The Microbiome of the Tsetse: Metabolic adaptation and the evolution of symbiosis

Rebecca Jane Hall

Doctor of Philosophy

University of York
Biology
September 2019

Abstract

Bacteria are found in symbiosis with insects of economic and medical importance. The symbionts provide a range of benefits for their hosts, including metabolic supplementation and protection from predation. Consequently, complex interactions can be observed between host and microbiome. The tsetse, genus *Glossina*, is the vector for the protozoan parasite *Trypanosoma brucei*. The tsetse has a unique microbiome, consisting of the obligate *Wigglesworthia glossinidia* and the facultative *Sodalis glossinidius*. The tsetse is thought to rely on *W. glossinidia* for the production of B group vitamins, but any benefit of *S. glossinidius* is yet to be determined. Elucidating key metabolic interactions within the tsetse microbiome, and understanding how these relationships evolved, may open up new avenues for disease control.

To this end, we have constructed and tested metabolic models for *S. glossinidius* and its free-living relative, *Sodalis praecaptivus*. The former revealed an intriguing network of metabolic dependencies within the microbiome. *S. glossinidius* was shown to depend on thiamine produced by *W. glossinidia*, the chitin monomer *N*-acetyl-D-glucosamine found in abundance within the tsetse, and amino acids from the blood meal. The *S. praecaptivus* model was evolved using a multi-objective evolutionary algorithm to explore possible evolutionary trajectories of the *Sodalis* genus. It was discovered that certain pseudogenisations in *S. glossinidius*, once thought to be pivotal, may have arisen early in the symbiosis with minimal effect. Finally, substrate binding proteins in *W. glossinidia* were examined. Whilst functional assays were inconclusive, phylogenetics unexpectedly revealed new information about the relationship of *W. glossinidia* to other Enterobacteriaeae.

The results presented here demonstrate the interconnected metabolic network within the tsetse microbiome, and provide a platform for investigating other problems in metabolism, adaptation, and evolution in insect-bacterial symbioses.

Contents

1 Introduction **1**

1.1 An introduction to symbiosis 1

1.2 Insect-microbe symbioses 3

1.2.1 Primary symbiosis 4

1.2.2 Secondary symbiosis 7

1.2.3 Parasitism 10

1.3 The tsetse and its microbiome 11

1.3.1 The tsetse and human African trypanosomiasis 11

1.3.2 The primary symbiont *Wigglesworthia glossinidia* 14

1.3.3 The secondary symbiont *Sodalis glossinidius* 15

1.4 The *Sodalis* genus 16

1.5 Tools for studying symbiosis 19

1.5.1 Traditional tools 19

1.5.2 Flux balance analysis 20

1.5.3 Defined media for physiological studies 23

1.5.4 *In silico* evolution to study the initiation and evolutionary trajectory of symbiosis 25

1.6 Chapter summary 26

1.6.1 Chapter 2 - A Tale of Three Species: Adaptation of *Sodalis glossinidius* to tsetse biology, *Wigglesworthia* metabolism and host diet . . . 26

1.6.2 Chapter 3 - A genome scale metabolic model for *Sodalis praecaptivus* reveals unusual biosynthetic pathways . . . 27

1.6.3 Chapter 4 - The *in silico* evolution of *Sodalis glossinidius* . . . 28

1.6.4 Chapter 5 - Phylogenetic and experimental analysis of amino acid transporters in *Wigglesworthia glossinidia* . . . 28

2 A Tale of Three Species: Adaptation of *Sodalis glossinidius* to tsetse biology, *Wigglesworthia* metabolism and host diet 30

2.1 Introduction . . . 30

2.2 Methods . . . 33

2.2.1 Refinement of the *S. glossinidius* metabolic network . . . 33

2.2.2 Flux balance analysis . . . 33

2.2.3 Bacterial strains, growth conditions and reagents . . . 34

2.2.4 Flow cytometry . . . 35

2.2.5 Statistical analysis . . . 35

2.3 Results . . . 35

2.3.1 The genome of *S. praecaptivus* enables an improved analysis of the *S. glossinidius* metabolic network . . . 35

2.3.2 A revised metabolic model, *i*LF517, for *S. glossinidius* . . . 38

2.3.3 A defined medium, SGM11, supports *S. glossinidius* growth . 39

2.3.4 *S. glossinidius* maintains a reliance on a sugar, namely the host-derived *N*-acetyl D-glucosamine . . . 40

2.3.5 *S. glossinidius* has adapted to thiamine produced by the primary tsetse symbiont . . . 42

2.3.6 *S. glossinidius* is dependent on external sources of L-glutamate or L-aspartate 44

2.3.7 *S. glossinidius* is not an L-arginine auxotroph 45

2.4 Discussion 47

3 A genome scale metabolic model for *Sodalis praecaptivus* reveals unusual biosynthetic pathways 51

3.1 Introduction 51

3.2 Methods 53

3.2.1 Bacterial strains, growth conditions and reagents 53

3.2.2 Construction of the *S. praecaptivus* metabolic network 54

3.2.3 Flux balance analysis 55

3.2.4 Reaction constraints 55

3.2.5 Solving futile cycles 57

3.2.6 Robustness analysis 57

3.3 Results 58

3.3.1 Construction of a metabolic model for *S. praecaptivus* 58

3.3.2 A large scale, iterative process of model testing 59

3.3.3 *S. praecaptivus* can grow on the unusual sugar alcohol xylitol . 61

3.3.4 Use of *N*-acetyl D-galactosamine by *S. praecaptivus* 63

3.3.5 A new model of *S. praecaptivus* metabolism, *i*RH830 65

3.3.6 Robustness analysis of the *S. praecaptivus* metabolic network . 68

3.3.7 Media provisioning affects individual reaction flux 72

3.4 Discussion 75

4 The *in silico* evolution of *Sodalis glossinidius* 78

4.1 Introduction . . . 78
4.2 Methods . . . 81
4.2.1 Flux balance analysis . . . 81
4.2.2 Implementation of multi-objective evolutionary algorithm . . . 81
4.2.3 MOEA variations . . . 83
4.2.4 Analysis of evolved populations . . . 84
4.3 Results . . . 84
4.3.1 Media provisioning affects evolutionary trajectories . . . 84
4.3.2 Temporal gene loss can be predicted . . . 88
4.3.3 A prediction of the evolutionary future of *S. glossinidius* as a symbiont . . . 91
4.3.4 Fluctuating the nutrient availability affects the trajectory and final population . . . 94
4.4 Discussion . . . 96

5 Phylogenetic and experimental analysis of amino acid transporters in *Wigglesworthia glossinidia* 100

5.1 Introduction . . . 100
5.2 Methods . . . 105
5.2.1 Phylogenetics . . . 105
5.2.2 Gene cloning . . . 105
5.2.3 Transformation . . . 106
5.2.4 Colony polymerase chain reaction . . . 106
5.2.5 Agarose gel electrophoresis . . . 107
5.2.6 Sodium dodecyl sulfate polyacrylamide gel electrophoresis . . . 107
5.2.7 Expression of *WIGMOR_0593* . . . 109

5.2.8 Sonication . . . 109
5.2.9 Native purification . . . 109
5.2.10 Denaturing purification . . . 110
5.2.11 Buffer exchange . . . 110
5.2.12 Calculating protein concentration . . . 111
5.2.13 Circular dichroism spectrophotometry . . . 111
5.2.14 Peptide mass fingerprinting . . . 111
5.2.15 Crystallisation trials . . . 111
5.2.16 Ligand binding assays . . . 112
5.3 Results . . . 112
5.3.1 Identification of significant genetic drift within the *W. glossinidia* species . . . 112
5.3.2 Phylogenetic analysis of ABC transporters in the Enterobacteriaceae . . . 114
5.3.3 The *W. glossinidia* SBPs are within a clade of unknown binding specificity . . . 117
5.3.4 Expression and purification of the SBPs from *W. glossinidia* . 119
5.3.5 No evidence of L-proline binding by WIGMOR_0593 . . . 127
5.3.6 Crystallisation trials . . . 128
5.4 Discussion . . . 129

6 Discussion **132**

6.1 The Spectrum of Secondary Symbioses: A study in *Sodalis* . . . 132
6.2 When is a pseudogene not a pseudogene? . . . 135
6.3 How good a tool is FBA? . . . 137
6.4 The initiation of symbiosis . . . 138

6.5 To primary symbiosis and beyond 141

6.6 Conclusions 143

7 Appendices **145**

7.1 Appendix A - Chapter 2 145

7.2 Appendix B - Chapter 4 150

7.3 Appendix C - Chapter 5 151

8 Abbreviations **152**

List of Tables

1.1 Primary symbionts discussed in Chapter 1 4

1.2 Members of the *Sodalis* genus 18

2.1 L-arginine biosynthesis pathway in *S. glossinidius* 46

3.1 Media used in Chapter 3 54

3.2 *In silico* blood medium 56

3.3 Positive results from the *S. praecaptivus* phenotypic screen 59

3.4 Negative results from the *S. praecaptivus* phenotypic screen and *in silico* test 60

3.5 Results that were negative in the *S. praecaptivus in vitro* screen but positive *in silico* 61

3.6 Candidate xylitol and GalNAc degradation pathway components in *E. coli* and *S. glossinidius* 68

4.1 *In silico* evolution conditions 84

5.1 Buffers and media used in Chapter 5 106

5.2 Colony PCR reaction mixture 107

5.3 Colony PCR cycle 107

5.4 SDS polyacrylamide gel recipe 108

5.5 Native nickel affinity purification protocol 110

5.6 Denaturing nickel affinity purification protocol 110

5.7 WIGMOR_0593 PMF 125

S1 Pantothenate biosynthesis pathway in *S. glossinidius* 146

S2 Biotin biosynthesis pathway in *S. glossinidius* 146

S3 Riboflavin biosynthesis pathway in *S. glossinidius* 146

S4 Protoheme biosynthesis pathway in *S. glossinidius* 146

S5 PLP biosynthesis pathway in *S. glossinidius* 146

S6 Thiamine biosynthesis pathway in *S. glossinidius* 147

S7 Tetrahydrofolate biosynthesis pathway in *S. glossinidius* 147

S8 Cobalamin biosynthesis pathway in *S. glossinidius* 147

S9 Nicotinamide biosynthesis pathway in *S. glossinidius* 147

S10 Amino acid biosynthesis and transport in *S. glossinidius* 148

S11 Fluxes through the uptake reactions in ancestral and evolved *i*RH830 models 150

S12 Fluxes through the uptake reactions in ancestral and evolved *i*LF517 models 150

List of Figures

1.1 The nature of symbiotic relationships 2

1.2 SEM of *B. aphidicola* in *A. pisum* bacteriocytes 5

1.3 SEM of *S. symbiotica* in the *A. pisum* bacteriome 9

1.4 *Wolbachia*-induced CI 11

1.5 The tsetse 12

1.6 *T. brucei*, *P. falciparum*, *Leishmania*, *Schistosoma* life cycles 13

1.7 The tsetse microbiome 15

1.8 Flow chart of metabolic model construction 22

2.1 Central metabolism in *S. glossinidius* 36

2.2 *S. glossinidius ppc* viewed in Artemis 37

2.3 Biomass output from *i*EB458 with changing oxygen 39

2.4 End-point *in vitro S. glossinidius* growth 40

2.5 Predicted PTS in *S. glossinidius* 42

2.6 Vitamin biosynthesis pathways in *Sodalis* species 43

2.7 Growth of *E. coli* knockout mutants from the Keio collection 47

3.1 Construction of *i*RH830 58

3.2 Growth of *S. praecaptivus* on xylitol 62

3.3 Putative xylitol degradation pathway in *S. praecaptivus* 63

3.4 Growth of *S. praecaptivus* on GalNAc 64

3.5 Putative GalNac degradation pathway in *S. praecaptivus* 65

3.6 Biomass output for *i*RH830 with varying oxygen levels 66

3.7 Robustness analysis of *i*RH830 69

3.8 Essential reactions for amino acid metabolism in *i*RH830 71

3.9 Robustness analysis of *i*LF517 72

3.10 Flux through central metabolism in *i*RH830 74

3.11 Flux through reactions for amino acid metabolism in *i*RH830 75

4.1 Process of the MOEA 82

4.2 *i*RH830 evolved under different media conditions 86

4.3 Core non-essential reactions in evolved *i*RH830 under different media conditions 88

4.4 Evolution of knockouts of *i*RH830 90

4.5 Core non-essential reactions in evolved knockouts of *i*RH830 91

4.6 Evolution of *i*LF517 92

4.7 Biomass output and number of reactions with flux in evolved *i*RH830 and evolved *i*LF517 93

4.8 Fluxes through uptake reactions in ancestral and evolved *i*RH830 and *i*LF517 94

4.9 *i*RH830 evolution under cycling media conditions 96

5.1 L-proline and L-glutamate transport and catabolism in *E. coli* 102

5.2 The oxidative branch of the pentose phosphate pathway 103

5.3 Schematic of the structure of ABC transporters 104

5.4 Inter- and intra-genus protein sequence comparison 114

5.5 Maximum likelihood phylogenetic tree of SBPs in the Enterobacteriaceae 116

5.6 Phylogenetic cluster containing the candidate *W. glossinidia* SBPs . . 118

5.7 Protein sequence and predicted structure of WIGMOR_0593 120

5.8 WIGMOR_0593 signal peptide prediction 121

5.9 *WIGMOR_0593* expression trials 122

5.10 Purified SBPs from *W, glossinidia morsitans*, *W. glossinidia brevipalpis*, and *Ca.* Sodalis pierantonius 124

5.11 Denaturing purification and buffer exchange for WIGMOR_0593 . . 126

5.12 CD spectra of native and refolded WIGMOR_0593 127

5.13 Thermal shift assay of amino acid binding by WIGMOR_0593 128

5.14 WIGMOR_0593 crystallisation trials 129

S1 Flow cytometer gating for *S. glossinidius* cell count 145

S2 Time-course growth of *E. coli* knockout mutants from the Keio collection 149

S3 pET100/D-TOPO vector map 151

Accompanying Materials

Four .xls files are provided as accompanying materials:

Supplementary File 1 : *i*LF517

Supplementary File 2 : *i*RH830

Supplementary File 3 : Core non-essential reactions in evolved *i*LF517 and evolved *i*RH830 solutions

Supplementary File 4: Fluxes through evolved *i*LF517 and evolved *i*RH830 solutions

Acknowledgements

First, I would like to thank my supervisors, Jamie Wood and Gavin Thomas. They provided me with excellent support throughout my PhD and during the thesis writing process, enabling me to develop as a scientist, and I would not be where I am now without them.

My thesis has benefited from help and insightful input by my TAP members, Pegine Walrad and James Moir, and I am grateful for their time and expertise.

The last four years have been made significantly easier and more enjoyable by the Thomas lab, past and present. Particular thanks go to Caroline, Reyme, Emm, Michelle and Sophie.

Thank you to the BBSRC for funding this project.

I would like to give a wholehearted thank you to some wonderful friends who have kept me going, particularly Harriet, Fiona, Amy, Nat and Sophie, to Miss Cook for inspiring my love of Biology, and to Sabrina, to whom this work is dedicated.

Finally, I would like to give special thanks to Steve for his unwavering belief, encouragement and support, both personal and technical.

Author Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

The following paper has been published and is presented in Chapter 2:

Hall RJ, Flanagan LA, Bottery MJ, Springthorpe V, Thorpe S, Darby AC, Wood AJ, Thomas GH (2019). A Tale of Three Species: Adaptation of *Sodalis glossinidius* to tsetse biology, *Wigglesworthia* metabolism and host diet. mBio 10 (1) e02106-18.

The work presented in Chapter 4 was conducted in collaboration with Stephen Thorpe (Biology, University of York). Chapters 3 and 4 form a submitted manuscript, available as a pre-print on bioRxiv (doi 10.1101/819946), as:

Hall RJ*, Thorpe S*, Thomas GH and Wood AJ. Simulating the evolutionary trajectories of metabolic pathways for insect symbionts in the *Sodalis* genus.

Author contributions are given at the start of chapters, where appropriate.

Chapter 1

Introduction

1.1 An introduction to symbiosis

Symbioses are ubiquitous. The term "symbiosis" comes from the Greek for "living together", and is used to describe two or more individuals of different species that live in close association, often to the benefit of all. One of the first uses of the term in a biological context was by Heinrich Anton de Bary in his 1878 speech *Die Erscheinung der Symbiose* (De Bary, 1879) (*The Phenomenon of Symbiosis*, translated into English by Oulhen et al. (2016)). There, De Bary used the word "symbiosis" to describe different organisms that live together in a relationship that is mutually beneficial. The nature of symbioses vary depending on the individuals involved. If all organisms benefit from the symbiosis, this relationship is termed mutualistic (Fig. 1.1). Commensal symbioses arise when one organism derives a positive benefit from the relationship with no gain or loss to their symbiotic partner. A relationship is described as parasitic if one organism benefits to the detriment of the other.

The beneficial nature of symbioses means that the associations can persist for millions of years. A symbiotic event 1.5 billion years ago, the engulfing of an α-proteobacterium by an archaeon, is thought to have kickstarted all complex life on Earth. This event formed the mitochrondria that are found in modern day eukaryotes. A similar acquisition, this time of a cyanobacterium, led to the formation of chloroplasts. The concept of endosymbiosis, where one organism lives inside another, was initially conceived by Constantin Merechowsky and was subsequently spearheaded by Lynn Margulis. Margulis was instrumental in driving the theory

of endosymbiosis when she published her keystone text *On the Origin of Mitosing Cells* in 1967 (Sagan, 1967). Margulis curated her Serial Endosymbiotic Theory to describe the origin of nucleated cells, whereby repeated endosymbiosis events led to the formation of all components of the eukaryotic cell (Sagan, 1967; Margulis, 1993). Whilst certain parts of her theory, notably regarding the origin of flagella, have not been verified (Lane, 2017), her work cemented the importance of symbiosis in the field of biology.

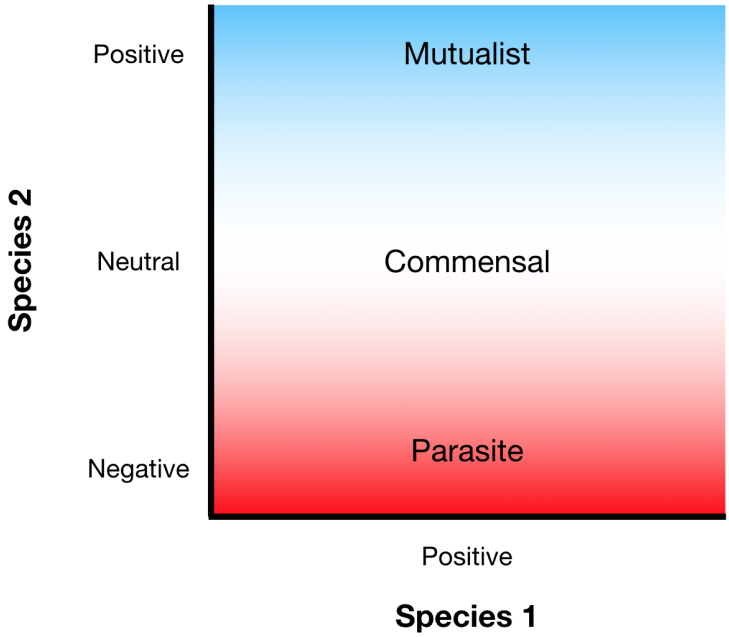


Figure 1.1: Symbioses can be mutualistic, commensal or parasitic, depending on their effect on the other organism.

Bacteria were crucial in the earliest symbioses, and they continue to form important symbiotic relationships today. Bacteria are found in symbiosis with individuals across the domains of life. The human microbiota has been a flourishing area of scientific research, epitomised by the National Institutes of Health Human Microbiome Project (HMP) (Methé et al., 2012; Huttenhower et al., 2012). The HMP was founded in part to characterise a healthy microbiome, acknowledging the close association between human and bacteria (Gevers et al., 2012). Bacteria are also found in symbiosis with plants. Legumes, for example, benefit from the nitrogen fixation carried out by the bacteria in the rhizosphere, with the latter receiving sugars from the plant in return (Franche et al., 2009). It is however the symbioses that bacteria form with insects that will form the basis of the work presented here. Bacteria are found on and within a wide range of different insects, with a variety of roles and adaptations. Relationships between insect and symbiont will be examined, inter-

preted, and discussed here, with a focus on metabolic adaptations and the evolution of the symbiotic relationship.

1.2 Insect-microbe symbioses

Insect-bacterial symbioses are an attractive area of research. The microbiomes of insects are often comprised of only a few individual species of bacteria. Insect life cycles can be controlled, monitored and replicated strictly. Insects are also of medical and economic importance. Some are vectors of diseases including malaria, dengue fever and yellow fever (Holt et al., 2002; Nene et al., 2007; Mellor et al., 2000), and others are major crop pests (Oerke, 2006). Building upon existing knowledge of insect microbiomes is therefore a field of research that is expanding rapidly.

The insect plays a vital and active role in the symbiotic relationship. The nature of the host influences, for example, the mechanism of symbiont transmission. For many insects, their symbionts are vertically transmitted from mother to offspring (Toju et al., 2010; Kaiwa et al., 2014; Ly Thao and Baumann, 2004a; Aksoy et al., 1997; Dale et al., 2001). The offspring of other, often social insects, are naturally symbiont-free and acquire their microbiome from their environment (Martinson et al., 2011). A bacterial microbiome provides several benefits to an insect, which will be discussed in depth later in this chapter. Regardless, it is important for the insect to constrain the bacterial population. This maintains the symbiotic nature of the relationship, preventing the microbiome from becoming pathogenic. Insect immune systems are comparable to innate immunity in mammals; non-specific and without memory. The insects cannot therefore rely on recognition of the commensal symbionts in order to either control the population size or protect them from elimination. Some insects, including the weevil *Sitophilus oryzae*, produce antimicrobial peptides that can selectively target the symbiotic bacteria to restrict the size of the population (Login et al., 2011). Other insects, such as the aphid *Acyrthosiphon pisum*, actively protect their symbionts. *A. pisum* degrades the immunogenic peptidoglycan fragments produced by its microbiota, preventing the symbionts from being eliminated by the immune system (Chung et al., 2018). The bacteria also play a role in ensuring their own persistence. In some instances, the bacteria produce modified outer membrane proteins that reduce their immunogenicity (Weiss et al., 2008).

Bacterial symbionts of insects can be broadly assigned into one of three categories; primary, secondary, or parasitic. These define the symbionts by the age of their association, their locality within the insect, and their genome structure, and are

linked to the role that the symbiont plays within the relationship. There is however no universal definition, with some symbionts displaying characteristics of more than one category. The appropriateness of these labels will be evaluated in Chapter 6, but they are nevertheless a useful way to consider insect microbiomes.

1.2.1 Primary symbiosis

Primary symbionts usually have an obligate relationship with their insect host, in which each species is mutually dependent on the other for their survival or ability to reproduce. Several primary symbionts are discussed here, with the key information detailed in Table 1.1. The relationship between host and primary symbiont is often ancient. *Buchnera aphidicola*, for example, is estimated to have initially infected the pea aphid *A. pisum* approximately 200-250 million years ago (MYA) (Moran et al., 1993; van Ham et al., 2003) (Table 1.1). The symbiosis between the leafhopper *Macrosteles quadrilineatus* and one of its obligate primary symbionts *Sulcia muelleri* is thought to be even older, with this association estimated to have initiated 260-280 MYA (Bennett and Moran, 2013).

Table 1.1: The primary symbionts discussed in Chapter 1.

Primary	Insect	Genome (kb)	A+T (%)	Association (MYA)
Buchnera aphidicola	*Acyrthosiphon pisum*	641	73.7	200-250
Ca. Portiera aleyrodidarum	*Bemisia tabaci*	351	76	Unknown
Carsonella ruddii	*Pachypsylla venusta*	160	83.4	Unknown
Sulcia muelleri	*Homalodisca vitripennis*	246	77.6	Unknown
Baumannia cicadellinicola	*H. vitripennis*	686	66.7	Unknown
Nasuia deltocephalinicola	*Macrosteles quadrilineatus*	112	83.4	Unknown
S. muelleri	*M. quadrilineatus*	190	76	260-280
S. muelleri	*Macrosteles quadripunctulatus*	191	75.6	Unknown
Ca. Hodgkinia cicadicola	*Diceroprocta semicincta*	144	41.6	Unknown
Ca. Tremblaya princeps	*Planococcus citri*	139	41.2	Unknown
Blochmannia floridanus	*Camponotus floridanus*	706	72.6	70
Wigglesworthia glossinidia	*Glossina* sp.	698	78	80

Primary symbionts usually localise intracellularly within specialised cells known as bacteriocytes, and are vertically transmitted to the insect progeny. Aphids, of the order Hemiptera, contain 60-80 large bacteriocytes that contain the primary symbiont *B. aphidicola* (Munson et al., 1991; Shigenobu et al., 2000) (Fig. 1.2). *Candidatus* Portiera aleyrodidarum, primary symbiont of the whitefly *Bemisia tabaci*, also resides in bacteriocytes, migrating out of them to the ovaries for transmission to the progeny (Ly Thao and Baumann, 2004b). Collections of bacteriocytes form an organ known as the bacteriome. The sequestration of bacteria into a compartmentalised bacteriome is mutually beneficial to the individuals within the symbiotic relationship (Normark, 2004). The bacteriome maintains a constant environment

for the symbiont whilst allowing it to avoid other microorganisms. It also sequesters the symbionts away from the systemic immune system of the insect. This is an important factor in enabling the relationship between insect and primary symbiont to persist.

Primary symbionts often display unusual genomes. This is due to population bottlenecking and selection pressures over prolonged periods of evolutionary time influencing the mutation rate of the individual (Moran, 1996). The evolution of symbiont genomes could be considered in terms of Muller's ratchet (Muller, 1964; Haigh, 1978; Pettersson and Berg, 2007). This is the process by which the genomes of an asexual population, that does not experience external admixture, accumulates deleterious mutations irreversibly. There are several parallels between the mutations and the reduction in size observed in the genomes of symbiotic bacteria, and the accumulation of mutations in a population as described by Muller (1964). For example, primary symbiont populations are small and, due to vertical transmission, are without external admixture. Deleterious mutations are therefore able to accumulate irreversibly over evolutionary time, whilst no new exogenous DNA can be acquired (Moran, 1996). Some of these mutations may result in the removal of DNA repair machinery, further accelerating the genome reduction (van Ham et al., 2003). The evolution of symbionts has been described as biphasic, with a period akin to Muller's ratchet followed by the accumulation of selfish mutations that ultimately lead to the disintegration of the symbiotic relationship (Rispe and Moran, 2000).

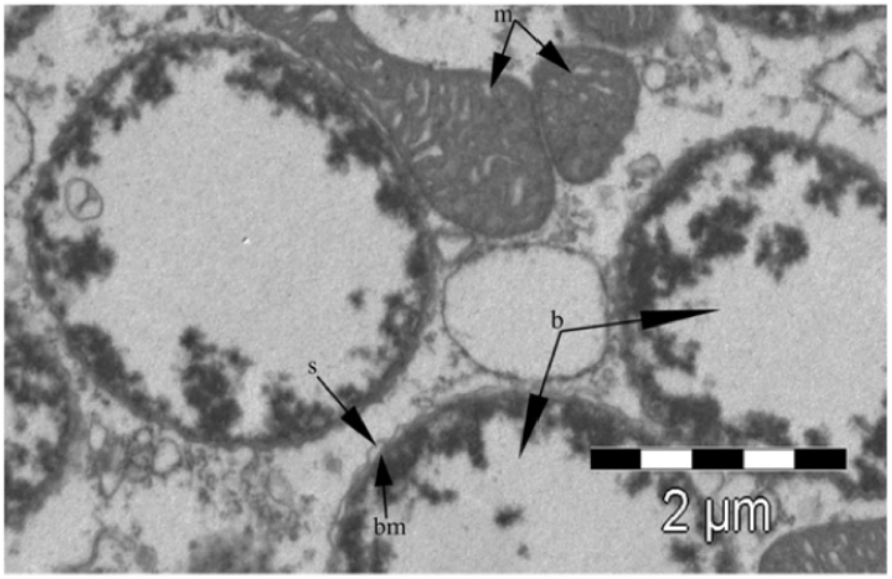


Figure 1.2: Scanning electron microscopy image of *B. aphidicola* within *A. pisum* bacteriocytes, taken from Brinza et al. (2009). s = host-derived membrane, bm = *B. aphidicola* membrane b = *B. aphidicola*, m = mitochondrion.

Symbiont genomes can also be subject to mutational bias. Primary symbiont genomes often display a bias towards the nucleotides adenine and thymine (A+T) (Moran, 1996, 2002). This shift to an A+T-rich genome is as a result of two common changes in the DNA composition, cytosine deamination and guanine oxidation, leading to conversions of cytosine and guanine to thymine and adenine (Michaels and Miller, 1992; Frederico et al., 1990; McCutcheon et al., 2009b). Guanine is the most easily oxidised of the four nucleotides (Kino et al., 2017), and adenine is the nucleotide most commonly incorporated opposite an abasic site that has arisen due to mutations (Shibutani et al., 1997). The propensity of genomes to incorporate mutations that shift the composition to A+T may lead to further gene inactivation and loss, and a subsequent reduction in genome size and bias in codon usage. There are several examples of primary symbionts with highly reduced, A+T-biased genomes. The *Ca.* Portiera aleyrodidarum genome is 76% A+T (Jiang et al., 2012; Ly Thao and Baumann, 2004b), for example, and the *Carsonella ruddii* genome is even more biased at 83.4% A+T (Nakabachi et al., 2006) (Table 1.1).

As a result of these A+T biases, homopolymers, or repeating units, are more likely to form (Medina and Sachs, 2010). This increases the likelihood of small insertions or deletions occurring, eventually leading to the inactivation and pseudogenisation of the gene (Tamas et al., 2008; Medina and Sachs, 2010). Genome reduction is seen not only in primary symbionts but also in pathogenic bacteria, emphasising that it is not one single method behind this process (Weinert and Welch, 2017). There are however some examples of symbiotic bacteria with highly reduced genome, but no A+T bias. *Candidatus* Hodgkinia cicadicola from the cicada *D. semicincta* has an A+T content of 41.6% (McCutcheon et al., 2009b) (Table 1.1). Similarly, *Candidatus* Tremblaya princeps, primary symbiont of the citrus mealybug *Planococcus citri* is 41.2% A+T (López-Madrigal et al., 2011). The lack of A+T bias in these symbiont genomes suggests that a loss of DNA repair systems is not the only driving force behind nucleotide composition biases, as repair enzymes have not been found in *Ca.* Hodgkinia cicadicola (McCutcheon et al., 2009b).

These unusual genomes can be used to make predictions about the role that a symbiont plays in its ecosystem. Symbiont genomes are highly reduced, with genes lost that are no longer required for a lifestyle entirely within an insect. Some may, for example, lose sigma factors that free-living species of bacteria rely on to respond rapidly to environmental changes (Moran, 2002). Such genes become superfluous within the more constant insect microenvironment. *B. aphidicola* has a reduced chromosome of 640,681 base pairs (bp) and two small plasmids (Shigenobu et al., 2000). *B. aphidicola* is a close relative of *Escherichia coli*, with its genome considered a subset of the *E. coli* genome (Shigenobu et al., 2000). Unlike *E. coli*, the

B. aphidicola cells contain approximately 120 copies of the genome per individual (Komaki and Ishikawa, 1999). The *B. aphidicola* genome contains 54 genes involved in the biosynthesis of essential amino acids (Shigenobu et al., 2000). Aphids are phloem sap-feeders, a diet that is rich in carbohydrates but deficient in essential amino acids. Strikingly, the amino acid biosynthesis genes retained by *B. aphidicola* are for those that are essential for the aphid host. The pathways to synthesise non-essential amino acids are almost entirely missing. This supports previous empirical investigations which suggested that *B. aphidicola* plays a nutritional role in this insect-bacterium symbiosis (Douglas, 1998). The *C. ruddii* genome also suggests a role for this symbiont within its phytophagic host, whereby genes involved in the biosynthesis of amino acids have been retained (Nakabachi et al., 2006). Primary symbionts may also enable their hosts to expand into a wider range of ecological niches. The ant *Camponotus floridanus* has an omnivorous diet, but is thought to be supplied essential amino acids by the primary symbiont *Blochmannia floridanus* (Gil et al., 2003; Feldhaar et al., 2007). This is believed to allow the ant to persist in less nutritionally-rich environments.

Sometimes the metabolic provisioning by symbiotic bacteria for their insect hosts may be undertaken in cooperation. There are two obligate symbionts found in the phloem-feeding leafhopper *M. quadrilineatus*; *S. muelleri* and *Nasuia deltocephalinicola*. These symbionts have highly reduced genomes of 190 kb and 112 kb, respectively (Bennett and Moran, 2013). *N. deltocephalinicola* is able to synthesise the essential amino acids L-histidine and L-methionine to complement the L-leucine, L-isoleucine, L-threonine, L-lysine, L-arginine, L-tryptophan, L-phenylalanine and L-valine produced by *S. muelleri* (Bennett and Moran, 2013). Complementation is often seen between obligate, primary symbionts, and co-infecting secondary symbionts. This has been observed in the sap-feeding glassy-winged sharpshooter *H. vitripennis*. In this example, *B. cicadellinicola* provides vitamins and cofactors whilst lacking the ability to synthesise the majority of the essential amino acids (Wu et al., 2006). This role is filled by *S. muelleri*, demonstrating both metabolic complementarity and the importance of secondary symbionts (Wu et al., 2006).

1.2.2 Secondary symbiosis

In comparison to primary symbionts, secondary symbionts usually have an evolutionarily younger association with their hosts. Their genomes are generally larger than that of primary symbionts, and without biases in nucleotide composition. The genome of *Hamiltonella defensa*, a facultative secondary symbiont of the aphid *A.*

pisum, is 2.1 Mb and 59.9% A+T (Degnan et al., 2009). This is smaller than free-living species of bacteria such as *E. coli* K-12 (4.64 Mb) (Blattner et al., 1997), *Yersinia pestis* (4.83 Mb) (Parkhill et al., 2001), and *Serratia marcescens* (5.24 Mb) (Chung et al., 2013), but over three times larger than that of the aphid's primary symbiont *B. aphidicola*.

Accompanying these larger genomes is the opportunity for external admixture. In contrast to primary symbionts, there have been some examples of the horizontal transmission of secondary symbionts (Sandstrom et al., 2001; Oliver et al., 2010). This has been demonstrated in the secondary aphid symbionts, with transmission more likely between closely-related insect species (Russell et al., 2003). The exact mechanism of how this horizontal transfer occurs remains to be elucidated, but parasitoid wasps or the transfer from plants have been suggested (Sandstrom et al., 2001). For example, *Rickettsia* species and *Candidatus* Cardinium hertigii have been horizontally transferred between whitefly (*B. tabaci*) and leafhopper (*Scaphoideus titanus*) hosts, respectively, via the phloem (Caspi-Fluger et al., 2012; Gonella et al., 2015). This therefore provides secondary symbionts with the opportunity to acquire new genes horizontally, reducing population bottlenecking. A key example of this is observed in *H. defensa*. The defensive properties displayed by *H. defensa* are thought to have been acquired from the presence of various bacteriophage (Degnan and Moran, 2008; Degnan et al., 2009; Oliver et al., 2010; Oliver and Higashi, 2019). It is not yet clear whether horizontal transmission of symbionts occurs in blood-feeding insects, or those that do not experience parasitism.

Secondary symbionts play a variety of roles within their host. Sometimes, they may not provide a known benefit. In other instances, the secondary symbiont may only be conditionally beneficial (Oliver et al., 2010). There are several examples of this in the aphid *A. pisum*. *H. defensa* provides protection for *A. pisum* against the parasitoid wasp *Aphidius ervi* (Oliver et al., 2003, 2008). The presence of *Serratia symbiotica* (shown in Fig. 1.3) improves the fecundity of *A. pisum* following heat stress (Montllor et al., 2002). A third secondary symbiont found in *A. pisum*, *Regiella insecticola*, enhances the resistance of its host to infection by the fungal pathogen *Pandora neoaphidis* (Scarborough et al., 2005). *R. insecticola* can also allow *A. pisum* to expand the range of plants on which it can reproduce successfully (Tsuchida et al., 2004). These symbionts are therefore not consistently advantageous to the insect. They only provide a benefit under certain conditions.

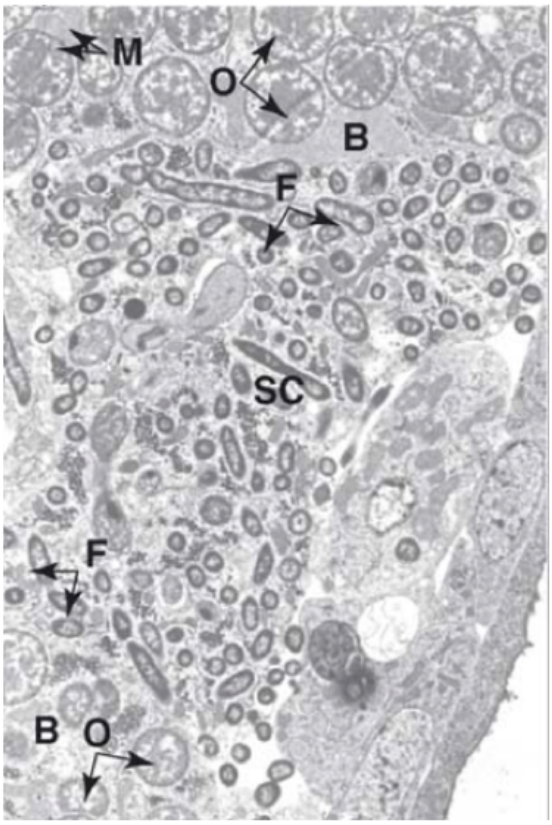


Figure 1.3: Scanning electron microscopy image of the *A. pisum* bacteriome, taken from Pontes and Dale (2006). B = bacteriocyte; F = *S. symbiotica*, M = mitochondrion, O = *B. aphidicola*, SC = sheath cell.

Secondary symbionts can also work cooperatively to provide a benefit for their host. For example, co-infection by *S. symbiotica* and *H. defensa* has been demonstrated to enhance the resistance of aphids to *A. ervi* infection in a laboratory environment (Oliver et al., 2006). Nested symbioses have been observed in the citrus mealybug *Planococcus citri*, in which a secondary γ-proteobacterium localises within the primary symbiont *Ca.* Tremblaya princeps (von Dohlen et al., 2001). This association is believed to have established on multiple occasions over evolutionary time (Thao et al., 2002). Here, the secondary symbiont is not providing a known benefit to *P. citri*, and may instead be using the primary symbiont for protection and to ensure its own transmission (von Dohlen et al., 2001). There may be other, as yet undiscovered, examples of primary and secondary symbionts in close association.

Secondary symbionts are not consistently found at 100% infection frequencies. *H. defensa* is found at intermediate frequencies in natural aphid populations, possibly because the cost of maintaining this symbiont may not outweigh the benefits if the insect is not consistently exposed to parasitoid infection (Oliver et al., 2008). It is possible that there may be a bias in data available for primary symbionts because

of the fact that secondaries of certain species are not found in all individuals. There may also be less of a desire to study secondary symbionts as they, by definition, do not play an essential role in the system. They are, however, potential candidates for disease vector and pest control measures, in part because of their *in vitro* culturability.

Secondary symbionts are not obligate, and therefore some have been successfully cultivated *in vitro*. *Arsenophonus triatominarum* of the triatomine bug, *Triatoma infestans* and *Candidatus* Arsenophonus arthropodicus, a secondary symbiont isolated from the tissues of the hippoboscid louse fly *Pseudolynchia canariensis*, have been isolated and maintained in an *Aedes albopticus* cell line (Hypsa and Dale, 1997; Dale et al., 2006). Other secondary symbionts have been grown without the use of cell lines, including the tsetse symbiont *Sodalis glossinidius* (Dale and Maudlin, 1999). This characteristic, and others observed in the *Sodalis* genus, will be discussed in depth in Section 1.4.

1.2.3 Parasitism

Certain insect-microbe symbioses could be described as parasitic, rather than mutualistic or commensal. One such species is *Wolbachia pipientis* (henceforth "*Wolbachia*"). The *Rickettsia*-like *Wolbachia* is one of the most promiscuous and well-studied insect-associated bacteria. It is a Gram negative bacterium that is estimated to be found in at least 20% of all insects (Hedges et al., 2008), including insects of the orders Lepidoptera, Diptera, Coleoptera, Hemiptera, Hymenoptera and Orthoptera (Werren et al., 1995). *Wolbachia* is maternally transmitted and consequently exerts a variety of effects on the reproduction of its host, including cytoplasmic incompatibility (CI) and male killing. As a result of CI, *Wolbachia*-infected males cannot produce viable offspring if mating with uninfected females (Fig. 1.4). These effects, which are often negative for the host species, mean that *Wolbachia* is not a traditional symbiont by strict definition. Its behaviour is parasitic, rather than neutral or beneficial. This manipulative phenotype is not limited to *Wolbachia*. *Arsenophonus nasoniae*, for example, has been observed to cause a son-killing phenotype in the parasitic wasp *Nasonia vitripennis* (Gherna et al., 1991). Importantly, the presence of *Wolbachia* has been linked to vector competence; the ability of a pathogen to persist within and spread through a population of insects. For example, laboratory strains of *Drosophila melanogaster* infected by *Drosophila* C virus have decreased mortality in the presence of *Wolbachia* (Hedges et al., 2008). *Wolbachia* has also been harnessed to great success in the control of dengue virus. Here, the *w*Mel

strain of *Wolbachia* from *D. melanogaster* has been introduced into the naturally *Wolbachia*-free dengue vector *Aedes aegypti* (Hoffmann et al., 2011; Walker et al., 2011). This avirulent strain of *Wolbachia* induces CI in the mosquitoes (Walker et al., 2011) and increases the resistance of the insect to the virus (Bian et al., 2010). This serves to highlight both the scientific interest and the potential value in researching the bacterial symbionts of insects.

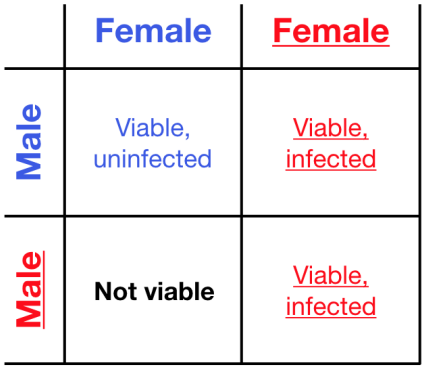


Figure 1.4: Cytoplasmic incompatibility induced by *Wolbachia*. Infected (red) males are not able to produce viable offspring when mating with uninfected (blue) females.

Wolbachia has been found in some species of tsetse (Diptera, genus *Glossina*), but not at 100% infection frequency (O'Neill et al., 1993; Cheng et al., 2000). It is not thought to have a metabolic role in the tsetse microbiome, instead exerting its influences on the insect's reproduction via CI (Fig. 1.4) (Alam et al., 2011). It is included here for completeness, but because of the lack of metabolic interest, coupled with the low infection frequencies, is not the focus of this thesis.

1.3 The tsetse and its microbiome

1.3.1 The tsetse and human African trypanosomiasis

The tsetse is a unique insect of great medical importance. The 366 Mb *Glossina morsitans morsitans* genome was published in 2014, containing 12,308 predicted protein-encoding genes (Attardo et al., 2014). It reproduces by adenotrophic viviparity, producing live larvae fed through milk glands. This behaviour is unusual within the Diptera order (Wall and Langley, 1993), and indeed for non-mammals. The tsetse has an obligate hematophagic diet of mammalian blood. This diet is highly unbalanced, with an abundance of amino acids but fewer saccharides. It is via these

blood meals that an infected tsetse can spread the protozoan parasite *Trypanosoma brucei*.

Figure 1.5: The tsetse. Image by Geoffrey Attardo from Heller (2011).

T. brucei is the causative agent of human African trypanosomiasis (HAT) and the wasting disease of cattle known as nagana (WHO, 1998). After feeding on an infected blood meal, procyclic *T. brucei* localise within the peritrophic matrix in the midgut of the insect, before differentiating into epimastigote and then metacyclic forms and migrating to the salivary glands (Fig. 1.6). From there, *T. brucei* can infect a new mammalian host, where they transform from slender to stumpy forms in the bloodstream. Metacyclic and stumpy forms are pre-adapted to their next hosts (McKean, 2003). This life cycle is comparable to that of *Leishmania* species that causes the neglected tropical disease (NTD) leishmaniasis (Fig. 1.6). *Leishmania* species also have stages within the midgut of their vector, the sandfly, before migrating to the salivary glands. The main difference between *T. brucei* and *Leishmania* life cycles is found in the mammalian host. *T. brucei* remains within the bloodstream, whereas *Leishmania* invades, and proliferates within, human macrophages (Handman, 2001). The *T. brucei* life cycle is arguably not as complex as that of the malarial parasite *Plasmodium falciparum*, which undergoes multiple stages of differentiation within the gut of its mosquito host (Fig. 1.6), as well as an erythrocytic phase within human erythrocytes that allows the parasite to proliferate (Lee et al., 2014). The life cycle of the trematode genus *Schistosoma*, causative agent of schistosomiasis, is arguably more intriciate still (Fig. 1.6). There are multiple stages of migration and differentiation within humans as the parasite translocates from the skin to the mesenteric vessles, via the lungs, heart and liver (Gray et al., 2011). The *Schistosoma* eggs are then released into the water in faeces, where they hatch and infect their intermediate snail host. The cercariae are then released back into the water.

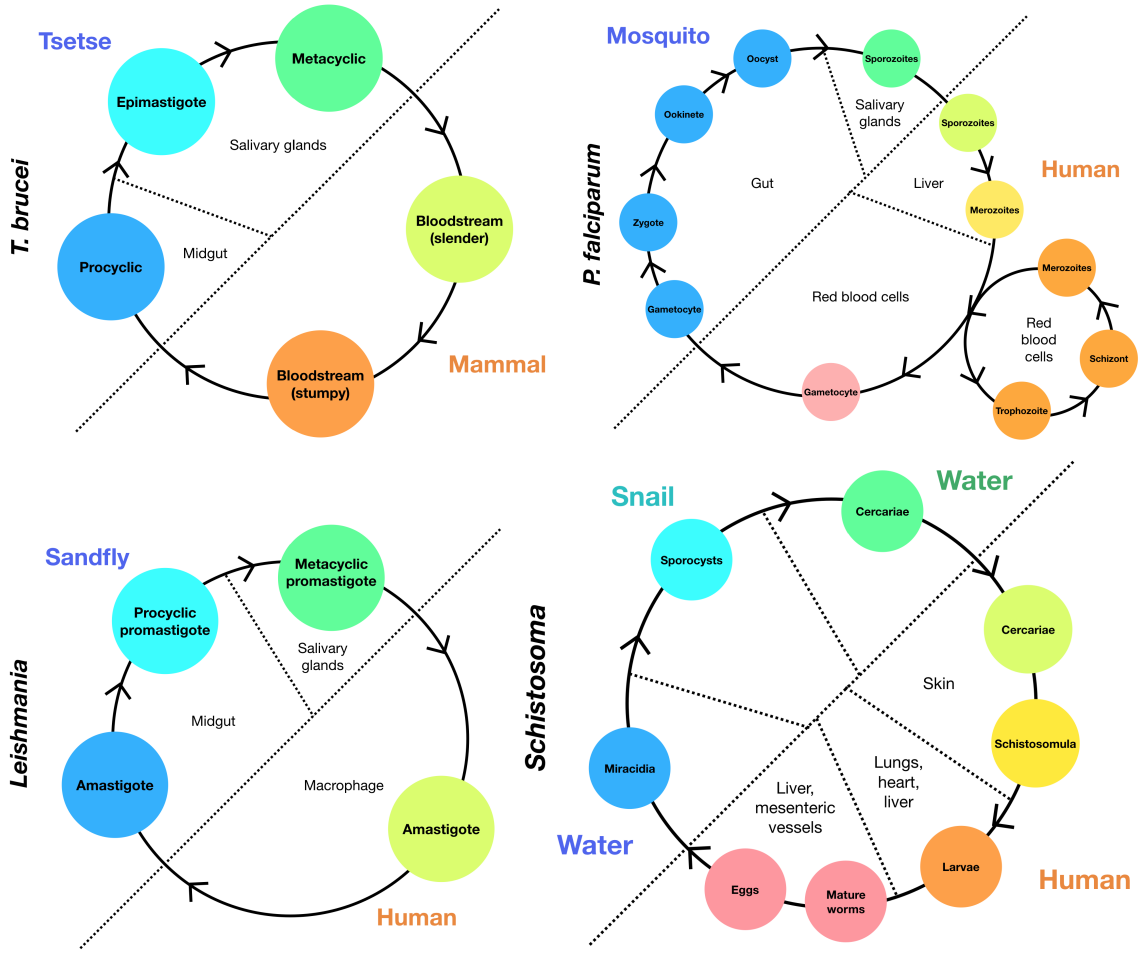


Figure 1.6: Parasitic life cycles. The life cycles of *T. brucei* in its tsetse and mammalian hosts, *P. falciparum* in its sandfly and human hosts, *Leishmania* in its sandfly and human hosts, and the *Schistosoma* trematode life cycle through humans, water, and snails. *T. brucei* life cycle adapted from Fenn and Matthews (2007) and McKean (2003). *P. falciparum* life cycle adapted from Lee et al. (2014). *Leishmania* life cycle adapted from Handman (2001). *Schistosoma* life cycle adapted from Gray et al. (2011).

HAT is currently endemic in 36 countries in sub-Saharan Africa, with an estimated 70 million people at risk of infection (Kennedy, 2004; Migchelsen et al., 2011; Roditi and Lehane, 2008; Simarro et al., 2011; Attardo et al., 2014). It is an NTD characterised by two stages of infection. The symptoms of the haemo-lymphatic stage are generally non-specific, including fever, headache and joint pain. The second, meningo-encephalitic stage occurs when *T. brucei* crosses the blood-brain barrier and enters the central nervous system. These symptoms are more severe, including severe neurological conditions and sleep cycle disruption. This stage is often fatal if left untreated.

HAT was responsible for millions of deaths in the twentieth century, but there has since been a steady decline in the number of cases. This is due to a combination of

successful control methods, and changes in the environment and climate (Büscher et al., 2017). This downward trajectory is in contrast to what is currently observed for malaria, leishmaniasis, and schistosomiasis. Estimates indicated there were approximately 212 million new cases of malaria in 2015 (Patouillard et al., 2017), with schistosomiasis affecting 250 million people (Utzinger et al., 2015). There are approximately 1.5-2 million new cases of visceral or cutaneous leishmaniasis annually (Desjeux, 2004). This may imply that HAT is possibly not a priority NTD. There are however still pockets of rural populations at significant risk of *T. brucei* infection. Poverty, political unrest, and other instabilities could cause a resurgence of the disease. There are also concerns that *T. brucei* may develop resistance to the current treatment options (Kennedy, 2004). The development of vaccines to protect against HAT is hindered by the antigenic variation employed by the parasites to evade the mammalian immune system (Aksoy, 2000; Xong et al., 1998). *T. brucei* is able to switch its variant surface glycoprotein coat to avoid recognition by antibodies produce by the immune system in response to specific proteins (Horn, 2014). This makes the development of vaccines against *T. brucei* extremely challenging. There is a need therefore for new, affordable and effective ways to control the disease and decrease the risk of a resurgence. One way that this could be achieved, as described here for the control of dengue, is by manipulating the tsetse's microbiome (Aksoy et al., 2001).

The tsetse is host to a unique microbiome that co-exists alongside *T. brucei*. The microbiome consists of the primary *Wigglesworthia glossinidia* and the secondary *S. glossinidius*. The tsetse is also vulnerable to the parasitic *Wolbachia*, which is not essential for the tsetse's survival or fecundity and not always present in natural tsetse populations (Cheng et al., 2000; Weiss and Aksoy, 2011). *W. glossinidia* and *S. glossinidius*, specifically their role within the tsetse microenvironment, will be discussed in depth here.

1.3.2 The primary symbiont *Wigglesworthia glossinidia*

W. glossinidia is an intracellular γ-proteobacterium. It localises within bacteriocytes in the anterior midgut of the tsetse, with an additional extracellular population found inside the milk glands (Pais et al., 2008; Snyder et al., 2010; Weiss and Aksoy, 2011; Aksoy, 1995) (Fig. 1.7). It is via these milk glands that *W. glossinidia* is transmitted vertically to the next generation. *W. glossinidia* is estimated to have co-evolved with the tsetse for approximately 80 million years (Chen et al., 1999; Snyder and Rio, 2013). It has a single, highly reduced chromosome of 697,724 bp

and a 5,200 bp plasmid (Akman et al., 2002), comparable in size to the obligate *B. aphidicola* (Shigenobu et al., 2000). The *W. glossinidia* genome composition displays a strong A+T bias, of 78% and a complete absence of insertion sequences or transposable elements (Akman et al., 2002). Noticeable in the *W. glossinidia* genome is the maintenance of pathways for vitamin biosynthesis (Akman et al., 2002). It is hypothesised that *W. glossinidia* has retained these in order to supply B group vitamins to its tsetse host, supplementing those that the insect is unable to synthesise or import from its blood meal. It has since been demonstrated empirically that pyridoxal 5'-phosphate and folate, synthesised *in vivo* by *W. glossinidia*, can affect the fecundity and fitness of the tsetse (Michalkova et al., 2014; Snyder and Rio, 2015). The *W. glossinidia* genome also encodes all components required to synthesise a complete flagellum (Akman et al., 2002). It is not known whether this is for motility or to assist in vertical transmission. It is not yet possible to cultivate *W. glossinidia in vitro*.

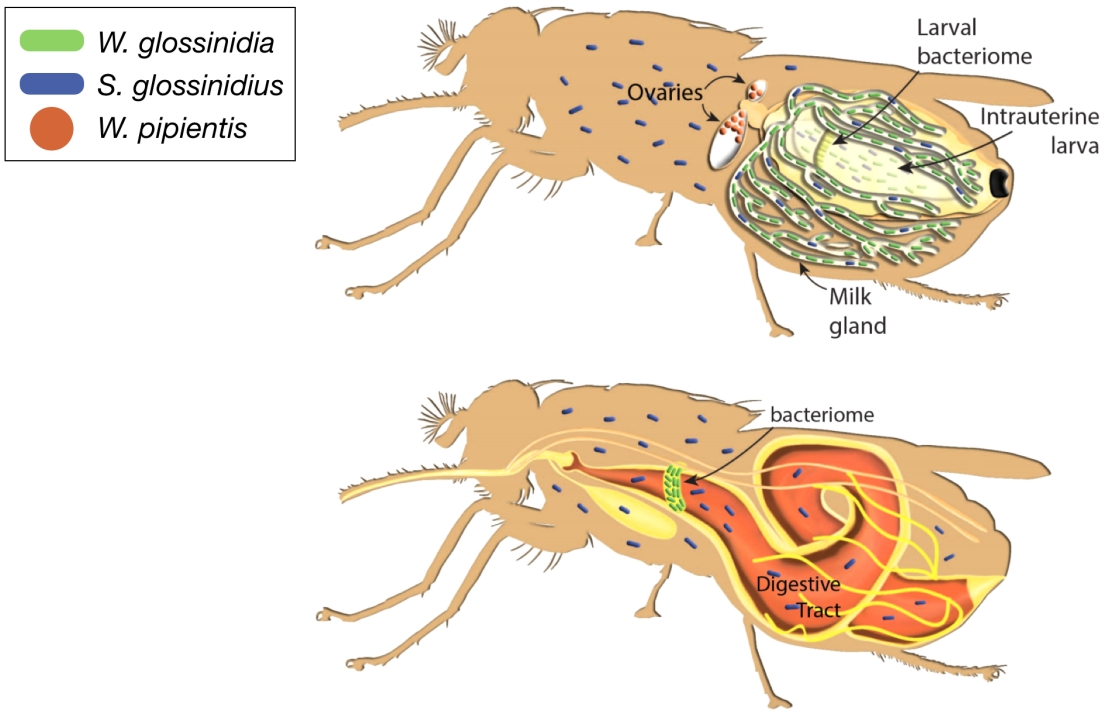


Figure 1.7: The localisation of *W. glossinidia* (green), *S. glossinidius* (blue) and *Wolbachia* (red) in the tsetse. Adapted from Attardo et al. (2014).

1.3.3 The secondary symbiont *Sodalis glossinidius*

The secondary endosymbiont *S. glossinidius* resides in the tsetse midgut, fat body and hemolymph (Cheng and Aksoy, 1999; Dale and Maudlin, 1999) (Fig. 1.7). The

genome of *S. glossinidius* is 4.17 Mb, with 2,432 protein encoding sequences (Toh et al., 2006). There have also been four circular elements of extrachromosomal DNA identified and sequenced; three plasmids and one bacteriophage-like element (Darby et al., 2005b). The extrachromosomal DNA shows evidence of symbiotic fitness factors, including siderophores and heat shock proteins, and of genes involved in conjugation (Darby et al., 2005b). These may have been important in the evolution of *S. glossinidius*. The 51% coding capacity of the *S. glossinidius* chromosome and the 972 pseudogenes found within it (Toh et al., 2006) reflect its symbiotic status. Many of these pseudogenes have been split into three or more fragments, or have lost a high percentage of their sequence. When categorised by function, a large proportion of the *S. glossinidius* pseudogenes are involved in defence, and the metabolism and transport of carbohydrates and inorganic ions (Toh et al., 2006). From this, it could be inferred that *S. glossinidius* has started to lose genes that are no longer required within the tsetse, given the insect's obligate hematophagic diet. The extensive complement of pseudogenes is indicative of a large process of genome reduction, and it is expected that these will be removed over time by ongoing evolutionary processes (Belda et al., 2010; Toh et al., 2006; Snyder and Rio, 2013).

As with *W. glossinidia*, the *S. glossinidius* genome retains the ability to encode a complete flagellum (Akman et al., 2002; Toh et al., 2006). Both symbionts are transmitted vertically to the tsetse larval progeny, suggesting a possible reason for the retention of the flagella apparatus in both. In contrast to *W. glossinidia*, however, *S. glossinidius* can be cultured *in vitro*. It is a microaerophile, demonstrating optimal growth in 5-10% oxygen (Dale and Maudlin, 1999) and at 25°C (Dale and Maudlin, 1999; Toh et al., 2006; Roma et al., 2019). Due to its genome size and culturability, the association of *S. glossinidius* with its host is thought to be more recent than that of *W. glossinidia* (Chen et al., 1999). *S. glossinidius* could be considered a snapshot in evolutionary time, an intermediate between a free-living and an obligate symbiont. It therefore provides an excellent platform for studying host-symbiont interactions and the evolution of symbiotic relationships.

1.4 The *Sodalis* genus

The *Sodalis* genus is particularly interesting and varied. *Sodalis*-like symbionts, as mentioned briefly, are found in a variety of insect species and orders (Table 1.2). Several, as exemplified by *S. glossinidius*, are found in blood-feeding insects (Dale and Maudlin, 1999; Toh et al., 2006; Novakova and Hypsa, 2007; Chrudimský et al., 2012; Boyd et al., 2016). These diets are low in sugars but rich in amino

acids. Insects that feed on plant sap, with diets high in sugars but low in amino acids, also contain symbionts of the *Sodalis* genus, in spite of their contrasting metabolite availability (Szklarzewicz et al., 2018; Koga et al., 2013; Koga and Moran, 2014; Michalik et al., 2014; Kaiwa et al., 2010; Hosokawa et al., 2015). *Sodalis*-like symbionts have also been identified in insects that feed on rice (Heddi et al., 1997; Oakeson et al., 2014), seeds (Toju et al., 2010; Toju and Fukatsu, 2011; Santos-Garcia et al., 2017), bark (Grünwald et al., 2010), and avian feathers (Fukatsu et al., 2007). This demonstrates the variety of different nutritional niches that members of the *Sodalis* genus can occupy.

Whilst some *Sodalis*-like symbionts have been identified from the sequencing of core genes (*groEL*, 16S rRNA) alone, others have complete genome sequences available (Table 1.2). These provide a valuable source of information about how they have adapted to their microenvironment. The 4.5 Mb genome of *Candidatus* Sodalis pierantonius is comparable to that of *S. glossinidius*, with a 43.9% A+T content and 1771 predicted pseudogenes (Oakeson et al., 2014). Unlike *S. glossinidius*, *Ca.* Sodalis pierantonius is described as a bacteriome-associated primary symbiont, with its removal impairing the flight and fecundity of its host (Oakeson et al., 2014; Heddi et al., 1997). It has been proposed that *Ca.* Sodalis pierantonius may play a nutritional role for *S. oryzae* via the provision of certain vitamins, namely pantothenate, biotin and riboflavin (Wicker, 1983).

Other *Sodalis*-allied symbionts have significantly smaller genomes than *S. glossinidius* (Table 1.2). The bacteriocyte-associated *Sodalis*-like symbiont of the hematophagic seal louse *Proechinophthirus fluctus* has a 2.18 Mb genome, and unlike *S. glossinidius* is hypothesised to be a replacement for a now extinct endosymbiont (Boyd et al., 2016). The genomes of *Candidatus* Sodalis baculum and SL-PSPU are even smaller, at 1.62 Mb and 1.38 Mb, respectively (Santos-Garcia et al., 2017; Koga et al., 2013; Koga and Moran, 2014). Both of these organisms localise to the bacteriome and are thought to provide their host insects with amino acids (SL-PSPU) and cofactors (*Ca.* Sodalis baculum).

A free-living species of the *Sodalis* genus, *Sodalis praecaptivus*, has recently been isolated from a human wound caused by an impalement by a crab apple tree (Clayton et al., 2012). This indicates that *S. praecaptivus* can survive on both plant and animal tissue. *S. praecaptivus* is a prototroph capable of growth in minimal media (Chari et al., 2015). It has a genome size of 5.16 Mb with only 61 pseudogenes (Oakeson et al., 2014). It is phylogenetically similar to several *Sodalis*-like symbionts of insects (Chari et al., 2015), including *Candidatus* Sodalis melophagi of the hematophagic sheep ked *Melophagus ovinus* (Chrudimský et al., 2012), *Ca.*

Table 1.2: Members of the *Sodalis* genus. — Indicates no sequenced genome.

Symbiont	Host	Host diet	Genome	Additional symbionts	Localisation	Description	Reference
S. praecaptivus	Free-living	N/A	5.16 Mb	N/A	Human wound	Free-living	(Chari et al., 2015; Clayton et al., 2012)
S. glossinidius	*Glossina* sp.	Blood	4.2 Mb	*W. glossinidia*, some incidences of *Wolbachia*	Widespread	Secondary mutualist	(Dale and Maudlin, 1999; Toh et al., 2006)
Sodalis-like	*Craterina melbae*	Blood	—	Not specified	Not specified	Not specified	(Novakova and Hypsa, 2007)
Ca. Sodalis melophagi	*Melophagus ovinus*	Blood	—	*Arsenophonus*, *Bartonella melophagi*	Bacteriome, gut	Early/intermediate state	(Chrudimský et al., 2012)
Sodalis-like	*Proechinophthirus fluctus*	Blood	2.18 Mb	Relative of *Rickettsia rickettsii*	Bacteriocyte, extracellular	Recent acquisition	(Boyd et al., 2016)
Ca. Sodalis pierantonius	*Sitophilus oryzae*	Rice	4.5 Mb	Candidatus *Nardonella* spp.	Bacteriome	Primary, recent acquisition	(Heddi et al., 1997; Oakeson et al., 2014)
Sodalis-like	*Puto superbus*	Plant sap	—	*Wolbachia*	Bacteriocyte	Primary	(Szklarzewicz et al., 2018)
SL-PSPU	*Philaenus spumarius*	Xylem sap	1.38 Mb	*Sulcia muelleri*	Bacteriocyte	Recent acquisition	(Koga et al., 2013; Koga and Moran, 2014)
Sodalis-like	*Cicadella viridis*	Plant sap	—	*Ca.* Sulcia muelleri, relative of *Pectobacterium*	Bacteriocyte, oocytes	Not specified	(Michalik et al., 2014)
Sodalis-like	*Ommatidiotus dissimilis*	Hare's-tail cottongrass	—	*Sulcia*, *Vidania*, *Wolbachia*, *Rickettsia*	Bacteriocyte	Obligate	(Michalik et al., 2018)
Sodalis-like	*Cantao ocellatus*	Plant sap	—	Gamma-proteobacterium	Gonads	Secondary, facultative	(Kaiwa et al., 2010)
Sodalis-like	Urostylididae	Plant sap	—	Not specified	Not specified	Not specified	(Hosokawa et al., 2015)
Sodalis-like	*Antestiopsis thunbergii*	Coffee plants	—	*Spiroplasma* spp., *Rickettsia* sp., gut symbiont	Various tissues	Secondary, facultative	(Matsuura et al., 2014)
Sodalis-like	*Curculio sikkimensis*	Seeds	—	*Candidatus* Curculioniphilus buchneri	Not specified	Secondary, facultative	(Toju et al., 2010; Toju and Fukatsu, 2011)
Sodalis-like	*Tetropium castaneum*	Bark	—	Ascomycetous yeast strains, gut bacteria	Epithelial cells	Not specified	(Grünwald et al., 2010)
Ca. Sodalis baculum	*Henestaris halophilus*	Seeds	1.62 Mb	Not specified	Bacteriome	Intermediate, mutualist	(Santos-Garcia et al., 2017)
Sodalis-like	*Columbicola columbae*	Avian feathers	—	Not specified	Bacteriocyte	Not specified	(Fukatsu et al., 2007)

Sodalis pierantonius (Heddi et al., 1997; Oakeson et al., 2014), and symbionts of the slender pigeon louse *Columbicola columbae* (Fukatsu et al., 2007), chestnut weevil *Curculio sikkimensis* (Toju et al., 2010; Toju and Fukatsu, 2011), and stinkbug *Cantao ocellatus* (Kaiwa et al., 2010). A dendrogram produced by matrix assisted laser desorption ionization - time of flight (MALDI-TOF) performed on *S. glossinidius* and *S. praecaptivus* suggests that the two species are closely related (Chari et al., 2015). Due to their close phylogenies but distinctive environments, *S. praecaptivus* is an excellent reference organism from which to investigate the evolution of insect-*Sodalis* symbioses.

1.5 Tools for studying symbiosis

1.5.1 Traditional tools

Prior to the advent of rapid genome sequencing, the study of insect symbionts was predominantly achieved by observing the effect of their removal. Aposymbiotic insects are usually produced by via treatment with antibiotics. This can have a deleterious effect on the behaviour of the insect. Distrupting the *A. pisum* microbiome, for example, results in insects that feed poorly (Wilkinson and Douglas, 1995a). This can be measured using electrical penetration graphs that record stylet penetration. Generating aposymbiotic insects does however allow the metabolic role of a symbiont to be investigated. Following antiobiotic treatment, *Myzus persicae* aphids grow better when their diet is supplemented with L-methionine (Douglas, 1988b). This suggests the *M. persicae* symbionts are provisioning this metabolite to their host. The same method has been applied to *A. pisum* aphids to investigate amino acid provisioning and use (Prosser and Douglas, 1991; Douglas and Prosser, 1992; Wilkinson and Douglas, 1995b).

Symbionts within bacteriocytes can also be dissected out of their insects, and microscopy performed to analyse their number and size (Douglas and Dixon, 1987). The dissected bacteriocytes can be supplied with labelled carbon sources to measure metabolite uptake by the symbionts, as demonstrated in *A. pisum* (Whitehead and Douglas, 1993). The microbiome of the leafhopper *Euscelis incisus* has also been examined following dissection. In this example, the gut of the insect was spread on to nutrient agar plates and incubated to isolate bacterial cultures (Douglas, 1988a). Dissecting insects is challenging due to the small size of the bacteriocytes, as well as the risk of contamination between the different species of symbionts (Douglas,

1988a).

The ability to sequence bacterial genomes cheaply and rapidly has progressed the study of insect symbionts. Symbiotic bacteria can be difficult to culture via traditional microbiology techniques. Obligate symbionts, by definition, cannot survive outside of their hosts. This may be due to a reliance on metabolites or other compounds produced by the host, as a product of their intracellularity, or a combination of the two. Mutualists may be cultured *in vitro* but there are often issues with slow growth and a reliance on rich media, leading to issues with opportunistic contamination. Because of this, alternative techniques to studying symbiotic bacteria are desirable.

1.5.2 Flux balance analysis

One such technique for studying symbiotic bacteria is flux balance analysis (FBA). FBA is a computational, constraint-based modelling technique that is used to examine the metabolic network of an organism (Orth et al., 2010). Constraint-based modelling assumes that a system will form a steady state. This means that all metabolites that are fed into a network will have a route through and out, and cannot accumulate in the system (Thiele and Palsson, 2010). Some may be used to form the *in silico* biomass of an organism via a pre-defined, species-specific biomass reaction. There are many outcomes that can be found within the solution space, but this space can be constrained by specifying the maximum and minimum fluxes allowed through each reaction (Kauffman et al., 2003). For example, the network could be constrained to only allow uptake and use of a single carbon source. The resulting flux through each reaction can then be measured. Not all reactions will carry flux under a given condition.

An overview of the process of constructing a flux balance model is shown in Fig. 1.8. The process begins by generating or accessing an annotated genome of the organism (Thiele and Palsson, 2010). All of the predicted metabolic genes encoded by the organism can then be identified (Orth et al., 2010; Kauffman et al., 2003). This is usually achieved by mining the genome annotation, either through identifying common gene names, via large-scale BLAST searches, or a combination of the two. This should identify all metabolic enzymes as well as proteins involved in the active or passive transport of metabolites. The reactions encoded by these proteins are then extracted, commonly from databases including BiGG Models (King et al., 2016), KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2019) and MetaCyc (Caspi et al., 2007), and compiled into a draft model.

Biosynthetic pathways are often incomplete at this stage, and the construction then requires an iterative process of manual gap-filling and *in vitro* testing (Thiele and Palsson, 2010). Several techniques could be used during the gap-filling process. The draft model could be compared to pre-existing metabolic models to identify commonly included reactions. Closely and more distantly related related organisms could serve as queries or templates when mining for specific genes. Here, *in vitro* testing is crucial for characterising the capability of the organism to grow on a variety of substrates, ensuring that this is reflected in the final model. One weakness of FBA and other computational tools is that the accuracy of the modelling depends upon the quality of the genome sequence and annotation. Extensive *in vitro* testing should help to mitigate this.

Metabolic modeling can be a powerful tool in the study of symbiont biology. These organisms, with their reduced genome capacity, are often difficult to culture. The models can therefore provide insight in how to optimise symbiont culture, or to make predictions that would not be possible *in vitro*. For example, multiple flux balance models have been linked together to simulate metabolic cross-feeding in the symbionts of cicadas, spittlebugs and sharpshooters (Ankrah et al., 2018). Metabolic models can however be limited by the cultivation challenges of symbiotic bacteria. An inability to conduct a thorough phenotypic screen may result in a model that is incomplete or that does not closely match the metabolism of the symbiont.

Once implemented, FBA models can be manipulated to examine a range of important characteristics of symbiotic bacteria *in silico*, include gene essentiality and metabolite production. Individual reactions, or a set of reactions encoded by a single gene, may be removed and the resulting effect on the network flux examined (Thiele and Palsson, 2010). This can inform as to whether a certain knockout carries a lethal phenotype.

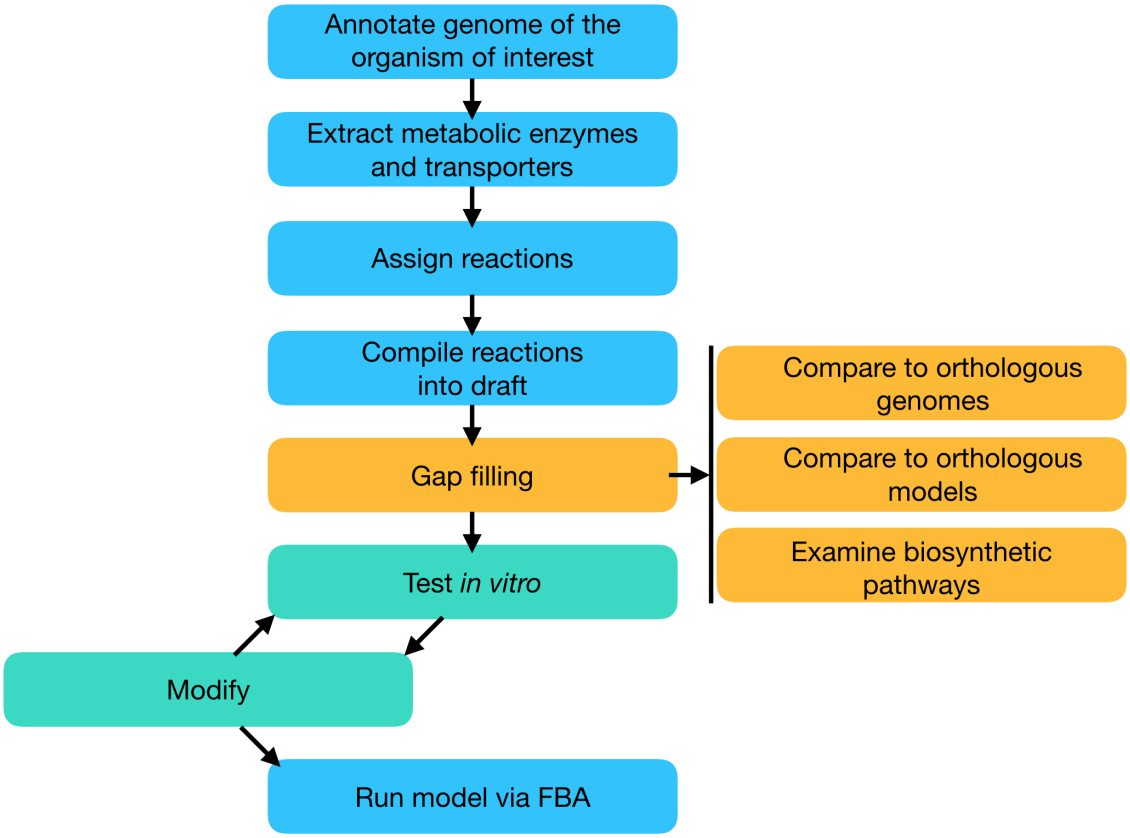


Figure 1.8: An overview of the process of constructing a metabolic model. Blue = key steps, orange = the process of gap filling, green = the iterative processing of testing and modifying.

FBA can also be used to illustrate and investigate the unusual nature of symbiont genomes. The relatively constant internal environment of an insect provides an opportunity for non-essential genes, sometimes entire metabolic pathways, to be lost over time. The patterns of gene pseudogenisation are therefore good predictors of both the insect microenvironment and the relationship between host and symbiont. The redundancy of a network can be established by the repeated removal of genes or reactions *in silico* (Thiele and Palsson, 2010). This information can be used as a comparison to both free-living and obligate models, to estimate the stage of symbiosis and genome reduction that an organism has reached.

An FBA model of *S. glossinidius* metabolism has been published by Belda et al. (2012). This offers access to a large amount of information about *S. glossinidius* metabolism. In the absence of a well-annotated free-living relative from which to compare, Belda et al. (2012) reconstruct an ancestral *S. glossinidius* model by converting predicted pseudogenes to functional reactions. *E. coli* K-12 was used as a comparison. An extant model, *i*EB458, was created by removing the pseudogenes. Their ancestral model was able to produce a viable biomass output with the ex-

ogenous addition of D-glucose alone, whereas the extant model was not. This was due to several key pseudogenisations. First, the extant network is unable to catalyse glycogen biosynthesis from D-glucose. Glycogen was a component of the biomass reaction and a positive biomass output could not be produced without the synthesis of all components. Second, their work included pseudogenisations in *argA*, *argC*, *argD*, and *argG*. These genes encode reactions in the pathway for L-arginine biosynthesis from L-glutamate. In the absence of these genes, the extant network was unable to synthesise L-arginine. They state that this prevented the subsequent biosynthesis of spermidine and putrescine (Belda et al., 2012). *i*EB458 is a useful starting point from which to understand the metabolism of *S. glossinidius* and how it may have adapted to a symbiotic lifestyle. It is however fundamentally limited by a lack of experimental validation.

There has also an FBA model constructed for *W. glossinidia* (Gavin Thomas, unpublished). It makes important, unusual predictions about the symbiont's metabolism. This model implies that *W. glossinidia* uses L-proline as its energy source via gluconeogenesis. L-proline is abundant within the tsetse microenvironent, relied upon by both the insect and *T. brucei* (Bursell, 1963, 1978; Hargrove, 1976; Mantilla et al., 2017). It would therefore seem logical for *W. glossinidia* to adapt to use a widely available metabolite. This model is however limited by the fact that no L-proline transporters have been defined in *W. glossinidia*. Experimental verification that the symbiont is capable of importing this amino acid is required to support the model.

1.5.3 Defined media for physiological studies

Research into bacterial metabolism *in vitro* requires simple, minimal media. This allows metabolites to be tested individually, and any change in growth observed can be attributed back to this compound with confidence. This is not currently possible with symbiotic bacteria. As symbionts lose coding capacity, their ability to survive in minimal media is compromised. They may have lost entire biosynthetic pathways or transporters as they became adapted to their insect's microenvironment. This is not an issue unique to symbiotic bacteria. Indeed, it has been estimated that only 1% of microbial life is culturable using currently available techniques (Staley and Konopka, 1985; Amann et al., 1995; Vartoukian et al., 2010).

There have been some examples of symbionts being cultured in insect cell lines (Hypsa and Dale, 1997; Dale et al., 2006; Darby et al., 2005a). Whilst a successful method for maintaining symbiont cell culture, this technique carries an increased risk of contamination and requires of specialist equipment and knowledge (Lynn,

2002). It can also be difficult to entirely tease apart the effect that altering different conditions has on an individual species; bacteria may respond differently if they are isolated from their co-culture. This is therefore not an optimal culture protocol for studying the metabolism of symbiotic bacteria.

Other symbionts have been cultured in isolation but in rich media, including brain heart infusion (BHI) broth (Azambuja et al., 2004; Hrusa et al., 2015) and Mitsuhashi-Maramorosch (M-M) medium (Weiss et al., 2008; Maltz et al., 2012). This has advantages over using insect cell lines as pure cultures of the symbiont in isolation are used. It is an adequate method for studying certain phenotypes, for example the effect of temperature on cell growth (Roma et al., 2019). It is not however useful for studying the intricacies of metabolism, for several reasons. First, the recipes are undefined, or contain ingredients that are themselves complex. BHI contains digests of beef hearts and calf brains (Sigma). It is therefore not possible to know exactly which metabolites are present here, or in what quantities. Second, the metabolites cannot be removed individually. Removing metabolites and measuring the resulting effect on the growth phenotype is a useful way to study their essentiality. This is not possible in BHI or M-M. It is also more difficult to measure the metabolites being produced by the bacteria if the starting medium is undefined.

The gold standard for *in vitro* tests of metabolism is to use a rationally designed, species-specific, defined medium. Rationally designed growth media was achieved for the causative agent of Whipple's disease *Tropheryma whipplei* (Renesto et al., 2003). This was specific to *T. whipplei*, but still contained undefined components. Defined media have been developed successfully for species of lactic acid bacteria (Jensen and Hammer, 1993; Elli et al., 2000), soil bacteria (Gonzalez-Lopez et al., 1983; Rodelas et al., 1993), and marine microorganisms (Woung Kim and Bok Lee, 2003; Hanna and Lilly, 1974; Woelkerling et al., 1983), enabling precise questions about their metabolism to be answered with confidence. This gold standard has yet to be achieved for the symbiotic bacteria of insects. *S. glossinidius* is currently cultured in BHI (Roma et al., 2019), BHI with added horse blood (Runyen-Janecky et al., 2010; Hrusa et al., 2015; Smith et al., 2013), Schneider's insect medium (Goodhead et al., 2018), and M-M medium (Dale et al., 2001; Haines et al., 2009, 2002; Weiss et al., 2006), amongst other rich media. Studies into the metabolism of *S. glossinidius* will remain limited until a defined medium can be designed to support the *in vitro* growth of this symbiont.

1.5.4 *In silico* evolution to study the initiation and evolutionary trajectory of symbiosis

There are several key stages in the formation of a symbiotic relationship. Much of the focus of scientific research is on the adaptations that arise once the symbiont has become internalised. These adaptations provide a wealth of information about the basis upon which the relationship has established. Less is known about the initial interaction between insect and bacterium. This crucial stage is interesting for several reasons. First, the bacterium must provide at least a neutral presence from the outset. It needs to avoid being eliminated by the immune system of the insect. The aphid *A. pisum* is thought to have acquired two genes from bacteria; *amiD* and *ldcA1* (Chung et al., 2018). These genes encode proteins that are thought to degrade immunogenic peptidoglycan fragments produced by *B. aphidicola*, thus protecting the symbiont from elimination (Chung et al., 2018). This could be a vital adaptation of the aphid to support its symbionts, but the temporal acquisition of this is not known.

The initial stages of symbiosis may also be the most important in shaping the evolution and reduction of the symbiont genome. For example, *B. aphidicola* is thought to have undergone 65-74% of its genome reduction early in its association with the aphid (van Ham et al., 2003). There is also a trend for the rate of genome reduction to decrease exponentially over time (Moran, 2002). This is another indication that studying the earliest timepoints in the symbiotic relationship is crucial for understanding how and why they occur. Belda et al. (2012) attempted to estimate the evolution of *S. glossinidius* by reconstructing an ancestral metabolic network from predicted pseudogenes. This work is however biased in that the end-point (the *S. glossinidius* genome) is informing the starting point. The discovery of *S. praecaptivus* provides an unbiased starting point from which to predict and evaluate the trajectory of the *S. glossinidius*-tsetse symbiosis.

Predicting what will happen to a symbiont in the future is as important as deciphering the key events that occurred at initiation. Using FBA to study the evolution of the minimal metabolic networks observed in symbiotic bacteria has been attempted previously. Pál et al. (2006) used an *E. coli* metabolic model (*i*JR904 (Reed et al., 2003)) to simulate the evolution of *B. aphidicola* and *W. glossinidia*. A reaction was selected at random, removed, and then deleted permanently if its removal did not result in a lethal phenotype. This was repeated until no further reactions could be removed, with the entire process replicated 500 times. Pál et al. (2006) were able to produce minimal networks with a high (80%) degree of similarity to *B. aphidicola*

and *W. glossinidia*. This is, however, not entirely surprising; they began with an *E. coli* model and aimed to produce a model for an organism with a genome described as a subset of *E. coli* (Shigenobu et al., 2000). This technique is also limited in its ability to explore the evolutionary space. Once a randomly selected reaction is removed, it can never be replaced. The removal of this reaction, and others that occupy the same niche within a model, will result in a set of reactions becoming essential. By not allowing reactions to be added back into the system further down the line, the number of possible solutions found by the evolution is limited. It does however provide a framework for harnessing FBA to study evolution *in silico* through gene (or reaction) knockouts.

This ability to conduct *in silico* gene knockouts is a strength of using FBA to study symbiotic bacteria, as these organisms are often not amenable to traditional molecular biology tools for genetic manipulation. Obligate symbionts cannot in their nature be transformed as they are not cultivatable outside of their insect hosts. For the tsetse symbiont *S. glossinidius*, there are only a few examples of successful molecular engineering. There is one example of a modification of *S. glossinidius* by lambda Red-mediated recombination (Pontes and Dale, 2011), but this is not replicated elsewhere. It has been reported that *S. glossinidius* was transformed with a plasmid that enabled the production of a single domain antibody (De Vooght et al., 2014), as well as a plasmid encoding antibiotic resistance genes (Beard et al., 1993), but these are the exceptions. Computational evolution experiments, in which key genes or reactions are removed and the effects monitored, could therefore act as a proxy for *in vitro* genetic manipulation in symbionts that are not amenable to traditional techniques. A tool that combines the strengths of FBA with a way to explore the evolutionary space could provide novel insights into the trajectories that symbiotic bacteria undergo.

1.6 Chapter summary

1.6.1 Chapter 2 - A Tale of Three Species: Adaptation of *Sodalis glossinidius* to tsetse biology, *Wigglesworthia* metabolism and host diet

The metabolic requirements and interactions of *S. glossinidius* with its tsetse host are poorly understood. To increase the understanding of how the symbiont interacts with its microoenvironment and the adaptations that it has undergone, we have con-

structed a metabolic model of *S. glossinidius* metabolism. The model enabled the design and experimental verification of a defined medium that supports *S. glossinidius* growth *in vitro*. This has subsequently been used to analyse aspects of *S. glossinidius* metabolism, revealing multiple unique adaptations of the symbiont to its environment. Continued dependence on a sugar, and the importance of the chitin monomer *N*-acetyl-D-glucosamine as a carbon and energy source, suggests adaptation to host-derived molecules. Adaptation to the amino acid-rich blood diet is revealed by a strong dependence on L-glutamate as a source of carbon and nitrogen. Growth of *S. glossinidius* is demonstrated in the absence of L-arginine, contradicting a previously assumed auxotrophy. Finally, the selective loss of thiamine biosynthesis, a vitamin provided to the host by *W. glossinidia*, reveals an inter-symbiont dependence. The reductive evolution of *S. glossinidius* to exploit environmentally derived metabolites has resulted in multiple weaknesses in the metabolic network. These weaknesses may become targets for reagents that inhibit *S. glossinidius* growth and aid the reduction of trypanosomal transmission.

1.6.2 Chapter 3 - A genome scale metabolic model for *Sodalis praecaptivus* reveals unusual biosynthetic pathways

There is a well-annotated genome available for the free-living *S. praecaptivus*. This provides the opportunity to study its metabolism in depth, as well as using it as a comparison to its symbiotic relative *S. glossinidius*. Here, we describe the first FBA model of *S. praecaptivus* metabolism, *i*RH830. This model describes *S. praecaptivus* as a protroph, capable of growth on an *in silico* minimal medium. The model is extensively tested, with the large-scale screen of carbon usage revealing two pathways rarely seen in FBA models of Gram negative bacteria. These pathways, for xylitol and *N*-acetyl-D-galactosamine usage, are incorporated into *i*RH830, highlighting the importance of *in vitro* testing of FBA models. Analysis of metabolic pathways in *S. glossinidius* suggests that the symbiont once had, and has subsequently lost, the ability to metabolise xylitol. This model can serve as a starting point for studies into the evolution of metabolic networks in the *Sodalis* genus.

1.6.3 Chapter 4 - The *in silico* evolution of *Sodalis glossinidius*

FBA has now been used to augment the understanding of how *S. glossinidius* and *S. praecaptivus* have adapted to their contrasting environments. Little is known, however, about the evolutionary trajectory that *S. glossinidius* has undertaken during its internalisation, or what may happen to the symbiont in the future. The temporal occurrence of key pseudogenisations is not known, and the extent of the effect of the tsetse-specific metabolite availability is not fully understood. This lack of understanding is not limited to the tsetse-*S. glossinidius* relationship, but to all insect-bacterial symbioses. We therefore present a tool to study the evolution of symbiosis *in silico*. We combine FBA with a multi-objective evolutionary algorithm, using the *Sodalis* genus as a model system. We describe how this has enabled predictions to be made about how early key genes were lost in evolutionary time, the impact of growth media, and the future course of genome reduction in *S. glossinidius*. This method could be applied to other symbioses for which a well-annotated genome is available. It may also be suited for questions in industrial biotechnology, in which directed evolution could be beneficial.

1.6.4 Chapter 5 - Phylogenetic and experimental analysis of amino acid transporters in *Wigglesworthia glossinidia*

FBA, whilst a useful tool in several ways, is limited without experimental validation of its predictions. A pre-existing model of *W. glossinidia* predicts a reliance on the abundant metabolite L-proline, but there have been no transporters to import this amino acid characterised in this organism. In this chapter we present a phylogeny of amino acid binding proteins across members of the Enterobacteriaceae. This has been used to categorise a candidate *W. glossinidia* binding protein as a member of a unique cluster without an *E. coli* orthologue. We present two strains of *W. glossinidia*, *W. glossinidia brevipalpis* and *W. glossinidia morsitans*, as more dissimilar at the protein sequence level than several inter- and intra-genus comparisons. A candidate amino acid binding protein from *W. glossinidia* is expressed and purified, and undergoes extensive structure and function tests with the aim of characterising its binding capability. The data presented here highlight the variability between strains of obligate symbionts, and the importance of acknowledging their deviation away from standard comparisons of free-living bacteria.

By understanding the metabolic role of the symbionts within their host we can also

understand how the internal environment is modified and in turn how this may affect the persistence of *T. brucei*. The tsetse microbiome is well-characterised and therefore provides a unique opportunity to study the evolution of symbiosis. It is vitally important to elucidate the relationships between host and symbiont because, as stated by Ed Yong (2016):

We cannot fully understand the lives of animals without understanding our microbes and our symbioses with them.

- Ed Yong, I Contain Multitudes: The microbes within us and a grander view of life

Chapter 2

A Tale of Three Species: Adaptation of *Sodalis glossinidius* to tsetse biology, *Wigglesworthia* metabolism and host diet

The following chapter has been published as:

Hall RJ, Flanagan LA, Bottery MJ, Springthorpe V, Thorpe S, Darby AC, Wood AJ, Thomas GH (2019). A Tale of Three Species: Adaptation of *Sodalis glossinidius* to tsetse biology, *Wigglesworthia* metabolism and host diet. mBio 10 (1) e02106-18.

RJH conducted the experiments and analysed the data. LAF and MJB conducted preliminary *in silico* and *in vitro* work, respectively. VS constructed DETOXbase. ST assisted with programming. AJW and GHT supervised the project. All authors contributed to the final manuscript.

Small changes have been made to the published manuscript to incorporate the supplementary figures into the main text.

2.1 Introduction

It has been estimated that only 1% of all microbial life is culturable (Staley and Konopka, 1985; Amann et al., 1995; Vartoukian et al., 2010). Included in this are

a vast array of symbiotic bacteria. Inter-species competition, as well as sensitivity to temperature, pH, oxygen and nutrient availability, means many species cannot be cultured using standard conditions (Vartoukian et al., 2010; Koch, 1997; Hugenholtz et al., 1998). The ability to culture medically significant microorganisms is an important tool in disease control. This includes the pathogens and the key symbionts within the system. Improved culture methods, combining microbiology with genomics, have been used to analyse the microbial flora of a number of disease vectors. Notable examples are *Paenibacillaceae* and *Serratia marcescens* in the Asian malarial vector *Anopheles stephensi* (Rani et al., 2009), the more complex flora of *Aedes aegypti* (Gusmão et al., 2010), and the defined microbiome of the tsetse fly, insect vector for the *Trypanosoma brucei* parasites that cause human African trypanosomiasis (HAT) (Aksoy, 2000).

HAT is endemic in 36 countries in sub-Saharan Africa, with an estimated 65 million people at risk of infection (Roditi and Lehane, 2008; Migchelsen et al., 2011; Simarro et al., 2011). The tsetse, genus *Glossina*, also hosts a limited bacterial microbiome alongside the parasitic *T. brucei*. Namely, this consists of a primary, obligate symbiont *Wigglesworthia glossinidia* and, typically, a secondary facultative symbiont *Sodalis glossinidius* (Aksoy, 1995; Dale and Maudlin, 1999). *S. glossinidius* is of medical importance as its presence correlates positively with the ability of the tsetse to be infected by *T. brucei* (Welburn and Maudlin, 1991; Welburn et al., 1993; Welburn and Maudlin, 1999; Geiger et al., 2006). Its complement of over 1500 pseudogenes, and large genome size of 4.17 Mb, is consistent with it making a rapid and recent movement from free-living to a host-restricted niche (Toh et al., 2006; Belda et al., 2010). The high rate of pseudogene accumulation is consistent with the loss of many cellular processes and metabolic pathways that are no longer needed for life in the tsetse. These include genes involved in the transport of carbohydrates not present in the blood meal (Toh et al., 2006) and of L-arginine biosynthesis (Belda et al., 2010). The recent discovery of a closely related, free-living species of *Sodalis*, *S. praecaptivus*, provides a useful, relevant comparison (Chari et al., 2015). It enables informed predictions to be made about the presence or absence of key metabolic genes in *S. glossinidius*. The hypothesis that *S. glossinidius* has specifically lost metabolic capabilities during its transition to symbiosis can also be tested.

Symbiotic bacteria often present with small, degraded genomes (McCutcheon and Moran, 2011). As a result of gene loss and inactivation, symbionts often cannot be grown outside of their host. *S. glossinidius* can however still be cultured, but it requires undefined rich media (Dale and Maudlin, 1999) and a longer incubation time than the free-living *S. praecaptivus* (Chari et al., 2015). This increases the risk of contamination by faster-growing organisms and limits *in vitro* study of metabolite

essentiality. Rationally designed growth media was achieved in a landmark paper for the causative agent of Whipple's disease *Tropheryma whipplei* (Renesto et al., 2003), but this still contained undefined components. An entirely defined medium will improve the culturing of *S. glossinidius* and the study of its physiology dramatically. This may then enable genetic manipulation of this organism to express antiparasitic molecules toward the elimination of *T. brucei* (Darby et al., 2005b; Weiss et al., 2006; Snyder and Rio, 2013). This process is already a consideration for the control of other vector-borne diseases (Hillesland et al., 2008; Coutinho-Abreu et al., 2010).

To define a *S. glossinidius*-specific growth medium, with the eventual aim of understanding more about the symbiont's biology and metabolic dependencies, an experimental approach was combined with whole genome metabolic modelling (GEM) and flux balance analysis (FBA) to model *S. glossinidius in silico*. This is a powerful method when based on a well-annotated genome and the ability to test *in silico* hypotheses experimentally Orth et al. (2010). Analysis of the metabolic network of *S. glossinidius* was first undertaken by Belda et al. (2012), who describe a network of 458 gene products and 560 reactions, *i*EB458. A key finding was the pseudogenisation of the phosphoenolpyruvate (PEP) carboxylase gene (*ppc*), preventing the conversion of PEP to oxaloacetate for the tricarboxylic acid (TCA) cycle. The pseudogenisation of components of the L-arginine biosynthesis pathway indicate the requirement of an external source of L-arginine to supplement growth *in silico*. They conclude that exogenous L-arginine is required both as a biomass component and to form succinate via putrescine in order to supplement the TCA cycle in the absence of *ppc*. The common hexose sugar D-glucose is given as the sole carbon source. Importantly, this construction of *i*EB458 was limited by a lack of a well-annotated relative from the same genus, which is no longer an issue since the discovery of *S. praecaptivus*.

We present here a significantly advanced and improved model, *i*LF517, and describe how it has enabled the development of an entirely defined medium that supports *S. glossinidius* growth *in vitro* (SGM11). Our data indicate the use of a carbon source lacking in the blood meal, namely *N*-acetyl-D-glucosamine (GlcNAc). This suggests a complex nutritional interaction of *S. glossinidius* with the tsetse chitinous peritrophic matrix. Degradation of this by a microbe-derived chitinase might explain the increased persistence of the trypanosome when *S. glossinidius* is present (Welburn and Maudlin, 1991; Welburn et al., 1993). Using SGM11, we demonstrate that *S. glossinidius* is not, as thought previously, a true auxotroph for L-arginine. Rather, it has a unique vitamin auxotrophy for thiamine, likely provided by the primary symbiont *W. glossinidia* (Snyder et al., 2012; Snyder and Rio, 2015) through an interaction currently undefined.

2.2 Methods

2.2.1 Refinement of the *S. glossinidius* metabolic network

The previously published GEM, *i*EB458 (Belda et al., 2012), was assessed for missing or potentially incorrect gene assignments. A reaction was removed if there could be no functional gene identified, either through absence or through pseudogenisation based on the size of the gene in comparison to its *Escherichia coli* or *S. praecaptivus* orthologues. Those reactions for which a gene assignment had been uncovered were added to the new model. Reactions were maintained if removing them resulted in a lethal phenotype, observed when the biomass output returned a value of zero. BiGG Models (King et al., 2016), KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2019) and EcoCyc (Keseler et al., 2017) databases were used to identify *E. coli* genes encoding the reactions for which a *S. glossinidius* gene assignment had not been found. Translated nucleotide and protein BLAST were used to search for known *S. glossinidius* proteins and confirmation of genes and pseudogenes was performed using the Artemis Genome Visualisation Tool (Rutherford et al., 2000). Candidate pseudogenes were aligned with functional orthologues using ClustalX 2.1 (Larkin et al., 2007).

2.2.2 Flux balance analysis

The FBA solutions were generated using the GNA linear programming kit (GLPK) integrated with custom software in Java. Oxygen uptake was constrained to 12 mmol gr DW^{-1} hr^{-1} in order to simulate a reduced oxygen environment. The uptake of ammonia, water, phosphate, sulphate, potassium, sodium, calcium, carbon dioxide, protons and essential transition metals was unconstrained. Uptake of all other metabolites was set to zero, with the exception of those used in the analyses which have been set at 2 mmol gr DW^{-1} hr^{-1} GlcNAc and L-glutamate, 0.5 mmol gr DW^{-1} hr^{-1} L-arginine and 0.01 mmol gr DW^{-1} hr^{-1} thiamine. Cofactor constraints were implemented by introducing these metabolites to the biomass functions at small fluxes (1 x 10^{-5} mmol gr DW^{-1} hr^{-1}) (Thomas et al., 2009). The phenotype was considered viable if the biomass production rate was greater than 1 x 10^{-4} gr DW (mmol glucose)$^{-1}$ hr^{-1}. Futile cycles, identified as reactions carrying biochemically unsustainable flux, were altered to the correct reaction stoichiometry where possible with guidance from EcoCyc (Keseler et al., 2017) and BiGG (King et al., 2016). Full description of the model is provided in Supp. File 1.

2.2.3 Bacterial strains, growth conditions and reagents

S. glossinidius strain GMM4 was obtained from the University of Liverpool. Working stocks were established by growing starter cultures on brain heart infusion (BHI) (Sigma-Aldrich) plates in microaerophilic conditions generated by Oxoid CampyGen sachets (Thermo Fisher Scientific) until growth was visible. Colonies were then transferred to liquid BHI media and incubated for four to seven days in cell culture flasks at room temperature until growth was visible. *S. glossinidius* from the working stock was then transferred to 5 ml fresh Luria-Bertani (LB) media (Sigma-Aldrich) and supplemented with either D-glucose or GlcNAc at concentrations of 83 mM (Sigma-Aldrich).

*i*LF517 and the *S. glossinidius* metabolic network were used to design *in silico* an entirely defined media, SGM11, in which to grow the bacterium. *S. glossinidius* from the working stock was transferred to 5 mL M9 minimal media (Neidhardt et al., 1974) containing the following supplements; 17 mM GlcNAc, 17 mM trehalose, 17 mM L-serine, 17 mM L-arginine, 4 mM L-proline, 17 mM L-glutamic acid monosodium salt hydrate, 17 mM L-aspartate, 4 mM nicotinamide, 9 mM α-ketoglutarate, 9 mM fumaric acid, 400 μM thiamine monophosphate (Sigma-Aldrich). Metabolites were omitted from SGM11 individually and in combination to test the model predictions.

The culture flasks were incubated for 48 hours (LB) or 72 hours (SGM11) at 25°C in a temperature-controlled water bath. Gentle agitation was achieved using magnetic stir bars, to achieve a suitable balance between oxygenation, settling and disturbance, was implemented at a stirring speed of 500 rpm. Intermediate time points were found to compromise the sterility of the cultures and therefore destructive sampling was the only reliable method of investigation, negating the possibility of higher resolution temporal data. End-point increase in *S. glossinidius* growth was measured at an optical density of 650 nm. Preliminary experiments using variable sampling times indicated that this was the most appropriate sampling time for *S. glossinidius* to repeatably capture the final steady state but still retain a proxy for growth rate to guide the modelling results.

E. coli gene deletion mutants were obtained from the Keio collection (Baba et al., 2006). Cells were cultured in M9 minimal media (Neidhardt et al., 1974) with 0.4% D-glucose, and either 20 or 100 mM L-glutamic acid monosodium salt hydrate or 20 mM L-arginine in a microplate reader (Epoch) for 24 hours at 37°C.

2.2.4 Flow cytometry

Cell count was generated using *S. glossinidius* taken from a starter culture in BHI and diluted in M9 salts. Cells were stained with DAPI at 2 μL/mL for 10 minutes at room temperature and measured on the CytoFLEX S (Fig. S1). Flow cytometer was calibrated using counting beads from Beckman Coulter (Miami USA).

2.2.5 Statistical analysis

All statistical analyses were performed using SciPy in Python (version 2.7.10). Error bars show standard error of the mean, statistical significance assessed using one-way analysis of variance (ANOVA).

2.3 Results

2.3.1 The genome of *S. praecaptivus* enables an improved analysis of the *S. glossinidius* metabolic network

S. praecaptivus is the only free-living member of the *Sodalis* genus to have been characterised. It has a 5.16 Mb genome with a 57.5% G+C content (Chari et al., 2015). Using this discovery, *S. praecaptivus* was compared to *S. glossinidius* to reassess the existing metabolic model of the symbiont. This additional information verified many of the important findings in *i*EB458, whilst others relating to carbon and nitrogen usage were not supported.

One central hypothesis derived from *i*EB458 comes as a consequence of the inactivation of the PEP carboxylase reaction encoded by *ppc* (Belda et al., 2012). This loss in *S. glossinidius* should pose a problem for its metabolism as it loses a route to replenish oxaloacetate from PEP. This represents an important anapleurotic reaction to maintain high flux through the TCA cycle in the related bacterium *E. coli*. To compensate for this loss, Belda et al. (2012) hypothesised a three-fold function for exogenous L-arginine for *S. glossinidius*; as a biomass component, as a biosynthetic precursor to putrescine and spermidine and as an anapleurotic substrate, via succinate (Fig. 2.1). A functional *ppc* is present in *S. praecaptivus* (Sant_3959) whereas the gene in *S. glossinidius* contains multiple frameshifts and premature stop codons (Fig. 2.2), suggesting loss as a result of selection pressures or genetic drift.

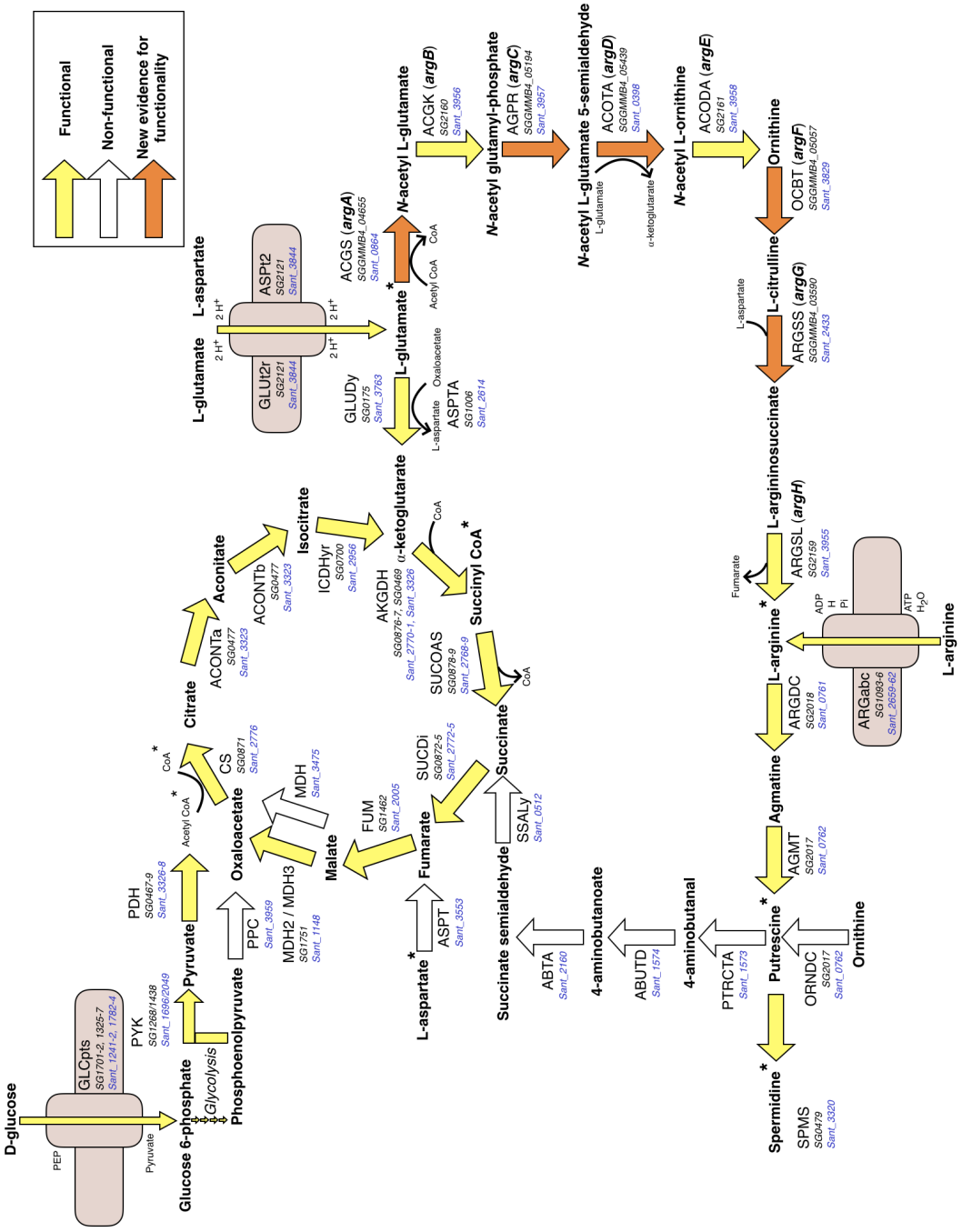


Figure 2.1: Overview of central metabolism in the *S. glossinidius* metabolic network. Functional (yellow) and non-functional (white) pathway components are highlighted, with the *S. glossinidius* and *S. praecaptivus* gene associations given in black and blue type, respectively. Reactions for which new evidence has suggested that they might be functional are given in orange. The TCA cycle is complete and transporters for L-arginine, L-glutamate and L-aspartate are present. PPC and the reactions connected ornithine with the TCA cycle are pseudogenised. There is evidence from the new genome annotation that the L-arginine biosynthesis pathway from L-glutamate may not be entirely pseudogenised, as previously thought (Belda et al., 2012). Selected co-reactants are included. * indicates biomass components.

The proposed anapleurotic link from L-arginine to succinate is not supported in our analysis, with little or no evidence for these genes being present in *S. glossinidius* (Fig. 2.1 & Supp. File 1). L-arginine must link to the TCA cycle to serve as an anapleurotic substrate. In *i*EB458 this linkage was proposed to occur via putrescine transaminase (PTRCTA), aminobutyraldehyde dehydrogenase (ABUTD) and 4-aminobutyrate transaminase (ABTA) (Belda et al., 2012) (Supp. File 1). PatA is required for this PTRCTA reaction but BLASTp searches find no evidence of an orthologue in *S. glossinidius* (Fig. 2.1). There is however a functional *patA* gene in *S. praecaptivus* (Sant_1573). Similarly, there is no functional orthologue of *patD* for ABUTD, nor of *puuE* or *gabT* for ABTA. L-arginine can still be converted to agmatine and then to putrescine via arginine decarboxylase (ARGDC, SG2018) and agmatinase (AGMT, SG2017). This is the only route to the synthesis of this biomass component, meaning a source of L-arginine in the cell is still predicted to be essential. It is not however supporting the proposed additional anapleurotic function.

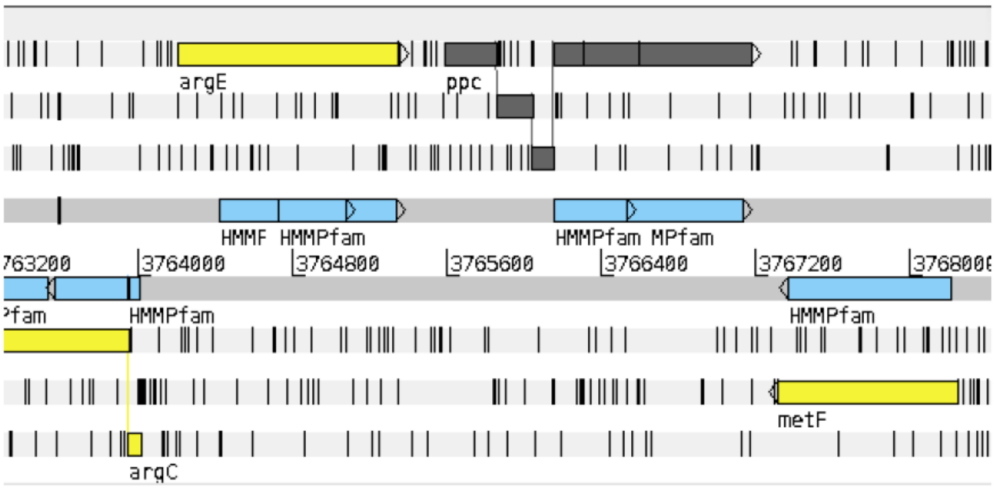


Figure 2.2: Analysis of a portion of the *S. glossinidius ppc* gene in Artemis. Intact coding sequences are indicated in yellow. Vertical lines represent stop codons. The *ppc* gene (dark gray) is split across three reading frames due to frameshift mutations and contains two premature stop codons.

An alternative organic compound must therefore serve the role of supplying TCA cycle intermediates in the absence of *ppc*. Removal of the ABTA reaction from *i*EB458, breaking the link of putrescine to succinate, results in zero biomass production. This can be rescued by the *in silico* addition of metabolites that can be introduced easily into the TCA cycle; the amino acids L-aspartate or L-glutamate or the organic acids succinate, fumarate or α-ketoglutarate. The loss of *ppc* therefore requires the addition of a second organic substrate in addition to glucose for growth.

This is not L-arginine as suggested by Belda et al. (2012). L-arginine is taken up for biomass production, either directly or for conversion to putrescine and spermidine. It is not subsequently converted to succinate to enter the TCA cycle. This specific loss of *ppc* in *S. glossinidius* suggests an adaptation to an amino acid rich environment that results from the tsetse blood diet. Important components, including L-aspartate and L-glutamate, are predicted to be present at high concentrations.

2.3.2 A revised metabolic model, *i*LF517, for *S. glossinidius*

A systematic reanalysis of the *S. glossinidius* genome enabled the construction of an independent metabolic model, *i*LF517. This model has significant differences to *i*EB458 (Belda et al., 2012). Growth was not supported in *i*LF517 using the uptakes of oxygen, D-glucose and L-arginine given in *i*EB458, indicating alternative carbon and nitrogen sources are used. *S. praecaptivus* was used as a comparator to assess the presence of important metabolic genes in *S. glossinidius*, a resource not available to Belda et al. (2012). Full details of all reactions removed from *i*EB458 and those added to *i*LF517 are highlighted in Supp. File 1. *i*LF517 contains 517 genes, 703 metabolites and 638 reactions (excluding pseudoreactions). This model can be viewed and analysed through a web-based FBA browser on DETOXbase (www.detoxbase.org/publications/iLF517).

*i*LF517 was analysed via FBA to investigate metabolite essentiality and the presence of predicted auxotrophies. 80 reactions are included in *i*LF517 that were absent in *i*EB458 and 32 have been removed (Supp. File 1). *i*EB458 simulates high oxygen transfer rates, using an uptake value of 20 mmol gr DW^{-1} hr^{-1}. This value was selected originally for *E. coli* and indicates highly aerated growth conditions in a chemostat (Andersen and von Meyenburg, 1980; Varma et al., 1993; Reed et al., 2003). *S. glossinidius* is however sensitive to high levels of oxygen (Dale and Maudlin, 1999). Cultures used here were grown in conditions of reduced aeration in comparison to *E. coli* or *S. praecaptivus*. The oxygen uptake rates given in *i*EB458 are therefore unrealistic for the simulations. Decreasing the oxygen supplied to *i*EB458 *in silico* results in a decrease in biomass output (Fig. 2.3), demonstrating that using unrealistic oxygen uptake rates exaggerate possible growth. The oxygen uptake rate in *i*LF517 was subsequently reduced to 12 mmol gr DW^{-1} hr^{-1}, guided by the value given in a model of another microaerophile, *Helicobacter pylori* (Thiele et al., 2005).

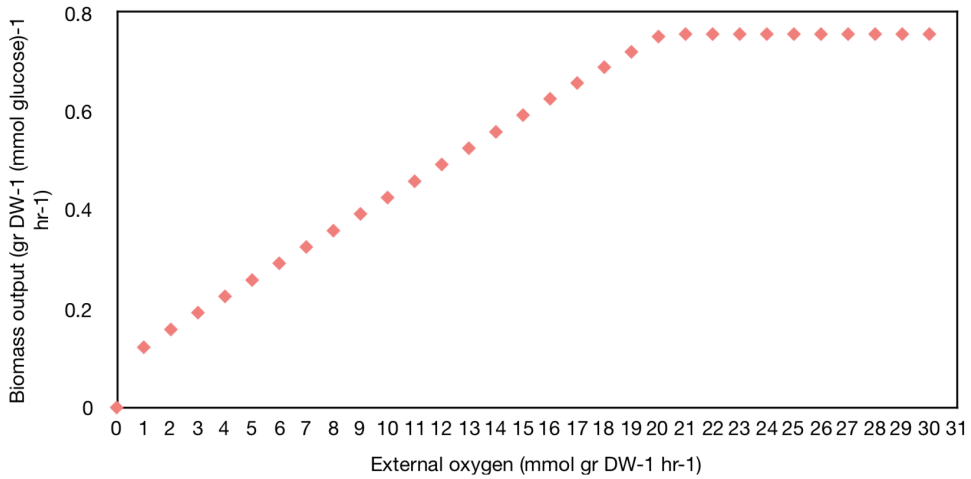


Figure 2.3: Change in biomass output with increasing oxygen in *i*EB458. Supplemented oxygen was increased until a plateau in biomass output was observed. *i*EB458 is supplemented with oxygen at 20 mmol gr DW^{-1} h^{-1}. This is reduced to 12 mmol gr DW^{-1} h^{-1} in *i*LF517 in line with the microaerophilic *H. pylori*.

2.3.3 A defined medium, SGM11, supports *S. glossinidius* growth

The use of complex media is insightful for examining certain aspects of bacterial physiology. It does however limit the ability to investigate all metabolic functions. A defined medium with components of known concentrations is therefore desirable. *i*LF517 was used to design a defined medium, SGM11, containing metabolites that the model predicts may enhance *S. glossinidius* growth or becoming limiting.

No growth was observed after 72 hours incubation in M9 medium (Neidhardt et al., 1974) supplemented with GlcNAc as a carbon source (Fig. 2.4A). The addition of trehalose, L-serine, L-arginine, L-proline, L-glutamate, L-aspartate, nicotinamide, α-ketoglutarate, fumarate and thiamine monophosphate (TMP), to create SGM11, resulted in higher yields than with LB alone, although this difference was not statistically significant (Fig. 2.4A). Cell concentration was estimated using flow cytometry as approximately 3.8 x 108 (SEM 9.7 x 107) for an OD_{650} value of 0.28 (Fig. S1) in order to gain an indication of the number of cells that growth in LB corresponds to. SGM11 provides a defined starting point to test the essentiality of key metabolites.

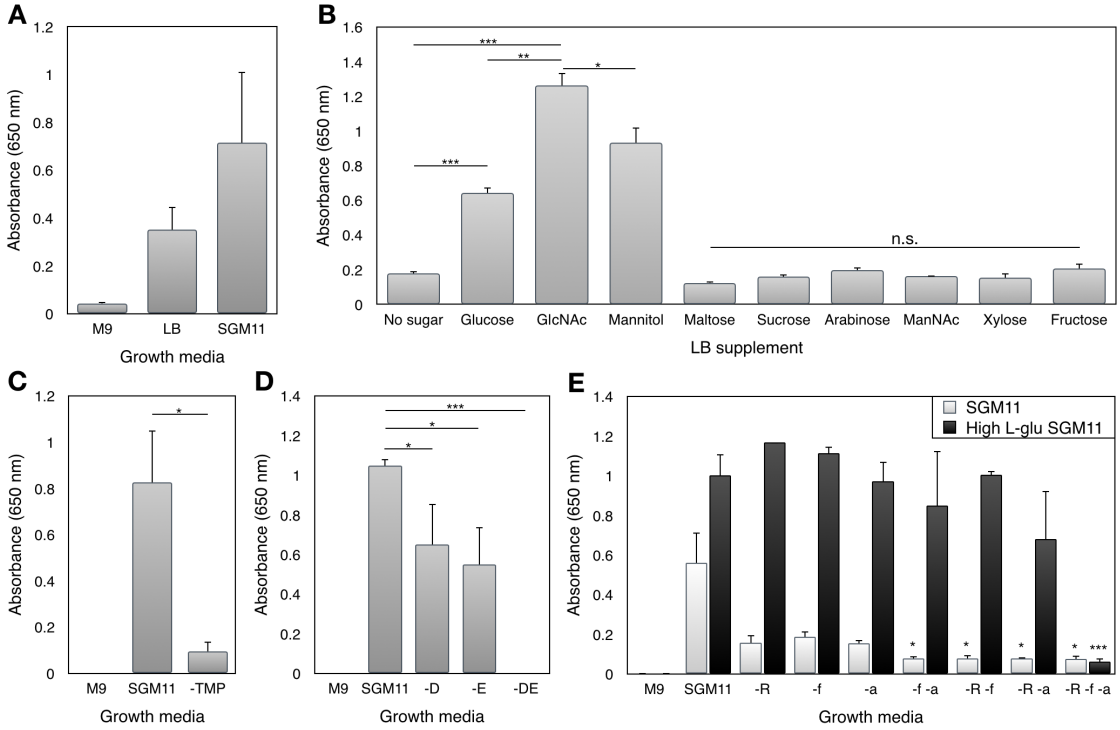


Figure 2.4: Testing *S. glossinidius* metabolism and *i*LF517 predictions experimentally. **(A)** The custom, defined growth medium SGM11 supports *S. glossinidius* growth to an average optical density (650 nm) of approximately 0.7. **(B)** Supplementation of LB with D-glucose, GlcNAc or mannitol results in significantly greater *S. glossinidius* growth than without supplementation of a carbon source. No other carbon source had a significant improvement. GlcNAc supplementation results in significantly greater growth in comparison to either D-glucose or mannitol. No sugar treatment represents pooled triplicates from two experiments, mannitol five replicates pooled from two experiments. **(C)** *S. glossinidius* cannot grow in SGM11 when thiamine monophosphate has been removed (-TMP). **(D)** The removal of L-aspartate (-D) or L-glutamate (-E) from SGM11 reduces *S. glossinidius* growth significantly ($p<0.05$). Removing both (-DE) abolishes growth entirely ($p<0.001$). **(E)** Removing L-arginine (-R), fumarate (-f) or α-ketoglutarate (-a) from SGM11 (light grey) impairs *S. glossinidius* growth. Removing two or all of these metabolites reduces growth significantly ($p<0.05$). In SGM11 with high L-glutamate (black), only one of these metabolites is required to support normal growth. Removal of L-arginine, fumarate and α-ketoglutarate abolishes growth ($p<0.001$). Measurements end-point growth in triplicate, unless specified. Error bars SEM. Significant differences to SGM11 indicated. *$p<0.05$, **$p<0.01$, ***$p<0.001$, one-way ANOVA.

2.3.4 *S. glossinidius* maintains a reliance on a sugar, namely the host-derived *N*-acetyl D-glucosamine

D-glucose and other saccharides were investigated for their potential to act as the main carbon source for *S. glossinidius*. The re-annotated *S. glossinidius* genome

showed that the glucose-specific phosphotransferase system (PTS) gene *ptsG* had been pseudogenised. It is intact in *S. praecaptivus* (Sant_2470), suggesting that there may be weak selection for its retention within the tsetse. Whilst the non-specific ManXYZ could substitute for this loss of function (Fig. 2.5), alternative carbon sources were examined computationally and experimentally.

This investigation highlighted immediately the presence of a GlcNAc-specific PTS transporter gene, *nagE*, in *S. glossinidius* (*SG0859*) (Toh et al., 2006) (Fig. 2.5). The maintenance of *nagE* alongside the promiscuous ManXYZ implies that GlcNAc could be an important carbon source. GlcNAc is also of particular interest with regards to tsetse biology. It is a breakdown product of the insect's chitinous peritrophic membrane and a potential link with the persistence of trypanosome infection (Welburn et al., 1993; Aksoy et al., 2001; Dyer et al., 2013). The mannitol-specific transporter encoded by *mtlA* (*SG0014*) is also retained (Toh et al., 2006) (Fig. 2.5). When used as the main carbon source, *i*LF517 produced biomass output values of 0.30, 0.35 and 0.32 gr DW (mmol glucose)$^{-1}$ hr^{-1} for D-glucose, GlcNAc and mannitol, respectively.

S. glossinidius was grown experimentally in LB and LB supplemented with a selection of carbon sources test the hypothesis that GlcNAc and mannitol may be suitable alternatives. Of the saccharides tested, only D-glucose, GlcNAc and mannitol increase growth significantly in comparison to LB alone (Fig. 2.4B) ($p<0.01$, one-way ANOVA). Approximately two times greater end-point growth is exhibited with an equimolar amount of GlcNAc in comparison to D-glucose ($p<0.05$, one-way ANOVA) (Fig. 2.4B). Normalising the carbon added from D-glucose with regard to GlcNAc has no significant effect on the optical density reached (data not shown). This demonstrates that the difference in growth between D-glucose and GlcNAc is not as a result of the additional carbon. *S. glossinidius* grows significantly better on GlcNAc than mannitol ($p<0.05$, one-way ANOVA), reflecting the *in silico* results qualitatively.

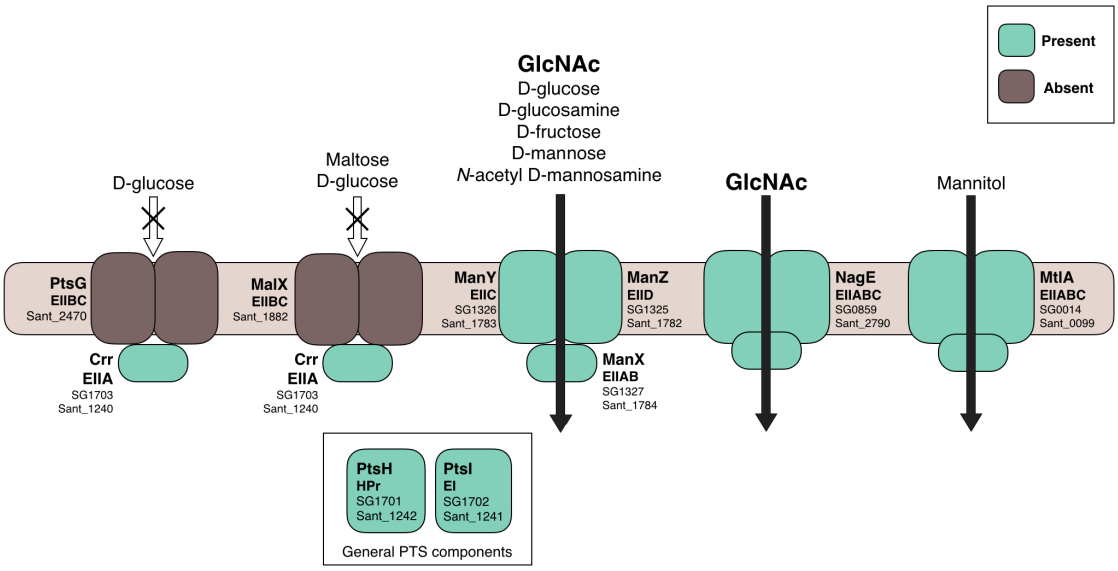


Figure 2.5: Predicted PTS transport in *Sodalis* species. The presence (green) or absence (brown) of PTS proteins in *S. glossinidius* are shown, with the corresponding orthologue in *S. praecaptivus* (Sant_) given for reference. The genes encoding the D-glucose-specific PtsG and MalX are likely pseudogenised. *S. glossinidius* has retained the ability to transport GlcNAc and mannitol through the specific NagE and MtlA systems, respectively. Other carbon sources can be imported via the promiscuous ManXYZ.

2.3.5 *S. glossinidius* has adapted to thiamine produced by the primary tsetse symbiont

The addition of L-arginine, L-glutamate and a carbon source, namely GlcNAc, does not produce a positive biomass output in *i*LF517. It is likely that *S. glossinidius* requires supplementation from certain vitamins that it cannot synthesise. During the transition to symbiosis, *S. glossinidius* may lose genes that encode vitamin and cofactor biosynthetic pathways in favour of retaining transporters. There is also thought to be a connection between aspects of the tsetse microbiome in terms of vitamin biosynthesis (Belda et al., 2010; Rio et al., 2016).

S. glossinidius appears to have retained the components of pantothenate (Table S1), biotin (Table S2), riboflavin (Table S3), protoheme (Table S4), NAD (Table S9), PLP (Table S5) and tetrahydrofolate (Table S7) biosynthesis pathways found in *S. praecaptivus* (Fig. 2.6). Neither *S. praecaptivus* nor *S. glossinidius* can synthesise cobalamin (Table S8). The key difference between the two species is in the pathway for thiamine biosynthesis (Table S6). The loss of this pathway in *S. glossinidius* has been noted previously (Rio et al., 2016), but here the complete pathway in *S.*

praecaptivus (Sant_3916-21) is used as a comparison. The loss of this pathway in *S. glossinidius* may be a specific adaptation to the tsetse. To investigate this further, three other *Sodalis*-allied symbionts were examined; *Candidatus* Sodalis pierantonius str. SOPE from the rice weevil *Sitophilus oryzai* (Gil et al., 2008; Oakeson et al., 2014), and *Sodalis*-like symbionts from the meadow spittlebug *Philaenus spumarius* (Koga and Moran, 2014) and the seed bug *Henestaris halophilus* (Santos-Garcia et al., 2017). The *P. spumarius* symbiont appears most similar to *S. praecaptivus*, with the fewest number of genes predicted to be absent or pseudogenised, whereas the *H. halophilus* symbiont has lost the ability to encode the entire biotin and protoheme biosynthetic pathways (Fig. 2.6). *Ca.* Sodalis pierantonius differs from *S. glossinidius* in that it cannot synthesise biotin (Fig. 2.6). This organism, along with the *H. halophilus* symbiont, does however share similarities with *S. glossinidius* in the pseudogenisation of components of the thiamine pathway including *thiF*, *thiG* and *thiH*. This suggests an adaptation to symbiosis with certain insects. It is important to note that *S. glossinidius* is unusual in that it can function as either a primary or a secondary symbiont depending on the insect host, and therefore inter-species comparisons should be treated with caution (Rosas-Pérez et al., 2017). Thiamine is a cofactor for many enzymes, including pyruvate dehydrogenase (Webb et al., 1998), that are essential in *i*LF517. The potential thiamine auxotrophy in *S. glossinidius* was assessed *in silico* and supplementation of thiamine or TMP was required to produce a positive biomass output in *i*LF517 (Supp. File 1).

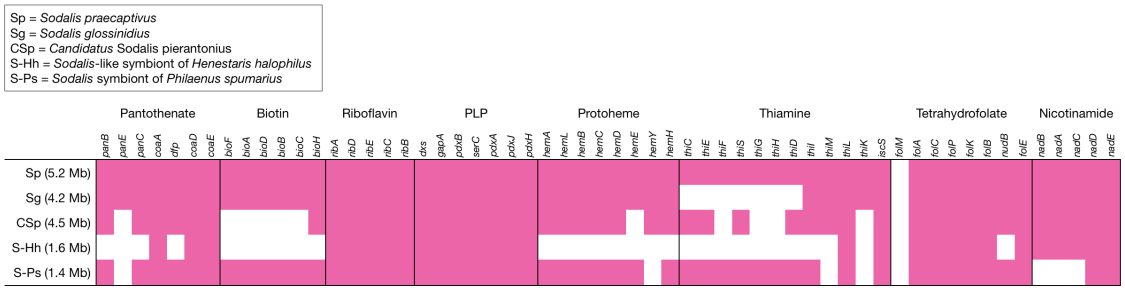


Figure 2.6: Completeness of vitamin biosynthesis pathways in *Sodalis* species. The presence (pink) or absence (white) of vitamin biosynthesis genes, found using tBLASTn and BLASTp, in *S. praecaptivus* (Sp), *S. glossinidius* (Sg), *Candidatus* Sodalis pierantonius (Csp), *Sodalis*-like symbiont of *Henestaris halophilus* (S-Hh), and *Sodalis* endosymbiont of *Philaenus spumarius* (S-Pf).

The reliance of *S. glossinidius* on an external source of thiamine in order to produce TMP was then investigated experimentally. TMP was removed from SGM11 and the ability of the symbiont to grow was measured. Removal of TMP from SGM11 resulted in a significant reduction in *S. glossinidius* growth ($p<0.05$, one-way ANOVA) (Fig. 2.4C). A thiamine ABC transporter (SG0431-3) that likely transports TMP

is also present (Hollenbach et al., 2002). This adds weight to the *in silico* evidence from *i*LF517 and the elegant empirical work of Snyder et al. (2010) that *S. glossinidius* relies on TMP from its environment. The only source of TMP in the tsetse microenvironment is that excreted by *W. glossinidia*, which is thought to supply it to the tsetse. *S. glossinidius* may therefore have adapted to not only its host but also to the metabolism of the primary symbiont in order to scavenge the available TMP.

2.3.6 *S. glossinidius* is dependent on external sources of L-glutamate or L-aspartate

Culturing *S. glossinidius* in SGM11 enables the thorough testing of amino acid usage in *i*LF517. Metabolites can be removed individually and the effect on biomass production measured. *i*LF517 requires either L-glutamate or L-aspartate to produce a positive biomass output. Both L-glutamate and L-aspartate likely enter the cell through the GltP transporter SG2121 (Fig. 2.1). *S. glossinidius* is predicted to have transporters for 14 amino acids (Table S10) and so this reliance on L-glutamate or L-aspartate is not merely due to the ability to transport only these amino acids.

Removing L-aspartate or L-glutamate from SGM11 individually resulted in a significant decrease in the growth yield achieved by *S. glossinidius* after incubating for 72 hours ($p<0.05$, one-way ANOVA) (Fig. 2.4D). Removing both L-aspartate and L-glutamate abolished growth completely ($p<0.001$, one-way ANOVA). This confirms that an exogenous source of one of these amino acids is essential for *S. glossinidius* growth. Examination of *i*LF517 reveals that L-glutamate feeds directly into the TCA cycle through deamination to α-ketoglutarate. The direct route to feed L-aspartate into the TCA cycle via fumarate (L-aspartase) is however missing in *S. glossinidius* (Fig. 2.1). When *i*LF517 is supplied with L-aspartate instead of L-glutamate, 69% of the available L-aspartate is channelled into the aspartate transaminase reaction (ASPTA), producing L-glutamate and oxaloacetate. The resulting biomass output was reduced from 0.35 to 0.31 gr DW $(\text{mmol glucose})^{-1}$ hr^{-1}, demonstrating that *S. glossinidius* can use L-aspartate if L-glutamate is not available although the latter may be preferred. L-glutamate is therefore likely an important energy source in *i*LF517 both to form L-aspartate and to replenish the TCA cycle at α-ketoglutarate (Fig. 2.1).

2.3.7 *S. glossinidius* is not an L-arginine auxotroph

Initial analysis of amino acid biosynthesis in *S. glossinidius* appeared to confirm existing opinion (Belda et al., 2012) that the only amino acid with an incomplete biosynthetic pathway is L-arginine (Table S10). A functional uptake system is also present, suggesting that *S. glossinidius* is indeed an L-arginine auxotroph. To assess this experimentally, *S. glossinidius* was grown in SGM11 with L-arginine removed. The growth yield decreased (Fig. 2.4E), but surprisingly was not totally abolished as expected for a true auxotroph. Excess L-glutamate was added to SGM11 to determine if it could rescue this reduction in growth, as L-glutamate appears to be a key metabolite to *S. glossinidius*. A 5-fold increase in L-glutamate concentration to 85 mM increased the bacterial yield (Fig. 2.4E), likely due to the extra carbon and nitrogen available. Remarkably, the excess L-glutamate rescued the growth defect caused by the removal of L-arginine completely (Fig. 2.4E and Fig. S2).

The *S. glossinidius* pathway for L-arginine biosynthesis was subsequently reanalysed using the latest genome annotation (LN854557) to assess its completeness in comparison to *S. praecaptivus*. This revealed that *argB*, *argE*, *argF*/*argI* and *argH* are full length and therefore likely functional (Fig. 2.1). It has been noted previously that *argC* is pseudogenised (Belda et al., 2012) but the new annotation indicates that this may not be the case. ArgC is also detectable by proteomics, suggesting this gene is indeed likely functional (Goodhead et al., 2018).

Table 2.1: Functionality of the L-arginine biosynthesis pathway. tBLASTn results for *S. glossinidius* orthologues of components of the L-arginine biosynthesis pathway in *E. coli* and *S. praecaptivus*. aa = amino acids, bp = base pairs. Those components for which the new *S. glossinidius* genome annotation has provided evidence for functionality are shown in bold type.

Gene	Size in *E. coli*	Size in *S. praecaptivus*	Reaction	*S. glossinidius*	Conclusion
argA	1,332 bp 443 aa	Sant_0864 418 aa	ACGS	*SGGMMB4_04654* (*argA_1*) *SGGMMB4_04655* (*argA_2*) 804 bp total	*argA_2* may produce a subunit that can function individually, using its GTG start codon
argB	777 bp 258 aa	Sant_3956 257 aa	ACGK	*SGGMMB4_05193* 765 bp	New annotation indicates full-length gene
argC	1,005 bp 334 aa	Sant_3957 334 bp	AGPR	*SGGMMB4_05194* 999 bp	New annotation indicates full-length gene and ArgC detected by proteomics
gabT	1,281 bp 426 aa	Sant_2160 425 aa	ACOTA	See *argD*	See *argD*
argD	1,221 bp 406 aa	Sant_0398 407 aa	ACOTA	*SGGMMB4_05438* (*argD_1*) 1-708 bp *SGGMMB4_05439* (*argD_2*) 681-828 bp	May use functional alternatives *bioA* (SG0902) or *hemL* (SG0500)
argE	1,152 bp 383 aa	Sant_3958 382 aa	ACODA	*SGGMMB4_05195* 1,146 bp	New annotation indicates full-length gene
argF	1,005 bp 334 aa	Sant_3829 338 aa	OCBT	*SGGMMB4_05057* 1,014 bp	New annotation indicates full-length gene
argI	1,005 bp 334 aa	Sant_3829 338 aa	OCBT	*SGGMMB4_05057* 1,014 bp	New annotation indicates full-length gene
argG	1,344 bp 447 aa	Sant_2433 445 aa	ARGSS	*SGGMMB4_03589* (*argG_1*) *SGGMMB4_03590* (*argG_2*) 1,341 bp total	*argG_2* fragment is almost full length
argH	1,374 bp 457 aa	Sant_3955 457 aa	ARGSL	*SGGMMB4_05192* 1,371	New annotation indicates full-length gene

The new annotation suggests that the *S. glossinidius argG* gene (*SGGMMB4_03590*) has a fragment (*argG_2*) that is almost full length in comparison to its *S. praecaptivus* orthologue. This indicates that *argG* is also likely functional in spite of its description as a pseudogene in the previous annotation.

The *argA* gene appears in two fragments; *SGGMMB4_04654* (*argA_1*) and *SGGMMB4_04655* (*argA_2*). It has been demonstrated in *Pseudomonas aeruginosa* that the two separate ArgA protein domains can be expressed individually (Sancho-Vaello et al., 2012; Shi et al., 2015). The C-terminal, acetyltransferase domain can also function as a stand-alone protein when a high concentration of L-glutamate is provided (Sancho-Vaello et al., 2012). The *SGGMMB4_04655* (*argA_2*) fragment of this gene has a GTG start codon and may therefore be functional in the conditions described in Fig. 2.4E.

The *S. glossinidius argD* orthologue also appears in two pieces; *SGGMMB4_05438* (*argD_1*) and *SGGMMB4_05439* (*argD_2*). Lal et al. (2014) show that an *argD* mutant of *E. coli* can still show some *N*-acetylornithine aminotransferase activity, demonstrating that other proteins can compensate for a loss of this gene. The hypothesis that the loss of certain genes in the L-arginine biosynthesis pathway is not lethal was then assessed *in vivo*. *E. coli argA*, *argD* and *argG* knockouts from the Keio collection (Baba et al., 2006) were grown in M9 minimal media (Neidhardt

et al., 1974) alone or with the addition of L-arginine. The *argD* knockout mutant can grow in the absence of L-arginine (Fig. 2.7, time course in Fig S2), confirming that the loss of this gene can be compensated for by alternative proteins in *E. coli*. Indeed, candidate aminotransferase genes exist in *S. glossinidius*, including *bioA* (*SG0902*) or *hemL* (*SG0500*), that may provide functional alternatives.

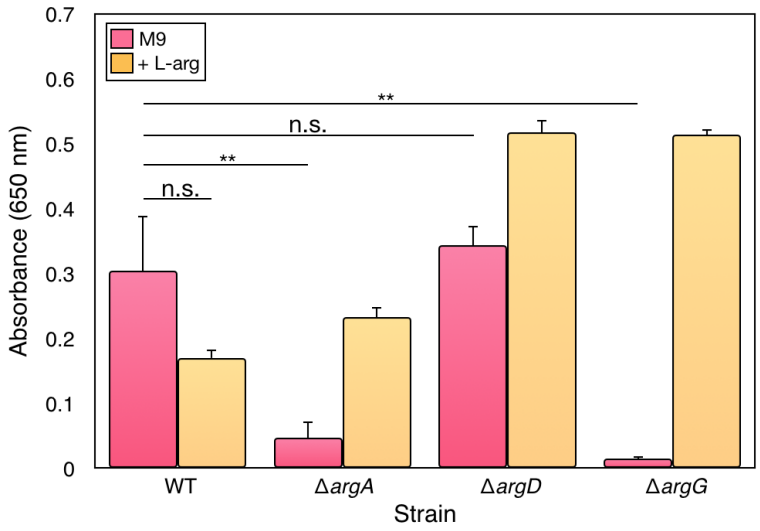


Figure 2.7: Using *E. coli* knockout mutants from the Keio collection to examine the essentiality of L-arginine biosynthesis genes. *E. coli argA*, *argD* and *argD* deletion mutants grown in M9 plus glucose (pink) with the addition of L-arginine (orange), in comparison to wild-type (WT) BW25113. *argA* and *argG* knockout mutants are not able to grow in the absence of exogenous L-arginine, whereas WT and the *argD* knockout mutant can. Measurements are end-point growth increase in triplicate, error bars SEM. **p<0.01, one-way ANOVA. Time course results found in Fig. S2.

The data here suggest that under certain conditions, *S. glossinidius* can synthesise L-arginine, surviving when it is not supplied exogenously. It is therefore not a true auxotroph and could instead be described as a relic of a prototroph transitioning to auxotrophy.

2.4 Discussion

Symbiotic bacteria are important components of medically significant microbiomes. The study of their physiology and metabolism is however limited frequently by culturability issues. Tools to improve this are therefore desirable. This work describes a new refined FBA model of the *S. glossinidius* metabolic network and demonstrates its application in designing defined growth media for the symbiont. The carbon source for *i*LF517 is GlcNAc, as opposed to D-glucose in *i*EB458. It has been veri-

fied empirically that *S. glossinidius* achieves a significantly better growth yield with GlcNAc in comparison to D-glucose (Fig. 2.4B). This is an important progression in the development of a metabolic model of *S. glossinidius*. The use of GlcNAc may be a result of both the pseudogenisation of the D-glucose-specific PTS transporter (Fig. 2.5) and the availability *in vivo* of this host-derived sugar. This inclusion of GlcNAc may also support the theory that *S. glossinidius* is connected to the persistence of the trypanosomes within the tsetse. GlcNAc can inhibit D-glucose uptake by procyclic trypanosomes, resulting in a metabolic switch to the more efficient oxidative phosphorylation with L-proline and a faster growth rate (Dyer et al., 2013; Ebikeme et al., 2008). The free GlcNAc may derive from a breakdown of the tsetse peritrophic membrane by a chitinase secreted by *S. glossinidius* (Welburn et al., 1993). The data here indicate that there may indeed be a link between the symbiont, the parasite and the availability of GlcNAc within the tsetse.

The experimental evidence demonstrates that *S. glossinidius* still requires a sugar for growth, even though it clearly relies on amino acids such as L-glutamate (Fig. 2.4D). This is important as other bacterial species have been shown to reduce their metabolic networks to grow on amino acids alone (Tomb et al., 1997; Parkhill et al., 2000; Schilling et al., 2002; Guccione et al., 2008; Leach et al., 1997). That this is not the case suggests an adaptation to use an abundant host-derived sugar, namely GlcNAc. Furthermore, this may have then allowed the loss of *ppc* to occur during the transition to a symbiotic lifestyle (Fig. 2.1). The result is a more constrained metabolic network that makes the organism less metabolically flexible than its free-living relative.

L-glutamate has been shown *in silico* and empirically to be an essential nutrient for *S. glossinidius* (Fig. 2.4D, Supp. File 1). It supplements the TCA cycle at α-ketoglutarate and forms L-arginine via ornithine (Fig. 2.1). An excess of this amino acid rescued the growth defect caused by the removal of L-arginine, thought previously to be an essential metabolite (Fig. 2.4E). The *argD* gene has become fragmented but all other genes thought previously to be pseudogenised appear functional in the new genome annotation (Fig. 2.1). *S. glossinidius* is therefore an L-arginine prototroph, not an auxotroph as thought previously, capable of growth in the absence of L-arginine when sufficient L-glutamate is available. Unusual amino acid biosynthesis pathways are not uncommon in symbiotic bacteria. Indeed, components of the L-arginine biosynthesis pathway can function differently in symbionts. One example is a potential fusion of ArgA and ArgG in *Sulcia muelleri*, symbiont of the sap-feeding sharpshooter *Homalodisca vitripennis* (McCutcheon and Moran, 2007). The experimental conditions used here aim to reflect the tsetse microenvironment; metabolite concentrations vary according the stage of the hunger cycle or the

tissue sampled but internal L-glutamate has been measured at 34 mM in the tsetse tissue (Bursell, 1963; Bursell and Slack, 1976). SGM11 could be considered 'low' L-glutamate at 18 mM, and therefore the 5x (85 mM) media subsequently removes any limitations caused by insufficient L-glutamate.

It may be that the L-arginine biosynthesis pathway is undergoing the process of inactivation and will become entirely pseudogenised over evolutionary time. This is supported by the complete pathway for L-arginine biosynthesis in *S. praecaptivus*, suggesting that these genes may have been lost within the tsetse environment as a result of selection pressure or drift. It implies that L-glutamate is not limiting inside the tsetse, allowing relaxed selection pressure on the L-arginine biosynthesis genes. It also emphasises strongly the importance of using *in vitro* experiments to test *in silico* assertions. This is particularly relevant in symbiotic bacteria where the functionality of broken or fragmented genes is not certain. Indeed, a recent report implies that some *S. glossinidius* genes thought to be pseudogenised are in fact under transcriptional control (Goodhead et al., 2018).

TMP has been described here as an essential external metabolite. *S. glossinidius* is reliant on an external source of thiamine, both *in silico* in *i*LF517 and experimentally in the form of TMP in SGM11 (Fig. 2.4C). *S. glossinidius* may use its intact transporter to obtain TMP *in vivo* from *W. glossinidia*, which has retained the ability to synthesise this vitamin (Toh et al., 2006; Rio et al., 2016). The results presented here provide the first clear experimental evidence of a potential metabolic linkage between the two important symbionts of the tsetse, *S. glossinidius* and *W. glossinidia*, and suggests that the TMP released by *W. glossinidia* is transported around the tsetse for use by both host cells and other symbionts.

*i*LF517 and SGM11 can now be used as a tool to predict with accuracy how *S. glossinidius* might respond to genetic manipulation. Using genomics to investigate and implement custom growth conditions is an area of research that is progressing rapidly, aided by advancements in gene sequencing and analysis. This includes the design of defined microbiological growth media (Blakemore et al., 1979; Rodriguez-Valera et al., 1980; Nautiyal, 1999), enabling metabolic and physiological investigations that would not be possible in complex or standard media (Jensen and Hammer, 1993; van de Rijn and Kessler, 1980; Adler and Templeton, 1967; De Man et al., 1960). This has implications in disease control, both for HAT and for other diseases where the insect vectors have characterised bacterial microbiomes (Rio et al., 2004; Azambuja et al., 2004). *Wolbachia*, for example, has been introduced successfully into the mosquito *Aedes aegypti*, with a notable reduction in infection by the pathogenic dengue and chikungunya viruses and the malaria parasite *Plasmodium*

(Moreira et al., 2009; Zabalou et al., 2004). *Wolbachia* is found naturally in a range of medically significant insects including *Phlebotomus chinensis* (visceral leishmaniasis) (Li et al., 2016) and *Aedes albopictus* (dengue, yellow fever, West Nile and chikungunya) (Blagrove et al., 2012; Zhou et al., 1998; Sinkins SP, 1995). While some tsetse populations do present with *Wolbachia* infection (Doudoumis et al., 2012, 2013; Alam et al., 2011; Cheng et al., 2000), the persistence of *S. glossinidius* and its colocalization with the parasitic *T. brucei* makes the latter an ideal candidate for novel disease control methods. It is hoped that the techniques described here may also translate to the microbiomes of other medically significant insects, including *Rhodococcus rhodnii* from the Chagas disease vector *Rhodnius prolixus* (Rodríguez et al., 2011), and *Acetobacteraceae* spp. and *Pseudomonadaceae* spp. in *Leishmania infantum*-infected sand flies (*Lutzomyia longipalpis*) (Kelly et al., 2017).

Use of *i*LF517 and the *S. glossinidius* metabolic network has enabled the design of a defined growth medium that supports growth of the symbiont. Whilst several FBA models for insect symbionts have been published (Thomas et al., 2009; Macdonald et al., 2011; Ankrah et al., 2017; González-Domenech et al., 2012), this study is the first example of using FBA to improve the *in vitro* culture of these organisms. SGM11 facilitated the discovery that *S. glossinidius* is not a true L-arginine auxotroph and demonstrates its reliance on exogenous sources of thiamine and L-glutamate. SGM11 will improve greatly the ability to test other aspects of *S. glossinidius* metabolism and growth kinetics that have until now been limited by the restrictions of rich media. The continued transition of *S. glossinidius* to a symbiotic lifestyle can now be predicted using this model. By comparing its dispensable, redundant genes to those in both free-living and symbiotic bacteria it is possible to assess the trajectory of this symbiosis.

Chapter 3

A genome scale metabolic model for *Sodalis praecaptivus* reveals unusual biosynthetic pathways

The following chapter, with Chapter 4, forms part of a submitted manuscript, available as a pre-print on bioRxiv (doi 10.1101/819946), as:

Hall RJ*, Thorpe S*, Thomas GH and Wood AJ. Simulating the evolutionary trajectories of metabolic pathways for insects symbionts in the *Sodalis* genus.

RJH constructed and tested the model. ST assisted with programming. AJW wrote the custom FBA solver. GHT and AJW supervised the project.

3.1 Introduction

Computational biology is now well established as a key tool of scientific discovery, now that vast amounts of data are generated quickly and cheaply from advancements in sequencing technology (Edwards et al., 2002; Kauffman et al., 2003). Genome scale metabolic modelling of microorganisms enables predictions to be made about metabolite preferences, transporter use, and the functionality of biosynthetic pathways (Edwards et al., 2002; Lewis et al., 2012). Microbial metabolism can be simulated using FBA, a constraint-based quantitative approach that reconstructs a metabolic network from a genome annotation (Kauffman et al., 2003; Orth et al., 2010). FBA is a powerful tool when based on a well annotated genome and with

the provision of *in vitro* experimental validation (Orth et al., 2010).

FBA is widely used for biotechnology applications, and this can be re-purposed to examine symbiosis. There are several examples in the literature of using FBA to analyse the metabolism of symbiotic bacteria, including for *Buchnera aphidicola* (Thomas et al., 2009; Macdonald et al., 2012, 2011), *Sodalis glossinidius* (Belda et al., 2012; Hall et al., 2019), *Portiera aleyrodidarum* (Ankrah et al., 2017), *Hamiltonella defensa*, (Ankrah et al., 2017) and strains of *Blattabacterium* (González-Domenech et al., 2012). There are also models published for the *Synechocystis* species used in the study of artificially induced symbiosis (Sørensen et al., 2016; Shastri and Morgan, 2005; Nogales et al., 2012; Knoop et al., 2010, 2013). FBA is useful in this instance, as experiments that would not be possible empirically, due to culturability issues, can be performed *in silico*. Furthermore, the genomes of symbiotic bacteria are often unusual, with large pathway deletions or widespread pseudogenisation (Shigenobu et al., 2000; Toh et al., 2006; Goodhead et al., 2018; Dale and Maudlin, 1999). Analysis of the resulting metabolic network via FBA can suggest whether these biosynthetic pathways are active, and predict which external metabolites might be required to support growth *in vitro*.

Studies of symbiont metabolism can be limited by the lack of a well-defined close relative with which to compare. Several bacterial symbionts, including *B. aphidicola* and *Wigglesworthia glossinidia*, are closely related to *Escherichia coli* (Shigenobu et al., 2000; Riley et al., 2006). The *E. coli* genome has been annotated extensively and a large proportion of its proteome has been characterized experimentally. Inferences from these symbiont genomes can therefore be made with reasonable confidence via direct comparison to *E. coli* orthologues (Thomas et al., 2009). Others, including *Uzinura diaspidicola* from Diaspididae scale insects (Sabree et al., 2013) and *Ishikawaella capsulata* from stinkbugs (Nikoh et al., 2011), do not have close relatives described to the same degree. The study of the relationship that these symbionts have with their host must therefore be conducted without access to a comparator organism.

A free-living organism within the *Sodalis* genus has been characterised and sequenced recently (Clayton et al., 2012; Chari et al., 2015). *Sodalis praecaptivus* was isolated from a human wound, caused by an impalement by a crab apple tree. It is assumed that the tree was the likely source of the *S. praecaptivus* infection. *S. praecaptivus* is a prototroph, capable of growth in minimal media and at 37°C (Chari et al., 2015). The annotated genome sequence for *S. praecaptivus* is also available (Clayton et al., 2012). This organism therefore provides a rich set of data from which to begin investigations into the origin of, and adaptations within, the tsetse-*S. glossinidius*

symbiosis.

It was shown in Chapter 2 that *S. glossinidius* has adapted metabolically to its microenvironment. It is important to improve upon current knowledge of the metabolism of its free-living relative, in order to understand how these adaptations have arisen. To this end, we present a metabolic model for *S. praecaptivus*. This model, *i*RH830, is a prototrophic model that can grow on rich and minimal media. Through the testing of *i*RH830, unusual pathways for the metabolism of xylitol and *N*-acetyl-D-galactosamine (GalNAc) were uncovered. This model can subsequently provide an ideal starting point from which to investigate the spectrum of possible evolutionary trajectories that *S. glossinidius* may have taken during its process of internalisation within its tsetse host.

3.2 Methods

3.2.1 Bacterial strains, growth conditions and reagents

S. praecaptivus was obtained from DMSZ. Working stocks were established by incubating starter cultures on LB (Sigma-Aldrich) (Table 3.1) agar plates overnight at 37°C. A single colony was then sub-cultured on to a fresh LB plate and incubated overnight at 37°C. A single colony was selected with a sterile pipette tip and used for downstream experimentation as per Biolog, Inc manufacturer protocol. Briefly, the colony was vortexed in IF-0 media before a redox dye was added (Biolog). Phenotypic microplates were used to screen for the ability of *S. praecaptivus* to grow on a range of carbon sources, using PM1 and PM2A microplates (Biolog). A 100 μL bacterial suspension in the relevant media was added per well. Optical density was measured at 590 nm and 730 nm in a microplate reader (Epoch), and incubated with double orbital shaking at 37°C for 24 hours.

Discrepancies between *in silico* and *in vitro* Biolog results were reexamined by establishing individual cultures of *S. praecaptivus* in M9 salts (Table 3.1) in 96-well microplates, and supplemented with the metabolite of interest at a range of concentrations from 25 mM to 50 μM. Cultures were incubated in a microplate reader with double orbital shaking at 37°C for 36 hours.

Table 3.1: Media used in Chapter 3. D-glucose not added to M9 if only the salts were required.

Medium	Recipe
LB	10 g NaCl (Fisher), 10 g tryptone (Melford), 5 g yeast extract (Oxoid) in 1 L dH_2O 15 g of agar powder (Oxoid) added to make agar plates
M9	6 g Na_2HPO_4, 3 g KH_2PO_4, 500 mg NaCl, 1 g NH_4Cl in 1 L dH_2O pH 7 using NaOH Autoclave, then once cool add: 2 mL 1M $MgSO_4$ (filter sterilised), 40 mL 500 mM D-glucose (filter sterilised)

3.2.2 Construction of the *S. praecaptivus* metabolic network

The annotated *S. praecaptivus* genome sequence, CP006569.1, was downloaded from NCBI in GenBank format. Genes in *S. praecaptivus* with the same annotation as genes in the *E. coli* str. K-12 substr. MG1655 genome (ASM584v2) were highlighted, and the reactions encoded by these genes extracted from the BiGG Models database (King et al., 2016). These processes were automated using custom scripts written in Python.

FBA models of *S. glossinidius* (*i*LF517, Chapter 2) and *E. coli* (*i*JO1366 (Orth et al., 2011; Orth and Palsson, 2012), *i*JR904 (Reed et al., 2003), *i*AF1260 (Feist et al., 2007)) were then used to aid the identification of missing reactions. The reactions and corresponding gene assignments in these published models were compared to the draft *S. praecaptivus* model. These gene assignments were then used to guide translated nucleotide and protein BLAST searches of the *S. praecaptivus* genome. KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2019) and EcoCyc (Keseler et al., 2017) databases were used to confirm the identity of the *E. coli* genes encoding each reaction. *S. glossinidius* gene assignments were taken from *i*LF517, detailed in Chapter 2. These orthologues in *E. coli* and *S. glossinidius*, with sequences taken from UniProt (Bateman et al., 2019), were used as BLAST search queries.

KEGG, BiGG Models, and MetaCyc (Caspi et al., 2007) were used to assign reaction stoichiometry. Candidate pseudogenes were aligned with known functional orthologues using ClustalX 2.1 (Larkin et al., 2007). Those with sequences missing or mutations in key residues were not included in the model. FBA and literature searches were used to identify and fill gaps in metabolic pathways appropriately (Thiele and Palsson, 2010).

3.2.3 Flux balance analysis

To use FBA to simulate the metabolism of an organism, each reaction in the metabolic network must first be represented mathematically. Each metabolite may be found in either one or multiple reactions. The direction of the flux (I) of these metabolites within the reactions is defined, and may be reversible or irreversible.

The objective function (Z) is then identified (Thiele and Palsson, 2010). Here, this is the production of biomass. A biomass reaction is then defined. This model uses the *E. coli* biomass reaction described by, for example, Reed et al. (2003), Feist et al. (2007), and Orth et al. (2011).

The fluxes through the exchange reactions are then constrained, and upper and lower limits are set for all other reactions (Thiele and Palsson, 2010). This limits what can ultimately move in and out of the network. This can be depicted as:

$$I^{\min} \leq I \leq I^{\max} \tag{3.1}$$

The flux through the biomass reaction is then maximised, an additional constraint (ZI), and the stoichiometric matrix defined (S) (Thiele and Palsson, 2010; Orth et al., 2010). As this is a steady state model, the total flux through the network must be equal to zero:

$$max(ZI) \tag{3.2}$$

$$SI = 0 \tag{3.3}$$

This can then be solved using linear programming.

Here, the FBA solutions were generated using the GNA linear programming kit (GLPK) integrated with custom software in Java (written by AJW).

3.2.4 Reaction constraints

Each reaction was given upper and lower bounds to specify the maximum and minimum fluxes allowable through the reaction (Thiele and Palsson, 2010; Orth et al.,

2010). Oxygen uptake was constrained to 20 mmol gr DW^{-1} hr^{-1}, comparable to other models of free-living Gram negative bacteria (Orth et al., 2011; Orth and Palsson, 2012; Reed et al., 2003; Feist et al., 2007). The uptake of ammonia, water, phosphate, sulphate, potassium, sodium, calcium, carbon dioxide, protons, and essential transition metals was unconstrained for all conditions. Cofactor constraints were implemented by introducing these metabolites to the biomass function at small fluxes (1 x 10^{-5} mmol gr DW^{-1} hr^{-1}) (Thomas et al., 2009). *i*RH830 was supplied with either a nutrient limited medium of 6 mmol gr DW^{-1} hr^{-1} GlcNAc, 1 mmol gr DW^{-1} hr^{-1} thiamine, and selected cofactors (henceforth "famine"), or a tsetse-specific blood media (henceforth "blood", full recipe detailed in Table 3.2 and adapted from information in Stein and Moore (1954)). The phenotype was considered viable if the biomass production rate was greater than 1 x 10^{-4} gr DW $(mmol\ glucose)^{-1}$ hr^{-1}.

Table 3.2: *In silico* blood medium for *i*RH830, showing the maximum permitted and actual flux values for each metabolite (mmol gr DW^{-1} hr^{-1}).

Metabolite	Name	Flux (permitted)	Flux (actual)
acgam	GlcNAc	6	6
ala_L	L-alanine	3.41	3.41
arg_L	L-arginine	1.54	1.14
asn_L	L-asparagine	0.58	0.58
asp_L	L-aspartate	0.03	0.03
cys_L	L-cysteine	1.18	0
gln_L	L-glutamine	8.3	0
glu_L	L-glutamate	0.7	0.7
gly	Glycine	1.54	1.54
his_L	L-histidine	1.15	0.11
ile_L	L-isoleucine	0.89	0.33
leu_L	L-leucine	1.69	0.52
lys_L	L-lysine	2.72	0.39
met_L	L-methionine	0.38	0.18
orn	Ornithine	0.72	0
phe_L	L-phenylalanine	0.84	0.21
pro_L	L-proline	2.36	2.36
ser_L	L-serine	1.12	1.12
taur	Taurine	0.55	0
thm	Thiamine	1	0.00001
thr_L	L-threonine	1.39	0
trp_L	L-tryptophan	1.11	0.07
tyr_L	L-tyrosine	1.03	0.16
val_L	L-valine	2.88	0.49

To investigate the concordance between the *in vitro* screen and the *in silico* outputs, *i*RH830 was, where possible, supplemented with the carbon sources analysed at an exogenous concentration of 6 mmol gr DW^{-1} hr^{-1}. A qualitative presence/absence of

a positive biomass output was noted. Full description of the model is provided in Supp. File 2.

3.2.5 Solving futile cycles

Futile cycles, closed loops of a number of reactions, were detected by the presence of unsustainably large fluxes. Futile cycles often occur when several reversible reactions are present in which the product (M) of one becomes the substrate of another (Equations 3.4, 3.5, 3.6). These reactions were examined individually, and solved by adjusting the reversability with guidance from EcoCyc (Keseler et al., 2017) and BiGG Models (King et al., 2016).

$$M_1 \rightleftharpoons M_2 \tag{3.4}$$

$$M_2 \rightleftharpoons M_3 \tag{3.5}$$

$$M_3 \rightleftharpoons M_1 \tag{3.6}$$

3.2.6 Robustness analysis

Robustness analysis of the *i*RH830 network was executed using COBRApy (Ebrahim et al., 2013) to conduct single reaction deletions. *i*RH830 was supplied with 6 mmol gr DW^{-1} hr^{-1} glucose under aerated conditions. The flux through reactions was set to zero individually and the resulting effect on biomass output measured. Reactions were categorised as essential if the resulting biomass output was less than 1 x 10^{-3} gr DW (mmol glucose)$^{-1}$ hr^{-1}.

3.3 Results

3.3.1 Construction of a metabolic model for *S. praecaptivus*

A metabolic model can provide intricate detail about the metabolism of an organism. A model for *S. praecaptivus* was therefore constructed in order to elucidate key differences between this organism and the symbiotic *S. glossinidius*. A summary of construction process can be found in Fig. 3.1. First, the metabolic genes were extracted from the *S. praecaptivus* genome annotation and compared to known proteins in *E. coli* K-12. This process was repeated using the *S. glossinidius* genome. Reactions were collected from the BiGG Models database (King et al., 2016) and compiled into a draft model. The draft model was then compared to FBA models of *S. glossinidius* (*i*LF517, Chapter 2) and *E. coli* (*i*JO1366 (Orth et al., 2011; Orth and Palsson, 2012), *i*JR904 (Reed et al., 2003), *i*AF1260 (Feist et al., 2007)). Where reactions were found in the pre-existing models but not in the draft model, the *S. praecaptivus* genome was mined manually to assess whether that gene could be found. Key pathways in central metabolism and for the biosynthesis of amino acids were checked manually.

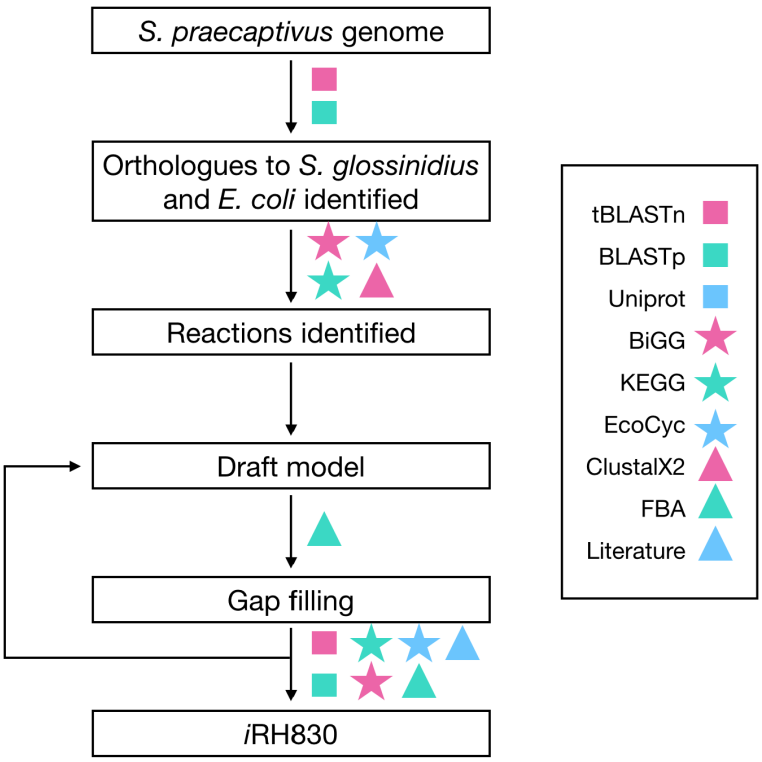


Figure 3.1: Overview of the construction process for *i*RH830.

3.3.2 A large scale, iterative process of model testing

Metabolic models are strengthened by thorough *in vitro* testing. A large scale phenotypic screen of metabolite use by *S. praecaptivus* was therefore conducted using Biolog phenotypic microplates in order to confirm the predictions made by the draft model. These plates are a qualitative measure of the ability of an organism to use a metabolite as a sole source of carbon. The presence ("positive") or absence ("negative") of visible growth above that of the manufacturer's reference was recorded. *S. praecaptivus* was able to use 19 of the 128 metabolites tested as a sole source of carbon (Table 3.3). Four metabolites produced borderline results and were therefore considered inconclusive as per the manufacturer's instructions (D-ribose, L-rhamnose, mucic acid and D-glucuronic acid). *S. praecaptivus* was unable to grow on 105 of the metabolites tested.

Table 3.3: Positive results from a phenotypic screen of individual metabolites. *i*RH830, *i*JO1366, and *i*LF517 Δbiomass outputs (gr DW (mmol glucose)$^{-1}$ hr^{-1}) following the addition of these metabolites to models currently not supplied with a sugar. Metabolites in **bold** initially did not produce a positive biomass output in early iterations of the model.

Metabolite	*i*RH830	*i*JO1366	*i*LF517
α-D-glucose	0.557	0.584	0.211
D-fructose	0.557	0.584	0.211
D-galactose	0.550	0.578	0.131
D-galacturonic acid	0.440	0.470	0
D-gluconic acid	0.507	0.533	0.158
D-glucosamine	0.558	0.584	0.211
D-mannitol	0.600	0.628	0.193
D-mannose	0.557	0.584	0.211
D-melibiose	1.050	1.175	0
D-sorbitol	0.600	0.628	0
D-trehalose	1.065	1.181	0.413
D-xylose	0.430	0.478	0
Glycerol	0.188	0.332	0
L-arabinose	0.430	0.478	0
***N*-acetyl D-galactosamine**	0.711	0	0
N-acetyl D-glucosamine	0.717	0.752	0.211
N-acetyl D-mannosamine	0.717	0.752	0.211
Pyruvic acid	0.181	0.211	0
Xylitol	0.502	0	0

The results of the phenotypic screen were then compared to the biomass outputs generated by a draft of the *S. praecaptivus* model. The individual metabolites were added to the draft model (including essential cofactors) at a flux of 6 mmol gr DW^{-1}

hr^{-1}, and the resulting biomass output recorded. Overall, there was 83.9% agreement between the *in vitro* screen and the biomass outputs of the draft model. Of the 19 metabolites that produced a positive growth phenotype *in vitro*, 17 resulted in a matched phenotype *in silico* (Table 3.3). The remaining two, *N*-acetyl-D-galactosamine (GalNAc) and xylitol will be discussed in more depth in Sections 3.3.3 and 3.3.4. There were 87 metabolites that produced a negative growth phenotype both *in silico* and *in vitro* (Table 3.4).

Table 3.4: All metabolites that produced a negative *in vitro* phenotype as well as a negative biomass output *in silico*.

Metabolite			
1,2-propanediol	2-aminoethanol	2-deoxyadenosine	2-hydroxy benzoic acid
4-hydroxy benzoic acid	Acetamide	Acetic acid	Acetoacetic acid
Adenosine	Adonitol	α-hydroxy butyric acid	α-keto-glutaric acid
α-methyl-D-glucoside	Arbutin	β-D-allose	β-methyl-D-glucoside
β-methyl-D-glucuronic acid	Butyric acid	Capric acid	Caproic acid
Chondroitin sulfate C	Citraconic acid	Citramalic acid	D-arabinose
D-arabitol	D-aspartic acid	D-cellobiose	D-fructose-6-phosphate
D-fucose	D-glucose-1-phosphate	D-glucose-6-phosphate	D-malic acid
D-raffinose	D-serine	D-tagatose	D-tartaric acid
D,L-carnitine	D,L-malic acid	Dextrin	Dihydroxyacetone
Dulcitol	Formic acid	Glycogen	Glycyl-L-proline
Inosine	Itaconic acid	L-arabitol	L-glutamine
L-histidine	L-homoserine	L-isoleucine	L-lactic acid
L-leucine	L-lysine	L-lyxose	L-methionine
Ornithine	L-phenylalanine	L-sorbose	L-tartaric acid
L-threonine	L-valine	m-hydroxy phenyl acetic acid	M-inositol
M-tartaric acid	Malonic acid	Maltose	Maltotriose
Mannan	Mono methyl succinate	*N*-acetyl-L-glutamic acid	*N*-acetyl-neuraminic acid
Oxalic acid	p-hydroxy phenyl acetic acid	Palatinose	Pectin
Phenylethylamine	Propionic acid	Quinic acid	Salicin
Sebacic acid	Stachyose	Sucrose	Thymidine
Tricarballylic acid	Tyramine	Uridine	

The 18 metabolites that produced a positive biomass output and a negative *in vitro* phenotype were checked by growing individually in universals with M9 plus the added metabolite. All produced no growth in *S. praecaptivus*, confirming the Biolog results (data not shown). Of these 18 metabolites, four were TCA cycle intermediates; citric acid, fumaric acid, malic acid, and succinic acid (Table 3.5). There are five transporters in the draft model encoded by *dcuB* (Sant_2332). Using Ecocyc as a guide, DcuB appears to function primarily during anaerobic respiration. *S. praecaptivus* is grown aerobically. The transporters encoded by *dcuB* were therefore removed (SUCFUMt, ASPt2_3, FUMt2_3, MALt2_3, SUCCt2_3), reducing the biomass output to zero on these TCA cycle intermediates. There are also nine amino acids in this group; D-alanine, glycine, L-alanine, L-arginine, L-asparagine, L-aspartic acid, L-glutamic acid, L-proline, and L-serine. The draft model is able to produce a viable biomass output with the addition of these amino acids due to the presence of functional transporters, but *in vitro* they may not provide sufficient energy to enable growth or the cells may not be able to import the metabolites at a high

enough concentration. The glycolaldehyde dehydrogenase reaction (Sant_3710) was made irreversible according to the data found in the BiGG Models database. This prevented a positive output from being produced on glycolic acid alone, reflecting the Biolog data. *S. praecaptivus* was also unable to grow on putrescine or D-saccharic acid. There is a functional putrescine transporter as well as a complete pathway to the TCA cycle, and a complete pathway from D-saccharic acid transport to pyruvate. These reactions were retained for completeness as there is no evidence from the model or genome to support their removal.

Table 3.5: Metabolites that produced a negative result on the *in vitro* phenotypic screen but a positive biomass output (gr DW (mmol glucose)$^{-1}$ hr^{-1}) *in silico*.

Metabolite	Biomass output	Reasoning
Citric acid	0.373	Anaerobic DcuB transporter *in silico*
Fumaric acid	0.226	Anaerobic DcuB transporter *in silico*
L-malic acid	0.226	Anaerobic DcuB transporter *in silico*
Succinic acid	0.243	Anaerobic DcuB transporter *in silico*
D-alanine	0.223	Possibly supply insufficient energy *in vitro*
Glycine	0.096	Possibly supply insufficient energy *in vitro*
L-alanine	0.223	Possibly supply insufficient energy *in vitro*
L-arginine	0.415	Possibly supply insufficient energy *in vitro*
L-asparagine	0.243	Possibly supply insufficient energy *in vitro*
L-aspartic acid	0.242	Possibly supply insufficient energy *in vitro*
L-glutamic acid	0.381	Possibly supply insufficient energy *in vitro*
L-proline	0.435	Possibly supply insufficient energy *in vitro*
L-serine	0.184	Possibly supply insufficient energy *in vitro*
α-D-lactose	1.050	Unknown, all pathways complete
D-saccharic acid	0.376	Unknown, all pathways complete
Glycolic acid	0.092	Incorrect reaction stoichiometry
L-fucose	0.533	Unknown, all pathways complete
Putrescine	0.423	Unknown, all pathways complete

3.3.3 *S. praecaptivus* can grow on the unusual sugar alcohol xylitol

Through the phenotypic screen, it was found that *S. praecaptivus* was able to grow on the sugar alcohol xylitol (Table 3.3), but xylitol was not initially included as a metabolite in the draft model. Xylitol is not found in any prokaryotic model in the BiGG Models database, hence it had not been considered for inclusion during construction.

To investigate the use of xylitol as a sole carbon source by *S. praecaptivus* further, cultures were established in a 96-well microplate with xylitol supplemented into M9

minimal media at concentrations of 12.5 mM, 5 mM, and 500 μM. *S. praecaptivus* was able to use xylitol equally well at 12.5 mM and 5 mM concentration (an optical density of approximately 0.27 at 650 nm (OD_{650})), with a small amount of growth observed at 500 μM xylitol (a final OD_{650} of approximately 0.14) (Fig. 3.2). There was no growth observed in the M9 salts with no added xylitol.

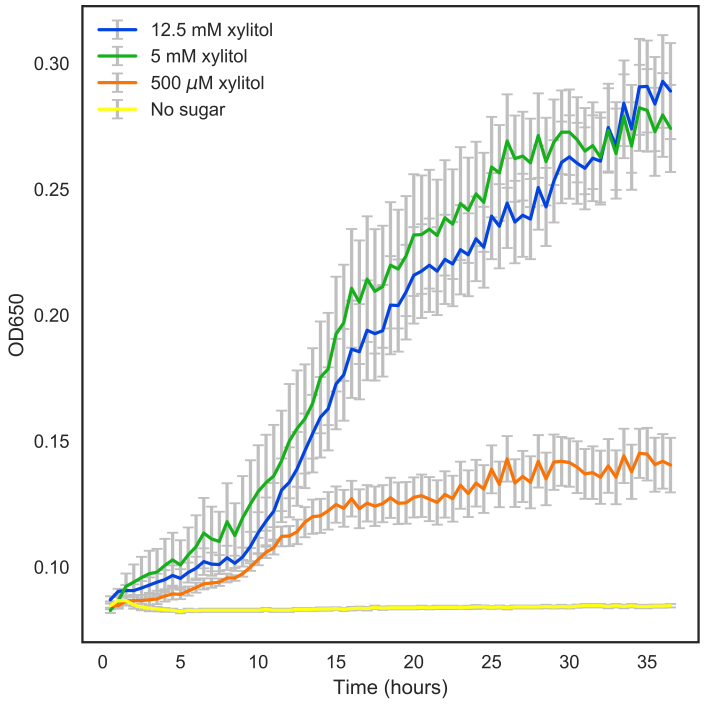


Figure 3.2: Growth of *S. praecaptivus* on 12.5 mM (blue), 5 mM (green) and 500 μM (orange) xylitol, and in M9 salts with no added sugar (yellow) over 36 hours. Measurements in sextuplicate, error bars SEM.

Due to the lack of pathways for xylitol degradation in the BiGG Models depository, alternative organisms known to degrade this metabolite were identified and used as a reference. KEGG was used to identify organisms with xylitol degradation pathways. The sequence for a xylitol dehydrogenase from *Morganella morganii* (MU9_3130) was extracted from UniProt and used in a protein BLAST search against *S. praecaptivus*. From this, a hypothetical xylitol dehydrogenase was identified (NCBI reference AFW03778) with 79.7% sequence identify to MU9_3130 (Fig. 3.3). This reaction converts xylitol into D-xylulose. The reaction to convert D-xylulose into D-xylulose 5-phosphate, leading to the pentose phosphate pathway, was already included in the draft model (Sant_3756). Orthologues to components of the xylitol transporter XltABC were also identified in *S. praecaptivus* (Sant_3104-6), using the periplasmic binding protein L580_2330 from *Serratia fonticola* as a search query.

This discovery enabled the inclusion of two new reactions into the draft model, XYLTD (xylitol dehydrogenase) and XYLTt (putative xylitol transporter) (Fig. 3.3).

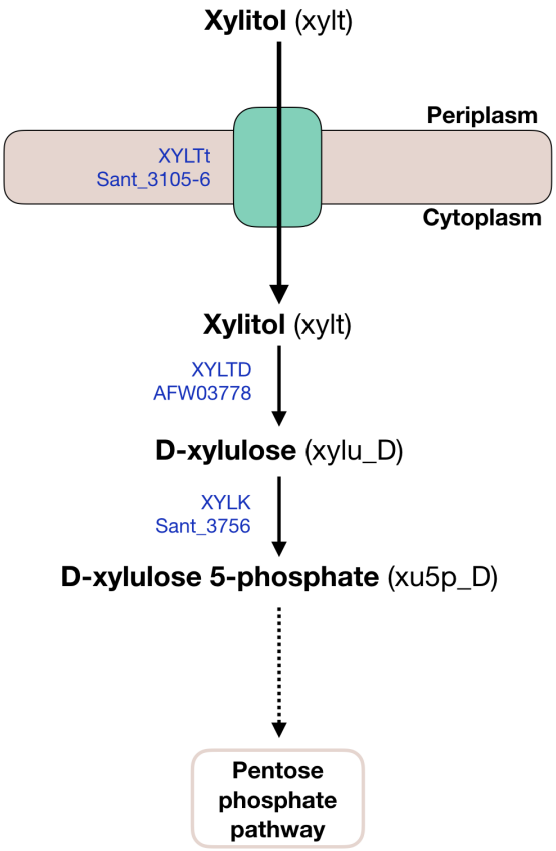


Figure 3.3: The putative xylitol degradation pathway in *S. praecaptivus*. Reaction names and *S. praecaptivus* gene assignments are given in blue.

3.3.4 Use of *N*-acetyl D-galactosamine by *S. praecaptivus*

The *in vitro* screen found that *S. praecaptivus* was able to use the amino acid derivative of galactose, GalNAc, as a sole source of carbon (Table 3.3). Several key species of free-living bacteria are not able to use GalNAc as an energy source, including *E. coli* K-12, *Klebsiella oxytoca*, *Klebsiella pneumoniae*, and species of *Salmonella* (Brinkkotter et al., 2000). This is as a result of not encoding functional phosphotransferase systems (PTS) to import the metabolite. To confirm the phenotypic screen quantitatively, *S. praecaptivus* was grown in a microplate reader for 36 hours in M9 salts alone and with the addition of 25 mM GalNAc. *S. praecaptivus*, as before, cannot grow in M9 salts alone, but reaches an OD_{650} of approximately 0.39 after 36 hours when GalNAc is added as the sole carbon source (Fig. 3.5). This confirms the qualitative observations of the large-scale screen.

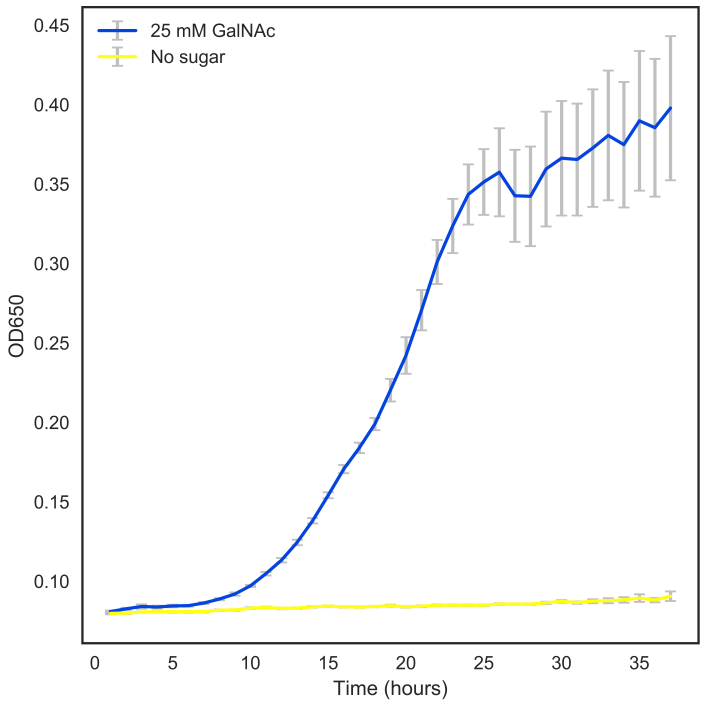


Figure 3.4: Growth of *S. praecaptivus* on 25 mM GalNAc (blue) and in M9 salts with no added sugar (yellow) over 36 hours. Measurements in sextuplicate, error bars SEM.

The ability of *S. praecaptivus* to metabolise GalNAc for use as a sole carbon source had been confirmed, but the draft model did not match this *in vitro* phenotype. Manual gap filling of a putative pathway from GalNAc to glyceraldehyde-3-phosphate (for the pentose phosphate pathway) was conducted using KEGG as a guide. Non-specific BLAST searches were conducted against the *S. praecaptivus* genome using sequences from the Enterobacteriaceae as queries. This identified a putative GalNAc PTS transporter (Sant_0021-24), a GalNAc-6-phosphate deacetylase (Sant_2792), a D-galactosamine-6-phosphate deaminase (Sant_2791), a kinase that converts D-tagatose-6-phosphate to D-tagatose-1,6-bisphosphate (Sant_1415), and an aldolase to then form glyceraldehyde 3-phosphate (Sant_2311, Sant_3725) (Fig. 3.5). The inclusion of these reactions enabled the draft model to produce a viable biomass output upon the addition of GalNAc.

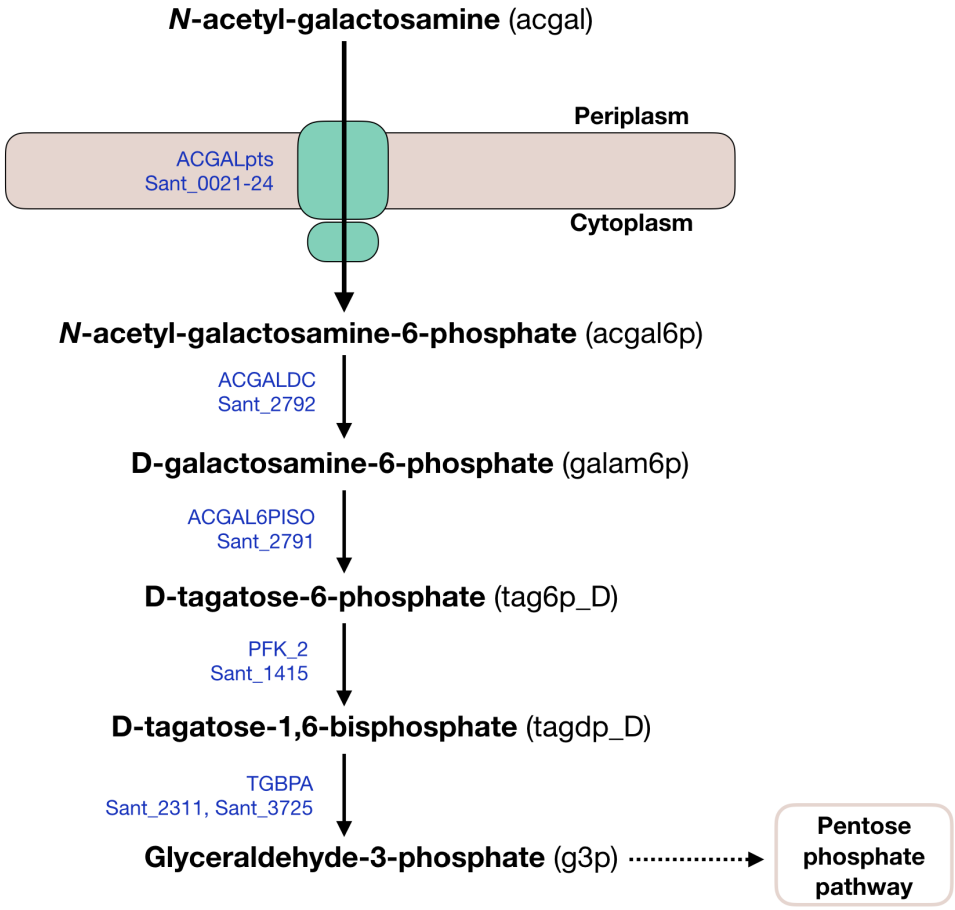


Figure 3.5: The putative GalNAc degradation pathway in *S. praecaptivus*. Reaction names and *S. praecaptivus* gene assignments are given in blue.

3.3.5 A new model of *S. praecaptivus* metabolism, *i*RH830

After thorough testing, a final model for *S. praecaptivus* is presented, *i*RH830. Full details are given in Supp. File 2. *i*RH830 contains 830 genes, 891 metabolites and 1248 reactions (excluding pseudoreactions). *i*RH830 uses an oxygen uptake value of 20 mmol gr DW^{-1} hr^{-1}, reflecting the highly aerated conditions the organism is grown in (Andersen and von Meyenburg, 1980; Varma et al., 1993) and to maintain consistency with models of *E. coli* metabolism (Orth et al., 2011; Orth and Palsson, 2012; Reed et al., 2003; Feist et al., 2007). This model is prototrophic for all essential amino acids and produces a range of biomass outputs when supplied with different carbon sources.

The performance of *i*RH830 was tested on gradient of oxygen concentrations. The model was supplied with either famine (GlcNAc and thiamine), blood (Table 3.2), or D-glucose-only media. The highest biomass output is observed in blood, followed by famine then D-glucose (Fig. 3.6). In D-glucose, the biomass output starts to

plateau at 13 mmol gr DW^{-1} hr^{-1} of oxygen. In famine and blood, it reaches the plateau at 18 and 40 mmol gr DW^{-1} hr^{-1} of oxygen, respectively. This represents the level at which oxygen is no longer the limiting factor in biomass production and when the nutrient availability is limiting instead. It is therefore higher in richer media.

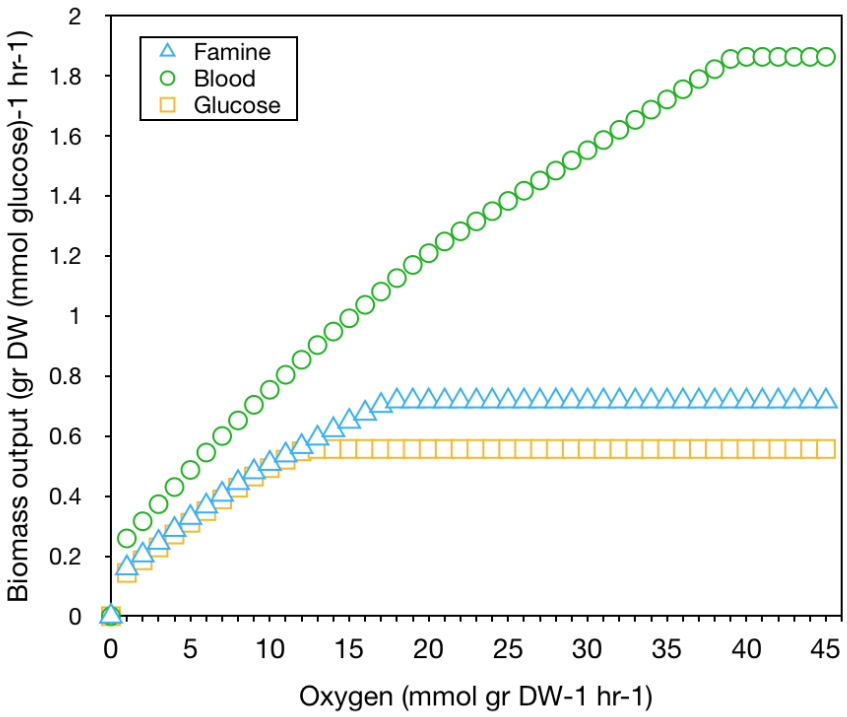


Figure 3.6: Biomass output for a range of oxygen concentrations when *i*RH830 was supplied with famine (blue triangle), blood (green circle), and D-glucose-only (orange square) media.

The xylitol and GalNAc degradation pathways included in *i*RH830 are unusual. It is not known whether these pathways are absent in other organisms, or omitted as a result of not testing for them. Models of *E. coli* (*i*JO1366 (Orth et al., 2011)) and *S. glossinidius* (*i*LF517, Chapter 2) metabolism were therefore compared to *i*RH830 for their ability to produce a viable biomass output on the carbon sources that tested positive with the *in vitro* screen. The change (Δ) in biomass output is given as *i*LF517 cannot produce a viable biomass output on these sugars alone (Hall et al., 2019). *i*RH830 produces a positive biomass output on all of these these carbon sources, as expected from the construction process (Table 3.3). *i*JO1366 is able to produce a viable biomass output on all carbon sources tested, with the exception of xylitol and GalNAc. This highlights the unusual nature of these metabolites in *i*RH830. As expected, there are several metabolites that do not support *i*LF517 *in silico*, due to the constraints of the smaller genome. It is unable to use D-galacturonic

acid, D-melibiose, D-sorbitol, D-xylose, glycerol, L-arabinose, or pyruvic acid. As with *i*JO1366, it is also unable to use xylitol or GalNAc.

To investigate whether *E. coli* and *S. glossinidius* also have the genomic capacity to metabolise xylitol and GalNAc, the *E. coli* str. K-12 substr. MG1655 and *S. glossinidius* str. *morsitans* genomes were mined using translated nucleotide BLAST searches with the *S. praecaptivus* proteins as queries. The *E. coli* genome does not contain xylitol or GalANc degradation genes orthologous to *S. praecaptivus* (Table 3.6). There are candidate genes for all except Sant_3725, but most of these are of low similarity. Two show a higher percentage identity, a deaminase (82.61%) and a kinase (86.22%). It is likely that these are involved in the transport and metabolism of other, similar carbon sources like GlcNAc. *S. glossinidius* appears to have retained all components of the xylitol transporter, xylitol dehydrogenase and xylulose kinase found in *S. praecaptivus* (Table 3.6). *S. glossinidius* contains many components for GalNAc transport and degradation, but several (SGGMMB4_05687, PTS subunit IIB, SGGMMB4_00446-7) are heavily truncated. There is also no orthologue for Sant_0021, with the top tBLASTn hit for this *S. praecaptivus* protein being the Sant_0022 orthologue (SGGMMB4_05685).

Table 3.6: Candidate components of the xylitol and GalNAc degradation pathways in *E. coli* and *S. glossinidius* based on tBLASTn searches using *S. praecaptivus* proteins. Percentage identity to *S. praecaptivus* sequences, predicted number of amino acids (aa), locus tags (SGGMMB4...) and NCBI descriptions (*italics*). Top tBLASTn results are given.

S. praecaptivus	*E. coli* K-12 substr. MG1655	*S. glossinidius*
Sant_3104	*RbsB*	SGGMMB4_01390
333 aa, SBP	30.69%, 281 aa	97.90%, 333 aa
Sant_3105	*D-ribose transporter*	SGGMMB4_01389
345 aa, permease	41.64%, 308 aa	92.17%, 345 aa
Sant_3106	*ATP binding protein*	SGGMMB4_01388
501 aa, ATP-binding	44.02%, 491 aa	96.81%, 501 aa
Sant_3107	*Xylulose kinase*	*SGMMB4_01387*
504 aa, xylulose kinase	29.70%, 483 aa	89.29%, 504 aa
AFW03778	*Alcohol dehydrogenase*	SGGMMB4_01386
344 aa, putative dehydrogenase	38.01%, 317 aa	92.73%, 344 aa
Sant_3756	*Xylulokinase*	*Xylulose kinase*
484 aa	62.86%, 482 aa	27.27%, 417 aa
Sant_0021	*PTS subunit IIA*	SGGMMB4_05685
144 aa	34.56%, 136 aa	76.24%, 101 aa
Sant_0022	*PTS subunit IIC*	SGGMMB4_05685
297 aa	45.80%, 261 aa	78.55% 295 aa
Sant_0023	*PTS subunit IIC component 1*	SGGMMB4_05687
255 aa	33.50%, 200 aa	53.85%, 86 aa
Sant_0024	*PTS subunit IIB*	*PTS subunit IIB*
160 aa	57.69%, 156 aa	97.20%, 107 aa
Sant_2792	*Deacetylase*	SGGMMB4_01931
380 aa	69.21%, 379 aa	92.37%, 380 aa
Sant_2791	*Deaminase*	*Deaminase*
267 aa	82.51%, 263 aa	95.51%, 267 aa
Sant_1415	*Kinase*	SGGMMB4_02147
312 aa	86.22%, 312 aa	90.12%, 324 aa
Sant_2311	*Aldolase*	SGGMMB4_00446-7
284 aa	63.38%, 284 aa	64.78%, 156 aa
Sant_3725 420 aa	*No significant similarity*	*No significant similarity*

3.3.6 Robustness analysis of the *S. praecaptivus* metabolic network

To test the ability of *S. praecaptivus* to grow on a range of metabolites, *i*RH830 was run on a tsetse-specific nutrient limited medium ("famine") and a blood medium simulating the internal tsetse environment and informed by *S. glossinidius* requirements (Hall et al., 2019) ("blood", Table 3.2). Robustness analysis was used to examine reaction essentiality and therefore redundancy in the *i*RH830 network. Reactions were removed individually and the resulting effect on biomass output noted. The same analysis was also run on *i*LF517 in blood as a comparison.

There are 282 essential reactions in *i*RH830 when the medium (famine) is nutritionally limited, and 228 in the tsetse-specific blood medium (Fig. 3.7). The overall

pattern for the two conditions is very similar. The subsystem most represented in either condition is for cofactor and prosthetic group biosynthesis, with 88 and 87 essential reactions for the famine and blood media, respectively. The main difference between blood and famine at the subsystem level can be attributed to amino acid metabolism. Of the 228 reactions essential in blood, 15.8% are involved in amino acid metabolism. In famine, which is not supplied with amino acids, 48 more essential reactions and a greater proportion of the total (29.8%) are used for amino acid metabolism. There are also three more essential transport reactions when the medium is limited.

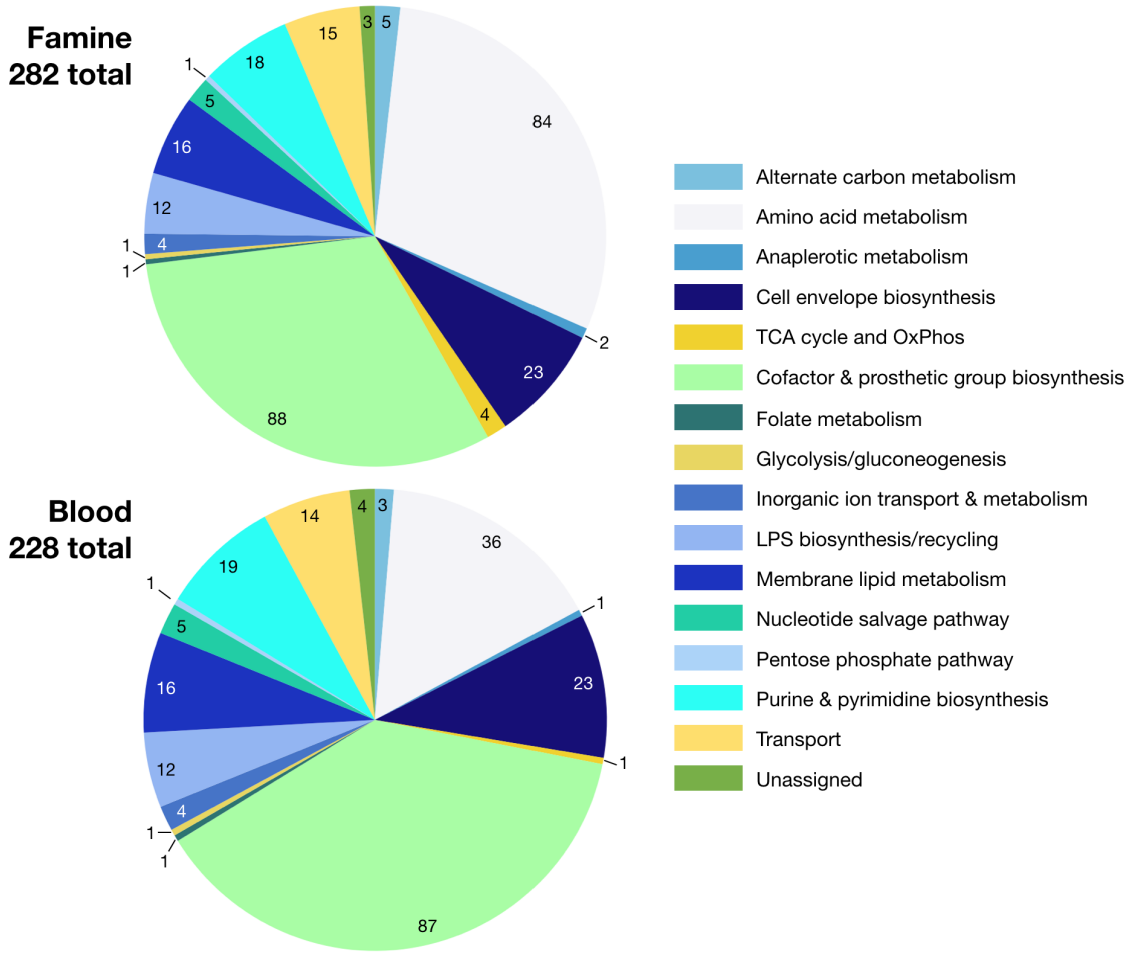


Figure 3.7: Robustness analysis of *i*RH830 in famine **(top)** and blood **(bottom)** media. Essential reactions are categorised by subsystem.

The reactions involved in amino acid metabolism were then analysed further to ascertain whether the two conditions differ in composition as well as size. Of the 84 essential reactions in the famine medium, the highest number (17) are involved in the metabolism of L-arginine and L-proline, followed by L-threonine and L-lysine (16) (Fig. 3.8, top). There are 13 essential reaction for the metabolism of L-valine, L-leucine and L-isoleucine. There are two subsystems with the fewest essential

reactions; L-glutamate (one), and glycine and L-serine metabolism (one). A single reaction in *i*RH830 (GHMT2r, glycine hydroxymethyltransferase) converts glycine to L-serine via a reversible reaction.

There are fewer essential reactions for amino acid metabolism when *i*RH830 is supplied with the blood medium, and the pattern of essentiality differs between the two media (Fig. 3.8, bottom). In blood, the two subsystems with the highest proportions mirror that seen in famine; L-threonine and L-lysine (eight), L-arginine and L-proline (seven). The pathways for L-tyrosine, L-tryptophan and L-phenylalanine also have seven essential reactions. The main difference is that there are no essential reactions involved in the metabolism of L-valine, L-leucine and L-isoleucine when *i*RH830 is supplied with a blood medium, compared to the 13 counted in famine. There is also only one reaction essential for L-histidine metabolism in blood, whereas there are 10 under the famine conditions.

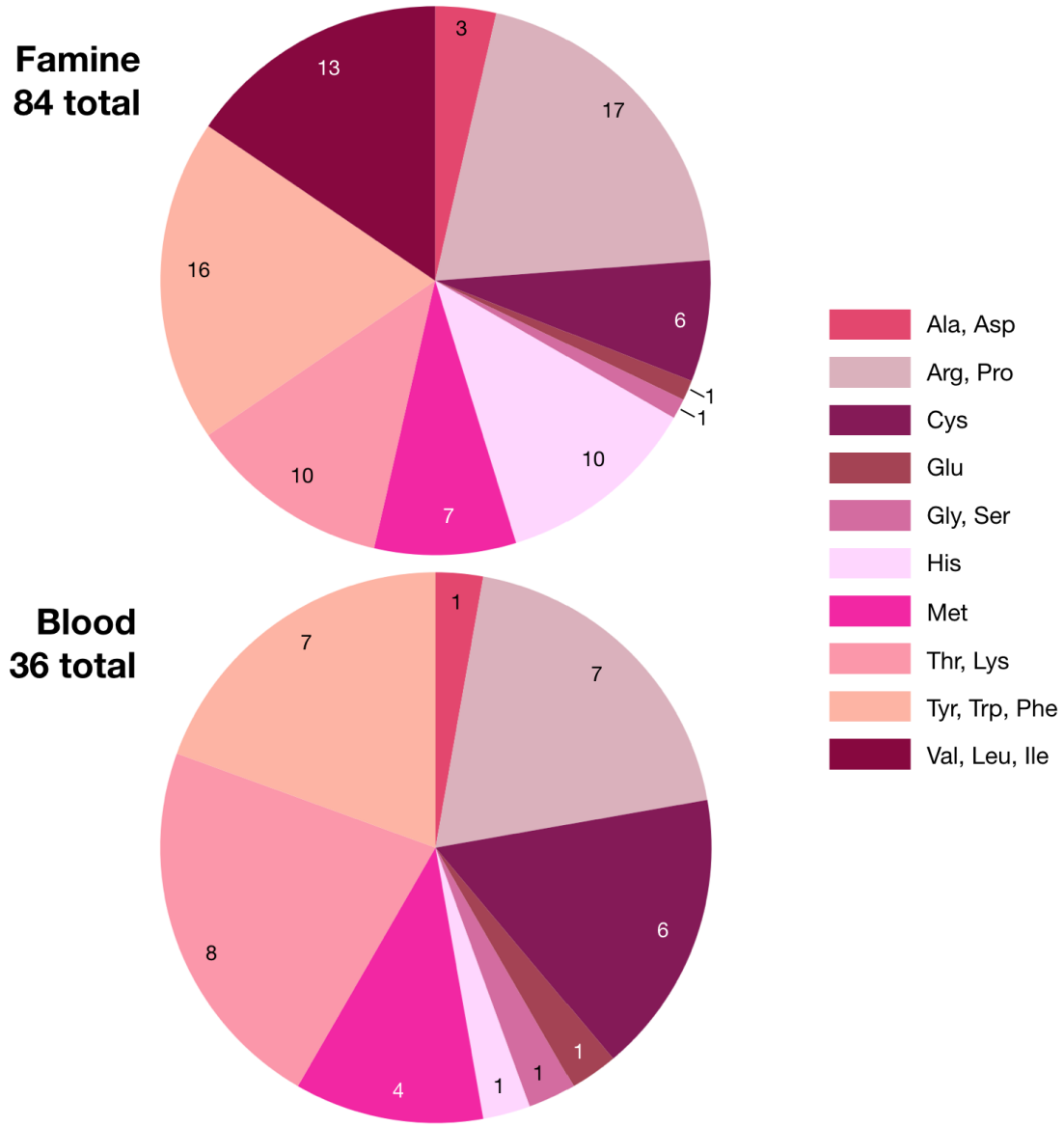


Figure 3.8: Essential reactions involved in amino acid metabolism in *i*RH830 in famine **(top)** and blood **(bottom)** media.

Robustness analysis was then run for *i*LF517 in the blood medium in order to compare to *i*RH830. *i*LF517 is not viable in famine. The overall pattern of reaction essentiality is highly similar for *i*LF517 and *i*RH830 (Fig. 3.9, top). For *i*LF517, there are 253 total essential reactions, 25 more than for *i*RH830 under the same conditions, highlighting the has greater redundancy in the *i*RH830 metabolic network. As with *i*RH830, the subsystem with the highest number of essential reactions in *i*LF517 is for cofactor and prosthetic group biosynthesis (78). The second largest is for amino acid biosynthesis (43), also mirroring *i*RH830.

Amino acid metabolism in *i*LF517 was then examined and compared to *i*RH830 (Fig. 3.9, bottom). There are seven more essential reactions involved in amino acid

metabolism in *i*LF517 compared to *i*RH830. As with *i*RH830 in blood, there are no essential reactions involved in L-valine, L-leucine and L-isoleucine metabolism in *i*LF517. This indicates that this network has maintained redundancy throughout its evolution. Unlike *i*RH830, the pathways for the metabolism of L-arginine and L-proline have significantly more essential reactions than for any other amino acid, with 13 compared to eight for the next largest (L-threonine and L-lysine).

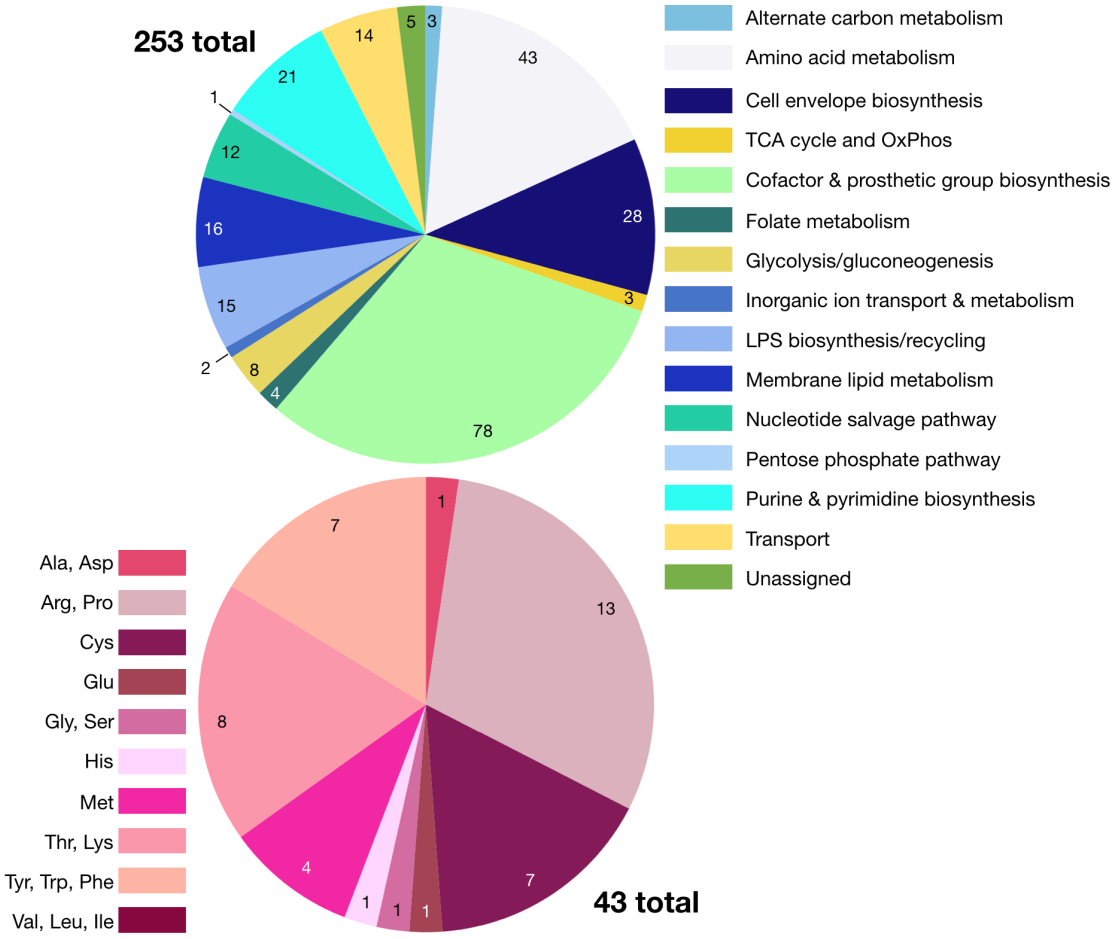


Figure 3.9: Robustness analysis for *i*LF517 in blood, showing the essential reactions grouped by subsystem **(top)** and those involved in amino acid metabolism specifically **(bottom)**.

3.3.7 Media provisioning affects individual reaction flux

After analysing reaction essentiality by subsystem, the flux through individual reactions was then examined. It was shown in Table 3.2 that several components of the blood medium show zero flux through the uptake reactions; L-cysteine, L-glutamine, ornithine, taurine, and L-threonine. This is in spite of transport reactions being present for these metabolites. Overall, there are larger fluxes through reac-

tions involved in central metabolism in the famine medium in comparison to the blood medium (Fig. 3.10). Central metabolism may be required to not only produce ATP, but to synthesise the essential amino acids required for biomass that are not provided in the medium. In famine, the greatest flux is seen through the pathways for oxidative phosphorylation, particularly ATP synthase (ATPS4r), cytochrome oxidase (CYTBO3), and NADH dehydrogenase (NADH6). In blood, the greatest flux is through another NADH dehydrogenase reaction, NADH7. There are greater fluxes through the reactions of the TCA cycle in famine than blood. The phosphoglycerate kinase (PGK) and phophoglycerate mutase (PGM) reactions have greater flux in blood than in famine.

There are small differences between fluxes through amino acid metabolism when *i*RH830 is provided with contrasting media (Fig. 3.11). The flux through reactions in L-alanine and L-aspartate, L-threonine and L-lysine, L-cysteine, and L-valine, L-leucine and L-isoleucine are comparable between the two media. There is greater flux through the glucosamine 6-phosphate deaminase (G6PDA) reaction involved in L-arginine and L-proline metabolism in famine than in blood. The pathways for glycine and L-serine and for L-glutamate metabolism also carry more flux in famine than in blood. For L-tyrosine, L-tryptophan and L-phenylalanine metabolism, the flux patterns are broadly similar between the two media with the exception of high flux through the enolase (ENO) reaction in famine.

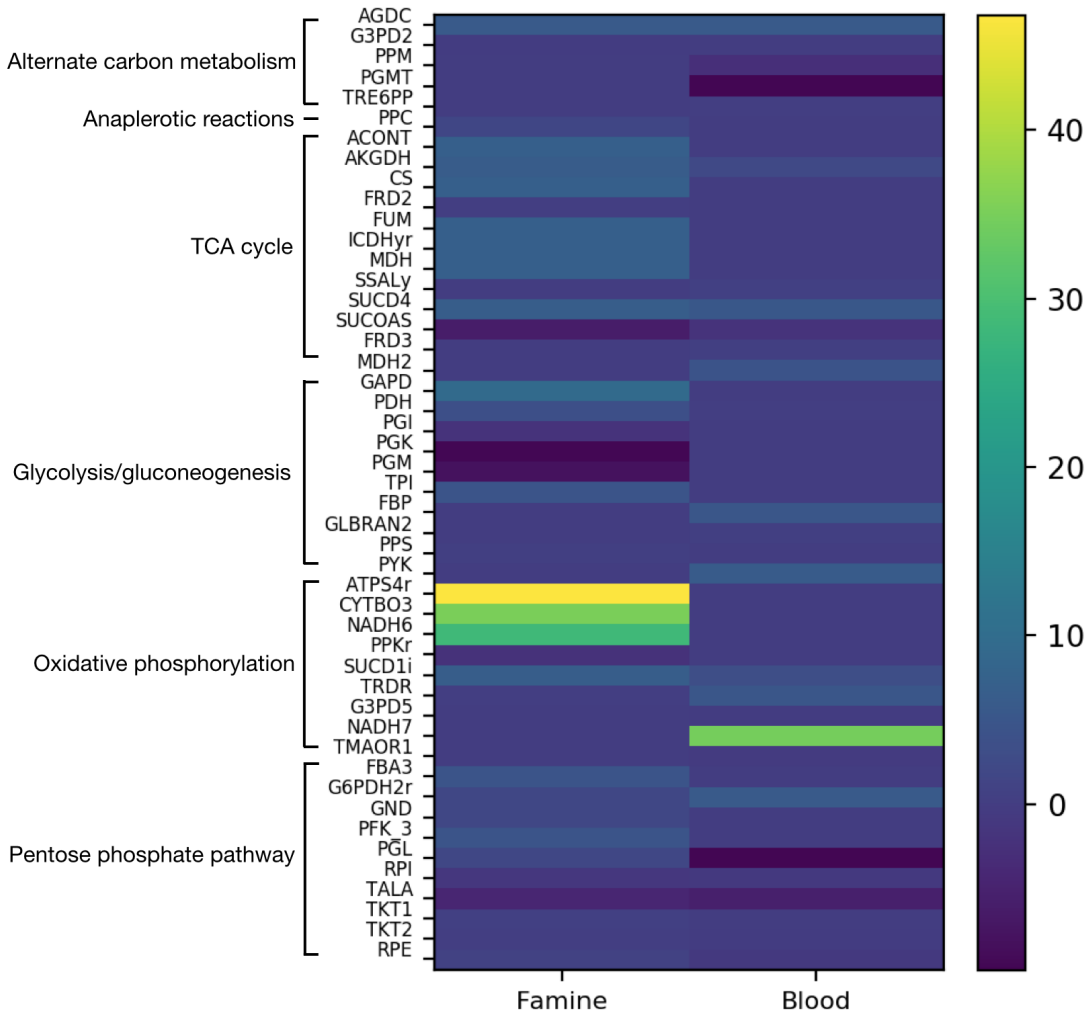


Figure 3.10: Flux (mmol gr DW^{-1} hr^{-1}) through the central metabolism reactions of *i*RH830 in famine and blood media.

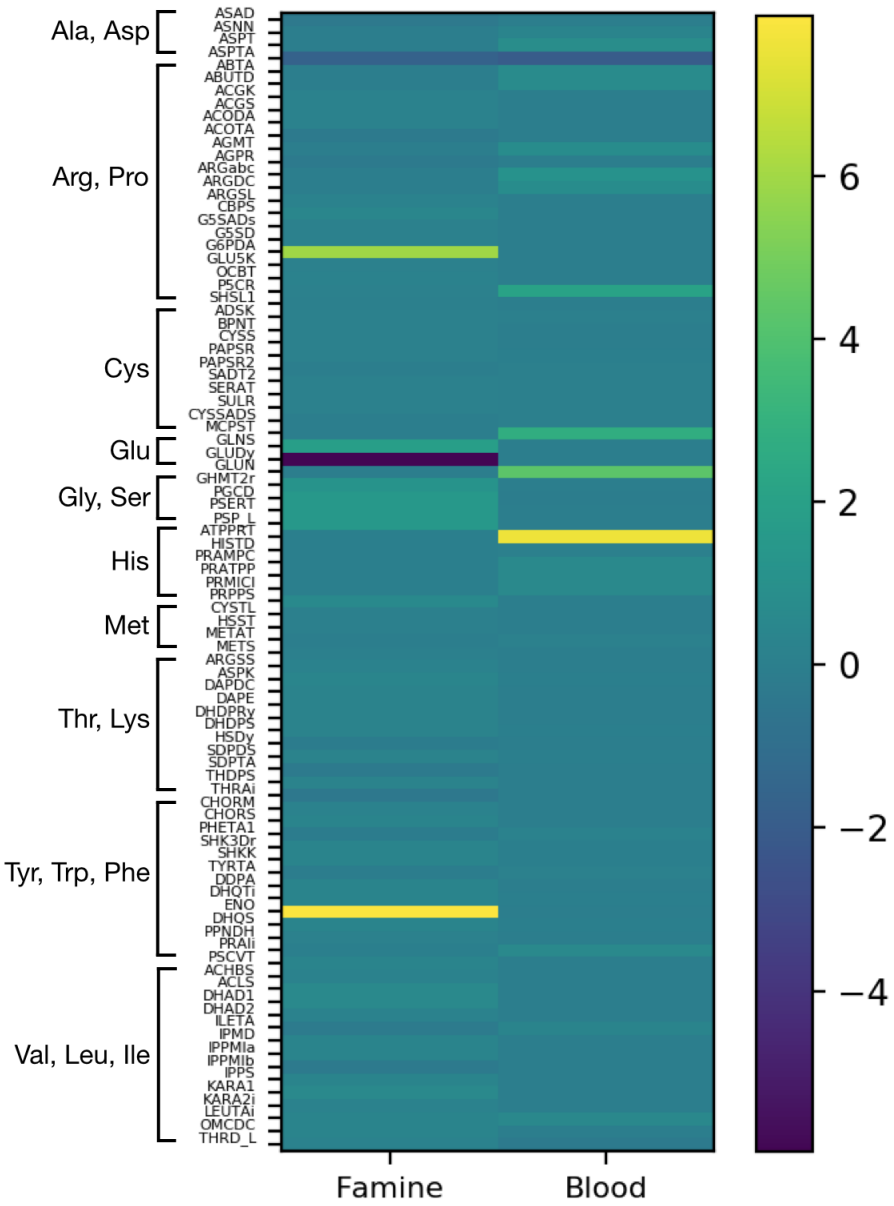


Figure 3.11: Flux (mmol gr DW^{-1} hr^{-1}) through the reactions involved in amino acid metabolism in *i*RH830 grown in famine and blood media.

3.4 Discussion

A flux balance model for *S. praecaptivus* metabolism has been presented here. This prototrophic model, *i*RH830, can produce a positive biomass output in both nutrient-limited and nutrient-rich conditions. As expected of a free-living organism, *i*RH830 exhibits a large amount of metabolic redundancy. The exact profile of redundancy varies depending on nutrient provisioning. Through the robustness analysis, it has been shown that more reactions are essential when the medium con-

tain fewer components. This is because more biosynthetic pathways, and therefore more reactions, are required to produce the metabolites needed for the biomass reaction as they cannot simply import them. There are also more essential reactions in the pathways for amino acid metabolism in the famine medium. This medium does not contain amino acids, and therefore the network is required to maintain more reactions to synthesise these metabolites in the absence of exogenous provisioning. Metabolic pathways for certain amino acids carry more flux in famine than in blood media, namely L-glutamate, glycine and L-serine, and L-tyrosine, L-tryptophan and L-phenylalanine. This suggests that these are key amino acids under nutrient-limited conditions. L-glutamate is a crucial amino acid for *S. glossinidius* (Chapter 2), and is a keystone metabolite for supplementing the TCA cycle and synthesising other amino acids. L-serine is also important for *S. glossinidius* (Chapter 2), and a single reaction (GHMT2r, glycine hydroxymethyltransferase) coverts L-serine to glycine. L-serine is also required to synthesise L-tryptophan (Keseler et al., 2017), with this and the other aromatic amino acids forming from the shikimate pathway via chorismate. Aromatic amino acid biosynthesis has been retained in other *Sodalis* symbionts. *Candidatus* Sodalis pierantonius, for example, is thought to synthesise L-tyrosine and L-phenylalanine for its weevil host (Vigneron et al., 2014). Maintaining these pathways in *S. praecaptivus*, and with high flux, is an indication that aromatic amino acids may be important to members of the *Sodalis* genus.

Exposure to the famine medium results in greater fluxes through reactions in *i*RH830 central metabolism compared to the blood medium. This is observed particularly in reactions for glycolysis/gluconeogenesis, the TCA cycle and oxidative phosphorylation. This suggests that under the famine conditions, the network must synthesise more of the components for the biomass reaction, increasing the flux through key pathways. These *in silico* results can subsequently be applied to understanding the pressures that may lead to a *S. praecaptivus*-like organism become symbiotic. In nutrient-limited conditions, such as those that *S. praecaptivus* may be exposed to, a large number of genes remain essential to synthesise a range of metabolites. Becoming internalised, and therefore experiencing a constant, nutrient-rich microenvironment, reduces the burden on central metabolism. This maintains the production of biomass at a reduced cost to the organism. With a decrease in the number of essential reactions, there are more candidate genes that could be lost as the genome becomes streamlined over time.

The importance of thoroughly testing FBA models has been demonstrated. Only through a large-scale phenotypic screen of carbon usage was the ability of *S. praecaptivus* to grow on xylitol and GalNAc alone identified. This had not been included in early iterations of *i*RH830 as it is a carbon source not often present in FBA models

of Gram negative bacteria. Xylitol is a polyol produced from the reduction of xylose and found in plant tissue (Ylikahri, 1979; Prakasham et al., 2009). It is possible that *S. praecaptivus*, due to the nature of its discovery, is able to metabolise components of plant tissue. The differences in xylitol transport and degradation pathways between *E. coli* and *S. glossinidius* are interesting. *S. glossinidius* appears to have orthologues to the components of the *S. praecaptivus* xylitol transporter and degradation pathway (Table 3.6). There were no orthologues identified in *E. coli*. The results here suggest that *E. coli* K-12 cannot metabolise xylitol, but that *S. glossinidius* can. One xylulose kinase orthologue in *S. glossinidius* appears truncated. Once becoming internalised by the tsetse, *S. glossinidius* would no longer require a transporter or metabolic pathway for xylitol. It is possible therefore that the symbiont is beginning to lose components of this pathway in the absence of an external supply of this carbon source.

Care must be taken when attributing metabolic differences between *S. praecaptivus* and *S. glossinidius* to symbiosis only. The assumption that *S. glossinidius* has lost its orthologue to the *S. praecaptivus* xylulose kinase Sant_3756 as a result of being internalised by the tsetse depends on the latter being a direct ancestor of the symbiont. Whilst it is known that *S. praecaptivus* is a close relative of *S. glossinidius* (Chari et al., 2015), it is in fact more closely related to other insect symbionts, namely those of the phytophagous weevil *Curculio sikkimensis* and shield bug *Cantao ocellatus* (Clayton et al., 2012). This is the only known free-living species of the *Sodalis* genus to be identified. That does not mean that it is the only one, nor that it is the closest ancestor to *S. glossinidius*. It is, however, still a useful comparison from which to investigate several aspects of the tsetse-*S. glossinidius* symbiosis. There is currently no information available about the route by which the ancestor of *S. glossinidius* took to become internalised. Analysing the differences between *S. praecaptivus* and *S. glossinidius* could provide interesting evolutionary details about the key pseudogenisations that have occurred in the symbiont.

Chapter 4

The *in silico* evolution of *Sodalis glossinidius*

The work presented in this chapter was conducted in collaboration with Stephen Thorpe, University of York. With Chapter 3, this work forms a submitted manuscript, also available as a pre-print on bioRxiv (doi 10.1101/819946), as:

Hall RJ*, Thorpe S*, Thomas GH and Wood AJ. Simulating the evolutionary trajectories of metabolic pathways for insect symbionts in the *Sodalis* genus.

RJH designed the experiments, conducted the data analysis and wrote the manuscript. ST constructed the evolutionary algorithm and generated the evolution plots. GHT and AJW supervised the project.

4.1 Introduction

Symbioses are both fundamental and ubiquitous in nature. Understanding their evolution poses an ongoing challenge, as well as an expanse of unresolved research questions. Bacterial symbionts of insects provide a range of benefits including stress tolerance (Wilcox et al., 2003; Dunbar et al., 2007), protection from predation (Wilcox et al., 2003; Oliver et al., 2003; Nakabachi et al., 2013), and the provision of metabolites (Aksoy, 1995; Shigenobu et al., 2000; Thomas et al., 2009; Snyder and Rio, 2015; Hrusa et al., 2015; Manzano-Marín et al., 2015). The latter forms arguably the strongest link within the symbioses. Host and symbiont frequently share metabolic substrates, as well as the products and components of individual

biosynthetic pathways (McCutcheon and Moran, 2007; Thomas et al., 2009; McCutcheon et al., 2009a,b; Wilson et al., 2010; McCutcheon and von Dohlen, 2011). These relationships typically enable the host to survive on a nutritionally restricted diet, such as the blood meal of the tsetse (Rio et al., 2003; Michalkova et al., 2014; Snyder and Rio, 2015) or the plant sap that feeds the aphid (Baumann et al., 1995; Akman Gündüz and Douglas, 2009; Richards et al., 2010).

Understanding the evolutionary pressures that affect the organisms within a symbiotic relationship is essential. This includes deciphering how the symbioses develop over time and the way in which the metabolism of the individuals is intertwined. It is, however, often hindered by biological difficulties. Symbiotic bacteria undergo genomic streamlining, have *in vitro* cultivation difficulties, may no longer express stress response genes and might lack a sound outer membrane (Nakabachi et al., 2013; Wu et al., 2004; Pérez-Brocal et al., 2006; Moran et al., 2008; Akman et al., 2002; Dale and Maudlin, 1999; Aksoy, 1995). It is therefore impossible in many cases to test hypotheses about host-symbiont interactions in controlled experimental conditions. In these circumstances, computational techniques offer a viable alternative. Computational techniques, including FBA, can be used to investigate metabolic potential and pseudogenisation in symbiotic bacteria.

FBA has been applied to several microbiological problems. Boolean logical operators have been incorporated into *E. coli* metabolic models to investigate the impact of gene regulation on a system (Covert and Palsson, 2002; Lee et al., 2006; Covert and Palsson, 2003; Covert et al., 2001). Dynamic FBA, where a rate of change in flux constraints is included, has successfully modelled diauxic growth in *Escherichia coli* (Mahadevan et al., 2002). FBA has been used to compare strains of *Blattabacterium* from separate cockroach lineages to assess their divergence (González-Domenech et al., 2012), and to predict the evolution of metabolism from *E. coli* experimental datasets (Harcombe et al., 2013). The evolution of metabolic networks in isolation has also been simulated with the aim of identifying key metabolites (Pfeiffer et al., 2005). FBA has not yet been harnessed to its full potential with regards to the investigation of symbiont evolution. This is perhaps surprising given that several models of *E. coli* metabolism are available as an evolutionary starting point (Edwards and Palsson, 2000; Feist et al., 2007; Reed et al., 2003; Orth et al., 2011; Orth and Palsson, 2012; Pál et al., 2006). The evolution of *Buchnera aphidicola* and *Wigglesworthia glossinidia* from *E. coli* has simulated using FBA (Pál et al., 2006). This work, whilst elegant, is limited two-fold. Primarily, reactions that are lost at the start have no chance of being reintroduced. This limits the evolutionary space that can be explored, as the loss of a key reaction at the start will fundamentally affect which reactions can be lost subsequently. Second, it assumes that *W.*

glossinidia, like *B. aphidicola*, is a subset of *E. coli* (Shigenobu et al., 2000), and that its evolution can therefore be simulated from an *E. coli* ancestor. This may not be the case, and will be discussed in depth in Chapter 5. A similar approach to that used by Pál et al. (2006) was used with dynamic FBA to study the evolution of cooperation and cross-feeding in *E. coli* (McNally and Borenstein, 2017). Using FBA in isolation to remove reactions successively may not therefore be the optimal way to simulate the evolution of symbiosis.

In silico evolution has been used increasingly in recent years to complement *in vivo* experimental evolution (Hindré et al., 2012). *In silico* evolution benefits from being able to test widely different ecological conditions whilst controlling key variables (Batut et al., 2013). For example, it allows the investigation of groups of mutations that lead to a specific phenotype, or mutations that are difficult to induce *in vitro* (François and Hakim, 2004). This has enabled the study of many aspects of evolution, including simulating the reduction of genome size in an individual (Batut et al., 2013). Multi-objective evolutionary algorithms (MOEA) have been used in many disciplines for solving problems that have two or more conflicting objectives. The use of MOEA in combination with metabolic models has been implemented for the design of minimal genomes (Wang and Maranas, 2018) and for the production of industrially relevant molecules (Fong et al., 2005; Garcia and Trinh, 2019). It has however seen only limited use for *in silico* evolution. The evolution of symbiosis can be considered as a multi-objective optimisation, as symbiotic bacteria undergo genome reduction whilst trying to maximise their individual growth.

There are metabolic models available for the free-living *Sodalis praecaptivus* (Chapter 3) and the tsetse-associated *Sodalis glossinidius* (Chapter 2). These models, *i*RH830 and *i*LF517, respectively, represent adaptations of the organisms to their contrasting environments. The *Sodalis* system is therefore an excellent candidate for assessing the ability of FBA to describe the evolution of symbioses. Here, a MOEA has been used to evolve *i*RH830 under various biological conditions. The aim was to investigate computationally the route that *S. glossinidius* may have taken in its transition to symbiosis. It is not known whether the solutions found by *S. glossinidius*, described in *i*LF517, are the only possible outcomes given the metabolic constraints of the microenvironment, or whether the symbiont's unusual metabolic network evolved by chance.

The application of the algorithm to *i*RH830 enabled the observation that certain key pseudogenisations may have occurred earlier in the symbiosis than previously thought. The effect of exposing the ancestral *Sodalis* to contrasting diets was also modelled, mirroring the different trajectories that this genus has taken within blood-

and sap-feeding insects. It is hoped that the techniques used here can be applied to other symbiotic systems to drive forward the discovery of novel relationship criteria.

4.2 Methods

4.2.1 Flux balance analysis

FBA solutions were generated as per Chapter 3. *i*RH830 was supplied with either the famine medium or blood medium detailed in Chapter 3, or a sap-inspired media (from Krishnan et al. (2011), henceforth "sap"). The phenotype was considered viable if the biomass production rate was greater than 1 x 10^{-4} gr DW (mmol glucose)$^{-1}$ hr^{-1}.

4.2.2 Implementation of multi-objective evolutionary algorithm

A MOEA was used to explore possible evolutionary trajectories in the *Sodalis* genus. This was constructed by Stephen Thorpe, University of York. An overview of the process is provided in Fig. 4.1. The non-dominated sorted genetic algorithm (NSGA-II) (Deb et al., 2002) from the Distributed Evolutionary Algorithms in Python (DEAP) package was used in combination with the COBRApy package (Ebrahim et al., 2013) for FBA evaluation. Equal weight was placed on reducing the number of reactions used in the model whilst maximising the biomass output.

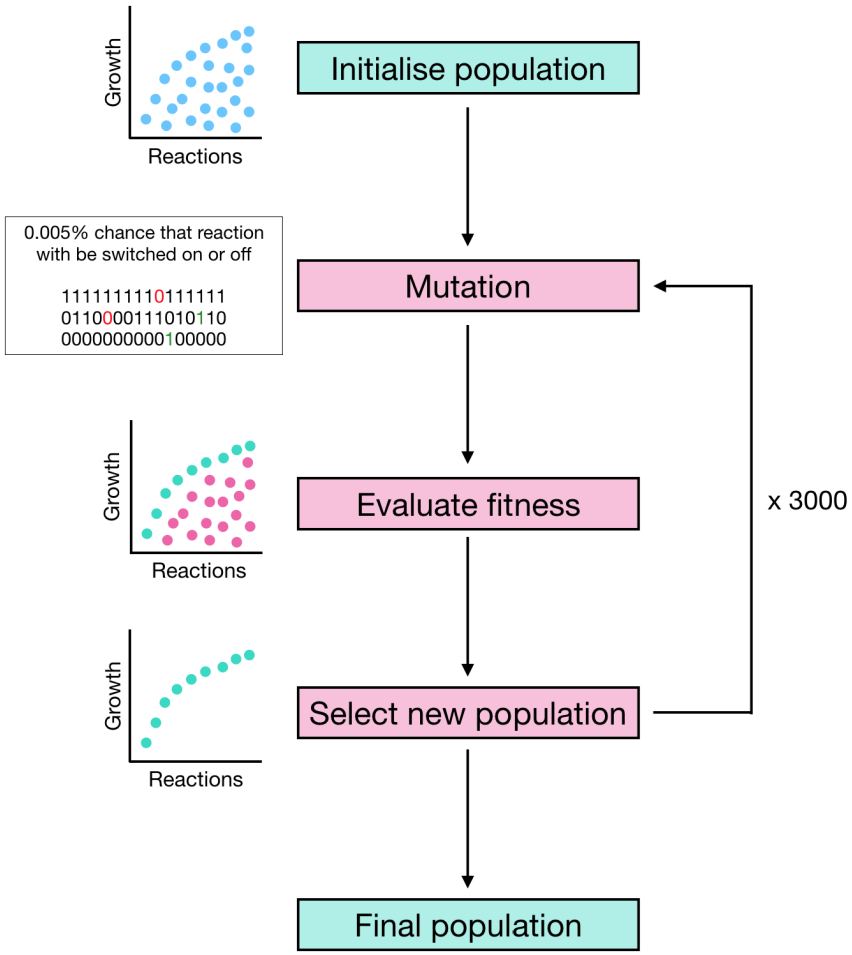


Figure 4.1: Process of the MOEA. A starting population of individuals is initialised, and the fitness calculated by solving the FBA model to calculate biomass output and the number of active reactions. The population is then copied, allowed to mutate and the fitness evaluated again. A new population is selected from the original and copy populations. Green boxes represent the start and final populations, pink boxes represent the iterative process of mutation and selection.

Population initiation

Prior to starting an evolutionary run, reactions essential to growth were identified using a single reaction knockout. Essential reactions were defined as those producing a biomass output of less than 1 x 10^{-3} gr DW (mmol glucose)$^{-1}$ hr^{-1}. Reactions that were identified as essential were not included in the subsequent mutation strategy, therefore reducing the solution space and computational time taken to run the MOEA. The essential reactions were added back to the evolved populations for downstream analysis.

A population of 100 genotype clones, where all non-essential reactions are active, was created (Fig. 4.1). Each genotype consisted of a binary number, where a 1

or 0 corresponded to the reaction being active or inactive, respectively. This is a proxy for gene loss, where a one-to-one gene-protein-reaction mapping is assumed. All post-evolution analysis focused on the reactions lost or retained.

Mutation

Mutation was performed on each genotype by flipping the value of each reaction with a probability of 0.005 (Fig. 4.1). The fitness of each individual is evaluated by solving the FBA model to calculate both its biomass output and the sum of number of active reactions.

Fitness evaluation and selection

The population was first evaluated for non-dominated individuals. This gave a population of individuals that has the highest biomass output for their current number of active reactions (Fig. 4.1). From the non-dominated population, the Euclidean distance between each individual was calculated. A greater priority was given to selecting individuals with a larger Euclidean distance. This prevented the clustering of similar potential solutions, thereby reducing the likelihood of becoming trapped in sub-optimal local minima within the search space. The resulting population maximised the convergence on the highest biomass output, lowest number of reactions, and the distribution of those solutions. There will be a set of solutions whereby the number of reactions cannot be minimised further without also reducing the corresponding biomass output. This set of solutions is known as a Pareto front. The algorithm was repeated for 3000 generations to produce genotypes that converged. This indicated that minimal new solutions were being found. The biomass output from the slim optimisation COBRApy function and the summation of the number of active reactions was used to evaluate the fitness.

4.2.3 MOEA variations

The MOEA was run under several conditions in order to investigate aspects of symbiont evolution. Full details are provided in Table 4.1. Scenario A investigated the trajectories taken when the *S. praecaptivus* model was provided with blood, sap, and famine growth media. In Scenario B, gene knockouts were simulated by removing individual reactions from the *S. praecaptivus* model prior to commencing the evolution. The reactions chosen were ASPTA, PDH, and PPC. In Scenario C,

the MOEA was applied to the model of *S. glossinidius* metabolism presented in Chapter 2, *i*LF517. Here, *i*LF517 was supplied with the blood medium for 3000 generations. In Scenario D, the *S. praecaptivus* model was exposed to cycling media conditions with the aim of simulating the possible fluctuation of nutrient availability inside the insect. Here, both the essential and non-essential reactions were available for mutation. The model was exposed to three conditions, each lasting for ten generations and repeated for a total of 3000 generations; five generations in blood and five in famine, nine generations in blood and one in famine, and the reverse of the latter.

Table 4.1: Conditions under which *i*RH830 and *i*LF517 were evolved, including wild-type (WT) or reaction knockouts, and media type.

Scenario	Test	Model	Media
A	Effect of growth media	*i*RH830 (WT)	Blood, sap, famine
B	Effect of gene loss	*i*RH830 (ΔASPTA, ΔPDH, ΔPPC)	Blood
C	Future of *S. glossinidius*	*i*RH830 (WT), *i*LF517 (WT)	Blood
D	Fluctuating nutrients	*i*RH830 (WT)	Blood:famine 9:1, 1:9, 5:5

4.2.4 Analysis of evolved populations

To identify key reactions in the evolved populations, individuals were selected from each condition and the remaining non-essential reactions extracted. The subset of reactions that were present in every individual selected were designated as "core non-essentials", and will be referred to hereafter as such. When examining the similarity between evolved models, exchange reactions and reactions carrying zero flux were discounted. To analyse the effect of cycling the growth media (Table 4.1, Scenario D), individuals from the evolved populations that experienced 90% of their generations in blood were selected. These individuals were then compared to evolved populations in the standard blood medium (Table 4.1, Scenario A) in order to assess the similarity between the results of the two conditions.

4.3 Results

4.3.1 Media provisioning affects evolutionary trajectories

Species of the *Sodalis* genus have been found in insects that feed on a variety of contrasting diets, including blood (e.g. tsetse (Dale and Maudlin, 1999) and ticks (Novakova and Hypsa, 2007; Chrudimský et al., 2012; Boyd et al., 2016)) and plant

tissue (e.g. weevils (Oakeson et al., 2014)). To replicate *Sodalis* evolution in different environments, the MOEA was applied to *i*RH830 that was supplied with a tsetse-specific blood medium, a medium that mirrors plant sap, and a nutritionally limited "famine" medium (Table 4.1, Scenario A). The algorithm underwent ten runs of 3000 generations and the resulting solutions collated.

In all conditions, the models evolved to completion, demonstrated by the convergence of solutions to the left of the plots (Fig. 4.2). The number of reactions decreases over evolutionary time, with the majority of solutions clustering at the maximum biomass output. This is an indication that sub-optimal solutions are being removed successfully. After 3000 generations, there are a range of solution sizes at the maximum biomass output found in sap, whereas in blood and famine all solutions at this time point cluster at the minimum number of reactions. The two rich media, blood and sap (Fig. 4.2, top, centre), produce a lot of metabolic flexibility, with a complete range of possible biomass outputs produced by the smallest models. When grown in the nutritionally limited famine medium, there is significantly less flexibility in terms of possible solutions found (Fig. 4.2, bottom). Here, the majority of the solutions cluster at the minimum reactions/maximum biomass output. This is as expected, given the fitness function of the MOEA. In blood and sap, the biomass outputs reach near zero, made possible by the variety of available substrates. In famine, the options for streamlining are limited, resulting in few solutions that are able to deviate away from what is selected by the fitness function.

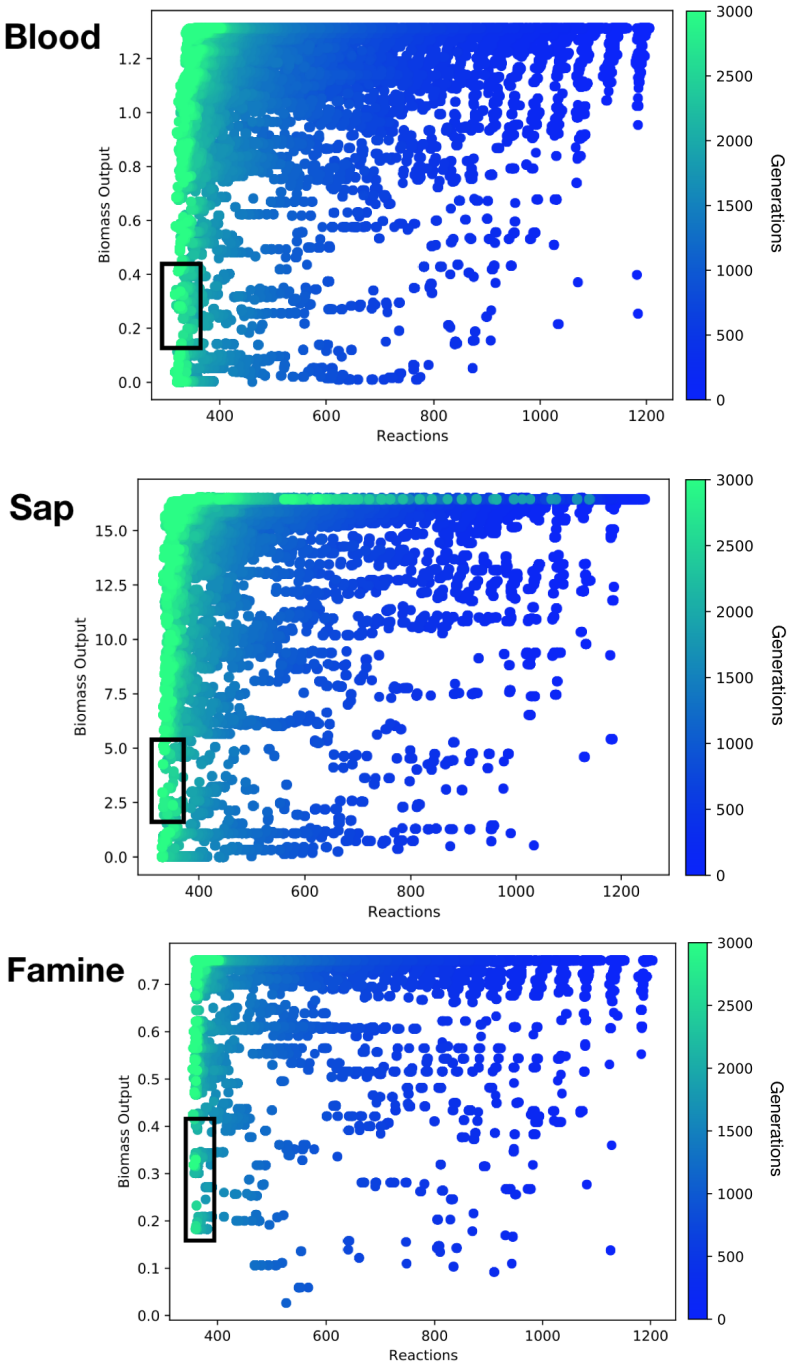


Figure 4.2: *i*RH830 evolved under different media conditions; a tsetse-species blood medium **(top)**, a medium mimicking plant sap **(centre)**, and a nutritionally limited famine medium **(bottom)**. The algorithm was run for 3000 generations, with the plot depicting new populations every 50 generations (blue to green). Black boxes indicate individual solutions selected for further analysis.

A number of individual solutions from each of these simulations were then selected (Fig. 4.2, black boxes). The raw, binary data were translated back into reaction names and this was subsequently processed to produce a list of "core non-essential reactions". These reactions are found in all individuals selected, and do not produce

a lethal phenotype when removed. A full list of all core non-essential reactions described in this Chapter can be found in Supp. File 3. There are 14 core non-essentials reactions found in all 1194 of the individuals examined when *i*RH830 was supplied with blood; AGDC, ARGabc, ASNt2r, G6PDA, H2Ot, HISt2r, ILEt2r, NH4t, RPE, TKT1, TKT2, TMK, TRPt2r and TYRt2r. In sap, only one non-essential reaction is found in all 1888 individuals; the L-arginine ABC transporter reaction ARGabc. As anticipated, when grown in the limited famine medium there are a higher number of core non-essential reactions (22 found in each of the 2989 individuals tested); ATPS4r, CO2t, ENO, FORt, GAPD, GHMT2r, GLUDy, ORNDC, PAPSR, PGCD, PGK, PGM, PPPGO3, PSERT, PSP_L, RPE, TALA, THRAi, TKT1, TPI, TRDR and TRPS1. The rare core non-essential reactions were then calculated. In famine, there are 73 unique reactions that occur in less than 0.1% of the 2989 evolved models. This is significantly more than for sap (13 in less than 0.1% of 1888 models) and blood (five in less than 0.1% of 1194 models).

These core non-essential reactions were then analysed by subsystem to assess themes across the different conditions. In blood, over half (eight of 14) of these are secondary transporter reactions (Fig. 4.3, top). This reflects what is observed in *S. glossinidius*, which has retained, for example, secondary amino acid transporters (Table S10). The loss of metabolic pathways and the maintenance of functional transporters is characteristic of symbiotic bacteria that are able to scavenge metabolites from their microenvironment. As mentioned previously, the only core non-essential reaction in sap is a transport reaction (Fig. 4.3, centre). In contrast, the set of core non-essentials are more varied when metabolites are limited (famine), with a particular emphasis towards central metabolism and amino acid metabolism (Fig. 4.3, bottom).

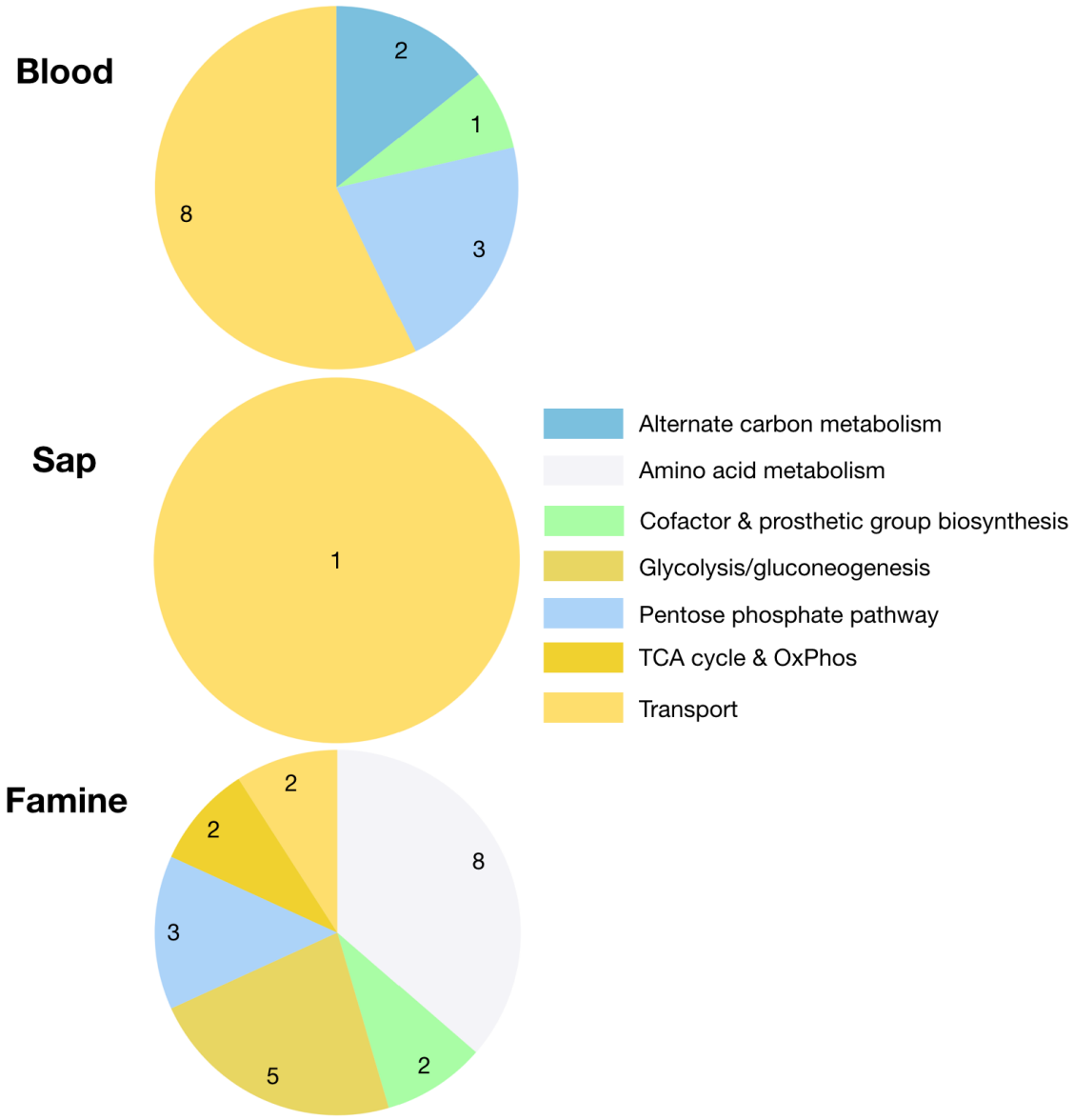


Figure 4.3: The proportion of core non-essential reactions per conditions by subsystem when the ancestral *i*RH830 is exposed to blood **(top)**, sap **(centre)**, or famine **(bottom)** media.

4.3.2 Temporal gene loss can be predicted

A characteristic of *S. glossinidius* and other symbiotic bacteria is their propensity to accumulate pseudogenes. It is not known whether key genes are lost early in the tsetse-*Sodalis* symbiosis to enable the initiation, or whether their loss is an inevitable consequence of genomic streamlining. To investigate the effect that pseudogenising key genes early in evolutionary time has on the trajectory of a symbiont, the MOEA was run on *i*RH830 with one of three reactions removed at the start, with the resulting solutions compared to wild-type (WT) (Table 4.1, Scenario B). The reactions

selected were PPC (phosphoenolpyruvate carboxylase), as a key pseudogenisation in *S. glossinidius* central metabolism (Chapter 2), and PDH (pyruvate dehydrogenase) and ASPTA (aspartate transaminase) as two other reactions that feed the TCA cycle.

When considering the population plots, there is minimal qualitative difference between ΔPDH and ΔPPC (Fig. 4.4, centre, bottom). ΔASPTA, in contrast, produces solutions with a much lower biomass output and with fewer individuals that deviate away from the optimum as defined by the fitness function (Fig. 4.4, top). A selection of individuals were then selected and the number of core non-essential reactions in the evolved models were then analysed as described previously (Fig. 4.4, black boxes). There are one, 11 and nine core non-essential reactions in the WT, ΔPDH and ΔPPC solutions, respectively, whereas there are 61 in ΔASPTA. These 61 reactions function in a variety of subsystems, particularly transport, central metabolism, amino acid metabolism and nucleotide salvage pathways. There are minimal differences between the core non-essential reactions at the subsystem level between ΔPDH and ΔPPC (Fig. 4.5, centre, bottom). The main difference of note is the presence of reactions involved in amino acid metabolism in the ΔASPTA (Fig. 4.5, top) but not the ΔPDH or ΔPPC solutions.

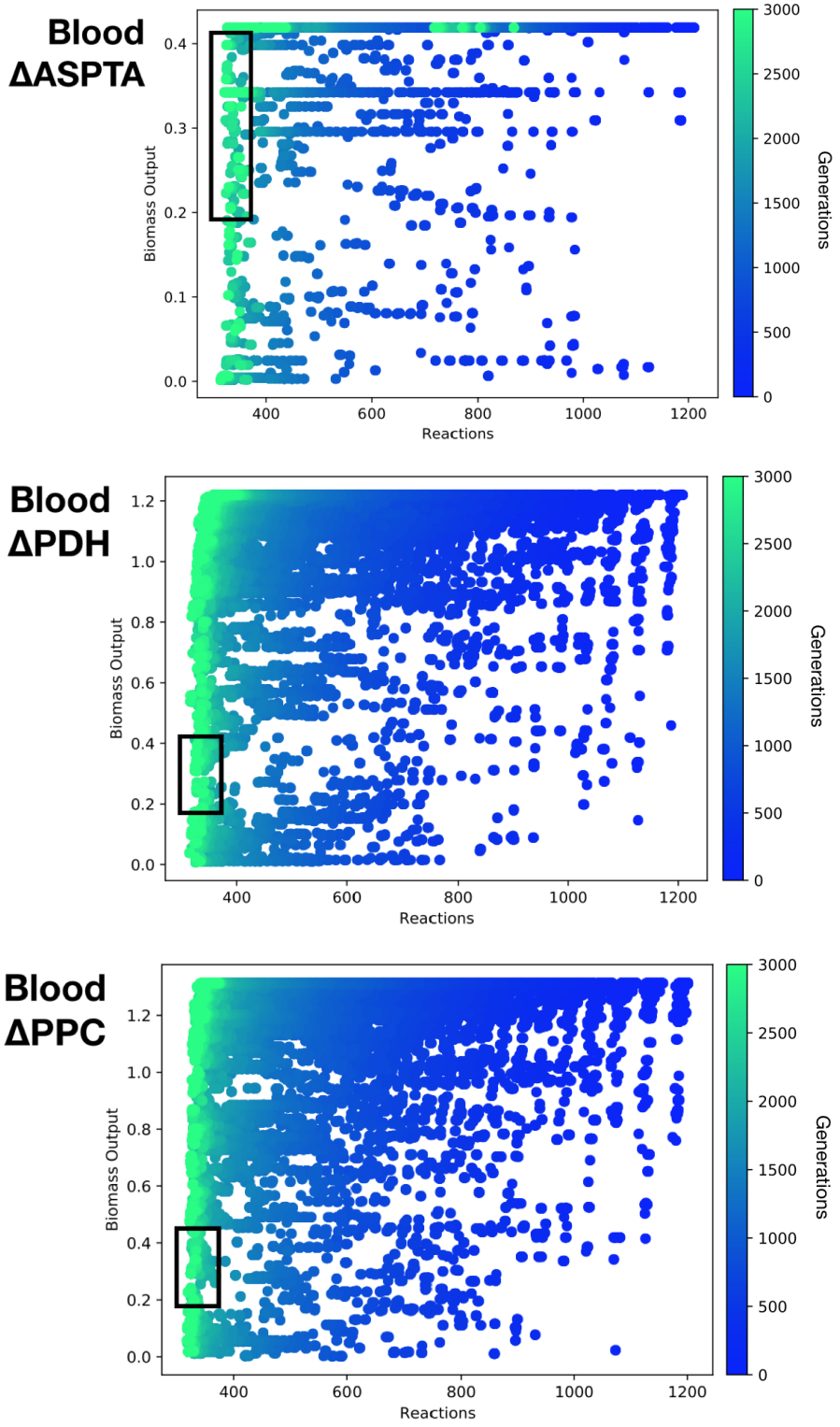


Figure 4.4: *i*RH830 evolution with the reactions ASPTA **(top)**, PDH **(centre)**, and PPC **(bottom)** removed at the start. MOEA was run for 3000 generations (blue to green).

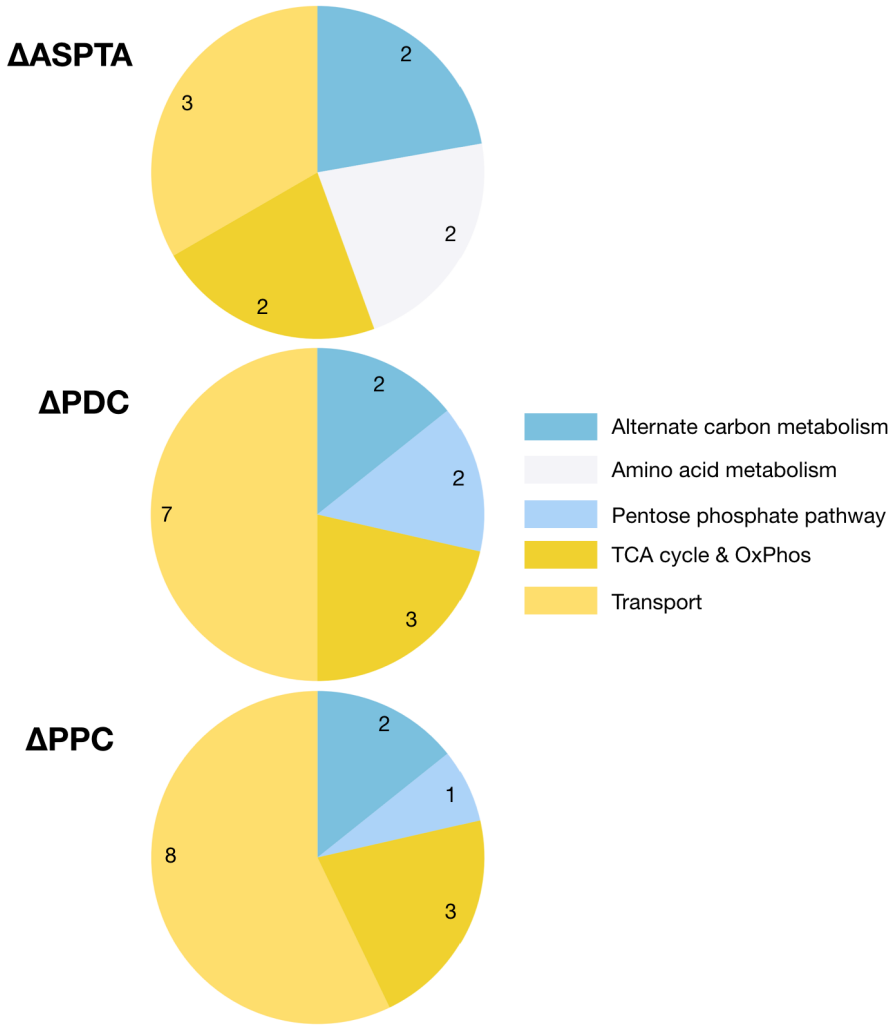


Figure 4.5: Core non-essential reactions in ΔASPTA **(top)**, ΔPDH **(centre)**, and ΔPPC **(bottom)** *i*RH830 models in a blood medium, grouped by subsystem.

4.3.3 A prediction of the evolutionary future of *S. glossinidius* as a symbiont

S. glossinidius is a secondary symbiont. Both bacterium and insect can survive independently of one another, and the former is likely a more recent acquisition. It is however unclear how recently *S. glossinidius* was captured, or, given the pseudogenisations already present, how much more streamlined its genome can become. The algorithm was therefore applied to *i*LF517 in a blood medium with the aim of evaluating potential future evolutionary trajectories within the bounds of its relationship with host and primary symbiont (Table 4.1, Scenario C).

There are a spread of biomass outputs found at the end of the evolution (Fig. 4.6), as observed when *i*RH830 was evolved in blood. The smallest solutions contain

approximately 300 reactions. The evolved solutions were then compared to the evolved *i*RH830 models to assess their similarity. Ten evolved models for both *i*RH830 and *i*LF517 were analysed. All exchange reactions and those that carried zero flux were removed from further analysis.

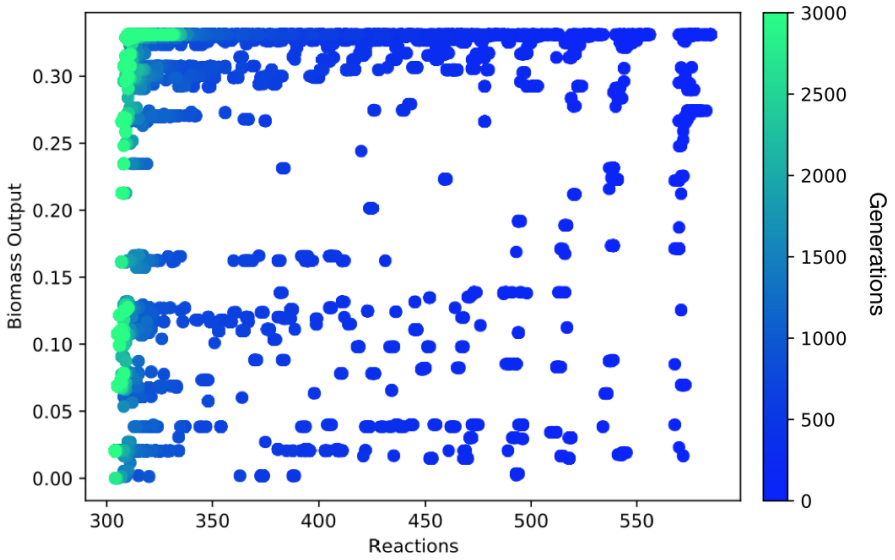


Figure 4.6: Evolution of *i*LF517 in a blood medium. MOEA was run for 3000 generations (blue to green).

Full evolved models with fluxes can be found in Supp. File 4. Of the 383 unique reactions that carry flux in the evolved *i*RH830 models, 289 (75.5%) are found in all ten. For the *i*LF517 solutions, 301 of the 316 (95.3%) unique reactions that carry flux are found in all ten. This suggests that the smaller *S. glossinidius* model has fewer viable trajectories compared to the larger *S. praecaptivus* model. Of the 441 unique reactions across the 20 evolved models, 225 (51%) were found in all of the *i*RH830 and *i*LF517 evolved solutions. The biomass outputs for the evolved *i*RH830 and *i*LF517 solutions range from 0.064 to 0.281 (gr DW (mmol glucose)$^{-1}$ hr^{-1}), and 0.075 to 0.331 (gr DW (mmol glucose)$^{-1}$ hr^{-1}), respectively (Fig. 4.7).

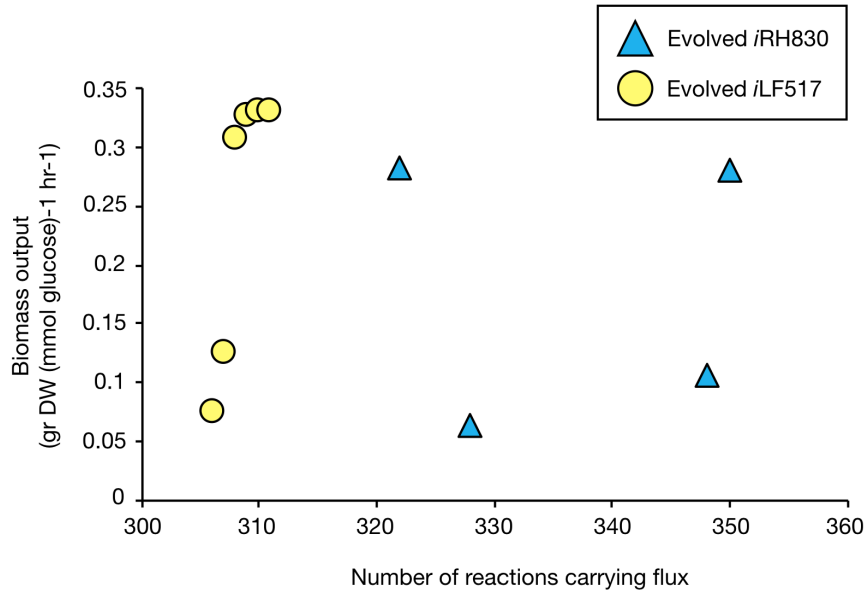


Figure 4.7: Biomass output and the numbers of reactions carrying flux in evolved *i*RH830 (blue triangle) and evolved *i*RH830 (yellow circle) models. The evolved solutions produce comparable biomass outputs. Ten evolved solutions are given for each, some duplicates are present.

The use of the *in silico* blood medium by ten evolved *i*RH830 and ten evolved *i*LF517 models in comparison to the ancestral models was then examined. Full details are given in Tables S11 and S12, and an overview in Fig. 4.8. GlcNAc (acgam) uptake shows the maximum allowed flux (6 mmol gr DW^{-1} hr^{-1}) in the ancestral *i*RH830, ancestral *i*LF517, and four evolved *i*RH830. All evolved *i*LF517 models had zero flux through the GlcNAc uptake reaction. Two evolved *i*RH830 models have zero uptake of L-alanine (ala). For the evolved *i*LF517 models, six show the maximum uptake (3.41 mmol gr DW^{-1} hr^{-1}) of L-alanine observed in the ancestral model. The remaining four have zero flux. All evolved *i*RH830 and *i*LF517 maintain flux of varying values through the L-arginine (arg), L-histidine (his), L-isoleucine (leu), L-leucine (leu), L-lysine (lys), L-methionine (met), L-phenylalanine (phe), L-serine, L-tryptophan (trp), L-tyrosine (tyr), and L-valine (val). The ancestral and evolved *i*RH830 models all have flux through the L-asparagine (asn), glycine (gly), and L-proline (pro) uptake reaction, whereas there is zero flux for these amino acids in the ancestral and evolved *i*LF517 models. For L-proline, only three of the evolved *i*RH830 models have ancestral-like uptake fluxes (2.36 mmol gr DW^{-1} hr^{-1}). L-cysteine (cys) is used in the ancestral and evolved *i*LF517 models, but none of the *i*RH830 models. L-aspartate (asp) shows similar patterns between the two species, with uptake seen in both ancestral and three each of the evolved models. L-glutamate (glu) is used in both ancestral models, but by only six of the evolved *i*RH830 models. None of the ancestral or evolved models have any flux through

the uptake reactions for L-glutamine (gln), L-ornithine (orn), taurine (taur) or L-threonine (thr).

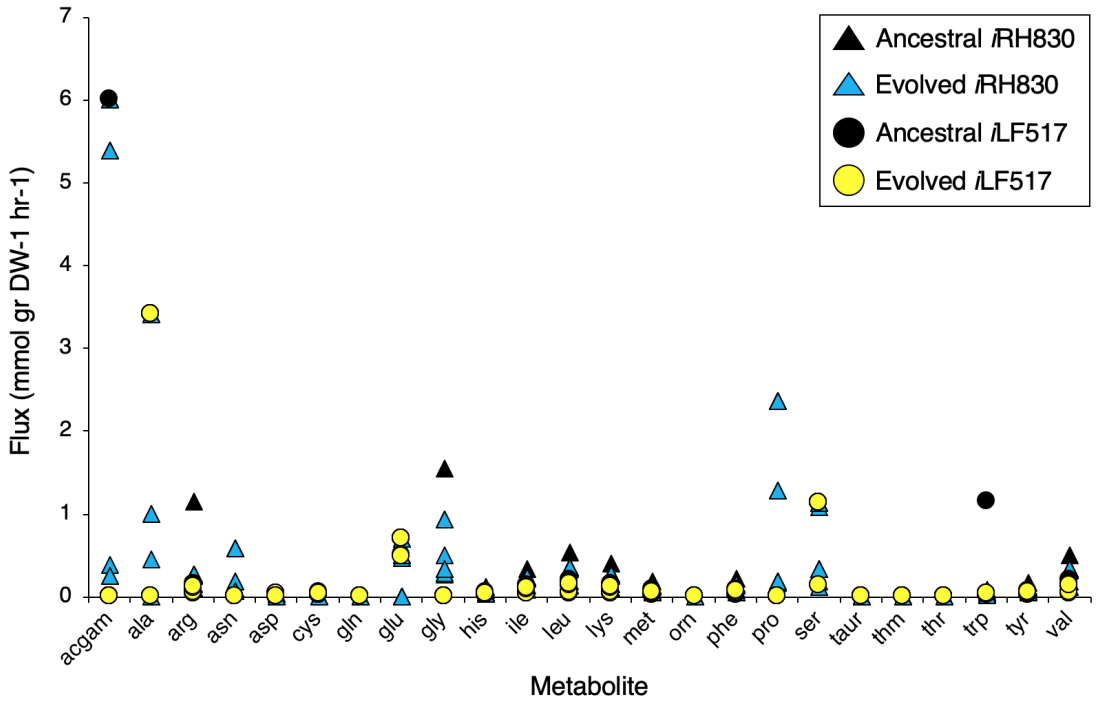


Figure 4.8: Fluxes through the uptake reactions for components of the blood medium in ancestral *i*RH830 (black triangle) and *i*LF517 (black circle), and evolved *i*RH830 (blue triangle) and *i*LF517 (yellow circle). Ten evolved solutions for *i*RH830 and *i*LF517 are given.

4.3.4 Fluctuating the nutrient availability affects the trajectory and final population

In order to simulate *S. glossinidius* evolution accurately, it is important to consider that insect feeding may result in the availability of metabolites fluctuating. To reflect this, the MOEA was modified in order to cycle the medium between blood and famine over different generational times (Table 4.1, Scenario D). The MOEA was run on *i*RH830 under three different, cycling conditions for a total of 3000 generations; (i) nine generations in blood, one in famine, (ii) five generations in blood, five generations in famine, and (iii) one generation in blood, nine in famine. The evolved models were then put back into both blood and famine to assess their fitness in the different media.

The three conditions produce three different evolutionary profiles (Fig. 4.9). There are a spread of solutions that are viable in blood, whereas there are only a spread of

solutions viable in famine when the model is exposed to at least equal numbers of generations in famine. When 90% of the generations are exposed to blood (Fig. 4.9, top), the majority of the solutions are adapted to this and have minimal flexibility when given the famine medium, finding only the maximum biomass outputs in the latter. When given an equal numbers of generations in each medium (Fig. 4.9, centre), there are a spread of solutions that are viable in both blood and famine media. Additionally, only the earliest generations are able to produce a high biomass output in the famine medium. This may represent some organisms that are averagely adapted to both conditions, without optimising for one. When 90% of the generations are in famine (Fig. 4.9, bottom), however, there are more possible solutions found that are successful in famine, as well as producing high biomass outputs in blood. This may represent some organisms becoming specifically adapted to one type of medium.

To examine whether cycling the medium increases the similarity of evolved *i*RH830 to evolved *i*LF517 models, ten evolved models from the condition exposed to blood for 90% of the generations were compared to the ten evolved *i*LF517 solutions. Full details can be found in Supp. File 4. The individual reactions were compared as before, and it was discovered that providing *i*RH830 with the cycling medium increased the number of reactions in all 20 models from 51% to 52.4% (222 reactions of a possible 424). Fluctuating the media may therefore produce solutions that have a higher degree of similarity to *S. glossinidius*.

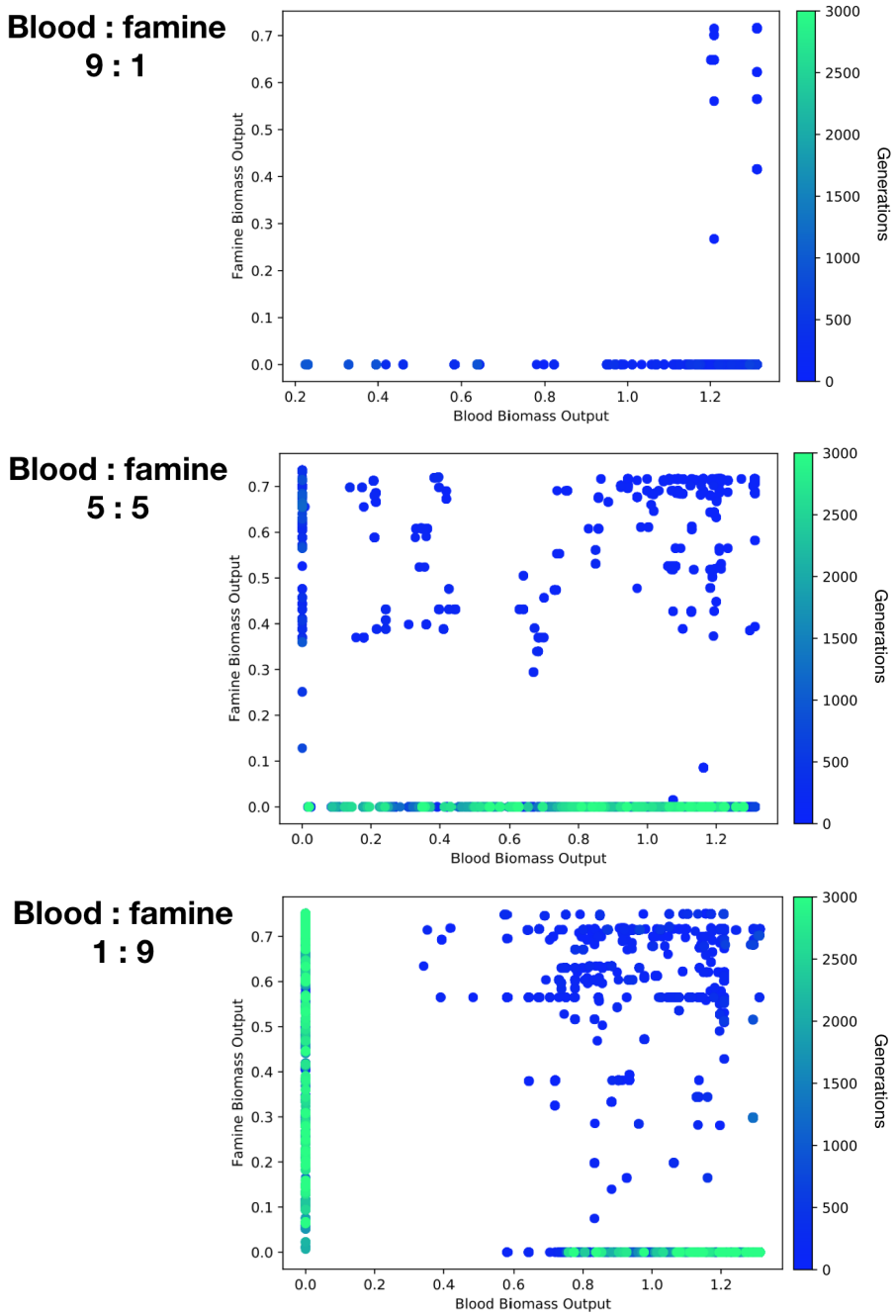


Figure 4.9: Cycling the media provided to *i*RH830 during its evolution. MOEA was run for a total of 3000 generations, with cyclical repeats of ten generations as: nine blood one famine **(top)**, five blood five famine **(centre)**, one blood nine famine **(bottom)**.

4.4 Discussion

Classical studies of microbial evolution, whilst useful, are ultimately limited by their inherent inability to replicate adaptations over large evolutionary time scales.

Here, we describe a solution to this issue and present a computational approach, by combining a MOEA with FBA. We present the *Sodalis* system as a model for this. The *Sodalis* genus is ideal for the study of the evolution of symbiosis in this way, as within the genus are a free-living and a host restricted species, both well defined with complete genome sequences and existing protocols for culture.

The work here demonstrates the power of evolutionary algorithms in the study of symbiont evolution. A strength of this system is that removal of a reaction from the model is not irreversible; it is possible for a reaction to be added back into an individual at any point. Whilst there is no evidence currently for horizontal gene transfer within *S. praecaptivus*, the NSGA-II algorithm is only intended to be used as a tool to explore the possible evolutionary space rather than as a biologically accurate model of genome reduction. Previous examples of evolving minimal metabolic networks do not allow for full exploration of the possible evolutionary space (Pál et al., 2006). Decisions that are made at the start of process persist which, whilst biologically relevant, do not allow the full complement of evolutionary routes to be examined. It is not possible to know at what point during the transition to symbiosis the genes in *S. glossinidius* became pseudogenised. It is therefore valid, using the technique described here, to not attach significance to the temporal sequence in which the mutations occur.

Supplying the ancestral *i*RH830 with contrasting growth media demonstrates the effect that nutrient provisioning may have on evolutionary trajectories of symbiotic bacteria. Exposure to famine reflects what might be expected in a nutrient-limited environment *in vivo*, in which evolutionary pressures result in the retention of pathways to synthesise key, essential metabolites. Here, this has shown to be particularly evident in the pathways retained for glycolysis/gluconeogenesis, the pentose phosphate pathway, and amino acid metabolism. This indicates that key pathways in central metabolism and for the synthesis of biomass components are being retained when the external environment is nutrient limited. The evolved solutions therefore reflect what is observed in symbiotic bacteria; the retention or loss of pathways can be used to inform about the microenvironment it resides within.

It was discussed in Section 4.3.1 that the evolved famine solutions contain a much greater number of core non-essential reactions that are present in a small percentage of the solutions. This suggests a lack of flexibility in the evolved network; either the reaction is found repeatedly, or not at all. This is not observed in the solutions provided with rich media (blood or sap), where a greater degree of flexibility is demonstrated by more reactions being included repeatedly across the evolutionary space. This implies that, *in vivo*, there are many possible trajectories for an early

symbiont if there are sufficient nutrients in the microenvironment.

This tool can produce biologically relevant simulations that accurately reflect the metabolic pressures that symbionts are exposed to. An example of this was demonstrated by the investigation of key knockouts in *S. glossinidius*. The symbiont has a pseudogenisation in *ppc*, a key gene in central metabolism (Chapter 2). It is not possible to deduce when this loss occurred from the genome annotation alone. By removing the PPC reaction from *S. praecaptivus* at the start of the *in silico* evolution, the resulting trajectories can be analysed and compared to WT. The loss of PPC appeared to have minimal effect on the resulting evolved populations compared to WT, in contrast to what was observed when the ASPTA reaction was removed at the start (Fig. 4.4). This would indicate that, *in vivo*, the loss of the gene encoding the ASPTA reaction would have a greater impact on a bacterial symbiont if it was lost early in the relationship in comparison to the lower burden that the loss of the genes encoding PDH or PPC would have. This is interesting when considered in the context of the tsetse-*S. glossinidius* symbiosis. *S. glossinidius* has lost the *ppc* gene, whereas it has retained the genes encoding the PDH (*SG0467-9*) and ASPTA (*SG1006*) reactions (Hall et al., 2019; Dale and Maudlin, 1999). As the profile of ΔPPC evolution is similar to that of WT, it could be suggested that the *ppc* gene could have been lost early in evolutionary time without heavily bottlenecking *S. glossinidius* evolution subsequently. The gene encoding the ASPTA reaction may have been retained by *S. glossinidius* because of the detrimental impact that its loss may have caused. This is therefore a useful tool for making general predictions about the temporal occurrence of key pseudogenisations in insect-bacterial symbioses.

It has been shown here that is it possible for *S. glossinidius* to reduce its metabolic network to approximately half of the size that it is currently. This provides support for the published hypothesis that *S. glossinidius* is a recent acquisition by the tsetse (Dale and Maudlin, 1999). The number of reactions remains slightly higher in evolved *i*RH830 models compared to evolved *i*LF517 solutions, possibly due to difficulties in finding the minima from a larger starting point. *i*RH830 can however be reduced down to look phenotypically similar to *i*LF517 at the level of biomass output, but with differences at the individual reaction level. Cycling the growth media to include a short period of low nutrient availability increased the similarity between evolved *i*RH830 and evolved *i*LF517 models. These results suggests therefore that the route that *S. glossinidius* has taken within the tsetse is perhaps just one of several possible routes. The differences also indicate that *S. praecaptivus* may not be the ancestor that initiated the tsetse-*S. glossinidius* symbiosis.

The *Sodalis* genus, with a well-characterised free-living organism and symbiont rel-

ative, is a useful system to investigate the evolution of symbiosis. Previous uses of metabolic models to simulate evolution have focused on *E. coli* as a proof of concept. The tool described here has augmented knowledge about the temporal loss of key genes in *S. glossinidius* central metabolism. It has also investigated how *S. glossinidius* may have adapted to fluctuations in nutrient availability over time. It is not limited to this system, however. Combining FBA with a MOEA could be used for any organism for which a well-annotated genome is available. It could be applied not only to the evolution of symbiosis but to the directed evolution of, for example, industrially relevant microorganisms or to the study of rapid genome evolution in pathogenic bacteria.

Chapter 5

Phylogenetic and experimental analysis of amino acid transporters in *Wigglesworthia glossinidia*

5.1 Introduction

Wigglesworthia glossinidia is the primary symbiont of the tsetse, localising intracellularly in bacteriocytes in the anterior gut (Aksoy, 1995). It is essential for the host (Nogge, 1978, 1976), demonstrated by the production of *W. glossinidia*-free tsetse by treatment with the antibiotic ampicillin (Pais et al., 2008). This clears the *W. glossinidia* population from the milk glands, resulting in progeny that do not have this symbiont. The female offspring were found to be infertile, highlighting the importance of *W. glossinidia* for the tsetse (Pais et al., 2008). *W. glossinidia* is also thought to influence the competency of the tsetse as a disease vector. This relationship has been shown to vary depending on the species of *Glossina* tested (Rio et al., 2019). *W. glossinidia* is therefore of scientific interest as a primary symbiont, as well as of medical importance.

The genomes of two subspecies of *W. glossinidia* have been published at the time of writing, from *Glossina morsitans morsitans* and *Glossina brevipalpis*. These subspecies are thought to be highly similar. The genomes of the two subspecies differ slightly in size, at 719,535 bp for *W. glossinidia morsitans* (Rio et al., 2012) and 703,004 bp for *W. glossinidia brevipalpis* (Akman et al., 2002). Both subspecies retain the capacity to synthesis the B vitamins biotin, thiamine, riboflavin, pan-

tothenate, and pyridoxal 5'-phosphate, and neither appear able to synthesise cobalamin or nicotinamide (Akman et al., 2002; Rio et al., 2012). There are 21 genes unique to *W. glossinidia morsitans* compared to *W. glossinidia brevipalpis*, and 19 in the latter compared to the former (Rio et al., 2012). Arguably the key difference between the two lies in the pathway for folate biosynthesis. This is complete in *W. glossinidia morsitans*, but absent in *W. glossinidia brevipalpis* (Rio et al., 2012). As with the tsetse, *Trypanosoma brucei* is also unable to synthesise folate, instead encoding a transporter that enables the parasite to scavenge this vitamin from its microenvironment (Berriman et al., 2005; Rio et al., 2012). This demonstrates the complex metabolic network within the tsetse microbiome, and the importance of fully understanding the role that *W. glossinidia* plays within this network. A deeper understanding of the metabolism of *W. glossinidia* is essential for elucidating the interactions between host, symbiont, and parasite.

It has been thought previously that, like *Buchnera aphidicola*, the *W. glossinidia* genome is a subset of the *Escherichia coli* genome (Shigenobu et al., 2000). It has been presumed that all metabolic genes present in the *W. glossinidia* genome would have an orthologue in *E. coli*. There are currently no studies that investigate this assumption. This may be due in part to the difficulties in studying an unculturable primary symbiont. A genome comparison could be used, but this relies on there being a high degree of sequence similarity between *W. glossinidia* and *E. coli*, and any predictions made can not be easily tested *in vitro*. A confirmation that the *W. glossinidia* genome is indeed a subset of *E. coli* would allow a well-annotated *E. coli* genome to be used as a reference with confidence. An inability to support it will, however, provide intriguing new insights into the phylogeny of *W. glossinidia*.

Metabolic modelling and FBA could be used to investigate the role of *W. glossinidia* in its ecosystem. Reconstructing its metabolic network would enable its intricacies to be explored in detail. It would also allow *in silico* experiments into the future of *W. glossinidia* as an obligate symbiont to be investigated, via analysis of the robustness of the network. A draft metabolic model of *W. glossinidia brevipalpis* has been developed by Sandy Macdonald and Gavin Thomas (University of York, data unpublished). When investigated by FBA, it was discovered that the amino acid L-proline could serve as the sole carbon, nitrogen, and energy source for *W. glossinidia*. In *E. coli*, exogenous L-proline is transported into the cell and is subsequently used in the downstream biosynthesis of L-glutamate and L-arginine (Fig. 5.1). There is also a direct route from L-proline and L-glutamate to the TCA cycle.

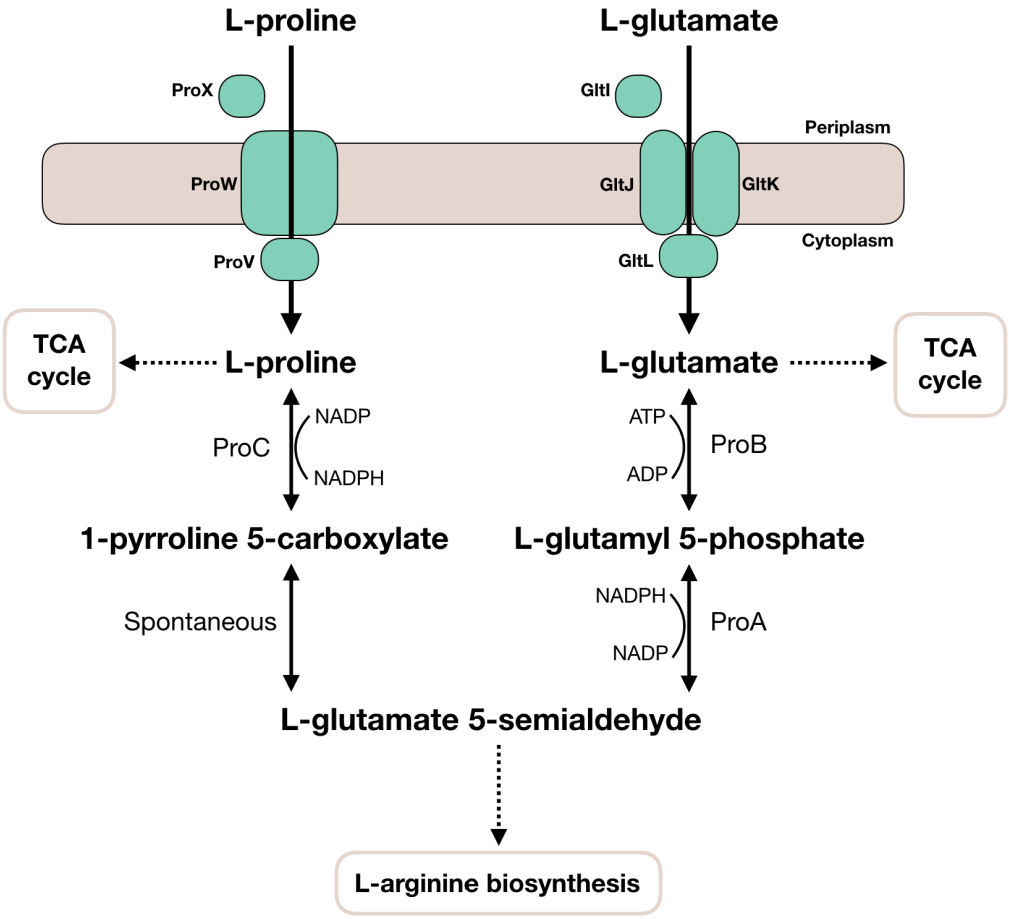


Figure 5.1: L-proline and L-glutamate transport and catabolism in *E. coli*. L-glutamate is imported using the GltIJKL ABC transporter and L-proline through ProXWV. Transporters for L-glutamate and L-proline have not been characterised in *W. glossinidia*.

Using FBA, L-proline alone was also found to be sufficient to redox balance nicotinamide adenine dinucleotide phosphate (NADP) and its reduced form (NADPH). *W. glossinidia* does not encode the oxidative branch of the pentose phosphate pathway. This branch, for the conversion of glucose-6-phosphate to ribulose-5-phosphate, produces two molecules of NADPH per molecule of glucose (Fig. 5.2). In the absence of the oxidative branch, the method of redox balancing by *W. glossinidia* was previously unknown. The use of L-proline provides a theoretical solution to this, via the conversion of L-proline and NADP to 1-pyrroline 5-carboxylate and NADPH (Fig. 5.1).

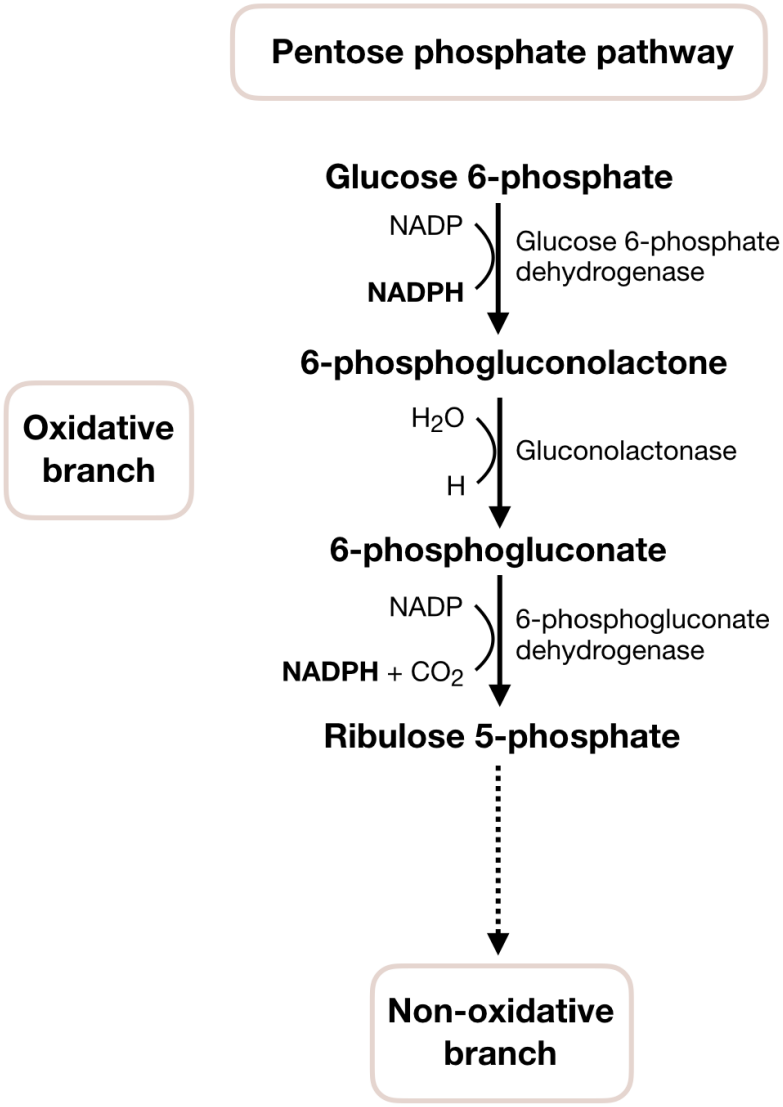


Figure 5.2: The oxidative branch of the pentose phosphate pathway.

L-proline is an integral metabolite within the tsetse microenvironment. It is essential for the insect as it is the main energy source for flight (Bursell, 1963, 1978; Hargrove, 1976; Dyer et al., 2013). The pyridoxal 5'-phosphate produced by *W. glossinidia* is thought to be critical for maintaining L-proline homeostasis within the tsetse (Michalkova et al., 2014; Rio et al., 2012). It is used as a cofactor by the enzyme alanine-glyoxylate aminotransferase, responsible for the conversion of L-alanine to L-proline in the fat bodies for use in flight (Michalkova et al., 2014). This therefore provides a clear metabolic link between host and symbiont. L-proline is also essential for the survival of *T. brucei* within the tsetse (Mantilla et al., 2017). That L-proline may also be essential for *W. glossinidia* is both intriguing and perhaps unsurprising.

It is believed that L-proline is being taken into the *W. glossinidia* cell by a trans-

porter as yet uncharacterised. Initial investigation identified one candidate transporter, annotated as FliY. In *E. coli*, FliY is a binding protein ATP binding cassette (ABC) transporter with a specificity for L-cystine (Butler et al., 1993). The specificity of this protein in *W. glossinidia* is not known. It is proposed that this transporter may import the L-proline that *W. glossinidia* can subsequently convert to L-glutamate, producing NADPH (Fig. 5.1).

ABC transporters are a large superfamily of proteins that couple transport to ATP hydrolysis (Higgins et al., 1986). Almost 5% of the *E. coli* genome is involved in encoding components of ABC transporters, reflecting their importance and abundance (Linton and Higgins, 1998). ABC uptake systems consist of five subunits (Higgins et al., 1986; Higgins, 2001) (Fig. 5.3). There are two transmembrane domains consisting of α-helices that span the membrane and two ATP-hydrolysing nucleotide binding domains (Hollenstein et al., 2007b). These transporters also associate with substrate binding proteins (SBPs) (Hollenstein et al., 2007a) that bind a variety of ligands at a range of affinities (Quiocho and Ledvina, 1996). It is often simpler to express, purify and characterise soluble proteins in comparison to those that are membrane-bound. Investigating SBPs is therefore a useful proxy for elucidating the role of the transporter as a whole.

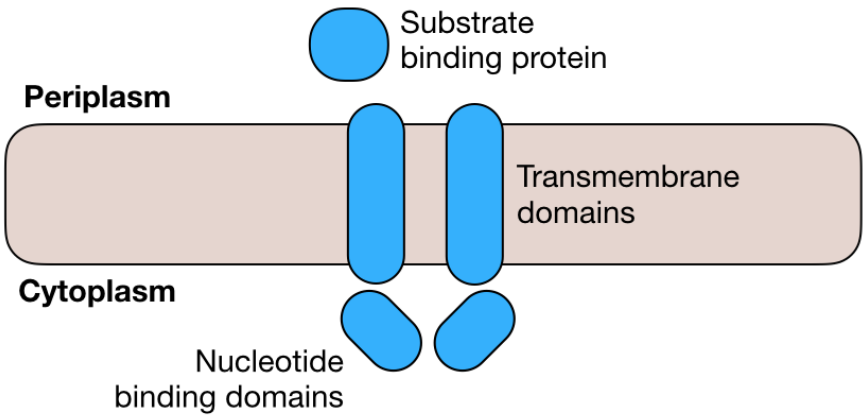


Figure 5.3: Schematic of the structure of ABC transporters.

Here, phylogenetics was used to investigate the relatedness of the two sequenced *W. glossinidia* subspecies, *W. glossinidia brevipalpis* and *W. glossinidia morsitans*. A large phylogenetic tree of amino acid binding proteins in the Enterobacteriaceae was constructed to assess the similarity of the *W. glossinidia* FliY to other SBPs. *W. glossinidia* FliY was found in a cluster with no *E. coli* orthologues, suggesting that *W. glossinidia* may not a subset of *E. coli* as previously thought. This candidate *W. glossinidia* L-proline SBP was expressed and purified, with the aim of establishing its binding specificity to provide *in vitro* support for the *in silico* FBA predictions.

This work provides a greater understanding of the phylogenetics of an important primary symbiont, as well as highlighting issues with using standard molecular biology techniques in the study of symbiotic bacteria.

5.2 Methods

5.2.1 Phylogenetics

Protein sequences were collected by conducting protein BLAST searches against the Enterobacteriaceae. *E. coli* K-12 sequences for ArgT, HisJ, ArtI, ArtJ, GlnH, GltI, FliY, ProX and YhdW were collected from the UniProt database (Bateman et al., 2019) and used as search queries. These proteins were selected as they are all of the amino acid SBPs in *E. coli*. All BLAST hits that produced full length proteins were collated. Any duplicates discovered through more than one search were removed, with only a single copy kept for further analysis. Locus tags were used as identifiers where possible. A sequence alignment was performed on the resulting 503 protein sequences using ClustalX 2.1 (Larkin et al., 2007). A maximum likelihood phylogenetic tree was then constructed in MEGA X (Kumar et al., 2018) using the resulting alignment file. The tree was constructed using the Jones-Taylor-Thornton matrix-based model (Jones et al., 1992). The bootstrap consensus tree was inferred from 100 replicates (Felsenstein, 1985). Branches corresponding to less than 50% of reproduced bootstrap replicates were collapsed. Further downstream processing of the tree was conducted using Interactive Tree of Life (iTOL) (Letunic and Bork, 2007).

5.2.2 Gene cloning

A *W. glossinidia* gene encoding a candidate SBP was identified, with the protein annotated as FliY (*WIGMOR_0593*). Phobius (Käll et al., 2004) and SignalP v5.0 (Juan et al., 2019) were used to establish that a signal peptide was present. The signal peptide was removed and the gene purchased pre-cloned into a pET100/D-TOPO (Thermo Fisher Scientific) expression vector. This included an N-terminal hexa-histidine tag, a T7 promoter, and a gene enabling resistance to ampicillin.

5.2.3 Transformation

The pET100/D-TOPO plasmid containing the *WIGMOR_0593* gene was transformed into calcium competent BL21 (DE3) *E. coli* cells (Agilent) using a standard heat shock protocol (Sambrook et al., 1989). Briefly, 100 μM competent cells were thawed on ice for 30 minutes before 10 ng plasmid was added and incubated on ice for an additional 30 minutes. Cells were heat shocked at 42°C for 90 seconds then 900 μL of LB (Table 5.1), pre-warmed to 37°C, was added. Cells recovered at 37°C for one hour at a shaking speed of 220 rpm. A 50 μL sample was then spread on to LB plates containing 100 μg/mL of ampicillin and incubated overnight at 37°C. Cultures were stored in 50% glycerol at -80°C until required.

Table 5.1: Buffers and media used in Chapter 5.

Buffer/medium	Recipe
LB	10 g NaCl (Fisher), 10 g tryptone (Melford), 5 g yeast extract (Oxoid) in 1 L dH_2O
AIM	5% v/v glycerol, 500 mg D-glucose, 2 g α-lactose, 3.3 g $(NH_4)_2SO_4$, 6.8 g KH_2PO_4 7.1 g Na_2PO_4, 1.2 mg $MgSO_4$ in 1 L dH_2O
1 M KPi (pH 7.8)	K_2HPO_4, KH_2PO_4
Tris	50 mM Tris-HCl (pH 8), 200 mM NaCl
TBE (5x)	54 g Tris, 27.5 g boric acid, 20 ml 500 mM EDTA (pH 8)
Coomassie blue	450 mL methanol, 2.5 g Brilliant Blue, 100 mL acetic acid, 450 mL dH_2O
Destain	10% v/v ethanol, 10% v/v acetic acid, 80% dH_2O
Sample buffer	5% 12 g glycerol, 3 mL dH_2O, 10 mL 10% SDS, 1 mL 1M Tris pH 7.2, 60 mg Bromophenol Blue. Add 3 μL β-mercaptoethanol per 100 μL before use
Running buffer	30 g/L Tris, 140 g/L glycine, 10 g/L SDS
Lysis buffer	2.5 mM KPi, 10 mM NaCl, 20% glycerol
Denaturing wash	50 mM KPi, 200 mM NaCl, 20% glycerol, 40 mM imidazole, 2 M GHCl
KPi wash buffer	50 mM KPi, 200 mM NaCl, 20% glycerol, 40 mM imidazole
Elution buffer	50 mM KPi, 200 mM NaCl, 20% glycerol, 500 mM imidazole
Storage buffer	20 mM KPi, 50 mM NaCl

5.2.4 Colony polymerase chain reaction

Five colonies were screened for successful transformation by polymerase chain reaction (PCR). Individual colonies were picked using a sterile pipette tip and mixed in 25 μL Milli-Q ultrapure water (Merck Millipore). Samples were boiled for 10 minutes at 95°C then stored on ice. The following primers were prepared at a 10 μM working stock concentration:

Forward primer: 5'-ATGGATAAGCTGAACGAC-3'
Reverse primer: 5'-TTACTGTTTGTATTTGTTGC-3'

PCR reaction mixtures were prepared as per Table 5.2. Positive control was 1 μL pET100-D/TOPO plasmid with *WIGMOR_0593* insert, negative control was 1 μL Milli-Q water. PCR was run in a thermal cycler (Bio-Rad) as per conditions detailed

in Table 5.3.

Table 5.2: Colony PCR reaction mixture.

Component	Volume (μL)
5x GoTaq reaction buffer (Promega)	10
Forward primer (10 μM) (IDT)	1
Reverse primer (10 μM) (IDT)	1
Milli-Q water	35.75
dNTP (10 mM) (Invitrogen)	1
GoTaq polymerase (Promega)	0.25
Colony in Milli-Q	1

Table 5.3: Colony PCR cycle. Denaturation, annealing and extension cycle repeated 35 times.

Stage	Temperature (°C)	Time
Initial denaturation	95	2 minutes
Denaturation	95	30 seconds
Annealing	55	30 seconds
Extension	72	1 minute
Final extension	72	5 minutes

5.2.5 Agarose gel electrophoresis

A 1% agarose gel was made using 500 mg agarose (Melford) and 50 mL Tris-Borate-EDTA (TBE) buffer (Table 5.1). The solution was microwaved for one minute to dissolve the agarose, then 5 μL SYBR Safe (Invitrogen) was added when the solution was warm to touch. Solution was poured into a gel tank to set before 50 mL TBE buffer was poured on top. Five μL of PCR product was added per well. A 5 μL volume of 1 Kb Plus ladder (Bioline) was used. PCR products were then separated using gel electrophoresis (Bio-Rad), run at 75 V for 50 minutes. Bands were visualised under ultraviolet light.

5.2.6 Sodium dodecyl sulfate polyacrylamide gel electrophoresis

A 1 mL sample from the expression trials was centrifuged for five minutes at 10,000 rpm in a microcentrifuge. A solution of BugBuster (Millipore) was made using 100 μL of 10x BugBuster and 900 μL PBS. The pellet was resuspended in BugBuster, with the pellet that had the midpoint OD value resuspended in 200 μL BugBuster

and the other pellets resuspended in a volume corrected to their OD. Samples were gently agitated for 30 minutes before being centrifuged for 10 minutes at 10,000 rpm in a microcentrifuge. The soluble and insoluble fractions were split, with 9 μL of the soluble fraction sample added to 3 μL of sample buffer (Table 5.1). For the purified samples, 9 μL of sample was added to 3 μL of sample buffer with no prior processing.

Sodium dodecyl sulfate (SDS) polyacrylamide gels were made according to the recipes in Table 5.4. Briefly, 4.5 mL resolving gel was pipetted between clean glass gel moulds and 1 mL ethanol added on top. The resolving gel was left to polymerise, then the ethanol was removed and the gel washed with Milli-Q water. Stacking gel was then added to just below the surface, a comb inserted, and the gel allowed to polymerise.

A 10 μL volume of the sample was loaded into each well and the tank filled with running buffer (Table 5.1). A 5 μL volume of PageRuler pre-stained protein ladder (Thermo Scientific) was also used. SDS polyacrylamide gel electrophoresis (SDS-PAGE) at 200 V for approximately 60 minutes on a Bio-Rad powerpack was used to separate the protein. After SDS-PAGE, the gels were stained with Coomassie Brilliant Blue (Table 5.1) and incubated on an orbital shaker for 45 minutes. The stain was removed and a detain (Table 5.1) added until the gel had destained and the bands were visible.

Table 5.4: Recipes for two SDS polyacrylamide gels.

Component	Volume
Resolving gel	
30% w/v acrylamide/0.8% w/v bisacrylamide	4 mL
1.5 M Tris (pH 8.8)	2.5 mL
10% SDS	100 μL
Milli-Q	3.4 mL
10% w/v APS in Milli-Q	100 μL
TEMED	10 μL
Stacking gel	
30% w/v acrylamide/0.8% w/v bisacrylamide	650 μL
500 mM Tris (pH 6.8)	1.25 mL
10% SDS	50 μL
Milli-Q	3.05 mL
10% w/v APS in Milli-Q	50 μL
TEMED	5 μL
Sample buffer (Table 5.1)	10 μL

5.2.7 Expression of *WIGMOR_0593*

Cultures of the transformed BL21 (DE3) cells from a glycerol stock were set up in 10 mL LB with 100 μg/ml ampicillin and incubated overnight at 37°C with aeration. This was used to inoculate 1 L LB/ampicillin to an OD_{650} of approximately 0.05. To optimise the expression of *WIGMOR_0593*, the plasmid was transformed into competent BL21 (DE3) and Tuner strains of *E. coli* and grown at either 30°C or 37°C in LB. Expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) once the cells had reached an OD_{650} of 0.2. Cells were harvested after three hours and centrifuged at 5000 rpm for 25 minutes using the JLA-8.1000 rotor on a Avanti J26 ultracentrifuge, before being resuspended in 35 mL lysis buffer (Table 5.1). Cells were then frozen at -80°C.

5.2.8 Sonication

Cells stored at -80°C were thawed, and 36 μL $MgCl_2$ (1000x stock), 360 μL PMSF protease inhibitor (100x stock), and a small amount of DNase powder (Merck) were added. Cells were incubated on ice prior to and during sonication. The sample was sonicated at 60 W for three seconds process and seven seconds rest for a total of three minutes using a Sonicator 3000 (Misonix). Cells were then centrifuged at 4°C for 25 minutes at 15,000 rpm using a Avanti ultracentrifuge and the JA 25.50 rotor, and stored on ice until purification.

5.2.9 Native purification

Protein was purified using a nickel affinity column on an AKTA Pure chromatography system (GE Life Sciences) using the method described in Table 5.5. Samples were separated by SDS-PAGE to confirm the presence of purified protein.

Table 5.5: Native nickel affinity purification of a his-tagged protein. CV = column volume, FV = fraction volume. Buffer compositions are defined in Table 5.1.

Method	AKTA setting
Transfer lines into 20% ethanol, perform a system wash	Standard manufacturer setting
Transfer all lines into dH_2O, perform a system wash	Standard manufacturer setting
Line 1 into KPi wash buffer	N/A
Line 2 into sample	N/A
Line B into elution buffer	N/A
Load 5 mL column, dripping dH_2O through the sample line	3 mL/min, 0.5 MPa
Equilibrate using KPi wash buffer (line 1)	5 mL/min, 0.5 MPa, 20 CV
Load sample on to the column (line 2)	5 mL/min, 0.5 MPa
Wash unbound protein using KPi wash buffer (line 1)	5 mL/min, 0.5 MPa, 10 CV
Elute bound protein with elution buffer (line B)	5 mL/min, 0.5 MPa, 3 mL FV, 5-10 CV

5.2.10 Denaturing purification

Denaturing purification was carried out on a nickel affinity column using an AKTA Pure chromatography system (GE Life Sciences) and guanidine hydrochloride gradient. The process is detailed in Table 5.6.

Table 5.6: Denaturing nickel affinity purification of a his-tagged protein. CV = column volume, FV = fraction volume. Buffer compositions are defined in Table 5.1.

Method	AKTA setting
Transfer lines into 20% ethanol, perform a system wash	Standard manufacturer setting
Transfer all lines into dH2O, perform a system wash	Standard manufacturer setting
Line 1 into KPi wash buffer	
Line 2 into sample	
Line B into elution buffer	
Load 5 mL column dripping dH_2O through the sample line	3 mL/min, 0.5 KPa
Equilibrate using KPi wash buffer (line 1)	5 mL/min, 0.5 MPa, 20 CV
Load sample on to the column (line 2)	5 mL/min, 0.5 MPa
Line 1 into denaturing wash buffer	
Wash with a gradient of denaturing wash buffer (line 1)	5 mL/min, 0.5 MPa, 30 CV
Elute bound protein with elution buffer (line B)	5 mL/min, 0.5 MPa, 3 mL FV, 5-10 CV

5.2.11 Buffer exchange

Buffer exchange was carried out on a 10 mL desalting column using the AKTA Pure chromatography system. A 2 mL sample loop was cleaned with 20 mL ethanol and 20 mL distilled water. A 2 mL volume of protein sample was loaded and fractionated into 500 μL volumes using storage buffer (Table 5.1).

5.2.12 Calculating protein concentration

Protein was concentrated using 10 kDa MWCO centrifugal filters (Sartorius). Filters were washed with 1 mL storage buffer (Table 5.1) at 4500 rpm and 4°C for 10 minutes, using the 19776-H rotor of a Sigma centrifuge. The flow through was removed and the elution fractions added. Samples were centrifuged at 4500 rpm and 4°C until concentrated sufficiently. The absorbance of the sample at 280 nm was measured and used with the protein extinction coefficient to calculate final concentration.

5.2.13 Circular dichroism spectrophotometry

Circular dichroism (CD) was performed under the supervision of Andrew Leech (York Technology Facility) to confirmed the correct refolding of WIGMOR_0593 following denaturing purification. A concentration of 0.2 mg/mL of WIGMOR_0593 in storage buffer (Table 5.1) was used. CD spectra were calculated at room temperature using a 1 mm path length quartz cuvette. A J810 CD spectrophotometer (Jasco) was used to measure CD spectra from 195 - 260 nm with a 1 nm data pitch, 2 nm band width, 2 second detector response time and a scanning speed of 100 nm per minute.

5.2.14 Peptide mass fingerprinting

Protein bands of interest were cut from an SDS-PAGE gel using a scalpel. The identity was confirmed by peptide mass fingerprinting (PMF), performed by the Rachel Bates (York Technology Facility). Briefly, a trypsin digest was used to generate peptide fragments. The fragments were analysed by matrix-assisted laser desorption ionization and tandem time-of-flight (MALDI-TOF/TOF) to calculate masses from which to compare to the reference WIGMOR_0593 sequence. The sequences were analysed in MASCOT (Matrix Science).

5.2.15 Crystallisation trials

PEG/Ion and JCSG crystal trays (Hampton Research) were set up under the supervision of Reyme Herman (University of York) using 10 μg/mL and 5 μg/mL of WIGMOR_0593 that had been dialysed overnight in Tris buffer (Table 5.1). Sam-

ples were centrifuged for one minute at 14,000 g in a microcentrifuge to remove any precipitate. A 3 μL sample of protein was added into eight guide wells, then 150 nL of sample and 150 nL Tris buffer (Table 5.1) was added per well of pre-set crystal trays (Hampton Research). Wells were checked periodically for the presence of crystallised protein. Candidate crystals were fished, and x-ray diffraction to assess candidate protein crystals performed by York Structural Biology Laboratory (YSBL) staff.

5.2.16 Ligand binding assays

Thermal shift assays were used to investigate amino acid binding. WIGMOR_0593 was dialysed overnight in potassium phosphate (KPi) buffer (Table 5.1) and tested at 3 μM concentration with 3 mM ligand (L-proline, L-arginine, L-serine, L-glutamate). SYPRO Orange dye (Applied Biosystems) was used. Assay was run for one hour using a StepOnePlus real-time PCR machine (Thermo Fisher Scientific). Eight replicates were taken for each conditions. D-glucose was used as a non-amino acid ligand control. Protein-only, ligand-only, and buffer-only controls were also used.

5.3 Results

5.3.1 Identification of significant genetic drift within the *W. glossinidia* species

A candidate SBP was identified in *W. glossinidia*. This protein is annotated as FliY in *W. glossinidia brevipalpis*, and WIGMOR_0593 in *W. glossinidia morsitans*. The *E. coli* FliY is specific for L-cystine (Butler et al., 1993), but it is not known whether it has been annotated correctly in *W. glossinidia*. It was discovered using protein BLAST that there is only 66.8% identity between FliY and WIGMOR_0593. It was not known whether this low sequence identity was unique to this protein, or whether it is indicative of genetic drift between the two subspecies.

To investigate this, the protein sequences of a group of core proteins, including DNA and RNA polymerase subunits and components of the flagella biosynthesis pathway, were used in a set of two-way comparisons; *W. glossinidia morsitans* against *W. glossinidia brevipalpis*, *Sodalis glossinidius* against *Sodalis praecaptivus*, *E. coli* against *E. fergusonii*, and *E. coli* against *Salmonella enterica subspecies enterica*

serovar Typhi. A selection of inner membrane, outer membrane, and cytoplasmic proteins were selected. It was hypothesised that the protein sequence identity between the *W. glossinidia* subspecies would be greater than the intra-genus comparisons.

As anticipated, the protein similarity between *Sodalis* species and between *Escherichia* species is high overall, with proteins involved in the formation of RNA polymerase conserved at over 97% identity (Fig. 5.4). Intriguingly, the protein sequence identity between *W. glossinidia morsitans* and *W. glossinidia brevipalpis* was found to be both low and variable. Some, like the elongation factor Tu (TufA), have a high sequence identity (94%). The RNA polymerase subunit proteins have sequence identities ranging from 75-90% between the two *W. glossinidia* subspecies, whereas the lowest value for any other comparison within this group is 98% (RpoA and RpoH for the *Sodalis* comparison, and RpoD for the comparison of *E. coli* to *Salmonella* ser. Typhi). In contrast, proteins that form DNA polymerase subunits (DnaENQX, HolAB, PolA) have low sequence identities between the *W. glossinidia* subspecies. These values range from 64% (DnaQ) to as low as 40% (HolA). They are variable in the other three comparisons, but the lowest value there is calculated to be 77% (DnaX between *S. glossinidius* and *S. praecaptivus*). The sequence for the outer membrane protein OmpF is highly conserved between *S. glossinidius* and *S. praecaptivus*, and between *E. coli* and *E. fergusonii*, with percentage identities of 96% and 94%, respectively. There is 59% sequence identity between *E. coli* and *Salmonella* ser. Typhi OmpF, but the identity between the two *W. glossinidia* subspecies is even lower than this inter-genus comparison, with only 40% sequence identity. These data suggest that the *W. glossinidia* strains, at the protein level, are not as similar as had been anticipated previously.

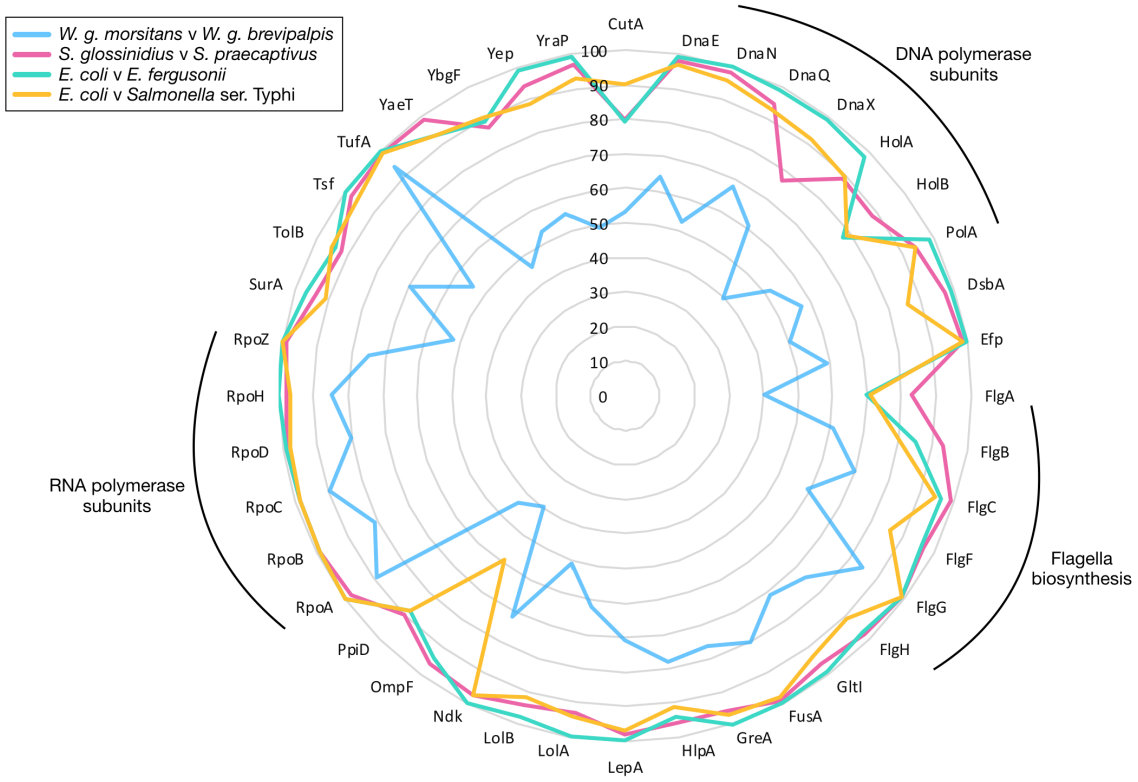


Figure 5.4: Protein sequence similarity depicted as percentage identity. Key protein pathways are highlighted. Comparisons between *W. glossinidia morsitans* and *W. glossinidia brevipalpis* (blue), *S. glossinidius* and *S. praecaptivus* (pink), *E. coli* and *E. fergusonii* (green), and *E. coli* and *Salmonella* ser. Typhi (orange).

5.3.2 Phylogenetic analysis of ABC transporters in the Enterobacteriaceae

A candidate amino acid binding protein is required in order to provide evidence for the ability of *W. glossinidia* to use L-proline as a sole energy source. *W. glossinidia*, however, lacks an obvious orthologue to known L-proline ABC transporters. It may be possible to use the sequences of FliY and WIGMOR_0593 to predict their binding specificity. A maximum likelihood phylogenetic tree was therefore constructed to investigate the phylogeny of amino acid binding proteins in the Enterobacteriaceae (Fig. 5.5). Known *E. coli* K-12 sequences served as protein BLAST search queries and their positions on the tree are indicated (*). The 503 proteins come from 35 different species. Some genera have multiple representations, including *Klebsiella*, *Sodalis*, and *Citrobacter*. In all instances, the closest orthologue of the *E. coli* proteins was found to be from *Shigella flexneri*. This is unsurprising as the *Shigella* genus is thought to be a subclade of *E. coli* (Luria and Burrous, 1957; Sims and Kim, 2011). YhdW is a putative *E. coli* binding protein of unknown specificity.

The YhdW clade is small, containing several highly similar proteins from *S. flexneri* and a putative *S. praecaptivus* protein, amongst others. The *E. coli* GlnH protein binds to L-glutamine, and it clusters close to a group of 11 identical proteins. The clade that contains the L-proline binding protein ProX also displays similar proteins found in the three species of the *Sodalis* genus, but intriguingly, given the premise of this work, not in the *W. glossinidia* subspecies. FliY (also called TcyJ) is an *E. coli* cystine binding protein (Butler et al., 1993). There are similar proteins found in *Klebsiella* and *Citrobacter* species. The three *Klebsiella* species also have highly similar proteins to the *E. coli* L-glutamate binding protein GltI. This protein was also a candidate of interest given the reversible relationship between L-glutamate and L-proline (Fig. 5.1). HisJ (L-histidine) and ArgT (L-lysine, L-arginine, L-ornithine) cluster closely to each other and to ArtI (putative). This is perhaps unsurprising, given the close sequence similarity between HisJ, ArgT, ArtI, and the L-arginine binding protein ArtJ. This therefore serves as a reference from which to infer binding specificities of amino acid binding proteins.

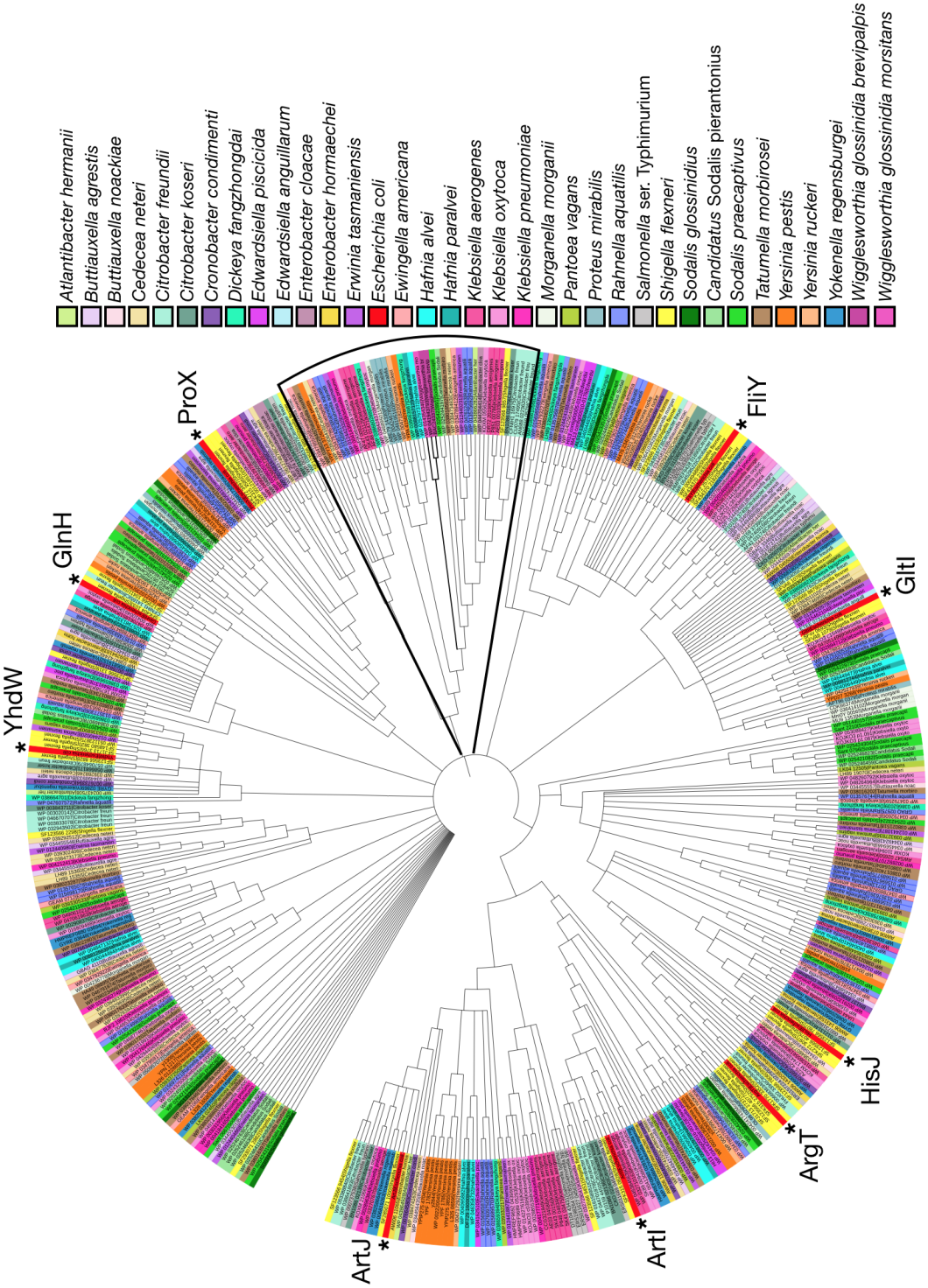


Figure 5.5: A maximum likelihood phylogenetic tree of amino acid binding proteins in the Enterobacteriaceae, found using protein BLAST with *E. coli* K-12 proteins as the search query. The position of the *E. coli* proteins are highlighted with an asterisk (*) and protein name. The cluster containing the *W. glossinidia* proteins is highlighted (and expanded in Fig. 5.6), with the node containing the *W. glossinidia* proteins labelled in bold.

5.3.3 The *W. glossinidia* SBPs are within a clade of unknown binding specificity

The clade containing the *W. glossinidia* proteins was then examined. If the *W. glossinidia* genome is indeed a subset of *E. coli*, it can be assumed that an *E. coli* SBP protein will be present within the *W. glossinidia* clade. The *E. coli* proteins locate around the tree, but several clades show no *E. coli* orthologues. This illustrates the diversity of subfamilies within the Enterobacteriaceae. One of these clades contains the candidate binding proteins from the two *W. glossinidia* subspecies (Fig. 5.6). The confidence in this varies between the sequences, but bootstrap values of one are calculated for the *W. glossinidia* sequences. This clade contains several other species of interest, including orthologues in *Candidatus* Sodalis pierantonius and *S. praecaptivus*, but not *S. glossinidius*. The NCBI entries for several proteins suggest a range of possible amino acid specificities, including L-glutamine (the *S. flexneri* protein SF123566_9857, WP_043866282 from *Atlantibacter hermannii* and WP_047612453 from *Rahnella aquatilis*) and L-cysteine (KOXM_05562 from *Klebsiella oxytoca*). Other proteins in this clade are accompanied by minimal information, with many described only as ABC transporter binding proteins. These *W. glossinidia* proteins therefore define a clade of SBPs with unknown binding specificity.

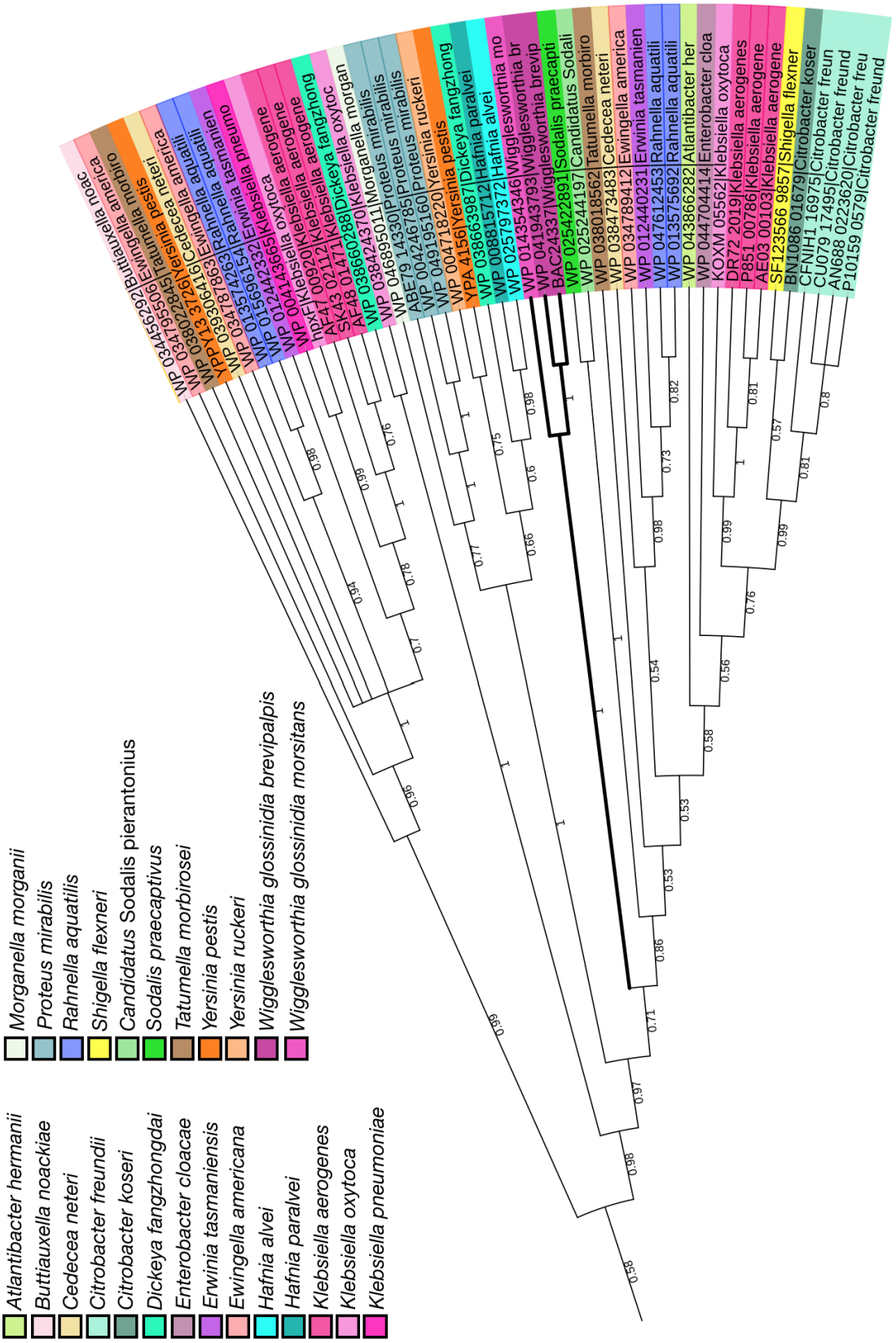
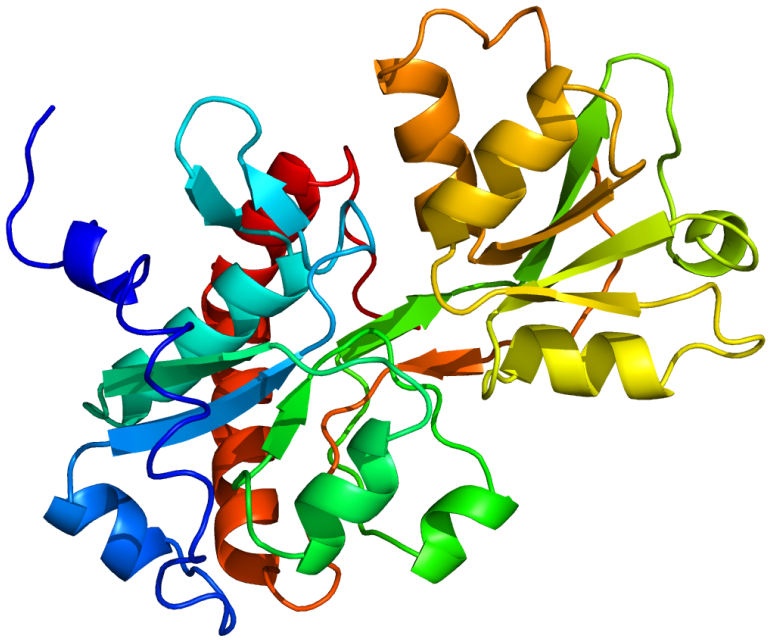


Figure 5.6: The cluster containing the candidate *W. glossinidia* SBPs. Number labels indicate bootstrap values from 100 replicates.

5.3.4 Expression and purification of the SBPs from *W. glossinidia*

Phylogenetics suggested that the SBPs in *W. glossinidia morsitans* and *W. glossinidia brevipalpis* are likely to be amino acid-specific. It was however unable to assist in identifying which amino acids they may be specific for. One of these proteins, WIGMOR_0593 from *W. glossinidia morsitans*, was subsequently selected for *in vitro* characterisation. Its orthologues in *W. glossinidia brevipalpis* and *Ca.* Sodalis pierantonius were also evaluated, producing comparable results. Only data from the WIGMOR_0593 investigation is presented here. WIGMOR_0593 is predicted to be a small SBP consisting predominately of α-helices (Fig. 5.7, top), with 264 amino acid residues (Fig. 5.7, bottom).



1	MFTKFFLNLS	RQITLLIILL	IPLKIVFADK	LNDIKKSGIL	KVAIFDSNPP
51	FGKIDFKSHQ	IIGYDADIAQ	ELAKMLKVQL	QLISTNPSNR	IPLLQSGKVD
101	LIIADITITP	ERKKVINFSI	PYFSTKQKIL	THISSSDNLQ	DYAKERIGVV
151	KGTTGQQTLS	HRFPEAKIIA	YNDIPMAFSA	LRNKNVKAIT	QDSTILEGLI
201	FGAPDRNQYK	ILSNSISEET	IGIGVKKGEE	RLLNEINKNL	MKLENSGIKN
251	NIYNKWFNLN	KYKQ			

Figure 5.7: The predicted structure of the small periplasmic binding protein WIGMOR_0593 produced by Phyre2 **(top)**. Reference sequence for WIGMOR_0593 SBP. **(bottom)**. Predicted signal peptide is underlined, matched peptides in PMF highlighted in red.

Phobius (Käll et al., 2004) and SignalP (Juan et al., 2019) programmes established that WIGMOR_0593 contained a signal peptide (Fig. 5.8). There are no cysteine residues found in the sequence. The signal peptide was therefore removed during the design of the synthetic gene. This enabled it to be expressed in high concentrations in the cytoplasm of the expression strain of *E. coli* (Maqbool et al., 2011). The resulting peptide contained 237 amino acid residues, and had a predicted molecular weight of 26.7 kDa.

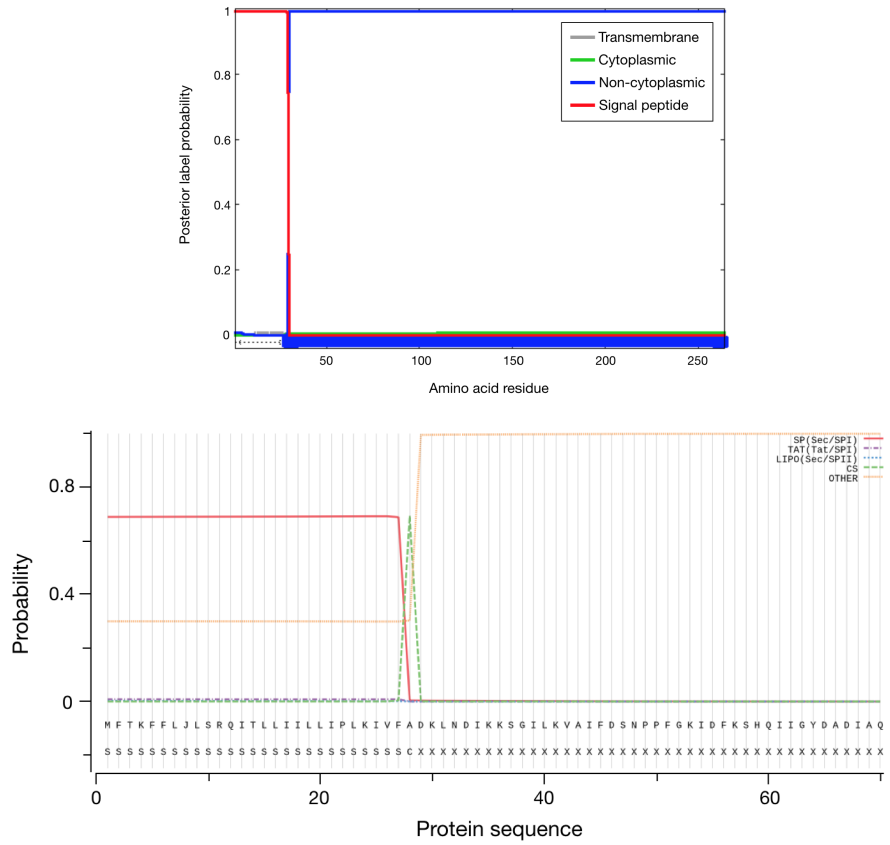


Figure 5.8: WIGMOR_0593 signal peptide prediction by Phobius **(top)** and SignalP **(bottom)**.

The pre-cloned WIGMOR_0593 protein was transformed into BL21 (DE3) and Tuner strains of *E. coli* and expressed under different conditions. BL21 (DE3) *E. coli* was grown at 37°C in LB, and 37°C and 30°C in autoinduction media (AIM) (Table 5.1). The Tuner strain of *E. coli* was grown at and 37°C and 30°C in AIM. Samples were collected at various time points and analysed for the presence of WIGMOR_0593. The expression trials did not produce large amounts of proteins under any of the conditions tested (representative SDS-PAGE images shown in Fig. 5.9).

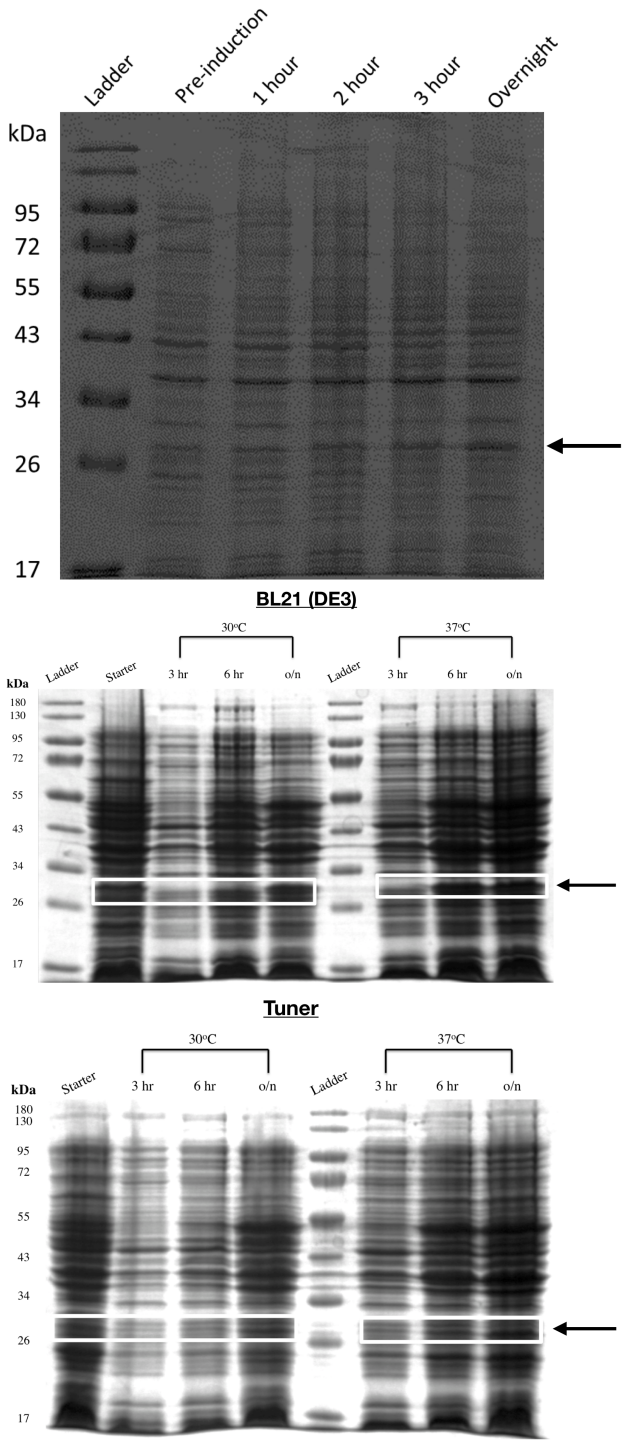


Figure 5.9: Expression of WIGMOR_0593, expected molecular weight 26.7 kDa. An SDS-PAGE gel representative of expression trials in various conditions. BL21 (DE3) in LB growth medium at 37°C **(top)**. Samples taken pre-induction with IPTG then at one, two, and three hours post-induction. An overnight sample was also taken. BL21 (DE3) at 30°C and 37°C in AIM **(middle)**, and Tuner at 30°C and 37°C in AIM **(bottom)**. Samples taken at three and six hours post-induction, and following an overnight incubation. Estimated position of the SBPs are marked with arrows.

It was hypothesised that a small amount of protein, not clear on the SDS-PAGE gels, was being produced. Purifying the sample on a nickel affinity column could isolate the his-tagged protein, even at low concentrations. A large-scale, 1 L trial was therefore conducted in LB using the BL21 (DE3) strain at 37°C. The cells were induced by the addition of 1 mM IPTG, harvested at three hours post-induction, and purified by nickel affinity purification. Purified protein at a molecular weight of approximately 26 kDa was then observed (Fig. 5.10, top). FliY from *W. glossinidia brevipalpis* and a close orthologue from *Ca.* Sodalis pierantonius, found using the phylogenetic tree (Fig. 5.6), were also expressed and purified using the same protocol (Fig. 5.10, centre, bottom). The functional assays for these two proteins were inconclusive, and therefore for clarity only the downstream data for WIGMOR_0593 will be presented here.

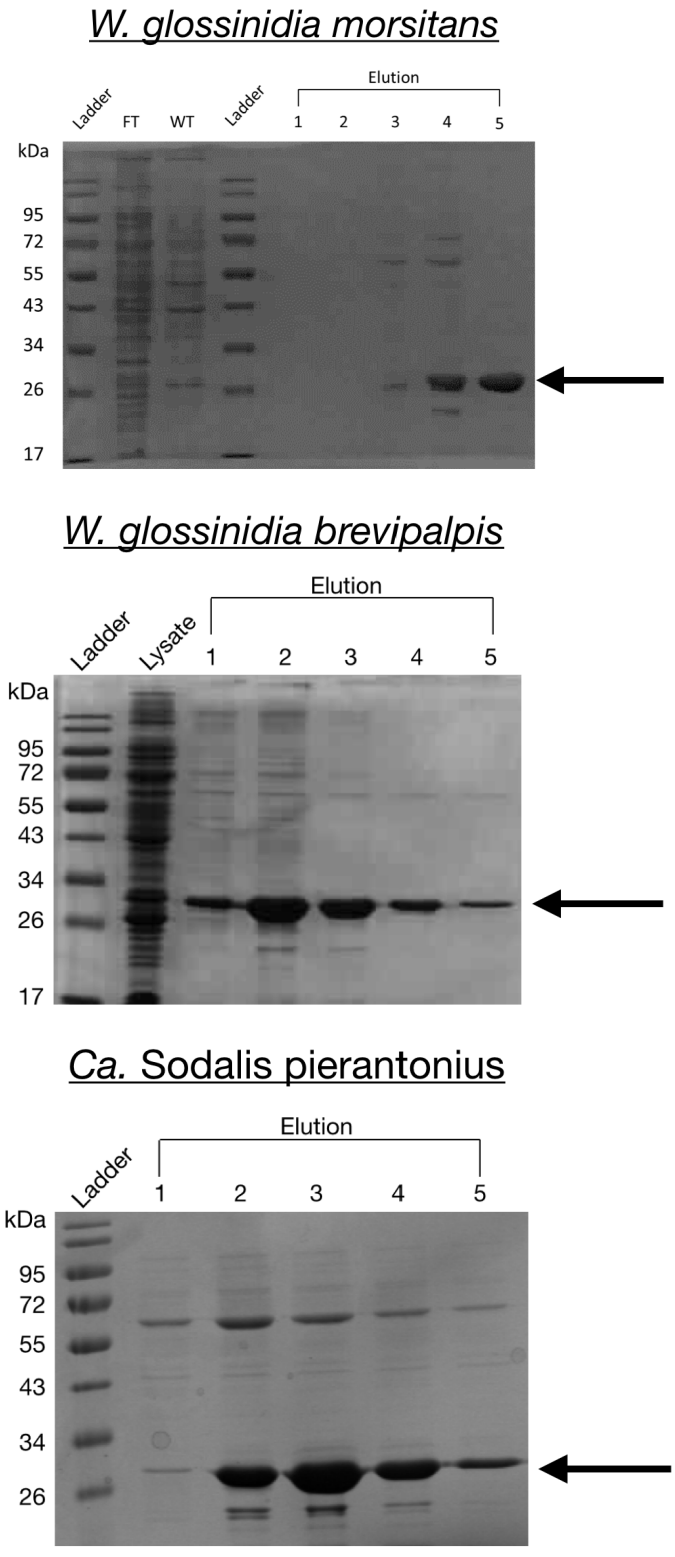


Figure 5.10: An SDS-PAGE gel showing purified protein from *W. glossinidia morsitans* **(top)**, *W. glossinidia brevipalpis*, **(centre)**, and *Ca.* Sodalis pierantonius **(bottom)** in five elution fractions following nickel affinity purification. Expected molecular weight 26.7 kDa, 26.7 kDa, and 26.2 kDa, respectively. FT = flow through, WT = wash through.

To confirm the identity of the protein, the band of interest was cut from the SDS-PAGE gel (Fig. 5.10, top) and peptide mass fingerprinting (PMF) subsequently performed by the York Technology Facility. This achieved 33% coverage of the WIGMOR_0593 reference sequence (Fig. 5.7, bottom, residues marked in red), producing seven peptide fragments (Table 5.7). From this, it was predicted that the eluted protein was a match to the candidate *W. glossinidia morsitans* SBP.

Table 5.7: PMF results showing the start and end amino acid from the reference sequence, and the observed, expected (expt) and calculated (calc) relative molecular mass (Mr) of the peptide fragment. The sequence of the fragment is given, with **bold** type indicating the oxidation of a residue.

Start residue	End residue	Observed	Mr (expt)	Mr (calc)	Peptide
42	53	1291.6151	1290.6078	1290.6608	K.VAIFDSNPPFGK.I
78	90	1469.7402	1468.7330	1468.7998	K.VQLQLISTNPSNR.I
99	112	1568.8246	1567.8173	1567.8821	K.VDLIIADITITPER.K
129	144	1804.8098	1803.8025	1803.9003	K.ILTHISSSDNLQDYAK.E
168	182	1694.8124	1693.8051	1693.8861	K.IIAYNDIPMAFSALR.N
168	182	1710.8146	1709.8073	1709.8810	K.IIAYNDIP**M**AFSALR.N
188	206	2016.9541	2015.9468	2016.0528	K.AITQDSTILEGLIFGAPDR.N

To ensure that the purified protein was folded correctly and to release any pre-bound ligand, a 1 L expression was performed as described previously, but with a denaturing purification achieved by a guanidine hydrochloride (GHCl) gradient (Fig. 5.11, top). A buffer exchange was then performed to remove the glycerol and imidazole (Fig. 5.11, bottom). This refolded protein was then analysed by circular dichroism (CD) to compare the refolded structure to the native protein sample. The α-helices and β-sheets that form the structure of a protein lead to characteristic signatures, known as CD spectra, when exposed to circularly polarised light. The CD spectra are highly similar between the two samples (Fig. 5.12). Both show traces at 210 - 220 nm that are characteristic of a protein with several α-helices, reflecting the structure predicted by Phyre2 from the protein sequence (Fig. 5.7, top). This suggests that they have the same structure, and that denaturing and refolding the protein does not impact its structure.

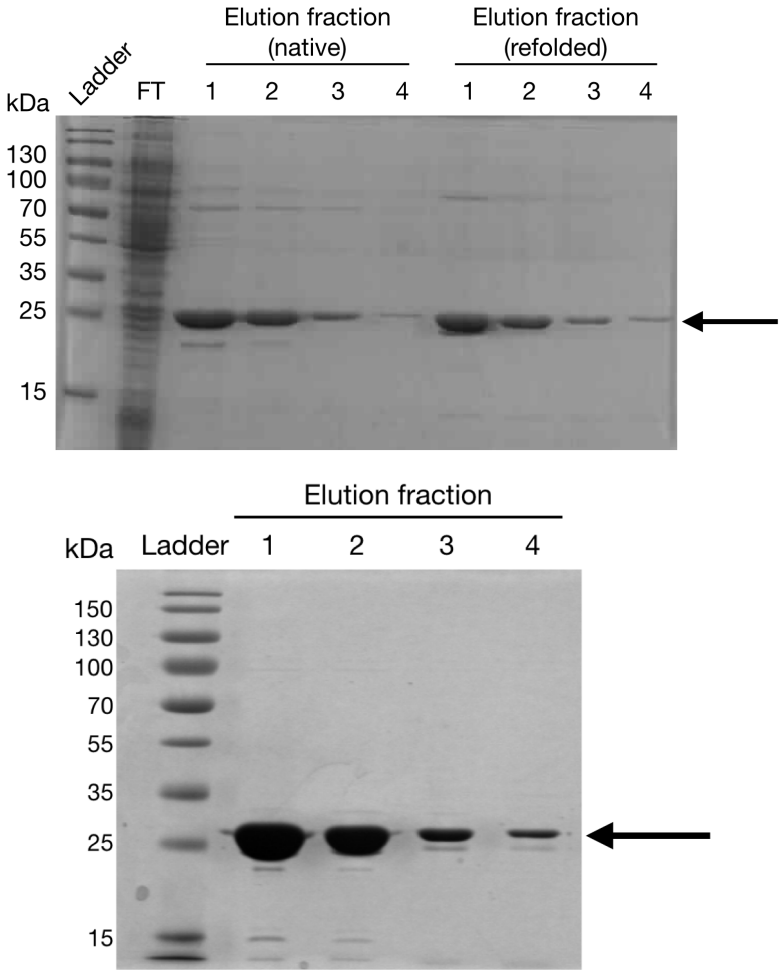


Figure 5.11: SDS-PAGE gel showing WIGMOR_0593 purified under native conditions **(top, left)**, and following denaturing with a GHCl gradient and subsequent refolded **(top, right)**. SDS-PAGE gel showing the refolded WIGMOR_0593 in four elution fractions following buffer exchange **(bottom)**.

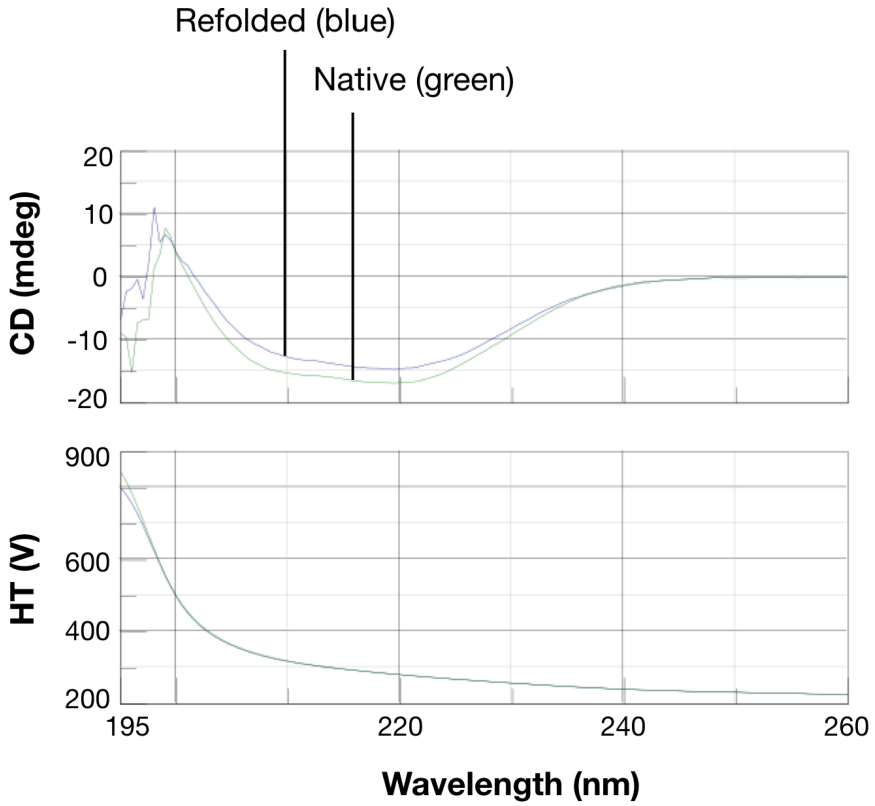


Figure 5.12: CD spectra of native (green) and refolded (blue) WIGMOR_0593. HT = high tension.

5.3.5 No evidence of L-proline binding by WIGMOR_0593

The potential of this protein to bind a selection of amino acids *in vitro* was then examined. Thermal shift assays were used as other techniques require the sequence to contain several tryptophan residues, whereas WIGMOR_0593 contains only one (Fig. 5.7, bottom). The melting temperature of the protein was measured. It was expected that the melting temperature would increase upon ligand binding as the protein becomes more stable (Huynh and Partch, 2015).

The melting temperature of WIGMOR_0593 was approximately 77.9°C in the absence of added ligand (Fig. 5.13). When L-proline, L-glutamate, L-arginine, or D-glucose were added at 1000 times the concentration of WIGMOR_0593, there was no difference in melting temperature, indicating an absence of ligand binding. The average melting temperature increased to approximately 79.5°C following the addition of L-serine, a shift of 1.6°C, suggesting the protein had become more stable. The theoretical use of L-serine by *W. glossinidia* was then examined. Sequences for the L-serine transporter SdaC, and the L-serine deaminases that can convert L-serine to pyruvate (TdcG, SdaA, TdcB, SdaB) in *E. coli* were taken from UniProt (Bateman

et al., 2019) and used as queries against *W. glossinidia morsitans* in a translated nucleotide BLAST search. An orthologue to SdaC was found, WIGMOR_0265. There was however no orthologues to any of the four L-serine deaminases. This suggests that *W. glossinidia* may not be able to metabolise L-serine, and therefore the thermal shift results should be treated with caution.

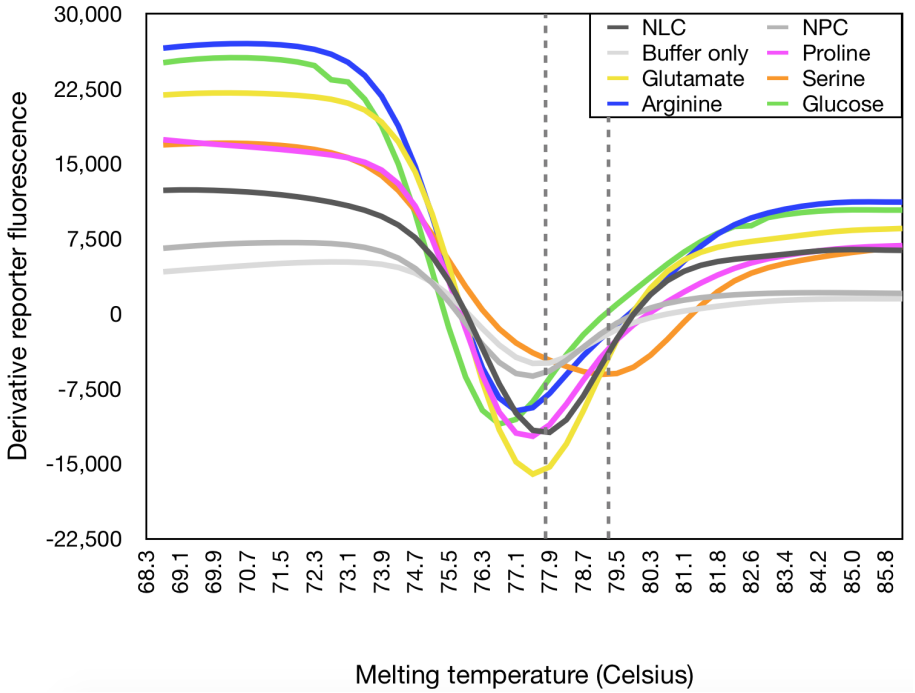


Figure 5.13: Melting temperature profile of WIGMOR_0593 in the presence of L-proline, L-glutamate, L-serine, L-arginine, and D-glucose. No-ligand controls (NLC), no-protein controls (NPC), and buffer-only controls are used. Dashed grey line shows approximate melting temperature of NLC (left) and of WIGMOR_0593 with added L-serine (right). Measurements an average of eight replicates.

5.3.6 Crystallisation trials

It is sometimes possible to determined the specificity of a binding protein by crystallising the native sample and analysing any co-crystallised, pre-bound ligand (Schlichting, 2005; Bhatt et al., 2010). An attempt was therefore made to crystallisation native samples of WIGMOR_0593. The protein was added to pre-cast JCSG and PEG/Ion crystallisation trays (Hampton Research). Protein crystals were observed in the JCSG (Fig. 5.14, top) and PEG/Ion (Fig. 5.14, centre, bottom) trays. These wells contain 10% w/v polyethylene glycol (PEG) 1000 with 10 μg/mL protein (top left), 4% v/v Tacsimate pH 6.0, 12% w/v PEG 3350 with 5 μg/mL protein (top right), and 200 mM ammonium citrate tribasic pH 7.0, 20% w/v PEG 3350 with 10 μg/mL protein (bottom). These crystals were fished, and initial in-house x-

ray diffraction subsequently performed by YSBL staff. None of the crystals tested diffracted at a high enough resolution to be sent for further testing. It has therefore not been possible to elucidate the binding specificity of this protein via crystallisation.

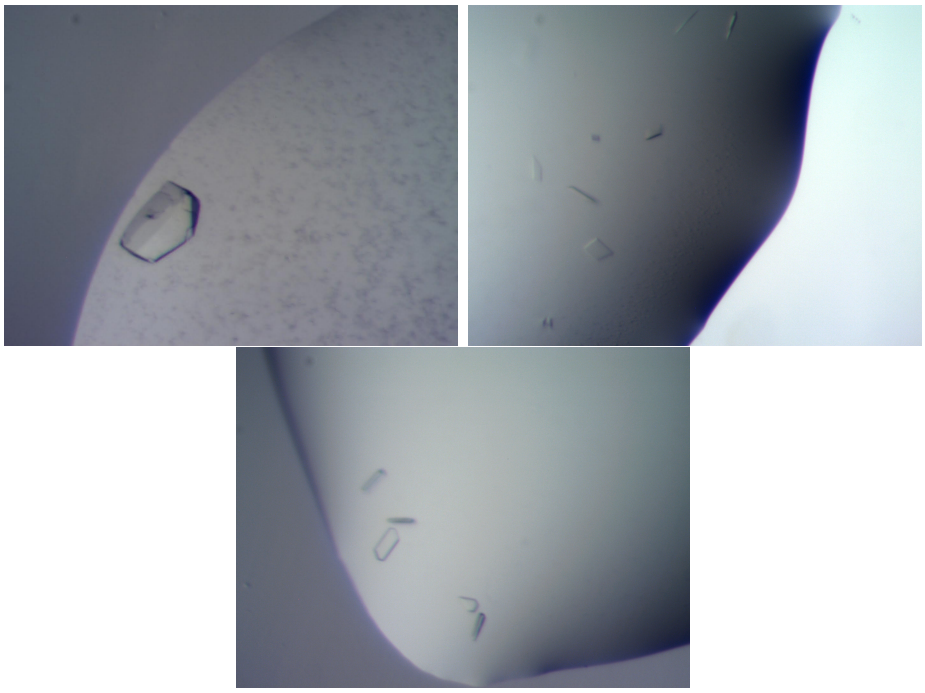


Figure 5.14: Examples from WIGMOR_0953 crystallisation trials in JCSG **(top left)** and PEG/Ion **(top right, bottom)** trays, from a starting concentration of 10 μg/mL **(top left, bottom)** and 5 μg/mL **(top right)** in Tris buffer.

5.4 Discussion

The association of *W. glossinidia* with the tsetse is ancient and obligate (Chen et al., 1999; Snyder and Rio, 2013). The symbiont is thought to provide B vitamins to its host (Michalkova et al., 2014; Snyder and Rio, 2015), but the exact nature of its own metabolic network remains to be elucidated. Building upon current knowledge of *W. glossinidia* metabolism and phylogeny will further not only the understanding of the tsetse microbiome, but also of the role of primary symbionts within their insect hosts.

The *W. glossinidia* genome was previously assumed to be a subset of the *E. coli* genome, as seen in the aphid primary symbiont *B. aphidicola* (Shigenobu et al., 2000). This assumption has not been tested previously, and places limits on understanding the metabolic adaptations of *W. glossinidia* fully. It is possible that by assuming it is highly similar to *E. coli*, important differences that have arisen as

a result of its association with the tsetse may be overlooked. It has been deduced here that the genome of *W. glossinidia* is not entirely a subset of *E. coli*. The candidate *W. glossinidia* amino acid binding proteins, when analysed phylogenetically, do not cluster with an *E. coli* orthologue. This contradicts what would have been anticipated if it was indeed a true *E. coli* subset.

It was also initially presumed that there would be minimal differences between the two subspecies of *W. glossinidia*. By comparing a range of cytoplasmic, periplasmic, and membrane-bound proteins from *W. glossinidia morsitans* and *W. glossinidia brevipalpis*, it was found that, at the protein sequence level, these two organisms are more dissimilar than equivalent comparisons between *E. coli* and *Salmonella* ser. Typhi (Fig. 5.4). This indicates that significant genetic drift has occurred between the *W. glossinidia* subspecies during their spatial separation. The association of *W. glossinidia* with the tsetse is estimated to be 80 million years old (Chen et al., 1999; Snyder and Rio, 2013). Population bottlenecking over prolonged periods of evolutionary time may have allowed the *W. glossinidia* subspecies to diverge (Moran, 1996; Muller, 1964; Pettersson and Berg, 2007). It also suggests that caution needs to be taken when extrapolating results from one subspecies to the other.

The first published expression and purification of a protein from *W. glossinidia* has been demonstrated here. Unpublished work has purified the *W. glossinidia* chorismate mutase/prephenate dehydratase PheA (Thomas lab, unpublished). Here, a pre-cloned synthetic gene was used due to the unculturable nature of the symbiont. The protein encoded by this gene, WIGMOR_0593, did not diffract to a high enough resolution to elucidate its structure. The binding assays suggest that, under these conditions, this protein is not able to bind L-proline and therefore cannot provide experimental support for the *W. glossinidia* FBA model. There is however tentative evidence that it might bind L-serine, highlighted in the small shift in melting temperature in the presence of this amino acid (Fig. 5.13). The protein that converts L-serine to pyruvate is missing in *W. glossinidia morsitans*. The lack of this reaction suggests that *W. glossinidia* may not be able to use L-serine as a sole source of carbon, nitrogen, and energy. It is however possible that *in situ* this protein may be capable of promiscuous binding to other amino acids.

This work highlights the issues with using standard microbiology and molecular biology tools to studying symbiotic bacteria. The use of synthetic genes circumvents the standard vector cloning route of using PCR to extract a gene of interest, something that is not possible without being able to grow the organism *in vitro*. This gene was pre-cloned into an expression vector, and although small-scale expression trials did not indicate protein over-expression there was evidence of this when a large

volume was used (Fig. 5.10). The CD spectra suggest that the refolded protein has the same structure as the native sample. If this was a standard protein from a free-living species of bacteria, it could be logical to assume that this reflects correct refolding. There are however possible issues with expressing symbiont proteins in *E. coli*. For example, high concentrations of the molecular GroEL produced by *W. glossinidia* have been measured in the tsetse midgut (Haines et al., 2002). GroEL assists protein refolding. It is possible that the lack of binding seen in the thermal shift assay is as a result of this SBP requiring an abundance of *W. glossinidia* chaperone to fold correctly. The CD spectra indicate that the refolded WIGMOR_0593 is similar to the native sample, not that it is in the conformation found *in vivo*. This should be considered in future work investigating symbiont proteins, and perhaps the co-expression of a symbiont chaperone to aid correct refolding should be investigated. The co-expression of *E. coli* GroEL has been suggested previously as a technique to improve the production of symbiont proteins (Fares et al., 2002).

The work presented in this chapter furthers current understanding about the phylogeny of SBP specificities in the Enterobacteriaceae. It highlighted the diversity of SBPs found in this family. Future computational work investigating the metabolism of *W. glossinidia morsitans* may uncover additional differences between the subspecies. Here, work has built upon the foundation provided by a *W. glossinidia brevipalpis* metabolic model. *W. glossinidia morsitans* is however arguably the more medically-relevant subspecies. *G. brevipalpis* is thought to be refractory to *T. brucei* infection (Moloo et al., 1994; Rotureau and Van Den Abbeele, 2013), whereas *G. morsitans morsitans* is susceptible. One of the key differences between *W. glossinidia brevipalpis* and *W. glossinidia morsitans* is that it is only the latter that can produce the vitamin folate (Snyder and Rio, 2015). Exposing *G. brevipalpis* to folate has been shown to increase the susceptibility of these insects to trypanosome infection (Rio et al., 2019). Constructing a metabolic model for *W. glossinidia morsitans* in order to compare to *W. glossinidia brevipalpis* may uncover additional differences between the two subspecies. Furthermore, the application of the evolutionary tool presented in Chapter 4 may enable the future of *W. glossinidia* as a symbiont to be examined. This is crucial for the tsetse microbiome as a whole, when the pivotal role played by *W. glossinidia* is considered.

Chapter 6

Discussion

The work presented in this thesis describes metabolic adaptations that have evolved within the tsetse microbiome. The response to metabolite availability, and how this has subsequently influenced the evolutionary trajectory of the symbionts, has been discussed in detail. In this chapter, key questions that have arisen as a result of the research presented here will be examined.

6.1 The Spectrum of Secondary Symbioses: A study in *Sodalis*

Classical definitions of primary and secondary symbionts have been discussed in depth. If considering *Sodalis glossinidius* in isolation, it is tempting to categorise it as an exemplar secondary symbiont. The *Sodalis* genus however, when viewed as a whole, exhibits huge variability in infection type and preference. This will be discussed in this chapter, and it raises several questions. Is *S. glossinidius* a typical secondary symbiont? If not, how does one define a secondary symbiont? Importantly, how is the *Sodalis* genus capable of existing not only within hugely variable insects, but externally in the environment as well?

When addressing these questions, of immediate note are the variable diets of the insects in which *Sodalis*-like symbionts have been identified, summarised in Chapter 1. Several, like the tsetse, are blood feeders, including the hippoboscoid fly *Craterina melbae* (Novakova and Hypsa, 2007) and the sheep ked *Melophagus ovinus* (Chrudimský et al., 2012). A *Sodalis*-like symbiont has been found in the slen-

der pigeon louse *Columbicola columbae* (Fukatsu et al., 2007), an insect that feeds on a keratin-rich diet of avian feathers. The majority identified to date, however, have been identified in insects with plant-based diets. Some are sap feeders, such as the stinkbug *Cantao ocellatus* (Kaiwa et al., 2010), the scale insect *Puto superbus*, (Szklarzewicz et al., 2018) or the green leafhopper *Cicadella viridis* (Michalik et al., 2014). Others gain nutrition from different parts of the plant, like the chestnut weevil *Curculio sikkimensis* (Toju et al., 2010; Toju and Fukatsu, 2011) or the longhorned beetle *Tetropium castaneum* (Grünwald et al., 2010). It is not yet known whether Sodalis infects plant feeders preferentially, or whether the interest in crop pests has skewed the available data. This is particularly pertinent when considering the phylogeny described by Boyd et al. (2016). They suggest that the closest relative to *S. glossinidius* was, at the time of writing and based on *groEL*, *ftsZ* and *dnaJ* sequences, a *Sodalis*-like symbiont they term PSPU, of the froghopper *Philaenus spumarius*. These symbionts must have shared a recent common ancestor, but their host insects have diets that are entirely different nutritionally. Interestingly, the vector of *T. cruzi*, causative agent of Chagas disease (American trypanosomiasis), is not thought to be infected by a *Sodalis*-allied organism (Gumiel et al., 2015; Díaz et al., 2016; da Mota et al., 2012). The *Arsenophonus* genus is also able to infect a broad array of insects with contrasting diets (Nováková et al., 2009), including blood (Hypsa and Dale, 1997; Šorfová et al., 2008; Grindle et al., 2003; Dale et al., 2006; Trowbridge et al., 2006), plant components (Ly Thao and Baumann, 2004a; Spaulding and Von Dohlen, 2001; Subandiyah et al., 2000; Hansen et al., 2007; Babendreier et al., 2007; Zreik et al., 1998; Karimi et al., 2019), and other insects (Gherna et al., 1991; Dunn and Stabb, 2005). This suggests that this feature is not unique to *Sodalis*, and that the ability to infect hosts with a broad range of diets could therefore be a key criterion to defining a secondary symbiont.

S. glossinidius is found extracellularly throughout the tsetse, whereas the obligate symbiont *Wigglesworthia glossinidia* is bacteriocyte-associated (Aksoy, 1995). In contrast to *S. glossinidius*, there are several *Sodalis*-like symbionts that have been found within the bacteriome. These insects include *P. superbus* (Szklarzewicz et al., 2018), *M. ovinus* (Chrudimský et al., 2012), *Ommatidiotus dissimilis* (Michalik et al., 2018), *C. columbae* (Fukatsu et al., 2007) and *C. viridis* (Michalik et al., 2014). In taking this route and becoming intracellular, the symbionts are likely to be more restricted in their ability to survive *ex vivo* and may therefore become obligate. *Candidatus* Sodalis melophagi, however, is bacteriome-associated but can be cultured *in vitro*, in contrast to *W. glossinidia* (Chrudimský et al., 2012). Some of these symbionts co-localise within the bacteriocytes with primary symbionts (Michalik et al., 2014). This hints at metabolite sharing between the two organisms, or the replace-

ment of a pre-existing primary symbiont. Comparing these *Sodalis*-like symbionts to the extracellular *S. glossinidius* suggests several points about the tsetse symbiont. *S. glossinidius* may be a more recent acquisition than those that are bacteriome-associated and may become intracellular later in evolutionary time. The evidence of *Candidatus* Sodalis pierantonius, symbiont of the rice weevil *Sitophilus oryzae*, implies however that *Sodalis*-allied symbionts can be both primary and a recent acquisition (Oakeson et al., 2014; Heddi et al., 1997). Alternatively, *S. glossinidius* may, as with the symbiont of *P. fluctus* (Boyd et al., 2016), provide the replacement for *W. glossinidia*. An argument against this lies in the dependence of *S. glossinidius* on *W. glossinidia* for the vitamin thiamine that the latter produces for the tsetse (Snyder et al., 2010). *S. glossinidius* is unable to synthesise thiamine but has retained an intact transporter (Chapter 2). Consequently, is not certain whether *S. glossinidius* would be capable of becoming the sole primary symbiont for the tsetse. It might instead progress down a spectrum of secondary symbioses as its genome becomes increasingly streamlined. It may not therefore be possible to define a secondary *Sodalis*-allied symbiont based on extracellular locality within the insect, or on the timing of its capture.

An alternative characteristic of secondary symbionts that might be used is their genome structure. *S. glossinidius* has a genome size comparable to free-living organisms, but a high proportion of pseudogenes (Dale and Maudlin, 1999). It has not yet undergone the extreme genome streamlining observed in typical obligate symbionts like *W. glossinidia* or the aphid primary *Buchnera aphidicola* (van Ham et al., 2003; Munson et al., 1991; Moran and Mira, 2001; Thomas et al., 2009). Curiously, *Ca.* Sodalis pierantonius has a genome size and structure consistent with *S. glossinidius*, but is believed to play the role of a primary symbiont. *S. glossinidius* also does not display an A+T bias in nucleotide composition, nor evidence of accelerated molecular evolution. Genetic information is not available for all *Sodalis*-like symbionts but there is no evidence for A+T biases for those that have been sequenced. An A+T bias is indicative of obligate symbiosis (Moran and Wernegreen, 2000). The symbionts identified at the time of writing may therefore be classified as secondary in this respect. It is however difficult to extrapolate the conclusions made from analysis of a limited number of genes (16S rRNA, *groEL*). Without complete genome sequences it is not possible to identify whether these organisms have undergone, or are in the process of, genome reduction.

Without using genetics to characterise a secondary *Sodalis* symbiont, another option is to examine their physiology and the role they play within their ecosystem. *S. glossinidius* has retained its ability to grow *ex vivo* in both rich and defined media (Dale and Maudlin, 1999; Toh et al., 2006; Hall et al., 2019; Beard et al., 1993;

Welburn et al., 1987). The benefit of *S. glossinidius* acquisition for the tsetse has not yet been elucidated (Toh et al., 2006); it is not thought to play a role in nutrient provisioning, for example. There may be a link between *S. glossinidius* and *T. brucei* infection, although this theory is still under construction (Farikou et al., 2010; Dale and Welburn, 2001; Geiger et al., 2006). An interesting observation of the *Sodalis*-like symbionts is the variation in their infection frequencies. The symbionts of *C. sikkimensis* and *C. ocellatus* have, for example, very low infection frequencies, indicating they may not provide an indispensable benefit to the insects (Kaiwa et al., 2010; Toju et al., 2010; Toju and Fukatsu, 2011). The pigeon louse *C. columbae*, in contrast, has a *Sodalis*-allied symbiont with over 95% infection frequency (Fukatsu et al., 2007), suggesting a more important role. From this, it cannot be said whether bacterial physiology or function could be used to define a secondary *Sodalis* symbiont explicitly.

S. glossinidius has been described previously as a model organism (Novakova and Hypsa, 2007). But a model for what is not clear. By comparing *S. glossinidius* to other *Sodalis*-like organisms, it is evident that *S. glossinidius* is neither a typical secondary symbiont nor a typical *Sodalis*. It is able to infect insects with completely contrasting diets, but it is not known what genomic or metabolic characteristics allow this promiscuity. The trajectory of *Sodalis* symbioses also remain to be elucidated. *Sodalis* may always begin its symbioses as a secondary, or extremely rapid genome reduction may result in the immediate occurrence of a primary symbiosis. It may be time to consider secondary symbiosis as a spectrum, rather than a discrete label. Acknowledging that so-called secondary symbionts may occupy different points on a continuous scale opens up possible roles within an ecosystem that may not have been considered otherwise.

6.2 When is a pseudogene not a pseudogene?

Through this work it has become clear that it is vital to test metabolic models thoroughly to avoid making incorrect assumptions about an organism. An example of this was shown in Chapter 3, with the discovery of the unusual ability of *S. praecaptivus* to metabolise xylitol and GalNAc. It is particularly important for symbiotic bacteria that have unusual genetic compositions. Through the testing of *i*LF517 detailed in Chapter 2, a key conclusion was made that genes assumed previously to be pseudogenised in fact appear to be part of functional pathways. This was underlined by the discovery that *S. glossinidius* could grow in the absence of L-arginine, in spite of an initial genome annotation categorising key genes in the

pathway from L-glutamate to L-arginine as pseudogenes (Toh et al., 2006; Belda et al., 2012).

This then raises the interesting question of how to characterise a pseudogene in symbiotic bacteria. It may be particularly challenging given the fast mutation rate and subsequent unusual genetic compositions in symbionts (Moran, 1996). Additionally, not all pseudogenes are created equally. Pseudogenes may be completely absent, or identifiable only by scars in the genome. This may be evidence of a gene lost early in the association, with a long period of time available for its erasure. Others may have become inactive as a result of single mutations, or several insertions or deletions. This may suggest that this particular pseudogenisation is a more recent one. Genes lost at an intermediate stage of the symbiosis may display large portions of the sequencing missing, either at the ends or in the middle. There have also been examples documented in *B. aphidicola* and *Blochmannia pennsylvanicus* in which the presence of frameshift mutations in poly(A) tracts, repeating occurrences of adenines within a sequence, can result in a gene being incorrectly annotated as a pseudogene (Tamas et al., 2008). This serves to highlight the care needed when analysing symbiont genomes. Determination of the completeness should not rely on the genome annotation but should instead use alignments to known full length genes. This should then be supported, where possible, by *in vitro* investigations, as demonstrated in Chapters 2 and 3.

It has been discussed here how symbiotic bacteria may be less amenable to genetic manipulation. One of the most widely used methods to investigate the function of a gene is to knock it out and measure the resulting phenotype. This requires the organism to be transformed, a process that may not be survivable for symbionts. Alternative techniques or proxies should therefore be considered. Here, the *E. coli* Keio collection (Baba et al., 2006) was used to examine whether the loss of single genes in the L-arginine biosynthesis pathway could cause an auxotrophic phenotype. It was found that the absence of *argD* did not prevent *E. coli* from growing without exogenous L-arginine. This suggested that there must be some redundancy in this role in *E. coli*, something that was then also discovered bioinformatically in *S. glossinidius* via the presence of alternative genes; *bio A* (*SG0902*) and *hemL* (*SG0500*). Another possible avenue for confirming pseudogenisations in symbiotic bacteria may be to synthetically clone the symbiont gene of interest into a knockout strain of *E. coli*. If the symbiont gene is unable to complement the lost gene, then this suggests a true pseudogenisation.

When the first *S. glossinidius* FBA model, *i*EB458, was published by Belda et al. (2012), it omitted several genes, presumed to be pseudogenised, that were subse-

quently proven functional following *in vitro* experimentation (Chapter 2). This then raises the possibility that other FBA models of symbiont metabolism may also have made incorrect assumptions about pseudogenes. This may be particularly true if the models have not been fully validated experimentally. For example, multiple models of the cyanobacterium *Synechocystis* sp. PCC 6803 have been published (Shastri and Morgan, 2005; Nogales et al., 2012; Knoop et al., 2010, 2013). Of these, one has used *in vitro* data to confirm gene essentiality in the network (Knoop et al., 2010). This is useful for verifying the reactions already included in the model, but does not directly consider those excluded during the construction process. This was improved in a model published more recently that uses transcriptome analysis for verification (Knoop et al., 2013). There is also a metabolic model for the aphid symbiont *B. aphidicola* (Thomas et al., 2009), as well as publications detailing the *B. aphidicola* genome and its evolution (Shigenobu et al., 2000; van Ham et al., 2003; Wilson et al., 2010). It is not possible to grow the obligate *B. aphidicola in vitro* in order to thoroughly test what has been inferred from its genome. In particular, the description of "pseudogene" by van Ham et al. (2003) is not defined. Some of the predictions made could be strengthened by using alternate techniques to verify key pseudogenisation. This could be done via the cloning of synthetic genes as discussed.

Only once metabolic models are tested and validated in depth can the predictions that they make be acknowledged with confidence. This is only possible when pseudogenes in symbionts have been classified accurately. This then leads to a subsequent consideration of the appropriateness of FBA when applied to symbiotic bacteria.

6.3 How good a tool is FBA?

FBA is a useful tool for laboratory strains and industrially relevant bacteria. It enables a spectrum of experiments that would not otherwise be feasible (Edwards et al., 2002; Lewis et al., 2012). An example of this has been presented Chapter 4, where it has been used to predict evolutionary trajectories within the *Sodalis* genus. It may however be limited intrinsically by its own assumptions when applied to symbiotic bacteria.

FBA uses linear programming to achieve an objective function. For prokaryotes, the objective function is often the production of biomass (a "biomass output") (Varma and Palsson, 1994). Through this, the metabolites within the system are converted to biomass (Segrè et al., 2002). By defining the objective function as biomass production, FBA assumes that the organism has maximised growth over

evolutionary time. Segrè et al. (2002) argue that this assumption is not appropriate for genetically engineering organisms, or for those that have not yet been exposed to evolutionary pressure for a long period of time. To overcome this, they developed the method of minimization of metabolic adjustment (MOMA) as an alternative (Segrè et al., 2002). MOMA is more suited for knockout mutants as it takes into consideration that a genetically modified organism may have flux distributions that are sub-optimal. In doing so, MOMA finds intermediate solutions that are less optimal than for wild-type organisms. This technique was developed in the context of genetically modified organisms, but it poses interesting questions about symbionts.

The symbiotic bacteria of insects are arguably not adapted to maximise biomass production. Insects actively control the size of their symbiont population (Login et al., 2011), and symbionts must avoid being immunogenic to prevent elimination (Weiss et al., 2008). This suggests that the objective function for symbionts may not be biomass production. Some symbionts have also not been exposed to prolonged periods of evolutionary pressure. *W. glossinidia* has been within the tsetse microenvironment for 80 million years (Chen et al., 1999; Snyder and Rio, 2013), and other primary symbionts are thought to have established over 200 million years ago (Moran et al., 1993; van Ham et al., 2003; Bennett and Moran, 2013). Other symbionts, like *S. glossinidius*, are more recent associations (Dale and Maudlin, 1999). It cannot be said therefore that all symbionts have been exposed to the prolonged evolutionary pressures that may be required to fulfil the assumptions of FBA.

Incorporating an allowance into the objective function may be a possible way to ensure that FBA is as applicable to symbiotic bacteria as possible. One possible way to do this is through regulatory on/off minimization (ROOM) (Shlomi et al., 2005). This algorithm predicts the steady state flux through a metabolic network following gene removal, and therefore may be more appropriate for symbionts that are undergoing extensive and rapid genome reduction. Alternatively, the objective function could be modified to reflect *in vitro* phenotypes, either measured or predicted (Schuetz et al., 2007). Acknowledging that symbiotic bacteria need to be treated differently from standard, free-living species of bacteria would hopefully lead to metabolic models that are more reflective of the symbionts.

6.4 The initiation of symbiosis

Symbioses are being studied increasingly intensively in insects that are of importance for disease control (Holt et al., 2002; Nene et al., 2007; Mellor et al., 2000) or

food security (Oerke, 2006). The majority of the research naturally focuses on the relationships as they appear in nature. This is largely due to the technical difficulties in altering essential relationships for the purpose of experimentation. Arguably, the events that led to the initiation of the symbiotic relationship are of equal importance. This includes estimating the nature of the earliest symbiotic ancestor, and understanding what adaptations host and symbiont would need to enable the relationship to progress and persist. Elucidating these details may open new avenues of research into harnessing these relationships for pest or vector control.

The model system for studying symbiosis induction is the Hawaiian bobtail squid *Euprymna scolopes* and the symbiotic, luminescent *Vibro fischeri* found in its light organ (Nyholm and McFall-Ngai, 2004). The colonisation by *V. fischeri* is advantageous from a technical perspective in that it is rapid and without co-infection by other microorganisms (Nyholm and McFall-Ngai, 2004). All interactions are therefore squid-*V. fischeri*-specific, allowing the roles of host and symbiont to be designated specifically (Koropatnick et al., 2007). The induction of symbiosis in this system has been split into three events; initiation, accommodation, and persistence (Ruby, 1996). This information is key to understanding this particular symbiotic relationship, but whether it can be extrapolated to other systems is unclear.

Symbiosis initiation has also been studied in microbe-microbe symbioses. The unicellular marine protozoan *Paramecium bursaria* harbours large numbers of symbiotic algae of the species *Chlorella* in its cytoplasm (Kodama and Fujishima, 2010b; Reisser, 1980). The discovery of a naturally aposymbiotic strain of *P. bursaria* (Tonooka and Watanabe, 2002) enabled subsequent experimentation that artificially induce a relationship between alga and protozoan (Kodama and Fujishima, 2009). This research identified some of the steps involved in the induction of symbiosis, and the requirement of *Chlorella* to have resistance to *P. bursaria* digestive enzymes (Kodama and Fujishima, 2010a).

Research into symbiosis initiation is in its infancy in insects. The study of insect-bacterial symbioses often only considers the established relationship. There is generally a practical reason for this. It is simpler to study a system that is already in place, and use this to make inferences about why these organisms came together, than it is to investigate the initiation of the relationship. Existing research into the tsetse microbiome focuses on *W. glossinidia* and *S. glossinidius* as they are currently. It is difficult to determine what may have caused the initial interaction between the tsetse and the infective symbiont ancestors. For *W. glossinidia*, this event is thought to have occurred 80 million years ago (Chen et al., 1999; Snyder and Rio, 2013). Prior to the work presented in Chapter 5, the *W. glossinidia* was

thought to be highly similar to that of *Escherichia coli*. The latter was considered to be a potential ancestor for the symbiont (Pál et al., 2006). It was demonstrated here however that there are proteins within *W. glossinidia* that do not appear to have an orthologue in *E. coli*. This suggests that more work needs to be undertaken to firmly establish the nature of the organism that first infected the tsetse. This could then be used in subsequent *in vivo* work to infect aposymbiotic tsetse. There are however additional difficulties with this. It is possible to produce tsetse that are *W. glossinidia*-free in order to reinfect with possible ancestral species (Dale and Welburn, 2001), but this antibiotic treatment has been shown to result in insects that are in poor health (Nogge, 1976). This problem is not specific to tsetse (Wilkinson and Douglas, 1995a), and can subsequently limit research into the initiation of symbiosis.

The discovery of *S. praecaptivus* (Clayton et al., 2012; Chari et al., 2015) provided an opportunity to study the trajectory that the *Sodalis* genus has taken from free-living to host-restricted. As a close relative of *S. glossinidius*, this species presents a possible starting point from which to investigate the changes that may have occurred within the symbiont genome. Genes that are present in *S. praecaptivus* but absent in *S. glossinidius* may be interpreted as not being required within the tsetse. Conversely, genes retained in the symbiont may be categorised as either being essential, or possibly as potential targets for future pseudogenisation. We have demonstrated here how this organism can be used to investigate the evolutionary trajectory of the *Sodalis* genus. We have shown *in silico* that certain genes can be lost early in the relationship with minimal effect, whereas others have severely detrimental outcomes for the symbiont if they become pseudogenised close to initiation. This tool has therefore opened up more avenues for studying the symbiosis initiation and evolution *in silico*. It is not however a perfect solution to the difficulties in studying symbiosis initiation. The ancestor that started this interaction may not be *S. praecaptivus*, but another relative. Infecting *S. glossinidius*-free tsetse with *S. praecaptivus* may enable the subsequent effects on the insects to be studied in comparison to wild-type. Not all species of tsetse contain *S. glossinidius* at 100% infection frequency (Dennis et al., 2014). It may therefore be possible to select a species with naturally lower levels of *S. glossinidius* for this, reducing or removing the need for antibiotic treatment. If this is successful then it may become a useful tool for investigating the events and requirements at the start of the symbiosis.

6.5 To primary symbiosis and beyond

Existing symbioses are merely snapshots in evolutionary time, and care should be taken when interpreting the nature of these relationships. It is challenging to decipher exactly what transitions each relationship has undergone historically, and even more so to predict what may happen in the future. It is not clear whether all species begin their symbioses as secondary symbionts, existing in a commensal state without a notable effect on the insects. Some species may remain as commensal symbionts for large periods of evolutionary time, as demonstrated by species of lactic acid bacteria that have lived in symbiosis with the Apidae family of bees for approximately 80 million years (Cardinal et al., 2010). Others progress to becoming mutualistic and then obligate primary symbionts relatively rapidly. *Ca.* Sodalis pierantonius is an example of this. It is thought that it lost over 50% of its genome over a short period of time (Clayton et al., 2012). It is also not known whether all symbionts, at the start of an association, have the potential to become obligate primary symbionts.

S. glossinidius is at an earlier stage in its symbiotic evolution than the primary *W. glossinidia*. Its genome, with the vast complement of pseudogenes, is undergoing large-scale reduction (Dale and Maudlin, 1999). Whether *S. glossinidius* provides a benefit to its host remains to be elucidated. It is unlikely, however, that *S. glossinidius* could become an obligate symbiont. It has been demonstrated in Chapter 2 that *S. glossinidius* likely relies on *W. glossinidia* for thiamine in the absence of its own complete biosynthetic pathway. This suggests that unless *S. glossinidius* was to horizontally acquire thiamine biosynthesis genes from *W. glossinidia*, it would be unable to exist in the absence of the tsetse's primary symbiont.

Studies into insect-microbe symbioses invariably conclude at the primary symbionts. Examples have been given here of how the symbionts become replaced when there are other organisms more capable of provisioning that particular benefit. There are also examples of two bacterial symbionts merging within an insect, specifically the nested symbiosis observed in the mealybug (von Dohlen et al., 2001). Here, a β-proteobacterium appears to have engulfed a γ-proteobacterium. The authors hypothesise that this may enable the exchange of genetic material, thus slowing the genome degradation process. It is not known how common this is in nature, but it nevertheless implies that at least one of the symbionts is performing its role sub-optimally. Less attention is directed towards what may happen if a primary symbiont continues to undergo genomic degradation whilst maintaining its essential benefit to the insect. *W. glossinidia*, for example, has a highly reduced genome (Akman

et al., 2002; Rio et al., 2012). Prior to this work, the benefits of *W. glossinidia* to its host, namely in the provision of B vitamins, was already known (Michalkova et al., 2014; Snyder and Rio, 2015). The benefit to *S. glossinidius*, via the production of thiamine, was presented in Chapter 2. This suggests that *W. glossinidia* is a keystone species within the tsetse microbiome, and therefore a possible candidate for unusual strategies to ensure its maintenance.

The earliest and most ancient symbiosis has been described here; the association between an α-proteobacterium and an archaeon that persisted and formed modern day eukaryotes. Bacteria undergo a transition within an insect, from a free-living species with a large genome that retains an extensive coding capacity and redundancy, to an obligate symbiont with a highly reduced genome and an inability to survive in isolation (Moran, 1996). Mitochondria, by comparison, retain a small amount of DNA that allows a rapid response to changes in energy demand. They cannot be separated from their hosts, have transferred DNA to the host chromosome, and are inherited maternally. This then raises the question of whether, at a certain point, an obligate symbiont may be defined as an organelle.

W. glossinidia could be on the path to being defined not as a symbiont, but as an organelle. As discussed, its genome is highly reduced (Akman et al., 2002; Rio et al., 2012) and it provides an essential function for both the tsetse (Michalkova et al., 2014; Snyder and Rio, 2015) and *S. glossinidius* (Chapter 2) (Akman et al., 2002). It is also maternally inherited (Aksoy, 2000). There is however no evidence of a transfer of genetic material from symbiont to host. The tsetse genome does display hallmarks of horizontal transfer events from *Wolbachia* (Attardo et al., 2014), indicating that DNA from *W. glossinidia* could conceivably become integrated into the tsetse genome in the future. This may enable the *W. glossinidia* genome to become reduced further, to the point of either elimination and eventual replacement, or persistence as an organelle.

Looking outside of the tsetse to other insects may provide clues as to whether the ultimate end-point of symbiosis is to become an organelle. *Ca.* Sodalis pierantonius, for example, is thoroughly intertwined within the biology of its host (Heddi et al., 1997). It provides an essential function, producing vitamins and amino acids for its weevil host. It is inherited maternally. Crucially, there is thought to be a form of genetic interaction between symbiont and host (Heddi et al., 1997). These characteristics that describe *Ca.* Sodalis pierantonius are also those used to describe organelles like mitochondria. Conversely, *Ca.* Sodalis pierantonius still retains a reasonably large genome size (4.5 Mb) that distinguishes it from an organelle (Oakeson et al., 2014).

It is therefore not clear what the final destination of insect symbionts is. Some may progress from secondary to primary, whereas others may never become the latter. What is evident however is that future research should focus not only on symbioses as they are currently, but also on what might happen to the symbionts in the future. *In silico* tools such as the FBA evolver developed in Chapter 4 are one such method to do so.

6.6 Conclusions

Insect-bacterial symbioses are widespread and borne out of necessity. Insects have health and economic importance as disease vectors and crop pests. Methods of control that do not involve pesticides and that can reach rural areas are desirable. For this to be achieved, a greater understanding of the biology of the insect is essential. This includes the insect's microbiota. The insects and their symbionts are intertwined metabolically and physiologically, and as such should be considered as a whole.

To this end, we have sought to enhance current understanding of the tsetse microbiome using a combination of computational and empirical techniques. First, metabolic adaptations that occur as a result of insect-microbe symbioses were investigated. We have discovered, using a new FBA model and *in vitro* testing, that *S. glossinidius* has unusual adaptations to the metabolism of *W. glossinidia*, and the physiology and hematophagic diet of their host (Chapter 2). It relies on *W. glossinidia* for a supply of the B vitamin thiamine due to multiple pseudogenisations in its biosynthetic pathway. The maintenance of two PTS transporters that can import GlcNAc, and an improvement of growth on this metabolite *in vitro*, suggests that *S. glossinidius* may be using a secreted chitinase to break down the tsetse's peritrophic membrane into monomers of GlcNAc that it can use as an energy source. *S. glossinidius* has also adapted to scavenge the amino acids found in the blood meal. It uses L-glutamate from this to supplement the TCA cycle, required as a result of a key pseudogenisation in the gene *ppc*.

This pseudogenisation, and other key events in the tsetse-*S. glossinidius* symbiosis have been investigated here through the use of FBA and a multi-objective evolutionary algorithm. The construction and testing of a model for *S. praecaptivus* metabolism uncovered pathways for the metabolism of xylitol and GalNAc not frequently observed in models of prokaryotes, including *E. coli* (Chapter 3). Applying the evolutionary algorithm to this model allowed the possible trajectories of *S.*

glossinidius once it became internalised to be simulated (Chapter 4). It was discovered that the loss of *ppc* in the symbiont could have occurred early in the symbiosis with minimal effect of the resulting trajectories. In contrast, it was found that had *S. glossinidius* lost the gene encoding the ASPTA reaction early, instead of retaining it, the resulting trajectories could have been poorer. This tool could be used in other symbioses for which a well-annotated relative is available, or to improve the directed evolution of industrially relevant microorganisms.

We have used extensive phylogenetics to dispute the previously held views that the *W. glossinidia* genome is a subset of *E. coli* (Chapter 5). We also highlighted the low level of sequence similarity between *W. glossinidia brevipalpis* and *W. glossinidia morsitans*. To our knowledge, this work is the first published example of a purified protein from *W. glossinidia*. This protein did not show any evidence of binding to L-proline, but this work did serve to highlight the difficulties in expressing symbiont proteins outside of their original host.

The work presented in this thesis furthers the understanding of insect-microbe symbioses, specifically the metabolic adaptations that arise as a result. It also provides a new way to examine the evolution of symbiosis. Future work applying this knowledge to other symbioses may help to harness the power of symbionts in the control of disease.

Chapter 7

Appendices

7.1 Appendix A - Chapter 2

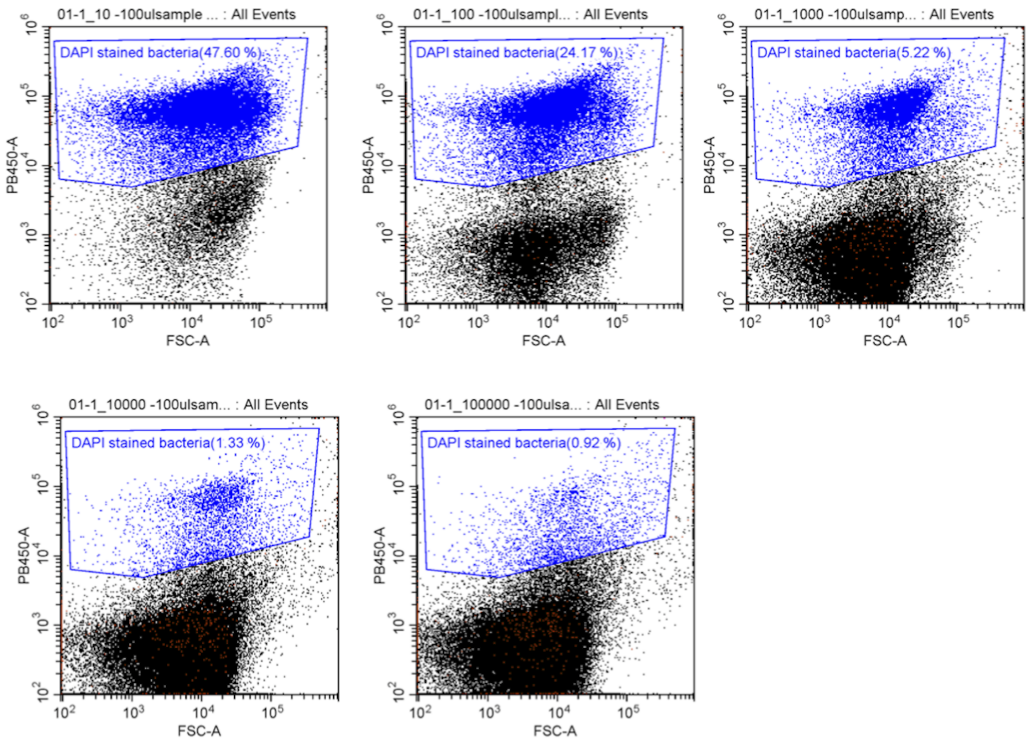


Figure S1: Flow cytometer gating for DAPI-stained *S. glossinidius* cell count.

Table S1: Pantothenate biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
panB	3-methyl-2-oxobutanoate hydroxymethyltransferase	MOHMT	SG0488	Sant_3309
panE	2-dehydropantoate 2-reductase	DPR	SG0661	Sant_3012
panC	Pantothenate synthase	PANTS	SG0487	Sant_3310
coaA	Pantothenate kinase	PNTK	SG0126	Sant_3935
dfp	Phosphopantothenate-cysteine ligase	PPNCL2	SG2209	Sant_0420
dfp	Phosphopantothenoyl-cysteine decarboxylase	PPCDC	SG0299	Snat_0420
coaD	Pantetheine-phosphate adenylyltransferase	PTPATi	SG2205	Sant_4015
coaE	Dephospho-CoA kinase	DPCOAK	SG0461	Sant_3340
panF	Pantothenate sodium symporter	PNTOt4pp	—	Sant_0493

Table S2: Biotin biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
bioF	8-amino-7-oxononanoate synthase	AOXSr	SG0904	Sant_2734
bioA	Adenosylmethionine-8-amino-7-oxononanoate transaminase	AMAOTr	SG0902	Sant_2736
bioD	Dethiobiotin synthase	DBTS	SG1466	Sant_2732
bioB	Biotin synthase	BTS2	SG0903	Sant_2735
bioC	Malonyl-CoA methyltransferase	MALCOAMT	SG0905	Sant_2733
bioH	Pimeloyl-[ACP] methyl ester esterase	PMEACPE	SG2324	Sant_0363

Table S3: Riboflavin biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
ribA	GTP cyclohydrolase II	GTPCII	SG1410	Sant_1871
ribD	5-amino-6-(5-phosphoribosylamino) uracil reductase	APRAUR	SG0651	Sant_3023
ribE	Riboflavin synthase	RBFSb	SG1439	Sant_3022
ribC	Riboflavin synthase	RBFSa	SG1439	Sant_2045
ribB	3,4-dihydroxy-2-butanone-4-phosphate synthase	DB4PS	SG0263	Sant_3614

Table S4: Protoheme biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
hemA	Glutamyl-tRNA reductase	GLUTRR	SG1877	Sant_2113
hemL	Glutamate-1-semialdehyde aminotransferase	G1SAT	SG0500	Sant_3285
hemB	Porphobilinogen synthase	PPBNGS	GS1529	Sant_2301
hemC	Hydroxymethylbilane synthase	HMBS	SG2366	Sant_0327
hemD	Uroporphyrinogen-III synthase	UPP3S	SG2367	Sant_0326
hemE	Uroporphyrinogen decarboxylase	UPPDC1	SG0138	Sant_3913
hemY	Protoporphyrinogen oxidase	PPPGO	SG0121	Sant_0261
hemH	Ferrochelatase	FCLT	SG0694	Sant_2970
ccmABCDE	Protoheme transport via ABC system	PHEMEabcpp	SG1635-7	Sant_1288-92

Table S5: PLP biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
dxs	1-deoxy-D-xylulose 5-phosphate synthase	DXPS	SG0656	Sant_3017
gapA	Erythrose 4-phosphate dehydrogenase	E4PD	SG1347	Sant_1815
pdxB	Erythronate 4-phosphate dehydrogenase	PERD	SG1621	Sant_1314
serC	O-phospho-4-hydroxy-L-threonine: 2-oxoglutarate aminotransferase	OHPBAT	SG0990	Sant_2634
pdxA, *pdxJ*	Pyridoxine 5'-phosphate synthase	PDX5PS	SG0425, SG1784	Sant_3384, Sant_1101
pdxH	Pyridoxine 5'-phosphate oxidase, pyridoxamine 5'-phosphate oxidase	PDX5POi, PYAM5PO	SG1447	Sant_2034

Table S6: Thiamine biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
thiC	4-amino-2-methyl-5- phosphomethylpyrimidine synthetase	AMPMS2	—	Sant_3961
thiE	Thiamine-phosphate diphosphorylase	TMPPP	—	Sant_3917
thiF	Sulfur carrier protein adenylyltransferase	THZPSN	—	Sant_3918
thiS	Sulfur carrier protein	THZPSN	—	Sant_3919
thiG	Thiazole synthase	THZPSN	—	Sant_3920
thiH	2-iminoacetate synthase	THZPSN	—	Samt_3921
thiD	Hydroxymethylpyrimidine kinase, phosphomethylpyrimidine kinase	HMPK1, PMPK	—	Sant_1177
thiI	tRNA sulfurtransferase	THZPSN	SG0659	Sant_3014
thiM	Hydroxyethylthiazole kinase	HETZK	SG1739	Sant_1176
thiL	Thiamine-phosphate kinase	TMPK	SG0654	Sant_3020
thiK	Thiamine kinase	TMK	SG1071	Sant_2466
iscS	Cysteine desulfurase	THZPSN, ICYSDS	SG1769	Sant_1125
tbpAthiPQ	Thiamine transport via ABC system	THMabcpp	SG0431-3	Sant_3371-3

Table S7: Tetrahydrofolate biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
folM	Dihydromonapterin reductase	DHMPTR	—	—
folA	Dihydrofolate reductase	DHFR	SG0421	Sant_3388
folC	Dihydrofolate synthase	DHFS	SG1616	Sant_1319
folP	Dihydropteroate synthase	DHPS2	SG0372	Sant_3460
folK	6-hydroxymethyl-dihydropterin pyrophosphokinase	HPPK2	SG0489	Sant_3308
folB	Dihydroneopterin aldolase, dihydroneopterin epimerase	DHNPA2r, DHNPTE	SG0256	Sant_3622
nudB	Dihydroneopterin triphosphate pyrophosphatase	DNTPPA	SG1258	Sant_1683
folE	GTP cyclohydrolase I	GTPCI	SG0957	Sant_1427

Table S8: Cobalamin biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
btuR	Cobinamide adenyltransferase, cob(I)alamin adenosyltransferase	CBIAT, CBLAT	—	Sant_1862
cobT	Nicotinate-nucleotide dimethylbenzimidazole phosphoribosyltransferase	NNDMBRT	—	—
cobC	Alpha-ribazole 5-phosphate phosphatase	RZ5PP	—	—
cobU	Adenosyl cobinamide phosphate guanyltransferase, adenosyl cobinamide kinase	ACBIPGT, ADOCBIK	—	—
cobS	Adenosylcobalamin 5'-phosphate synthase	ADOCBLS	—	—

Table S9: Nicotinamide biosynthesis capabilities. BLASTp results showing orthologues of *S. praecaptivus* vitamin biosynthesis proteins in *S. glossinidius*.

Gene	Reaction	BiGG reaction	*S. glossinidius*	*S. praecaptivus*
nadB	L-aspartate oxidase	ASPO6	SG1794	Sant_1091
nadA	Quinolinate synthase	QULNS	SG0889	Sant_2754
nadC	Quinolinate phosphoribosyltransferase	NNDPR	SG0464	Sant_3335
nadD	Nicotinate-mononucleotide adenylyltransferase	NMNAT	SG0800	Sant_3420
nadE	NAD+ synthetase	NADS1	SG1866	Sant_2100

Table S10: Amino acid biosynthesis and transport capabilities. Transporter families (TransportDB); dicarboxylate/amino acid:cation symporter (DAACS), amino acid-polyamine-organocation family (APC), branched chain amino acid symporter (LIVSC), alanine or glycine:cation symporter (AGCS), hydroxyl/aromatic amino acid permease (HAAAP). *Cystine is the oxidised form of cysteine and transported via the specific FliY, of which there is no apparent orthologue in *S. glossinidius* or *S. praecaptivus*. [1]This is in contrast to Toh et al. (2006) who state that the pathway of alanine biosynthesis is not complete.

Amino acid	Pathway completeness	Reaction	Family	Gene	*S. praecaptivus* transporter
Glutamate	From α-ketoglutarate	GLUt2r	DAACS	*SG2121*	GLUt2r (Sant_3844) GLUabc (Sant_2814-7)
Glutamine	From α-ketoglutarate	—	—	—	—
Arginine	**Incomplete**	ARGabc	ABC	*SG1093-6*	ARGabc (Sant_1325-8/Sant_2659-62)
Proline	From α-ketoglutarate	—	—	—	PROt2r (Sant_0028) PROabc (Sant_0944, 46-7)
Lysine	From aspartate	LYSt2r	APC	*SG0955*	LYSt2r (Sant_1419)
Asparagine	From aspartate	—	—	—	—
Aspartate	From oxaloacetate	ASPt2	DAACS	*SG2121*	ASPt3_3 (Sant_2332) ASPt2r (Sant_3844) ASPabc (Sant_2814-7)
Isoleucine	From aspartate	ILEt2r	LIVCS	*SG0640*	ILEt2r (Sant_3037)
Threonine	From aspartate	—	—	—	—-
Methionine	From aspartate	METabc	ABC	*SG1915-7*	METabc (Sant_0923-5)
Valine	From pyruvate	VALt2r	LIVCS	*SG0640*	VALt2r (Sant_3037)
Leucine	From pyruvate	LEUt2r	LIVCS	*SG0640*	LEUt2r (Sant_3037) LEUabc (Sant_3573-7)
Alanine	From pyruvate[1]	ALAt2r	AGCS	*SG0408*	ALAt2r (Sant_3406/Sant_2281)
Serine	From 3-phospho D-glycerate	SERt2r	HAAAP	*SG0922*	SERt2r (Sant_2673)
Cysteine*	From 3-phospho D-glycerate	—	—	—	—
Glycine	From 3-phospho D-glycerate	—	—	—	GLYt2r (Sant_3406/Sant_2281)
Phenylalanine	From D-erythrose 4-phosphate	PHEt2r	APC	*SG0465*	PHEt2r (Sant_3330)
Tryptophan	From D-erythrose 4-phosphate	TRPt2r	APC	*SG0465*	TRPt2r (Sant_3330)
Tyrosine	From D-erythrose 4-phosphate	TYRt2r	APC	*SG0465*	TYRt2r (Sant_3330)
Histidine	From ribose 5-phosphate	HISt2r	APC	*SG0465*	HISt2r (Sant_3330) HISabc (Sant_1325-8)

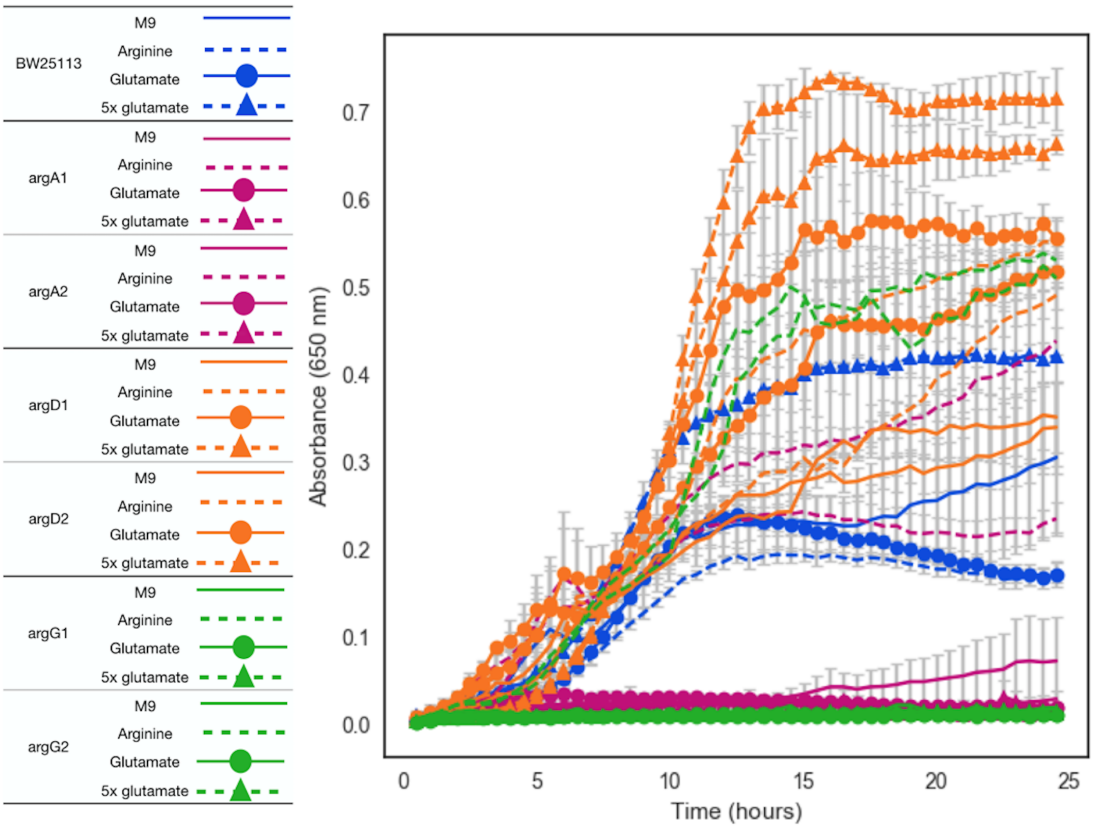


Figure S2: Effects of L-arginine biosynthesis gene deletions on *E. coli* growth *in vitro*. Wild-type *E. coli* strain BW25113 (blue) was grown in M9 medium alone (solid line) or with added L-arginine (dashed line), L-glutamate (solid line, circle marker) or 5x concentration L-glutamate (dashed line, triangle marker) for 24 hours in a microplate reader. This was compared to growth in the same medium combinations for *argA* (pink), *argD* (orange), and *argG* (green) gene deletion mutants from the Keio collection (Baba et al., 2006). Two isolates for each mutant are shown. The *argD* knockout strain is able to grow under all conditions, whereas strains lacking *argA* and *argG* can grow only when exogenous L-arginine is provided. Measurements for triplicate experiments are shown; error bars depict SEM. The graph was produced using Plate Whisperer software for microplate data analysis, developed by Stephen Thorpe.

7.2 Appendix B - Chapter 4

Table S11: Fluxes (mmol gr DW^{-1} hr^{-1}) through the uptake reactions for components of the blood medium for ancestral *i*RH830 and ten evolved (E1-E10) models.

Metabolite	*i*RH830	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
acgam	6	6	5.49	6	5.39	5.39	0.24	0.38	6	6	0.24
ala	3.41	0.44	3.41	0.44	3.41	3.41	0	3.41	0.99	0.99	0
arg	1.14	0.25	0.26	0.25	0.26	0.26	0.09	0.14	0.26	0.26	0.09
asn	0.58	0.57	0.58	0.57	0.57	0.57	0.06	0.1	0.18	0.18	0.06
asp	0.03	0	0.03	0	0.03	0.03	0	0	0	0	0
cys	0	0	0	0	0	0	0	0	0	0	0
gln	0	0	0	0	0	0	0	0	0	0	0
glu	0.7	0.46	0.49	0.46	0.49	0.49	0	0.7	0	0	0
gly	1.54	0.26	0.28	0.26	0.28	0.28	0.32	0.5	0.93	0.93	0.32
his	0.11	0.07	0.07	0.07	0.07	0.07	0.02	0.04	0.07	0.07	0.02
ile	0.33	0.21	0.22	0.21	0.22	0.22	0.08	0.12	0.22	0.22	0.08
leu	0.52	0.33	0.34	0.33	0.34	0.34	0.12	0.18	0.34	0.34	0.12
lys	0.39	0.25	0.26	0.25	0.26	0.26	0.09	0.14	0.26	0.260	0.09
met	0.18	0.11	0.12	0.11	0.12	0.12	0.04	0.06	0.12	0.12	0.04
orn	0	0	0	0	0	0	0	0	0	0	0
phe	0.21	0.13	0.14	0.13	0.14	0.14	0.05	0.08	0.14	0.14	0.05
pro	2.36	0.16	0.17	0.16	0.17	0.17	2.36	2.26	1.27	1.27	2.36
ser	1.12	1.07	1.12	1.07	1.12	1.12	0.11	0.17	0.32	0.32	0.11
taur	0	0	0	0	0	0	0	0	0	0	0
thm	0.00001	0.000008	0.000008	0.000008	0.000008	0.000008	0	0	0	0	0
thr	0	0	0	0	0	0	0	0	0	0	0
trp	0.07	0.04	0.04	0.04	0.04	0.04	0.01	0.02	0.04	0.04	0.01
tyr	0.16	0.1	0.1	0.1	0.1	0.1	0.04	0.06	0.11	0.11	0.04
val	0.49	0.31	0.32	0.31	0.32	0.32	0.11	0.17	0.32	0.32	0.11

Table S12: Fluxes (mmol gr DW^{-1} hr^{-1}) through the uptake reactions for components of the blood medium for ancestral *i*LF517 and ten evolved (E1-E10) models.

Metabolite	*i*LF517	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10
acgam	6	0	0	0	0	0	0	0	0	0	0
ala	3.41	0	3.41	0	3.41	3.41	3.41	0	0	3.41	3.41
arg	0.15	0.02	0.11	0.04	0.1	0.11	0.11	0.02	0.04	0.1	0.11
asn	0	0	0	0	0	0	0	0	0	0	0
asp	0.03	0	0.03	0	0	0.03	0.03	0	0	0	0
cys	0.04	0.01	0.03	0.01	0.03	0.03	0.03	0.01	0.01	0.03	0.03
gln	0	0	0	0	0	0	0	0	0	0	0
glu	0.7	0.7	0.7	0.7	0.49	0.7	0.7	0.7	0.7	0.48	0.7
gly	0	0	0	0	0	0	0	0	0	0	0
his	0.04	0.01	0.03	0.01	0.03	0.03	0.03	0.01	0.01	0.03	0.03
ile	0.13	0.02	0.09	0.03	0.08	0.09	0.09	0.02	0.03	0.08	0.09
leu	0.2	0.03	0.14	0.05	0.13	0.14	0.14	0.03	0.05	0.13	0.14
lys	0.15	0.02	0.11	0.04	0.1	0.11	0.11	0.02	0.04	0.1	0.11
met	0.07	0.01	0.05	0.02	0.05	0.05	0.05	0.01	0.02	0.05	0.05
orn	0	0	0	0	0	0	0	0	0	0	0
phe	0.08	0.01	0.06	0.02	0.05	0.06	0.06	0.01	0.02	0.05	0.06
pro	0	0	0	0	0	0	0	0	0	0	0
ser	1.12	0.12	1.12	1.12	1.12	1.12	1.12	0.12	1.12	1.12	1.12
taur	0	0	0	0	0	0	0	0	0	0	0
thm	0.000005	0.0000007	0.000003	0.000001	0.000003	0.000003	0.000003	0.0000007	0.000001	0.000003	0.000003
thr	0	0	0	0	0	0	0	0	0	0	0
trp	1.14	0.004	0.02	0.01	0.02	0.02	0.02	0.004	0.01	0.02	0.02
tyr	0.06	0.01	0.04	0.02	0.04	0.04	0.04	0.01	0.02	0.04	0.04
val	0.19	0.03	0.13	0.05	0.12	0.12	0.12	0.03	0.05	0.12	0.13

7.3 Appendix C - Chapter 5

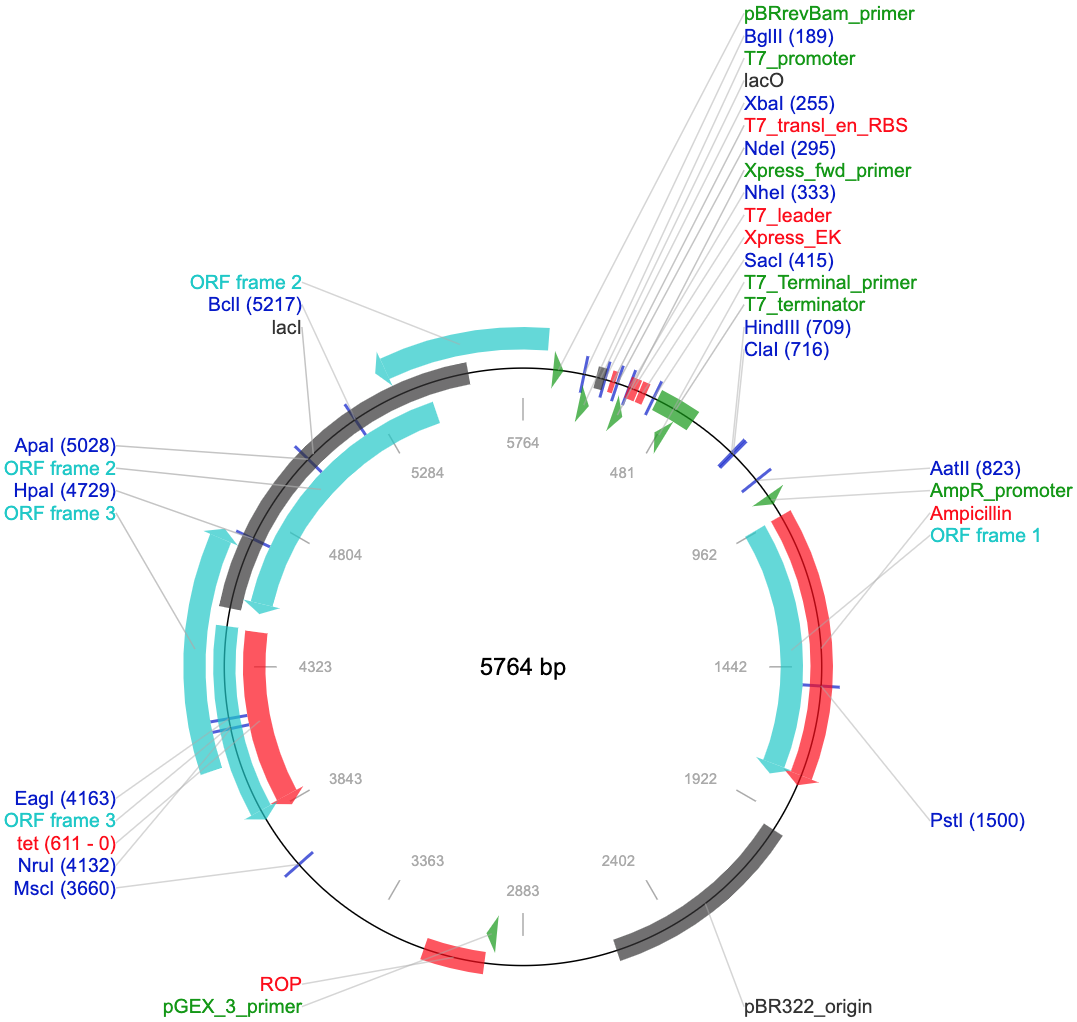


Figure S3: Vector map for pET100/D-TOPO (Thermo Fisher Scientific) expression vector, accessed via the Addgene repository.

Chapter 8

Abbreviations

ABC	ATP-binding cassette
ANOVA	Analysis of variance
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BHI	Brain heart infusion
bp	Base pair
CD	Circular dichroism
DNA	Deoxyribonucleic acid
FBA	Flux balance analysis
GalNAc	*N*-acetyl-D-galactosamine
GlcNAc	*N*-acetyl-D-glucosamine
HAT	Human African trypanosomiasis
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	Kilobase
kDa	Kilodalton
KPi	Potassium phosphate
LB	Luria-Bertani
Mb	Megabase
(m)L	(Milli)litre
(m)M	(Milli)mole
MOEA	Multi-objective evolutionary algorithm
MOMA	Minimization of metabolic adjustment
NaCl	Sodium chloride
NAPD	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced NADP
nm	Nanometer
NSGA-II	Non-dominated sorting genetic algorithm-II
NTD	Neglected tropical disease
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMF	Peptide mass fingerprinting
PTS	Phosphotransferase system
RNA	Ribonucleic acid
rpm	Revolutions per minute
SBP	Substrate binding protein
SDS	Sodium dodecyl sulfate
SEM	Standard error (of the) mean
TCA	Tricarboxylic acid
TEMED	Tetramethylethylenediamine
TMP	Thiamine monophosphate
V	Volts
W	Watts
WT	Wild-type

Bibliography

Adler, J. & Templeton, B. (1967). The effect of environmental conditions on the motility of Escherichia coli. *Journal of General Microbiology*, 46(2):175 – 184.

Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., & Aksoy, S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia. *Nature Genetics*, 32(3):402–407.

Akman Gündüz, E. & Douglas, A. E. (2009). Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proceedings of the Royal Society B*, 276:987–991.

Aksoy, S. (1995). Wigglesworthia gen. nov. and Wigglesworthia glossinidia sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *International Journal of Systematic and Evolutionary Microbiology*, 45(4):848–851.

Aksoy, S. (2000). Tsetse - A haven for microorganisms. *Parasitology Today*, 16(3):114–118.

Aksoy, S., Chen, X., & Hypsa, V. (1997). Phylogeny and potential transmission routes of midgut-associated endosymbionts of tsetse (Diptera:Glossinidae). *Insect Molecular Biology*, 6(2):183–190.

Aksoy, S., Maudlin, I., Dale, C., Robinson, A. S., & Neill, S. L. O. (2001). Prospects for control of African trypanosomiasis by tsetse vector manipulation. *Trends in Parasitology*, 17(1):29–35.

Alam, U., Medlock, J., Brelsfoard, C., Pais, R., Lohs, C., Balmand, S., Carnogursky, J., Heddi, A., Takac, P., Galvani, A., & Aksoy, S. (2011). Wolbachia symbiont infections induce strong cytoplasmic incompatibility in the tsetse fly Glossina morsitans. *PLOS Pathogens*, 7(12):e1002415.

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, 59(1):143–69.

Andersen, K. B. & von Meyenburg, K. (1980). Are growth rates of Escherichia coli in batch cultures limited by respiration? *Journal of Bacteriology*, 144(1):114–23.

Ankrah, N. Y. D., Chouaia, B., & Douglas, A. E. (2018). The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. *mBio*, 9:01433–18.

Ankrah, N. Y. D., Luan, J., & Douglas, A. E. (2017). Cooperative metabolism in a three-partner insect-bacterial symbiosis revealed by metabolic modeling. *Journal of Bacteriology*, 199(15):00872–16.

Attardo, G. M., Abila, P. P., Auma, J. E., Baumann, A. A., Benoit, J. B., Brelsfoard, C. L., Ribeiro, J. M. C., Cotton, J. A., Pham, D. Q. D., Darby, A. C., & et al (2014). Genome sequence of the tsetse fly (Glossina morsitans): Vector of African Trypanosomiasis. *Science*, 344(6182):380–386.

Azambuja, P., Feder, D., & Garcia, E. (2004). Isolation of Serratia marcescens in the midgut of Rhodnius prolixus: impact on the establishment of the parasite Trypanosoma cruzi in the vector. *Experimental Parasitology*, 107(1-2):89–96.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2(1):2006.0008.

Babendreier, D., Joller, D., Romeis, J., Bigler, F., & Widmer, F. (2007). Bacterial community structures in honeybee intestines and their response to two insecticidal proteins. *FEMS Microbiology Ecology*, 59(3):600–610.

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Alpi, E., Bely, B., Bingley, M., Britto, R., Bursteinas, B., Busiello, G., & et al (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1):D506–D515.

Batut, B., Parsons, D. P., Fischer, S., Beslon, G., & Knibbe, C. (2013). In silico experimental evolution: a tool to test evolutionary scenarios. *BMC Bioinformatics*, 14(Suppl 15):S11.

Baumann, P., Baumann, L., Lai, C.-Y., Rouhbakhsh, D., Moran, N. A., & Clark, M. A. (1995). Genetics, physiology, and evolutionary relationships of the genus Buchnera: Intracellular symbionts of aphids. *Annual Review of Microbiology*, 49:55–94.

Beard, C. B., O'Neill, S. L., Mason, P., Mandelco, L., Woese, C. R., Tesh, R. B., Richards, F. F., & Aksoy, S. (1993). Genetic transformation and phylogeny of bacterial symbionts from tsetse. *Insect Molecular Biology*, 1(3):123–131.

Belda, E., Moya, A., Bentley, S., & Silva, F. J. (2010). Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of Sodalis glossinidius, the secondary endosymbiont of tsetse flies. *BMC Genomics*, 11:449.

Belda, E., Silva, F. J., Peretó, J., & Moya, A. (2012). Metabolic networks of Sodalis glossinidius: A systems biology approach to reductive evolution. *PLoS ONE*, 7(1):e30652.

Bennett, G. M. & Moran, N. A. (2013). Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biology and Evolution*, 5(9):1675–1688.

Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D. C., Lennard, N. J., Caler, E., Hamlin, N. E., Haas, B., & et al (2005). The genome of the African trypanosome Trypanosoma brucei. *Science*, 309(5733):416–422.

Bhatt, T. K., Yogavel, M., Wydau, S., Berwal, R., & Sharma, A. (2010). Ligand-bound structures provide atomic snapshots for the catalytic mechanism of D-amino acid deacylase. *Journal of Biological Chemistry*, 285(8):5917–5930.

Bian, G., Xu, Y., Lu, P., Xie, Y., & Xi, Z. (2010). The endosymbiotic bacterium Wolbachia induces resistance to dengue virus in Aedes aegypti. *PLoS Pathogens*, 6(4):e1000833.

Blagrove, M. S. C., Arias-Goeta, C., Failloux, A.-B., & Sinkins, S. P. (2012). Wolbachia strain wMel induces cytoplasmic incompatibility and blocks dengue transmission in Aedes albopictus. *Proceedings of the National Academy of Sciences*, 109(1):255 – 260.

Blakemore, R. P., Maratea, D., & Wolfe, R. S. (1979). Isolation and pure culture of a freshwater magnetic spirillum in chemically defined medium. *Journal of Bacteriology*, 140(2):720 – 729.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science*, 277(5331):1453–62.

Boyd, B. M., Allen, J. M., Koga, R., Fukatsu, T., Sweet, A. D., Johnson, K. P., & Reed, D. L. (2016). Two bacterial genera, Sodalis and Rickettsia, associated with the seal louse Proechinophthirus fluctus (Phthiraptera: Anoplura). *Applied and Environmental Microbiology*, 82(11):3185 – 3197.

Brinkkotter, A., Kloss, H., Alpert, C.-A., & Lengeler, J. W. (2000). Pathways for the utilization of N-acetyl-galactosamine and galactosamine in Escherichia coli. *Molecular Microbiology*, 37(1):125–135.

Brinza, L., Viñuelas, J., Cottret, L., Calevro, F., Rahbé, Y., Febvay, G., Duport, G., Colella, S., Rabatel, A., Gautier, C., Fayard, J.-M., Sagot, M.-F., & Charles, H. (2009). Systemic analysis of the symbiotic function of Buchnera aphidicola, the primary endosymbiont of the pea aphid Acyrthosiphon pisum. *Comptes Rendus Biologies*, 332(11):1034–1049.

Bursell, E. (1963). Aspects of the metabolism of amino acids in the tsetse fly, Glossina (Diptera). *Journal of Insect Physiology*, 9(4):439–452.

Bursell, E. (1978). Quantitative aspects of proline utilization during flight in tsetse flies. *Physiological Entomology*, 3(4):265–272.

Bursell, E. & Slack, E. (1976). Oxidation of proline by sarcosomes of the tsetse fly, Glossina morsitans. *Insect Biochemistry*, 6(2):159 – 167.

Büscher, P., Cecchi, G., Jamonneau, V., & Priotto, G. (2017). Human African trypanosomiasis. *The Lancet*, 390(10110):2397–2409.

Butler, J. D., Levin, S. W., Facchiano, A., Miele, L., & Mukherjee, A. B. (1993). Amino acid composition and N-terminal sequence of purified cystine binding Protein of Escherichia coli. *Life Sciences*, 52(14):1209–1215.

Cardinal, S., Straka, J., & Danforth, B. N. (2010). Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proceedings of the National Academy of Sciences*, 107(37):16207–11.

Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S. Y., Shearer, A. G., Tissier, C., & et al (2007). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 36(Database):D623–D631.

Caspi-Fluger, A., Inbar, M., Mozes-Daube, N., Katzir, N., Portnoy, V., Belausov, E., Hunter, M. S., & Zchori-Fein, E. (2012). Horizontal transmission of the insect symbiont Rickettsia is plant-mediated. *Proceedings of the Royal Society B*, 279(1734):1791–1796.

Chari, A., Oakeson, K. F., Enomoto, S., Jackson, D. G., Fisher, M. A., & Dale, C. (2015). Phenotypic characterisation of Sodalis praecaptivus sp. nov., a close non-insect associated member of the Sodalis-allied lineage of insect endosymbionts. *International Journal of Systematic and Evolutionary Microbiology*, 65:1400–1405.

Chen, X., Li, S., & Aksoy, S. (1999). Concordant evolution of a symbiont with its host insect species: Molecular phylogeny of genus Glossina and its bacteriome-associated endosymbiont, Wigglesworthia glossinidia. *Journal of Molecular Evolution*, 48(1):49–58.

Cheng, Q. & Aksoy, S. (1999). Tissue tropism, transmission and expression of foreign genes in vivo in midgut symbionts of tsetse flies. *Insect Molecular Biology*, 8(1):125 – 132.

Cheng, Q., Ruel, T. D., Zhou, W., Moloo, S. K., Majiwa, P., O'Neill, S. L., & Aksoy, S. (2000). Tissue distribution and prevalence of Wolbachia infections in tsetse flies, Glossina spp. *Medical and Veterinary Entomology*, 14(1):44–50.

Chrudimský, T., Husník, F., Nováková, E., & Hypša, V. (2012). Candidatus Sodalis melophagi sp. nov.: Phylogenetically independent comparative model to the tsetse fly symbiont Sodalis glossinidius. *PLoS ONE*, 7(7):e40354.

Chung, S. H., Jing, X., Luo, Y., & Douglas, A. E. (2018). Targeting symbiosis-related insect genes by RNAi in the pea aphid- Buchnera symbiosis. *Insect Biochemistry and Molecular Biology*, 95:55–63.

Chung, W.-C., Chen, L.-L., Lo, W.-S., Kuo, P.-A., Tu, J., & Kuo, C.-H. (2013). Complete genome sequence of Serratia marcescens WW4. *Genome Announcements*, 1(2):00126–13.

Clayton, A. L., Oakeson, K. F., Gutin, M., Pontes, A., Dunn, D. M., von Niederhausern, A. C., Weiss, R. B., Fisher, M., & Dale, C. (2012). A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect–bacterial symbioses. *PLoS Genetics*, 8(11):e1002990.

Coutinho-Abreu, I. V., Zhu, K. Y., & Ramalho-Ortigao, M. (2010). Transgenesis and paratransgenesis to control insect-borne diseases: Current status and future challenges. *Parasitology International*, 59(1):1–8.

Covert, M. W. & Palsson, B. O. (2002). Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *The Journal of Biological Chemistry*, 277(31):28058–28064.

Covert, M. W. & Palsson, B. O. (2003). Constraints-based models: regulation of gene expression reduces the steady-state solution space. *Journal of Theoretical Biology*, 221:309–325.

Covert, M. W., Schilling, C. H., & Palsson, B. (2001). Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, 213(1):73–88.

da Mota, F. F., Marinho, L. P., Moreira, C. J. d. C., Lima, M. M., Mello, C. B., Garcia, E. S., Carels, N., & Azambuja, P. (2012). Cultivation-Independent Methods Reveal Differences among Bacterial Gut Microbiota in Triatomine Vectors of Chagas Disease. *PLoS Neglected Tropical Diseases*, 6(5):e1631.

Dale, C., Beeton, M., Harbison, C., Jones, T., & Pontes, M. (2006). Isolation, pure culture, and characterization of Candidatus Arsenophonus arthropodicus, an intracellular secondary endosymbiont from the hippoboscid louse fly Pseudolynchia canariensis. *Applied and Environmental Microbiology*, 72(4):2997–3004.

Dale, C. & Maudlin, I. (1999). Sodalis gen. nov. and Sodalis glossindius sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly Glossina morsitans morsitans. *International Journal of Systematic Bacteriology*, 49:267–275.

Dale, C. & Welburn, S. (2001). The endosymbionts of tsetse flies: manipulating host–parasite interactions. *International Journal for Parasitology*, 31(5-6):628–631.

Dale, C., Young, S. A., Haydon, D. T., & Welburn, S. C. (2001). The insect endosymbiont Sodalis glossinidius utilizes a type III secretion system for cell invasion. *Proceedings of the National Academy of Sciences*, 98(4):1883–1888.

Darby, A. C., Chandler, S. M., Welburn, S. C., & Douglas, A. E. (2005a). Aphid-symbiotic bacteria cultured in insect cell lines. *Applied and Environmental Microbiology*, 71(8):4833–4839.

Darby, A. C., Lagnel, J., Matthew, C. Z., Bourtzis, K., Maudlin, I., & Welburn, S. C. (2005b). Extrachromosomal DNA of the symbiont Sodalis glossinidius. *Journal of Bacteriology*, 187(14):5003–5007.

De Bary, A. (1879). *Die erscheinung der symbiose: Vortrag gehalten auf der versammlung deutscher naturforscher und aerzte zu cassel*. Trübner.

De Man, J. C., Rogosa, M., & Sharpe, M. E. (1960). A medium for the cultivation of Lactobacilli. *Journal of Applied Bacteriology*, 23(1):130–135.

De Vooght, L., Caljon, G., De Ridder, K., & Van Den Abbeele, J. (2014). Delivery of a functional anti-trypanosome Nanobody in different tsetse fly tissues via a bacterial symbiont, Sodalis glossinidius. *Microbial Cell Factories*, 13(1):156.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.

Degnan, P. H. & Moran, N. A. (2008). Diverse phage-encoded toxins in a protective insect endosymbiont. *Applied and Environmental Microbiology*, 74(21):6782–6791.

Degnan, P. H., Yu, Y., Sisneros, N., Wing, R. A., & Moran, N. A. (2009). Hamiltonella defensa, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences*, 106(22):9063–9068.

Dennis, J. W., Durkin, S. M., Horsley Downie, J. E., Hamill, L. C., Anderson, N. E., & MacLeod, E. T. (2014). Sodalis glossinidius prevalence and trypanosome presence in tsetse from Luambe National Park, Zambia. *Parasites & Vectors*, 7(1):378.

Desjeux, P. (2004). Leishmaniasis: current situation and new perspectives. *Comparative Immunology, Microbiology and Infectious Diseases*, 27(5):305–318.

Díaz, S., Villavicencio, B., Correia, N., Costa, J., & Haag, K. L. (2016). Triatomine bugs, their microbiota and Trypanosoma cruzi: asymmetric responses of bacteria to an infected blood meal. *Parasites & Vectors*, 9(1):636.

Doudoumis, V., Alam, U., Aksoy, E., Abd-Alla, A. M., Tsiamis, G., Brelsfoard, C., Aksoy, S., & Bourtzis, K. (2013). Tsetse-Wolbachia symbiosis: Comes of age and has great potential for pest and disease control. *Journal of Invertebrate Pathology*, 112:S94–S103.

Doudoumis, V., Tsiamis, G., Wamwiri, F., Brelsfoard, C., Alam, U., Aksoy, E., Dalaperas, S., Abd-Alla, A., Ouma, J., Takac, P., Aksoy, S., & Bourtzis, K. (2012). Detection and characterization of Wolbachia infections in laboratory and natural populations of different species of tsetse flies (genus Glossina). *BMC Microbiology*, 12(Suppl 1):S3.

Douglas, A. (1988a). Experimental studies on the mycetome symbiosis in the leafhopper Euscelis incisus. *Journal of Insect Physiology*, 34(11):1043–1053.

Douglas, A. (1988b). Sulphate utilization in an aphid symbiosis. *Insect Biochemistry*, 18(6):599–605.

Douglas, A. & Dixon, A. (1987). The mycetocyte symbiosis of aphids: Variation with age and morph in virginoparae of Megoura viciae and Acyrthosiphon pisum. *Journal of Insect Physiology*, 33(2):109–113.

Douglas, A. & Prosser, W. (1992). Synthesis of the essential amino acid tryptophan in the pea aphid (Acyrthosiphon pisum) symbiosis. *Journal of Insect Physiology*, 38(8):565–568.

Douglas, A. E. (1998). Nutritional interactions in insect-microbe symbioses: Aphids and their symbiotic bacteria Buchnera. *Annual Review of Entomology*, 43:17–37.

Dunbar, H. E., Wilson, A. C. C., Ferguson, N. R., & Moran, N. A. (2007). Aphid thermal tolerance Is governed by a point mutation in bacterial symbionts. *PLoS Biology*, 5(5):e96.

Dunn, A. K. & Stabb, E. V. (2005). Culture-independent characterization of the microbiota of the ant lion Myrmeleon mobilis (Neuroptera: Myrmeleontidae). *Applied and Environmental Microbiology*, 71(12):8784–8794.

Dyer, N. A., Rose, C., Ejeh, N. O., & Acosta-Serrano, A. (2013). Flying tryps: survival and maturation of trypanosomes in tsetse flies. *Trends in Parasitology*, 29(4):188–196.

Ebikeme, C. E., Peacock, L., Coustou, V., Riviere, L., Bringaud, F., Gibson, W., & Barrett, M. (2008). N-acetyl D-glucosamine stimulates growth in procyclic forms of Trypanosoma brucei by inducing a metabolic shift. *Parasitology*, 135(5):585–594.

Ebrahim, A., Lerman, J. A., Palsson, B. O., & Hyduke, D. R. (2013). COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology*, 7(1):74.

Edwards, J. S., Covert, M., & Palsson, B. (2002). Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology*, 4(3):133–140.

Edwards, J. S. & Palsson, B. O. (2000). Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics*, 1(1):1.

Elli, M., Zink, R., Rytz, A., Reniero, R., & Morelli, L. (2000). Iron requirement of Lactobacillus spp. in completely chemically defined growth media. *Journal of Applied Microbiology*, 88(4):695–703.

Fares, M. A., Ruiz-González, M. X., Moya, A., Elena, S. F., & Barrio, E. (2002). GroEL buffers against deleterious mutations. *Nature*, 417(6887):398–398.

Farikou, O., Njiokou, F., Mbida Mbida, J. A., Njitchouang, G. R., Djeunga, H. N., Asonganyi, T., Simarro, P. P., Cuny, G., & Geiger, A. (2010). Tripartite interactions between tsetse flies, Sodalis glossinidius and trypanosomes—An epidemiological approach in two historical human African trypanosomiasis foci in Cameroon. *Infection, Genetics and Evolution*, 10(1):115–121.

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., & Palsson, B. O. (2007). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, 3(1):121.

Feldhaar, H., Straka, J., Krischke, M., Berthold, K., Stoll, S., Mueller, M. J., & Gross, R. (2007). Nutritional upgrading for omnivorous carpenter ants by the endosymbiont Blochmannia. *BMC Biology*, 5(1):48.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39(4):783–791.

Fenn, K. & Matthews, K. R. (2007). The cell biology of Trypanosoma brucei differentiation. *Current Opinion in Microbiology*, 10(6):539–546.

Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., Blattner, F. R., Maranas, C. D., & Palsson, B. O. (2005). In silico design and adaptive evolution of Escherichia coli for production of lactic acid. *Biotechnology and Bioengineering*, 91(5):643–648.

Franche, C., Lindström, K., & Elmerich, C. (2009). Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. *Plant and Soil*, 321(1-2):35–59.

François, P. & Hakim, V. (2004). Design of genetic networks with specified functions by evolution in silico. *Proceedings of the National Academy of Sciences*, 101(2):580–585.

Frederico, L. A., Kunkel, T. A., & Shaw, B. R. (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry*, 29(10):2532–2537.

Fukatsu, T., Nikoh, N., Smith, W. A., Tanaka, K., Nikoh, N., Sasaki-Fukatsu, K., Yoshizawa, K., Dale, C., & Clayton, D. H. (2007). Bacterial endosymbiont of the slender pigeon louse Columbicola columbae, allied to endosymbionts of grain weevils and tsetse flies. *Applied and Environmental Microbiology*, 73:6660–6668.

Garcia, S. & Trinh, C. (2019). Comparison of multi-objective evolutionary algorithms to solve the modular cell design problem for novel biocatalysis. *bioRxiv*, page 10.1101/616078.

Geiger, A., Ravel, S., Mateille, T., Janelle, J., Patrel, D., Cuny, G., & Frutos, R. (2006). Vector competence of Glossina palpalis gambiensis for Trypanosoma brucei s.l. and genetic diversity of the symbiont Sodalis glossinidius. *Molecular Biology and Evolution*, 24(1):102–109.

Gevers, D., Knight, R., Petrosino, J. F., Huang, K., McGuire, A. L., Birren, B. W., Nelson, K. E., White, O., Methé, B. A., Huttenhower, C., & et al (2012). The Human Microbiome Project: A community resource for the healthy human microbiome. *PLoS Biology*, 10(8):e1001377.

Gherna, R. L., Werren, J. H., Weisberg, W., Cote, R., Woese, C. R., Mandelco, L., & Brenner, D. J. (1991). Arsenophonus nasoniae gen. nov., sp. nov. the causative agent of the son-killer trait in the parasitic wasp Nasonia vitripennis. *International Journal of Systematic Bacteriology*, 41(4):563–565.

Gil, R., Belda, E., Gosables, M., Delaye, L., Vallier, A., Vincent-Monegat, C., Heddi, A., Silva, F., Moya, A., & Latorre, A. (2008). Massive presence of insertion sequences in the genome of SOPE, the primary endosymbiont of the rice weevil Sitophilus oryzae. *International Microbiology*, 11:41–48.

Gil, R., Silva, F. J., Zientz, E., Delmotte, F., Gonzá lez Candelas, F., Latorre, A., Rausell, C., Kamerbeek, J., rgen Gadau, J., Hö lldobler, B., H J van Ham, R. C., Gross, R., & Moya, A. (2003). The genome sequence of Blochmannia floridanus: Comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences*, 100(16):9388–9393.

Gonella, E., Pajoro, M., Marzorati, M., Crotti, E., Mandrioli, M., Pontini, M., Bulgari, D., Negri, I., Sacchi, L., Chouaia, B., Daffonchio, D., & Alma, A. (2015). Plant-mediated interspecific horizontal transmission of an intracellular symbiont in insects. *Scientific Reports*, 5:15811.

González-Domenech, C., Belda, E., Patiño-Navarrete, R., Moya, A., Peretó, J., & Latorre, A. (2012). Metabolic stasis in an ancient symbiosis: Genome-scale metabolic networks from two Blattabacterium cuenoti strains, primary endosymbionts of cockroaches. *BMC Microbiology*, 12(Suppl 1):S5.

Gonzalez-Lopez, J., Salmeron, V., Moreno, J., & Ramos-Cormenzana, A. (1983). Amino acids and vitamins produced by Azotobacter vinelandii ATCC 12837 in

chemically-defined media and dialysed soil media. *Soil Biology and Biochemistry*, 15(6):711–713.

Goodhead, I., Blow, F., Brownridge, P., Hughes, M., Kenny, J., Krishna, R., McLean, L., Pongchaikul, P., Beynon, R., & Darby, A. C. (2018). Large scale and significant expression from pseudogenes in Sodalis glossinidius - a facultative bacterial endosymbiont. *bioRxiv*, page 124388.

Gray, D. J., Ross, A. G., Li, Y.-S., & McManus, D. P. (2011). Diagnosis and management of schistosomiasis. *British Medical Journal*, 342:d2651.

Grindle, N., Tyner, J. J., Clay, K., & Fuqua, C. (2003). Identification of Arsenophonus-type bacteria from the dog tick Dermacentor variabilis. *Journal of Invertebrate Pathology*, 83(3):264–266.

Grünwald, S., Pilhofer, M., & Höll, W. (2010). Microbial associations in gut systems of wood- and bark-inhabiting longhorned beetles [Coleoptera: Cerambycidae]. *Systematic and Applied Microbiology*, 33(1):25–34.

Guccione, E., del Rocio Leon-Kempis, M., Pearson, B. M., Hitchin, E., Mulholland, F., van Diemen, P. M., Stevens, M. P., & Kelly, D. J. (2008). Amino acid-dependent growth of Campylobacter jejuni: Key roles for aspartase (AspA) under microaerobic and oxygen-limited conditions and identification of AspB (Cj0762), essential for growth on glutamate. *Molecular Microbiology*, 69(1):77–93.

Gumiel, M., da Mota, F. F., Rizzo, V. d. S., Sarquis, O., Castro, D. P. d., Lima, M. M., Garcia, E. d. S., Carels, N., & Azambuja, P. (2015). Characterization of the microbiota in the guts of Triatoma brasiliensis and Triatoma pseudomaculata infected by Trypanosoma cruzi in natural conditions using culture independent methods. *Parasites & Vectors*, 8(1):245.

Gusmão, D. S., Santos, A. V., Marini, D. C., Bacci, M., Berbert-Molina, M. A., & Lemos, F. J. A. (2010). Culture-dependent and culture-independent characterization of microorganisms associated with Aedes aegypti (Diptera: Culicidae) (L.) and dynamics of bacterial colonization in the midgut. *Acta Tropica*, 115(3):275–281.

Haigh, J. (1978). The accumulation of deleterious genes in a population—Muller's Ratchet. *Theoretical Population Biology*, 14(2):251–267.

Haines, L., Haddow, J., Aksoy, S., Gooding, R., & Pearson, T. (2002). The major protein in the midgut of teneral Glossina morsitans morsitans is a molecular chaperone from the endosymbiotic bacterium Wigglesworthia glossinidia. *Insect Biochemistry and Molecular Biology*, 32(11):1429–1438.

Haines, L. R., Thomas, J. M., Jackson, A. M., Eyford, B. A., Razavi, M., Watson, C. N., Gowen, B., Hancock, R. E. W., & Pearson, T. W. (2009). Killing of trypanosomatid parasites by a modified bovine host defense peptide, BMAP-18. *PLoS Neglected Tropical Diseases*, 3(2):e373.

Hall, R. J., Flanagan, L. A., Bottery, M. J., Springthorpe, V., Thorpe, S., Darby, A. C., Wood, A. J., & Thomas, G. H. (2019). A tale of three species: Adaptation of Sodalis glossinidius to tsetse biology, Wigglesworthia metabolism, and host diet. *mBio*, 10(1):02106–18.

Handman, E. (2001). Leishmaniasis: current status of vaccine development. *Clinical Microbiology Reviews*, 14(2):229–43.

Hanna, B. A. & Lilly, D. M. (1974). Growth of Uronema marinum in chemically defined medium. *Marine Biology*, 26(2):153–160.

Hansen, A. K., Jeong, G., Paine, T. D., & Stouthamer, R. (2007). Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California. *Applied and Environmental Microbiology*, 73(23):7531–7535.

Harcombe, W. R., Delaney, N. F., Leiby, N., Klitgord, N., & Marx, C. J. (2013). The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. *PLoS Computational Biology*, 9(6):e1003091.

Hargrove, J. (1976). Amino acid metabolism during flight in tsetse flies. *Journal of Insect Physiology*, 22(2):309–313.

Heddi, A., Charles, H., Khatchadourian, C., Bonnot, G., & Nardon, P. (1997). Molecular characterization of the principal symbiotic bacteria of the weevil Sitophilus oryzae: A peculiar G + C content of an endocytobiotic DNA. *Journal of Molecular Evolution*, 47:52–61.

Hedges, L. M., Brownlie, J. C., O'Neill, S. L., & Johnson, K. N. (2008). Wolbachia and virus protection in insects. *Science*, 322(5902):702.

Heller, K. (2011). Tsetse flies rely on symbiotic Wigglesworthia for immune system development. *PLoS Biology*, 9(5):e1001070.

Higgins, C. F. (2001). ABC transporters: Physiology, structure and mechanism - An overview. *Research in Microbiology*, 152(3-4):205–210.

Higgins, C. F., Hiles, I. D., Salmond, G. P. C., Gill, D. R., Downie, J. A., Evans, I. J., Holland, I. B., Gray, L., Buckel, S. D., Bell, A. W., & Hermodson, M. A. (1986). A family of related ATP-binding subunits coupled to many distinct biological processes in bacteria. *Nature*, 323(6087):448–450.

Hillesland, H., Read, A., Subhadra, B., Hurwitz, I., McKelvey, R., Ghosh, K., Das, P., & Durvasula, R. (2008). Identification of aerobic gut bacteria from the kala azar vector, Phlebotomus argentipes: A platform for potential paratransgenic manipulation of sand flies. *American Journal of Tropical Medicine and Hygiene*, 79(6):881–886.

Hindré, T., Knibbe, C., Beslon, G., & Schneider, D. (2012). New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Reviews Microbiology*, 10(5):352–365.

Hoffmann, A. A., Montgomery, B. L., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P. H., Muzzi, F., Greenfield, M., Durkan, M., Leong, Y. S., Dong, Y., & et al (2011). Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature*, 476(7361):454–457.

Hollenbach, A. D., Dickson, K. A., & Washabaugh, M. W. (2002). Overexpression, purification, and characterization of the periplasmic space thiamin-binding protein of the thiamin traffic ATPase in Escherichia coli. *Protein Expression and Purification*, 25(3):508–518.

Hollenstein, K., Dawson, R. J., & Locher, K. P. (2007a). Structure and mechanism of ABC transporter proteins. *Current Opinion in Structural Biology*, 17(4):412–418.

Hollenstein, K., Frei, D. C., & Locher, K. P. (2007b). Structure of an ABC transporter in complex with its binding protein. *Nature*, 446:213–216.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M. C., Wides, R., & et al (2002). The genome sequence of the malaria mosquito Anopheles gambiae. *Science*, 298(5591):129–149.

Horn, D. (2014). Antigenic variation in African trypanosomes. *Molecular & Biochemical Parasitology*, 195:123–129.

Hosokawa, T., Kaiwa, N., Matsuura, Y., Kikuchi, Y., & Fukatsu, T. (2015). Infection prevalence of Sodalis symbionts among stinkbugs. *Zoological Letters*, 1(1):5.

Hrusa, G., Farmer, W., Weiss, B. L., Applebaum, T., Roma, J. S., Szeto, L., Aksoy, S., & Runyen-Janecky, L. J. (2015). TonB-dependent heme iron acquisition in the

tsetse fly symbiont Sodalis glossinidius. *Applied and Environmental Microbiology*, 81(February):04166–14.

Hugenholtz, P., Goebel, B. M., & Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology*, 180(18):4765 – 4774.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., Creasy, H. H., Earl, A. M., FitzGerald, M. G., Fulton, R. S., & et al (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214.

Huynh, K. & Partch, C. L. (2015). Analysis of protein stability and ligand interactions by thermal shift assay. In: *Current Protocols in Protein Science*, volume 79, pages 1–28. John Wiley & Sons, Inc.

Hypsa, V. & Dale, C. (1997). In vitro culture and phylogenetic analysis of Candidatus Arsenophonus triatominarum, an intracellular bacterium from the triatomine bug, Triatoma infestans. *International Journal of Systematic Bacteriology*, 47(4):1140–1144.

Jensen, P. R. & Hammer, K. (1993). Minimal requirements for exponential growth of Lactococcus lactis. *Applied and Environmental Microbiology*, 59(12):4363–6.

Jiang, Z.-F., Xia, F., Johnson, K. W., Bartom, E., Tuteja, J. H., Stevens, R., Grossman, R. L., Brumin, M., White, K. P., & Ghanim, M. (2012). Genome sequences of the primary endosymbiont "Candidatus Portiera aleyrodidarum"; in the whitefly Bemisia tabaci B and Q biotypes. *Journal of Bacteriology*, 194(23):6678–6679.

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. Technical Report 3.

Juan, J., Armenteros, A., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37:420–423.

Kaiwa, N., Hosokawa, T., Kikuchi, Y., Nikoh, N., Meng, X. Y., Kimura, N., Ito, M., & Fukatsu, T. (2010). Primary gut symbiont and secondary, Sodalis-allied symbiont of the Scutellerid stinkbug Cantao ocellatus. *Applied and Environmental Microbiology*, 76(11):3486–94.

Kaiwa, N., Hosokawa, T., Nikoh, N., Tanahashi, M., Moriyama, M., Meng, X.-Y., Maeda, T., Yamaguchi, K., Shigenobu, S., Ito, M., & Fukatsu, T. (2014).

Symbiont-supplemented maternal investment underpinning host's ecological adaptation. *Current Biology*, 24(20):2465–2470.

Käll, L., Krogh, A., & Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, 338(5):1027–1036.

Kanehisa, M. & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30.

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., & Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Research*, 47(D1):D590–D595.

Karimi, S., Askari Seyahooei, M., Izadi, H., Bagheri, A., & Khodaygan, P. (2019). Effect of Arsenophonus endosymbiont elimination on fitness of the date palm hopper, Ommatissus lybicus (Hemiptera: Tropiduchidae). *Environmental Entomology*.

Kauffman, K. J., Prakash, P., & Edwards, J. S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14(5):491–496.

Kelly, P. H., Bahr, S. M., Serafim, T. D., Ajami, N. J., Petrosino, J. F., Meneses, C., Kirby, J. R., Valenzuela, J. G., Kamhawi, S., & Wilson, M. E. (2017). The gut microbiome of the vector Lutzomyia longipalpis is essential for survival of Leishmania infantum. *mBio*, 8(1):01121–16.

Kennedy, P. G. E. (2004). Human African trypanosomiasis of the CNS: Current issues and challenges. *Journal of Clinical Investigation*, 113(4):496–504.

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., & et al (2017). The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Research*, 45(D1):D543–D550.

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., & Lewis, N. E. (2016). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, 44(D1):D515–D522.

Kino, K., Hirao-Suzuki, M., Morikawa, M., Sakaga, A., & Miyazawa, H. (2017). Generation, repair and replication of guanine oxidation products. *Genes and Environment*, 39:21.

Knoop, H., Gründel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., & Steuer, R. (2013). Flux balance analysis of cyanobacterial metabolism: The netabolic network of Synechocystis sp. PCC 6803. *PLoS Computational Biology*, 9(6):e1003081.

Knoop, H., Zilliges, Y., Lockau, W., & Steuer, R. (2010). The metabolic network of Synechocystis sp. PCC 6803: Systemic properties of autotrophic growth. *Plant Physiology*, 154(1):410–422.

Koch, A. L. (1997). Microbial physiology and ecology of slow growth. *Microbiology and Molecular Biology Reviews*, 61(3):305 – 318.

Kodama, Y. & Fujishima, M. (2009). Infection of Paramecium bursaria by symbiotic Chlorella species. In: *Endosymbionts in Paramecium*, pages 31–55. Springer, Berlin, Heidelberg.

Kodama, Y. & Fujishima, M. (2010a). *Induction of secondary symbiosis between the ciliate Paramecium and the green alga Chlorella.*

Kodama, Y. & Fujishima, M. (2010b). Secondary symbiosis between Paramecium and Chlorella cells. *International Review of Cell and Molecular Biology*, 279:33–77.

Koga, R., Bennett, G. M., Cryan, J. R., & Moran, N. A. (2013). Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. *Environmental Microbiology*, 15(7):2073–2081.

Koga, R. & Moran, N. A. (2014). Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *The ISME Journal*, 8:1237.

Komaki, K. & Ishikawa, H. (1999). Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. *Journal of Molecular Evolution*, 48(6):717–722.

Koropatnick, T. A., Kimbell, J. R., & Mcfall-Ngai, M. J. (2007). Responses of host hemocytes during the initiation of the squid-Vibrio symbiosis. *The Biological Bulletin*, 212:29–39.

Krishnan, H. B., Natarajan, S. S., Bennett, J. O., & Sicher, R. C. (2011). Protein and metabolite composition of xylem sap from field-grown soybeans (Glycine max). *Planta*, 233:921–931.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, 35(6):1547–1549.

Lal, P. B., Schneider, B. L., Vu, K., & Reitzer, L. (2014). The redundant aminotransferases in lysine and arginine synthesis and the extent of aminotransferase redundancy in Escherichia coli. *Molecular Microbiology*, 94(4):843–856.

Lane, N. (2017). Serial endosymbiosis or singular event at the origin of eukaryotes? *Journal of Theoretical Biology*, 434:58–67.

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., & et al (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21):2947–2948.

Leach, S., Harvey, P., & Wait, R. (1997). Changes with growth rate in the membrane lipid composition of and amino acid utilization by continuous cultures of Campylobacter jejuni. *Journal of Applied Microbiology*, 82(5):631 – 640.

Lee, A. H., Symington, L. S., & Fidock, D. A. (2014). DNA repair mechanisms and their biological roles in the malaria parasite Plasmodium falciparum. *Microbiology and Molecular Biology Reviews*, 78(3):469–486.

Lee, J. M., Gianchandani, E. P., & Papin, J. A. (2006). Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics*, 7(2):140–150.

Letunic, I. & Bork, P. (2007). Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23(1):127–128.

Lewis, N. E., Nagarajan, H., & Palsson, B. O. (2012). Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305.

Li, K., Chen, H., Jiang, J., Li, X., Xu, J., & Ma, Y. (2016). Diversity of bacteriome associated with Phlebotomus chinensis (Diptera: Psychodidae) sand flies in two wild populations from China. *Scientific Reports*, 6:36406.

Linton, K. J. & Higgins, C. F. (1998). The Escherichia coli ATP-binding cassette (ABC) proteins. *Molecular Microbiology*, 28(1):5–13.

Login, F. H., Balmand, S., Vallier, A., Vincent-Monégat, C., Vigneron, A., Weiss-Gayet, M., Rochat, D., & Heddi, A. (2011). Antimicrobial peptides keep insect endosymbionts under control. *Science*, 334(6054):362 – 365.

López-Madrigal, S., Latorre, A., Porcar, M., Moya, A., & Gil, R. (2011). Complete genome sequence of Candidatus Tremblaya princeps strain PCVAL, an intriguing translational machine below the living-cell status. *Journal of Bacteriology*, 193(19):5587–5588.

Luria, S. E. & Burrous, J. W. (1957). Hybridization between Escherichia coli and Shigella. *Journal of Bacteriology*, 74(4):461–476.

Ly Thao, M. & Baumann, P. (2004a). Evidence for multiple acquisition of Arsenophonus by whitefly species (Sternorrhyncha: Aleyrodidae). *Current Microbiology*, 48:140–144.

Ly Thao, M. & Baumann, P. (2004b). Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Applied and Environmental Microbiology*, 70(6):3401–3406.

Lynn, D. E. (2002). Methods for maintaining insect cell cultures. *Journal of Insect Science*, 2:9.

Macdonald, S. J., Lin, G. G., Russell, C. W., Thomas, G. H., & Douglas, A. E. (2012). The central role of the host cell in symbiotic nitrogen metabolism. *Proceedings of the Royal Society B*, 279:2965–2973.

Macdonald, S. J., Thomas, G. H., & Douglas, A. E. (2011). Genetic and metabolic determinants of nutritional phenotype in an insect-bacterial symbiosis. *Molecular Ecology*, 20(10):2073–2084.

Mahadevan, R., Edwards, J. S., & Doyle, F. J. (2002). Dynamic flux balance analysis of diauxic growth in Escherichia coli. *Biophysical Journal*, 83(3):1331–1340.

Maltz, M. A., Weiss, B. L., O'Neill, M., Wu, Y., & Aksoy, S. (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):7760–7768.

Mantilla, B. S., Marchese, L., Casas-Sánchez, A., Dyer, N. A., Ejeh, N., Biran, M., Bringaud, F., Lehane, M. J., Acosta-Serrano, A., & Silber, A. M. (2017). Proline metabolism is essential for Trypanosoma brucei brucei survival in the tsetse vector. *PLoS Pathogens*, 13(1):e1006158.

Manzano-Marín, A., Oceguera-Figueroa, A., Latorre, A., Jiménez-García, L. F., & Moya, A. (2015). Solving a bloody mess: B-vitamin independent metabolic convergence among gammaproteobacterial obligate endosymbionts from blood-feeding arthropods and the leech Haementeria officinalis. *Genome Biology and Evolution*, 7(10):2871–2884.

Maqbool, A., Levdikov, V. M., Blagova, E. V., Hervé, M., Horler, R. S. P., Wilkinson, A. J., & Thomas, G. H. (2011). Compensating stereochemical changes allow

murein tripeptide to be accommodated in a conventional peptide-binding protein. *Journal of Biological Chemistry*, 286(36):31512–31521.

Margulis, L. (1993). *Symbiosis in cell evolution: Microbial communities in the Archean and Proterozoic eons*, (second ed.). Freeman New York.

Martinson, V. G., Danforth, B. N., Minckley, R. L., Rueppell, O., Tingek, S., & Moran, N. A. (2011). A simple and distinctive microbiota associated with honey bees and bumble bees. *Molecular Ecology*, 20(3):619–628.

Matsuura, Y., Hosokawa, T., Serracin, M., Tulgetske, G. M., Miller, T. A., & Fukatsu, T. (2014). Bacterial symbionts of a devastating coffee plant pest, the stinkbug Antestiopsis thunbergii (Hemiptera: Pentatomidae). *Applied and Environmental Microbiology*, 80(12):3769–3775.

McCutcheon, J. P., McDonald, B. R., & Moran, N. A. (2009a). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proceedings of the National Academy of Sciences*, 106(36):15394–9.

McCutcheon, J. P., McDonald, B. R., & Moran, N. A. (2009b). Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. *PLoS Genetics*, 5(7):e1000565.

McCutcheon, J. P. & Moran, N. A. (2007). Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proceedings of the National Academy of Sciences*, 104(49):19392–19397.

McCutcheon, J. P. & Moran, N. A. (2011). Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, 10:13.

McCutcheon, J. P. & von Dohlen, C. D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Current Biology*, 21(16):1366–1372.

McKean, P. G. (2003). Coordination of cell cycle and cytokinesis in Trypanosoma brucei. *Current Opinion in Microbiology*, 6(6):600–607.

McNally, C. P. & Borenstein, E. (2017). Metabolic model-based analysis of the emergence of bacterial cross-feeding through extensive gene loss. *bioRxiv*, page 10.1101/180208.

Medina, M. & Sachs, J. (2010). Symbiont genomics, our new tangled bank. *Genomics*, 95(3):129–137.

Mellor, P. S., Boorman, J., & Baylis, M. (2000). Culicoides biting midges: their role as arbovirus vectors. *Annual Review of Entomology*, 45:307–340.

Methé, B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., Gevers, D., Petrosino, J. F., Abubucker, S., Badger, J. H., & Al, E. (2012). A framework for human microbiome research. *Nature*, 486(7402):215–221.

Michaels, M. L. & Miller, J. H. (1992). The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *Journal of Bacteriology*, 174(20):6321–6325.

Michalik, A., Jankowska, W., Kot, M., Gołas, A., & Szklarzewicz, T. (2014). Symbiosis in the green leafhopper, Cicadella viridis (Hemiptera, Cicadellidae). Association in statu nascendi? *Arthropod Structure & Development*, 43(6):579–587.

Michalik, A., Szwedo, J., Stroiński, A., Świerczewski, D., & Szklarzewicz, T. (2018). Symbiotic cornucopia of the monophagous planthopper Ommatidiotus dissimilis (Fallén, 1806) (Hemiptera: Fulgoromorpha: Caliscelidae). *Protoplasma*, 255(5):1317–1329.

Michalkova, V., Benoit, J. B., Weiss, B. L., Attardo, G. M., & Aksoy, S. (2014). Vitamin B6 generated by obligate symbionts is critical for maintaining proline homeostasis and fecundity in tsetse flies. *Applied and Environmental Microbiology*, 80(18):5844–5853.

Migchelsen, S. J., Büscher, P., Hoepelman, A. I. M., Schallig, H. D. F. H., & Adams, E. R. (2011). Human African trypanosomiasis: a review of non-endemic cases in the past 20 years. *International Journal of Infectious Diseases*, 15(8):e517–e524.

Moloo, S. K., Kabata, J. M., & Sabwa, C. L. (1994). A study on the maturation of procyclic Trypanosoma brucei brucei in Glossina morsitans centralis and G. brevipalpis. *Medical and Veterinary Entomology*, 8(4):369–374.

Montllor, C. B., Maxmen, A., & Purcell, A. H. (2002). Facultative bacterial endosymbionts benefit pea aphids Acyrthosiphon pisum under heat stress. *Ecological Entomology*, 27(2):189–195.

Moran, N. A. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences*, 93:2873–2878.

Moran, N. A. (2002). Microbial minimalism: Genome reduction in bacterial pathogens. *Cell*, 108(5):583–586.

Moran, N. A., Mccutcheon, J. P., & Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annual Review of Genetics*, 42:165–190.

Moran, N. A. & Mira, A. (2001). The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. *Genome Biology*, 2:0054.1–0054.12.

Moran, N. A., Munson, M. A., Baumann, P., & Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society B: Biological Sciences*, 253:167–171.

Moran, N. A. & Wernegreen, J. J. (2000). Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends in Ecology & Evolution*, 15(8):321–326.

Moreira, L. A., Iturbe-Ormaetxe, I., Jeffery, J. A., Lu, G., Pyke, A. T., Hedges, L. M., Rocha, B. C., Hall-Mendelin, S., Day, A., Riegler, M., & et al (2009). A Wolbachia symbiont in Aedes aegypti limits infection with Dengue, Chikungunya, and Plasmodium. *Cell*, 139(7):1268–1278.

Muller, H. (1964). The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1):2–9.

Munson, M. A., Baumann, P., & Kinsey, M. G. (1991). Buchnera gen. nov. and Buchnera aphidicola sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. *International Journal of Systematic Bacteriology*, 41(4):566–568.

Nakabachi, A., Ueoka, R., Oshima, K., Teta, R., Mangoni, A., Gurgui, M., Oldham, N. J., van Echten-Deckert, G., Okamura, K., Yamamoto, K., & et al (2013). Defensive bacteriome symbiont with a drastically reduced genome. *Current Biology*, 23(15):1478–1484.

Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H. E., Moran, N. A., & Hattori, M. (2006). The 160-Kilobase genome of the bacterial endosymbiont Carsonella. *Science*, 314:267.

Nautiyal, C. (1999). An efficient microbiological growth medium for screening phosphate solubilizing microorganisms. *FEMS Microbiology Letters*, 170(1):265–270.

Neidhardt, F. C., Bloch, P. L., & Smith, D. F. (1974). Culture medium for enterobacteria. *Journal of Bacteriology*, 119(3):736–47.

Nene, V., Wortman, J. R., Lawson, D., Haas, B., Kodira, C., Tu, Z., Loftus, B., Xi, Z., Megy, K., Grabherr, M., & Al, E. (2007). Genome sequence of Aedes aegypti, a major arbovirus vector. *Science*, 316(5832):1718–1723.

Nikoh, N., Hosokawa, T., Oshima, K., Hattori, M., & Fukatsu, T. (2011). Reductive evolution of bacterial genome in insect gut environment. *Genome Biology and Evolution*, 3:702–714.

Nogales, J., Gudmundsson, S., Knight, E. M., Palsson, B. O., & Thiele, I. (2012). Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. *Proceedings of the National Academy of Sciences*, 109(7):2678–83.

Nogge, G. (1976). Sterility in tsetse flies (Glossina morsitans Westwood) caused by loss of symbionts. *Experientia*, 32(8):995–996.

Nogge, G. (1978). Aposymbiotic tsetse flies, Glossina morsitans morsitans obtained by feeding on rabbits immunized specifically with symbionts. *Journal of Insect Physiology*, 24(4):299–304.

Normark, B. B. (2004). The strange case of the armored scale insect and its bacteriome. *PLoS Biology*, 2(3):e43.

Novakova, E. & Hypsa, V. (2007). A new Sodalis lineage from bloodsucking fly Craterina melbae (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont Sodalis glossinidius. *FEMS Microbiology Letters*, 269(1):131–135.

Nováková, E., Hypša, V., & Moran, N. A. (2009). Arsenophonus, an emerging clade of intracellular symbionts with a broad host distribution.

Nyholm, S. V. & McFall-Ngai, M. (2004). The winnowing: establishing the squid–vibrio symbiosis. *Nature Reviews Microbiology*, 2(8):632–642.

Oakeson, K. F., Gil, R., Clayton, A. L., Dunn, D. M., von Niederhausern, A. C., Hamil, C., Aoyagi, A., Duval, B., Baca, A., Silva, F. J., Vallier, A., Jackson, D. G., Latorre, A., Weiss, R. B., Heddi, A., Moya, A., & Dale, C. (2014). Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biology and Evolution*, 6(1):76–93.

Oerke, E.-C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, 144(1):31–43.

Oliver, K. M., Campos, J., Moran, N. A., & Hunter, M. S. (2008). Population dynamics of defensive symbionts in aphids. *Proceedings of the Royal Society B*, 275:293–299.

Oliver, K. M., Degnan, P. H., Burke, G. R., & Moran, N. A. (2010). Facultative symbionts in aphids and the horizontal transfer of ecologically important traits. *Annual Review of Entomology*, 55:247–266.

Oliver, K. M. & Higashi, C. H. (2019). Variations on a protective theme: Hamiltonella defensa infections in aphids variably impact parasitoid success. *Current Opinion in Insect Science*, 32:1–7.

Oliver, K. M., Moran, N. A., & Hunter, M. S. (2006). Costs and benefits of a superinfection of facultative symbionts in aphids. *Proceedings of the Royal Society B*, 273:1273–1280.

Oliver, K. M., Russell, J. A., Moran, N. A., & Hunter, M. S. (2003). Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, 100(4):1803–1807.

O'Neill, S. L., Gooding, R. H., & Aksoy, S. (1993). Phylogenetically distant symbiotic microorganisms reside in Glossina midgut and ovary tissues. *Medical and Veterinary Entomology*, 7:377–383.

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. O. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism–2011. *Molecular Systems Biology*, 7(1):535.

Orth, J. D. & Palsson, B. O. (2012). Gap-filling analysis of the iJO1366 Escherichia coli metabolic network reconstruction for discovery of metabolic functions. *BMC Systems Biology*, 6(1):30.

Orth, J. D., Thiele, I., & Palsson, B. (2010). What is flux balance analysis? *Nature Biotechnology*, 28(3):245–248.

Oulhen, N., Schulz, B. J., & Carrier, T. J. (2016). English translation of Heinrich Anton de Bary's 1878 speech, 'Die Erscheinung der Symbiose' ('De la symbiose'). *Symbiosis*, 69:131–139.

Pais, R., Lohs, C., Wu, Y., Wang, J., & Aksoy, S. (2008). The obligate mutualist Wigglesworthia glossinidia influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and Environmental Microbiology*, 74(19):5965–5974.

Pál, C., Papp, B., Lercher, M. J., Csermely, P., Oliver, S. G., & Hurst, L. D. (2006). Chance and necessity in the evolution of minimal metabolic networks. *Nature*, 440(7084):667–670.

Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., & et al (2000). The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature*, 403(6770):665–668.

Parkhill, J., Wren, B. W., Thomson, N. R., Titball, R. W., Holden, M. T. G., Prentice, M. B., Sebaihia, M., James, K. D., Churcher, C., Mungall, K. L., &

et al (2001). Genome sequence of Yersinia pestis, the causative agent of plague. *Nature*, 413(6855):523–527.

Patouillard, E., Griffin, J., Bhatt, S., Ghani, A., & Cibulskis, R. (2017). Global investment targets for malaria control and elimination between 2016 and 2030. *BMJ Global Health*, 2(2):e000176.

Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J. M., Silva, F. J., Moya, A., & Latorre, A. (2006). A small microbial genome: The end of a long symbiotic relationship? *Science*, 314(5797):312–313.

Pettersson, M. E. & Berg, O. G. (2007). Muller's ratchet in symbiont populations. *Genetica*, 130(2):199–211.

Pfeiffer, T., Soyer, O. S., & Bonhoeffer, S. (2005). The evolution of connectivity in metabolic networks. *PLoS Biology*, 3(7):e228.

Pontes, M. H. & Dale, C. (2006). Culture and manipulation of insect facultative symbionts. *Trends in Microbiology*, 14(9):406–412.

Pontes, M. H. & Dale, C. (2011). Lambda red-mediated genetic modification of the insect endosymbiont Sodalis glossinidius. *Applied and Environmental Microbiology*, 77(5):1918–1920.

Prakasham, R. S., Sreenivas Rao, R., & Hobbs, P. J. (2009). Current trends in biotechnological production of xylitol and future prospects. *Current Trends in Biotechnology and Pharmacy*, 3(1):8–36.

Prosser, W. & Douglas, A. (1991). The aposymbiotic aphid: An analysis of chlortetracycline-treated pea aphid, Acyrthosiphon pisum. *Journal of Insect Physiology*, 37(10):713–719.

Quiocho, F. A. & Ledvina, P. S. (1996). Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Molecular Microbiology*, 20(1):17–25.

Rani, A., Sharma, A., Rajagopal, R., Adak, T., & Bhatnagar, R. K. (2009). Bacterial diversity analysis of larvae and adult midgut microflora using culture-dependent and culture-independent methods in lab-reared and field-collected Anopheles stephensi-an Asian malarial vector. *BMC Microbiology*, 9(1):96.

Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology*, 4(9):R54.

Reisser, W. (1980). The metabolic interactions between Paramecium bursaria Ehrbg. and Chlorella spec. in the Paramecium bursaria-symbiosis. *Archives of Microbiology*, 125(3):291–293.

Renesto, P., Crapoulet, N., Ogata, H., La Scola, B., Vestris, G., Claverie, J.-M., & Raoult, D. (2003). Genome-based design of a cell-free culture medium for Tropheryma whipplei. *The Lancet*, 362(9382):447–449.

Richards, S., Gibbs, R. A., Gerardo, N. M., Moran, N., Nakabachi, A., Richards, S., Stern, D., Tagu, D., Wilson, A. C. C., & et al (2010). Genome Sequence of the Pea Aphid Acyrthosiphon pisum. *PLoS Biology*, 8(2):e1000313.

Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K., Blattner, F. R., Chaudhuri, R. R., Glasner, J. D., Horiuchi, T., Keseler, I. M., Kosuge, T., & et al (2006). Escherichia coli K-12: a cooperatively developed annotation snapshot–2005. *Nucleic Acids Research*, 34(1):1–9.

Rio, R. V., Attardo, G. M., & Weiss, B. L. (2016). Grandeur alliances: symbiont metabolic integration and obligate arthropod hematophagy. *Trends in Parasitology*, 32(9):739 –749.

Rio, R. V., Hu, Y., & Aksoy, S. (2004). Strategies of the home-team: symbioses exploited for vector-borne disease control. *Trends in Microbiology*, 12(7):325–336.

Rio, R. V. M., Jozwick, A. K. S., Savage, A. F., Sabet, A., Vigneron, A., Wu, Y., Aksoy, S., & Weiss, B. L. (2019). Mutualist-provisioned resources impact vector competency. *mBio*, 10(3):00018–19.

Rio, R. V. M., Lefevre, C., Heddi, A., & Aksoy, S. (2003). Comparative genomics of insect-symbiotic bacteria: influence of host environment on microbial genome composition. *Applied and Environmental Microbiology*, 69(11):6825–6832.

Rio, R. V. M., Symula, R. E., & Wang, J. (2012). Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont Wigglesworthia. *mBio*, 3(1):1–13.

Rispe, C. & Moran, N. A. (2000). Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. *The American Naturalist*, 156(4):425–441.

Rodelas, B., Salmerón, V., Martinez-Toledo, M. V., & González-López, J. (1993). Production of vitamins by Azospirillum brasilense in chemically-defined media. *Plant and Soil*, 153(1):97–101.

Roditi, I. & Lehane, M. J. (2008). Interactions between trypanosomes and tsetse flies. *Current Opinion in Microbiology*, 11:345–351.

Rodríguez, J., Pavía, P., Montilla, M., & Puerta, C. J. (2011). Identifying triatomine symbiont Rhodococcus rhodnii as intestinal bacteria from Rhodnius ecuadoriensis (Hemiptera: Reduviidae) laboratory insects. *International Journal of Tropical Insect Science*, 31(1-2):34–37.

Rodriguez-Valera, F., Ruiz-Berraquero, F., & Ramos-Cormenzana, A. (1980). Isolation of extremely halophilic bacteria able to grow in defined inorganic media with single carbon sources. *Microbiology*, 119(2):535–538.

Roma, J. S., D'Souza, S., Somers, P. J., Cabo, L. F., Farsin, R., Aksoy, S., Runyen-Janecky, L. J., & Weiss, B. L. (2019). Thermal stress responses of Sodalis glossinidius, an indigenous bacterial symbiont of hematophagous tsetse flies. *bioRxiv*, 10.1101/63.

Rosas-Pérez, T., de León, A. V.-P., Rosenblueth, M., Ramírez-Puebla, S. T., Rincón-Rosales, R., Martínez-Romero, J., Dunn, M. F., Kondorosi, , & Martínez-Romero, E. (2017). The symbiome of Llaveia Cochineals (Hemiptera: Coccoidea: Monophlebidae) includes a gammaproteobacterial cosymbiont Sodalis TME1 and the known Candidatus Walczuchella monophlebidarum. In: *Insect Physiology and Ecology*, page InTech. InTech.

Rotureau, B. & Van Den Abbeele, J. (2013). Through the dark continent: African trypanosome development in the tsetse fly. *Frontiers in Cellular and Infection Microbiology*, 3:53.

Ruby, E. G. (1996). Lessons from a cooperative, bacteria-animal association: The Vibrio fischeri-Euprymna scolopes light organ symbiosis. *Annual Review of Microbiology*, 50:591–624.

Runyen-Janecky, L. J., Brown, A. N., Ott, B., Tujuba, H. G., & Rio, R. V. M. (2010). Regulation of high-affinity iron acquisition homologues in the tsetse fly symbiont Sodalis glossinidius. *Journal of Bacteriology*, 192(14):3780–3787.

Russell, J. A., Latorre, A., Sabater-Munoz, B., Moya, A., & Moran, N. A. (2003). Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Molecular Ecology*, 12(4):1061–1075.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., & Barrell, B. (2000). Artemis: sequence visualization and annotation. *Bioinformatics*, 16(10):944–945.

Sabree, Z. L., Huang, C. Y., Okusu, A., Moran, N. A., & Normark, B. B. (2013). The nutrient supplying capabilities of Uzinura, an endosymbiont of armoured scale insects. *Environmental Microbiology*, 15(7):1988–1999.

Sagan, L. (1967). On the origin of mitosing cells. *Journal of Theoretical Biology*, 14(3):225–IN6.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: a laboratory manual*. Number Ed. 2. Cold Spring Harbor Laboratory Press.

Sancho-Vaello, E., Fernández-Murga, M. L., & Rubio, V. (2012). Functional dissection of N-Acetylglutamate synthase (ArgA) of pseudomonas aeruginosa and restoration of its ancestral n-acetylglutamate kinase activity. *Journal of Bacteriology*, 194(11):2791–2801.

Sandstrom, J. P., Russell, J. A., White, J. P., & Moran, N. A. (2001). Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular Ecology*, 10(1):217–228.

Santos-Garcia, D., Silva, F. J., Morin, S., Dettner, K., & Kuechler, S. M. (2017). The all-rounder Sodalis: A new bacteriome-associated endosymbiont of the lygaeoid bug Henestaris halophilus (Heteroptera: Henestarinae) and a critical examination of its evolution. *Genome Biology and Evolution*, 9(10):2893–2910.

Scarborough, C. L., Ferrari, J., & Godfray, H. C. J. (2005). Aphid protected from pathogen by endosymbiont. *Science*, 310(5755):1781.

Schilling, C. H., Covert, M. W., Famili, I., Church, G. M., Edwards, J. S., & Palsson, B. O. (2002). Genome-scale metabolic model of Helicobacter pylori 26695. *Journal of bacteriology*, 184(16):4582–4593.

Schlichting, I. (2005). X-ray crystallography of protein–ligand interactions. In: *Protein-Ligand Interactions*, pages 155–166. Humana Press.

Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, 3(1):119.

Segrè, D., Vitkup, D., & Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117.

Shastri, A. & Morgan, J. (2005). Flux balance analysis of photoautotrophic metabolism. *Biotechnology Progress*, 21(6):1617–1626.

Shi, D., Allewell, N. M., & Tuchman, M. (2015). The N-acetylglutamate synthase family: Structures, function and mechanisms. *International Journal of Molecular Sciences*, 16(6):13004–13022.

Shibutani, S., Takeshita, M., & Grollman, A. P. (1997). Translesional synthesis on DNA templates containing a single abasic site. *The Journal of Biological Chemistry*, 272(21):13916–13922.

Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., & Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. *Nature*, 407:81–86.

Shlomi, T., Berkman, O., & Ruppin, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences*, 102(21):7695–7700.

Simarro, P. P., Diarra, A., Ruiz Postigo, J. A., Franco, J. R., & Jannin, J. G. (2011). The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000–2009: the way forward. *PLoS Neglected Tropical Diseases*, 5(2):e1007.

Sims, G. E. & Kim, S.-H. (2011). Whole-genome phylogeny of Escherichia coli/Shigella group by feature frequency profiles (FFPs). *Proceedings of the National Academy of Sciences*, 108(20):8329–8334.

Sinkins SP, Braig HR, O. S. (1995). Wolbachia superinfections and the expression of cytoplasmic incompatibility. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 261(1362):325 – 330.

Smith, C. L., Weiss, B. L., Aksoy, S., & Runyen-Janecky, L. J. (2013). Characterization of the achromobactin iron acquisition operon in Sodalis glossinidius. *Applied and Environmental Microbiology*, 79(9):2872–2881.

Snyder, A. K., Deberry, J. W., Runyen-Janecky, L., & Rio, R. V. M. (2010). Nutrient provisioning facilitates homeostasis between tsetse fly (Diptera: Glossinidae) symbionts. *Proceedings of the Royal Society B: Biological Sciences*, 277(1692):2389–2397.

Snyder, A. K., McLain, C., & Rio, R. V. M. (2012). The tsetse fly obligate mutualist Wigglesworthia morsitans alters gene expression and population density via exogenous nutrient provisioning. *Applied and Environmental Microbiology*, 78(21):7792–7.

Snyder, A. K. & Rio, R. V. M. (2013). Interwoven biology of the tsetse holobiont. *Journal of Bacteriology*, 195(19):4322–4330.

Snyder, A. K. & Rio, R. V. M. (2015). Wigglesworthia morsitans folate (Vitamin B9) biosynthesis contributes to tsetse host fitness. *Applied and Environmental Microbiology*, 81(16):5375 – 5386.

Sørensen, M. E. S., Cameron, D. D., Brockhurst, M. A., & Wood, A. J. (2016). Metabolic constraints for a novel symbiosis. *Royal Society Open Science*, 3(3):150708.

Šorfová, P., Škeříková, A., & Hypša, V. (2008). An effect of 16S rRNA intercistronic variability on coevolutionary analysis in symbiotic bacteria: Molecular phylogeny of Arsenophonus triatominarum. *Systematic and Applied Microbiology*, 31(2):88–100.

Spaulding, A. W. & Von Dohlen, C. D. (2001). Psyllid endosymbionts exhibit patterns of co-speciation with hosts and destabilizing substitutions in ribosomal RNA. *Insect Molecular Biology*, 10(1):57–67.

Staley, J. T. & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Reviews in Microbiology*, 39:321–346.

Stein, W. H. & Moore, S. (1954). The free amino acids of human blood plasma. *The Journal of Biological Chemistry*, 211(2):915–926.

Subandiyah, S., Nikoh, N., Tsuyumu, S., Somowiyarjo, S., & Fukatsu, T. (2000). Complex endosymbiotic microbiota of the citrus psyllid Diaphorina citri (Homoptera: Psylloidea). *Zoological Science*, 17(7):983–989.

Szklarzewicz, T., Kalandyk-Kołodziejczyk, M., Michalik, K., Jankowska, W., & Michalik, A. (2018). Symbiotic microorganisms in Puto superbus (Leonardi, 1907) (Insecta, Hemiptera, Coccomorpha: Putoidae). *Protoplasma*, 255(1):129–138.

Tamas, I., Wernegreen, J. J., Nystedt, B., Kauppinen, S. N., Darby, A. C., Gomez-Valero, L., Lundin, D., Poole, A. M., & Andersson, S. G. E. (2008). Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly (A) tracts. *Proceedings of the National Academy of Sciences*, 105(39):14934–14939.

Thao, M. L., Gullan, P. J., & Baumann, P. (2002). Secondary (gamma-proteobacteria) endosymbionts infect the primary (beta-proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. *Applied and Environmental Microbiology*, 68(7):3190–3197.

Thiele, I. & Palsson, B. O. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1):93–121.

Thiele, I., Vo, T. D., Price, N. D., & Palsson, B. (2005). Expanded metabolic reconstruction of Helicobacter pylori (iIT341 GSM/GPR): An in silico genome-scale characterization of single- and double-deletion mutants. *Journal of Bacteriology*, 187(16):5818–5830.

Thomas, G. H., Zucker, J., Macdonald, S. J., Sorokin, A., Goryanin, I., & Douglas, A. E. (2009). A fragile metabolic network adapted for cooperation in the symbiotic bacterium Buchnera aphidicola. *BMC Systems Biology*, 3:24.

Toh, H., Weiss, B. L., Perkin, S. A. H., Yamashita, A., Oshima, K., Hattori, M., & Aksoy, S. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of Sodalis glossinidius in the tsetse host. *Genome Research*, 16(2):149–156.

Toju, H. & Fukatsu, T. (2011). Diversity and infection prevalence of endosymbionts in natural populations of the chestnut weevil: relevance of local climate and host plants. *Molecular Ecology*, 20(4):853–868.

Toju, H., Hosokawa, T., Koga, R., Nikoh, N., Meng, X. Y., Kimura, N., & Fukatsu, T. (2010). "Candidatus Curculioni bucheri", a novel clade of bacterial endocellular symbionts from weevils of the genus Curculio. *Applied and Environmental Microbiology*, 76:275–282.

Tomb, J.-F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., & et al (1997). The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature*, 388(6642):539–547.

Tonooka, Y. & Watanabe, T. (2002). A natural strain of Paramecium bursaria lacking symbiotic algae. *European Journal of Protistology*, 38(1):55–58.

Trowbridge, R. E., Dittmar, K., & Whiting, M. F. (2006). Identification and phylogenetic analysis of Arsenophonus- and Photorhabdus-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *Journal of Invertebrate Pathology*, 91(1):64–68.

Tsuchida, T., Koga, R., & Fukatsu, T. (2004). Host plant specialization governed by facultative symbiont. *Science*, 303(5666):1989.

Utzinger, J., Becker, S., van Lieshout, L., van Dam, G., & Knopp, S. (2015). New diagnostic tools in schistosomiasis. *Clinical Microbiology and Infection*, 21(6):529–542.

van de Rijn, I. & Kessler, R. E. (1980). Growth characteristics of group A streptococci in a new chemically defined medium. *Infection and Immunity*, 27(2):444 – 448.

van Ham, R. C. H. J., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J. M., Jiménez, L., Postigo, M., Silva, F. J., & et al (2003). Reductive genome evolution in Buchnera aphidicola. *Proceedings of the National Academy of Sciences*, 100(2):581–586.

Varma, A., Boesch, B. W., & Palsson, B. O. (1993). Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates. *Applied and Environmental Microbiology*, 59(8):2465–73.

Varma, A. & Palsson, B. . (1994). Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Biotechnology*, 12:994–998.

Vartoukian, S. R., Palmer, R. M., & Wade, W. G. (2010). Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters*, 309(1):1–7.

Vigneron, A., Masson, F., Vallier, A., Balmand, S., Rey, M., Vincent-Monégat, C., Aksoy, E., Aubailly-Giraud, E., Zaidman-Rémy, A., & Heddi, A. (2014). Insects recycle endosymbionts when the benefit is over. *Current Biology*, 24(19):2267–2273.

von Dohlen, C. D., Kohler, S., Alsop, S. T., & McManus, W. R. (2001). Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature*, 412(6845):433–436.

Walker, T., Johnson, P. H., Moreira, L. A., Iturbe-Ormaetxe, I., Frentiu, F. D., McMeniman, C. J., Leong, Y. S., Dong, Y., Axford, J., Kriesner, P., & et al (2011). The wMel Wolbachia strain blocks dengue and invades caged Aedes aegypti populations. *Nature*, 476(7361):450–453.

Wall, R. & Langley, P. A. (1993). The mating behaviour of tsetse flies (Glossina): a review. *Physiological Entomology*, 18(2):211–218.

Wang, L. & Maranas, C. D. (2018). MinGenome: An n silico top-down approach for the synthesis of minimized genomes. *ACS Synthetic Biology*, 7:462–473.

Webb, E., Claas, K., & Downs, D. (1998). thiBPQ encodes an ABC transporter required for transport of thiamine and thiamine pyrophosphate in Salmonella Typhimurium. *Journal of Biological Chemistry*, 273(15):8946–8950.

Weinert, L. A. & Welch, J. J. (2017). Why might bacterial pathogens have small genomes? *Trends in Ecology & Evolution*, 32(12):936–947.

Weiss, B. & Aksoy, S. (2011). Microbiome influences on insect host vector competence. *Trends in Parasitology*, 27(11):514–522.

Weiss, B. L., Mouchotte, R., Rio, R. V. M., Wu, Y., Wu, Z., Heddi, A., & Aksoy, S. (2006). Interspecific transfer of bacterial endosymbionts between tsetse fly species: infection establishment and effect on host fitness. *Applied and Environmental Microbiology*, 72(11):7013–7021.

Weiss, B. L., Wu, Y., Schwank, J. J., Tolwinski, N. S., & Aksoy, S. (2008). An insect symbiosis is influenced by bacterium-specific polymorphisms in outer-membrane protein A. *Proceedings of the National Academy of Sciences*, 105(39):15088–15093.

Welburn, S. C., Arnold, K., Maudlin, I., & Gooday, G. W. (1993). Rickettsia-like organisms and chitinase production in relation to transmission of trypanosomes by tsetse-flies. *Parasitology*, 107:141–145.

Welburn, S. C. & Maudlin, I. (1991). Rickettsia-like organisms, puparial temperature and susceptibility to trypanosome infection in Glossina morsitans. *Parasitology*, 102(02):201.

Welburn, S. C. & Maudlin, I. (1999). Tsetse – trypanosome interactions : rites of passage. *Parasitology Today*, 15(10):399–403.

Welburn, S. C., Maudlin, I., & Ellis, D. S. (1987). In vitro cultivation of Rickettsia-like-organisms from Glossina spp. *Annals of Tropical Medicine & Parasitology*, 81(3):331–335.

Werren, J. H., Windsor, D., & Guo, L. R. (1995). Distribution of Wolbachia among neotropical arthropods. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 262(1364):197–204.

Whitehead, L. F. & Douglas, A. E. (1993). A metabolic study of Buchnera, the intracellular bacterial symbionts of the pea aphid Acyrthosiphon pisum. *Microbiology*, 139:821–826.

WHO (1998). Control and surveillance of African trypanosomiasis. *WHO Technical Report Series*, 881:1–113.

Wicker, C. (1983). Differential vitamin and choline requirements of symbiotic and aposymbiotic S. oryzae (Coleoptera: Curculionidae). *Comparative Biochemistry and Physiology Part A: Physiology*, 76(1):177–182.

Wilcox, J. L., Dunbar, H. E., Wolfinger, R. D., & Moran, N. A. (2003). Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Molecular Microbiology*, 48(6):1491 – 1500.

Wilkinson, T. & Douglas, A. (1995a). Aphid feeding, as influenced by disruption of the symbiotic bacteria: an analysis of the pea aphid (Acyrthosiphon pisum). *Journal of Insect Physiology*, 41(8):635–640.

Wilkinson, T. & Douglas, A. (1995b). Why pea aphids (Acyrthosiphon pisum) lacking symbiotic bacteria have elevated levels of the amino acid glutamine. *Journal of Insect Physiology*, 41(11):921–927.

Wilson, A. C. C., Ashton, P. D., Calevro, F., Charles, H., Colella, S., Febvay, G., Jander, G., Kushlan, P. F., Macdonald, S. J., Schwartz, J. F., & et al (2010). Genomic insight into the amino acid relations of the pea aphid, Acyrthosiphon pisum, with its symbiotic bacterium Buchnera aphidicola. *Insect Molecular Biology*, 19:249–258.

Woelkerling, W., Spencer, K. G., & West, J. A. (1983). Studies on selected corallinaceae (Rhodophyta) and other algae in a defined marine culture medium. *Journal of Experimental Marine Biology and Ecology*, 67(1):61–77.

Woung Kim, K. & Bok Lee, S. (2003). Growth of the hyperthermophilic marine archaeon Aeropyrum pernix in a defined medium. *Journal of Bioscience and Bioengineering*, 95(6):618–622.

Wu, D., Daugherty, S. C., Van Aken, S. E., Pai, G. H., Watkins, K. L., Khouri, H., Tallon, L. J., Zaborsky, J. M., Dunbar, H. E., Tran, P. L., Moran, N. A., & Eisen, J. A. (2006). Metabolic Complementarity and Genomics of the Dual Bacterial Symbiosis of Sharpshooters. *PLoS Biology*, 4(6):e188.

Wu, M., Sun, L. V., Vamathevan, J., Riegler, M., Deboy, R., Brownlie, J. C., McGraw, E. A., Martin, W., Esser, C., Ahmadinejad, N., & et al (2004). Phylogenomics of the reproductive parasite Wolbachia pipientis wMel: A streamlined genome overrun by mobile genetic elements. *PLoS Biology*, 2(3):e69.

Xong, H. V., Vanhamme, L., Chamekh, M., Chimfwembe, C. E., Abbeele, J. V. D., Pays, A., Meirvenne, N. V., Hamers, R., Baetselier, P. D., Pays, E., & Gene, B.-R. S. (1998). A VSG expression site–associated gene confers resistance to human serum in Trypanosoma rhodesiense. *Cell*, 95:839–846.

Ylikahri, R. (1979). Metabolic and nutritional aspects of xylitol. *Advances in Food Research*, 25:159–180.

Yong, E. (2016). *I contain multitudes: The microbes within us and a grander view of life*. Random House.

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

Zhou, W., Rousset, F., & O'Neill, S. (1998). Phylogeny and PCR–based classification of Wolbachia strains using wsp gene sequences. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1395):509 – 515.

Zreik, L., Bove, J. M., & Garnier, M. (1998). Phylogenetic characterization of the bacterium-like organism associated with marginal chlorosis of strawberry and proposition of a Candidatus taxon for the organism, "Candidatus Phlomobacter fragariae". *International Journal of Systematic Bacteriology*, 48(1):257–261.