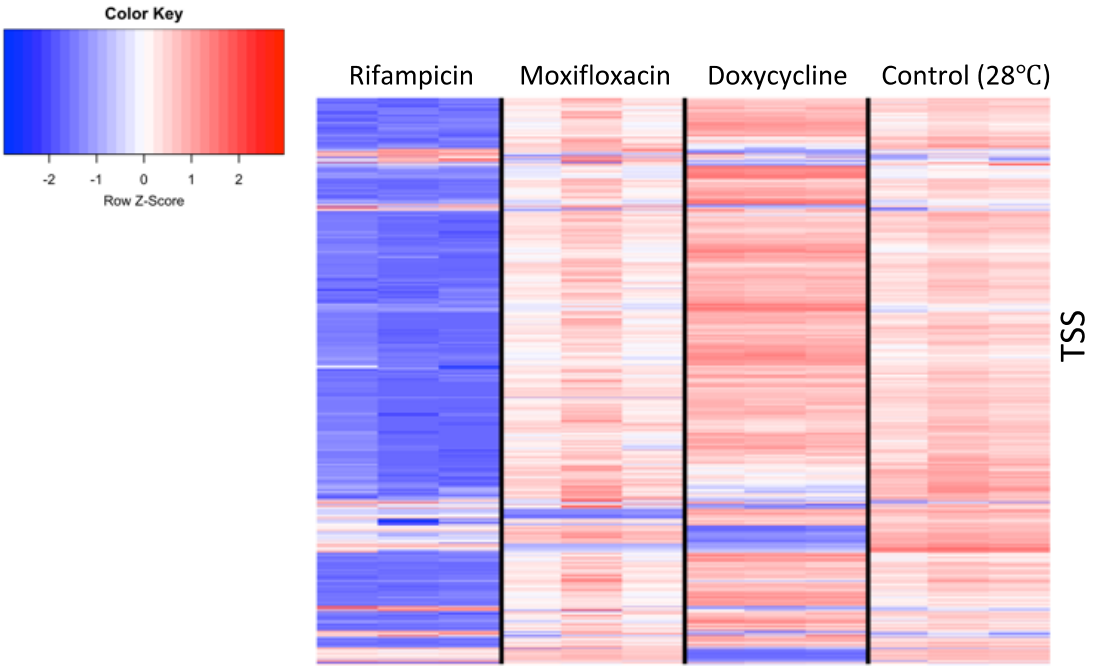


Figure 20. Heatmap of wAlbB differentially expressed total TSS under antibiotic exposure. TSS are normalized to CPM within each replicate (n = 3). Expression levels displayed in Z-scores (CPM minus mean CPM of associated row, divided by standard deviation of CPMs of that row).

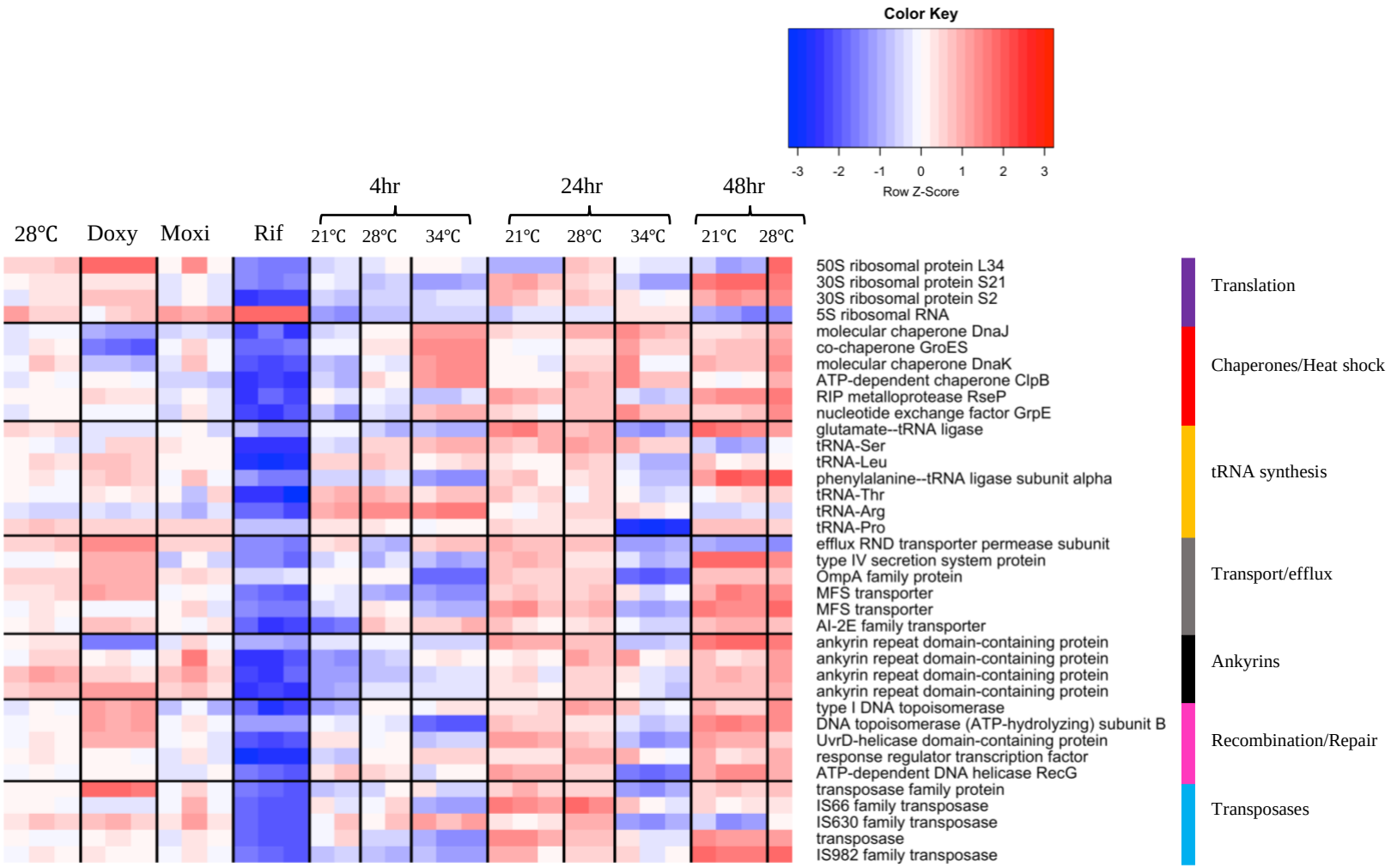


Figure 21. Heatmap of wAlbB DE pTSS under all stress conditions. A selection of various DE pTSS based on their function categories. TSS are normalized to CPM within each replicate (n = 3, minimum of 2). Expression levels displayed in Z-scores (CPM minus mean CPM of associated row, divided by standard deviation of CPMs of that row).

For *w*AlbB, pTSS exhibited overlapping and condition specific DE between 21°C, 34°C, doxycycline, and rifampicin (Figure 22). One common DE pTSS was found between 21°C and doxycycline which was associated to a T4SS protein (DEJ70_04590). Two common DE pTSS was found between 34°C and rifampicin associated to 5S rRNA (DEJ70_05240) and RNase P RNA component class A (DEJ70_01630). No other upregulated pTSS were common between stress conditions.

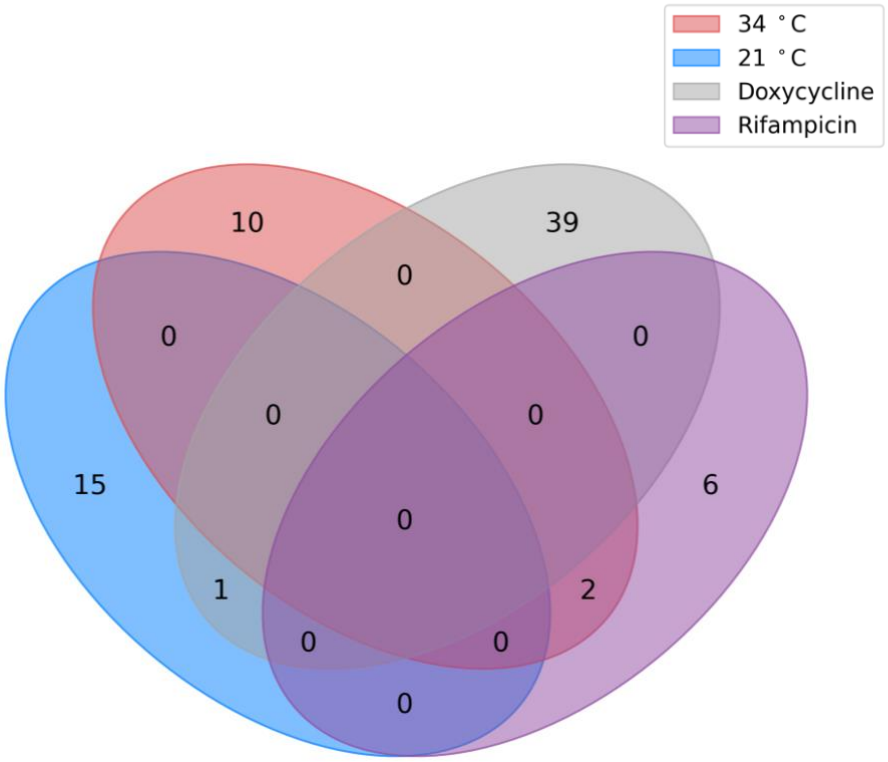


Figure 22. Venn diagram of stress induced upregulated pTSS in *w*AlbB. Summary of differentially expressed TSS and their overlapping occurrence under different antibiotic exposures. Moxifloxacin is excluded due to lack of upregulated TSS.

Discussion

Primary transcriptional start sites are not exclusively regulated within *Wolbachia*

To determine whether the TSS discovered were subject to regulation under temperature and antibiotic stress, total TSSs were analysed via edgeR to determine the extent/capacity for differential expression. This study successfully detected statistically significant differential expression amongst all known TSS types for both *w*MelPop-CLA and *w*AlbB. The initial hypothesis was for pTSS to encompass the vast majority of regulated TSS since the definition

of pTSS implies the "primary" transcriptional unit for gene expression. This study found that the proportion of differentially expressed TSS types resembled that of the total TSS types discovered, this suggests that all TSS types are capable of regulation. This surprising outcome suggests that pTSS are not exclusively regulated within the *Wolbachia* strains tested. The fact that the various types of TSS are regulated whatsoever emphasizes the limitations of current transcriptomic approaches, such as RNA-seq, to approach the resolution required to distinguish between transcriptional units. The clarity of single-nucleotide cappable-seq prevents the merging of distinct transcriptional units. Therefore, assigning sequencing reads to specific TSS types reveals and accurately quantifies transcript diversity within the *Wolbachia* strains in question.

Wolbachia strains show transcriptional differences in thermal stress

Heat shock proteins are amongst the most ancestral response mechanisms of all cellular life that counter protein instability in response to increased environmental temperature shifts. Assuming that the pTSSs are key drivers for gene expression, the initial hypothesis was for heat shock related pTSSs to be upregulated upon an increase thermal stress to 34°C. After being exposed to 34°C heat stress for a minimum of 4 hours, both *w*MelPop-CLA and *w*AlbB were capable of upregulating stress related HSPs such as *DnaK, grpE, groES* and *htpG* . Heat shock proteins play a major role in maintaining protein stability during heat stress, once bound to exposed hydrophobic regions of denatured proteins they prevent their aggregation and promote refolding. The DnaK-DnaJ-GrpE (KJE) system is one of the key chaperone systems in bacteria, it is composed of DnaK, its co-chaperone DnaJ, and the nucleotide-exchange factor GrpE. The co-chaperone DnaJ delivers denatured proteins to DnaK for refolding whilst GrpE stimulates ADP-ATP turnover releasing the restabilized protein (Brehmer *et al.*, 2001). The GroEL/GroES system (GroELS) functions to encapsulate unfolded proteins within its ring dome-like chamber enabling protein refolding within a stable environment (Chaudhry *et al.*, 2003). Along with the heat shock chaperones, both strains also displayed upregulation of protease components such as HslV and ClpB which degrade irreversibly damaged proteins.

All the previously mentioned heat shock related pTSSs generally showed peak upregulation at 24hrs. However, after 48hr exposure to 34°C *w*MelPop-CLA exhibited in mass downregulation of all pTSS including HSPs which may be indicative of cell death after extreme prolonged heat stress. Upregulation of *Wolbachia* HSPs during changes to the host cell environment have been

observed previously in response to DENV infection (Leitner and Bishop, 2021) and host development (Grote *et al.*, 2017; Chung *et al.*, 2019). Similarly, the obligate endosymbiont of the tsetse fly *Sodalis glossinidius* also upregulates components of the KJE system (Roma *et al.*, 2019), the similarity of these findings supports the use of pTSS to quantify endosymbiont gene expression in a similar manor to RNA-seq.

Capable-seq was able to detect the pTSS for sigma-32 for both *w*MelPop-CLA and *w*AlbB under thermal stress, however only *w*MelPop-CLA exhibited differential expression of this seemingly essential heat shock components. Based on the differential TSS data obtained in this study, it appears only *w*MelPop-CLA is capable of upregulating sigma-32 whilst *w*AlbB appears to upregulate HSPs without, or with minimal contribution from, sigma-32. It may be possible that *w*AlbB is less dependent on typical heat shock regulators than *w*MelPop-CLA possibly due to evolutionary adaptions to a mosquito host that regularly encounters higher temperatures than the drosophila host of *w*MelPop-CLA. The host-symbiont temperature co-adaption hypothesis has been supported by the observation that temperate localised *Drosophila*-*w*Mel infections have higher rates of vertical transmission that those of more tropical *Drosophila*-*w*Mel associations in regions of Australia (Hague *et al.*, 2021).

It is possible *w*AlbB could be receiving signals from its natively adapted host that supersedes the regulatory function of sigma-32 to induce HSP expression, whereas *w*MelPop-CLA in its non-native host behaves in a more independent/pathogenic and energy intensive manor. When exposed to 21°C cold stress for a minimum of 4hr exposure *w*MelPop-CLA displayed contrasting regulation of HSPs as components of both the KJE and GroELS systems, such as *dnaK* and *htpG* respectively, were downregulated. At 24hr and 48hr exposure to 21°C *w*AlbB also downregulated of heat shock related proteins such as Hsp20, DnaJ, and GroES. The observed upregulation of HSPs supports the notion that endosymbionts of reduced genome size still retain the capacity to regulate genes in response to temperature fluctuation despite a relatively constant intracellular environment. Our observations demonstrate that both strains are equally capable of upregulating heat shock related genes in response to heat stress.

This data implies that the differences in thermal susceptibility observed between *w*MelPop-CLA, *w*AlbB, and presumably *w*Mel (Ross *et al.*, 2017) are unlikely due to a failure to induce HSPs in response to temperature induced protein instability. Differences in host background,

specifically hosts that share a longer phylogenetic history with their symbiont, may impact host-symbiont communication during stressed conditions. Understanding the essentiality of regulatory components such as sigma-32 under different host backgrounds may help understand host-symbiont communications under stress. *Wolbachia* infected mosquito release strategies would benefit by better understanding climate-host-symbiont co-adaptive relationships to choose the best suited strains for successful and sustainable deployment.

Seeking shade in times of stress: Evidence for stress induced translocation

During the immediate heat shock at 34°C for 4hrs, *w*MelPop-CLA upregulated the hypothetical protein WD0811. Interestingly the upregulated hypothetical WD0811 contained a eukaryotic WH2 motif involved in cytoskeleton interactions. An NCBI blastp search found that WD0811 had closest non-*wolbachia* homology (75.56% identity, 7e-68 E-value) to the WH2 domain protein of the *Trichonephila clavate* spider of the golden orb-web genus. Eukaryotic domains related to those of spiders have previously been associated to the EAM of phage origin coincidentally harbouring the black widow latrotoxin-CTD (Bordenstein and Bordenstein, 2016).

The WH2 domain is commonly found in WASP proteins which are typically involved in cell motility, trafficking and signalling (Veltman and Insall, 2010). Many bacterial pathogens manipulate host actin in order to traverse the intracellular environment (Dramsi and Cossart, 1998). *R. conorri* and *R. japonica* induce bacterial entry by manipulating host clathrin and actin (Chan *et al.*, 2009). Other bacterial pathogens that encode WH2 domain-containing proteins include TARP in *Chlamydia* (Jewett *et al.*, 2006), SipC in *S. enterica* (Hayward and Koronakis, 1999), and VopL in *Vibrio parahaemolyticus* (Liverman *et al.*, 2007; Namgoong *et al.*, 2011). *Wolbachia* has been confirmed to utilize host dynamin and clatherin for clatherin-mediated endocytosis (White *et al.*, 2017). A previous study also predicted WD0811 to be a secreted effector that demonstrated punctate localization and was capable of inducing growth defects when expressed in yeast (Rice, Sheehan and Newton, 2017). The WD0811 effector was upregulated across both 4 and 24hr 34°C timepoints. It's possible that under heat stress *w*MelPop-CLA, and thus *w*Mel, may attempt to rearrange host actin to translocate to niches in the host that are less susceptible to heat stress.

Membrane remodelling and biofilm potential

In support of membrane remodelling in response to stress, the outer membrane protein (OMP) family pTSS *ompA* was upregulated under doxycycline exposure. Unlike other porins such as OmpF and OmpC from *E. coli*, OmpA functions primarily as a structural OMP contributing to the integrity of the cell membrane (Dupont *et al.*, 2004). Overexpression of the structurally related OmpX from *Enterobacter aerogenes* inversely decreased expression of OmpF/C porins and increased antibiotic resistance (Dupont *et al.*, 2004). Although *w*AlbB does not encode homologs for OmpF/C porins, doxycycline exposure may still promote an increase in membrane structural integrity. Interestingly the OmpA-like protein of *w*AlbB (DEJ70_02485) lacks the transmembrane domain present in *E. coli* OmpA and only contains the peptidoglycan binding domain towards the C-terminus. The PG-binding domain of OmpA has been shown to directly bind to diaminopimelic acid (DAP) found in the peptide linkages of PG (Park *et al.*, 2012). OmpA deficient *Acinetobacter baumannii* have displayed diminished virulence and outer membrane integrity (Choi *et al.*, 2005). Although *Wolbachia* retains genes required for lipid II synthesis there is currently no evidence of a typical PG layer, however a PG-like structures may still benefit structurally from a PG-binding OmpA-like protein. Since the immune deficiency (IMD) pathway of insects implements peptidoglycan recognition proteins (PGRP) to target the cell wall of gram-negative bacteria, components of a PG-like layer may still evade immune recognition whilst reinforcing structural integrity with the use of OmpA-like proteins. It would be of interest to see whether mosquito PGRPs are able to detect OmpA-bound PG in comparison to unbound OmpA to determine if OmpA also functions to mask immunogenic components.

Besides contributing to membrane stability, OmpA has been implicated in aspects of host interaction. Deletion of *ompA* from the gram-negative gammaproteobacterial *Yersinia pestis* has been shown to decrease intracellular survival (Verghese, 2011). Mutations in *ompA* within the facultative symbiont *Sodalis glossinidius* results in poor colonization of the Tsetse fly host due to reduced biofilm formation (Weiss *et al.*, 2008; Maltz *et al.*, 2012). Knockout mutations of *ompA* in the mosquito gut symbiont *Cedecea neteri* also resulted in impaired biofilm production and colonization (Hegde *et al.*, 2019). Although the potential for *Wolbachia* to produce biofilms has not been studied, an exopolysaccharide biosynthesis protein (EBP) (DEJ70_03710) was upregulated in the presence of doxycycline albeit just below the 2-FC threshold. It is also interesting to note that the EBP is not only retained in both *w*MelPop-CLA

and *w*AlbB but has also been duplicated in both strains which may suggest that its function is beneficial.

The alphaproteobacterial *Rhizobium leguminosarum* forms mutualistic symbiotic relationships with legumes in the formation of nitrogen-fixing nodules. *R. leguminosarum* has been found to secrete exopolysaccharides via a T1SS which are implicated in the colonization and adhesion of root nodules (Russo *et al.*, 2006). Subpopulations of extracellular *Wolbachia* may secrete biofilm-like compounds for extracellular persistence or host cell adhesion. It is tempting to speculate that *Wolbachia* could use a biofilm-like adhesions to maintain proximity to each other, this could permit the uptake of multiple *Wolbachia* in a single engulfment event and increase the density of *Wolbachia* within intracellular niches and increase the likelihood of vertical transmission.

A pTSS associated to the *plsC* gene (DEJ70_02635) encoding 1-acyl-sn-glycerol-3-phosphate acyltransferase in the glycophospholipid metabolism pathway was the highest upregulated (pTSS in response to doxycycline exposure. PlsC is an integral membrane protein responsible for the catalytic conversion of 1-acyl-sn-glycerol-3-phosphate into 1,2-Diacyl-sn-glycerol 3-phosphate also known as phosphatidic acid (PA) (Coleman, 1990). PA is the major precursor for membrane phospholipids such as phosphatidylethanolamine, phosphatidylglycerol, and cardiolipin. Doxycycline exposure to *w*MelPop-CLA was also found to upregulate components involved in phospholipid metabolism such as *gspA* (Darby *et al.*, 2014).The *gspA* gene encodes Glycerol-3-phosphate dehydrogenase which catalysis the conversion of glycerone-P to sn-glycerol-3P. Once acylated by glycerol-3P acyltransferase to 1-Acyl-sn-glycerol-3P, PlsC subsequently uses 1-Acyl-sn-glycerol-3P as substrate for synthesis of key phospholipids. Upregulation of the phospholipid biosynthesis pathway may indicate an attempt to promote membrane stability and decrease the permeability for antibiotics to enter the cell.

The behaviour of Wolbachia secretion systems in response to stress

Of the four putatively functional secretion systems (T1SS, T4SS, Sec, Tat) observed in *Wolbachia*, the present study found upregulation of pTSS involved in all four systems. The T1SS translocates unfolded proteins from the cytoplasm across the OM in a one step process. The T1SS is composed of 3 main proteins: an ATP-binding cassette protein (ABC), membrane fusion protein (MFP), and an OMP forming the channel in which unfolded proteins are

exported into the extracellular space. Of the three required T1SS components, evidence of upregulation in wAlbB was found for the MFP HylD and a pTSS associated to a TolC family protein (DEJ70_01430) under doxycycline exposure. The functional role of the T1SS in *Wolbachia* has yet to be revealed, however other members of the Rickettsiales family have been found to use the T1SS for secretion of ANKs such as *Ehrlichea* (Wakeel *et al.*, 2011), *Rickettsia* (Kaur *et al.*, 2012) and *Orientia* (VieBrock *et al.*, 2014). The present study found an ANK gene (DEJ70_01550) that was upregulated alongside HlyD MFP under doxycycline exposure. Although the *Wolbachia* T1SS has yet to be studied extensively, upregulation of the *w*AlbB ANK (DEJ70_01550) should be considered as a possible candidate T1SS effector for future studies.

The T4SS is involved in the transport of macromolecules across the cell envelope. The essentiality of the T4SS has been associated with environmental adaption, pathogenicity, and symbiotic establishment. Upregulation of T4SS subunits was observed for virB8 (DEJ70_04590) and VirB3 (DEJ70_01260) in *w*AlbB under 21°C and doxycycline conditions respectively. The absence of upregulation for the remaining core subunits of the T4SS suggests that either upregulation of the entire T4SS machinery is not required for increased macromolecular transport, or that VirB3 and VirB3 is able to function independently. Although speculative, an independent function for individual T4SS components may be possible. The VirB2 subunit of *Agrobacterium tumefaciens* T4SS has been suggested to be involved in host cell adhesion (Backert, Fronzes and Waksman, 2008). *Arabidopsis thaliana* treated with RNA interference (RNAi) targeting Virb2 interacting proteins in *A. tumefaciens* were found to be less susceptible to transformation by *A. tumefaciens*. *Wolbachia* T4SS components could therefore be involved in adhesion to the enveloping host cell membrane supporting the previously discussed stress induced motility. Overexpression of T4SS components during periods of stress may enhance the association to the host cell membrane to increase overall membrane stability or enable translocation to more hospitable host tissues/cells.

ANKs are exported via the T4SS in various Rickettsiales such as the ER associated *Rickettsia* ANK protein RARP-2 (Lehman *et al.*, 2018). It is possible that the *w*AlbB ANK (DEJ70_01550) is also exported via the T4SS comparable to RARP-2. If so, *w*AlbB ANK (DEJ70_01550) may also associate to host derived membranes such as the ER and host membrane in which *Wolbachia* resides. Assuming the *w*AlbB ANK (DEJ70_01550) does

associate to host membranes, upregulation may reflect a demand for host lipids to reinforce the permeability barrier and/or to increase nutritional uptake under stress.

The behaviour of *Wolbachia* efflux pumps in response to stress

As efflux pumps are a key method of drug expulsion in gram negative bacteria their presence in response to antibiotics is to be expected. Upregulation of efflux pumps in response to rifampicin exposure has been observed in several bacteria. In *M. tuberculosis* the major facilitator superfamily (MFS) efflux pump gene *Rv1258c* has been significantly correlated with clinical isolates resistant to multiple antibiotics which include rifampicin (Siddiqi *et al.*, 2004). Besides *Rv1258c*, a further 15 efflux genes have been implicated in rifampicin resistant strains of *M. tuberculosis* (Li *et al.*, 2015). Resistance to rifamycins has also been observed in *Neisseria gonorrhoeae* coinciding with the overexpression of the *mtr* operon that encodes *mtrCDE* efflux pump system (Hagman *et al.*, 1995).

The effectiveness of efflux pumps is determined by the competitive balance between drug influx across the outer membrane (OM) relative to drug efflux. Antibiotics, such as doxycycline, are relatively lipophilic and able to penetrate both the OM and inner membrane (IM) by diffusion or through OM bound porins. Coincidentally the current study found *w*AlbB to upregulated both an MFS transporter (dej70_00025) and an efflux RND transporter (DEJ70_04560) in response to doxycycline exposure. Also, the porin (DEJ70_01090) was in the 75th percentile of downregulated pTSS in rifampicin stressed *w*AlbB, this may be an effort to limit to influx of rifampicin through the OM. Most MFS efflux pumps are located within the IM and expel antibiotics from the cytosol into the periplasm. However, antibiotics are readily diffused back into the cytosol hence why MFS-type efflux pumps do not usually play a dominant role in antibiotic resistance in gram-negative bacteria (Li, Plésiat and Nikaido, 2015). It appears that even though *w*MelPop-CLA and *w*AlbB are capable of upregulating efflux pumps in response to drug exposure, their susceptibility to doxycycline and rifampicin may partly be explained by inefficient drug efflux. These data emphasise the key role efflux pumps play in antibiotic expulsion in intracellular pathogens and supports their involvement in the ability for *Wolbachia* to export chemical stressors via transcriptional regulation of efflux pumps.

Efflux pumps have also been implicated in the regulation of peptidoglycan in response to cell wall stress. Zeevi, *et al.* (2013) found that deletions in the transporter genes of *Listeria monocytogenes* resulted in an inability to produce and shed peptidoglycan under stress. The facultative intracellular parasite *Yersinia enterocolitica* was found to harbour temperature regulated MFS transporters at 37°C (Bengoechea and Skurnik, 2000). *Wolbachia* lack transglycosylases responsible for the linear glycan chains of peptidoglycan (Foster *et al.*, 2005), however *Wolbachia* possess a subset of peptidoglycan genes such as lipid II suggesting a peptidoglycan-like cell wall. As temperature stress is known to cause an increase in membrane permeability and *w*MelPop-CLA is more thermally sensitive than *w*AlbB, *w*MelPop-CLA may be increasing support for membrane repair/stability.

Stress induced transposase and phage induction

Resistance mechanisms in intracellular bacteria are largely attributable to spontaneous genomic mutations as opposed to horizontal gene transfer. Transposable elements like transposases are capable of inducing adaptive mutagenesis that may lead to environmental adaptations as first proposed by McClintock (1984). Despite their highly reduced genome, *Wolbachia* harbour an unusually high number insertion sequences (IS) of which include transposases (Wu *et al.*, 2004; Sinha *et al.*, 2019). Insertion sequences in *Wolbachia* contribute to 11% of their genomic content whereas most other prokaryotic genomes are composed of only 3% (Siguier, Filée and Chandler, 2006). Both *w*MelPop-CLA and *w*AlbB were observed to regulate the expression of various transposases under stress. The study observed *w*MelPop-CLA upregulating the transposase (WD_RS03395) under 34°C with progressively increased expression throughout exposure. This data supports the hypothesis that *Wolbachia* harbour temperature sensitive mobile genetic elements similar to what has been observed in regards to the WO phage in *w*Mel (Bordenstein and Bordenstein, 2011). Besides temperature, antibiotic exposure was also observed to regulate the expression of transposases in *w*AlbB. Of the transposases that were upregulated in *w*AlbB in response to antibiotic exposure, doxycycline induced the expression of the transposase (DEJ70_06920) which neighboured a pseudogenized IS110 family transposase (DEJ70_RS06900) of possible origin.

The fact that each regulated transposase was dependent on a specific stressor suggests that *Wolbachia* transposable elements are not induced by a general stress signal but can be expressed by specific stressors, this suggests that given a wider range of stressors more

transposases could be transcriptionally regulated. It has previously been predicted that >70% of *Wolbachia* insertion sequences are non-functional (Cerveau *et al.*, 2011), it would be of interest to confirm whether transcriptionally active transposons are also functional. If non-functional transposons are being expressed then they could be either: the transcriptional "waste" signature of degrading genome, a signalling molecule, or a function independent of translation i.e. ncRNA.

Regarding phage, *w*AlbB surprisingly displayed upregulation of a phage portal protein (DEJ70_RS00855) in response to 34°C heat stress. Phage in *w*AlbB have been determined to be inactivate due to pseudogenized and absent components required for lytic expression, the portal protein that has been upregulated is also a pseudogene. Upregulation of a phage pseudogene may be a remnant of a previously activate phage in the *w*AlbB evolutionary history which has yet to be completely removed.

Protein and nucleotide salvage

Doxycycline induced ribosomal stalling produces incompletely synthesized polypeptides which are then recycled via the use of proteases. Doxycycline exposure induced upregulation of the inner membrane bound protease FtsH in wAlbB , FtsH contributes to the degradation of proteins for AA recycling (Langklotz, Baumann and Narberhaus, 2012). FtsH proteolysis is also involved in the post-translational regulation of the heat shock regulator RpoH (Tomoyasu *et al.*, 1995). Proteolytic degradation of *rpoH* by FtsH would contribute to explaining the downregulation of previously mentioned heat shock genes whilst providing an AA pool for nutritional uptake.

Rifampicin exhibits it's bactericidal activity by binding to the β subunit pocket of RNAP to sterically block elongating RNA and thus inhibit transcription (Campbell *et al.*, 2001). Rifampicin induced transcriptional inhibition explains why the majority of DE pTSS were downregulated, thus it is difficult to separate intentional downregulation compared to rifampicin induced transcriptional inhibition. Based on the crystal structure of rifampicin bound to the RNAP of *Thermus aquaticus*, extension of the RNA transcript is hindered when the transcript is 2-3 nucleotides in length (Campbell *et al.*, 2001). These short abortive transcripts are eventually bound for nucleotide salvage and metabolically recycled. In support

of nucleotide salvage, the current study observed upregulation Nucleoside-diphosphate kinase in nucleotide metabolism. Upregulation of nucleotide metabolism pathways in response to rifampicin has been observed in other bacteria such as *E. coli* (Shaw *et al.*, 2003). These data may imply that *Wolbachia* responds to stalled RNA synthesis by increasing the supply pool of RNA precursors for energy metabolism to maintain survival. Several pTSS associated genes related to energy and nucleotide metabolism were also found to be upregulated in response to doxycycline. Since *Wolbachia* have limited capacity for carbohydrate metabolism, energy consumption relies mainly on *de novo* nucleotide synthesis and AA metabolism (Foster *et al.*, 2005). Nucleotide metabolism associated pTSS such as nucleoside-diphosphate kinase (NDK, DEJ70_00725) and dCTP deaminase were both significantly upregulated. It is possible that as doxycycline blocks ribosomal translation *w*AlbB may breakdown the build-up of stunted RNA for energy utilization.

Ribosomal rescue

Doxycycline functions primarily by binding to the 30S ribosomal subunit which subsequently prevents amino acyl-tRNA binding to the acceptor A-site (Griffin *et al.*, 2010). Given the known mechanism of doxycycline, it was expected that doxycycline exposed *w*AlbB would attempt to mitigate these effects by upregulating ribosomal subunits. Upregulation of ribosomal subunits in response to doxycycline exposure has been observed in *w*MelPop-CLA (Darby *et al.*, 2014). This data suggests that arthropod *Wolbachia* of supergroups A and B share an ability to upregulate translational machinery in response to ribosomal inhibition. A previous study conducted by Rao, *et al.* (2012) exposed the filarial (*Brugia malayi*) derived *Wolbachia* (*w*Bm) to doxycycline. Contrary to what has been noted in arthropod *Wolbachia,* the study noted an absence of ribosomal protein upregulation in treated worms.

Chapter 4: Investigating candidate ncRNA and TSS assisted operon structure

Introduction

Chapters 2 and 3 reported that a significant portion of TSS do not appear to be primary drivers of gene expression. A significant proportion of both *w*MelPop-CLA and *w*AlbB genomes

harbour an abundance of intergenic and antisense transcripts, these non-canonical TSS could be indicative of alternative regulatory mechanisms. The following chapter explores the regulation of non-canonical TSS and examines the possible function of asTSS, oTSS, and iTSS in the context of antisense regulation, intergenic ncRNA, and alternative transcriptional units (ATUs) respectively. The presence and function of ncRNA in arthropod *Wolbachia* are starting to become apparent, ncRNA have previously been implicated in roles involving host germline stem cell regulation (Ote, Ueyama and Yamamoto, 2016) emphasising their importance in host interactions. The resolution provided by Cappable-seq allows for a detailed examination of transcription within predicted operons. The aim of this chapter is to provide a foundation for ATU/sub-operon investigations in *Wolbachia.* This study reports that ATUs are relevant to the *Wolbachia* thermal susceptibility and that ATUs within endosymbionts may be a strategy to increase genomic plasticity within streamlined genomes.

Methods

Bioinformatics: RNA secondary structure analysis

To evaluate the likelihood in which a TSS could be associated with an RNA encoded secondary structure, the 100 nucleotide upstream sequences of all TSS types were analysed for computationally predicted RNA secondary strcture. RNA secondary structures and minimum free energy (MFE) values were obtained using the 100 nucleotide downstream sequences for all discovered TSS types as input into RNAfold (version 2.4.13) of the ViennaRNA package (version 2.0) using default settings. A total of 2000 (1000 per strand) randomly generate TSS locations were created with a custom python3 script as a comparative control. MFE values of all predicted secondary structures were plotted onto a violin plot to assess MFE distributions between TSS types and randomly generated TSS.

Bioinformatics: Operon structure analysis

To assess the structure of predicted operons within *w*MelPop-CLA and *w*AlbB, gene annotation for both strains were used as input into the operon prediction tool Rockhopper (version 1.7.0.11). BAM alignment files for 28°C 48hr conditions were used as input into Rockhopper to decipher whether the intergenic gaps between genes displayed evidence of co-expression. Operons and associated TSS were visualised using the Artemis genome browser and annotation tool (version 18.1.0). Expression of an operon was determined if the initiating gene of a

predicted operon harboured a pTSS. Operons were considered to contain an ATU if they displayed a moderately expressed intergenic TSS upstream to a gene within the associated operon.

Results

ncRNA and secondary structure analysis

To assess the likelihood that an expressed RNA transcript can form a secondary structure, the 100 nt downstream sequence of all TSS were subject to RNA folding analysis to retrieve the MFE of its predicted secondary structure. A total of 2,000 (1,000 per strand) randomly generated TSS sites were used as a comparative baseline to compare the distribution of MFE values per TSS type within each strain (Figure 23). The mean MFE values for randomly generated sequences were -14.68 and -14.23 for *w*MelPop-CLA and *w*AlbB respectively (Table 8). The highest mean MFE values for *w*MelPop-CLA was associated to gTSS (-13.73) whereas for *w*AlbB pTSS were the highest median MFE at -13.85. The lowest MFE values were observed for oTSS (-18.33) and iTSS (-15.93) for *w*MelPop-CLA and *w*AlbB respectively. The lowest observed MFE values were associated to iTSS and oTSS transcripts for both *w*MelPop-CLA and *w*AlbB. Compared to randomly generated TSS, MFE values for iTSS and oTSS from *w*MelPop-CLA and *w*AlbB were significantly ($p < 0.05$) lower.

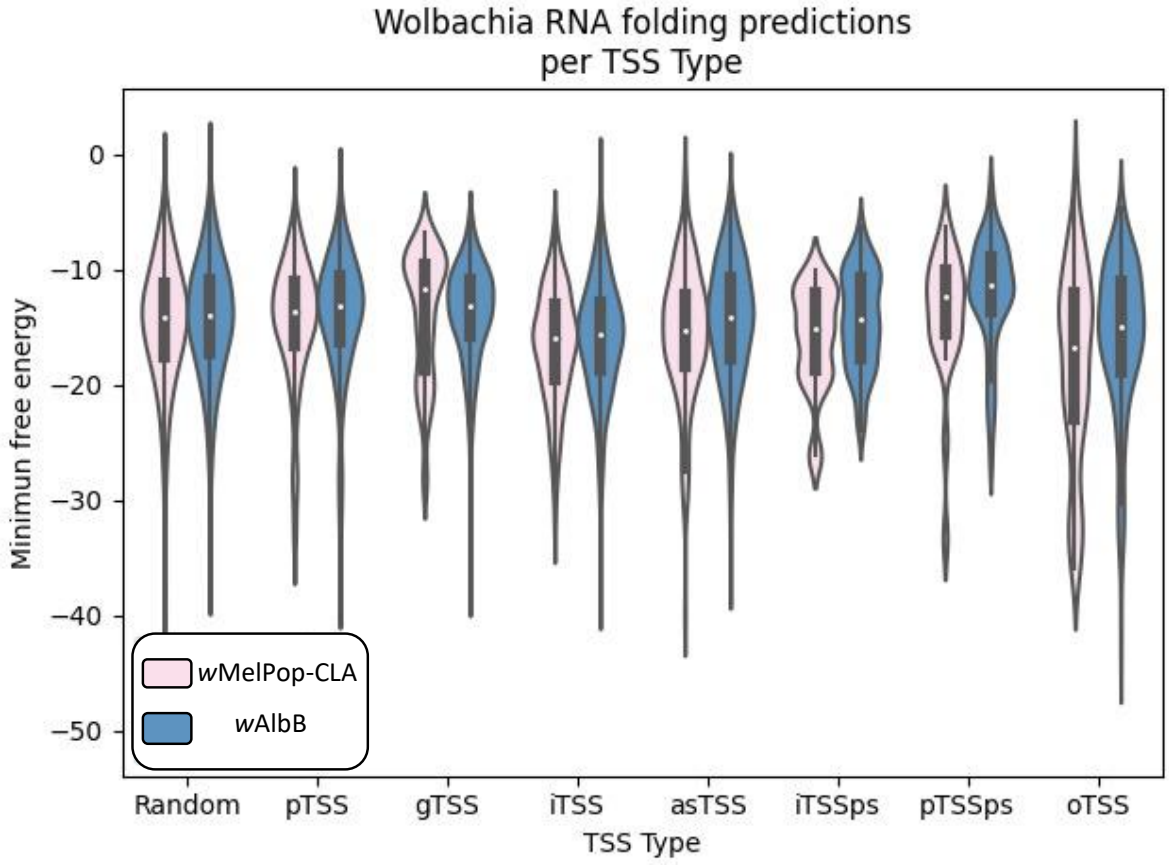


Figure 23) Violin plot of RNA folding MFE distributions per TSS type. MFE values for each TSS type plotted on a violin plot, central point represents medium value. Pink colour represents *w*MelPop-CLA, blue represent *w*AlbB.

Table 8) Summary of TSS RNA folding probability. Pink colour represents *w*MelPop-CLA, blue represent *w*AlbB. Asterisk '*' indicates independent students t-test significance ($p < 0.05$) relative to random control.

		Random	pTSS	gTSS	iTSS	asTSS	iTSSps	pTSSps	oTSS
	Mean MFE	-14.68	-14.56	-13.73	-16.66	-15.52	-15.92	-13.95	-18.33
***w* MelPop-CLA**	***n***	2000	355	25	214	135	11	22	52
	signif		ns	ns	*	ns	ns	ns	*
	Mean MFE	-14.23	-13.85	-13.94	-15.93	-14.45	-14.32	-12.08	-15.9
***w* AlbB**	***n***	2000	568	155	1229	785	86	54	199
	signif		ns	ns	*	ns	*	ns	*

The *fsa* associated iTSS in both *w*MelPop-CLA and *w*AlbB was consistently the most highly expressed TSS along with the 6S RNA associated iTSS. Architecture of the local region of *fsa* in both *w*MelPop-CLA and *w*AlbB was visually inspected using the Artemis genome viewer tool. Both *w*MelPop-CLA and *w*AlbB *fsa* genes contained a highly expressed iTSS located precisely 326 nt downstream the CDS with an average normalized expression of 292,750 and

34,453 CPM respectively at 28°C (Figure 24). The *w*AlbB *fsa* gene contained an upstream pTSS whilst *w*MelPop-CLA contained a pTSS associated to the upstream neighbouring gene WD_RS02460 albeit with considerably lower expression (Figure 24).

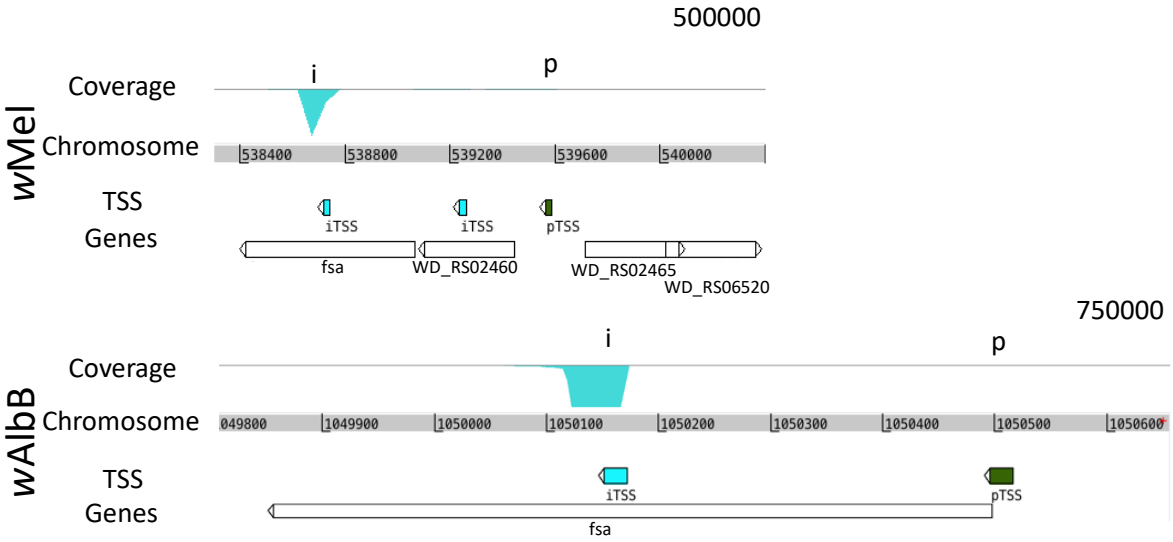


Figure 24) Local region of highly expressed iTSS of the *fsa* gene in *w*Mel and *w*AlbB. Coverage (value top right) represents read coverage relative to displayed genomic region. TSS of interest highlighted on coverage plot: 'i' iTSS, 'p' pTSS.

The full amino acid sequence of the *fsa* CDS was inspected with the InterPro (v88.0) protein classification tool to analyse the protein domain active sites relative to the TSS site. Both *w*MelPop-CLA and *w*AlbB Fsa proteins contained a transaldolase active site at residues 22-30 and 23-31 respectively (Figure 25). The iTSS of both *w*MelPop-CLA and *w*AlbB is located at residue 110 of the *fsa* CDS and downstream of the Fsa active site (Figure 25). Structural RNA folding probability of the 100 nt sequence downstream to the TSS predicted an MFE of -24.7 and -28.1 for *w*MelPop-CLA and *w*AlbB respectively. The folded structure of both iTSS fsa transcript predicted three hairpin loops with high base-pairing probability (Figure 25).

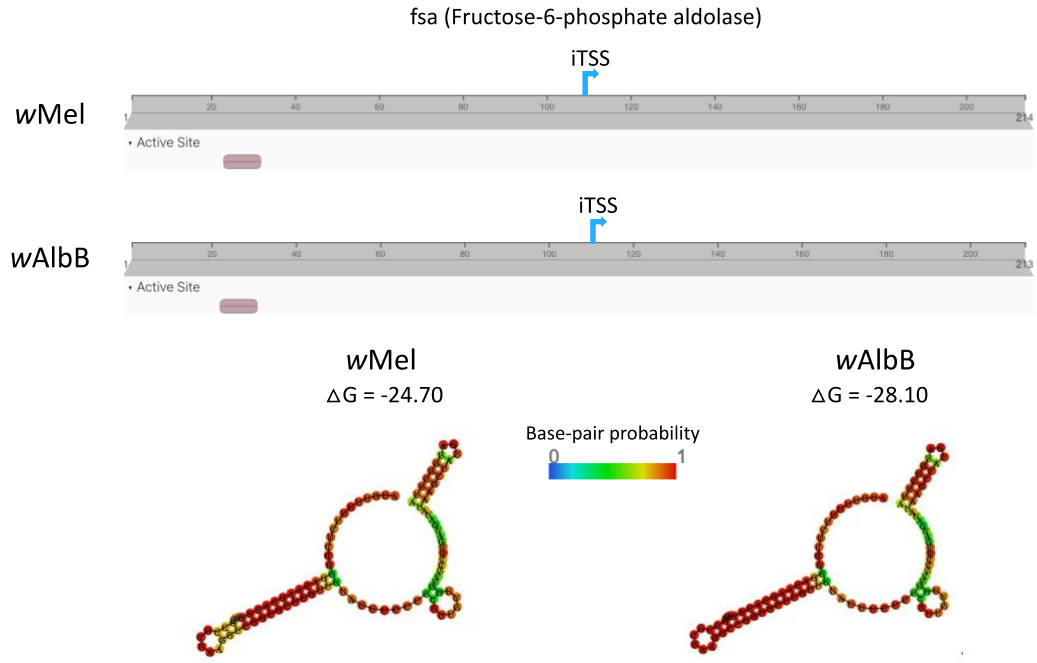


Figure 25) Protein domain comparison (Interpro) of the highly expressed iTSS associated *fsa* gene in *w*Mel and *w*AlbB. iTSS location displayed onto amino acid residues. Active site represents transaldolase activity.

The top 20 expressed *w*MelPop-CLA oTSS and their associated MFE values were sorted according to their average CPM value (Table 9). Of the top 20 expressed *w*MelPop-CLA oTSS transcripts, 16 oTSS exhibited differential regulation under at least one temperature stress condition. Average expression ranged from 46 to 2,543 CPM whilst MFE values ranged from -9.3 to -37.1

Table 9) Top 20 expressed *w*MelPop-CLA candidate intergenic ncRNA. Top 20 expressed oTSS for candidate intergenic ncRNA. "id" indicates TSS type, strand, genomic coordinate, and closest coding sequence. Average CPM values are taken from 28°C 48hr samples (n = 3). Minimum free energy (MFE) predicted from RNAfold server using 100 nts downstream the TSS. Evidence of regulation highlighted by blue (downregulated) or red (upregulated) coloured cells. Any ncRNA that have been previously reported are referenced in "Reported".

id	Avg CPM	MFE	21°C	34°C	Reported
oTSS_+_1039620_WD_RS04920	2453	-31.1	-4.8	-3.3	ncrwMel02 (Woolfit et al., 2015)
oTSS_-_915471_WD_RS04305	1233	-19.8	-2.2		
oTSS_+_587738_WD_RS02740	973	-21.5			(Woolfit et al., 2015)
oTSS_+_1247436_WD_RS05925	720	-17.4		2.6	
oTSS_+_168335_WD_RS00865	243	-27.6		-22.8	
oTSS_+_1129557_WD_RS06825	202	-32.5			
oTSS_-_388537_WD_RS01820	201	-9.3		-18.5	
oTSS_-_12612_WD_RS00065	152	-35.3		2.9	
oTSS_+_934857_WD_RS04410	140	-28.4		-2.6	
oTSS_+_163736_WD_RS00825	111	-19.6		-10.3	
oTSS_-_352761_WD_RS01620	100	-14	-2.8	5.4	
oTSS_+_360587_WD_RS01715	91	-9.8		-8.7	
oTSS_+_458017_WD_RS06470	85	-40.8			
oTSS_+_679584_WD_RS03195	72	-37.1			
oTSS_+_372453_WD_RS01755	61	-13.7		-7.2	
oTSS_+_360568_WD_RS01715	51	-9.8	-4.8		
oTSS_+_980572_WD_RS04630	51	-24.3	-5.3		
oTSS_+_432096_WD_RS06450	49	-19.7	2.5	3.2	
oTSS_+_130667_WD_RS00655	48	-30.3		-5	
oTSS_+_166797_WD_RS00865	46	-17		-4.4	

The two most highly expressed oTSS transcripts in *w*MelPop-CLA (ncrwMel02 and oTSS_-_915471.0_WD_RS04305) were expressed at an average of 2453 and 1233 CPM at 28°C. The ncrwMel02 transcript was located at position 1039620 (+ strand) overlapping a pTSS transcript on the opposite strand and upstream a cation: proton antiporter (WD_RS04890). The predicted RNAfold secondary structure for ncrwMel02 contained two hairpin loops, two internal bulges, and a multibranched loop with a relatively low MFE of -31.1 compared to randomly generated transcripts (Figure 26). The ncrwMel02 associated TSS was downregulated at -3.4-fc and -2.3-fc under 21°C and 34°C respectively. The oTSS_-_915471.0_WD_RS04305 transcript, hence forth referred to as ncrwMelP471, was located at position 915471 (- strand) upstream an IS110 transposase (WD_RS06115). The predicted RNAfold secondary structure for ncrwMelP471 displayed near bilateral symmetry with two outer hairpin loops, two larger internal loops, and two smaller internal loops with a relatively low MFE of -19.8 compared to randomly generated transcripts (Figure 26).

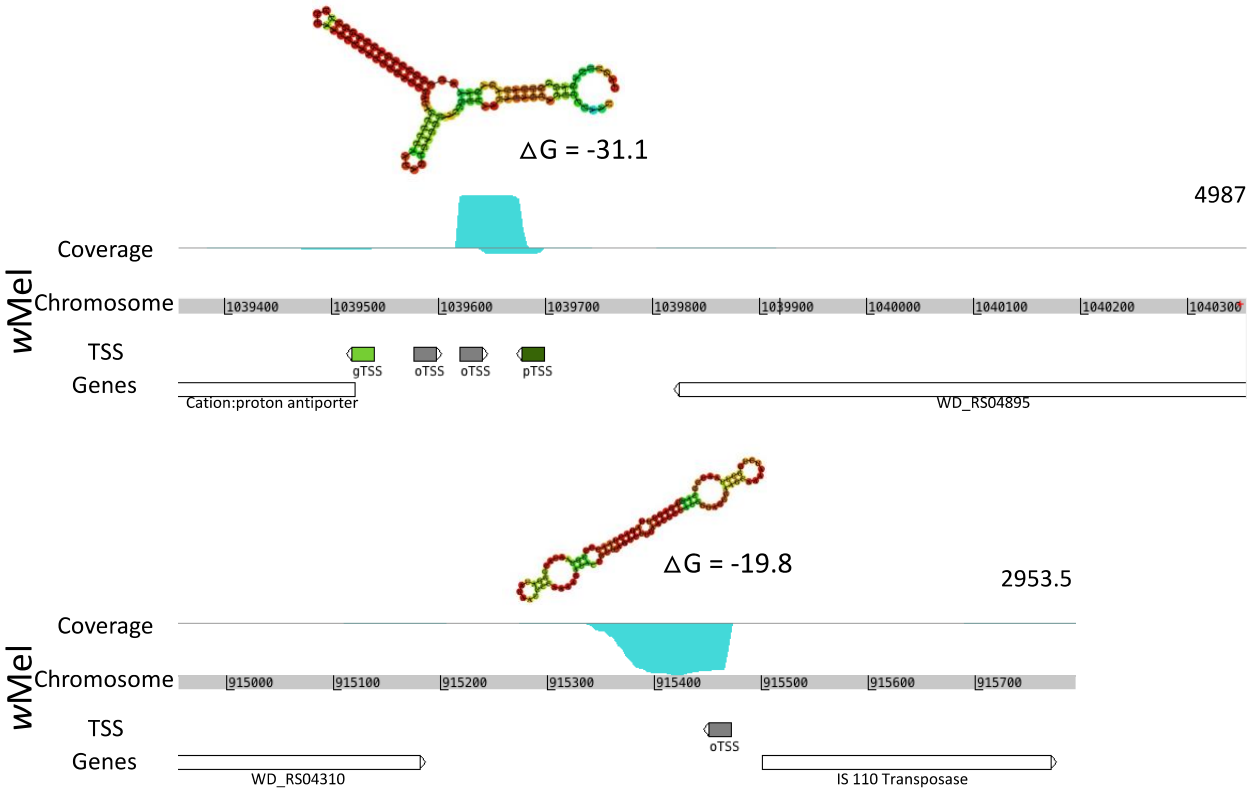


Figure 26) Local region and RNA folding prediction of the two most highly expressed *w*MelPop-CLA oTSS ncRNA candidate. Coverage (value top right) represents read coverage relative to displayed genomic region. Predicted folded RNA secondary structure displayed above associated TSS and coverage peak.

The top 20 expressed *w*AlbB oTSS and their associated MFE values were sorted according to their average CPM value (Table 10). Of the top 20 expressed *w*AlbB oTSS transcripts 14 exhibited differential regulation under at least one antibiotic stressor whilst none of the top 20 oTSS were regulated under temperature stress. Average expression ranged from 276 to 4,986 CPM whilst MFE values ranged from -8.6 to -39.6.

Table 10) Top 20 expressed *w*AlbB candidate intergenic ncRNA. Top 20 expressed oTSS for candidate intergenic ncRNA. "id" indicates TSS type, strand, genomic coordinate, and closest coding sequence. Average CPM values are taken from 28°C 48hr samples (*n* = 3). Minimum free energy (MFE) predicted from RNAfold server using 100 nts downstream the TSS. Evidence of regulation highlighted by blue (downregulated) or red (upregulated) coloured cells.

id	Avg CPM	MFE	21°C	34°C	Doxy	Rif
oTSS_-_261019.0_DEJ70_RS01320	4986	-12.6				2.3
oTSS_-_652774.0_DEJ70_RS03120	3445	-8.6				2.6
oTSS_+_1263788.0_DEJ70_RS06135	2808	-10.8				2.1
oTSS_+_628168.0_DEJ70_RS03015	1314	-11.3				2.6
oTSS_+_402709.0_DEJ70_RS02005	1262	-11.1				3.7
oTSS_-_1115509.0_DEJ70_RS05400	1073	-24.6				
oTSS_-_974456.0_DEJ70_RS04640	953	-13.7				
oTSS_+_803042.0_DEJ70_RS03840	806	-17.3				
oTSS_-_940682.0_DEJ70_RS04490	745	-19.6				
oTSS_-_874644.0_DEJ70_RS04165	606	-16.3			-1.8	-3.5
oTSS_+_584385.0_DEJ70_RS02810	565	-31.7				2.7
oTSS_+_1235358.0_DEJ70_RS06005	525	-10.9				2.6
oTSS_+_874154.0_DEJ70_RS04255	464	-10.5				3.7
oTSS_-_1094527.0_DEJ70_RS05330	446	-15.6				-2.8
oTSS_-_824252.0_DEJ70_RS03905	442	-12.5				2.5
oTSS_+_948372.0_DEJ70_RS04540	432	-39.6				-2
oTSS_+_1015048.0_DEJ70_RS04890	329	-35.7				-5.4
oTSS_+_48416.0_DEJ70_RS00250	326	-12.4				2.2
oTSS_+_186087.0_DEJ70_RS01010	284	-28.6				
oTSS_+_1472560.0_DEJ70_RS07120	276	-19				

The top 20 *w*MelPop-CLA asTSS were sorted according to their average expression (Table 11). Of the top 20 asTSS expressed in *w*MelPop-CLA, average expression ranged from 54 to 4986 CPM whilst MFE values ranged from -7.8 to -29.1 at 28°C. Regarding differential expression, 3 of the top 20 *w*MelPop-CLA asTSS displayed evidence of regulation with WsnRNA-59 upregulated 5.67-fc at 21°C and asTSSs targeting WD_RS02715 and the RNA polymerase beta subunit (WD_RS06155) downregulated -2.36 and -2.69 under 34°C exposure respectively.

Table 11) Top 20 expressed *w*MelPop-CLA candidate antisense ncRNA. Top 20 expressed asTSS for candidate antisense ncRNA. "id" indicates TSS type, strand, genomic coordinate, and closest coding sequence. Average CPM values are taken from 28°C 48hr samples (n = 3). MFE predicted from RNAfold server using 100 nts downstream the TSS. Evidence of regulation highlighted by blue (downregulated) or red (upregulated) coloured cells. "TSS Relative" indicates the asTSS position relative to the sense target gene, relative positions are split into 10th percentile segments, "extended" indicates the asTSS is located within 100 nt upstream antisense to the target CDS. Any ncRNA that have been previously reported are noted in "Reported".

Candidate antisense ncRNA	Target	TSS Relative	Avg CPM	MFE	21°C	34°C	Reported
asTSS_-_1246854.0_WD_RS05920	DNA mismatch repair endonuclease MutL		4986	-29.1			
asTSS_-_259390.0_WD_RS01225	major capsid protein	extended	1503	-20.3			(Darby et al., 2014)
asTSS_-_440655.0_WD_RS02035	hypothetical protein	extended	1014	-27.3			
asTSS_-_582226.0_WD_RS02715	major capsid protein	extended	522	-20.4	5.67		WsnRNA-59 (Mayoral et al., 2014, Darby et al., 2014)
asTSS_-_1103869.0_WD_RS05230	hypothetical protein		324	-22.1		-2.36	
asTSS_+_168335.0_WD_RS07050	hypothetical protein	extended	232	-27.6			
asTSS_+_934857.0_WD_RS04405	tRNA-Ala	extended	177	-28.4			
asTSS_+_1059321.0_WD_RS04995	P-type conjugative transfer protein VirB9		126	-17.5			
asTSS_+_956901.0_WD_RS04510	metalloprotease TldD		121	-7.8			
asTSS_+_163736.0_WD_RS00810	Mrp/NBP35 family ATP-binding protein	extended	104	-19.6			
asTSS_-_22945.0_WD_RS06155	DNA-directed RNA polymerase subunit beta/beta'		92	-10.7		-2.69	
asTSS_+_938801.0_WD_RS04425	threonine--tRNA ligase		86	-13.4			
asTSS_+_1263363.0_WD_RS05975	efflux RND transporter permease subunit		83	-15.2			
asTSS_-_220303.0_WD_RS06900	hypothetical protein		83	-17.7			
asTSS_-_1238108.0_WD_RS05865	IS110 family transposase		83	-13.2			
asTSS_-_480828.0_WD_RS02220	type II secretion system protein		68	-13.8			
asTSS_-_582139.0_WD_RS02715	major capsid protein		64	-16			WsnRNA-59 (Mayoral et al., 2014, Darby et al., 2014)
asTSS_+_166797.0_WD_RS00835	oxidoreductase 2-nitropropane dioxygenase	extended	59	-17			
asTSS_-_958066.0_WD_RS04525	uroporphyrinogen-III synthase		55	-15.6			
asTSS_-_374359.0_WD_RS01765	lipoyl synthase	extended	54	-18.9			

The top 20 *w*AlbB asTSS were sorted according to their average expression (Table 12). Of the top 20 asTSS expressed in *w*AlbB, the average expression ranged from 173 to 9909 CPM whilst MFE values ranged from -5.7 to -27.9. Evidence for differential expression was observed for 10 out of the top 20 wAlbB asTSS under at least one stressor.

Table 12) Top 20 expressed *w*AlbB candidate antisense ncRNA. Top 20 expressed asTSS for candidate antisense ncRNA. "id" indicates TSS type, strand, genomic coordinate, and closest coding sequence. Average CPM values are taken from 28°C 48hr samples (n = 3). MFE predicted from RNAfold server using 100 nts downstream the TSS. Evidence of regulation highlighted by blue (downregulated) or red (upregulated) coloured cells. "TSS Relative" indicates the asTSS position relative to the sense target gene, relative positions are split into 10th percentile segments, "extended" indicates the asTSS is located within 100 nt upstream antisense to the target CDS.

Candidate antisense ncRNA	Target	TSS Relative	Avg CPM	MFE	21°C	34°C	Doxy	Rif
asTSS_-_1184959.0_DEJ70_RS05750	DNA mismatch repair endonuclease MutL		9909	-26.2				
asTSS_+_1030608.0_DEJ70_RS04965	F0F1 ATP synthase subunit alpha		754	-20				3.28
asTSS_-_508637.0_DEJ70_RS02510	hypothetical protein		734	-10.3				
asTSS_-_184944.0_DEJ70_RS07175	hypothetical protein	extended	689	-12.1		2.14		-2.41
asTSS_-_167849.0_DEJ70_RS00885	lipoyl(octanoyl) transferase LipB		566	-15.9				
asTSS_-_304163.0_DEJ70_RS01520	aminopeptidase P family protein		487	-24.3				2.08
asTSS_+_1210934.0_DEJ70_RS05880	DUF4815 domain-containing protein		480	-20.8				2.28
asTSS_-_451998.0_DEJ70_RS02235	IS982 family transposase	extended	480	-11.5				
asTSS_+_691989.0_DEJ70_RS03320	nitronate monooxygenase		453	-5.9			-1.91	3.75
asTSS_-_1020134.0_DEJ70_RS04910	AAA family ATPase		451	-17.4				
asTSS_+_422942.0_DEJ70_RS02075	IS481 family transposase	extended	369	-16.1				4.61
asTSS_-_1371079.0_DEJ70_RS06620	major capsid protein		349	-17.2			-1.97	2.27
asTSS_+_1421138.0_DEJ70_RS06860	hypothetical protein		342	-27.9		2.02		
asTSS_-_886197.0_DEJ70_RS04235	hypothetical protein		294	-23.2				
asTSS_-_978100.0_DEJ70_RS04685	CTP synthase		282	-13.2				2.85
asTSS_-_568308.0_DEJ70_RS02755	ankyrin repeat domain-containing protein	extended	213	-9.2				2.93
asTSS_+_1328835.0_DEJ70_RS06440	ankyrin repeat domain-containing protein		210	-9				
asTSS_+_1179204.0_DEJ70_RS05705	RNA polymerase sigma factor RpoD		210	-15.5				
asTSS_-_1271465.0_DEJ70_RS06165	kinase/pyrophosphorylase		190	-7.5				
asTSS_-_532256.0_DEJ70_RS02610	AI-2E family transporter	extended	173	-5.7				

An asTSS opposite the *mutL* gene was the most highly expressed asTSS for both *w*MelPop-CLA and *w*AlbB with an average expression of 4986 and 9909 CPM respectively at 28°C (Figure 27). Both strains predict *mutL* to be operon associated with *purE* directly upstream, *w*AlbB has an additional downstream hypothetical (DEJ70_RS05755) within its predicted operon. Both strains predicted an RNA secondary structure with relatively low MFE values at -29.1 and -26.2 for *w*MelPop-CLA and *w*AlbB respectively (Figure 27). The *w*MelPop-CLA *mutL* asTSS RNA secondary structure displayed bilateral symmetry with 2 distal hairpin loops and 3 central internal loops. The *w*AlbB *mutL* asTSS displayed bilateral symmetry with 1 distal hairpin loop, 4 internal loops, and 1 minor internal bulge.

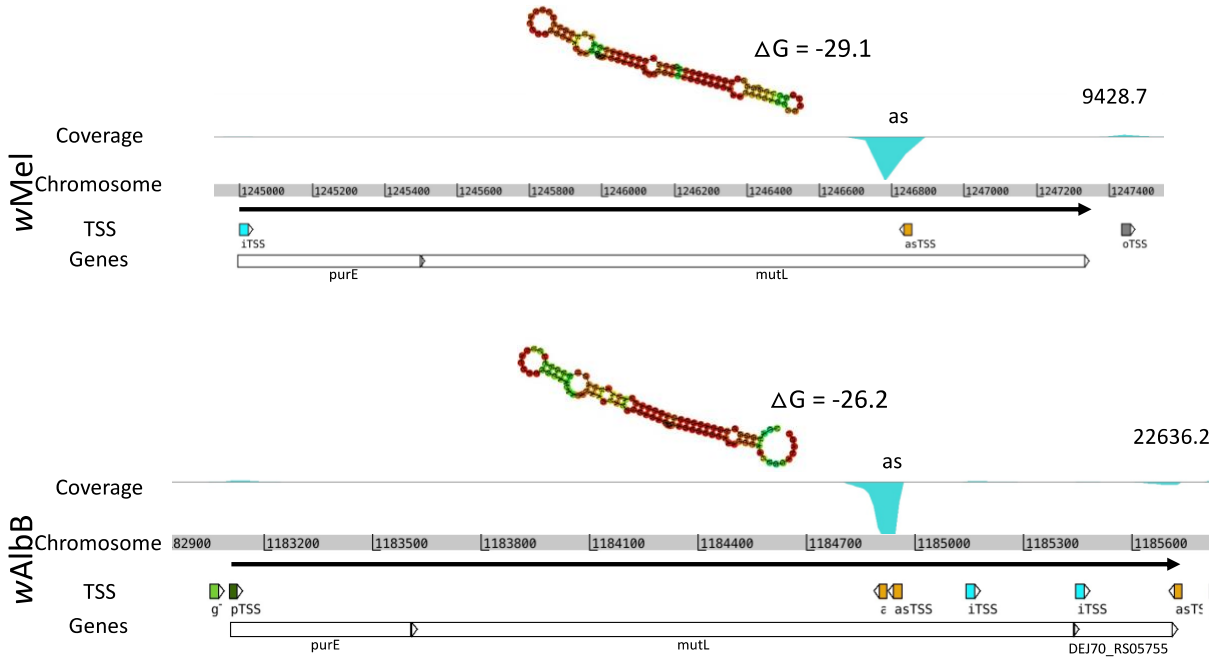


Figure 27) local region of highly expressed *mutL* asTSS in *w*Mel and *w*AlbB. Coverage (value top right) represents read coverage relative to displayed genomic region. Predicted folded RNA secondary structure displayed above associated TSS and coverage peak. TSS of interest highlighted on coverage plot: 'as' asTSS. Black arrow indicates predicted operon.

Both *w*MelPop-CLA and *w*AlbB harboured an asTSS associated to the phage major capsid protein (WD_RS01225 and DEJ70_RS06620 respectively) with an average expression of 1503

and 349 CPM respectively at 28°C (Figure 28 and Tables 11-12). RNA secondary structure predicted a low MFE value relative to randomly generated TSS of -20.3 and -17.2 for *w*MelPop-CLA and *w*AlbB respectively. The structural features of the predicted secondary structure for the *w*MelPop-CLA major capsid asTSS displayed a single hairpin loop at the distal end with 2 internal loops. For the *w*AlbB major capsid asTSS, secondary structure prediction displayed a single distal hairpin loop with several low base-pairing probability internal loops (Figure 28).

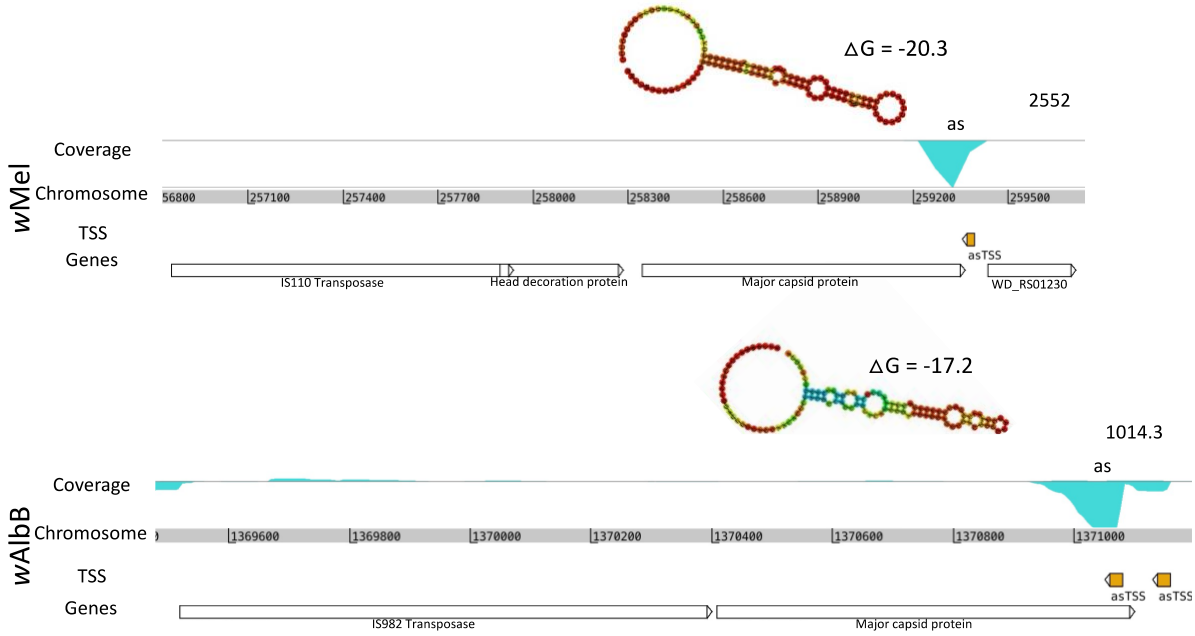


Figure 28) local region of major capsid asTSS in *w*Mel and *w*AlbB. Predicted folded RNA secondary structure displayed above associated TSS and coverage peak. TSS of interest highlighted on coverage plot: 'as' asTSS.

TSS assisted operon structure analysis

Of the 1,294 and 1,406 genes in *w*MelPop-CLA and *w*AlbB respectively, 233 (18%) and 212 (15%) were predicted to be arranged into operons of two or more genes with the use of TSS data in Rockhopper operon prediction tools (Table 13). The majority of operon associated genes were predicted to be situated in 2 gene pairs representing 70% of all operons both

*w*MelPop-CLA and *w*AlbB. The largest predicted operons contained 28 and 23 genes for *w*MelPop-CLA and *w*AlbB respectively. Operons with a confidently associated pTSS accounted for 17% (*n* = 40) and 27% (*n* = 58) of all predicted operons for *w*MelPop-CLA and *w*AlbB respectively (Table 13). The least frequent TSS type associated to predicted operons were the alternate gTSS occurring for 2 and 10 operons in *w*MelPop-CLA and *w*AlbB respectively.

Table 13) Summary of TSS assisted operon prediction. Occurrence of predicted operons with number of associated genes. *w*MelPop-CLA (Pink), *w*AlbB (Blue).

n# Genes in operon	*w* Mel	*w* AlbB
28	1	
23		1
10	1	
9		1
8	1	
7	4	2
6	3	4
5	8	2
4	20	21
3	32	32
2	163	149
Total genes in operons	233	212

Table 14) TSS types associated with TSS assisted operon prediction. Summary of operons with associated TSS types. *w*MelPop-CLA (Pink), *w*AlbB (Blue).

	Operons	pTSS	gTSS	iTSS	asTSS
w Mel	233	40	2	24	13
w AlbB	212	58	10	84	56

The local region of the conserved ribosomal operon of both *w*MelPop-CLA and *w*AlbB was inspected (Figure 29). The ribosomal operon of both strains contained a total of 12 genes involving proteins related to both 30S and 50S ribosomal subunits, translational elongation, secretion, and transcription. Due to the requirement of overlapping reads between gene junctions, TSS reads could not correctly assign the RNA polymerase subunit as the 12^{th} gene due to an intergenic gap separating the RNA polymerase to the 50S L7/12 subunit. A pTSS was identified for both *w*MelPop-CLA and *w*AlbB 35 and 33 nucleotides upstream respectively from the initiation codon of the 30S S12 subunit, average expression was 60 and 751 CPM respectively at 28°C. The *w*MelPop-CLA ribosomal operon contained a total of 17 TSS consisting of 1 pTSS, 14 iTSS, and 2 asTSS. The *w*AlbB ribosomal operon contained a total of 30 TSS consisting of 2 pTSS, 24 iTSS, and 4 asTSS. The additional pTSS of the *w*AlbB ribosomal operon is situated 10 nucleotides upstream the elongation factor Tu and downstream elongation factor G. The highest TSS expression originated from the iTSS located in the 50S L11 gene of both *w*MelPop-CLA (WD_RS00105) and *w*AlbB (DEJ70_RS01790) at 234 and 1,564 CPM respectively at 28°C.

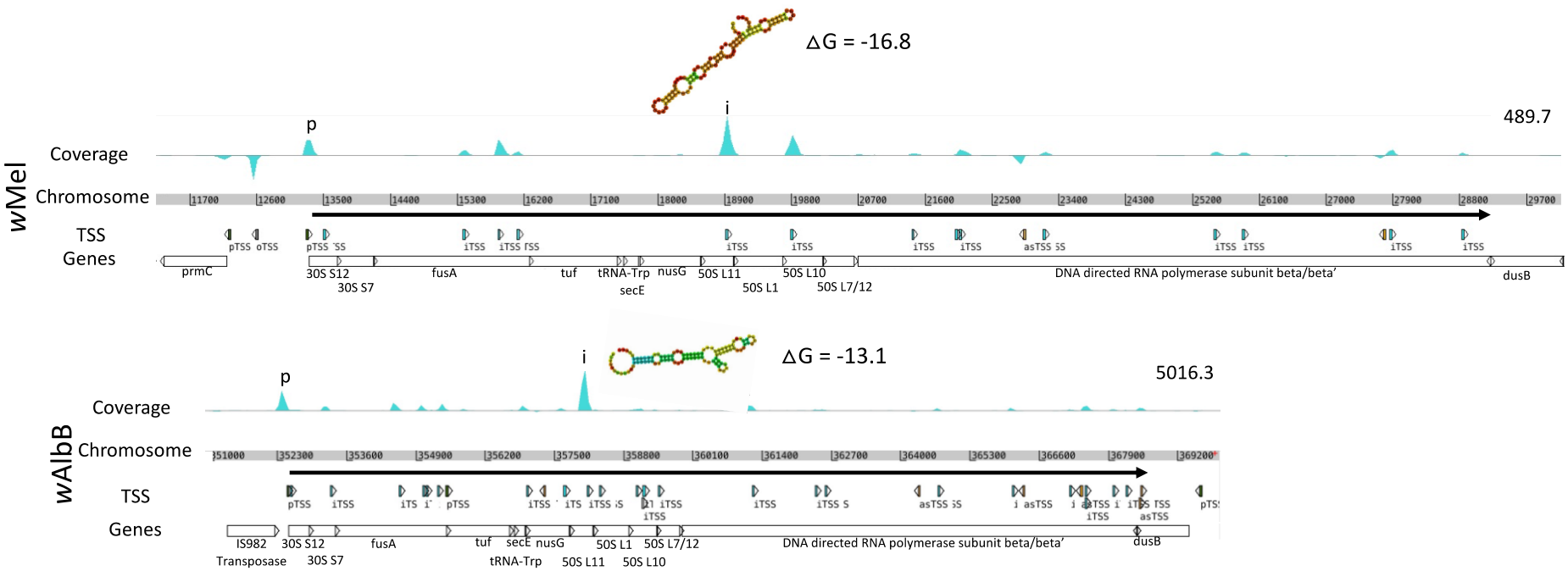


Figure 29) Local region of the 30S/50S ribosomal operon of *w*Mel and *w*AlbB. Coverage (value top right) represents read coverage relative to displayed genomic region. TSS of interest highlighted on coverage plot: 'p' pTSS, 'i' iTSS. Black arrow indicates predicted operon.

The local region of the conserved oxidoreductase operon of both *w*MelPop-CLA and *w*AlbB was inspected (Figure 30). The oxidoreductase operon of both strains contained a total of 6 genes involving proteins related to oxidative phosphorylation and biotin metabolism. A pTSS was identified for both *w*MelPop-CLA and *w*AlbB at approximately 22 and 23 nucleotides upstream respectively from the initiation codon of the NADH-quinone oxidoreductase subunit J, average expression was 9 and 78 CPM respectively at 28°C. The *w*MelPop-CLA oxidoreductase operon contained a total of 4 TSS consisting of 1 pTSS, 2 iTSS, and 1 asTSS. The *w*AlbB oxidoreductase operon contained a total of 7 TSS consisting of 1 pTSS, 5

iTSS, and 1 asTSS. The highest expressed TSS originated from the iTSS located within the *nuoL* gene of both *w*MelPop-CLA (WD_RS04365) and *w*AlbB (DEJ70_RS02705) at 15 and 92 CPM respectively at 28°C.

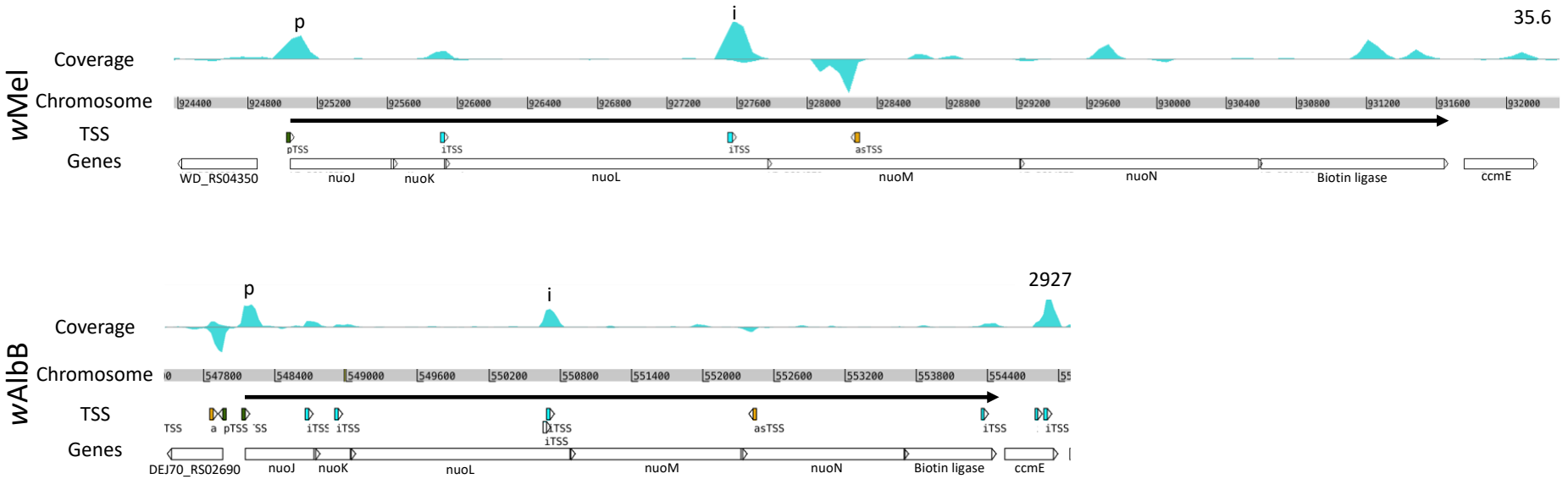


Figure 30) Local region of the oxidoreductase operon of *w*Mel and *w*AlbB. Coverage (value top right) represents read coverage relative to displayed genomic region. TSS of interest highlighted on coverage plot: 'p' pTSS, 'i' iTSS. Black arrow indicates predicted operon.

The local region of the T4SS operon that contains 8 genes from $tRNA^{Leu}$ – *wspB* (T4SS-1) was inspected for both *w*MelPop-CLA and *w*AlbB (figure 31). A pTSS was identified for both *w*MelPop-CLA and *w*AlbB approximately 37 nucleotides upstream the initiation codon of $tRNA^{Leu}$, average expression was 94 and 1703 CPM for *w*MelPop-CLA and *w*AlbB respectively at 28°C. The *w*MelPop-CLA T4SS-1 operon contains a total of 8 TSS consisting of 2 pTSS, 2 iTSS, and 2 asTSS. The second pTSS for *w*MelPop-CLA is located between the intergenic region of *virD4* and *wspB* approximately 20 nucleotides upstream the *wspB* initiation codon. The *w*MelPop-CLA *wspB* associated pTSS is expressed at 81 CPM at 28°C. The *w*AlbB-CLA T4SS-1 operon contains a total of 19 TSS consisting of 3 pTSS, 8 iTSS, 7 asTSS, and 1 pseudogene associated pTSS (pTSSps). The 3 pTSS associated to the *w*AlbB T4SS-1 operon are upstream to $tRNA^{Leu}$, *virB8*, and *virD4* expressed at 1703, 83, and 12 CPM

respectively at 28°C. The pTSSps associated to the pseudogenized *wspB* gene was 20 nucleotides upstream the *wspB* initiation codon expressed at 60 CPM at 28°C.

The local region of the CI two gene operon was inspected for *w*MelPop-CLA and *w*AlbB (Figure 32). The CI operon contains 2 genes, *cifA* and *cifB*, in which TSS analysis of both genomes detected a single pTSS 25 and 32 nucleotides upstream the *cifA* initiation codon for *w*MelPop-CLA and *w*AlbB respectively. The *w*MelPop-CLA CI operon contained a total of 4 TSS consisting of 1 pTSS and 3 iTSS. The *w*AlbB CI operon contained a total of 7 TSS consisting of 1 pTSS, 2 iTSS, and 4 asTSS. The expression of the CI operon pTSS at 28°C was 593 and 439 CPM for *w*MelPop-CLA and *w*AlbB respectively.

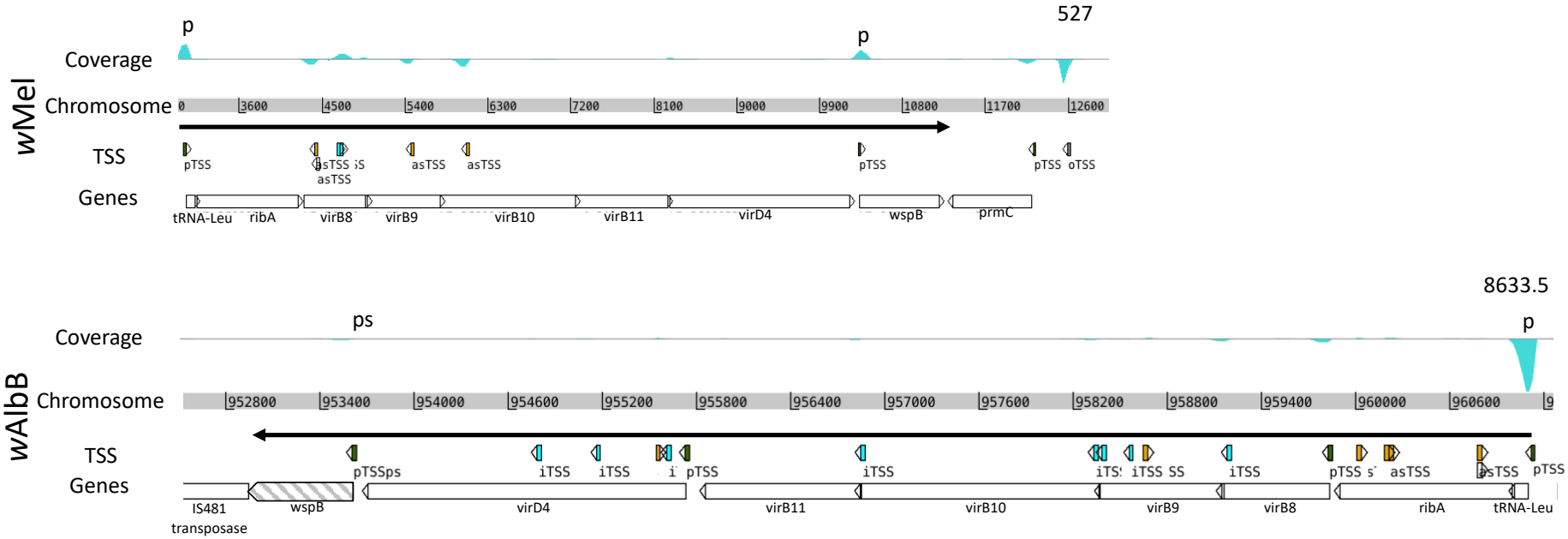


Figure 31) Local region of the T4SS-1 operon of *w*Mel and *w*AlbB. Coverage (value top right) represents read coverage relative to displayed genomic region. TSS of interest highlighted on coverage plot: 'p' pTSS, 'ps' pTSSps. Black arrow indicates predicted operon.

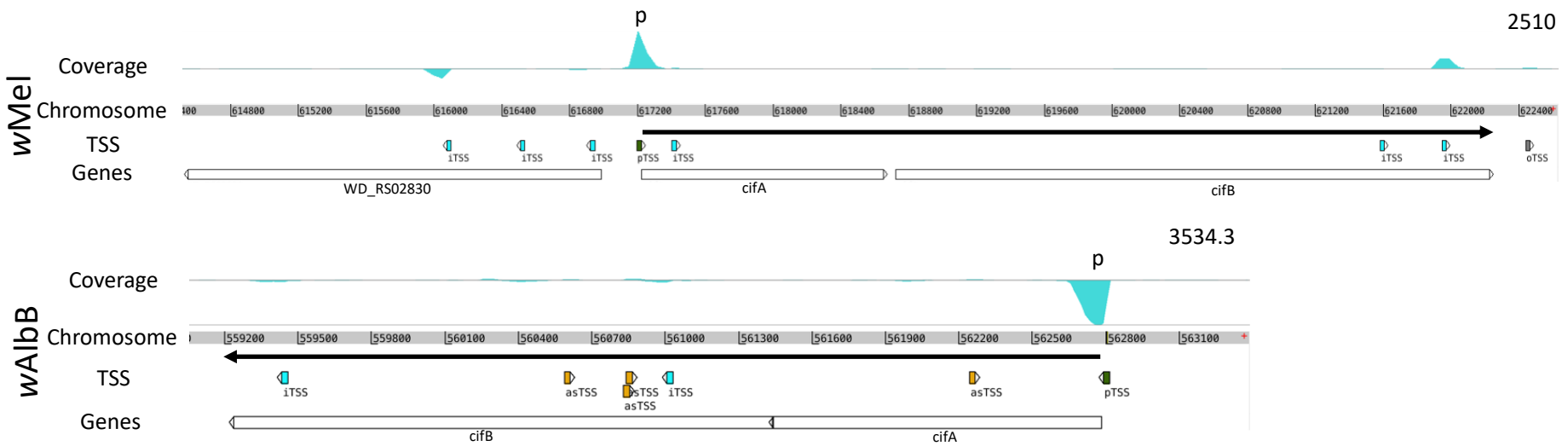


Figure 32) Local region of the CI operon in *w*Mel and *w*AlbB. Coverage (value top right) represents read coverage relative to displayed genomic region. TSS of interest highlighted on coverage plot: 'p' pTSS. Black arrow indicates predicted operon.

The local region of the response regulator (RR) transcription factor for *w*MelPop-CLA (WD_RS03320) and *w*AlbB (DEJ70_RS01625) was inspected (figure 33). A pTSS was detected for the RR of both *w*MelPop-CLA and *w*AlbB 31 nucleotides upstream the initiation codon, the pTSS were expressed at 147 and 1562 CPM respectively at 28°C. A total of 3 TSS were detected for the *w*MelPop-CLA RR consisting of 1pTSS, 1 iTSS, and 1 asTSS. For the *w*AlbB RR a total of 6 TSS were detected consisting of 1 pTSS, 2 gTSS, and 3 iTSS. The gTSS directly upstream the pTSS of the RR were located 155 and 75 nucleotides upstream from the initiation codon and expressed at 15 and 55 CPM for *w*MelPop-CLA and *w*AlbB respectively at 28°C. The iTSS directly downstream the pTSS was located exactly 18 nucleotides downstream the initiation codon and expression at 28°C was 16 and 48 CPM for *w*MelPop-CLA and *w*AlbB respectively.

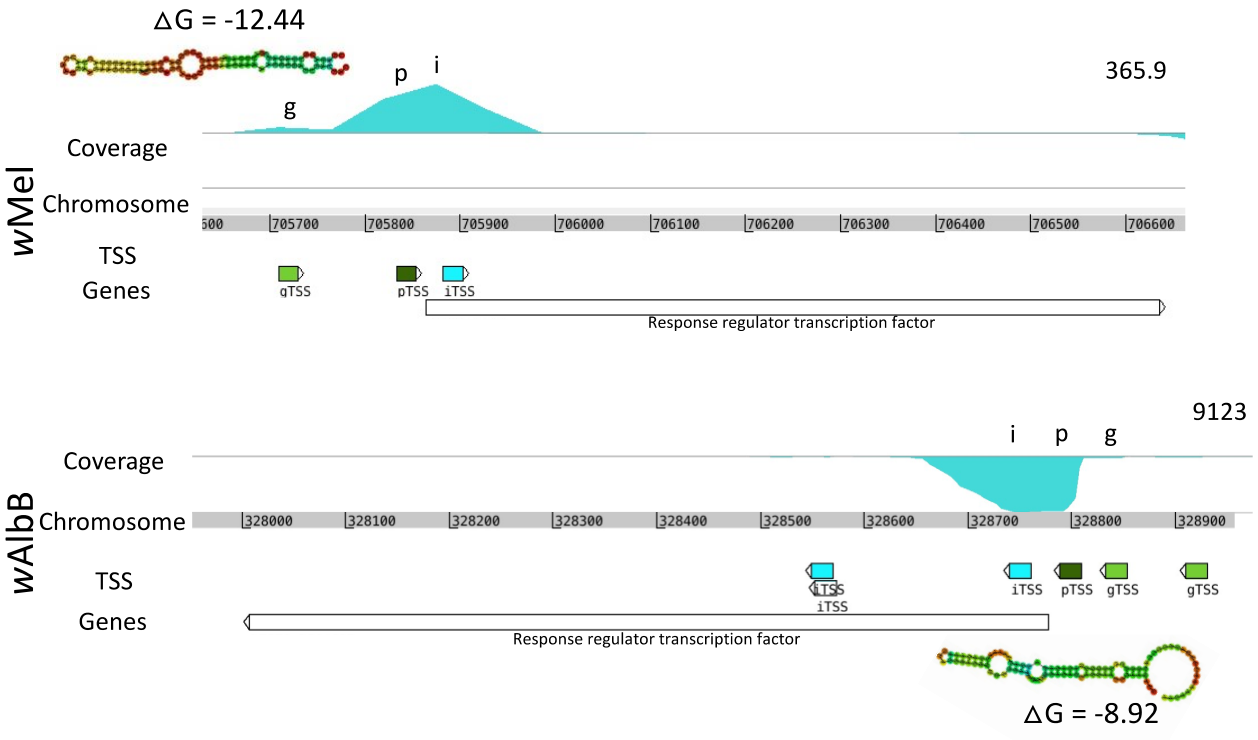


Figure 33) Local region of the response regulator gene in *w*Mel and *w*AlbB. Coverage (value top right) represents read coverage relative to displayed genomic region. TSS of interest highlighted on coverage plot: 'i' iTSS, 'p' pTSS, 'g' gTSS. Black arrow indicates predicted operon.

Discussion

Probability of RNA secondary structures within non-canonical TSS

An advantage that capable-seq enables by capturing the 5' end of nascent transcripts is that small RNA, intragenic transcripts, and potential ncRNA signals are not hidden amongst the sequence data as seen in conventional RNA-seq analysis. One of the aims of the current study was to use the 5' retaining ability of cappable-seq to detect the presence of ncRNA on a global scale. Once all detected TSS had been categorized based on their location relative to neighbouring genes, non-canonical TSS (gTSS, iTSS, asTSS, oTSS) were further inspected for their ncRNA potential. The catalytic and regulatory functions of ncRNA are based on the interaction between molecular structures such as RNA-RNA and RNA-protein. Using RNA secondary structure prediction tools such as RNAfold, MFE values of non-canonical TSS could be analysed and compared to predict the likelihood of ncRNA within *w*MelPop-CLA and *w*AlbB. Since MFE is based on thermodynamic stability (Zuker and Stiegler, 1981), the RNA structures with the lowest MFE should predict the most stable and thus likely structure.

The initial hypothesis was for iTSS and oTSS transcripts to have an average lower MFE value than pTSS if they were more likely to form secondary structures. Since any sufficiently long RNA sequence may fold into a structure resembling an RNA secondary structure with hairpin loops, it was important to normalise nucleotide length and compare values against a randomised group of transcripts within the same genome. For both *w*MelPop-CLA and *w*AlbB a randomly generated set of 2000 transcripts (1000 per strand) were used for a comparative control. The current study found that for both *w*MelPop-CLA and *w*AlbB, iTSS and oTSS transcripts on average had statistically lower MFE values, and thus greater thermodynamic stability, than randomly generated transcripts. In contrast, the secondary structure of pTSS transcripts were not statistically different from random transcripts for both *w*MelPop-CLA and *w*AlbB suggesting a less likely probability of functional secondary structures from primary transcripts. This outcome supported the categorization of oTSS as candidate intergenic ncRNA and the further investigation of iTSS for secondary structure analysis. In a similar manor to pTSS, transcripts associated to asTSS also did not produce significantly different MFEs to random transcript, this outcome also supported the categorization of asTSS as candidate cis-acting ncRNA that may function by complementary base pairing instead of forming functional secondary structures.

High intragenic expression within a glycolytic enzyme

Of particular interest throughout the study was the presence of a highly expressed iTSS located within the *fsa* gene of both *w*MelPop-CLA and *w*AlbB. The *fsa* iTSS was consistently in the top two most expressed TSS for both *w*MelPop-CLA and *w*AlbB under all tested conditions. Although no evidence of differential expression was observed, the transcriptional resources used to express the iTSS *fsa* was striking. The location of the *fsa* iTSS was identical for both *w*MelPop-CLA and *w*AlbB suggesting high conservation and thus functional retention. The *fsa* gene supposedly encodes fructose-6-phosphate aldolase capable of catalysing the reversible formation of fructose 6-phosphate (F6P) from glyceraldehyde 3-phosphate and sedoheptulose 7-phosphate (Schürmann and Sprenger, 2001). Although the chemical reaction of FSPA is understood it's physiological role has yet to be determined (Schürmann and Sprenger, 2001). However, for F6P to be utilized for glycolysis requires phosphorylation to F16BP by phosphofructokinase which is absent in *Wolbachia.* For the F6PA protein to be functional it requires translation of its active site. The highly expressed iTSS of *fsa* is located downstream from the active site which suggests that the translated product of the *fsa* iTSS transcript is incapable of contributing to F6P synthesis.

When analysing the 100 nt downstream sequence of *fsa* iTSS, a low MFE secondary structure was predicted for both *w*MelPop-CLA and *w*AlbB. It's possible that the intended product of *fsa* is not a protein but instead a ncRNA to which its function is unknown. A limitation in the current study is that the 3' end of transcripts is not included in the analysis and prevents identification of the full-length transcript. The full length of the *fsa* iTSS transcript could be shorter or longer resulting in varying RNA folding structures. The observation that both *w*MelPop-CLA and *w*AlbB harbour a highly expressed *fsa* iTSS in identical locations brings forth further questions to be addressed; Do filarial *Wolbachia* also harbour a highly expressed *fsa* iTSS and what could the functional role of the *fsa* iTSS be? Performing immunostaining on *fsa* in a similar manor to wsp may show its localization and potential involvement in host-symbiont communication if expressed in specific host tissues.

Regulation of previously detected intergenic ncRNA upon thermal stress

The 89 and 356 oTSS detected in the present study for *w*MelPop-CLA and *w*AlbB respectively were further inspected for potential ncRNA features. Of the 89 oTSS detected for *w*MelPop-CLA, the top 20 most expressed oTSS successfully detected two known *Wolbachia* ncRNAs. The *w*MelPop-CLA oTSS that were detected at genomic positions 1039620 (+ strand) and 587738 (+ strand) were previously detected by Woolfit, *et al.* (2015). The *w*MelPop-CLA oTSS at position 1039620 was identified as *ncrwMel02* and was the most highly expressed oTSS in *w*MelPop-CLA throughout the study. The initial study that detected *ncrwMel02* (Woolfit *et al.*, 2015) found that the *w*Mel *ncrwMel02* was upregulated in testes compared to ovaries which the authors suggest might have a role in host reproductive manipulation. The current study observed that *ncrwMel02* was downregulated when exposed to temperatures of 21°C and 34°C conditions. If *ncrwMel02* is indeed involved in host reproductive manipulation, downregulation in response to temperature extremes may negatively impact *Wolbachia*-vector release strategies. Future research should determine the functional role of *ncrwMel02* to better comprehend its implications. In addition to *ncrwMel02*, the small ncRNA *ncrwMel01* was also detected in the current study. The small ncRNA *ncrwMel01* was detected at genomic position 587738 (+ strand) but did not show evidence of differential regulation under thermal stress. The functional role of *ncrwMel01* has also not been determined. Another small ncRNA detected in both the present study and by Woolfit, *et al.* (2015) was located at genomic position 67860 (- strand) but also did not show evidence of regulation under thermal stress. The detection of three small ncRNAs that have previously been reported emphasises the ability of cappable-seq to combine transcriptomic information that would usually require both 5'RACE and RNA-seq methodologies.

The study conducted by Woolfit, *et al.* (2015) used intergenic regions of conservation between *w*Mel and *w*Pip, therefore the present study may also detect additional ncRNA that could be unique to the *w*Mel lineage. In addition to the detection of previously identified ncRNA, the current study found a regulated wMelPop-CLA oTSS at genomic position 915471 (-) which displayed upregulation under 34°C thermal stress. The combined characteristics of intergenic location, upregulation, relatively high expression, and energetically favourable secondary structure confidently places this TSS as a strong candidate ncRNA within *w*MelPop-CLA.

Further experiments utilizing reproductive tissue specific RT-qPCR may further reveal whether this highly expressed candidate ncRNA has potential for host communication. The study found that 12 of the 20 top expressed candidate intergenic ncRNA in *w*MelPop-CLA displayed regulation under thermal stress. These regulatory observations imply that multiple *Wolbachia* ncRNAs are sensitive to thermal fluctuations emphasizing the need to further investigate their functional role and potential impact for *Wolbachia*-vector release strategies.

Intergenic candidate ncRNAs were also detected in *w*AlbB amongst the 356 oTSS discovered. In terms of regulatory capacity, it is interesting to note that none of the top 20 *w*AlbB intergenic ncRNA candidates were regulated under thermal stress. However, rifampicin exposure resulted in regulation of 13 candidate ncRNA 10 of which were upregulated. Since rifampicin functions to inhibit RNA synthesis, it is surprising to observe upregulation of ncRNA in contrast to the global downregulation observed for the majority of *w*AlbB pTSS. Lack of regulation amongst the top 20 *w*AlbB ncRNA in response to thermal stress may reflect its thermal resilience relative to *w*MelPop-CLA. Comparing conserved intergenic regions between *w*Mel and *w*AlbB may reveal conserved ncRNA between supergroups A and B respectively. It would be of interest to further expand the analysis to regions conserved between both filarial and arthropod *Wolbachia* with the goal of revealing highly conserved and thus functionally important ncRNA. As stated previously, a limitation of the current study is the lack of 3' end identification thus preventing the full length of each ncRNA from being observed. Although MFE values for RNA secondary structures can support the identification of ncRNA, predictions based on MFE alone should be interpreted with caution. Functional ncRNA may have multiple conformations that are condition specific. Secondary structure prediction software may not detect non-canonical structures which are limited by the databases on which they depend.

The presence of ncRNA in obligate intracellular bacteria are rare but not totally absent as streamlined genomes would appear to eliminate non-functional intergenic regions. Using a similar 5'P-dependent exonuclease RNA-seq approach, *Chlamydia trachomatis* was found to harbour two small ncRNA within intergenic regions (Albrecht *et al.*, 2009). Like *C. trachomatis, Wolbachia* lack the small RNA-binding protein Hfq. It may still be possible that a functional homolog of Hfq may be present or that *Wolbachia* small ncRNA do not require additional chaperones to function with their targets. Under a greater variety of conditions, it's possible that both *w*AlbB and *w*MelPop-CLA harbour additional intergenic ncRNA yet to be

detected. Both *w*AlbB and *w*MelPop-CLA contain and regulate ncRNA but to what extent they impact symbiont and host biology remains to be seen.

The *mutL* duplication harbours a highly expressed antisense ncRNA

In addition to oTSS as candidate intergenic ncRNA, asTSS were highlighted as prime candidates for cis-acting antisense RNA. A total of 173 and 1,126 asTSS were discovered in *w*MelPop-CLA and *w*AlbB respectively, the top 20 highly expressed asTSS were further examined as candidate antisense ncRNA. Surprisingly, an asTSS was amongst the most highly expressed transcripts for both *w*MelPop-CLA and *w*AlbB. This TSS was associated with the *mutL* gene encoding the DNA mismatch repair protein MutL. The function of MutL is to bind to the MutS recognition protein and recruit MutH, thus invoking incision of the DNA mismatch (Li, 2008) critical for minimising mutation rates. *Wolbachia* was the first known bacterium to harbour two *mutL* paralogs (Wu et al., 2004), one of which has been associated with the eukaryotic association module (EAM) of bacteriophage origin (Bordenstein and Bordenstein, 2016). The *mutL* gene in question is not the EAM-associated copy but part of a two- or three-gene operon, for *w*Mel and *w*AlbB respectively, starting with purE.

Studies on regulatory antisense RNAs are still in their infancy and their mechanisms of action are varied. Possible mechanisms involve the targeted degradation of genes bound to the cis-encoded antisense RNAs by base complementarity. The antisense-sense dsRNA complex then recruits particular RNases, such as dsRNA-specific RNase III for degradation, transcriptional termination of the sense mRNA, and translational inhibition (Thomason and Storz, 2010). Given the current understanding of cis-antisense RNAs, if the *mutL* asTSS represents a cis-antisense RNA that is capable of inducing degradation/translational inhibition of the mRNA strand, translational inhibition of *mutL* may permit the accumulation of mutations and thus increase the mutation rate as seen in Δ*mutL* E. coli mutants (Lenski, Sniegowski and Gerrish, 1997). Also, a defective MutL protein may prevent DNA repair by homologous recombination. A possible RNA secondary structure was predicted within the antisense sequence of the *mutL* gene for both *w*MelPop-CLA and *w*AlbB which may act in cis or trans locations. The *w*AlbB *mutL* asTSS was 43-fold higher (9,565 average CPM) than its associated pTSS (101 CPM), which supports the possibility that the antisense transcripts may be negatively regulating the pTSS transcript via targeted degradation. The pTSS to the other *mutL* *w*AlbB copy was detected

(DEJ70_RS07000) albeit with a significantly lower expression value. An associated pTSS for the *w*MelPop-CLA *mutL* was not detected which could be due to lower sequencing depths inadequate enough to detect a negatively regulated pTSS, or that the half-life of the dsRNA is too short for detection, or that a *mutL* pTSS is legitimately not present. Also, a pTSS was not detected for the other *mutL* copy in *w*MelPop-CLA which seems to suggest that only the antisense transcript of *mutL* is expressed and/or functional.

Mutations in *mutL* have also been shown to increase conjugal recombination frequencies between E. coli and S. typhimurium by 1,000-fold (Matic, Rayssiguier and Radman, 1995). Enhanced permissive integration of non-native DNA may explain the unusually high abundance of transposons within both the *w*Mel and *w*AlbB genomes and the existence of WO phage and the EAM. The combined effect of an increase in mutation rate and recombinational tolerance may promote the evolution of innovative adaptions, thus supporting the widespread establishment of *Wolbachia* amongst its numerous hosts and phenotypic diversity, albeit at the expense of genomic stability. Alternatively, the *mutL* asTSS may encode an antisense RNA that instead stabilizes the sense *mutL* transcript; however, the low abundance of the associated pTSS does not favour such a model in *Wolbachia*. Experimental validation, which awaits genetic tractability, would be needed to assess the true function of these transcripts.

Antisense regulation of phage and mobile genetic elements

A surprisingly high level of expression was observed in *w*MelPop-CLA for a pTSS associated phage tail protein (WD_RS02845, 3/841). In contrast, no pTSS related to phage exhibited high expression in *w*AlbB. This disparity in phage expression between *w*MelPop-CLA and *w*AlbB is not entirely unexpected, as the *w*AlbB phage is predicted to be inactive due to pseudogenization of essential phage components that are intact in *w*Mel (Sinha *et al.*, 2019). Interestingly, an asTSS was also amongst the top 20 expressed TSS in *w*MelPop-CLA, which was associated with the phage major capsid protein. Expression of phage tail and antisense transcript of the major capsid protein were both previously observed by Darby, *et al.* (2014), who conducted the first global gene expression analysis of *w*MelPop-CLA in the RML-12 cell line. The agreement between these two studies for the expression of phage-related transcripts supports the suggestion that *w*MelPop-CLA expresses phage at basal conditions. However, the presence of an antisense transcript of a phage component suggests that *w*MelPop-CLA may be actively supressing phage activity by blocking or degrading phage transcript. This hypothesis

assumes that the antisense transcript is working in accordance with antisense-sense complementary degradation. The use of asRNA to repress the lytic cycle of phage has been observed in *Salmonella*, in which the lytic phase of the P22 phage is repressed by the C2 protein, which blocks transcription of proteins required for lytic cycle development. Lytic growth of the P22 phage is activated by the anti-repressor protein Ant which prevents C2 repression. However, the asRNA Sar can prevent the lytic cycle by blocking the *ant* ribosome binding site, inhibiting the translation of *ant* and thus preventing the lytic cycle of the P22 phage (Schaefer and McClure, 2021). *Salmonella enterica* has also been found to harbour the antisense ncRNA *isrg* expression of which was inversely correlated with a phage tail component (Padalon-Brauch *et al.*, 2008). The asRNA discovered here may indicate a strategy of lytic phage repression by inhibiting the translation of phage structural components via asRNA-mRNA complimentary binding; the additional presence of the dsRNA-specific RNase III in both *w*MelPop-CLA and *w*AlbB theoretically permits such a strategy.

Antisense transcripts were also detected for another major capsid protein (WD_RS02175) located at genomic regions 582226 and 582139 on the negative strand. These antisense transcripts have previously been detected and given the name *WsnRNA-59* (Darby *et al.*, 2014; Mayoral, *et al.*, 2014). The study conducted by Mayoral, *et al.* (2014) found two antisense ncRNA (*WsnRNA-46* and *WsnRNA-59*). Both ncRNA were present in mosquito fractions purified of *Wolbachia*, *WsnRNA-46* was shown to downregulate its complementary target gene *murD* and able to upregulate the host-encoded dynein heavy chain gene *dhc*. As *Wolbachia* have been reported to use host microtubules and dynein to localise to the anterior pole of developing oocytes (Ferree *et al.*, 2005), *Wolbachia* ncRNA like *WsnRNA-46* may be key to the *Wolbachia*-host relationship. Given the known mechanism of complementary antisense downregulation, it's possible that *WsnRNA-59* is also capable of downregulating its complementary major capsid target gene in a similar fashion to *WsnRNA-46* with *murD*. The present study also found upregulation of *WsnRNA-59* at 21°C. As *WsnRNA-59* was reported to be overexpressed in mosquito ovaries, temperature fluctuations could impact host-symbiont communication if the role of *WsnRNA-59* is related to host reproduction. Injecting *Wolbachia* infected insect cell lines with an expression vector, as was done for *WsnRNA-46*, may detect *WsnRNA-59* regulated host transcripts and shed light on potential function.

Cold-activated phage have been studied in the past, phage generally maintain their lysogenic state in colder temperatures (Shan *et al.*, 2014; Egilmez *et al.*, 2018). If antisense transcripts in *Wolbachia* function by antisense-sense complementary binding to the genes in which they overlap, it is conceivable that *Wolbachia* uses temperature dependant antisense transcripts to regulated phage expression and repress lytic activation. Since WO-B has been suggested to be inactive/defective, it is possible that *W*snRNA-59 has evolved an alternative role in regulating non-phage related aspects of *Wolbachia* biology such as host communication under stressed conditions. Conservation of *W*snRNA-59 within different *Wolbachia* supergroups appears to be restricted to those of supergroup A hence the absence of such a transcript in *w*AlbB (Mayoral, Hussain, Joubert, Iturbe-Ormaetxe, O'Neill, *et al.*, 2014). The pseudogenized *w*MelPop-CLA phage portal protein (WD_RS01205) also exhibited antisense upregulation (4.7-fold) at 21°C 24hr. In contrast to *w*MelPop-CLA, *w*AlbB also harboured an antisense associated phage portal protein but which displayed upregulation (2.4-fold) at 34°C 48hr exposure. Since *w*AlbB has an inactive phage, the presence of phage suppressive transcripts may be an evolutionary remnant derived from the *w*Mel-*w*AlbB ancestor or capable of an altogether unknown different function. Phage portal proteins are key initiators of capsid assembly (Cuervo and Carrascosa, 2012), antisense transcripts targeting such genes may support *w*MelPop-CLA suppression of phage.

In addition to phage components, antisense transcripts associated to transposable elements were detected in both *w*MelPop-CLA and *w*AlbB. Antisense transcripts were associated to a transposase of the IS110 family (WD_RS05865) in *w*MelPop-CLA and IS481 (DEJ70_RS02075) and IS982 (DEJ70_RS02235) in *w*AlbB. Only the IS481 transposase in *w*AlbB was found to be downregulated under rifampicin exposure (2.6 fold) whilst the previously mentioned transposase associated antisense were constitutively expressed. Constitutive expression of transposase associated antisense ncRNA may be a strategy in which *Wolbachia* minimise transposase induced genomic disruption. Whilst transposases can act as a source of adaptive variability, *Wolbachia* must balance adaptive potential against genomic instability which may in part rely on antisense repression. Expression of the antisense snRNA *isrN* in *S. enterica* has been associated with steady state expression to its associated transposase STM2765 (Padalon-Brauch *et al.*, 2008).

Intragenic TSS reveals sub-operon architecture in *Wolbachia*

Another goal of the present study was to improve the understanding of operon structure in *Wolbachia.* Genes within an operon usually share similar functions such as the *Lac* operon encoding proteins involved in lactose metabolism (Jacob and Monod, 1961; Jacob *et al.*, 2005). Genes within operons are usually transcribed together (polycistronic) or individually (monocistronic). With the advent of high throughput next generation sequencing, alternative transcriptional units (ATUs) are redefending conventional perspectives on operon structure. The conventional distinction between an operon and an ATU is that a gene can be transcribed by multiple transcriptional units but belong to only one operon (Mejía-Almonte *et al.*, 2020). The present study applied the categorization of TSSs in combination with operon prediction tools to shed light on the operon architecture of *w*MelPop-CLA and *w*AlbB under varying stress conditions. Internal transcriptional start sites within operons supports ATUs in the form of sub-operons within *Wolbachia* genomes. The present study supports the distinction suggested by Mejía-Almonte, *et al.* (2020) to distinguish between 'simple' and 'complex' operons, simple operons referring to those with a single transcriptional unit whereas complex operons are those with multiple ATU.

Using the bioinformatic operon prediction tool Rockhopper, a total of 558 and 634 predicted polycistronic operons with a minimum of two genes were precited for *w*MelPop-CLA and *w*AlbB respectively. Of these predicted operons 40 and 58 indicated expression via a primary pTSS which assumes full length polycistronic operon expression. Interestingly evidence for 40 and 58 complex operons harbouring internal iTSS expression was found for *w*MelPop-CLA and *w*AlbB respectively. Further inspection of highly conserved operons was compared between the two strains, the first operon to be compared was the 12 gene 30S/50S ribosomal operon. However, the use of TSS in combination with Rockhopper operon prediction is not ideal. Intergenic junctions may not be able to be bridged with small fragment TSS data and may require either full length trabscript data or conventional RNA-seq to bridge such gaps. An example of the weakness in TSS assisted operon structure is that the ribosomal operons only included 11 genes when they have 12, this is because TSS was unable to bridge the junction between the 12^{th} gene due to an intergenic gap greater in length than the TSS containing transcript separating the RNA polymerase to the 50S L7/12 subunit.

Wolbachia ribosomal operons contain ATUs

Both strains successfully identified a pTSS upstream from the 30S S12 gene which based on operon prediction ends with the beta-subunit of the DNA-directed RNAP. Interestingly both operons display an internal iTSS located upstream to the 50S L1 ribosomal subunit and internal to the 50S L11 with higher expression than the associated pTSS. Downstream the iTSS are genes for the 50S subunits and RNAP beta subunit whilst those upstream the iTSS are for the 30S subunits, translation elongation, and protein secretion. Assuming the primary pTSS of the ribosomal operon and the internal iTSS within the 50S are separate transcriptional units, the operon can produce at least 2 distinct ATUs. Ribosomal ATUs have been detected in previouse studies using SMRT-cappable-seq which enables capture of the full length transcript from the TSS to the transcription termination site (Yan *et al.*, 2018). The study conducted by Yan, *et al.* (2018) found that the *rplB* gene which encodes the 50S L2 subunit was present in 13 different ATUs when grown in Rich media compared to 3 in M9 media, genes more common in multiple ATUs were involved in translation suggesting that ATUs are regulated by environmental conditions and functional plasticity. Unlike SMRT-cappable-seq, standard cappable-seq can only retrieve the TSS without knowledge of the TTS, the present study is still able to present evidence for the TSS of such ATUs to assist full length verification in future studies. *Wolbachia* may utilize ATUs depending on the requirements of the cell. The ATUs detected in the ribosomal operon may indicate a strategy to preferentially express genes involved in translation (pTSS downstream ribosomes) or transcription (iTSS downstream RNAP). Processing of the ATUs could be attributed to specific endoribonucleases such as RNase III which is capable of cleaving hairpin-loops (Oliva, Sahr and Buchrieser, 2015), RNA structures resembling classic hairpin loops were detected within the 100 nt downstream the iTSS which may represent a viable cleavage site. The ATUs may be simultaneously transcribed within a single cell or by specific sub-populations dependant on environmental condition or cellular location.

ATUs within an oxidoreductase operon may regulate biotin synthesis

Another homologous operon in which its structure was investigated was the 6 gene proton-translocating oxidoreductase operon involved in the respiratory electron transport chain. The operon contains the NADH-quinone oxidoreductase subunits *nuoJKLMN* ending with a biotin ligase. Both *w*MelPop-CLA and *w*AlbB operons contain a pTSS upstream the *nuoJ* subunit

whilst also containing a prominent iTSS within 3' end of the *nuoL* gene. NADH-quinone oxidoreductase is involved in the first step of the electron transport chain for oxidative phosphorylation with the end goal of ATP synthesis. The importance of symbiont derived B vitamins in insects with restricted diets such a blood has been well established, for example the tsetse fly *Glossina morsitans* is dependent on its obligate symbiont *Wigglesworthia* for B6 (Michalkova *et al.*, 2014) and the planthopper *Laodelphax striatellus* is supplemented by its *Wolbachia* symbiont with biotin for egg laying fecundity (Ju *et al.*, 2019). Biotin ligases are known to have bifunctional properties involving the enzymatic attachment of biotin to its cognate enzyme and repression of the biotin operon by binding to the biotin operon operator (Sarah K. Henke and John E. Cronan, 2016).

It is interesting that a protein capable of biotin suppression should be transcribed with those involved in ATP synthesis. Biotin synthesis is energetically costly since up to 19 molecules of ATP are required for synthesis of a single biotin molecule (Feng, Zhang and Cronan, 2013). The oxidoreductase operon in *w*MelPop-CLA and *w*AlbB may encode two ATUs which could balance the downstream synthesis of ATP with the ATP intensive synthesis of biotin. In *E. coli* the biotin ligase BirA binds to biotinoyl-adenylate which is a preceding product of biotin synthesis, bound adenylate then triggers a BirA monomer to dimer transition which enables DNA-binding activity of BirA to the biotin operator (Streaker, Gupta and Beckett, 2002). Under conditions of high biotin concentration BirA inhibits transcription of the biotin synthesis operon, conversely under low biotin concentrations adenylate is synthesised in low amounts and consequentially unable to promote the BirA DNA-binding dimer and thus derepressed the operator-bound BirA (Beckett, 2007).

The present study found that the iTSS within the *nuoJ* gene in both *w*MelPop-CLA and *w*AlbB displayed regulation in response to stress induced conditions. For *w*MelPop-CLA the *nuoJ* iTSS was upregulated at 21°C whereas it was upregulated under rifampicin exposure for *w*AlbB. Cold stress and RNAP inhibition may lower metabolic capacities of the cell and thus require *Wolbachia* to conserve and focus energy on stress resistance mechanism. It is plausible both *w*MelPop-CLA and *w*AlbB would preferentially upregulate the ATU containing the biotin repressor in favour of oxidoreductase genes and ATP synthesis to minimise costly biotin synthesis in unfavourable conditions. If biotin synthesis is indeed repressed during conditions of low temperature, strains such a *w*MelPop-CLA may not be able to supplement their hosts

with the required vitamin. Inadequate supplementation of biotin may lower the fitness of the host and shift the dynamic of the symbiont-host relationship from mutualism to pathogenic, consequentially the success of *Wolbachia*-vector release strategies may be compromised as the fitness cost of a *Wolbachia* infection may impact host lifespan or result in an evolutionary pressure to vacate the residing symbiont as it will no longer be able to provide the necessary vitamins to justify its occupancy.

The T4SS is cotranscribed with *ribA* and contains an ATU exclusive to *wspB*

One of the key aspects of *Wolbachia* biology is the ability to secrete proteins and compounds outside of the cell to communicate and interact with the host. *Wolbachia* harbour several secretion systems but the dominant and most well understood machinery involved appears to be the T4SS. The T4SS is arranged into two main operons, one operon encodes the *virB8-11, virD4,* and *wspB* (T4SS-1) whilst the second encodes *virB3, virB4,* and *virB6* genes (T4SS-2). Previous studies have disputed the transcriptional architecture of the T4SS operon, evidence supporting a model in which the T4SS-1 operon is cotranscribed with the downstream *wspB* gene is in agreement however upstream cotranscription of the *ribA* gene involved in riboflavin biosynthesis is conflicting (Rancès *et al.,* 2008; Li and Carlow, 2012). The present study found evidence for a pTSS located upstream to the tRNA-Leu and *ribA* gene pair in both *w*MelPop-CLA and *w*AlbB, the tRNA-Leu gene is directly adjacent to the riboflavin *ribA* upstream of *virB8.* An additional two pTSS were located within the intergenic regions between *ribA*-*virB8* and *wspB*-*virD4* in *w*AlbB however its expression was significantly lower than that associated to the tRNA-Leu and *ribA* gene pair. The relatively high expression of the tRNA-Leu pTSS supports a model in which the T4SS-1 operon is transcribed starting from tRNA-Leu through to *wspB* for both *w*MelPop-CLA and *w*AlbB.

The study conducted by Li and Carlow (2012) also found evidence for two transcription factors (*w*BmxR1 and *w*BmxR2) capable of regulating the T4SS-1 operon. Li and Carlow (2012) were able to verify a region 109 nt upstream *ribA* capable of binding to both *w*BmxR1 and *w*BmxR2. Although the study was conducted on the filarial *Wolbachia w*Bm, orthologs of *w*BmxR1 and *w*BmxR2 were detected in *w*Mel (WD1304 and WD0931 respectively). Cappable-seq successfully detected a pTSS associated to the *w*BmxR1 ortholog WD1304 in *w*MelPop-CLA (WD_RS05910) and *w*AlbB (DEJ70_RS05740). Interestingly differential regulation was observed for the *w*AlbB ortholog (DEJ70_RS05740) where it was downregulated 2.7-fold

during rifampicin exposure, however given that rifampicin inhibits RNAP transcription this observation was not entirely surprising.

The *wspB* gene within the T4SS-1 operon is annotated to be pseudogenized in *w*AlbB whilst intact in *w*Mel and *w*MelPop-CLA. Interestingly a pTSS was associated to *wspB* for both strains suggesting an ATU within the T4SS-1 operon that transcribes *wspB* independently. The current study found that only *w*MelPop-CLA displayed regulation of the *wspB* associated pTSS by being downregulated 2.9-fold under 34°C thermal stress. The present data suggests that the pseudogenization of the *w*AlbB *wspB* gene is accompanied by loss of regulation whilst still retaining a remnant of its pTSS, in addition the *wspB* of *w*MelPop-CLA is still able to regulate the *wspB* as a separate ATU from the T4SS-1 operon.

TSS supports a single bicistronic operon model of CI genes with possible antisense regulation

Another cluster of genes whose architecture is of particular importance is the gene pair *cifA* and *cifB* responsible for cytoplasmic incompatibility. In the present study a pTSS was identified upstream *cifA* for both *w*MelPop-CLA and *w*AlbB with no pTSS identified within the intergenic region in *w*MelPop-CLA or upstream *cifB* in *w*AlbB. The presence of a single dominant pTSS upstream *cifA* supports a model in which *cifA* and *cifB* are cotranscribed as a bicistronic transcript. Previous research conducted by Lindsey, *et al.* (2017) used an RNA-seq approach encompassing 24 *D. melanogaster* life cycle stages to determine whether *w*Mel *cifA* and *cifB* are coregulated. The authors findings revealed that *cifA* is expressed 8-fold higher than its *cifB* counterpart, differential regulation between *cifA* and *cifB* does not support a cotranscribed model. The authors also identified a Rho-independent transcription terminator within the *w*Mel *cifA*-*cifB* intergenic region which may provide a mechanism to regulate the two genes. Another possible mechanism in which *cifB* can be cotranscribed with *cifA* but individually regulated could be the presence of a Puf family-like RNA-binding domain within *cifA* allowing it to theoretically regulate the cotranscribed *cifB* transcript. The lack of a *cifB* associated TSS in combination with previous evidence for transcriptional regulation between the two genes supports the model in which *cifA* and *cifB* are cotranscribed but regulated by an intergenic termination signal. The current study also found evidence of regulation of the *cifA* pTSS in *w*AlbB with downregulation at 34°C thermal stress and interestingly upregulation of

an antisense asTSS under rifampicin exposure. Although expression of the *w*AlbB *cifB* asTSS was relatively weak (8 CPM) and not conserved in *w*MelPop-CLA, antisense regulation could be another mechanism in which CI genes may be regulated.

A two-component system devoid of a histidine-kinase harbours a regulated alternative TSS with possible 5'UTR regulatory region

TCSs are stimulus-response systems which bacteria use to sense and respond to various environmental ques. TCSs generally involve a membrane-bound sensor histidine kinase (HK) which senses an environmental signal and a response regulator (RR) which when phosphorylated by the HK mediates the cellular response. The two TCSs which are present in nearly all *Wolbachia* are the PleC/PleD and CckA/CtrA systems. The PleC/PleD is known to regulate concentrations of c-di-GMP whereas the CckA/CtrA system is a master regulator of cell cycle progression. Interestingly CckA has been pseudogenized in both *w*Mel and *w*AlbB (Amelia R.I. Lindsey, 2020) which would suggest that either CckA can no longer phosphorylate its RR CtrA or that another HK, such as PleC, is able to replace the role of CckA. The RR CtrA remains intact in both *w*MelPop-CLA and *w*AlbB and so the present study further inspected TSS surrounding its local region. Both strains displayed a pTSS, iTSS, and gTSS associated to *ctrA*. Although the pTSS from both strains did not show any evidence of regulation, the gTSS directly upstream the pTSS was downregulated at 34°C thermal stress by 141-fold and 26-fold for *w*MelPop-CLA and *w*AlbB respectively. Evidence of regulation of an alternative TSS associated to a RR led to inspection of the downstream sequence which supported evidence for an RNA secondary structure for both strains. Although the MFE values for each of the secondary structures was not drastically different relative to random sequences, the fact that a seemingly conserved alternative TSS of both strains was capable of downregulation and a secondary structure formation deserves further study especially in the absence of its HK. Although the exact biological function of CtrA in *Wolbachia* is unknown, its role as a master regulator of cell cycle in the majority of alphaproteobacteria assumes a like-wise role for cell cycle progression within *Wolbachia* (Brilli *et al.*, 2010). Downregulation at 34°C could indicate an attempt for *w*MelPop-CLA and *w*AlbB to reduce replication during thermal stress and simultaneously reduce the burden upon its host. The predicted RNA secondary structure associated to the gTSS of both strains may be evidence for a mechanism

in which a 5'UTR element regulates intracellular replication in response to thermal fluctuations.

Chapter 5: General discussion and perspectives

Efficacy of cappable-seq and primary TSS as minority transcripts

The central objective of this study was to compare the transcriptional responses of two *Wolbachia* strains, *w*MelPop-CLA and *w*AlbB, to better guide the choice of strain used in current *Wolbachia*-vector release strategies. The application of cappable-seq was selected in the prospect of providing an effective enrichment of primary *Wolbachia* transcripts at single-nucleotide resolution to uncover the diversity of TSSs. In chapter 2 the effectiveness of cappable-seq for the enrichment of *Wolbachia* primary transcripts was successfully demonstrated, an enrichment of 7-fold was achieved without the need for a methodological control. The success of cappable-seq provided the deepest sequenced TSS study amongst UIPs to date, the first global TSS report in *Wolbachia*, and the first comparative transcriptional study between any two *Wolbachia* strains. This study reports that *w*MelPop-CLA and *w*AlbB harbour the full diversity of known TSS. A surprising outcome was for pTSS, assumed to be the primary transcript of gene expression, constituted a minor proportion of the total TSS. In terms of transcript abundance, internal iTSS and antisense asTSS appear to be major features in both *w*MelPop-CLA and *w*AlbB transcriptomes. This study succeeded in detecting promoter motifs for the σ^{70} housekeeping factor in both strains but with little indication of alternative promotor motifs, the major prevalence of intergenic and antisense transcripts appears to be a sign of alternative mechanism of gene regulation.

Heat shock response

Chapter 3 investigated the regulation of TSSs during temperature and antibiotic stress. Comparable to the composition of TSS discovered in chapter 2, this study reports that pTSS are not exclusively regulated units of transcription within the *w*MelPop-CLA and *w*AlbB transcriptomes. This finding alone suggests caution should be taken when interpretating RNA-seq data as the full capacity for an organism to regulate its genes, the resolution cappable-seq provides revealed non-canonical units of transcription which may be obscured by conventional RNA-seq analysis and provide key biological information. Based on the regulation of pTSS alone, this study was able to report that both *w*MelPop-CLA and *w*AlbB are capable of upregulating HSPs in response to 34°C thermal stress. However, *w*AlbB appears to induce a heat shock response with minimal involvement of sigma-32 relative to *w*MelPop-CLA.

This study reports that metabolic pathways of *w*MelPop-CLA, specifically relating to nucleotide metabolism and ATP synthesis, are upregulated when exposed to thermal heat stress. The lack of comparable metabolic genes regulated in *w*AlbB under thermal heat stress suggests that *w*AlbB, relative to *w*MelPop-CLA, does not require as much energy or that constitutive expression of metabolic genes sufficient to tolerate thermal stress. Since *Wolbachia* will acquire its resources from the host, the higher metabolic requirements of *w*MelPop-CLA, and likely *w*Mel, may place a significantly greater fitness cost to the host than *w*AlbB. Infection density will be a factor in determining which strain has a higher metabolic demand on its host, experiments where both strains can be maintained at equivalent densities will be valuable in determining the fitness cost to the host under equivalent thermal conditions.

Reproductive localization and transmission

One of the crucial features of the *Wolbachia* lifestyle is to locate to the germ-cell line of the host to successfully transmit to the next generation of offspring. This study reports that under thermal heat stress *w*MelPop-CLA regulates key genes related to transmission, the first of which is the predicted effector WD0811. The WD0811 gene encodes the eukaryotic-like WH2

domain commonly involved in pathogen host-cell entry. The *w*MelPop-CLA strain upregulates WD0811 under thermal stress, this could suggest that *w*MelPop-CLA may attempt to disseminate to less temperature sensitive areas within the host.

Response to antibiotics

Regarding the transcriptional response of *w*AlbB to antibiotics, this study reports expression of genes mainly involved in phospholipid metabolism, translation, efflux, and nucleotide metabolism in response to doxycycline and rifampicin exposure. Upregulation of genes related to phospholipid metabolism and fatty acid synthesis indicates that *w*AlbB prioritises genes which function in membrane remodelling. Remodelling the membrane in response to antibiotic exposure suggests that *w*AlbB decreases membrane permeability minimizing the diffusion of doxycycline and rifampicin. Membrane remodelling during exposure to antibiotics may also relate to the introduction of efflux pumps, within the IM to reduce intracellular antibiotic concentrations.

Studies have compared the impact of environmental concentrations of tetracyclines on *w*Mel, *w*MelPop, and *w*AlbB on infection density and CI (Endersby-Harshman, Axford and Hoffmann, 2019). The authors reported that at comparable concentrations of chlortetracycline (50ug\l) *w*Mel and *w*MelPop strains exhibited significantly reduced *Wolbachia* density and loss of CI compared to *w*AlbB, *w*AlbB was still able to induce CI. The study conducted by Darby, *et al.* (2014) reported that *w*MelPop-CLA was also capable of upregulating phospholipid, translation, and nucleotide synthesis genes but unlike the current study did not report upregulation of any efflux pumps. It is possible that the upregulation of efflux system reported in this study is a key factor in the antibiotic susceptibility between *w*MelPop-CLA and *w*AlbB. Environmental assessment studies report that maximum field concentrations of tetracyclines do not exceed 0.10–0.97 µg/litre which is inadequate to cause lethal effects of genetically modified mosquitos (Oxitec, 2016). However bioaccumulation of tetracycline, especially at larval stages, may incrementally disrupt *w*Mel variants to successfully induce CI and could hamper vector-release strategies (Curtis *et al.*, 2015). Given the evidence collected to data, *Wolbachia*-vector release strategies may benefit from selecting *w*AlbB over *w*Mel variants if environmental antibiotic concentrations are significant enough to detrimentally impact field releases. Relative to temperature wAlbB prioritises the extrusion of antibiotics via secretion

and efflux systems. In regard to temperature, *w*AlbB prioritises protein folding machinery I order to maintain protein homeostasis.

Induction of MGEs in response to stress

The present study reports that MGEs in the form of transposases and phage components are upregulated in response to stressors in *w*MelPop-CLA and *w*AlbB. The *w*MelPop-CLA strain exhibited upregulation of transposases within the WO-A region under 34°C heat stress. Upregulation of transposases under stress is a common behaviour and theoretically could allow *Wolbachia* to acquire novel adaptive mutations over time. Upregulation of WO-phage under low and high temperature stress has been previously reported (Bordenstein and Bordenstein, 2011) although the current study only found expression of phage tail pTSS which was not itself regulated under temperature stress within *w*MelPop-CLA. Interestingly *w*AlbB displayed upregulated of a pseudogenized phage portal protein which may be a remnant of a once activate phage, the function of a such a transcript is inconclusive if it is functional at all.

An overlooked perspective on WO phage, which this study highlights, is the possibility of an alternate function of phage tail components as PTLBs. PTLBs may be acting on behave of the symbiont for communicating with the host as seen in other previously mentioned bacteria residing in invertebrates. Although speculative if WO particles have been adapted in *Wolbachia*-host communication, host specificity would need to be considered when using artificial *Wolbachia*-host infection such as *w*Mel in mosquito hosts. Further research into alternative functions of WO phage components would of interest to *Wolbachia* biology and the greater field of symbiosis.

Role of ncRNAs in *Wolbachia* stress response

Another objective of the current study was to detect the presence of ncRNA and their regulatory capacity between *w*MelPop-CLA and *w*AlbB under varying stressors. This study reports that a large proportion of both *w*MelPop-CLA and *w*AlbB transcriptomes express antisense and intergenic transcripts that could encode ncRNA. Cappable-seq successfully detected previously reported intergenic ncRNA and identifies new intergenic and antisense candidate ncRNA capable of regulation. Of the known detected intergenic ncRNA, this study reports the first

regulatory observation of the ncRNA *ncrwMel02* under temperature extremes of 21°C and 34°C. As previously reported by Woolfit, *et al.* (2015), *ncrwMel02* was overexpressed in testes compared to ovaries which may indicate its involvement in reproductive manipulation. Although the function of *ncrwMel02* has yet to be understood, downregulation of a possible reproductive factor in the *Wolbachia*-host association may impact vector release strategies if found to be significant in CI. Understanding the functional role of *ncrwMel02* should be a point of focus in future research of *Wolbachia* ncRNA. Interestingly, the highly expressed intergenic ncRNA detected in *w*AlbB did not exhibit regulation under temperature stress which may add to the growing evidence of its thermal resilience in comparison to *w*MelPop-CLA and *w*AlbB.

Antisense transcripts appear to be a major component in *w*MelPop-CLA and *w*AlbB genomes and may have important roles in *Wolbachia* biology. This study confirmed the presence of the *w*MelPop-CLA antisense transcript *WsnRNA-59* and found it to be upregulated under 21°C cold stress. Since *WsnRNA-59* overlaps with a phage major capsid protein it could be involved in antisense repression of temperature sensitive phage. Interestingly this study reports a regulated antisense transcript of a phage portal protein for *w*MelPop-CLA and *w*AlbB. Since the portal protein in question is pseudogenized in *w*MelPop-CLA (WD_RS01205) and phage is deemed inactive in *w*AlbB, antisense regulation of these transcripts is puzzling. It is possible that during genome reduction symbionts may first utilize antisense repression before pseudogenization and degradation.

Using the 100 nt sequences of the oTSS and asTSS transcript, the present study attempted to construct predicted RNA secondary structures. Whilst RNA secondary structure prediction is fundamentally based on thermodynamic probability, this study was able to demonstrate likely secondary structures for TSS based ncRNA. Using randomised sequences to compare against average MFE value to each TSS type, this study supported to likelihood that oTSS and iTSS transcripts on average can form stable secondary structures. Predicted secondary structures of interest were those associated to the *mutL* antisense transcript, *fsa* iTSS and the regulated RR gTSS. The *mutL* asTSS was the most highly expressed antisense transcript in both *w*MelPop-CLA and *w*AlbB transcriptomes. The *mutL* is of particular interest due to its duplication has only been found in *Wolbachia*, the associated asTSS is conserved in both strains which suggests functional importance. It is possible that repression *mutL* mismatch repair gene may serve to increase mutational innovation over successive generations. Furthur research into why *mutL*

might have been duplicated and whether it indeed encodes a functional secondary structure will be of interest, the possibility of the secondary structure acting as an RNA thermomsensor could be proposed. The *fsa* internal iTSS transcript was the most highly expressed TSS in both transcriptomes which suggest major importance to *Wolbachia* biology. Since the *fsa* iTSS is outside of the *fsa* encodes active site, the function of this transcript does not seem to be related to its annotated gene. A similar and low MFE secondary structure was predicted for both *w*MelPop-CLA and *w*AlbB which may give clues to its true role. The regulated gTSS associated to the RR of both strains suggests a possible regulatory mechanism in which TCSs can utilize 5'UTR secondary structures in response to stress, further examination into how such a structure is utilized in regulating the RR will also be on interest.

Implications of sub-operon architecture

With the resolution provided by cappable-seq, the transcriptional architecture of operons has now been greatly improved in *Wolbachia.* This study not only supports a model in which the T4SS is cotranscribed with riboflavin *ribA* gene but also supports a model in which the *wspB* gene can form an ATU capable of independent regulation under thermal stress. The *wspB* has recently become a key candidate in the thermal susceptibility between *w*AlbB and *w*Mel variants. A recent study conducted by Gu, *et al.* (2022) found a *w*Mel variant (*w*MelM) which was capable of higher thermal tolerance compared to *w*Mel, the only genomic difference in which they were able to discern was the pseudogenization of the *wspB* gene. The *wspB* gene is intact in *w*Mel and *w*MelPop-CLA whilst being pseudogenized in the thermostable *w*AlbB strain. Another recent study by Hague, *et al.* (2021) reported a strong correlation with pseudogenization of the *wspB* gene and tropical *Wolbachia* variants. This study reports that under heat stress, the *wspB* ATU is downregulated by *w*MelPop-CLA. The pseudogenization of the WspB protein appears to impact the extracellular loops which is suggested to be involved in host cell recognition (Hague *et al.*, 2021). Another recent publication has implicated *wspB* and thermal tolerance in *w*AlbB variants, the variant *w*AlbB-II was shown to be more susceptible to heat stress and also encoded a pseudogenized *wspB* (Martinez, 2022). Given the evidence of a downregulated *wspB* ATU in *w*MelPop-CLA and its previously recognised role in thermal tolerance, this study supports a model in which *wspB* is a crucial determinant in thermal resilience. Therefore, this study suggests the preference for *w*AlbB over *w*Mel when selecting thermally stable *Wolbachia* variants for *Wolbachia*-release strategies to prevent loss of infection. Continuing research into the function of wspB and its possibly involvement in

host communication will help better improve the understanding of why *wspB* appears to be a major determining factor in thermal reliance.

The current perspective is that intracellular symbionts undergo genome reduction via ongoing mutations leading to pseudogenization and eventual loss of redundant genes, those genes whose function can be compensated by the host are no longer required and therefore lost (McCutcheon and Moran, 2012). As the intracellular environment of the host is considered relatively stable, regulatory mechanism become excessive and are also lost. The new operon model in which genes within a single operon can be part of multiple ATUs may contradict the conventional genome reduction perspective. The current study reports that internal and antisense transcripts are major features in the *Wolbachia* transcriptome. Operons which seem to contain genes of unrelated function, such as biotin ligase found in the ATU within the oxidoreductase operon, may contain ATUs that can be independently regulated. It may be possible that as genomes undergo reduction and their proteins become unstable, transcriptional start sites may become progressively more common. Protein instability may also affect essential proteins such as RNAP, it could be imagined that additional TSS may arise via transfer of promoters by transposases, weak RNAP binding, and accumulation of mutations leading to weak promotor-like regions. Genome reduction is usually summed as an ongoing evolutionary trajectory toward "simplicity", however, internal, and antisense TSS may be features which allow increasing complexity by enabling multiple transcriptional units to form. This view essentially implies that as genomes become progressively smaller, the transcriptional capability become increasingly denser. Comparing the TSS density and the total number of ATUs per operon within obligate symbionts relative to their free-living counterparts would be an informative assessment of this speculative hypothesis.

Implications for *Wolbachia*-vector release strategies

The *w*Mel strain has now been released into mosquito populations in several countries including Australia, Brazil, Indonesia, and Vietnam (Ross, Turelli and Hoffmann, 2019). The *w*AlbB strain is less prevalent in vector release strategies but has been successfully released in Malaysia (Nazni *et al.*, 2019). Maintaining high infection density of *Wolbachia* within released mosquitos is essential for effective CI, too high of an infection becomes a substantial fitness cost to the host and may prevent vector-release strategies as seen in *w*MelPop (Nguyen *et al.*,

2015). Maintaining a balance of environmental resilience, high infection, host fitness, and complete CI is key for continuing success of vector-release strategies.

Previous reports of heatwaves in Australia have resulted in temporary reductions in *w*Mel density especially within early larval stages (Ross *et al.*, 2020). Season loss of *w*Mel was also reported in Vietnam which strongly correlated with host dry seasons (Hien *et al.*, 2021). The *w*Mel strain has remained in the city of Cairns Australia for over a decade and has undergone little genomic changes (Ross *et al.*, 2022). The *w*AlbB, since its deployment in Malaysia, has also maintained high density and dengue inhibition for 2 years (Ahmad *et al.*, 2021). Future releases and current infected mosquito populations may be at risk as global temperature extremes are predicted to increase (Donat *et al.*, 2013), more severe temperatures can expect to further decrease *Wolbachia* infection density resulting in loss of CI and viral protection. Interestingly, the choice between strains *w*Mel and *w*AlbB may not have to be made. A superinfection of both strains termed *w*Mel*w*AlbB appears to show high infection, complete CI, and DENV inhibition (Joubert *et al.*, 2016). Concerns regarding non-native communication between *w*Mel and the *Aedes* host, possibly via the ncRNA that have been regulated in the current study, may be complemented by coinfection with *w*AlbB especially under thermal stress.

Successful establishment of *Wolbachia* within mosquito population is dependent of maternal transmission of the symbiont to successive generations. The current study reports that genes predicted to be involved in transmission and motility are of importance when *Wolbachia*, particularly *w*MelPop-CLA and by implication *w*Mel, is confronted to thermal extremes. The study suggests that genes such as the WH2 domain contain gene should be monitored in mosquito populations for changes which may affect their function. The *wspB* gene should also be monitored in wAlbB populations as reversion back to a functional gene may result in thermal susceptibility.

Limitations of the study and future considerations

The current study admittedly is not without its limitation which could be improved upon to provide a more informative analysis. The *w*MelPop-CLA was used in the RML12 cell line as a model for exploring the transcriptome of *w*Mel, all analyses involved mapping reads from *w*MelPop-CLA on the *w*Mel genome. Using *w*MelPop-CLA as a model for *w*Mel enabled a

higher quantity of RNA to be obtained due to *w*MelPop-CLA ability to grow at higher densities than *w*Mel. Genomic difference between *w*MelPop-CLA relative to *w*Mel are minimal but do include loss of an IS5 insertion sequence, duplication of an ankyrin gene, a 143 kb inversion, and 156 SNPs/indels (Woolfit *et al.*, 2013). Besides the differences stated, all *w*Mel genes have orthologs in *w*MelPop-CLA which was suitable for using *w*MelPop-CLA as a model for *w*Mel transcriptomics. It is possible that within the known small genomic difference in *w*MelPop-CLA may harbour non-canonical TSS that encode ncRNA or ATUs that could functionally distinguish between the two closely related strains. Use of a *w*Mel infected cell line might not have given adequate RNA the *w*MelPop-CLA which would have results in a less informative analysis.

Another limitation was for origin of the host cell lines used. The RML12 and Aa23 cell lines are derived from sex unspecific larval and embryonic stages respectively. Although both are derived from early developmental stages of the mosquito host, transcriptional differences may occur between larval and embryonic stages in both the host and the symbiont. Lack of sex specificity may also mask any transcriptional differences that are sex specific. Ideally the use of whole host *Aedes aegypti* mosquitos would better simulate real-world conditions, separating the abdomen and head would also allow for transcriptional differences that may be tissue specific. Although ideal, using whole mosquitos requires a heavily maintained insectary which would have exponentially complicated the entire project. Using cell lines are more convenient that whole mosquitos and in turn give higher RNA yields. For future studies, it would be of interest to compare *w*Mel and *w*AlbB under thermal stress in whole mosquitos. Methods such as RT-qPCR or dual-RNA-seq would allow comparative expression of both host and symbionts gene as has been done for *Wigglesworthis* in its Tsetse fly host (Bing *et al.*, 2017).

Although cappable-seq enabled single nucleotide resolution of TSS, the 3’ TTS was unavailable to be identified. Being able to identify the TSS would have enabled the identification of the full-length transcript. Knowing the full length of each transcript would have identified to boundaries for each ATU and the entire length of candidate ncRNA. RNA secondary structure analysis will also benefit from full length identification, using the exact full-length sequence to construct predicted secondary structures instead of using a 100-nucleotide estimate providing a more accurate model.

References

‘6S RNA, a Global Regulator of Transcription’ (2018) in *Regulating with RNA in Bacteria and Archaea*. American Society of Microbiology, pp. 355–367. doi: 10.1128/microbiolspec.rwr-0019-2018.

Ahmad, N. A. *et al.* (2021) ‘Wolbachia strain wAlbB maintains high density and dengue inhibition following introduction into a field population of Aedes aegypti’, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 376(1818), p. 20190809. doi: 10.1098/rstb.2019.0809.

Akman, L. *et al.* (2002) ‘Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia’, *Nature Genetics*, 32(3), pp. 402–407. doi: 10.1038/ng986.

Albrecht, M. *et al.* (2009) ‘Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome’, *Nucleic Acids Research*. Oxford University Press, 38(3), pp. 868–877. doi: 10.1093/nar/gkp1032.

Albrecht, M. *et al.* (2011) *The transcriptional landscape of Chlamydia pneumoniae*. doi: 10.1186/gb-2011-12-10-r98.

Ant, T. H. *et al.* (2018) ‘The Wolbachia strain wAu provides highly efficient virus transmission blocking in Aedes aegypti’, *PLOS Pathogens*. Edited by D. S. Schneider. Public Library of Science, 14(1), p. e1006815. doi: 10.1371/journal.ppat.1006815.

Atwal, S. *et al.* (2021) ‘Discovery of a diverse set of bacteria that build their cell walls without the canonical peptidoglycan polymerase apbp’, *mBio*, 12(4), pp. 1–12. doi: 10.1128/mBio.01342-21.

Backert, S., Fronzes, R. and Waksman, G. (2008) ‘VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems?’, *Trends in Microbiology*, 16(9), pp. 409–413. doi: 10.1016/j.tim.2008.07.001.

Balvín, O. *et al.* (2018) ‘Co-speciation in bedbug Wolbachia parallel the pattern in nematode hosts’, *Scientific Reports*. Nature Publishing Group, 8(1), p. 8797. doi: 10.1038/s41598-018-25545-y.

Bandi, C. *et al.* (1998) ‘Phylogeny of Wolbachia in filarial nematodes’, *Proceedings of the Royal Society B: Biological Sciences*. The Royal Society, 265(1413), pp. 2407–2413. doi: 10.1098/rspb.1998.0591.

Baumann, P., Baumann, L. and Clark, M. A. (1996) ‘Levels of Buchnera aphidicola chaperonin GroEL during growth of the aphid Schizaphis graminum’, *Current Microbiology*, 32(5), pp. 279–285. doi: 10.1007/s002849900050.

Beckett, D. (2007) ‘Biotin sensing: Universal influence of biotin status on transcription’, *Annual Review of Genetics*, 41, pp. 443–464. doi: 10.1146/annurev.genet.41.042007.170450.

Beckmann, J. F. and Fallon, A. M. (2014) ‘Detection of the Wolbachia Protein WPIP0282 in Mosquito Spermathecae : Implications for Cytoplasmic Incompatibility’, 43(9), pp. 1–22. doi: 10.1016/j.ibmb.2013.07.002.Detection.

Beckmann, J. F., Ronau, J. A. and Hochstrasser, M. (2017) ‘A Wolbachia deubiquitylating enzyme induces cytoplasmic incompatibility’, *Nature Microbiology*, 2(March), p. 17007. doi:

10.1038/nmicrobiol.2017.7.

Bengoechea, J. A. and Skurnik, M. (2000) ‘Temperature-regulated efflux pump/potassium antiporter system mediates resistance to cationic antimicrobial peptides in Yersinia’, *Molecular Microbiology*, 37(1), pp. 67–80. doi: 10.1046/j.1365-2958.2000.01956.x.

Bhatt, S. *et al.* (2013) ‘The global distribution and burden of dengue’, *Nature*, 496(7446), pp. 504–507. doi: 10.1038/nature12060.

Bian, G. *et al.* (2010) ‘The endosymbiotic bacterium Wolbachia induces resistance to dengue virus in Aedes aegypti’, *PLoS Pathogens*, 6(4), pp. 1–10. doi: 10.1371/journal.ppat.1000833.

Bing, X. *et al.* (2017) ‘Unravelling the relationship between the tsetse fly and its obligate symbiont Wigglesworthia: transcriptomic and metabolomic landscapes reveal highly integrated physiological networks.’, *Proceedings. Biological sciences*. The Royal Society, 284(1857), p. 20170360. doi: 10.1098/rspb.2017.0360.

Bleidorn, C. and Gerth, M. (2017) ‘A critical re-evaluation of multilocus sequence typing (MLST) efforts in Wolbachia’, *doi.org*, p. 133710. doi: 10.1101/133710.

Bordenstein, S. R. and Werren, J. H. (1998) ‘Effects of A and B Wolbachia and host genotype on interspecies cytoplasmic incompatibility in Nasonia’, *Genetics*, 148(4), pp. 1833–1844. doi: 10.1093/genetics/148.4.1833.

Bordenstein, Sarah R. and Bordenstein, Seth R. (2016) ‘Eukaryotic association module in phage WO genomes from Wolbachia’, *Nature Communications*. Nature Publishing Group, 7. doi: 10.1038/ncomms13155.

Bordenstein, Sarah R and Bordenstein, Seth R (2011) ‘Temperature affects the tripartite interactions between bacteriophage WO, Wolbachia, and cytoplasmic incompatibility’, *PLoS ONE*, 6(12). doi: 10.1371/journal.pone.0029106.

Brehmer, D. *et al.* (2001) ‘Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange’, *Nature Structural Biology*, 8(5), pp. 427–432. doi: 10.1038/87588.

Brilli, M. *et al.* (2010) ‘The diversity and evolution of cell cycle regulation in alpha-proteobacteria: A comparative genomic analysis’, *BMC Systems Biology*. BioMed Central, 4, p. 52. doi: 10.1186/1752-0509-4-52.

Callaini, G., Dallai, R. and Riparbelli, M. G. (1997) ‘Wolbachia-induced delay of paternal chromatin condensation does not prevent maternal chromosomes from entering anaphase in incompatible crosses of Drosophila simulans’, *Journal of Cell Science*, 110(2), pp. 271–280. doi: 10.1242/jcs.110.2.271.

Campbell, E. A. *et al.* (2001) ‘Structural mechanism for rifampicin inhibition of bacterial RNA polymerase’, *Cell*, 104(6), pp. 901–912. doi: 10.1016/S0092-8674(01)00286-0.

Caturegli, P. *et al.* (2000) ‘ankA:an Ehrlichia phagocytophila Group Gene Encoding a Cytoplasmic Protein Antigen with Ankyrin Repeats’, 68(9), pp. 1–7. Available at: papers2://publication/uuid/A0E6F0EE-F592-498D-BA77-6A48D2415967.

Cavalazzi, B. *et al.* (2021) ‘Cellular remains in a ~3.42-billion-year-old subseafloor hydrothermal environment’, *Science Advances*, 7(29). doi: 10.1126/sciadv.abf3963.

Cerveau, N. *et al.* (2011) 'Short- and long-term evolutionary dynamics of bacterial insertion sequences: Insights from Wolbachia endosymbionts', *Genome Biology and Evolution*, 3(1), pp. 1175–1186. doi: 10.1093/gbe/evr096.

Chan, Y. G. Y. *et al.* (2009) 'Rickettsial outer-membrane protein B (rOmpB) mediates bacterial invasion through Ku70 in an actin, c-Cbl, clathrin and caveolin 2-dependent manner', *Cellular Microbiology*, 11(4), pp. 629–644. doi: 10.1111/j.1462-5822.2008.01279.x.

Chaudhry, C. *et al.* (2003) 'Role of the γ-phosphate of ATP in triggering protein folding by GroEL-GroES: Function, structure and energetics', *EMBO Journal*, 22(19), pp. 4877–4887. doi: 10.1093/emboj/cdg477.

Choi, C. H. *et al.* (2005) 'Outer membrane protein 38 of Acinetobacter baumannii localizes to the mitochondria and induces apoptosis of epithelial cells', *Cellular Microbiology*, 7(8), pp. 1127–1138. doi: 10.1111/j.1462-5822.2005.00538.x.

Christensen, S. and Serbus, L. R. (2010) 'Comparative Analysis of Wolbachia Genomes Reveals Streamlining and Divergence of Minimalist Two-Component Systems'. doi: 10.1534/g3.115.017137doi:10.1534/g3.115.017137/-/DC1.

Chrostek, E. *et al.* (2013) 'Wolbachia Variants Induce Differential Protection to Viruses in Drosophila melanogaster: A Phenotypic and Phylogenomic Analysis', *PLoS Genetics*, 9(12). doi: 10.1371/journal.pgen.1003896.

Chung, M. *et al.* (2018) 'Targeted enrichment outperforms other enrichment techniques and enables more multi-species RNA-Seq analyses', *Scientific Reports*. Springer US, 8(1), pp. 1–12. doi: 10.1038/s41598-018-31420-7.

Chung, M. *et al.* (2019) 'Drug Repurposing of Bromodomain Inhibitors as Potential Novel Therapeutic Leads for Lymphatic Filariasis Guided by Multispecies Transcriptomics', *mSystems*. Edited by A. Fodor, 4(6). doi: 10.1128/mSystems.00596-19.

Clark, M. A. *et al.* (2000) 'Cospeciation between bacterial endosymbionts (Buchnera) and a recent radiation of aphids (Uroleucon) and pitfalls of testing for phylogenetic congruence', *Evolution*, 54(2), pp. 517–525. doi: 10.1111/j.0014-3820.2000.tb00054.x.

Coleman, J. (1990) 'Characterization of Escherichia coli cells deficient in 1-acyl-sn-glycerol-3-phosphate acyltransferase activity', *Journal of Biological Chemistry*. Â© 1990 ASBMB. Currently published by Elsevier Inc; originally published by American Society for Biochemistry and Molecular Biology., 265(28), pp. 17215–17221. doi: 10.1016/s0021-9258(17)44891-5.

Cordaux, R. *et al.* (2008) 'Intense transpositional activity of insertion sequences in an ancient obligate endosymbiont', *Molecular Biology and Evolution*, 25(9), pp. 1889–1896. doi: 10.1093/molbev/msn134.

Courcelle, J. *et al.* (2001) 'Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient Escherichia coli', *Genetics*, 158(1), pp. 41–64. doi: 10.1093/genetics/158.1.41.

Cuervo, A. and Carrascosa, J. L. (2012) 'Viral connectors for DNA encapsulation', *Current Opinion in Biotechnology*, 23(4), pp. 529–536. doi: 10.1016/j.copbio.2011.11.029.

Curtis, Z. *et al.* (2015) 'Assessment of the impact of potential tetracycline exposure on the phenotype of Aedes aegypti OX513A: Implications for field use', *PLoS Neglected Tropical*

Diseases, 9(8), pp. 1–15. doi: 10.1371/journal.pntd.0003999.

Dale, C. and Moran, N. A. (2006) 'Molecular Interactions between Bacterial Symbionts and Their Hosts', *Cell*, 126(3), pp. 453–465. doi: 10.1016/j.cell.2006.07.014.

Darby, A. C. *et al.* (2012) 'Analysis of gene expression from the Wolbachia genome of a filarial nematode supports both metabolic and defensive roles within the symbiosis', *Genome Research*, 22(12), pp. 2467–2477. doi: 10.1101/gr.138420.112.

Darby, A. C. *et al.* (2014) 'Integrated transcriptomic and proteomic analysis of the global response of Wolbachia to doxycycline-induced stress', *The ISME Journal*, 8(10), pp. 925–937. doi: 10.1038/ismej.2013.192.

Dobson, S. L. and Rattanadechakul, W. (2001) 'A novel technique for removing Wolbachia infections from Aedes albopictus (Diptera: Culicidae)', *Journal of Medical Entomology*, 38(6), pp. 844–849. doi: 10.1603/0022-2585-38.6.844.

Donat, M. G. *et al.* (2013) 'Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset', *Journal of Geophysical Research Atmospheres*, 118(5), pp. 2098–2118. doi: 10.1002/jgrd.50150.

Dougan, D. A., Mogk, A. and Bukau, B. (2002) 'Protein folding and degradation in bacteria: To degrade or not to degrade? That is the question', *Cellular and Molecular Life Sciences*, 59(10), pp. 1607–1616. doi: 10.1007/PL00012487.

Dramsi, S. and Cossart, P. (1998) 'Intracellular pathogens and the actin cytoskeleton', *Annual Review of Cell and Developmental Biology*, 14, pp. 137–166. doi: 10.1146/annurev.cellbio.14.1.137.

Dunbar, H. E. *et al.* (2007) 'Aphid thermal tolerance is governed by a point mutation in bacterial symbionts', *PLoS Biology*. Edited by C. Godfray. Public Library of Science, 5(5), pp. 1006–1015. doi: 10.1371/journal.pbio.0050096.

Dupont, M. *et al.* (2004) 'Enterobacter aerogenes OmpX, a cation-selective channel mar- and osmo-regulated', *FEBS Letters*, 569(1–3), pp. 27–30. doi: 10.1016/j.febslet.2004.05.047.

Durvasula, R. V *et al.* (1997) *Prevention of insect-borne disease: An approach using transgenic symbiotic bacteria*, *Medical Sciences*. Available at: www.pnas.org. (Accessed: 1 January 2019).

Egilmez, H. I. *et al.* (2018) 'Temperature-dependent virus lifecycle choices may reveal and predict facets of the biology of opportunistic pathogenic bacteria', *Scientific Reports*. Springer US, 8(1), pp. 1–13. doi: 10.1038/s41598-018-27716-3.

Endersby-Harshman, N. M., Axford, J. K. and Hoffmann, A. A. (2019) 'Environmental Concentrations of Antibiotics May Diminish Wolbachia infections in Aedes aegypti (Diptera: Culicidae)', *Journal of Medical Entomology*, 56(4), pp. 1078–1086. doi: 10.1093/jme/tjz023.

Ettwiller, L. *et al.* (2016) 'A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome', *BMC Genomics*. BMC Genomics, 17(1), p. 199. doi: 10.1186/s12864-016-2539-z.

Fallon, A. M. (2021) 'DNA recombination and repair in Wolbachia: RecA and related proteins', *Molecular Genetics and Genomics*. Springer Science and Business Media Deutschland GmbH, 296(2), pp. 437–456. doi: 10.1007/s00438-020-01760-z.

Fan, Y. and Wernegreen, J. J. (2013) ‘Can’t Take the Heat: High Temperature Depletes Bacterial Endosymbionts of Ants’, *Microbial Ecology*. Springer US, 66(3), pp. 727–733. doi: 10.1007/s00248-013-0264-6.

Fattouh, N., Cazevieille, C. and Landmann, F. (2018) ‘Wolbachia endosymbionts subvert the endoplasmic reticulum to acquire host membranes without triggering ER stress’, *PLoS Neglected Tropical Diseases*. Public Library of Science, 13(3), p. e0007218. doi: 10.1371/journal.pntd.0007218.

Faucher, S. P. *et al.* (2010) ‘Legionella pneumophila 6S RNA optimizes intracellular multiplication’, *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), pp. 7533–7538. doi: 10.1073/pnas.0911764107.

Feder, M. E. *et al.* (1999) ‘Interaction of Drosophila and its endosymbiont Wolbachia: Natural heat shock and the overcoming of sexual incompatibility’, *American Zoologist*. Oxford University Press, 39(2), pp. 363–373. doi: 10.1093/icb/39.2.363.

Feng, Y., Zhang, H. and Cronan, J. E. (2013) ‘Profligate biotin synthesis in α-proteobacteria - a developing or degenerating regulatory system?’, *Molecular Microbiology*, 88(1), pp. 77–92. doi: 10.1111/mmi.12170.

Fenollar, F., Maurin, M. and Raoult, D. (2003) ‘Wolbachia pipientis growth kinetics and susceptibilities to 13 antibiotics determined by immunofluorescence staining and real-time PCR’, *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology (ASM), 47(5), pp. 1665–1671. doi: 10.1128/AAC.47.5.1665-1671.2003.

Ferree, P. M. *et al.* (2005) ‘Wolbachia Utilizes Host Microtubules and Dynein for Anterior Localization in the Drosophila Oocyte’, *PLoS Pathogens*. Public Library of Science, 1(2), p. e14. doi: 10.1371/journal.ppat.0010014.

Flores, H. A. *et al.* (2020) ‘Multiple Wolbachia strains provide comparative levels of protection against dengue virus infection in Aedes aegypti’, *PLOS Pathogens*. Edited by F. M. Jiggins. Public Library of Science, 16(4), p. e1008433. doi: 10.1371/journal.ppat.1008433.

Foster, J. *et al.* (2005) ‘The Wolbachia genome of Brugia malayi: Endosymbiont evolution within a human pathogenic nematode’, *PLoS Biology*, 3(4), pp. 0599–0614. doi: 10.1371/journal.pbio.0030121.

Francez-Charlot, A. *et al.* (2009) ‘Sigma factor mimicry involved in regulation of general stress response’, *Proceedings of the National Academy of Sciences of the United States of America*, 106(9), pp. 3467–3472. doi: 10.1073/pnas.0810291106.

Francez-Charlot, A. *et al.* (2015) ‘The general stress response in Alphaproteobacteria’, *Trends in Microbiology*. Elsevier Current Trends, pp. 164–171. doi: 10.1016/j.tim.2014.12.006.

Fujii, Y. *et al.* (2004) ‘Isolation and characterization of the bacteriophage WO from Wolbachia, an arthropod endosymbiont’, *Biochemical and Biophysical Research Communications*, 317(4), pp. 1183–1188. doi: 10.1016/j.bbrc.2004.03.164.

Gabriela Flores-Ramirez, Barbora Jankovicova, Zuzana Bilkova, Jan A. Miernyk, L. S. (2014) ‘Identification of Coxiella burnetii surface-exposed and cell envelope associated proteins using a combined bioinformatics plus proteomics strategy’, *Proteomics*, pp. 1–25. doi: 10.1002/jssc.201200569.

Gamer, J. *et al.* (1996) 'A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the Escherichia coli heat shock transcription factor σ32', *EMBO Journal*, 15(3), pp. 607–617. doi: 10.1002/j.1460-2075.1996.tb00393.x.

Gao, R., Mack, T. R. and Stock, A. M. (2007) 'Bacterial response regulators: versatile regulatory strategies from common domains', *Trends in Biochemical Sciences*, 32(5), pp. 225–234. doi: 10.1016/j.tibs.2007.03.002.

Garwood, R. J. and Edgecombe, G. D. (2011) 'Early Terrestrial Animals, Evolution, and Uncertainty', *Evolution: Education and Outreach*, 4(3), pp. 489–501. doi: 10.1007/s12052-011-0357-y.

Geoghegan, V. *et al.* (2017) 'Perturbed cholesterol and vesicular trafficking associated with dengue blocking in Wolbachia-infected Aedes aegypti cells', *Nature Communications*, 8(1). doi: 10.1038/s41467-017-00610-8.

Ghosh, Suparna *et al.* (2017) 'Identification of novel small RNAs in burkholderia cenocepacia KC-01 expressed under iron limitation and oxidative stress conditions', *Microbiology (United Kingdom)*, 163(12), pp. 1924–1936. doi: 10.1099/mic.0.000566.

Giannoukos, G. *et al.* (2012) *Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes*, *Genome Biology*. Available at: http://genomebiology.com/2012/13/3/r23.

Griffin, M. O. *et al.* (2010) 'Tetracyclines: A pleitropic family of compounds with promising therapeutic properties. Review of the literature', *American Journal of Physiology - Cell Physiology*, 299(3). doi: 10.1152/ajpcell.00047.2010.

Grote, A. *et al.* (2017) 'Defining Brugia malayi and Wolbachia symbiosis by stage-specific dual RNA-seq', *PLoS Neglected Tropical Diseases*, 11(3). doi: 10.1371/journal.pntd.0005357.

Gu, X. *et al.* (2022) 'A wMel Wolbachia variant in Aedes aegypti from field-collected Drosophila melanogaster with increased phenotypic stability under heat stress', *Environmental Microbiology*, 24, pp. 2119–2135. doi: 10.1111/1462-2920.15966.

Gutzwiller, F. *et al.* (2015) 'Dynamics of Wolbachia pipientis Gene Expression Across the Drosophila melanogaster Life Cycle'. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4683655/pdf/2843.pdf (Accessed: 14 June 2018).

Haas, B. J. *et al.* (2012) *How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?* doi: 10.1186/1471-2164-13-734.

Hagman, K. E. *et al.* (1995) 'Resistance of Neisseria gonorrhoeae to antimicrobial hydrophobic agents is modulated by the mtrRCDE efflux system', *Microbiology*, 141(3), pp. 611–622. doi: 10.1099/13500872-141-3-611.

Hague, M. T. J. *et al.* (2021) 'Temperature effects on cellular host-microbe interactions explain continent-wide endosymbiont prevalence', *Current Biology*. Elsevier Inc., pp. 1–11. doi: 10.1016/j.cub.2021.11.065.

Hansen, A. K. and Moran, N. A. (2011) 'Aphid genome expression reveals host-symbiont cooperation in the production of amino acids', *Proceedings of the National Academy of*

Sciences of the United States of America, 108(7), pp. 2849–2854. doi: 10.1073/pnas.1013465108.

Hayward, R. D. and Koronakis, V. (1999) 'Direct nucleation and bundling of actin by the SipC protein of invasive Salmonella', *EMBO Journal*, 18(18), pp. 4926–4934. doi: 10.1093/emboj/18.18.4926.

Hegde, S. *et al.* (2019) 'CRISPR/Cas9-mediated gene deletion of the ompA gene in symbiotic Cedecea neteri impairs biofilm formation and reduces gut colonization of Aedes aegypti mosquitoes', *PLoS neglected tropical diseases*, 13(12), p. e0007883. doi: 10.1371/journal.pntd.0007883.

Henrikus, S. S., van Oijen, A. M. and Robinson, A. (2018) 'Specialised DNA polymerases in Escherichia coli: roles within multiple pathways', *Current Genetics*. Springer Berlin Heidelberg, 64(6), pp. 1189–1196. doi: 10.1007/s00294-018-0840-x.

Hermans, P. G., Hart, C. A. and Trees, A. J. (2001) 'In vitro activity of antimicrobial agents against the endosymbiont Wolbachia pipientis.', *The Journal of antimicrobial chemotherapy*, 47(5), pp. 659–63. doi: 10.1093/JAC/47.5.659.

Hertig, M. and Wolbach, S. B. (1924) 'Studies on Rickettsia-Like Micro-Organisms in Insects.', *The Journal of medical research*, 44(3), pp. 329-374.7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19972605%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2041761.

Hien, N. T. *et al.* (2021) 'Environmental factors influence the local establishment of Wolbachia in Aedes aegypti mosquitoes in two small communities in central Vietnam', *Gates Open Research*, 5(May), p. 147. doi: 10.12688/gatesopenres.13347.1.

Hoffmann, A. A. *et al.* (2011) 'Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission', *Nature*. Nature Publishing Group, 476(7361), pp. 454–459. doi: 10.1038/nature10356.

Iturbe-Ormaetxe, I. *et al.* (2005) 'Distribution, expression, and motif variability of ankyrin domain genes in Wolbachia pipientis', *Journal of Bacteriology*, 187(15), pp. 5136–5145. doi: 10.1128/JB.187.15.5136-5145.2005.

Jacob, F. *et al.* (2005) 'The operon: A group of genes with expression coordinated by an operator', *Comptes Rendus - Biologies*, 328(6 SPEC. ISS.), pp. 514–520. doi: 10.1016/j.crvi.2005.04.005.

Jacob, F. and Monod, J. (1961) 'Genetic regulatory mechanisms in the synthesis of proteins', *Journal of Molecular Biology*, 3(3), pp. 318–356. doi: 10.1016/S0022-2836(61)80072-7.

Jewett, T. J. *et al.* (2006) 'Chlamydial TARP is a bacterial nucleator of actin', *Proceedings of the National Academy of Sciences of the United States of America*, 103(42), pp. 15599–15604. doi: 10.1073/pnas.0603044103.

Johnston, K. L. *et al.* (2014) 'Repurposing of approved drugs from the human pharmacopoeia to target Wolbachia endosymbionts of onchocerciasis and lymphatic filariasis', *International Journal for Parasitology: Drugs and Drug Resistance*. Australian Society for Parasitology, 4(3), pp. 278–286. doi: 10.1016/j.ijpddr.2014.09.001.

Joubert, D. A. *et al.* (2016) 'Establishment of a Wolbachia Superinfection in Aedes aegypti Mosquitoes as a Potential Approach for Future Resistance Management', *PLoS Pathogens*, 12(2), p. 1005434. doi: 10.1371/journal.ppat.1005434.

Ju, J.-F. *et al.* (2019) 'Wolbachia supplement biotin and riboflavin to enhance reproduction in planthoppers', *The ISME Journal*. Nature Publishing Group, pp. 1–12. doi: 10.1038/s41396-019-0559-9.

Kambris, Z. *et al.* (2009) 'Immune activation by life-shortening wolbachia and reduced filarial competence in mosquitoes', *Science*, 326(5949), pp. 134–136. doi: 10.1126/science.1177531.

Kambris, Z. *et al.* (2010) 'Wolbachia stimulates immune gene expression and inhibits plasmodium development in anopheles gambiae', *PLoS Pathogens*, 6(10). doi: 10.1371/journal.ppat.1001143.

Kampfraath, A. A. *et al.* (2019) 'Genome expansion of an obligate parthenogenesis-associated Wolbachia poses an exception to the symbiont reduction model', *BMC Genomics*. BMC Genomics, 20(1), pp. 1–14. doi: 10.1186/s12864-019-5492-9.

Kaur, S. J. *et al.* (2012) 'TolC-dependent secretion of an ankyrin repeat-containing protein of Rickettsia typhi', *Journal of Bacteriology*, 194(18), pp. 4920–4932. doi: 10.1128/JB.00793-12.

Langklotz, S., Baumann, U. and Narberhaus, F. (2012) 'Structure and function of the bacterial AAA protease FtsH', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier B.V., 1823(1), pp. 40–48. doi: 10.1016/j.bbamcr.2011.08.015.

Lefoulon, E. *et al.* (2012) 'A new type F Wolbachia from Splendidofilariinae (Onchocercidae) supports the recent emergence of this supergroup', *International Journal for Parasitology*. Australian Society for Parasitology Inc., 42(11), pp. 1025–1036. doi: 10.1016/j.ijpara.2012.09.004.

Lefoulon, E. *et al.* (2016) 'Breakdown of coevolution between symbiotic bacteria Wolbachia and their filarial hosts', *PeerJ*, 2016(3). doi: 10.7717/peerj.1840.

Lehman, S. S. *et al.* (2018) 'The rickettsial ankyrin repeat protein 2 is a type IV secreted effector that associates with the endoplasmic reticulum', *mBio*, 9(3). doi: 10.1128/mBio.00975-18.

Leitner, M. and Bishop, C. (2021) 'Transcriptional Response of Wolbachia to Dengue Virus Infection in Cells of the Mosquito Aedes aegypti'.

Lenski, R. E., Sniegowski, P. and Gerrish, P. (1997) 'Evolution of high mutation rates in experimental populations of E. coli', *Nature*, 387(June), pp. 703–705.

LePage, D. P. *et al.* (2017) 'Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility', *Nature*. Nature Publishing Group, 543(7644), pp. 243–247. doi: 10.1038/nature21391.

Li, G. *et al.* (2015) 'Study of efflux pump gene expression in rifampicin-monoresistant Mycobacterium tuberculosis clinical isolates', *Journal of Antibiotics*. Nature Publishing Group, 68(7), pp. 431–435. doi: 10.1038/ja.2015.9.

Li, G. M. (2008) 'Mechanisms and functions of DNA mismatch repair', *Cell Research*, 18(1), pp. 85–98. doi: 10.1038/cr.2007.115.

Li, X. Z., Plésiat, P. and Nikaido, H. (2015) 'The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria', *Clinical Microbiology Reviews*, 28(2), pp. 337–418. doi: 10.1128/CMR.00117-14.

Li, Z. and Carlow, C. K. S. (2012) 'Characterization of Transcription Factors That Regulate the Type IV Secretion System and Riboflavin Biosynthesis in Wolbachia of Brugia malayi', *PLoS ONE*, 7(12). doi: 10.1371/journal.pone.0051597.

Lindsey, A. R. I. *et al.* (2017) 'Evolutionary Genetics of Cytoplasmic Incompatibility Genes cifA and cifB in Prophage WO of Wolbachia', *bioRxiv*. doi: 10.1093/gbe/evy012.

Lindsey, Amelia R.I. (2020) 'Sensing, signaling, and secretion: A review and analysis of systems for regulating host interaction in wolbachia', *Genes*, 11(7), pp. 1–21. doi: 10.3390/genes11070813.

Liverman, A. D. B. *et al.* (2007) 'Arp2/3-independent assembly of actin by Vibrio type III effector VopL', *Proceedings of the National Academy of Sciences of the United States of America*, 104(43), pp. 17117–17122. doi: 10.1073/pnas.0703196104.

Luck, A. N. *et al.* (2014) 'Concurrent transcriptional profiling of Dirofilaria immitis and its Wolbachia endosymbiont throughout the nematode life cycle reveals coordinated gene expression', *BMC Genomics*. BioMed Central, 15(1), p. 1041. doi: 10.1186/1471-2164-15-1041.

Luck, A. N., Slatko, B. E. and Foster, J. M. (2017) 'Removing the needle from the haystack: Enrichment of Wolbachia endosymbiont transcripts from host nematode RNA by Cappable-seq???', *PLoS ONE*, 12(3), pp. 1–11. doi: 10.1371/journal.pone.0173186.

Maltz, M. A. *et al.* (2012) 'OmpA-mediated biofilm formation is essential for the commensal bacterium Sodalis glossinidius to colonize the tsetse fly gut', *Applied and Environmental Microbiology*, 78(21), pp. 7760–7768. doi: 10.1128/AEM.01858-12.

Martinez, J. *et al.* (2017) 'Symbiont strain is the main determinant of variation in Wolbachia-mediated protection against viruses across Drosophila species', *Molecular Ecology*, 26(15), pp. 4072–4084. doi: 10.1111/mec.14164.

Martinez, J. (2022) 'Genomic and phenotypic comparisons reveal distinct variants of Wolbachia strain wAlbB', pp. 1–46.

Masui, S. *et al.* (2001) 'Bacteriophage WO and virus-like particles in Wolbachia, an endosymbiont of arthropods', *Biochemical and Biophysical Research Communications*, 283(5), pp. 1099–1104. doi: 10.1006/bbrc.2001.4906.

Matic, I., Rayssiguier, C. and Radman, M. (1995) 'Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species', *Cell*, 80(3), pp. 507–515. doi: 10.1016/0092-8674(95)90501-4.

Mayoral, J. G., Hussain, M., Joubert, D. A., Iturbe-Ormaetxe, I., O'Neill, S. L., *et al.* (2014) 'Wolbachia small noncoding RNAs and their role in cross-kingdom communications.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(52), pp. 18721–6. doi: 10.1073/pnas.1420131112.

McClintock, B. (1984) 'The significance of responses of the genome to challenge', *Science*, 226(4676), pp. 792–801. doi: 10.1126/science.15739260.

McCutcheon, J. P. and Moran, N. A. (2012) 'Extreme genome reduction in symbiotic bacteria', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 13–26. doi: 10.1038/nrmicro2670.

McMeniman, C. J. *et al.* (2008) 'Host adaptation of a Wolbachia strain after long-term serial

passage in mosquito cell lines', *Applied and Environmental Microbiology*, 74(22), pp. 6963–6969. doi: 10.1128/AEM.01038-08.

Van Meer, M. M. M., Witteveldt, J. and Stouthamer, R. (1999) 'Phylogeny of the arthropod endosymbiont Wolbachia based on the wsp gene', *Insect Molecular Biology*, 8(3), pp. 399–408. doi: 10.1046/j.1365-2583.1999.83129.x.

Mejía-Almonte, C. *et al.* (2020) 'Redefining fundamental concepts of transcription initiation in bacteria', *Nature Reviews Genetics*. Springer US, 21(11), pp. 699–714. doi: 10.1038/s41576-020-0254-8.

Michalkova, V. *et al.* (2014) 'Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies', *Applied and Environmental Microbiology*, 80(18), pp. 5844–5853. doi: 10.1128/AEM.01150-14.

Min, K. T. and Benzer, S. (1997) 'Wolbachia, normally a symbiont of Drosophila, can be virulent, causing degeneration and early death', *Proceedings of the National Acadamy of Sciences U.S.A.*, 94(20), pp. 10792–10796. doi: 10.1073/pnas.94.20.10792.

Mittenhuber, G. (2002) 'An inventory of genes encoding RNA polymerase sigma factors in 31 completely sequenced eubacterial genomes', *Journal of Molecular Microbiology and Biotechnology*, 4(1), pp. 77–91.

Molloy, J. C. *et al.* (2016) 'Wolbachia modulates lipid metabolism in Aedes albopictus mosquito cells', *Applied and Environmental Microbiology*. American Society for Microbiology, 82(10), pp. 3109–3120. doi: 10.1128/AEM.00275-16.

Montllor, C. B., Maxmen, A. and Purcell, A. H. (2002) 'Facultative bacterial endosymbionts benefit pea aphids Acyrthosiphon pisum under heat stress', *Ecological Entomology*, 27(2), pp. 189–195. doi: 10.1046/j.1365-2311.2002.00393.x.

Moran, N. A. (1996) 'Accelerated evolution and Muller's rachet in endosymbiotic bacteria.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 93(7), pp. 2873–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8610134 (Accessed: 13 June 2018).

Moran, N. A., Tran, P. and Gerardo, N. M. (2005) 'Symbiosis and insect diversification: An ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes', *Applied and Environmental Microbiology*, 71(12), pp. 8802–8810. doi: 10.1128/AEM.71.12.8802-8810.2005.

Musolin, D. L., Tougou, D. and Fujisaki, K. (2010) 'Too hot to handle? Phenological and life-history responses to simulated climate change of the southern green stink bug Nezara viridula (Heteroptera: Pentatomidae)', *Global Change Biology*, 16(1), pp. 73–87. doi: 10.1111/j.1365-2486.2009.01914.x.

Namgoong, S. *et al.* (2011) 'Mechanism of actin filament nucleation by Vibrio VopL and implications for tandem W domain nucleation', *Nature Structural and Molecular Biology*. Nature Publishing Group, 18(9), pp. 1060–1067. doi: 10.1038/nsmb.2109.

Nazni, W. A. *et al.* (2019) 'Establishment of Wolbachia Strain wAlbB in Malaysian Populations of Aedes aegypti for Dengue Control', *Current Biology*. Elsevier Ltd., 29(24), pp. 4241-4248.e5. doi: 10.1016/j.cub.2019.11.007.

Nguyen, T. H. *et al.* (2015) 'Field evaluation of the establishment potential of wmelpop Wolbachia in Australia and Vietnam for dengue control', *Parasites and Vectors*. BioMed

Central, 8(1), p. 563. doi: 10.1186/s13071-015-1174-x.

O'Leary, N. A. *et al.* (2016) 'Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation', *Nucleic Acids Research*, 44(D1), pp. D733–D745. doi: 10.1093/nar/gkv1189.

O'Neill, S. L. *et al.* (1997) 'In vitro cultivation of Wolbachia pipientis in an Aedes albopictus cell line.', *Insect molecular biology*, 6, pp. 33–39. doi: 10.1046/j.1365-2583.1997.00157.x.

Oliva, G., Sahr, T. and Buchrieser, C. (2015) 'Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: Impact on metabolism and virulence', *FEMS Microbiology Reviews*, pp. 331–349. doi: 10.1093/femsre/fuv022.

Van Opijnen, T. and Breeuwer, J. A. J. (1999) 'High temperatures eliminate Wolbachia, a cytoplasmic incompatibility inducing endosymbiont, from the two-spotted spider mite', *Experimental and Applied Acarology*, 23(11), pp. 871–881. doi: 10.1023/A:1006363604916.

Organization, W. (2009) 'Dengue: Guidelines for Diagnosis, Treatment, Prevention'. doi: 10.1176/pn.41.1.0029b.

Osborne, S. E. *et al.* (2009) 'Variation in antiviral protection mediated by different Wolbachia strains in Drosophila simulans', *PLoS Pathogens*, 5(11). doi: 10.1371/journal.ppat.1000656.

Ote, M., Ueyama, M. and Yamamoto, D. (2016) 'Wolbachia Protein TomO Targets nanos mRNA and Restores Germ Stem Cells in Drosophila Sex-lethal Mutants', *Current Biology*, 26(17), pp. 2223–2232. doi: 10.1016/j.cub.2016.06.054.

Oulhen, N., Schulz, B. J. and Carrier, T. J. (2016) 'English translation of Heinrich Anton de Bary's 1878 speech, "Die Erscheinung der Symbiose" ("De la symbiose")', *Symbiosis*. Symbiosis, 69(3), pp. 131–139. doi: 10.1007/s13199-016-0409-8.

Oxitec (2016) 'Aedes aegypti OX513A', pp. 1–140.

Padalon-Brauch, G. *et al.* (2008) 'Small RNAs encoded within genetic islands of Salmonella typhimurium show host-induced expression and role in virulence', *Nucleic Acids Research*, 36(6), pp. 1913–1927. doi: 10.1093/nar/gkn050.

Pagès, V. and Fuchs, R. P. (2003) 'Uncoupling of leading- and lagging-strand DNA replication during lesion bypass in vivo', *Science*, 300(5623), pp. 1300–1303. doi: 10.1126/science.1083964.

Pan, X. *et al.* (2008) 'Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors', *Science*, 320(5883), pp. 1651–1654. doi: 10.1126/science.1158160.

Park, J. S. *et al.* (2012) 'Mechanism of anchoring of OmpA protein to the cell wall peptidoglycan of the gram-negative bacterial outer membrane', *The FASEB Journal*, 26(1), pp. 219–228. doi: 10.1096/fj.11-188425.

Prado, S. S. and Almeida, R. P. P. (2009) 'Phylogenetic placement of pentatomid stink bug gut symbionts', *Current Microbiology*, 58(1), pp. 64–69. doi: 10.1007/s00284-008-9267-9.

Rancès, E. *et al.* (2008) 'Genetic and functional characterization of the type IV secretion system in Wolbachia', *Journal of Bacteriology*. American Society for Microbiology, 190(14), pp. 5020–5030. doi: 10.1128/JB.00377-08.

Rao, R. U. *et al.* (2012) 'Effects of Doxycycline on gene expression in Wolbachia and Brugia

malayi adult female worms in vivo', *Journal of Biomedical Science*, 19(1), p. 21. doi: 10.1186/1423-0127-19-21.

Reed, K. M. and Werren, J. H. (1995) 'Induction of paternal genome loss by the paternal-sex-ratio chromosome and cytoplasmic incompatibility bacteria (Wolbachia): A comparative study of early embryonic events', *Molecular Reproduction and Development*, 40(4), pp. 408–418. doi: 10.1002/mrd.1080400404.

Reynolds, K. T., Thomson, L. J. and Hoffmann, A. A. (2003) 'The effects of host age, host nuclear background and temperature on phenotypic effects of the virulent wolbachia strain popcorn in Drosophila melanogaster', *Genetics*, 164(3), pp. 1027–1034. doi: 10.1126/science.8511587.

Rice, D. W., Sheehan, K. B. and Newton, I. L. G. (2017) 'Large-scale identification of Wolbachia pipientis effectors', *Genome Biology and Evolution*, 9(September), pp. 143–149. doi: 10.1093/gbe/evx139.

Richardson, K. M. *et al.* (2013) 'A replicated comparison of breeding-container suitability for the dengue vector Aedes aegypti in tropical and temperate Australia', *Austral Ecology*, 38(2), pp. 219–229. doi: 10.1111/j.1442-9993.2012.02394.x.

Richter, K., Haslbeck, M. and Buchner, J. (2010) 'The Heat Shock Response: Life on the Verge of Death', *Molecular Cell*, pp. 253–266. doi: 10.1016/j.molcel.2010.10.006.

Roma, J. S. *et al.* (2019) 'Thermal stress responses of Sodalis glossinidius, an indigenous bacterial symbiont of hematophagous tsetse flies', *PLoS Neglected Tropical Diseases*, 13(11), pp. 1–20. doi: 10.1371/journal.pntd.0007464.

Ross, P. A. *et al.* (2017) 'Wolbachia Infections in Aedes aegypti Differ Markedly in Their Response to Cyclical Heat Stress', *PLoS Pathogens*, 13(1), pp. 1–17. doi: 10.1371/journal.ppat.1006006.

Ross, P. A. *et al.* (2020) 'Heatwaves cause fluctuations in wMel Wolbachia densities and frequencies in Aedes aegypti', *PLOS Neglected Tropical Diseases*. Edited by A. Kohl, 14(1), p. e0007958. doi: 10.1371/journal.pntd.0007958.

Ross, P. A. *et al.* (2022) 'A decade of stability for wMel Wolbachia in natural Aedes aegypti populations', *PLoS Pathogens*, 18(2), pp. 1–18. doi: 10.1371/journal.ppat.1010256.

Ross, P. A., Turelli, M. and Hoffmann, A. A. (2019) 'Evolutionary Ecology of Wolbachia Releases for Disease Control'. doi: 10.1146/annurev-genet-112618.

Russo, D. M. *et al.* (2006) 'Proteins exported via the PrsD-PrsE type I secretion system and the acidic exopolysaccharide are involved in biofilm formation by Rhizobium leguminosarum', *Journal of Bacteriology*, 188(12), pp. 4474–4486. doi: 10.1128/JB.00246-06.

Sagan, L. (1967) 'On the origin of mitosing cells', *Journal of Theoretical Biology*, 14(3). doi: 10.1016/0022-5193(67)90079-3.

Santangelo, M. D. L. P. *et al.* (2011) 'Glycolytic and non-glycolytic functions of Mycobacterium tuberculosis fructose-1,6-bisphosphate aldolase, an essential enzyme produced by replicating and non-replicating bacilli', *Journal of Biological Chemistry*, 286(46), pp. 40219–40231. doi: 10.1074/jbc.M111.259440.

Sarah K. Henke and John E. Cronan (2016) 'The Staphylococcus aureus group II biotin

protein ligase BirA is an effective regulator of biotin operon transcription and requires the DNA binding domain for full enzymatic activity', *Mol Microbiol*, 176(5), pp. 139–148. doi: 10.1111/mmi.13470.The.

Schlüter, J. P. *et al.* (2013) 'Global mapping of transcription start sites and promoter motifs in the symbiotic α-proteobacterium Sinorhizobium meliloti 1021', *BMC Genomics*, 14(1). doi: 10.1186/1471-2164-14-156.

Schroeder, C. L. C. *et al.* (2016) 'Identification and characterization of novel small RNAs in Rickettsia prowazekii', *Frontiers in Microbiology*, 7(JUN), pp. 1–14. doi: 10.3389/fmicb.2016.00859.

Schürmann, M. and Sprenger, G. A. (2001) 'Fructose-6-phosphate Aldolase is a Novel Class I Aldolase from Escherichia coli and is Related to a Novel Group of Bacterial Transaldolases', *Journal of Biological Chemistry*, 276(14), pp. 11055–11061. doi: 10.1074/jbc.M008061200.

Shan, J. *et al.* (2014) 'Temperature dependent bacteriophages of a tropical bacterial pathogen', *Frontiers in Microbiology*, 5(NOV), pp. 1–7. doi: 10.3389/fmicb.2014.00599.

Sharma, C. M. *et al.* (2010) 'The primary transcriptome of the major human pathogen Helicobacter pylori', *Nature*, 464(7286), pp. 250–255. doi: 10.1038/nature08756.

Shaw, K. J. *et al.* (2003) 'Comparison of the changes in global gene expression of Escherichia coli induced by four bactericidal agents', *Journal of Molecular Microbiology and Biotechnology*, 5(2), pp. 105–122. doi: 10.1159/000069981.

Sheehan, K. B. *et al.* (2016) 'Identification and Characterization of a Candidate *Wolbachia pipientis* Type IV Effector That Interacts with the Actin Cytoskeleton', *mBio*, 7(4), pp. e00622-16. doi: 10.1128/mBio.00622-16.

Shropshire, J. D. and Bordenstein, S. R. (2019) 'Two-By-One model of cytoplasmic incompatibility: Synthetic recapitulation by transgenic expression of cifA and cifB in Drosophila', *PLOS Genetics*. Edited by A. A. Hoffmann. Public Library of Science, 15(6), p. e1008221. doi: 10.1371/journal.pgen.1008221.

Siddiqi, N. *et al.* (2004) 'Mycobacterium tuberculosis isolate with a distinct genomic identity overexpresses a tap-like efflux pump', *Infection*, 32(2), pp. 109–111. doi: 10.1007/s15010-004-3097-x.

Siguier, P., Filée, J. and Chandler, M. (2006) 'Insertion sequences in prokaryotic genomes', *Current Opinion in Microbiology*, 9(5), pp. 526–531. doi: 10.1016/j.mib.2006.08.005.

Sinha, A. *et al.* (2019) 'Complete Genome Sequence of the Wolbachia wAlbB Endosymbiont of Aedes albopictus', *Genome Biology and Evolution*. Oxford University Press, 11(3), pp. 706–720. doi: 10.1093/gbe/evz025.

Sinkins, S. P. *et al.* (2005) 'Wolbachia variability and host effects on crossing type in Culex mosquitoes', *Nature*, 436(7048), pp. 257–260. doi: 10.1038/nature03629.

Streaker, E. D., Gupta, A. and Beckett, D. (2002) 'The biotin repressor: Thermodynamic coupling of corepressor binding, protein assembly, and sequence-specific DNA binding', *Biochemistry*, 41(48), pp. 14263–14271. doi: 10.1021/bi0203839.

Tamas, I. *et al.* (2002) '50 million years of genomic stasis in endosymbiotic bacteria.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 296(5577),

pp. 2376–9. doi: 10.1126/science.1071278.

Taylor, M. J., Bandi, C. and Hoerauf, A. (2005) 'Wolbachia bacterial endosymbionts of filarial nematodes', *Advances in Parasitology*, 60, pp. 245–284. doi: 10.1016/S0065-308X(05)60004-8.

Teixeira, L., Ferreira, Á. and Ashburner, M. (2008) 'The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster', *PLoS Biology*. Edited by L. Keller. Plenum Press, 6(12), pp. 2753–2763. doi: 10.1371/journal.pbio.1000002.

Thomason, M. K. and Storz, G. (2010) 'Bacterial Antisense RNAs: How Many Are There, and What Are They Doing?', *Annual Review of Genetics*, 44(1), pp. 167–188. doi: 10.1146/annurev-genet-102209-163523.

Tomoyasu, T. *et al.* (1995) 'Escherichia coli FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor σ32', *EMBO Journal*, 14(11), pp. 2551–2560. doi: 10.1002/j.1460-2075.1995.tb07253.x.

Trotochaud, A. E. and Wassarman, K. M. (2004) '6S RNA function enhances long-term cell survival', *Journal of Bacteriology*, 186(15), pp. 4978–4985. doi: 10.1128/JB.186.15.4978-4985.2004.

Ulrich, J. N. *et al.* (2016) 'Heat Sensitivity of wMel Wolbachia during Aedes aegypti Development', *PLoS Neglected Tropical Diseases*, 10(7). doi: 10.1371/journal.pntd.0004873.

Veltman, D. M. and Insall, R. H. (2010) 'WASP family proteins: Their evolution and its physiological implications', *Molecular Biology of the Cell*, 21(16), pp. 2880–2893. doi: 10.1091/mbc.E10-04-0372.

Verghese (2011) 'The Outer Membrane Protein A (OmpA) of Y. pestis promotes intracellular survival and virulence in mice', *Bone*, 23(1), pp. 1–7. doi: 10.1016/j.micpath.2011.09.009.The.

VieBrock, L. *et al.* (2014) 'Orientia tsutsugamushi ankyrin repeat-containing protein family members are type 1 secretion system substrates that traffic to the host cell endoplasmic reticulum', *Frontiers in Cellular and Infection Microbiology*, 4(DEC), pp. 1–20. doi: 10.3389/fcimb.2014.00186.

Voronin, D. *et al.* (2010) 'Transinfection and growth discrepancy of Drosophila Wolbachia strain wMel in cell lines of the mosquito Aedes albopictus', *Journal of Applied Microbiology*, 108(6), pp. 2133–2141. doi: 10.1111/j.1365-2672.2009.04621.x.

Wakeel, A. *et al.* (2011) 'Ehrlichia chaffeensis Tandem Repeat Proteins and Ank200 are Type 1 Secretion System Substrates Related to the Repeats-in-Toxin Exoprotein Family', *Frontiers in Cellular and Infection Microbiology*, 1(December), p. 22. doi: 10.3389/fcimb.2011.00022.

Walker, G. C. (1984) 'Mutagenesis and inducible responses to deoxyribonucleic acid damage in Escherichia coli', *Microbiological Reviews*, 48(1), pp. 60–93. doi: 10.1128/mmbr.48.1.60-93.1984.

Walker, T. *et al.* (2011) 'The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations', *Nature*. Nature Publishing Group, 476(7361), pp. 450–453. doi: 10.1038/nature10355.

Wassarman, K. M. and Storz, G. (2000) '6S RNA regulates E. coli RNA polymerase

activity', *Cell*, 101(6), pp. 613–623. doi: 10.1016/S0092-8674(00)80873-9.

Weinert, L. A. *et al.* (2015) 'The incidence of bacterial endosymbionts in terrestrial arthropods', *Proceedings of the Royal Society B: Biological Sciences*, 282(1807), pp. 3–8. doi: 10.1098/rspb.2015.0249.

Weiss, B. L. *et al.* (2008) *An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A*. Available at: www.pnas.orgcgidoi10.1073pnas.0805666105 (Accessed: 1 January 2020).

White, P. M. *et al.* (2017) 'Mechanisms of horizontal cell-to-cell transfer of Wolbachia spp. in Drosophila melanogaster', *Applied and Environmental Microbiology*, 83(7). doi: 10.1128/AEM.03425-16.

Wilcox, J. L. *et al.* (2003) 'Consequences of reductive evolution for gene expression in an obligate endosymbiont', *Molecular Microbiology*, 48(6), pp. 1491–1500. doi: 10.1046/j.1365-2958.2003.03522.x.

Wiwatanaratanabutr, S. and Kittayapong, P. (2006) 'Effects of temephos and temperature on Wolbachia load and life history traits of Aedes albopictus', *Medical and Veterinary Entomology*, 20(3), pp. 300–307. doi: 10.1111/j.1365-2915.2006.00640.x.

Woolfit, M. *et al.* (2013) 'Genomic evolution of the pathogenic Wolbachia strain, wMelPop', *Genome Biology and Evolution*, 5(11), pp. 2189–2204. doi: 10.1093/gbe/evt169.

Woolfit, M. *et al.* (2015) 'Discovery of putative small non-coding RNAs from the obligate intracellular bacterium Wolbachia pipientis', *PLoS ONE*, 10(3), pp. 1–19. doi: 10.1371/journal.pone.0118595.

Wu, M. *et al.* (2004) 'Phylogenomics of the Reproductive Parasite Wolbachia pipientis wMel: A Streamlined Genome Overrun by Mobile Genetic Elements', *PLoS Biology*. Edited by Nancy A. Moran. Oxford University Press, 2(3), p. e69. doi: 10.1371/journal.pbio.0020069.

Yan, B. *et al.* (2018) 'SMRT-Cappable-seq reveals complex operon variants in bacteria', *Nature Communications*, 9(1). doi: 10.1038/s41467-018-05997-6.

Yen, J. H. and Barr, A. R. (1971) 'New hypothesis of the cause of cytoplasmic incompatibility in Culex pipiens L.', *Nature*, 232, pp. 657–658. doi: 10.1038/232657a0.

Zeevi, M. K. *et al.* (2013) 'Listeria monocytogenes multidrug resistance transporters and cyclic Di-AMP, which contribute to type I interferon induction, play a role in cell wall stress', *Journal of Bacteriology*, 195(23), pp. 5250–5261. doi: 10.1128/JB.00794-13.

Zhukova, M. V., Voronin, D. A. and Kiseleva, E. V. (2008) 'High temperature initiates changes in Wolbachia ultrastructure in ovaries and early embryos of Drosophila melanogaster', *Cell and Tissue Biology*. SP MAIK Nauka/Interperiodica, 2(5), pp. 546–556. doi: 10.1134/S1990519X08050131.

Zilber-Rosenberg, I. and Rosenberg, E. (2008) 'Role of microorganisms in the evolution of animals and plants: The hologenome theory of evolution', *FEMS Microbiology Reviews*. Oxford University Press, pp. 723–735. doi: 10.1111/j.1574-6976.2008.00123.x.

Zuker, M. and Stiegler, P. (1981) 'Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information', *Nucleic Acids Research*, 9(1), pp. 133–148. doi: 10.1093/nar/9.1.133.