Facultad de Ciencias Biológicas

Institut Cavanilles de Biodiversitat i Biologia Evolutiva

Programa Oficial de Postgrado en Biotecnología

VNIVERSITAT DE VALÈNCIA

Evolution and Comparative Genomics of Di-symbiotic Systems in Aphids from the Lachninae Subfamily and Genome Reduction in *Serratia symbiotica*

Memoria presentada por Alejandro Manzano Marín para optar al grado de Doctor por la Universitat de València.

Directora:

Dra. Amparo Latorre Castillo, catedrática de genética de la Universitat de València.

Mayo 2016

AMPARO LATORRE CASTILLO, Catedrática del Departamento de Genética de la Universidad de Valencia,

CERTIFICA que el trabajo para optar al grado de Doctor en Biotecnología, y que lleva por título "Evolution and Comparative Genomics of Di-symbiotic Systems in Aphids from the Lachninae Subfamily and Genome Reduction in *Serratia symbiotica*", ha estado realizado bajo su dirección en el Instituto Cavanilles de Biodiversidad y Biología Evolutiva por ALEJANDRO MANZANO MARÍN.

Y para que así conste, en el cumplimiento de la legislación vigente, firmo el presente certificado en:

Valencia, a de de 2016.

Amparo Latorre Castillo

Table of Contents

Acknowledgements ix

General Introduction 1

I.1 About symbiosis 1
I.1.1 Definitions and terms 1
I.1.2 Symbiosis as an evolutionary driving force 2
I.1.3 Symbiosis in insects 3
I.1.4 Genome reduction in endosymbionts 5
I.2 The Aphids (Hemiptera: Aphididae) 8
I.2.1 Aphid biology 8
I.2.2 Evolutionary relationships of aphid subfamilies . . . 10
I.2.3 The Lachninae subfamily 12
I.3 Aphids' endosymbionts 14
I.3.1 *Buchnera aphidicola* 15
I.3.2 Secondary endosymbionts 17
I.3.3 *Serratia symbiotica* 19

Objectives 25

Materials and Methods 27

M.1 Aphid collection and DNA extraction 27
M.2 Genome assembly and annotation 28
M.2.1 Preassembly 28

M.2.2 *S. symbiotica* SCt-VLC 28

M.2.3 *B. aphidicola* strain BTs-Pazieg and *S. symbiotica* strain STs-Pazieg 30

M.2.4 Annotation of sequences and metabolic reconstruction 31

M.3 Fluorescence *in situ* hybridisation 32

M.4 16S rRNA gene PCR, cloning, and sequencing 33

M.5 Protein clustering and genome rearrangements 34

M.6 Phylogenetic analyses 34

Chapter 1: The "early" co-obligate *Serratia symbiotica* endosymbiont of the aphid *Cinara* (*Cupressobium*) *tujafilina* 37

1.1 Introduction 38

1.2 Results & Discussion 39

1.2.1 Localisation of *S. symbiotica* in bacteriomes of *C.* (*Cu.*) *tujafilina* 39

1.2.2 The *S. symbiotica* strain SCt-VLC genome 40

1.2.3 Metabolic capabilities 43

1.2.4 Genome rearrangements and mobile elements . . . 47

1.2.5 Decay of EAAs biosynthetic operons and genes . . 51

1.3 Conclusion 53

Chapter 2: Evolutionary convergence of co-obligate endosymbiotic consortia between the aphids *Tuberolachnus salignus* and *Cinara* (*Cinara*) *cedri* 57

2.1 Introduction 58

2.2 Results & Discussion 60
2.2.1 Genomic features of *Buchnera* and *S. symbiotica* from *Tu. salignus* 60
2.2.2 Evolutionary Relationships and Genome reduction in *Buchnera* 61
2.2.3 Evolutionary Relationships and Genome Reduction in *S. symbiotica* 65
2.2.4 Metabolic capabilities of BTs and STs 71
2.2.5 Metabolic and transporter-function convergence between *Tu. salignus* and *C.* (*Ci.*) *cedri*'s co-obligate endosymbiotic systems 76
2.3 Conclusion 78

Chapter 3: Identity and localisation of secondary endosymbionts in aphids from the Lachninae subfamily 81
3.1 Introduction 82
3.2 Results and Discussion 86
3.2.1 Phylogenetic analyses and fluorescence *in situ* hybridisation of Lachninae aphids' secondary endosymbionts 86
3.2.2 Evolutionary succession of secondary endosymbionts within the Lachninae 94
3.3 Conclusion 100

Chapter 4: Genome reduction in *Serratia symbiotica* 103
4.1 Introduction 104

4.2 Results and Discussion 106
4.2.1 *S. symbiotica* strains and their shrinking genomes . 106
4.2.2 Erosion of essential amino acid biosynthetic routes . 109
4.2.3 Decay of RNA features and the loss of regulation . . 111
4.2.4 Informational machinery 113
4.2.5 Dwindling genes 117
4.3 Conclusion 120

General conclusions **121**

Resumen en Español **123**

Bibliography **135**

Appendix **167**
S.1 Additional figures 167
S.2 Additional tables 179

Abbreviations **181**

Acknowledgements

Getting to this point has not been easy (at all), and of course there has been many people without whom I would not have been able to make it. First of all, I have to thank my parents, (María Luisa Marín Gutiérrez and Alejandro Manzano Ramírez). Being themselves scientists, I will always remember the scientific discussions about my dad's work-relates issues at supper time. Also, my parents always encouraged me to watch The Discovery Channel, which was still about science dissemination back when I was a kid (not any more). Both of these things, I think, had a lot to do with me choosing to follow a scientific career. Next, I have to thank many of my science teachers, but not all. Especially, I have to thank my computer science teacher at St. Theresa's Secondary School, Mr. X (I can't for the life of me remember his last name). Through him, I had my first approach to programming, and my growing interest in it had a lot to do with me choosing to study Genomic Sciences. In University I ran into many great teachers, which continued to spark my interest in genomics and evolution. Among them, I particularly remember Julio Collado Vides, Guillermo Dávila, Miguel Ángel Ramírez, and Julio Augusto Freyre. I also have to thank my supervisor Amparo Latorre Castillo, for blindly giving me the opportunity to come Join her lab during my last year at University, as well as for providing me with the opportunity to develop my Ph.D. thesis with a Marie Curie Fellowship. From my current lab, I have two very important people to thank. Firstly, Rafael Patiño Navarrete, for I owe him my first experiences in the lab holding a pipette and performing PCRs. Secondly, I owe a great bunch to Diego Santos García. His expertise in wet-lab techniques greatly complemented my bioinformatic skills to go beyond the experiments I could think of (a part of this thesis was possible thanks to him). Regarding other labs, I must deeply thank Jean-Christophe Simon and his team at the INRA (for aphid collection and support with with a couple of works) and the team at DOME (Matthias

Horn, Gitta Szabo, and a ton of other people there that helped me out with the microscope, laboratory, etc...). On the non-lab-related side, there has also been some people that helped me make both my stay in València enjoyable: Segio López Madrigal (he taught me some Catalon... sorry Valencian), Fina and Jorge (from the New Orleans/Route 66 bar/second home), the physicists (including Eduardo, Luis David, Lotfi, Marco, etc... [great parties]), and Manuel Serra (responsible for making me a mountain-addict), and Stephan Köstlicher from Vienna for making my stay there very enjoyable. I want to particularly thank Jorge Mariano Collantes Alegre for his very good cartoons of aphids, which undoubtedly made this work just a little bit cooler. Last but not least, I want to thank my current couple Anouk Willemsen for her great support during my Ph.D. time. She was always there to criticise and help me rehearse my presentations, review my articles, and provide me with support during stressful times.

Thank you all.

"Throughout the world, numerous organisms are infected with a myriad of microorganisms, potentially leading to the establishment of obligate and mutualistic symbiotic associations. It is plausible that such symbiotic associations are coming into being and evolving at any time, everywhere." (Hosokawa *et al.*, 2016)

General Introduction

I.1 About symbiosis

I.1.1 Definitions and terms

The term "symbiosis" (from Greek σύν "together" and βίωσις "living"), in its biological context, was coined by Heinrich Anton de Bary in his work entitled "Erscheinung der Symbiose". He defined "symbiosis" as the "phenomena of cohabitation of unlike organisms", without restricting it to the effects of the interaction in either one of the players (de Bary, 1879). Although more restrictive uses of the term have been used (reviewed in Saffo, 1992), de Bary's definition holds as the most inclusive, considering the whole spectrum from neutral, to harmful partnerships. Within this spectrum, Martin and Schwab (2012) have recognised 6 broad categories of symbiotic interactions, based on the possible combinations of neutral (0), beneficial(+), and harmful(-) effects between two partners (Figure I.1). In the current work, we will mainly focus on one type of interaction: mutualism. This term is reserved for the symbiotic relationships where each partner benefits from the activity of the other. The term host is used to refer to the biggest partner, while the smaller one is called the symbiont.

Following Martin and Schwab (2012), symbiotic interactions can be further categorised based on the localisation of the symbiont. When the symbiont is located on the host's external surface, it is considered to be an ectosymbiont. On the contrary, if the symbiont is located inside the host's cells or tissues, the symbiont is considered to be an endosymbiont. At the same time, endosymbionts can be both intra- and extracellular, depending

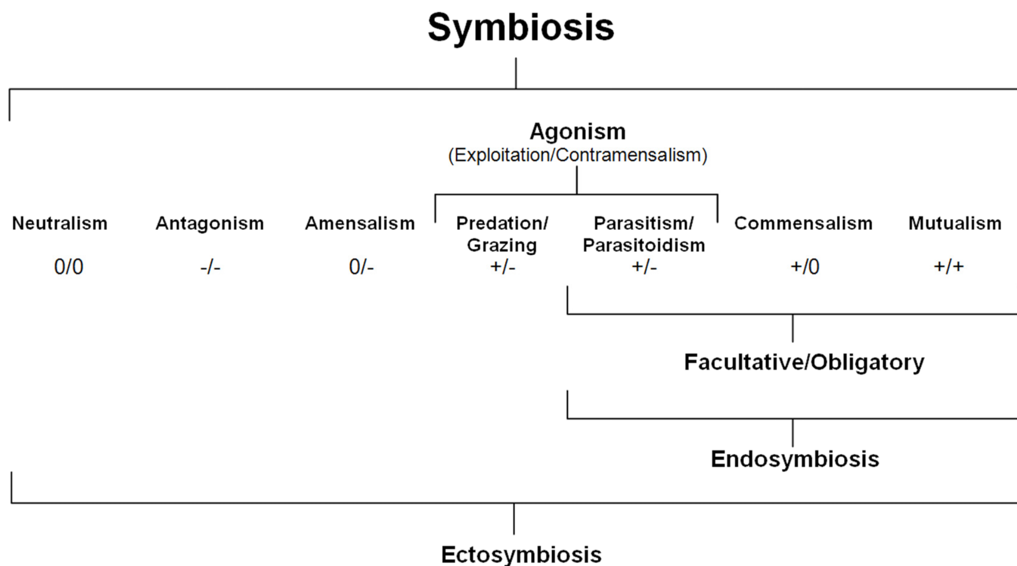


Figure I.1 Simplified and inclusive scheme of symbiosis. For any particular pair of interacting species: += beneficial effect, -= harmful effect, and 0= neutral effect. This scheme recognizes that endosymbiosis can be subdivided into intracellular/extracellular. Reproduced from Martin and Schwab (2012).

on the specific localisation within the host. Extracellular symbionts are those which are located within host's cavities and between cells in the host tissue (intercellular), while intracellular ones are those which are located inside the host's cells (within membrane-derived vacuoles or free in the cytoplasm).

Lastly, symbiotic relations can also be categorised into obligate or facultative, based on the degree of dependency of both partners. Contrary to obligate symbiotic interactions, facultative ones refer to those where the direct interaction is not required for the correct development, reproduction, and survival of one or both of the partners.

I.1.2 Symbiosis as an evolutionary driving force

In 1967, the American Evolutionary biologist, Lynn Margulis, published her highly influential paper *On the Origin of Mitotic Cells*. In this work,

she presented a theory of the origin of the discontinuity between eukaryotic and prokaryotic cells (Sagan, 1967). More specifically, she referred to the origin of "the mitochondria, the (9 + 2) basal bodies of the flagella, and the photosynthetic plastids" as having been derived from free-living cells. Although these ideas were not new, she synthesised them in a way that was consistent with the scientific knowledge of the time, thus advancing, substantiating, and invigorating the theory of symbiogenesis, first put forward by Mereschkowsky (1910). Although her theory of the spirochaetal origin of the flagella and cillia never received much support, the theory of the free-living prokaryotic origin of the mitochondria (from an aerobic bacterium) and chloroplast (from Cyanobacteria Mereschkowsky, 1905) was proven by phylogenetic evidence (Schwartz and Dayhoff, 1978).

Beyond the origin of evolutionary leaps such as the eukaryotic cell, symbiotic interactions between prokaryotes and eukaryotes are both widespread and diverse (Moya *et al.*, 2008). Eukaryotes, contrary to most prokaryotes, hold a rather limited metabolic repertoire, hence, symbiosis between these two types of organims has helped eukaryotes to widen their range of metabolic resources. Examples of this include dinitrogen fixation (Kneip *et al.*, 2007), methanogenesis (Schink, 1997), nitrogen and sulphur assimilation (Nakagawa *et al.*, 2014), and essential-nutrient anabolism (Douglas, 1998; Hansen and Moran, 2014). This expanded set of metabolic capabilities could have allowed the colonisation of new ecological niches.

I.1.3 Symbiosis in insects

Symbiotic associations between insects (class Insecta Linnaeus, 1758) and microorganisms are both diverse and widespread. The reason behind this can be attributed to the nutrient-deficient diet some insects have, such as plant phloem (deficient in essential amino acids [**EAAs**]

and some B-vitamins) (Ziegler, 1975) and mammalian blood (deficient in B-vitamins) (Lehane, 2005). The first big compendium describing these type of associations was written by Paul Buchner in 1953. He explored the endosymbionts present in many hemipteran families within the Coccoidea, Aphidoidea, Aleyrodoidea, Psylloidea, and Membracoidea superfamilies. In many of these, Buchner observed consistent infections across various individuals, as well as detected many of these microorganisms being hosted inside specific cells (bacteriocytes) within a specific organ-like structure, termed the bacteriome (called a mycetome in his book). Early metabolic and sequencing studies of some of these endosymbionts, confirmed that these indeed had the capacity to synthesise essential compounds for their insect hosts, and even developed metabolic interdependences with these (reviewed in Zientz *et al.*, 2004). Additionally, experimental studies directed towards the elimination of these obligate endosymbionts in the pea aphid *Acyrthosiphon pisum* (Aphidinae: Macrosiphini) (Koga *et al.*, 2003; Ohtaka and Ishikawa, 1991), the tsetse fly *Glossina morsitans* (Diptera: Glossinidae) (Nogge, 1976), and different cockroach families (Blattodea) (Sacchi *et al.*, 1993), have shown that aposymbiotic females generally display a lack of reproduction and decrease in fertility or complete sterility, corroborating the obligate nature of these associations.

Apart form bacteriocte-associated endosymbionts, insects have also been found to harbour symbionts in their digestive tract (reviewed in Engel and Moran, 2013). Although most insect guts harbour relatively simple microbiotas compared to mammalian guts, some insects like cockroaches or termites house large communities of specialised bacteria that seem to be host-specific (regardless of diet or lifestyle) (Sabree and Moran, 2014). Stink bugs (Hemiptera: Pentatomoidea) from the Plataspidae and Acanthosomatidae families posses midgut gammaproteobacterial extracellular symbionts that are vertically transmitted and that have been found to be of obligate nature, causing developmental, growth, and colouration defects when eliminated

(Hosokawa *et al.*, 2006; Kikuchi *et al.*, 2009).

As mentioned above, the microorganisms associated to insects are quite diverse. Insects have been found to harbour obligate vertically-transmitted eukaryotic (Cheng and Hou, 2001; Fukatsu *et al.*, 1994) and prokaryotic symbionts (commented in [Fig. 1] Moran *et al.*, 2008). These last did not originated from a single infection event, but rather have different origins, some belonging to different branches within the Bacteroidetes, Betaproteobacteria, and Gammaproteobacteria. Although the exact origin of many of these symbiotic bacterial strains is not yet resolved, recent investigations into Japanese populations of the stink bug *Plautia stalii* (Hemiptera: Pentatomidae) has revealed that this insect species keeps associations with different obligate bacterial symbionts (Hosokawa *et al.*, 2016). Also, it is capable of acquiring, and even replacing, not-yet-cultured reduced bacterial symbionts with cultivable strains of symbionts, pointing towards the environmental acquisition of these. This study implies that obligate and mutualistic interactions could potentially be "evolving at any time, anywhere" from microorganisms that are infecting numerous individuals throughout the world.

I.1.4 Genome reduction in endosymbionts

One common feature from vertically transmitted obligate endosymbionts from insects is a tendency to evolve reduced genomes (reviewed in McCutcheon and Moran, 2012; Moran *et al.*, 2008). This change is triggered by two main evolutionary mechanisms: the relaxation of natural selection and the continuous bottlenecks triggered by vertical transmission. The former results from a combination of the symbiont now residing in a more stable environment (inside the host), rendering certain "free-living" functions unnecessary and generating genetic/metabolic redundancy between the host and the symbiont (and/or another

symbiont), which together promote substantial gene-loss (reviewed in Moran *et al.*, 2008; Moya *et al.*, 2008). The latter is a result of the transmission mode of endosymbionts, where only a small subpopulation will be inherited to the next generation, thus increasing the effect of genetic drift and the associated Muller's ratchet (Moran, 1996). These mechanisms are further accentuated by the loss of DNA recombination, repair, and uptake mechanisms observed in currently sequenced bacterial symbionts (reviewed in Moran and Wernegreen, 2000; Moran *et al.*, 2008). In this vein, mitochondria and plastids have just been "the luckiest of a longstanding series of doomed endosymbionts who were saved by transfer of genes to the nucleus" (Keeling *et al.*, 2015).

Based on available genomic data from different host-associated bacteria, Toft and Andersson (2010) have divided the general genome reduction process of host-associated bacteria (namely intracellular) into five stages of host adaptation: from being free-living (**first** stage) to becoming an organelle (**fifth** stage) (Figure I.2). In the **second** stage, a free-living bacterium has effectively become a facultative intracellular (still being able to grow on artificial medium by standard techniques), such as *Legionella* spp. (Pine *et al.*, 1979) and *Bartonella* spp. (Wong *et al.*, 1995). These symbionts undergo drastic changes in outer-surface structural genes, horizontal gene acquisition, and rapid modification of these exogenous genes due to duplication and recombination events. In the **third** step, the facultative intracellular symbiont becomes an obligate one, loosing its ability to grow on artificial medium (by standard techniques). This is exemplified by *Rickettsia* spp. (Winkler, 1990) and *Coxiella* spp. (Baca and Paretsky, 1983), which typically display extensive pseudogenisation and gene-loss, along with a higher host specialisation. Once an intracellular habitat is secured and the interaction becomes stronger, genetic and metabolic losses or modifications can enable the transition to an obligate intracellular mutualist, the **fourth** stage. On this stage, the ontogeny of a host can undergo reprogramming to produce specialised cells (bacteriocytes) to

house these symbionts. Examples of bacteria in this stage include *Buchnera* spp. (van Ham *et al.*, 2003) and *Wiggelsworthia* spp. (Rio *et al.*, 2012). These late stage symbionts have lost most of their genetic repertoire, non-coding DNA, and mobile elements (**MEs**), having developed small and highly compact genomes (reviewed in Wernegreen, 2002). Lately, a variety extremely reduced genomes have been sequenced, which start to defy the limits of bacterial genome reduction (reviewed in McCutcheon, 2010; McCutcheon and Moran, 2012; Moran and Bennett, 2014). Nonetheless, this scheme of genome reduction could also prove true for obligate extracellular vertically-transmitted symbionts, such as has been exemplified by *Candidatus* Ishikawaella capsulata (Gammaproteobacteria), obligate extracellular endosymbionts from the Plataspidae family of stink bugs (Hosokawa *et al.*, 2006; Nikoh *et al.*, 2011).

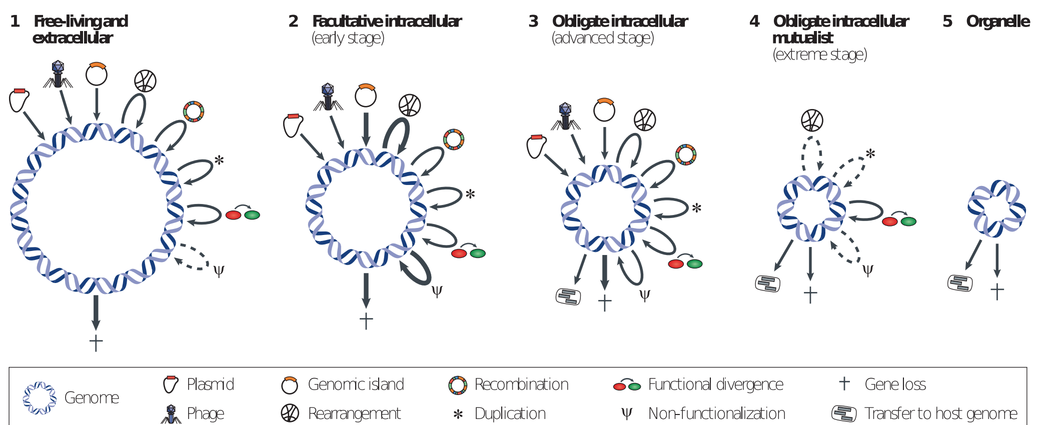


Figure I.2 Stages of host adaptation. The genome dynamics for different host-adaptation stages: free-living (1), facultative intracellular (2), obligate intracellular (3), obligate intracellular mutualist (4) and organelle (5). Arrows that point directly to the genomes indicate the acquisition of genes by horizontal gene transfer. Arrows that loop back to the genome indicate changes within the genome. Arrows that point away from the genome indicate gene loss or gene transfer to the host genome. The relative influence of each of these types of events at the different intracellular stages is shown by the weight of the arrow. Reprinted by permission from Macmillan Publishers Ltd: *Nature reviews. Genetics*. Toft and Andersson, copyright (2010).

Another general feature from vertically transmitted obligate

endosymbionts is the bias towards an Adenine- and Thymine-rich genome, with the notable exceptions of *Candidatus* Hodgkinia cicadicola (McCutcheon *et al.*, 2009) and *Candidatus* Tremblaya spp. (Husnik *et al.*, 2013; López-Madrigal *et al.*, 2011; McCutcheon and von Dohlen, 2011). Non-exclusive explanations for the basis of the "AT-rich genome" phenomenon have been put forward: **i)** There is a higher energetic cost in the biosynthesis of guanosine triphosphate (**GTP**) and cytosine triphosphate (**CTP**) compared to the one of adenosine triphosphate (**ATP**) and thymidine triphosphate (**TTP**) (Rocha and Danchin, 2002), **ii)** The loss of DNA replication and repair genes can cause an increase in frequency of cytosine deaminations, leading to **C** to **T** transitions in the leading strand (Klasson, 2006).

I.2 The Aphids (Hemiptera: Aphididae)

I.2.1 Aphid biology

Aphids (Hemiptera: Aphidoidea: Aphididae) are a highly diverse group of insects, with over 5,000 extant species (Favret, 2016). The Aphididae belong to the Aphidomorpha infraorder, which also includes the Adelgidae (Hemiptera: Adelgoidea) and Phylloxeridae (Hemiptera: Phylloxeroidea) families. The common ancestor of these probably lived some 280 million years ago (**Mya**) (Dixon, 1985; Heie, 1967), in the early Permian period of the Paleozoic era. The earliest fossil assigned to date to the Aphidoidea superfamily belongs to the extinct species †*Grimmenaphis magnifica* (†Oviparosiphidae), and dates back to 182.0-183.0 Mya in the Jurassic period of the Mesozoic era (Ansorge, 1996). It is hypothesised that the Aphidoidea underwent a rapid radiation during the Cretaceous period, having fossils of several families dating to this, followed by a massive extinction by the end of this period (reviewed in von Dohlen and Moran, 2000). However, fossils of virtually all extant

tribes are found in Baltic amber, dating back to the Paleogene period (35-45 Mya from the Eocene epoch) of the Cenozoic era.

Aphids feed on the phloem sap of plants, which is generally poor in EAAs (Douglas, 2006; Sandstrom and Moran, 1999) and some B-vitamins, namely riboflavin (vitamin B_2), biotin (vitamin B_7), folic acid (vitamin B_9), cobalamin (vitamin B_{12}) (Ziegler, 1975). Aphid life cycles can be quite complicated, and involve a series of morphologically distinct morphs. Although the various subfamilies of Aphididae have life cycles with characteristic features, most aphids present an holocylic life cycle (Figure I.3). This consists of one generation of sexual morphs (sexuales), which give rise to overwintering eggs, and several generations in which only parthenogenetic females are produced (see Blackman and Eastop, 1994). Briefly, the overwintering eggs resulting from sexual reproduction will hatch in spring, giving rise to viviparous females that will reproduce by parthenogenesis during spring and summer. Then, with the dropping temperatures of fall, sexuales will appear, producing the overwintering eggs that will start the cycle over again.

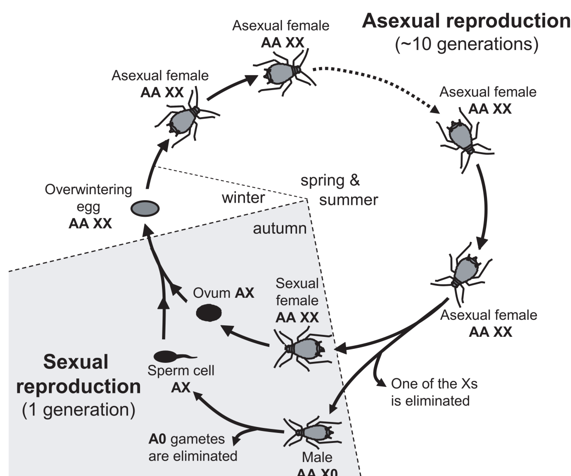


Figure I.3 Scheme of an holocyclic life cycle. Holocyclic life cycle of the aphid *Ac. pisum*. Reproduced from Jaquiéry *et al.* (2013).

Also, there are other more complex life cycles that involve host alternation, where first, the sexuales mate and lay fertilised eggs on a tree or shrub (primary host). This is followed by a regular migration (at some stage of the life cycle or another) of the aphids originating from the

hatched eggs to a totally unrelated herbaceous or woody plant (secondary host). In this, the partenogenetic generations (exules) occur. Then, before the next sexual generations, the aphids migrate back to the primary host.

I.2.2 Evolutionary relationships of aphid subfamilies

Aphids are organised into subfamilies, tribes, and subtribes. According to currently accepted taxonomy, 24 extant subfamilies are recognised (Favret, 2016). Four major hypothesis for aphid phylogeny have been put forward. The first three studies are summarised in (Figure S.1). Using molecular data from the *Buchnera* endosymbiont, Nováková *et al.* (2013) have performed the most comprehensive molecular analysis to date on the phylogenetic relationships of aphids (Figure I.4). This reconstruction contrasts previous hypothesis in some major aspects. Firstly, It confidently recovers the Pemphigini (Eriosomatinae) as sister taxa to the rest of aphid subfamilies, as well as a monophyletic clade including the Aphidinae and Calaphidinae. Also, it confidently recovers the Lachninae within an unresolved clade sister to the Fordini+Eriosomatini.

Some authors, such as Heie (1980), have raised most subfamilies to family level, but in the current work we will follow the subfamily classification. Also, we will base all following evolutionary assumptions based on the phylogenetic reconstruction of Nováková *et al.* (2013).

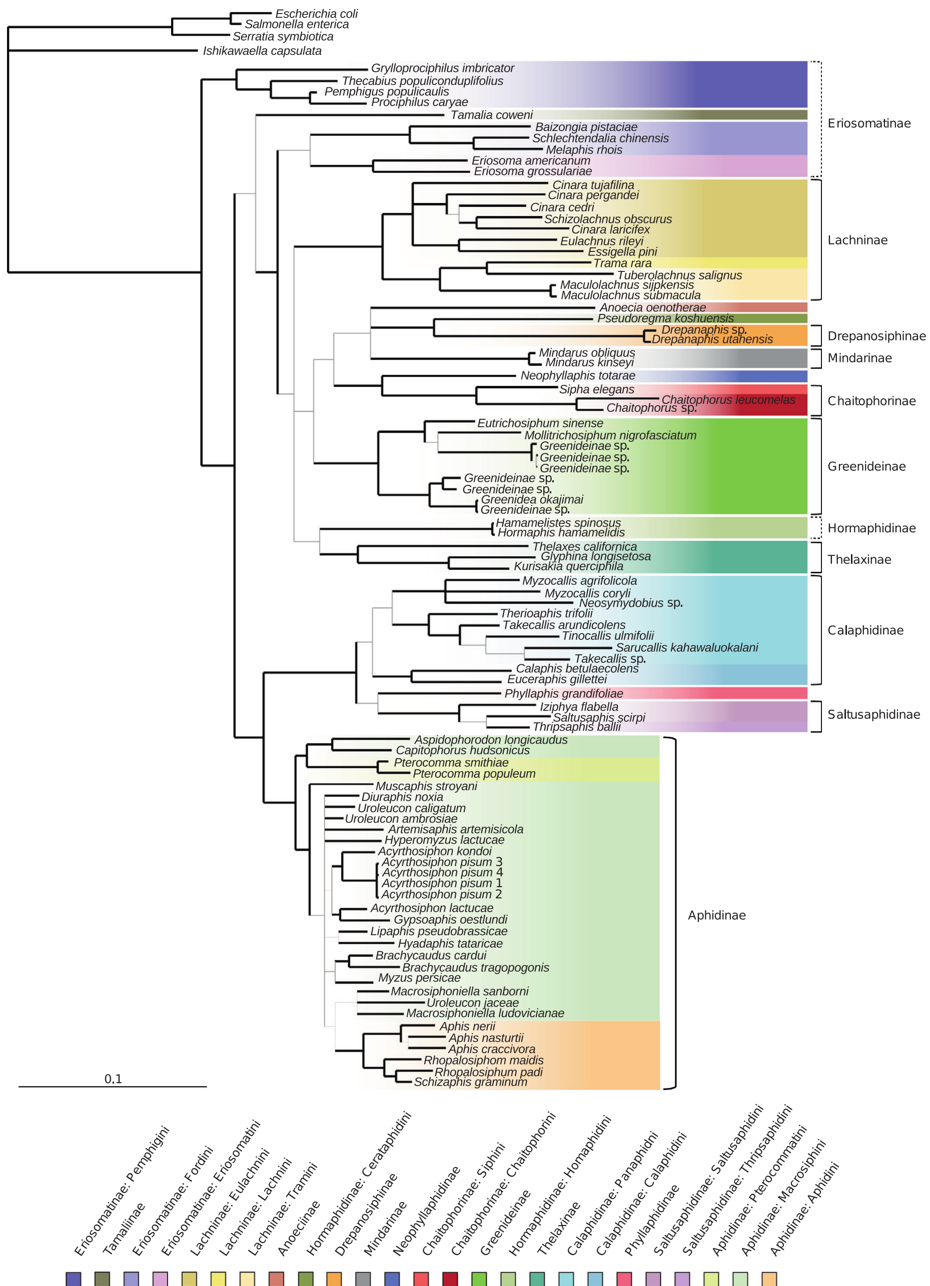


Figure I.4 Bayesian inference topology of aphids based on *Buchnera*-derived sequences. Bayesian inference topology based on the concatenated matrix including 1st and 2nd codon positions of *Buchnera trpB*, *dnaB*, *groEL*, and *ilvD* genes. Thicker lines designate branches with a posterior probability above 95%. Solid vertical lines refer to monophyletic subfamilies. Dashed vertical lines designate paraphyletic/polyphyletic taxa. Reprinted from Nováková *et al.* (2013), with permission from Elsevier.

I.2.3 The Lachninae subfamily

The Lachninae Herrich-Schaeffer, 1854 subfamily of aphids currently groups 403 extant species (Favret, 2016). These are mostly associated only with trees and none host-alternate (commented in Heie, 2015). Lachnind fossils are few, dating to the Neogene period (Miocene epoch) from the Cenozoic era, and most belong to extant genera (*Cinara*, *Stomaphis*, and *Longistigma*) (Heie and Wegierek, 2011). The oldest Lachninae fossil belongs to †*Lachnus pectorosus*, dating back to 15.97-20.43 Mya.

Species within this subfamily have been classically organised into three tribes: the **Eulachnini** Baker, 1920; the **Lachnini** Herrich-Schaeffer, 1854; and the **Tramini** Herrich-Schaeffer, 1854. In the first big effort to build a molecular phylogeny of the Lachninae, Normark (2000) already noticed that the Lachnini were recovered as paraphyletic, showing the Tramini as sister clade to *Tuberolachnus*+*Nippolachnus* genera (both belonging to the then proposed Tuberolachnini tribe). This Tramini+Tuberolachnini clade was recovered as sister to *Stomaphis* spp. (classified into the Stomaphidini). Recent and extensive molecular work on the Lachninae calls for the division of this subfamily into five strongly supported monophyletic tribes, adding the **Stomaphidini** Mordvilko, 1914 and **Tuberolachnini** Mordvilko, 1942 (Chen *et al.*, 2015a) (Figure I.5). In the current work, we will follow the latter five-tribe classification.

The **Eulachnini**, which feed exclusively on conifers, are classified into four genera, ***Cinara*** (the largest within the Lachninae, comprising 253 species), ***Essigella***, ***Eulachnus***, and ***Pseudessigella***. It is important to note that, based on molecular data, *Schizolachnus* spp. have been repeatedly found nested within the *Cinara* (Chen *et al.*, 2015a; Meseguer *et al.*, 2015; Nováková *et al.*, 2013). Therefore, *Schizolachnus* spp. have recently been transferred to the *Cinara* (*Schizolachnus*) subgenus (Chen *et al.*, 2015a). The *Cinara* genus has been traditionally taxonomically

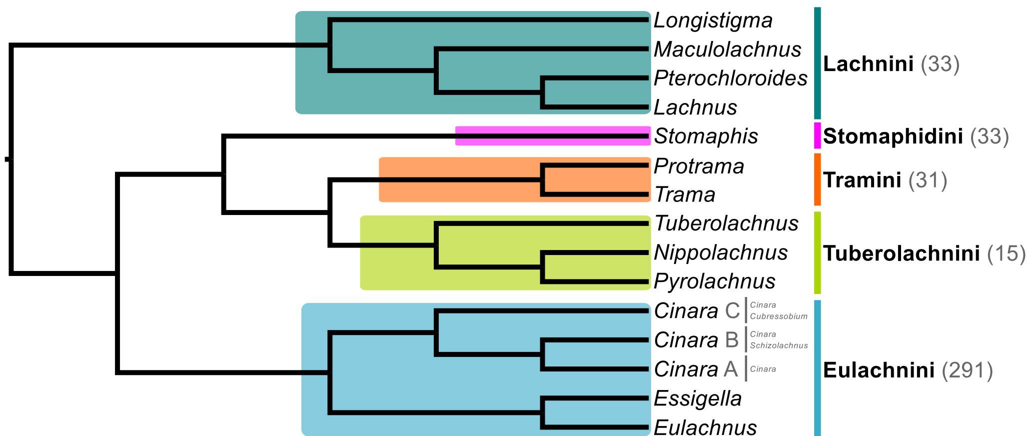


Figure I.5 Dendogram displaying the current molecular phylogeny of the Lachninae. Phylogenetic relationships of the five strongly supported monophyletic Lachninae tribes based on Chen *et al.* (2015a). *Cinara* groups A, B, and C are based on the strongly supported clades recovered by Meseguer *et al.* (2015). In parenthesis, the number of currently valid species is shown, including the missing genera *Eotrama* (Tramini), *Neonippolachnus* (Tuberolachnini), *Sinolachnus* (Tuberolachnini), and *Pseudessigella* (Eulachnini), for which no molecular data is available. Modified from Manzano-Marín *et al.* (2016).

classified into two main subgenera, *Cinara* (*Cinara*) and *Cinara* (*Cupressobium*). The third subgenus, *Cinara* (*Cedrobium*), has not been molecularly analysed, and thus, the phylogenetic placement of its only species (*Cinara* (*Cedrobium*) *laportei*) remains uncertain. Recent extensive molecular work in *Cinara* divides the genus into three major phylogenetic clades, termed simply **A**, **B**, **C** (Meseguer *et al.*, 2015). **Clade C** includes some members of the *Cinara* (*Cinara*) subgenus and all *Cinara* (*Cupressobium*) species, while **clade B** groups some *Cinara* (*Cinara*) species and all *Cinara* (*Schizolachnus*). Finally, **clade A** solely include members of the *Cinara* (*Cinara*) subgenus. The rest of the subfamilies are organised as follows: the **Lachnini** groups four genera, ***Lachnus***, ***Longistigma***, ***Maculolachnus*** and ***Pterochloroides***; the **Tuberolachnini** groups five, ***Neonippolachnus***, ***Nippolachnus***, ***Pyrolachnus***, ***Sinolachnus***, and ***Tuberolachnus***); the **Tramini** groups three, ***Eotrama***, ***Protrama***, and ***Trama***; and the **Stomaphidini** groups

only one, ***Stomaphis***. Even though the latter four tribes are composed of aphid genera that feed on deciduous trees (that lose their leaves seasonally), mainly angiosperms, the Tramini are unique in that they solely feed on roots of herbaceous plants, mostly composites. Also, they are mainly anholocyclic, being largely asexual (Blackman *et al.*, 2001). Finally, the Stomaphidini, that feed both on angiosperm and gymnosperm trees (bark-trunk and root), have developed a particularly large mouth part to accommodate bark-feeding (Blackman and Eastop, 1994).

I.3 Aphids' endosymbionts

As mentioned before, aphids' diet is deficient in EAAs and some B-vitamins (see section I.2.1: Aphid biology). Regarding the former, studies on the aphid *Myzus persicae* (Aphidinae: Macrosiphini) reared on an artificial diet lacking amino acids (Mittler and Dadd, 1963) have been performed Dadd and Mittler (1965). In this work, it was revealed that larvae could survive for considerable periods, and adult survival was hardly affected by the complete omission of amino acids from the diet. However, growth of larvae, and the numbers of larvae produced by adults, depended upon dietary supplies of amino acids. Concerning the need of B-vitamins, in the aphid *My. persicae*, a diet lacking biotin did not prevented larvae from developing into adults, but did adversely affected growth (Dadd *et al.*, 1967); and in *Neomyzus circumflexus* (Aphidinae: Macrosiphini), the author claimed to have been able to rear aphids "without influencing normal development" for at least five generations on a diet where both biotin and riboflavin were omitted (Ehrhardt, 1968). These works pointed towards these aphids not having a need for these essential compounds, which are indeed lacking from their diet. Viewed in the light of the work by Buchner (1953), these results suggested a putative role of the endosymbiotic bacteria found in the bacteriomes of aphids as nutrient suppliers.

I.3.1 *Buchnera aphidicola*

Advanced microscopic studies into the obligate endosymbiotic bacteria of the aphid *Ac. pisum*, confirmed this displayed the typical two membranes of the Gram-negatives (Griffiths and Beck, 1975) (reviewed in Houk and Griffiths, 1980) (Figure I.6A). Also, these studies revealed the existence of a third host-derived membrane (the **symbiosomal membrane**). Evidence for this being host-derived, comes from observations that show this membrane is lost during the infection process and then a new one is formed after entry into the embryonic bacteriocytes. Besides, two or more symbiont cells are sometimes found to be enveloped by the same symbiosomal membrane.

Molecular studies of the primary and secondary bacterial associates of *Ac. pisum* immediately showed that, while the secondary symbiont was found nested within free-living bacteria, the primary one consitituted a distinct lineage within the Gammaproteobacteria (Figure I.6B)

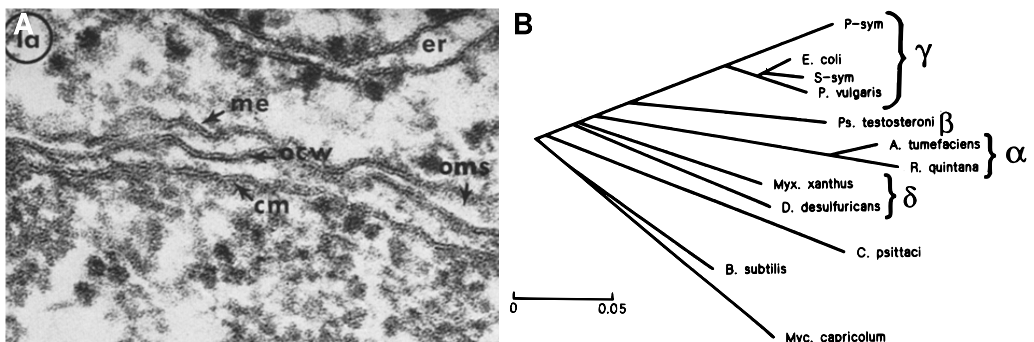


Figure I.6 Microscopy of primary endosymbiont of *Ac. pisum* and phylogenetic positioning of both primary and secondary endosymbionts. **A)** High magnification micrograph of primary symbiont membranes, cm = cytoplasmic membrane; er = endoplasmic reticulum; me = membrane envelope; ocw = outer cell wall layer; oms = outer membrane space. x158000. **B**). Phylogenetic tree illustrating the relationships of the primary and secondary endosymbionts to selected species of eubacteria. Scale represents estimated number of nucleotide substitutions per sequence position. α, β, γ, δ, subdivisions within the class Proteobacteria. Reproduced from Griffiths and Beck (1975); Unterman *et al.* (1989).

(Unterman *et al.*, 1989). Subsequent molecular studies confirmed that primary endosymbionts from aphids, belonging to four different subfamilies, formed a monophyletic clade, and that within this clade, endosymbionts paralleled the phylogenetic relationships of the aphid hosts (Munson *et al.*, 1991b). Based on this evidence, the genus *Buchnera* and species *Buchnera aphidicola* were proposed for the lineage of bacterial primary endosymbionts from aphids, with the type strain being that of the aphid *Schizaphis graminum* (Aphidinae: Aphidini) (Munson *et al.*, 1991a). Early experimental studies based on both antibiotic treatment and artificial diets lacking amino acids, showed an increased time to reach adulthood and a decrease in weight of the adult and in the relative growth rate of the aphid (reviewed in Douglas, 1989; Ishikawa, 1989). The first direct experimental study corroborating the role of *Buchnera* as a provider of an essential amino acid was put forward by Douglas (1990). In her study, she proved that isolated preparations of *Buchnera* can utilize sulphate as a sulphur source for methionine biosynthesis synthesis (quantified by the incorporation of ^{35}S inorganic sulphate).

The determination of the first *Buchnera* genome (Shigenobu *et al.*, 2000) along with that of its aphid host, *Ac. pisum* (The International pea aphid genomics consortium, 2010), revealed that not only *Buchnera* had indeed evolved a genome dedicated mainly to the production of EAAs, but also that it had established a metabolic complementation with its aphid host for the production of these (Shigenobu and Wilson, 2011). This was also corroborated by transcriptome analysis of aphid bacteriocytes, where it was found that genes involved in synthesising both non-essential amino acids (**NEAAs**) and the last steps of synthesis of many EAAs (missing from the *Buchnera* genome) were overexpressed (Hansen and Moran, 2011). Further sequencing of different *Buchnera* strains revealed that these preserved genome-wide synteny to a high degree, pointing towards a rapid genome reduction and loss of MEs before the diversification of extant aphids (Tamas *et al.*, 2002; van Ham *et al.*, 2003). Currently, nine

full genomes of *Buchnera* from different aphid species are available in the databases.

I.3.2 Secondary endosymbionts

Apart from *Buchnera*, many aphids have found to harbour additional endosymbiotic bacteria (**secondary endosymbionts**). These symbionts can be of obligate or facultative nature, relative to their host's dependence on them. Contrary to obligate symbionts, facultative ones are not required for the correct development, reproduction, and survival of their host. Still, they can provide a benefit under certain environmental or ecological conditions (conditional mutualism). The role of these secondary endosymbionts may not only be limited to a nutritional one, as is usually the case for currently described primary ones, but it can range from acting as defensive symbionts against parasitoid wasps (*Serratia symbiotica*, *Hamiltonella defensa*, and *Regiella insecticola*) (Hansen *et al.*, 2012; Oliver *et al.*, 2003, 2005), fungal parasites (*Re. insecticola*, *Rickettsia* sp., *Rickettsiella* sp., and *Spiroplasma* sp.) (Scarborough, 2005; Łukasik *et al.*, 2013), relating to plant utilization (*Re. insecticola*) (Ferrari *et al.*, 2007; Leonardo and Muiru, 2003; Tsuchida *et al.*, 2004), and to resistance after heat stress in the form of reproductive advantage (*Rickettsia* sp. and *S. symbiotica*) (Chen *et al.*, 2000; Montllor *et al.*, 2002) (reviewed in Oliver *et al.*, 2010, 2014, see also Henry *et al.*, 2013).

Most secondary endosymbonts identified and described to date have been found first in the model aphid *Ac. pisum*. Hitherto, eight different genera of these bacterial associates have been reported in the pea aphid, these being: *Re. insecticola* (also known as [**aka**] **U-type** and **PAUS**) (Moran *et al.*, 2005; Tsuchida *et al.*, 2005), *H. defensa* (aka **T-type** and **PABS**) (Moran *et al.*, 2005; Sandström *et al.*, 2001), *Rickettsia* sp. (aka **PAR**) (Chen *et al.*, 1996; Sakurai *et al.*, 2005), *Rickettsiella viridis* (Tsuchida *et al.*, 2010, 2014), *Wolbachia* sp.

(Gomez-Valero *et al.*, 2004) (for a large-scale screening see Augustinos *et al.*, 2011), the so-called "**X-type**" or **PAXS** (Guay *et al.*, 2009), *Spiroplasma* sp. (Fukatsu *et al.*, 2001) (lacking morphological description and localisation of the symbiont), and *S. symbiotica* (aka **R-type** and **PASS**) (Fukatsu *et al.*, 2000; Moran *et al.*, 2005). Contrary to *Buchnera*, these putatively facultative endosymbionts generally present a bacillar or filamentous morphology and are not confined to bacteriocytes, but rather can be co-inhabiting *Buchnera*-infected bacteriocytes, in different ones (**secondary bacteriocytes**), in cells located at the periphery of the bacteriome and found closely associated to bacteriocytes (**sheath cells**), and in the haemocoel (Figure S.2).

Most of the early studies and first genomic sequences belonging to these secondary associates came from facultative endosymbiotic bacteria from the pea aphid. The genomes of both *H. defensa* strain 5AT (Degnan *et al.*, 2009b) and *Re. insecticola* strain LSR1 (Degnan *et al.*, 2009a) revealed that, contrary to the highly reduced genomes of nutritional obligate endosymbionts from insects (Gil *et al.*, 2003), they presented moderately A+T biased and "mildly" reduced genomes (*circa* half the size of closest free-living relatives) rich in mobile DNA, mainly insertion-sequence elements and phage-derived genes. These two bacteria were found to form a monophyletic clade sister to *Yersinia* (Husník *et al.*, 2011), however they do not present genome-wide synteny, which points towards them having diverged early in their establishment as aphid endosymbionts. Further sequencing of different facultative endosymbiotic bacteria from insects, revealed that these genomic features seemed to be quite common among them (Burke and Moran, 2011; Degnan *et al.*, 2011; Penz *et al.*, 2012; Toh *et al.*, 2005; Wu *et al.*, 2004). Interestingly, these features have also been found to be shared with "recent" obligate mutualistic endosymbiotic bacteria such as *Sodalis pierantonius* strain SOPE (Oakeson *et al.*, 2014) and *Wolbachia* sp. strain WCle (Nikoh *et al.*, 2014). These can be interpreted as these bacteria finding themselves in intermediate stages of genome reduction

and accommodation to an obligate intracellular lifestyle.

I.3.3 *Serratia symbiotica*

As is the case for other endosymbionts, facultative strains of *S. symbiotica* are vertically transmitted (Chen and Purcell, 1997) (for a detailed description see Koga *et al.*, 2012). The secondary symbiont *S. symbiotica* has been found in a variety of different aphids, mainly belonging to the Aphidinae subfamily. Strains of this endosymbiotic bacteria present in members of this subfamily have been suspected to be of facultative nature, since they do not infect 100% of the individuals (Arneodo and Ortego, 2014; Chen and Purcell, 1997; Fukatsu *et al.*, 2000), having naturally occurring monosymbiotic individuals that develop normally without the need of *S. symbiotica* (Chen *et al.*, 2000). However, evidence from wide-scale endosymbiont screening (Henry *et al.*, 2013; Montllor *et al.*, 2002), as well as from artificial infections of aphids with *S. symbiotica* strains (Chen *et al.*, 2000; Montllor *et al.*, 2002), has unveiled that this bacterial species provide their host with a reproductive advantage at higher temperatures.

Sequencing of the facultative *S. symbiotica* strain Tucson from *Ac. pisum* (**SAp**) revealed that its genome displayed a massive gene loss, compared to free-living *Serratia*, and possessed a high ammount of MEs and pseudogenes, similarly to *H. defensa* and *Re. insecticola* (Burke and Moran, 2011). Also, it was determined that *S. symbiotica* had developed an obligate dependence on the aphid host and *Buchnera*, while these last two are not dependent on SAp, confirming its facultative status.

Within the Lachninae subfamily of aphids, a great number of species have been found to be associated with *S. symbiotica* strains (Burke *et al.*, 2009; Chen *et al.*, 2015b; Jousselin *et al.*, 2016; Lamelas *et al.*, 2008). Interestingly, previous phylogenetic analyses of the 16S rRNA gene sequences of these endosymbionts have revealed that the strains group

into two distinct clades, termed **clade A** and **B** (Burke *et al.*, 2009; Lamelas *et al.*, 2008) (Figure I.7). For the purpose of the current work, and to avoid confusion with clades A, B, and C of *Cinara*, from this point on, we will be referring to **clade FL** and **clade OL**, when alluding to clade A and B of *S. symbiotica*, respectively. Interestingly, while clade FL displayed short branches and was made up of mostly facultative *S. symbiotica* strains from Aphidinae aphids, clade OL displayed long branches and consisted in its entirety of strains from Lachninae aphids

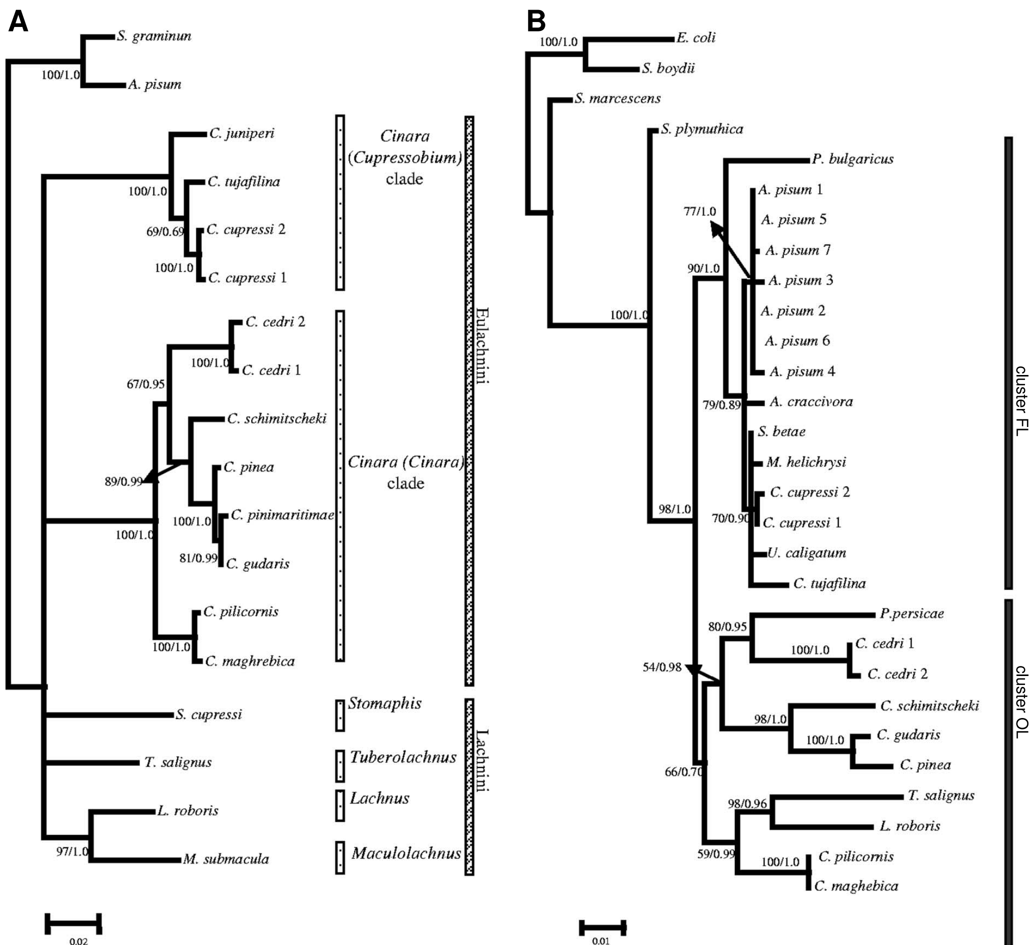


Figure I.7 *S. symbiotica* facultative-like (FL) and obligate-like (OL) clusters. A) Phylogenetic reconstruction of *Buchnera* 16S rDNA sequences from selected species. **B).** Phylogenetic reconstruction of *S. symbiotica* 16S rDNA sequences displaying the facultative-like and obligate like clusters, termed FL and OL respectively. Modified from Lamelas *et al.* (2008).

belonging to the Lachnini, Tuberolachnini, and Eulachnini tribes [*Cinara* spp.]). Additionally, contrary to clade FL, clade OL showed some level of phylogenetic congruency with the aphid hosts', similar to the obligate mutualist *Buchnera*. Most intriguing was the phylogenetic position of *Cinara* spp. from the *Cupressobium* subgenus (*Cinara* clade C), as they were recovered within *S. symbiotica* clade FL. In the view of the phylogenetic evidence, Lamelas *et al.* (2008) performed electron microscopies on *Cinara* (*Cupressobium*) *tujafilina* (clade FL, *Cinara* clade C) and *Cinara* (*Cinara*) *cedri* (clade OL, *Cinara* clade B) bacteriocytes. These microscopies revealed that, while in *C.* (*Cu.*) *tujafilina S. symbiotica* presented a cellular shape and tissue tropism resemblant to facultative endosymbiotic bacteria, *S. symbiotica* from *C.* (*Ci.*) *cedri* displayed a coccoid shape and it was housed exclusively within secondary bacteriocytes, similarly to *Buchnera* (Figure I.8). These results led Lamelas *et al.* (2008) to propose that there had been an infection by a *S. symbiotica* strain in the common ancestor of the Lachnini and Eulachnini tribes, followed by divergence. Subsequently, within members of *Cinara* (*Cinara*), *S. symbiotica* would have established a deep association due to the loss of some essential functions of *Buchnera*, which were taken over by *S. symbiotica*.

The sequencing of both *Buchnera* (Gosalbes *et al.*, 2008; Pérez-Brocal *et al.*, 2006) and *S. symbiotica* (Lamelas *et al.*, 2011b) from *C.* (*Ci.*) *cedri* (**BCc** and **SCc**, respectively) confirmed that these two have indeed established a co-obligate association with the aphid host. BCc, compared against *Buchnera* from *Ac. pisum* (**BAp**), was shown to have lost the ability to synthesise two essential nutrients, riboflavin and tryptophan. While the biosynthesis of the former is now performed by SCc (Lamelas *et al.*, 2011b), the production of the EAA tryptophan is now shared between both partners, BCc and SCc (Gosalbes *et al.*, 2008). Complementary gene-losses in the pathway leading to the synthesis of this compound are such that BCc preserves only the necessary genes to synthesise anthranilate from chorismate (*trpE* and

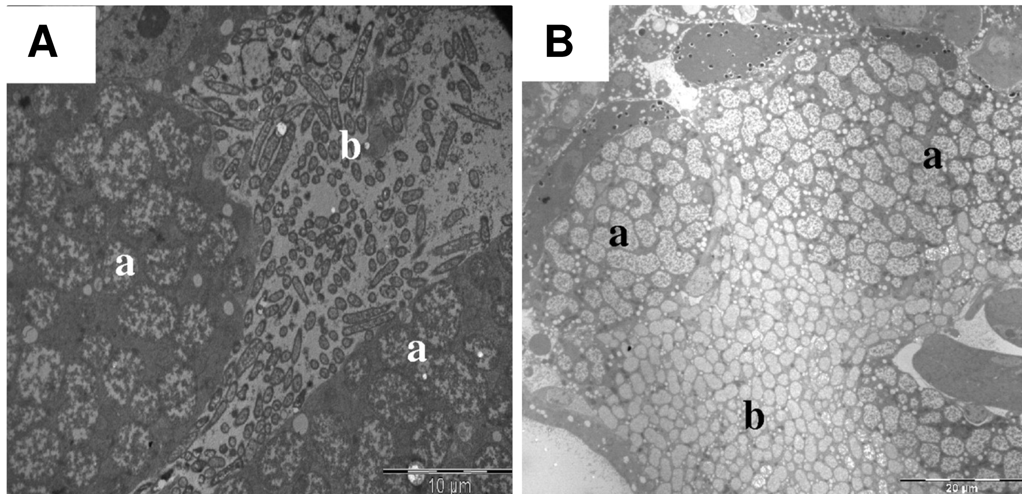


Figure I.8 Electron microscopy of endosymbiotic bacteria from *C. tujafilina* and *C. cedri*. A) Electron microscopy of *C. tujafilina* bacteriocytes displaying (a) *Buchnera* harboured only inside bacteriocytes and (b) *S. symbiotica* in secondary bacteriocytes, displaying a filamentous cell shape. **B).** Electron microscopy of *C. cedri* bacteriocytes displaying both (a) *Buchnera* and (b) *S. symbiotica* housed exclusively in distinct bacteriocytes displaying similar cell shapes. Reproduced from Lamelas *et al.* (2008).

trpG coded in a plasmid), while *S. symbiotica* retains exclusively the genes to synthesise tryptophan from anthranilate (*trpD*, *trpC*, *trpB*, and *trpA* present in the chromosome) (Figure I.9).

Comparative analyses of SCc against the genome of the facultative SAp revealed that these two genomes find themselves in different stages of genome reduction (Manzano-Marín *et al.*, 2012). SCc holds a reduced genome of *ca.* 1.7 megabase pairs (**Mbp**), around one Mbp smaller than that of SAp. However, about half of SCc's genome is composed of "junk" DNA, with no genes or regulatory features. Contrasting what is seen in *Buchnera*, a massive amount of rearragement was found between the two *S. symbiotica* genomes. Nevertheless, no MEs were found, neither active nor pseudogenised, in the genome of SCc, greatly contrasting what is observed in SAp. Consequently, it was concluded that SCc represented the intermediate stages of genome reduction between a facultative endosymbont like SAp, and a highly reduced obligate one, like

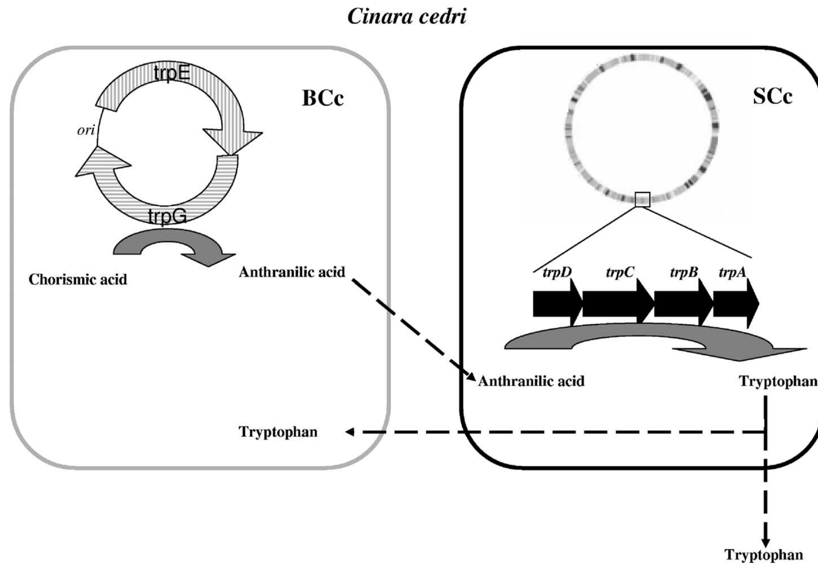


Figure I.9 Proposed metabolic complementation between BCc and SCc for the synthesis of tryptophan. Schematic representation of the predicted split biosynthetic pathway and tryptophan flux in the endosymbiotic system of the aphid *C.* (*Ci.*) *cedri*. Bacterial symbiont cell boundaries are displayed as rounded-corner rectangles. The space around these depicts the host environment. Reproduced from Gosalbes *et al.* (2008).

Buchnera. Additionally, recent pure culture (Sabri *et al.*, 2011) and genome sequencing (Foray *et al.*, 2014) of the *S. symbiotica* strain CWBI-2.3^T, isolated from the Macrosiphini aphid *Aphis fabae* (**SAf**), has revealed a surprising genomic and phenotypic variety among aphid-associated *S. symbiotica* strains. SAf holds a surprisingly large genome (3.58 Mbp), when compared to both SAp and SCc, and compared to SAp, still retains the ability to grow axenically in culture.

Objectives

One crucial question in evolutionary biology is how mutualistic associations are established and how do they evolve. Within the Lachninae subfamily of aphids (Hemiptera: Aphididae), secondary endosymbionts co-exist with the aphid-*Buchnera* symbiotic consortia. While most hosts harbour *S. symbiotica* endosymbionts, some seem to house secondary bacterial associates belonging to different bacterial taxa (Burke *et al.*, 2009; Chen *et al.*, 2015b; Jousselin *et al.*, 2016; Lamelas *et al.*, 2008). Additionally, *S. symbiotica* strains have been found to display contrasting histological and genomic features. In this thesis, we sought to better understand the evolutionary process behind the putative establishment of *S. symbiotica* as a co-obligate endosymbiont within the Lachninae subfamily of aphids and its putative replacement in specific lineages. For this purpose, we defined four main objectives:

1. To explore the genomic changes that occur in the very first stages of the settlement process undergone by a co-obligate endosymbiont. We approached this through the sequencing and assembly of *S. symbiotica* from *C.* (*Cu.*) *tujafilina*. We selected this organism given the histological differences found when compared against the co-obligate *S. symbiotica* from the closely-related *C.* (*Ci.*) *cedri*.

2. To contrast the hypothesis of an ancient acquisition, followed by divergence, of *S. symbiotica* as a co-obligate endosymbiont in the Lachninae *vs.* the replacement of this in the branch leading to the *Cinara* (*Cupressobium*). This was tackled through the comparison of the co-obligate endosymbiotic system found in *C.* (*Ci.*) *cedri* against that of the distantly related *Tu. salignus*, whose *S. symbiotica* both belong to the obligate-like phylogenetic cluster of *S. symbiotica*.

3. To understand the putative establishment and subsequent replacements of *S. symbiotica* within the Lachninae. This was addressed through both microscopic and molecular characterisation of the secondary endosymbionts of distantly-related representative Lachninae aphids.

4. To dissect the genome reduction process in *S. symbiotica* secondary endosymbionts and contrast the results against other naturally reduced endosymbiotic genomes, thereby assessing the generality of these observations. This was approached through the comparison of free-living *Serratia marcescens* strain Db11 *vs.* the divergent *S. symbiotica* strains isolated from different Aphidinae and Lachninae aphids.

Materials and Methods

M.1 Aphid collection and DNA extraction

C. (*Cu.*) *tujafilina* aphids were collected during two consecutive years from a single *Platycladus orientalis* tuja host plant located at 39.51488 north latitude 0.42412 west longitude in the municipality of Paterna, Valencian Community in Spain. Bacteriocyte enrichment from the sample was obtained as in Gil *et al.* (2003), and total DNA extraction was performed following a cetyltrimethylammonium bromide method (Wilson, 2002). DNA from the first year was sent for sequencing to Macrogen Inc. (Korea) where both single-end and paired-end 3kbp libraries were sequenced using 454 FLX and 454 FLX Titanium, respectively. Also an Illumina HiSeq2000 2x100bp 3kbp mate-pair library was prepared with DNA from the second year and sequenced also at Macrogen Inc. (Korea).

Tu. salignus aphids were collected from a single *Salix* sp. tree on September 5th 2013. The collection site is located at the Pacé (Pazieg, in Breton) commune in the Ille-et-Vilaine department of Brittany in north-western France (48.135161 N 1.786938 W). 15 aphids were dissected under a microscope to remove cuticle, gut, legs and head, in order to enrich bacteriome tissues. Mostly embryos, bacteriocytes and other organs that were dragged along were used for DNA extraction using the *JetFlex Genomic DNA Purification Kit* (GENOMED). Extracted DNA was sent to Macrogen Inc. (Korea), where one lane of HiSeq2000 2x100bp paired-end library was sequenced (mean insert size of 417 bps).

Aphid species analysed in Chapter 3 were collected at several sites in

Austria, France, and Spain. Details of the collection can be found on-line.

M.2 Genome assembly and annotation

M.2.1 Preassembly

For 454 reads,we first performed an extraction of the RAW reads using the program **sff_extract** v0.3.0 (developed by the COMAV Institute), with the -l option for removing the 454 FLX titanium linker sequence. Afterwards, both sides of the paired-end reads were rejoined into a single read using a linker of ten undefined nucleotides ("N"). All reads shorter than 100bp were discarded.

HiSeq2000 paired-end reads were first right-tailed trimmed (using a minimum quality threshold of 20) using **FASTX-Toolkit** and reads shorter than 75 bps where removed. Additionally, **PRINSEQ** (Schmieder and Edwards, 2011) was used to remove reads containing undefined nucleotides and to separate the resulting paired-end reads from the ones left without a pair.

M.2.2 *S. symbiotica* SCt-VLC

The remaining 454 reads were taxonomically assigned using **PhymmBL** v3.2 (Brady and Salzberg, 2011) with added genomes of various representatives from the class Insecta (*Atta cephalotes*, *Ac. pisum*, *Drosophila melanogaster*, and *Tribolium castaneum*) and *Homo sapiens* GRCh37.p5, along with their corresponding mitochondrial genomes. We determined that a total of approximately 16% of the 1,033,846 reads corresponded to the *Serratia* genus (161,796 reads), as visualized using **Krona** v2.2 (Ondov *et al.*, 2011). The high-level of assembly of this genome was possible thanks to the use of various bioinformatic and

experimental techniques. Briefly, the 454 reads were assembled using **wgs-assembler** v7.0 (Myers *et al.*, 2000) with option utgGenomeSize set to 2.5 Mbp and batRebuildRepeats turned on. This assembly yielded a total of 187 contigs ordered in 93 scaffolds with an N50 of 77,705 bps and a span of 2,623,798 bps (2,617,736 non-gap bps). After this assembly, the scaffolds were broken and used to map reads to them using **MIRA** v3.4.0 (Chevreux *et al.*, 1999). This process helped to both extend contig ends and to manually inspect each one of the built contigs for inconsistencies or misassemblies resulting in 200 contigs. A custom modified version of **SSPACE** v2.0 (Boetzer *et al.*, 2011) was used to scaffold the contigs using also the HiSeq2000 3kbp valid mate-pair data. This pipeline lead to the ordering of the aforementioned contigs into 105 scaffolds with 96 gaps. Given the repetitive nature of this genome, as is the case for many other facultative or "recently" acquired intracellular endosymbionts, many of the gaps were flanked by repetitive regions or had them near the gaps, and manual prediction of primers avoiding repetitive sequence was time consuming. For this purpose, we developed an *ad hoc* program called **primeScaff**. We retrieved 39 primers that did not overlap at all any masked repetitive sequence and used them for PCR amplification. Out of these, 30 amplifications were positive having 24 producing reads that bridged gaps, 4 that extended contig ends but did not bridged gaps, and 2 which produced multiple amplicons probably due the fact that they fall in repetitive regions that were not identified. From the nine PCRs that failed to produce an amplicon, seven helped us identify wrongly scaffolded gaps and led to gap bridging, whereas the other two led to scaffold breaking. After this, another round of 454 read mapping on the scaffolds using **MIRA** v3.4.0 and visualized using **Gap4** from the **Staden package** (Staden *et al.*, 1999) resulted in 16 more gaps being closed. When performing the same mapping on the contigs of the scaffolds, we identified a great number of small contigs (>200bp and <2kbp) that showed clear signs of being misassemblies of repetitive-region reads, so they were removed from the database. The resulting contigs were scaffolded again, and

GapFiller v1.11 (Boetzer and Pirovano, 2012) was run using the mate-pair HiSeq2000 reads resulting in 34 contigs. These contigs were scaffolded using **SSPACE** v2.0 into 24 scaffolds that were then submitted to **primeScaff** limiting the product size to be between 100 and 3,000bp. Our script then designed six pairs of primers for the ten remaining gaps. Of these, five pairs produced amplicons, which were sequenced by Sanger and then used to close five more gaps. These last 29 contigs were then used for iterative mapping of the 454 reads until no further extension was possible. Finally, after manual comparison of missing data compared with the genome of *S. symbiotica* strain Tucson, three more contigs were assembled and ordered into 22 scaffolds, using **SSPACE** v2.0, with a 454 average coverage of 13.9x. To correct the resulting reference sequence, we iteratively ran **Polisher** v2.0.8 on the 32 contigs using the pretreated HiSeq2000 reads until no more corrections to the reference were made. Finally, we mapped these reads to the "polished" reference using **Bowtie** v2.1.0 (Langmead and Salzberg, 2012) and visualized the result using **Tablet** (Milne *et al.*, 2013) to check for signs of misassemblies or remaining sequencing errors. None were found.

The sequences of *S. symbiotica* SCt-VLC have been deposited at DDBJ/EMBL/GenBank under the accession numbers FR904230-FR904248 and HG934887-HG934889.

M.2.3 *B. aphidicola* strain BTs-Pazieg and *S. symbiotica* strain STs-Pazieg

The remaining HiSeq2000 paired-ends reads were used to perform a *de novo* assembly with **SOAPdenovo** r240 (Luo *et al.*, 2012). The resulting contigs were then filtered by minimum length of 300 and coverage of 50x and where taxonomically assigned using **PhymmBL** v4.0. From these, we separated the contigs into two sets: **1)** the ones that were assigned to *Buchnera*, and **2)** the ones that were assigned to the rest of

Gammaproteobacteria (excluding *Buchnera*). These two sets were used for paired-end read mapping on **Bowtie** v2.2.0. The reads assigned to each contig set where then separately used to perform *de novo* assembly on **SOAPdenovo** r240, with a post-assembly filtering step as explained above. Then, **SSPACE** v3.0 was used to scaffold contigs with a minimum number of links of 30. The resulting gaps in scaffolds where filled using **GapFiller** v1.11 as well as manually curated. One circular contig was obtained for *S. symbiotica* and three for *B. aphidicola*, belonging to the chromosome and the leucine (circular) and tryptophan (linear) plasmids. Finally, to correct base errors in the assembled consensus sequences and manually check for misassembled regions, we used **Polisher** v2.0.8 and **Tablet**.

The sequences of *Buchnera* strain BTs-Pazieg and *S. symbiotica* strain STs-Pazieg have been deposited at DDBJ/EMBL/GenBank under the accession numbers LN890285-LN890287 and LN890288, respectively.

M.2.4 Annotation of sequences and metabolic reconstruction

For circular replicons, the origin of replication was predicted with **originX** (Worning *et al.*, 2006), which in both cases was very near to a putative DnaA-box (used to define the origin). Next, the genomes underwent a round of open reading frame (ORF) prediction using **Prodigal** v2.6.1 (Hyatt *et al.*, 2010), and were then annotated using the **BASys** server (Van Domselaar *et al.*, 2005). Second, a step of manual curation of the annotation was done on **UGENE** v1.18.0 (Okonechnikov *et al.*, 2012) through **blastx** (Altschul, 1997) searches of the intergenic regions as well as through **blastp** and **DELTA-BLAST** (Boratyn *et al.*, 2012) searches of the predicted ORFs against NCBI's nr database. Priority for the BLAST searches was as follows: **1)** against *Escherichia coli* K-12

substrain MG1655, **2)** against *Yersinia pestis* strain CO92, and **3)** against the whole nr database. The resulting coding sequences were considered to be putatively functional if all essential domains for the function were found, or if a literature search supported the truncated version of the protein as functional (details of the literature captured in the annotation file). RNAs were annotated using **tRNAscan-SE** v1.3.1 (Lowe and Eddy, 1997) (tRNAs, with the option -B for bacteria), **TFAM** v1.4 (Ardell, 2006) (to check aminoacyl charging potential for each tRNA species), **ARAGORN** v1.2.36 (Laslett, 2004) (tmRNAs) and **Infernal** v1.1.1 (Nawrocki and Eddy, 2013) (rRNAs and other non-coding RNAs usign the **Rfam** database by Nawrocki *et al.*, 2015). Ribosomal binding sites (RBSs) were predicted with **RBSfinder** v1.0 (Suzek *et al.*, 2001) to aid with the proper prediction of the translation start site. Metabolic reconstruction was performed in **Pathway Tools** (Karp *et al.*, 2015) using the **EcoCyc** (Keseler *et al.*, 2013) and **MetaCyc** (Caspi *et al.*, 2014) databases, followed by manual curation. Visual plotting of the inferred metabolism was done by hand using **Inkscape** v0.91.

Re-annotation of genomes, selected genes, and other RNA features was performed as described above. Annotation of specific domains of proteins for Chapter 4 was performed through **InterProScan** (Jones *et al.*, 2014) searches.

M.3 Fluorescence *in situ* hybridisation

Aphids used for fluorescence *in situ* hybridisation (**FISH**) experiments were dissected in absolute ethanol to extract embryos. These were then directly transferred to modified Carnoy's fixative (6 chloroform : 3 absolute ethanol : 1 glacial acetic acid) and left overnight, following Koga *et al.* (2009) protocol to quench autofluorescence. Briefly, fixed embryos were washed with absolute ethanol and transferred into a 6% solution of H_2O_2 diluted in absolute ethanol and were then left in this solution for two to

six weeks (changing the solution every three days). When bleached, they were washed twice with absolute ethanol and stored at -20℃. Staining was performed overnight at 28 °C in standard hybridisation buffer (20mM Tris-HCl [pH 8.0], 0.9 M NaCl, 0.01% SDS, and 30% formamide) and then washed (20mM Tris-HCl [pH 8.0], 5mM EDTA, 0.1 M NaCl, and 0.01% SDS) before slide preparation. The slides were examined using a confocal laser scanning microscope (TCS SP5 X, Leica; and FV1000, Olympus). A list of specific probes used for each aphid species is available on-line. The embryos from at least 10 individuals were analysed per sample.

M.4 16S rRNA gene PCR, cloning, and sequencing

Since all endosymbionts detected in Lachninae members so far are bacteria, we used the primers 16SA1 and 16SB1 (Fukatsu and Nikoh, 1998) to amplify partial 16rRNA genes for cloning. This cloning stategy was adopted in selected cases to facilitate phylogenetic reconstruction. Resulting amplicons were cloned into the pGEM-T Easy Vector (Promega) and SP6 (5'-ATTTAGGTGACACTATAG-3') and T7 (5'-TAATACGACTCACTATAGGG-3') primers were used for amplification and sequencing of the cloned DNA (at least 5 clones from each species). Specific primers for either *Buchnera* or *S. symbiotica* were designed based on the FISH probes. Specific PCR reactions and sequencing were done mainly to confirm the presence of the secondary endosymbionts. In the case of the SLSS and "X-type", specific primers were designed based on Attardo *et al.* (2008) ("Sodalis specific") and Guay *et al.* (2009) ("PAXSF"), respectively. A full list of primers pairs and PCR conditions is available on-line.

M.5 Protein clustering and genome rearrangements

For both performing phylogenetic inferences and understanding the genetic differences in both *Buchnera* and *Serratia* from the different aphids, we first ran a homologous protein clustering analysis using **OrthoMCL** v2.0.9 (Chen *et al.*, 2007; Li, 2003) using the afore-mentioned endosymbiotic bacteria and a set of closely-related free-living bacteria. These protein clusters were then manually curated to join in a single group proteins such as flagellum and outer membrane proteins which do not tend to cluster together given their low protein identity. Based on these manually curated groups we then created subsets containing single copy-core proteins shared by all selected strains. Using the latter set, we recoded the gene arrangements in each *S. symbiotica* genome for use with **MGR** v2.03 (Bourque and Pevzner, 2002) to infer the phylogeny that absolutely minimises (no heuristics) the number of rearrangements undergone among the strains.

Tables containing all species and accessions used in this study can be found on-line.

M.6 Phylogenetic analyses

For phylogenomic reconstructions, full protein sets of organisms of interest were retrieved from NCBI. Alignments were performed using **MAFFT** v7.220 (Katoh and Standley, 2013) (L-INS-i algorithm). We then removed divergent and ambiguously aligned blocks using **Gblocks** v0.91b (Talavera and Castresana, 2007) and concatenated the resulting alignments for following phylogenetic inference. For both *Buchnera* plus its free-living relatives and *S. symbiotica* plus its free-living relatives, we used the LG+I+G amino acid substitution model, which incorporates the

variability of evolutionary rates across sites in the matrix estimation (Le and Gascuel, 2008). Bayesian phylogenetic inference was performed in **MrBayes** v3.2.5 (Ronquist *et al.*, 2012), running two independent analyses with four chains each. In order to alleviate long-branch attraction artefacts commonly seen in endosymbionts (Husník *et al.*, 2011; Philippe and Roure, 2011), **Phylobayes** v4.1 (Lartillot *et al.*, 2009) was also run under the CAT+GTR+G (4 discrete categories) (under 4 independent runs) using dayhoff6-recoded concatenated amino acid alignments. Chains were run and compared using the **tracecomp** and **bpcomp** programs, and were considered converged at a maximum discrepancy of <0.16 and minimum effective size of 50. The resulting trees were visualised and exported with **FigTree** v1.4.1 and edited in **Inkscape**.

For the phylogenomic reconstruction shown in Chapter 1, a maximum-likelihood tree was calculated using 1,000 full bootstrap replicates with **RAxML** v7.7.6 (Stamatakis, 2006) using the PROTGAMMAWAGF substitution matrix.

For divergence time estimations, we performed two likelihood tests implemented in **MEGA** v6 (Tamura *et al.*, 2013) on the 24 *Buchnera* conserved genes that had been previously identified to follow the molecular clock hypothesis (Pérez-Brocal *et al.*, 2006). From these, we identified 8 single-copy conserved genes shared by all *Buchnera* genomes available to date (*ssb*, *rbfA*, *cspE*, *rpsS*, *dapF*, *leuB*, *infA*, and *rpmJ*) which still follow this hypothesis under Figure 2.3A topology. The amino acid sequences from these genes were aligned independently using **MAFFT** v7.220 and then back-translated to their nucleotide sequences. Concatenated alignments were used for divergence time estimations in **Phylobayes** v4.1 running two independent analyses with two chains each under the GTR nucleotide substitution model and an underlying across-site rate variations sampled from a discrete gamma distribution (4 categories) of mean 1 and variance 1/alpha (exponential of mean 1) (Phylobayes options -gtr -ln -dgam 4). A fixed topology derived

from the Bayesian reconstruction done with the single-copy core genes of the currently available *Buchnera* genomes and free-living relatives was used as input (since Phylobayes requires a fixed topology). A log normal autocorrelated relaxed clock was chosen with a root prior of 100 million years plus/minus 100, this last based on previous dating hypothesis of the origin of the family Aphididae (von Dohlen and Moran, 2000) (between 84-164 million years ago), which should be close to that of the splitting of Aphidinae and Lachninae subfamilies (Nováková *et al.*, 2013). We used the time divergence estimate between the tribes Aphidini and Macrosiphini (50-70 million years ago) (Clark *et al.*, 1999) as the calibration point for the molecular dating. In addition, a chain was run under the prior (-prior option) and was checked to assess if the resulting distributions were sufficiently wide, finding them to be so. Briefly, regarding the phylogenetic analyses of the *trpD*, *trpC*, *trpB*, and *trpA* genes, we aligned them using **MAFFT**, then fed the alignments into **Gblocks**, followed by phylogenetic reconstruction in **MrBayes**, as described above.

For Chapter 3, all phylogenetic analyses were performed as follows. First **SSU-ALIGN** v0.1 was used to align 16S rRNA sequences, followed by visual inspection of the alignments in **AliView** v1.17.1 (Larsson, 2014).Then, **GBlocks** v0.91b was used to eliminate poorly aligned positions and divergent regions with the option '-b5=h' to allow half of the positions with a gap. The final alignments were transformed into nexus format for phylogenetic analysis in **MrBayes** v3.2.5 under the GTR+I+G model. Two independent runs, each with four chains (three "heated", one "cold"), were run for 5,000,000 generations discarding the first 25% as burn-in and checked for convergence. Visualisation and tree-editing was done in **FigTree** v1.4.1 and **Inkscape** v0.91, respectively.

Chapter 1

The "early" co-obligate *Serratia symbiotica* endosymbiont of the aphid *Cinara* (*Cupressobium*) *tujafilina*

Figure 1.1 ***C.*** **(*Cu.*)** ***tujafilina*** **apterous female and nymphs.** Reproduced from www.nbair.res.in.

1.1 Introduction

C. (*Cu.*) *tujafilina* (Figure 1.1) feeds primarily on *Platycladus orientalis* and commonly presents an anholocyclic life cycle, meaning it produces no sexuales and reproduces solely through parthenogenetic females. However, a recent work by Durak and Durak (2015) found that *C.* (*Cu.*) *tujafilina* may have not lost the genetic ability to produce males, and under favourable conditions, can even develop in an holocyclic manner. This supported previous findings of a male (Zhuravlev, 2003) and an oviparous female of this species (Remaudière and Binazzi, 2003). From rearing experiments using a polish population of *C.* (*Cu.*) *tujafilina*, it was found that the optimal temperature for the development of this aphid species was 25 ° C, which was characterised by the shortest pre-reproduction period, the highest fecundity, and the highest demographic parameters (Durak and Borowiak-Sobkowiak, 2013).

As mentioned before, *C.* (*Cu.*) *tujafilina* harbours two endosymbiotic bacteria: a pleomorphic *Buchnera* inside bacteriocytes and a filamentous *S. symbiotica* with a broad tissue tropism (Lamelas *et al.*, 2008). The sequencing of *Buchnera* from *C.* (*Cu.*) *tujafilina* (**BCt**) revealed that, after *C.* (*Ci.*) *cedri*'s (BCc), it held the second smallest *Buchnera* genome sequenced so far (Lamelas *et al.*, 2011a). BCt holds a genome of just 444,930 kilobase pairs (**kbp**), just about 24 kbp larger than that of BCc. However, the *S. symbiotica* partners of BCc and BCt present a different cell morphology and tissue tropism (Lamelas *et al.*, 2008). Comparative genomics of the different *Buchnera* genomes available revealed that, unlike BCc, BCt still retains the chromosomal genes *trpD*, *trpC*, *trpB*, and *trpA*, meaning it codes for the necessary enzymes to synthesise tryptophan from chorismate and L-glutamine. This metabolic capability is shared by *Buchnera* of mono-endosymbiotic systems from other aphids. However, its genome, just as BCc's, has lost all the genes coding for the enzymes in charge of synthesising riboflavin, proven to be essential for *Ac.* *pisum* aphids (Nakabachi and Ishikawa, 1999). Since this

biosynthetic loss was only found in the *Buchnera* genomes from both *C.* (*Ci.*) *cedri* and *C.* (*Cu.*) *tujafilina*, it was hypothesised that the latter aphid could either afford this loss, or a second bacterium (*S. symbiotica*) was doing the job, as is the case for *C.* (*Ci.*) *cedri*.

To solve the questions of whether or not *S. symbiotica* is providing any essential nutrients to the aphid-*Buchnera* consortium in *C.* (*Cu.*) *tujafilina* and to infer the status of this symbiont in the association, the genome of *S. symbiotica* strain SCt-VLC from the aphid *C.* (*Cu.*) *tujafilina* (**SCt**) was sequenced. We have compared it against the genomes of both SAp (*Ac. pisum*) and SCc (*C.* [*Ci.*] *cedri*), as well as with the ones of free-living *Serratia*. We were able to determine its phylogenetic positioning and elucidate the process of genome reduction undergone, not only by SCt, but also by the other *S. symbiotica*. We also describe the genome re-ordering, gene inactivation and the role that MEs have played in these processes. Finally, and most significantly, We propose that a metabolic inactivation resulting from gene erosions in *B. aphidicola*, and not by losses in the secondary endosymbiont, is behind the establishment of *S. symbiotica* as an obligate endosymbiont in at least the *Cinara* genus.

1.2 Results & Discussion

1.2.1 Localisation of *S. symbiotica* in bacteriomes of *C.* (*Cu.*) *tujafilina*

Fluorescence *in situ* hybridisation (**FISH**) of *C.* (*Cu.*) *tujafilina* aphid embryos using a specific probe targeting the 16S rRNA molecule of *S. symbiotica* confirmed that this secondary endosymbiotic bacterium does indeed possess a filamentous-shaped cell (Figure 1.2). It is distributed along the bacteriome mainly surrounding *Buchnera* bacteriocytes, and sometimes inhabiting the cytoplasm of its own distinct ones. This cell

morphology and tissue tropism highly resembles that of facultative endosymbionts from *Ac. pisum* (Fukatsu *et al.*, 2000; Moran *et al.*, 2005; Sakurai *et al.*, 2005; Sandström *et al.*, 2001; Tsuchida *et al.*, 2005, 2010), including SAp. However, contrary to what is generally observed for facultative endosymbionts, 100% of individuals analysed harbour this associate, pointing towards an obligate association between this endosymbiont and its aphid host.

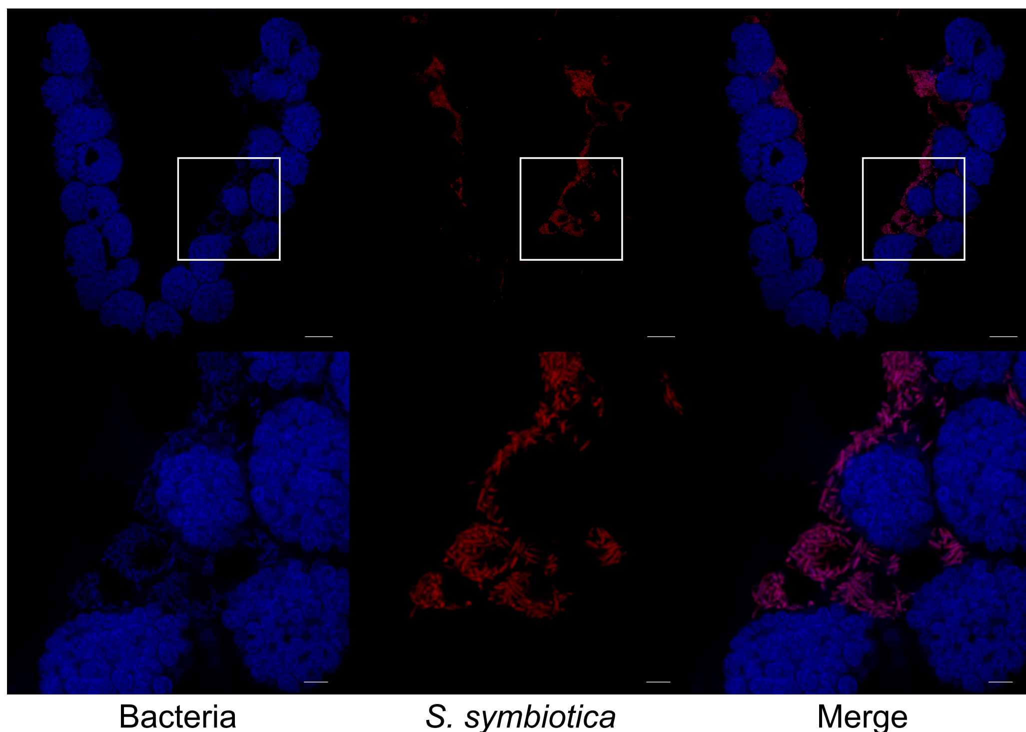


Figure 1.2 *C.* (*Cu.*) *tujafilina* bacteriomes displaying the localisation of *S. symbioyica*. FISH microscopy false coloured images displaying the filamentous-shaped *S. symbiotica* both surrounding primary bacteriocytes and infecting separate ones.

1.2.2 The *S. symbiotica* strain SCt-VLC genome

The genome of *S. symbiotica* strain SCt-VLC has been assembled to 32 contigs organized into 22 scaffolds spanning 2,494,579 base pairs (**bps**) (Figure 1.3), with an average G+C content of ~52%. Both the genome

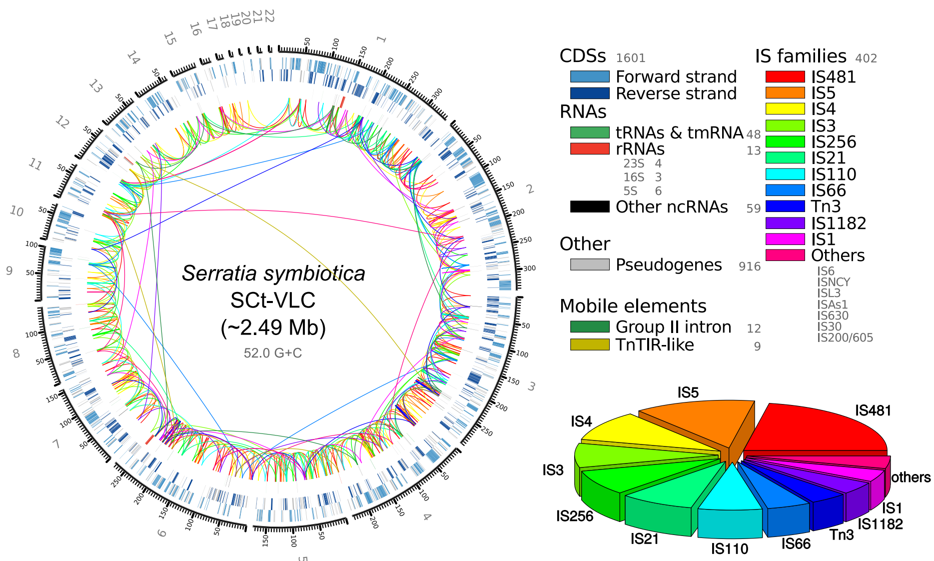


Figure 1.3 ***S. symbiotica*** **strain SCt-VLC genome.** SCt's circular genome diagram displaying the variety and genome-wide invasion by MEs. **Left:** Circular representation of the scaffolds composing the genome of SCt. From outer to inner, the rings show features on the forward strand, on the reverse strand, RNA features, and lines connecting different types of MEs. **Right top:** Colour-coding for the various elements displayed in the circular representation of the genome and their absolute abundance. **Right bottom:** Pie-chart depicting the relative abundances of IS element proteins in the genome. Reproduced from Manzano-Marín and Latorre (2014).

size and G+C content of SCt are quite similar to the ones calculated for SAp, a facultative endosymbiont of *Ac. pisum* (**table** 1.1). Additionally, their metabolic profiles have been quite similarly affected by genome reduction (Figure S.3). However, SCt's genome evidences a greater gene inactivation and a slighly smaller genome size, although this latter might differ from that of SAp given the highly-fragmented genome assembly of this (Burke and Moran, 2011). SCt's genome codes for 1,601 intact putative protein-coding sequences (**CDSs**) and 916 putative pseudogenes, compared to the 550 in SAp. It codes for one tmRNA and 47 tRNAs, with amino acid charging potential for all 20 standard amino acids, plus two tRNAs with amino acid charging potential for formylmethionine and one for lysylated isoleucine. Compared to SAp, it has lost the tRNA coding for selenocysteine. Out of the 13 rRNA genes it

possesses, two copies of the 23S rRNA gene run-off from contig ends, being unable to determine their completeness. SCt presents two intact copies of the full ribosomal RNA operon, and as seen in other "recently" established endosymbionts (including SAp) (Burke and Moran, 2011; Degnan *et al.*, 2009a,b; Oakeson *et al.*, 2014), its genome presents a high amount of mobile DNA (~13.4% of the total genome), with a variety of insertion sequence (**IS**) proteins belonging to ~18 different families that are scattered throughout the genome.

Features	***S. marcescens* Db11**	***S. symbiotica***		
		SAp	**SCt**	**SCc**
Chromosome (Mbp)	5.11	2.76	2.49	1.76
Mean G+C (%)	59.5	48.4	52.0	29.2
Predicted CDSs	4,709	2,098	1,601	677
Pseudogenes	12	550	916	108
rRNAs (23S,16S,5S)	7,7,8	5,5,5	4,3,6	1,1,1
tRNAs	88	44	47	36
Other rRNAs	69	57	59	6
Mobile elements	few	many	many	none
Cell shape	rod	filamentous	filamentous	spherical
Lifestyle	free living	facultative	co-obligate	co-obligate

Table 1.1 Comparison of different strains of *S. symbiotica* genomes. Comparison of genomic features from *S. symbiotica* strains contrasting with characteristics of a free-living relative (*S. marcescens* Db11).

Through a phylogenetic reconstruction using 354 single-copy proteins shared among selected organisms it was corroborated that the three *S. symbiotica* form a monophyletic clade nested within the *Serratia* genus (Figure 1.4). Similarly to SAp, the branch-length leading to SCt suggests a small genetic divergence from its free-living relatives. This means the two genomes are "closely" related, thus the differences in the gene loss-state must be related to the biological characteristics of their hosts or the roles their *Buchnera* partners play in nutrient provisioning. Regarding SCc (co-obligate endosymbiont with *Buchnera* in *C.* [*Ci.*] *cedri*), the long branch leading to it evidences an accelerated evolutionary process, reflected in its further genome reduction and accommodation as an obligate intracellular endosymbiont.

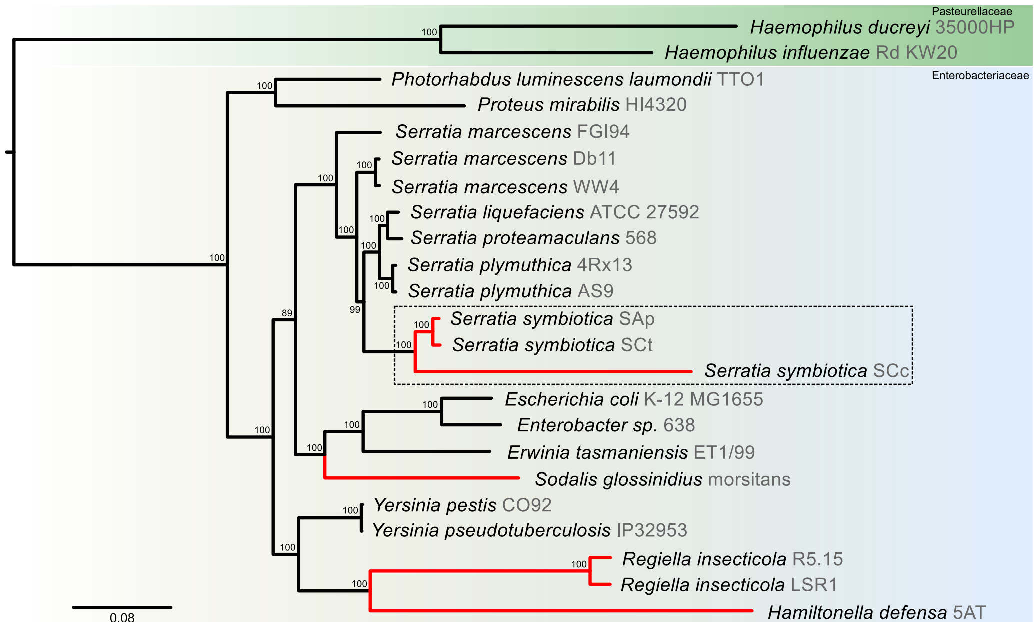


Figure 1.4 SCt's phylogenetic positioning. *Serratia* maximum-likelihood phylogenetic reconstruction using concatenated single-copy orthologs shared by all selected strains. SCt, as SAp, has diverged less than its endosymbiotic relative in *C.* (*Ci.*) *cedri* (SCc). In grey the strain designation is showed. Reproduced from Manzano-Marín and Latorre (2014).

1.2.3 Metabolic capabilities

Figure 1.5 shows the metabolic reconstruction for SCt. It is an aerobe bacterium having a complete cytochrome *bo* oxidase and, as many other endosymbionts, it can use acetyl-CoA to produce acetate and energy under oxygen-limiting conditions. It retains a complete ATP synthase and it can grow on different carbon sources such as glucose, fructose, and mannitol, also retaining complete PTS transport systems for these sugars. Unlike SAp, it has lost the mannose PTS system and the ability to grow on trehalose and N-acetylglucosamine. It has the *treC* gene pseudogenised (responsible for the conversion of trehalose 6-phosphate to b-D-glucose 6-phosphate) in addition to a pseudogenised version of *nagA* and a deletion of the *nagB* and *nagE* genes, impairing the import and conversion of N-acetylglucosamine to fructose-6-P. As other highly

reduced endosymbionts, it presents great pseudogenisation in the genes involved in the TCA cycle. This is in clear contrast with SAp, which putatively still presents a complete cycle (Burke and Moran, 2011). In this respect, it closer resembles SCc that, as all *Buchnera*, has lost almost all the genes involved in this pathway. Regarding the cell wall, SCt retains the ability to synthesize peptidoglycan (**lipid II**), enterobacterial common antigen (**ECA** or **lipid III**) and lipopolysaccharides (**LPS**), like other facultative endosmbionts.

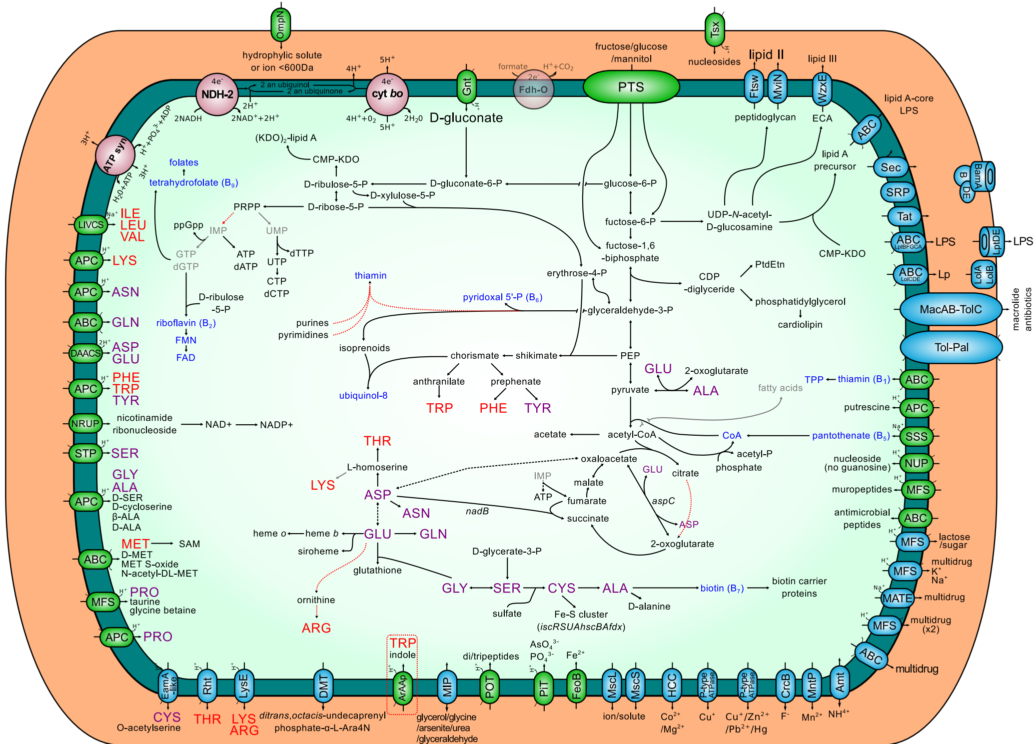


Figure 1.5 SCt metabolic reconstruction. SCt metabolic reconstruction as done by **PathwayTools**. Intact pathways are shown in black lines, unclear pathways are shown in grey lines, and the ones that are already represented elsewhere with another line are shown as dotted lines. Importers are represented with green ovals while exporters and exporters/importers are represented with blue ovals. EAAs and NEAAs are shown in red and purple lettering, respectively. Grey compounds represent those for which biosynthesis cannot be accounted for by the genomic data of SCt. Formate dehydrogenase-O is shown semi-transparent, given that the pseudogenisation of one of its genes is unclear, it being a selenoprotein. Red-dotted lines and boxes represent metabolic elements present in SAp but not SCt. Modified from Manzano-Marín and Latorre (2014).

As for amino acid biosynthesis, it retains the capability of synthesizing 4 EAAs and 9 NEAAs. It preserves complete pathways to synthesize the three aromatic amino acids: phenylalanine, tryptophan, and tyrosine from phosphoenol pyruvate and D-erythrose 4-phosphate. Most importantly, SCt can indeed synthesize tryptophan, not having the *trpE* and *trpG* genes pseudogenised nor lost, as SAp and SCc respectively. SCt is able to: interconvert aspartate to glutamate (through oxaloacetate and 2-oxoglutarrate, respectively), both of which could be imported; synthesize asparagine and threonine from L-aspartate; glutamine from L-glutamate; serine from D-glycerate-3-P; and glycine, cysteine and alanine from L-serine. The lysine biosynthetic pathway (DAP group) presents pseudogenization in both *astC* and *argD* genes, but still retains the rest of the pathway. It could be speculated that another transaminase might be doing the job, as in the highly reduced SCc, where these two genes have also been lost, while the rest of the pathway (except for the last step) is still encoded on the genome (Lamelas *et al.*, 2011b). Similarly, in both BCc and BCt, neither the *argD* nor *astC* genes are present. SCt retains specific transporters for importing most amino acids it cannot synthesize, except for arginine and histidine, as is the case for both SCc and SAp.

With respect to cofactors and vitamins, SCt preserves complete pathways for the biosynthesis of tetrahydrofolate (**THF**), flavin mononucleotide (**FMN**), flavin adenine dinucleotide (**FAD**), pyridoxal 5'P (vitamin B_6), ubiquinone, biotin (vitamin B_7), and riboflavin (vitamin B_2). This last vitamin is of special interest, since most *Buchnera* endosymbionts sequenced to date are able to synthesise it (from ribulose-5-P and GTP) and provide it to their host, which in *Ac. pisum* has been described as essential (Nakabachi and Ishikawa, 1999). The only exceptions to this are the *Buchnera* from both Lachninae subfamily representatives, *C.* (*Cu.*) *tujafilina* and *C.* (*Ci.*) *cedri*, which are missing all the genes involved in this pathway, rendering *Buchnera* unable to provide riboflavin to the aphid host, and turning its *Serratia* partner

indispensable for the biosynthesis of this vitamin. Biosynthesis of other vitamins and cofactors such as thiamin pyrophosphate (**TPP**), biotin and coenzyme A (**CoA**) could be possible given an external supply of the required intermediaries. SCt, as SCc but unlike SAp, preserves a thiamin ABC transporter (made up from the products of the *thiQ*, *thiP*, and *thiB* genes), which would in turn put a selective pressure to preserve only the genes *thiE* and *thiL* to be able to convert the imported thiamin into TPP. This is exactly the case in the more drastically reduced SCc. All this points towards the fact that it may be losing the selective pressure to retain the thiamin biosynthetic genes, hence it is possible to speculate that the thiamin biosynthesis pathway has recently started to undergo erosion. Finally, regarding CoA, it could be synthesised from pantothenate, which would be imported to the cell through a pantothenate/sodium SSS symporter encoded by the *panF* gene.

Concerning the synthesis of nucleotides, a very interesting decay pattern was observed. Contrary to SAp but similarly to SCc, the genes to synthesize inosine monophosphate (**IMP**) from 5-phosphoribosyl 1-pyrophosphate (**PRPP**) have been lost or pseudogenized, whilst the genes to synthesize uridine monophosphate (**UMP**) from PRPP have been retained. It is worth noticing that although the *pyrB* gene is pseudogenized by a frameshift, it might still produce a functional protein. This would confer SCt the capacity to synthesize pyrimidines *de novo* but not purines, which would have to be imported.

SCt, as SAp, codes for a variety of translocation systems such as Sec, twin-arginine (**Tat**) and the signal-recognition particle (**SRP**). It presents a complete MacAB-TolC macrolide efflux transport system which could provide resistance via active drug efflux. It also encodes for an intact Tol-Pal cell envelope complex and a diversity of export systems for a variety of compounds. This richer repertoire of translocators is in clear contrast with SCc, which only retains a small subset of these systems (Lamelas *et al.*, 2011b).

1.2.4 Genome rearrangements and mobile elements

The role MEs play in the genome reduction process undergone by bacteria transitioning from a free-living lifestyle to an obligate MEs-deprived endosymbiontic one, has typically been derived from comparative studies between "ancient" endosymbionts (lacking MEs) and distantly related free-living bacteria (with a "controlled" number of these). Some examples of recently-acquired endosymbionts include that of SAp (a member of the *Serratia*) (Burke *et al.*, 2009) and *Sodalis pierantonius* strain SOPE (belonging to the same genus as the recently isolated *Sodalis praecaptivus* strain HS^T) (Chari *et al.*, 2015; Clayton *et al.*, 2012; Oakeson *et al.*, 2014). However, a most interesting comparison is the one between two phylogenetically closely related strains enriched in MEs, as is the case of SAp and SCt. Through a minimal number of rearrangements phylogeny using *Serratia* shared single copy genes (Figure 1.6), it becomes evident that while no rearrangements are

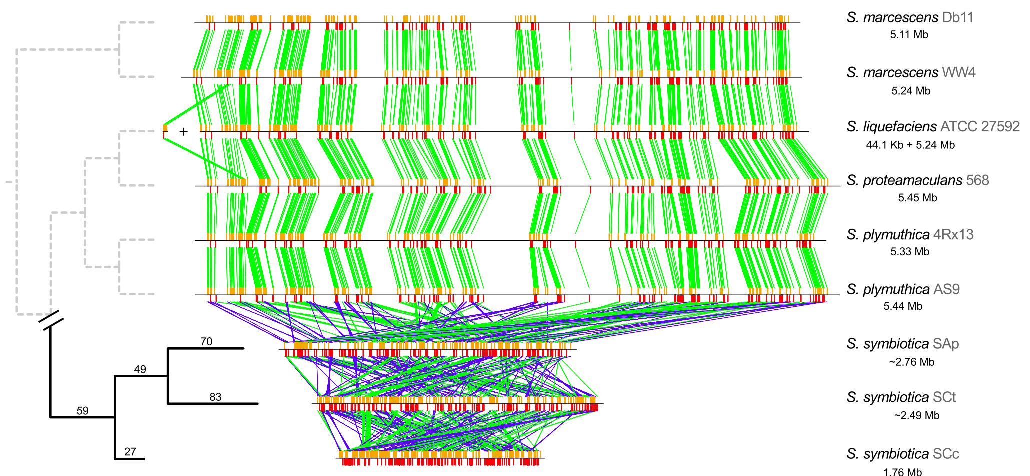


Figure 1.6 *S. symbiotica* minimum number of rearrangements tree based on the single-copy core genes of the selected *Serratia* spp. Left: rooted minimum number of rearrangements tree as calculated by **MGR**. **Right:** pairwise synteny plots of free-living *Serratia* along with endosymbiotic relatives SAp, SCt, and SCc. Species names (the strain designation is showed in grey) are shown along with its genomic size in Mbp. Reproduced from Manzano-Marín and Latorre (2014).

observed among the free-living *Serratia*, a great number of them have occurred on the branch leading to the *S. symbiotica* clade. Surprisingly, a great number of rearrangements are observed even between SCt and SAp. Although this result comes from the ordering of both SAp and SCt scaffolds in a way that minimises the rearrangements between these and free-living *Serratia*, a similar number is obtained when SAp's scaffolds are arranged taking SCt's as reference (Figure S.4). This result is in clear contrast with the fact that, phylogenetically, these two strains find themselves extremely close together, not having diverged greatly since the last common ancestor. Therefore, it can be postulated that SCt, SAp, and SCc are three divergent lineages of *S. symbiotica*, each of which has undergone a particular genome rearrangement process, resulting in different architectures for each endosymbiotic strain.

Among the different MEs SCt possesses (Figure 1.3), IS481 is the most abundant. Nevertheless, many other types of MEs are also shared by SAp. Among these, a putative quorum-sensing system mobile element, similar to the termed Tn*TIR* in *S. marcescens* strain SS-1 (Wei *et al.*, 2006a,b), was identified. This mobile element is also present, but not identified, in SAp, as well as a group II intron mobile element (**GIIME**) (retrotransposon). This latter is also found in the genomes of other two facultative endosymbionts of aphids: *Re. insecticola* strain LSR1 and *H. defensa* strain 5AT.

In order to explore how the high presence of MEs have impacted genome rearrangement in *S. symbiotica*, the genomes of SCt and SAp were compared. SCc's was not taken into account for this analysis, since it significantly diverges from the other two strains and lacks any traces of MEs. A total of 165 syntenic clusters were identified between SCt and SAp, with an average size of 9,596.90 bps (for SCt). The majority of these clusters (129) are flanked (+/-3 kbp), on at least one side, by an IS gene, 22 by putative phage elements, 11 by GIIMEs, and 4 by Tn*TIR*-like elements (Figure 1.7: inner rings). In some instances (when information available from SAp) both SCt and SAp displayed the same

MEs flanking the syntenic clusters. This suggests that the same MEs have mediated different rearrangements in these two organisms. Hence, we

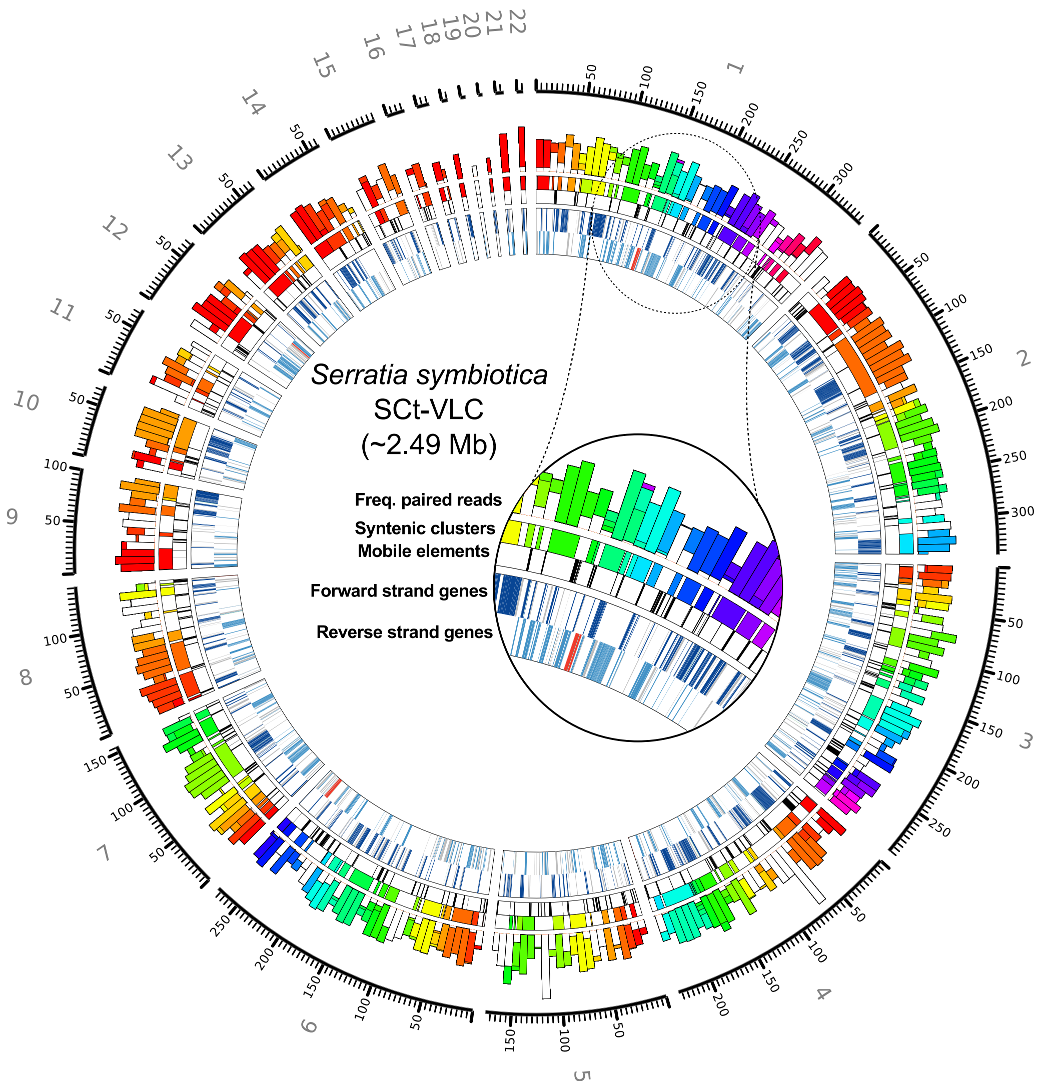


Figure 1.7 *S. symbiotica* discordant paired-end reads and syntenic clusters. Discordant paired-end alignments for *S. symbiotica* SCt-VLC scaffolds displaying no detectable intra-population rearrangements. The outer ring represents the absolute abundance of paired-end reads colour-coded according to the syntenic cluster on the same scaffold their mate occurs at. From outer to inner, the rest of the rings represent the colour-coded synteny clusters, the mobile-element proteins in black, the forward strand features, and the reverse strand features (colour-coded as in Figure 1.3), respectively. Modified from Manzano-Marín and Latorre (2014).

can confidently determine that IS MEs have been the key factor promoting rearrangements.

Given the above results, and to inquire into intra-population rearrangements, all HiSeq2000 non-repetitive paired-end reads were mapped, and those that showed discordant alignments (wrong orientation or too big or small insert size) were extracted (Figure 1.7: outer rings). Surprisingly, no major indication of intra-population rearrangements was found, as determined by the lack of important clusters of orientation-discordant paired reads. Thus, while SCt's genome harbours a high amount of mobile DNA, it is quite stable in a particular population and point in time, despite the great availability of mobile substrate. This would indicate that rearrangements might be happening in a slow fashion, although periodic and/or single-cell resequencing would be needed to determine the fixation rates of new gene orders and the variation within a population.

In addition to frameshifts, MEs have been proposed as a minor driving force of gene pseudogenisation and genome size reduction in young endosymbiont genomes (Belda *et al.*, 2010; Oakeson *et al.*, 2014). In SCt, at least 11 gene-sequences are split by MEs onto separate parts across the genome. In some cases, while SAp presents an intact CDS or pseudogene, SCt presents a translocated (Figure 1.8A) or even an interrupted (Figure 1.8B) version, caused by the insertion/action of one or more MEs. Additionally, 7 cases of this type of pseudogenisation were found, but given the nature of the inactivated protein or the lack of comparable sequence in SAp, we were unable to determine if these pseudogenisations were unique to SCt strain. Additionally, many pseudogenized mobile-element proteins form "genomic wastelands" composed of a variety of inactivated proteins ordered in tandem.

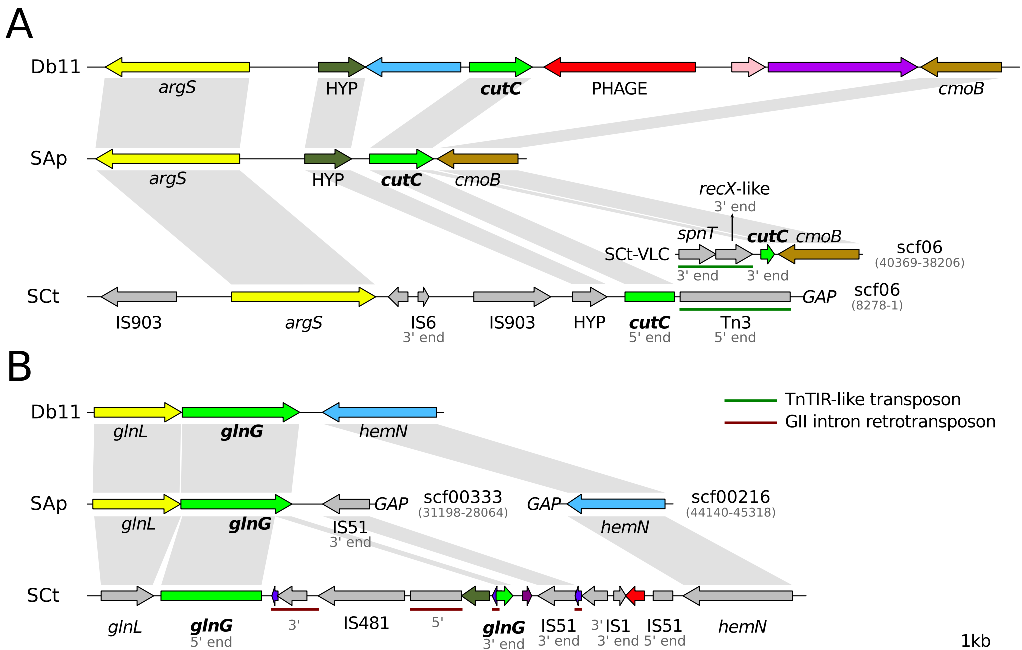


Figure 1.8 Mobile element-driven genetic inactivation in *S. symbiotica*. Examples of *S. symbiotica* loci where mobile element-driven gene inactivation and genome rearrangement is evidenced. **(A)** Translocated *cutC* gene. **(B)** Interrupted *glnG* gene. Comparison of *S. symbiotica* strains is shown against *S. marcescens* Db11. Coloured arrows represent intact CDS, while grey ones pseudogenes. Genes of interest are always coloured. Modified from Manzano-Marín and Latorre (2014).

1.2.5 Decay of EAAs biosynthetic operons and genes

Since the main role of *Buchnera* is to supply EAAs lacking from the aphid diet, the degradation of these routes in endosymbionts coexisting with *Buchnera* should serve as a "thermometer" of their level of accommodation to it. Such a difference is evident between the facultative SAp and the obligate SCc (Burke and Moran, 2011; Lamelas *et al.*, 2011b), the latter having lost most genes involved in the biosynthesis of EAAs. By comparing free-living *S. marcescens* Db11's EAAs' biosynthetic genetic repertoire against that of SAp, SCt, and SCc, a gradual degradation (from Db11 to SCc) of these genes becomes evident (Figure 1.9). This is particularly pronounced in the genes

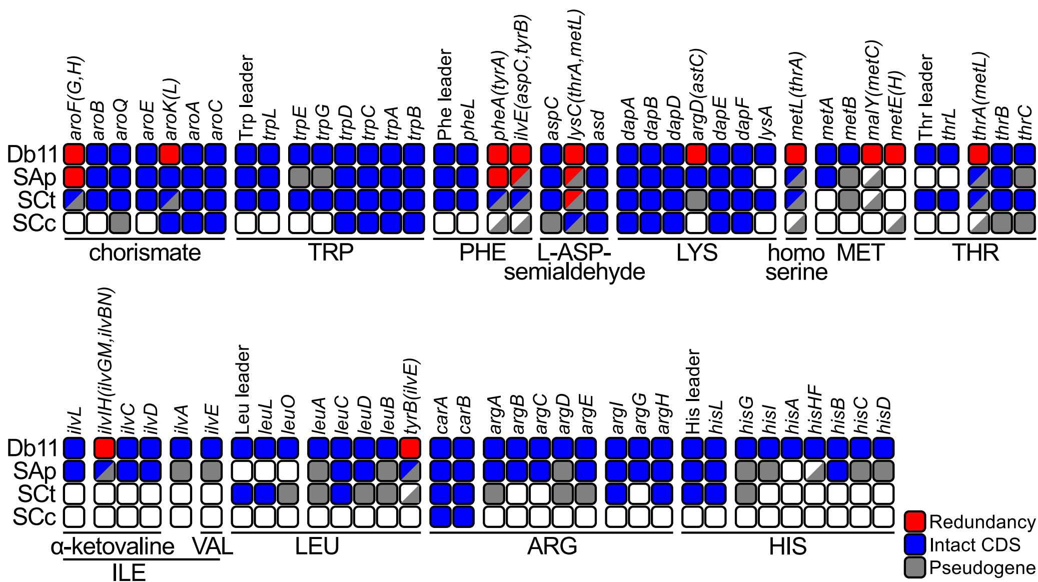


Figure 1.9 Erosion of EAAs' biosynthetic genes in *S. symbiotica* endosymbionts. Inactivation tables representing the erosion of EAAs-biosynthetic genes in *S. symbiotica* endosymbionts compared to the free-living *S. marcescens* Db11. In the bottom-right, colour code of box filling. Modified from Manzano-Marín and Latorre (2014).

involved in the biosynthesis of arginine, isoleucine, and valine, all of which can be synthesized by *Buchnera* in cooperation with the aphid host in *Ac. pisum* (Hansen and Moran, 2011). This provides additional evidence towards SCt being one step further than SAp in becoming a settled down co-obligate aphid endosymbiont along with *Buchnera*.

Some operons and transcription units have their structures differentially-disrupted or -shortened in the *S. symbiotica* endosymbionts, sometimes eliminating their regulation by the absence of leader attenuators, peptides, and/or regulatory proteins. An example of this is the *thr* (threonine) operon, which in *S. marcescens* Db11, holds upstream its respective leader sequence and peptide (*thrL*). However, in SAp both of these elements seem to have eroded, while in SCt, the biosynthetic genes have been physically separated from its attenuator elements into different locations in the genome (Figure 1.10A). Another example lies with the contiguous *argE* and *argCBGH* (arginine)

biosynthetic transcription units: while intact in *S. marcescens* Db11 and SAp, in SCt the *argCBGH* unit has undergone great erosion, retaining only the *argE* and *argH* genes (Figure 1.10B). Additionaly, the Ilv (isoleucine), *hisGDCBHAFI* (histidine), and *leuABCD* (leucine) transcriptional units also present disrupted and shortened structures in the SAp and SCt endosymbionts, missing or having pseudogenized versions of some genes (Figure S.5A-C).

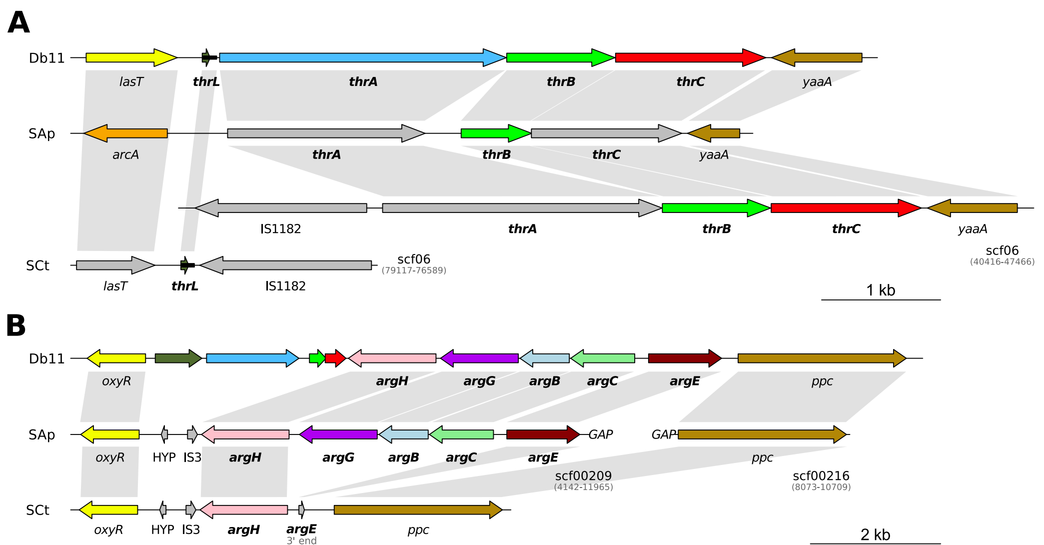


Figure 1.10 Thr and Arg biosynthetic operons and transcription units' degradation in *S. symbiotica*. Diagrams displaying the erosion of biosynthetic operons and transcription units in the facultative SAp and SCt, relative to Db11. **(A)** *thrABC* operon. **(B)** *argHGBC* operon and *argE* transcription unit. Modified from Manzano-Marín and Latorre (2014).

1.3 Conclusion

While SCt is phylogenetically, genomically, and metabolically very similar to SAp, various other lines of evidence point towards these two bacteria having diverged early on. In general, SCt reflects a very similar but slightly more reduced set of metabolic functions as compared to SAp. Additionally, as indicated by the phylogenomic reconstruction, these two

genomes are very closely related. Therefore, the differences in terms of their obligate/facultative status must be mainly related to the different environment they reside in (aphid host) and/or the role their *Buchnera* partners play the nutrient provisioning. Addressing this fundamental question, we have been able to identify a common loss in BCt and BCc, not shared by BAp (from the mono-endosybmiotic system of *Ac. pisum*), of the genes involved in the production of riboflavin (vitamin B_2), which could be the key to the persistent presence of a secondary obligate endosymbiont (mainly *S. symbiotica*) at least in the *Cinara*. While in BAp the genes involved in the pathway for the biosynthesis of riboflavin are all present, the complete inactivation of these has only been found so far in the functionally similar and reduced *Buchnera* endosymbionts from *C.* (*Ci.*) *cedri* and *C.* (*Cu.*) *tujafilina* (Lamelas *et al.*, 2011a; Pérez-Brocal *et al.*, 2006), both belonging to the Lachninae subfamily. Both of these present an association with a *S. symbiotica* endosymbiont, which is able to synthesize this essential compound, and thus making the symbiosis obligatory. Given both the phylogenetic evidence pointing towards an obligate status for *S. symbiotica* in many members from the Lachninae subfamily (Lamelas *et al.*, 2008), and the persistent presence of secondary endosymbionts in most lachnids (Burke *et al.*, 2009; Chen *et al.*, 2015b; Lamelas *et al.*, 2008), we can speculate that this loss happened in the *Buchnera* genome from the Lachninae last common ancestor (**LLCA**). This loss could have been propelled by the constant association with a *Serratia*, or other, facultative endosymbiont, holding an intact set of genes for this route. At some point, *S. symbiotica* would have become mandatorily present, further driving gene losses in both *Buchnera* and *S. symbiotica* partners. In the particular case of *C.* (*Ci.*) *cedri*, *Buchnera* would then have experienced even greater losses, putatively deriving from this constant association, particularly the one rendering it unable to supply the EAA tryptophan (Gosalbes *et al.*, 2008). This could have reinforced the establishment of *S. symbiotica* as a co-obligate intracellular endosymbiont, while at the same time further driven its genome degradation and possibly its intracellularisation into

separate bacteriocytes. On the contrary, neither in BAp nor in BCt this is observed, both preserving all genes necessary to synthesize tryptophan (from phosphoenol pyruvate and D-erythrose 4-phosphate). It is worth mentioning that BCt is the only *Buchnera* so far reported to have a chimeric Leucine/Trpyptohan plasmid (Gil *et al.*, 2006) instead of the typical high-copy number/*trpEG* expansion, postulated to serve for amplification of tryptophan synthesis (Lai *et al.*, 1994). Along with this fact, SCt is the only *S. symbiotica* sequenced so far that presents intact copies of the *trpEG* genes. This displays an evident co-evolution of these two endosymbionts in their aphid host.

In summary, we have found evidence that let us conclude that the genome of SCt finds itself in a stage of accommodation between that of facultative SAp and obligate SCc. This represents an especially interesting case in which two genomically very similar bacteria have different dispensability status, dictated mainly by their obligate partner, *Buchnera*. SCt could thus represent the very first stages of the settling down process from a facultative to a reduced obligate intracellular endosymbiont, not having yet experienced the massive losses leading to a deeply rooted co-obligate endosymbiosis, as witnessed in the symbiotic system of *C.* (*Ci.*) *cedri*.

Chapter 2

Evolutionary convergence of co-obligate endosymbiotic consortia between the aphids *Tuberolachnus salignus* and *Cinara* (*Cinara*) *cedri*

Figure 2.1 ***Tu. salignus*** **apterous female on** ***Salix matsudana*****.** Reproduced from imagenesangel.blogspot.com.es.

2.1 Introduction

Tu. salignus (Figure 2.1) feeds on numerous *Salix* spp. (willows and sallows), being found in the stems and branches of these. It presents an anholocyclic life cycle, with no males ever recorded, and its distribution is almost cosmopolitan wherever willows grow or are planted (Blackman and Eastop, 1994). *Tu. salignus* is the largest of all aphid species, currently known, feeding on *Salix* spp., and it has a distinctive large dark brown tubercle in the centre of the dorsum (Blackman and Eastop, 1994). Through different rearing experiments, it has been determined that the optimal temperature for the development of *Tu. salignus* varies depending on the host plant, being 20 ° C on *Salix alba* (from Turkey) (Özder and Sağlam, 2008) and *Salix viminalis* (from the UK) (Collins and Leather, 2001), and 25 ° C on *Salix babylonica* and *Salix matsudana* (from Turkey) (Özder *et al.*, 2007). This species has been used in classic studies of aphid feeding and nutrition, exploring the uptake of phloem sap (Mittler, 1957), the nitrogen and sugar composition of ingested phloem sap and excreted honeydew (Mittler, 1958a), and in terms of its nitrogen economy (Mittler, 1958b).

Similarly to many Lachninae aphids, *Tu. salignus* has been found to be associated, additionally from *Buchnera*, to a strain of *S. symbiotica* (Lamelas *et al.*, 2008). In this study, the strain in question was recovered in the obligate-like (**OL**) cluster of this species, similarly to the strain recovered from *C.* (*Ci.*) *cedri* (Figure I.7). Currently, the *Buchnera*-*S. symbiotica* di-symbiotic systems of both *C.* (*Ci.*) *cinara* and *C.* (*Cu.*) *tujafilina* (Eulachnini tribe) have been determined to be of co-obligate nature (Lamelas *et al.*, 2011b; Manzano-Marín and Latorre, 2014). However, the genome, cell shape, and tissue tropism of the co-obligate *S. symbiotica* endosymbiont is quite different in both aphids, closely resembling a "recent" or facultative endosymbiont in *C.* (*Cu.*) *tujafilina* and an ancient obligate one in *C.* (*Ci.*) *cedri*. Therefore, it was hypothesised that the endosymbiotic system from *Tu. salignus* was

putatively closer to that of *C.* (*Ci.*) *cedri*. Since this aphid belongs to the distantly related Tuberolachnini tribe, the study of its endosymbiotic consortium was expected to provide a broader evolutionary perspective into the *Buchnera*-*S. symbiotca* relationship within the Lachninae aphids, as well as into the establishment of endosymbiotic microbial consortia.

Based on this, to corroborate the loss-of-riboflavin-biosynthesis hypothesis for the establishment of *S. symbiotica* as a co-obligate endosymbiont in the Lachninae, and to infer the degree of dependence of *Buchnera* upon *S. symbiotica* in *Tu. salignus*, whole genomes for both *Buchnera* and *S. symbiotica* from this aphid species (**BTs** and **STs**) were determined. We have conducted FISH on aphid embryos to locate *S. symbiotica*. Also, we have performed a complete reannotation of both *Buchnera* (INSDC:CP000263) and *S. symbiotica* (INSDC:CP002295) endosymbionts from the aphid *C.* (*Ci.*) *cedri* (BCc and SCc, respectively). Through comparative genomics and metabolic inference, we have described both the putative genetic and the metabolic convergence found between the endosymbiotic systems of *C.* (*Ci.*) *cedri* and *Tu. salignus*. We have inferred presumably convergent losses through the comparison against the endosymbiotic system of *C.* (*Cu.*) *tujafilina*, as well as through comparisons against *Buchnera*-only systems. By means of molecular dating, phylogenetic analyses, and comparative genomics, we propose a timing and scenario for *S. symbiotica*'s establishment and further independent genome shrinkage within the Lachninae following putative convergent losses in the branches leading to *Tu. salignus* and *C.* (*Ci.*) *cedri*. Finally, and most relevantly, we provide further evidence for the aforementioned loss-of-riboflavin-biosynthesis hypothesis and propose that a convergent secondary biosynthetic-capability-loss by *Buchnera* in the aphids *C.* (*Ci.*) *cedri* and *Tu. salignus* could be behind the triggering of the intracellular establishment inside bacteriocytes and further genomic reduction of *S. symbiotica* in these two members of the Lachninae.

2.2 Results & Discussion

2.2.1 Genomic features of *Buchnera* and *S. symbiotica* from *Tu. salignus*

The genomes of both *Buchnera* strain BTs isolate Pazieg (**BTs**) and *S. symbiotica* strain STs isolate Pazieg (**STs**) from *Tu. salignus* have been assembled to one circular chromosome of 421 kbps (375 CDSs) with two plasmids (coding for genes involved in the biosynthesis of tryptophan [plasmid pTrp, 2.5 kb] and leucine [plasmid pLeu, 6.5 kb], respectively) for the former, and one circular chromosome of 650 kbps (492 CDSs) for the latter, with a mean Illumina HiSeq2500 coverage of 1,473x and 182x, respectively (Figure 2.2). The genomes have been deposited in the European Nucleotide Archive under the project numbers PRJEB10875 and PRJEB10876. As for many other *Buchnera*, the pTrp plasmid of BTs was not closed, since it presents long stretches of tandem repeats flanking the *trpE* and *trpG* genes (Gil *et al.*, 2006; Gosalbes *et al.*, 2008; Lai *et al.*, 1994, 1995, 1996; Rouhbakhsh *et al.*, 1996; van Ham *et al.*, 1999), complicating PCR-directed sequencing methods. The two major tandem repeats present in the plasmid pTrp are located upstream the *trpE* gene (18 nucleotides in length) and downstream the *trpG* gene (55 nucleotides in length), both consisting of at least 3 units. The latter repeat contains in its reverse complement strand a putative DnaA-box with the sequence 'TTATCCACA'. This repeat could act as an origin of replication for the plasmid. Regarding the pLeu plasmid, the origin of replication was identified using the **originX** software for bioinformatic prediction together with the identification of the conserved nucleotide motif described by Gil *et al.* (2006). We located this origin between the genes *repA1* and *leuB*. The location of the origin of replication for this plasmid is similar to the one found in the pLeu plasmids of Buchnera from the aphids *Thelaxes suberi* (Thelaxinae Baker, 1920) (van Ham *et al.*, 1997) and *C.* (*Ci.*) *cedri* (Gil *et al.*, 2006). The pLeu plasmid also presents a putative regulatory

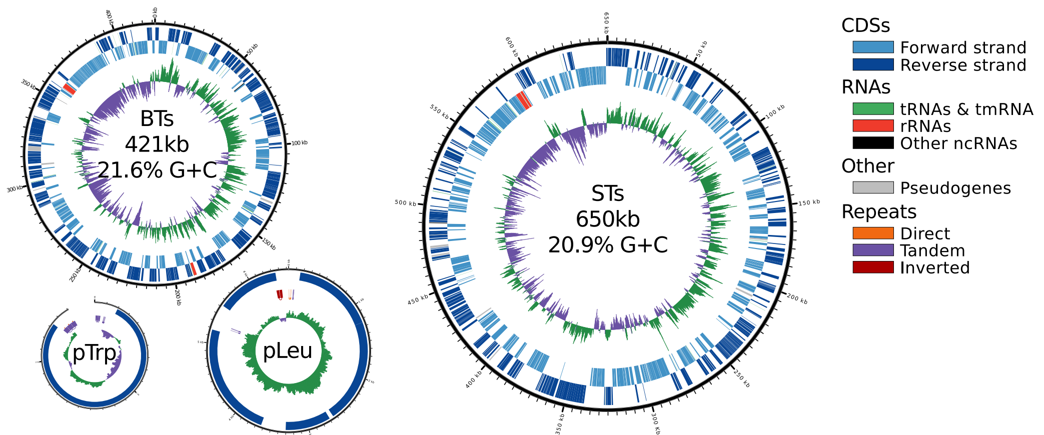


Figure 2.2 Circular representation of *Buchnera* and *S. symbiotica* genomes of *Tu. salignus*. Circular plots of *Buchnera* and *S. symbiotica* genomes displaying, from outermost to innermost ring, the features on the direct strand, the reverse strand, and GC-skew plot. In the case of the pLeu and pTrp plasmids, an additional ring before the GC-skew is used to represent repeat-elements. Reproduced from (Manzano-Marín *et al.*, 2016).

element partially overlapping the 5' of the *repA1* gene consisting of an inverted repeat which could be forming a long secondary structure. This would be consistent with the presence of a small similar inverted repeat found in the pLeu plasmid of BCc. Both chromosomes of BTs and STs retained a putative DnaA-box in the close vicinity of the replication origin predicted by the aforementioned software. This motif was used as the putative origin of replication for both genomes. No transposable elements of any kind have been found in any of the four replicons (both BTs and STs chromosomes and the pLeu and pTrp plasmids), being consistent with general observations of genomically reduced endosymbionts (Moran *et al.*, 2008).

2.2.2 Evolutionary Relationships and Genome reduction in *Buchnera*

Previous studies have shown that the smallest known *Buchnera* genomes were of those harboured by *C.* (*Ci.*) *cedri* and *C.* (*Cu.*)

tujafilina, both members of the Lachninae subfamily (Lamelas *et al.*, 2011a; Pérez-Brocal *et al.*, 2006). Consistent with previous observations for the *Buchnera* from Lachninae representatives, BTs displays a greatly reduced genome, and the second smallest after that of BCc (**table** 2.1). Through a Bayesian phylogenetic reconstruction, we found that, as expected from a previous study by Nováková *et al.* (2013), *Buchnera* from the Lachninae indeed form a monophyletic group, with a sister relationship to the sole representative from the Fordini tribe (Eriosomatinae subfamily), *Buchnera* from *Baizongia pistaciae* (**BBp**) (Figure 2.3A). At the same time, this Fordini/Lachninae clade maintains a sister relationship with the Aphidinae Latreille, 1802 clade, which is further divided into the Aphidini Latreille, 1802 and Macrosiphini Wilson, 1910 tribes. This result, along with an analysis on the clustering of orthologous proteins for currently available *Buchnera* genomes, confirmed previous findings that have proposed a great genetic-repertoire reduction in the *Buchnera* common ancestor of the Lachninae (Figure 2.3B). We found that the largest *Buchnera* pangenome was that of the Aphidinae (597 CDSs), which has at least 86 unique CDSs. Thus, the *Buchnera* common ancestor of the Lachninae would have already undergone a loss of at least 208 CDSs (86 unique to

Subfamily	**Aphidinae**		**Eriosomatinae**[a]	**Lachninae**		
Tribe	**Macrosiphini**	**Aphidini**	**Fordini**	**Tuberolachnini**	**Eulachnini**	
Buchnera strain	BAp	BSg	BBp	BTs	BCt	BCc
Genome size (Kbp)[b]	656[c]	654[c]	618	430[c]	453	425[c]
Chr. G+C content (%)	26.3	25.3	25.3	21.6	23.0	20.1
CDSs (chr.+plasmid[s])	562+7+2[c]	547+7+2[c]	504+3	375+5+2[c]	362+6	360+5+2[c]
tRNAs	32	32	32	31	31	31
rRNAs	3	3	3	3	3	3

[a] Recent work by Zhang and Chen (1999) and Li *et al.* (2014) support the split of the Eriosomatinae into up to three subfamilies (Eriosomatinae, Pemphiginae, and Fordinae), given their paraphyly. *Ba. pistaciae* would then belong to the Eriosomatinae (if divided into 2 subfamilies) or the Fordinae (if divided into 3 subfamilies).

[b] Reported genome sizes include both the plasmids.

[c] pTrp plasmid numbers are reported from the assembled contigs, since they are estimated to contain repeated units of the trpE and trpG genes.

Table 2.1 Comparison of genomic characteristics of *Buchnera* genomes from different aphid subfamilies. The genomic and genetic reduction undergone by the Lachninae's bacteria is evident. BAp= *Buchnera* from *Ac. pisum* strain APS, and BSg= *Buchnera* from *Sc. graminum*

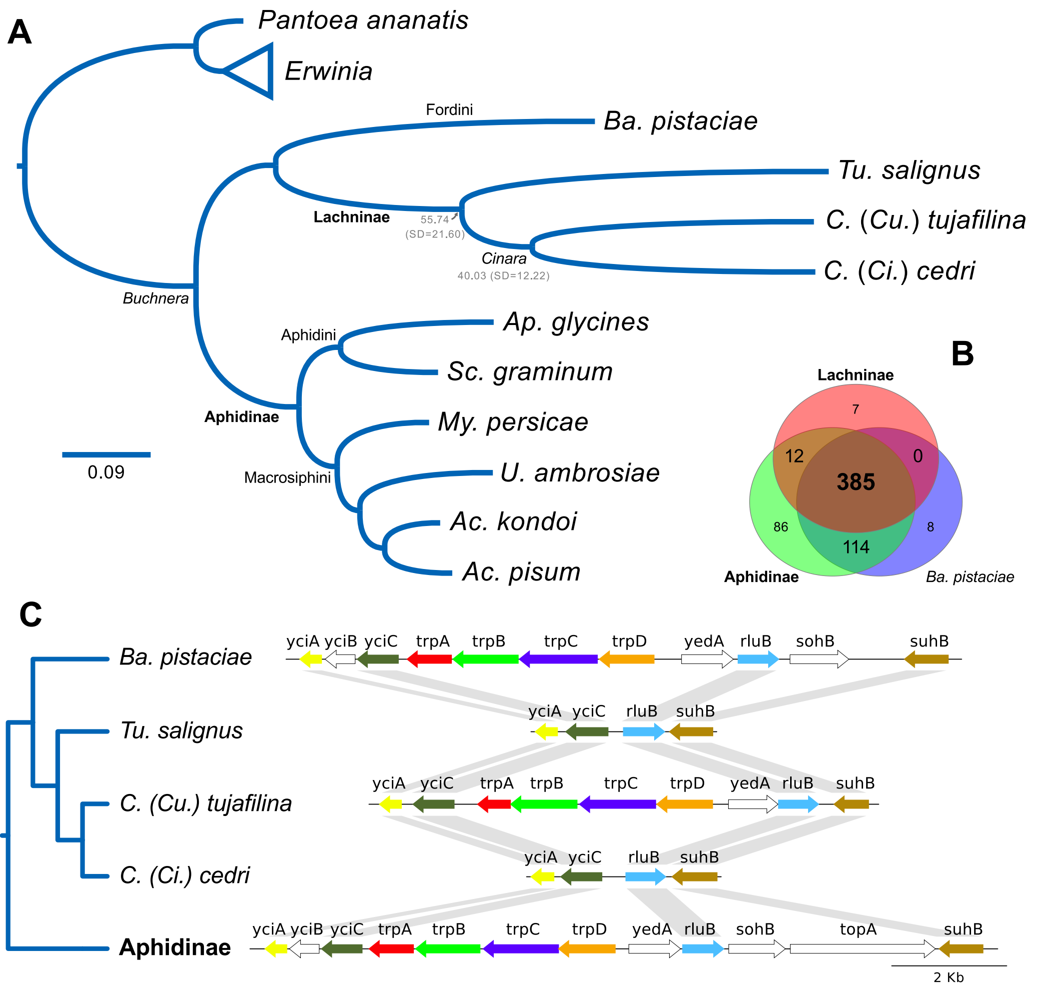


Figure 2.3 ***Buchnera* phylogenomic reconstruction, genetic-repertoire reduction in the Lachninae, and convergent loss of tryptophan biosynthetic genes.** **(A)** Subfamily names are displayed in bold lettering. Mean divergence time estimates for selected nodes are shown in grey under the subfamily or genus clade name. *Erwinia* spp. and *Pantoea ananatis* are the outgroups. Since all posterior probabilities were equal to 1, they have been excluded form the tree **(B)** Venn-like diagram displaying the results of the clustering of orthologous proteins. The Lachninae and Aphidinae are represented as pangenomes reconstructed from currently available sequences. **(C)** Genetic maps of the chromosomal *trp* genes of different *Buchnera* strains. Non-conserved genes are displayed in white, while the conserved ones are displayed in different colouring. The Aphidinae plot was done using *Buchnera* strain APS. It was chosen since it would represent the ancestral gene order and content for the Aphidini-Macrosiphini common ancestor. Modified from (Manzano-Marín *et al.*, 2016).

the Aphidinae, 114 shared by the Aphidinae and BBp, and 8 unique to BBp). This would provide further evidence of for the ancient acquisition of a secondary endosymbiont, whose presence would have relaxed the selection on various *Buchnera* genes, thereby promoting further gene loss. Most importantly, we found a convergent tryptophan biosynthetic gene loss in BTs and BCc, which is not shared by BCt (Figure 2.3C). This gene loss corresponds to the genomic deletion of the genes *trpA*, *trpB*, *trpC*, and *trpD*, which code for the enzymes required to synthesize tryptophan from anthranilate. Besides, a phylogenetic analysis of the aforementioned genes showed no indication of these being acquired from another source in BCt (Figure S.6). This loss would render both BTs and BCc unable to provide tryptophan to their respective hosts, suggesting a dependence on an additional associate.

To date both the gene-loss in *S. symbiotica* and the maximum age for the loss of the tryptophan biosynthetic capabilities of *Buchnera*, we used a molecular clock approach, using a subset of *Buchnera* genes previously inferred to follow this hypothesis (Pérez-Brocal *et al.*, 2006), in combination with the time estimate of 50-70 Mya for the divergence of the Aphidini and Macrosiphini tribes (Clark *et al.*, 1999). We obtained a mean date of 55.74(SD=21.60) (replicate: 52.62[SD=16.50]) Mya for the split of *Tu. salignus* and the *Cinara* lineage, and 40.03(SD=12.22) (replicate: 37.98[SD=12.16]) Mya for the split of *C.* (*Cu.*) *tujafilina* and *C.* (*Ci.*) *cedri* (Figure 2.3A). These estimates are highly congruent with the dates calculated using a combination of aphid nuclear and *Buchnera* gene amplicons by Meseguer *et al.* (2015). The full table of mean divergence times, standard deviations and other statistics, along with the replicate, can be found in Table S.1.

2.2.3 Evolutionary Relationships and Genome Reduction in *S. symbiotica*

As mentioned before, *Buchnera* has established a co-obligate association with a *S. symbiotica* bacterium in the Lachninae aphids *C.* (*Ci.*) *cedri* and *C.* (*Cu.*) *tujafilina* (Lamelas *et al.*, 2011b; Manzano-Marín and Latorre, 2014). This change in lifestyle was probably triggered by an ancient loss of the *Buchnera* riboflavin biosynthetic genes, and in the case of *C.* (*Ci.*) *cedri*, it was followed by a loss of the *trpA*, *trpB*, *trpC*, and *trpD* genes, resulting in tryptophan biosynthetic-role now being split between BCc and SCc (Gosalbes *et al.*, 2008). However, SCc and SCt display very contrasting genomic characteristics, representing two different stages of genome reduction (Lamelas *et al.*, 2011b; Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2012) (Table 2.2). While SCt presents a cellular shape and size as well as a tissue tropism similar to that of the facultative SAp (from the aphid *Ac. pisum*) and other facultative endosymbionts (Fukatsu *et al.*, 2000; Manzano-Marín and Latorre, 2014; Moran *et al.*, 2005), SCc (as *Buchnera*) presents a pleomorphic shape and is confined to specific

Subfamily		**Aphidinae**		**Lachninae**		
Tribe		**Aphidini**	**Macrosiphini**	**Eulachnini**		**Tuberolachnini**
***Serratia* strain**	**Db11**	**SAf**	**SAp**	**SCt**	**SCc**	**STs**
Lifestyle	Free-living	Facultative endosymbiont	Facultative endosymbiont	Co-obligate endosymbiont	Co-obligate endosymbiont	Co-obligate endosymbiont
Genome size (Mbp)	5.11	3.58[a]	2.76[a]	2.49[a]	1.76	0.65
Plasmids	-	Unknown	Unknown	Unknown	Unknown	Unkown
Chromosome G+C content (%)	59.5	52.1	52.1	52.1	29.2	20.7
CDSs	4,709	3,398	2,098	1,601	677	492
Coding density (%)	87.9	78.2	56.8	53.4	38.8	77.2
Pseudogenes	12	126	550	916	100	7
tRNAs	88	74	44	47	36	33
tRNAs	22	22	15	13	3	3
MEs	Few	Many	Many	Many	None	None

[a] Genome size is reported as the total size of the assembled scaffolds for unclosed genomes.

Table 2.2 Genomic characteristics of *Serratia* strains with different lifestyles. Comparison of genomic characteristics of free-living *Serratia* Db11 and *S. symbiotica* genomes from distinct aphid subfamilies. The various degrees of genomic reduction are evident and ordered from left to right. Db11= *S. marcescens* strain Db11. Modified from (Manzano-Marín *et al.*, 2016).

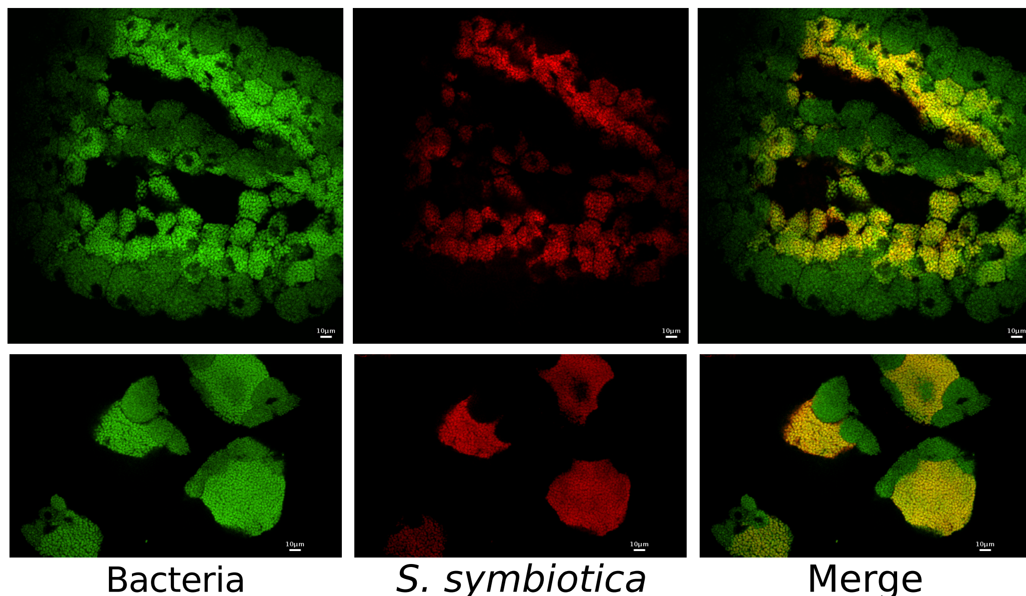


Figure 2.4 ***S. symbiotica*** **localization in bacteriomes of *Tu. salignus* embryos.** Whole-mount FISH of *Tu. salignus* embryos using 16S rRNA-directed probes. **(Left)** Bacterial staining using EUB338 6-FAM-labelled probe. **(Center)** *S. symbiotica* staining using STs DY-405 labelled probe. **(Right)** Merged image showing double-labelling of *S. symbiotica* and single-labelling of *Buchnera* with bacterial probe in bacteriocytes of **(Bottom)** early and **(Top)** later embryos. Modified from (Manzano-Marín *et al.*, 2016).

bacteriocytes (Lamelas *et al.*, 2008). Through FISH using specific probes against STs's 16S rRNA, we determined that similarly to the endosymbiotic system present in *C.* (*Ci.*) *cedri*, STs is present in distinct bacteriocytes, which are surrounded by *Buchnera* bacteriocytes. (Figure 2.4). Strikingly, STs holds an even more degraded genome than that of SCc, and hence represents a further stage in genome shrinkage, having already lost all the "junk" DNA that is still present in the reduced genome of SCc.

Through a Bayesian phylogenomic reconstruction, using single-copy conserved genes, we found *S. symbiotica* to form a monophyletic clade with a sister relation to the *S. marcescens* clade (which includes *Serratia ureilytica* nested within *S. marcescens*) (supplementary Figure S3, available online). However, given the very long branches leading to STs

and SCc, and their apparent sister relation, we also performed a more "sophisticated" phylogenetic reconstruction under Phylobayes using dayhoff6-recoded alignments under the CAT+GTR+G (4 discrete categories) model (in order to minimise the effects of long-branch attraction) (Figure 2.5A). Contrasting the MrBayes reconstruction, the sister relation of STs and SCc is weakly-supported, and the relations among *S. symbiotica* strains become less clear. It is worth mentioning that the "early" co-obligate SCt clusters together with facultative SAf and SAp endosymbionts, indicating a lack of high sequence divergence, as that seen in STs and SCc. *Buchnera* shows general co-cladogenesis with its aphid hosts (Baumann *et al.*, 1995; Funk *et al.*, 2000; Jousselin *et al.*, 2009; van Ham *et al.*, 2003), resulting from a putative massive genome reduction and loss of MEs before the diversification of aphids. Contrary to this, STs and SCc do not display co-cladogenesis, but rather cluster together, although with a low support. This could be explained by the highly accelerated branches leading to this, possibly causing a long-branch attraction artefact, as is seen in other highly accelerated endosymbiotic lineages (Husník *et al.*, 2011). The contrary case would be exemplified by SCt, whose lack of niche-change (compared to the facultative SAp) (Manzano-Marín and Latorre, 2014) would be reflected in the absence of strong evolutionary forces acting on its gene sequences, manifested in its short branch.

To infer whether both STs and SCc have indeed undergone an early common genomic reduction, as seen in *Buchnera*, or if both have undergone independent events of genome shrinkage before the loss of mobile genetic elements, we performed a minimal number of rearrangements phylogeny. If both STs and SCc would have undergone an early drastic common genome reduction and loss of MEs (as is the case of *Buchnera*), it would be expected that they both show a remarkable degree of synteny and phylogenetic congruency. On the contrary, if STs and SCc would have undergone independent events of genome reduction and loss of MEs, we would expect to see much

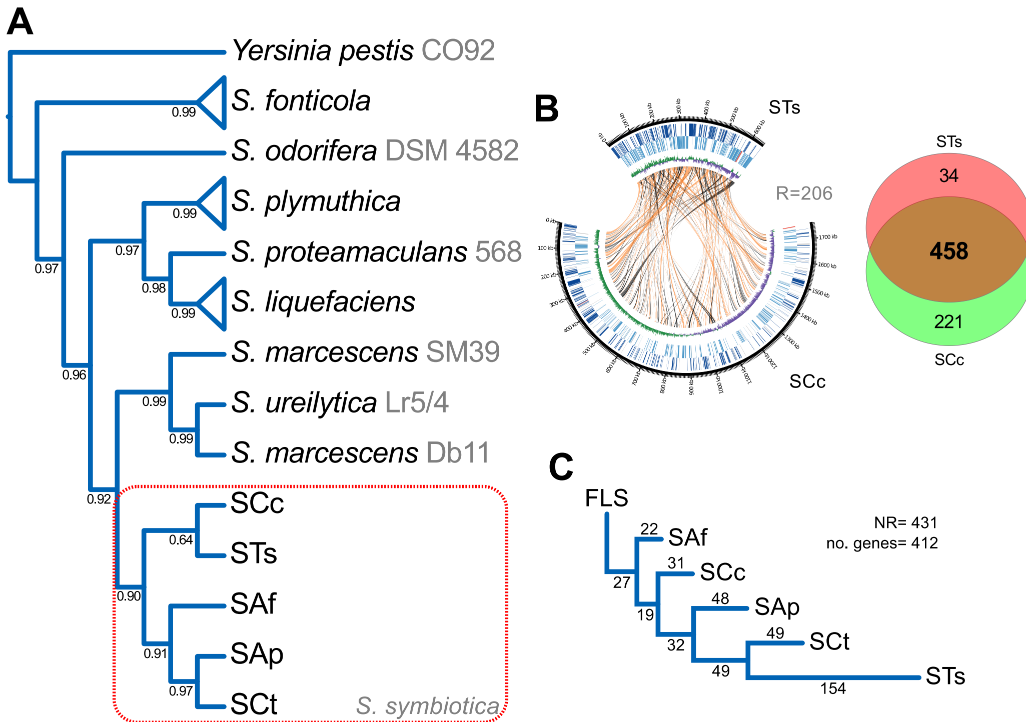


Figure 2.5 *Serratia* phylogenomic reconstruction, gene-order rearrangements, and shared genetic repertoire between STs and SCc. **(A)** Bayesian phylogenomic reconstruction of different Serratia strains using *Yersinia pestis* strain CO92 as an outgroup. Strain names are in grey shown after species names in. Asteriks at nodes stand for a posterior probability equal to 1 **(B)** On the left, circular plot displaying the chromosomes of STs and SCc. From outermost to innermost ring, the features on the direct strand, the reverse strand, and GC-skew plot. Lines going from one genome to another represent orthologous genes in direct (orange) or reverse (black) orientation. On the right, Venn-like diagram displaying the shared (core) and unshared genes between STs and SCc. **(C)** Minimum number of rearrangement phylogeny for *Serratia* strains as calculated by MGR. Numbers on top of branches indicate the number of inferred rearrangements undergone in each branch. NR= Number of rearrangements, FLS= free-living *Serratia*. Modified from (Manzano-Marín *et al.*, 2016).

genome reordering between the two, as seen among SCc, SCt, and SAp (Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2012). We found that at least 206 rearrangements have occurred between STs and SCc (Figure 2.5B), supporting the hypothesis of early independent events genome reduction. Through a minimum rearrangement phylogeny

using a subset of these genes (shared among *S. marcescens* Db11 and currently available *S. symbiotica* strains as single-copy protein-coding genes), we found that that the rearrangements undergone in STs is highest for all *S. symbiotica* strains (Figure 2.5C). In contrast, SAf displays the least rearrangements from the ancestral gene order, which is conserved among free-living Serratia strains (Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2012). Disagreeing with the concatenated gene phylogeny, SCc is, rearrangement-wise, more closely related to facultative *S. symbiotica* strains. Also, STs displays the largest number of rearrangements when compared to all *S. symbiotica* strains and *S. marcescens* Db11. This extreme lack of genome-wide synteny greatly contrasts *Buchnera* and other endosymbionts that underwent a rapid genome reduction before the diversification of their hosts (Degnan *et al.*, 2004; Patiño-Navarrete *et al.*, 2013; Tamas *et al.*, 2002; van Ham *et al.*, 2003), and thus strongly support the early divergence of SCc and STs.

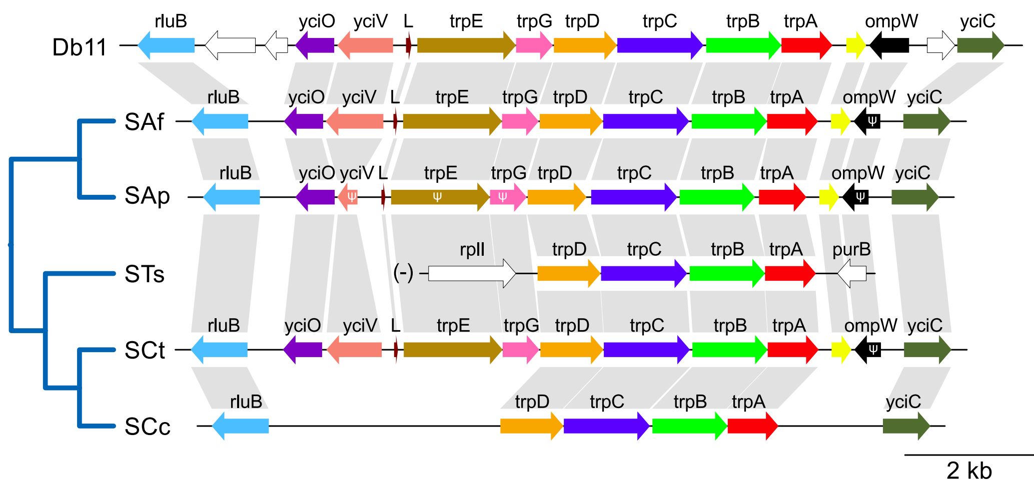


Figure 2.6 Genetic maps of the trp operon in different *Serratia* genomes. Dendrogram on the left displays phylogenetic relationships of the aphid hosts associated to each *S. symbiotica* strain. Non-conserved genes among all strains are displayed in white, while the conserved ones are shown in different colouring. The trp locus of STs is marked with a "(-)" to indicate it has been reverse-complemented (from its original orientation in the genome). L= trp operon leader peptide. Modified from (Manzano-Marín *et al.*, 2016).

Given the observed potential auxotrophy for tryptophan in *Buchnera*, we analysed the tryptophan biosynthetic loci in the different *S. symbiotica* strains. Not surprisingly, we observed the same pseudogenisation of the *trpE* and *trpG* genes seen in SCc in STs (Figure 2.6). In the case of STs, an inversion and translocation of the trp locus has occurred. On the contrary, in SCc, this locus has retained the flanking genes and large "junk" DNA regions putatively corresponding to the degenerated gene-sequences, suggesting a closer relationship to SCt. These *trpE* and *trpG*-losses have led to a metabolic co-dependence of *Buchnera* and the aphid with their *S. symbiotica* partner for the production of tryptophan, an essential compound supplied by the *Buchnera* in mono-endosymbiotic systems in other aphids (Douglas and Prosser, 1992). Interestingly, this pathway is also found split in a similar fashion in some *Carsonella*-secondary di-symbiotic systems found in psyllids (Sloan and Moran, 2012), which could indicate that anthranilate is a metabolite able to pass through the different membranes of the two endosymbionts and hosts. This could be reflected in the apparent selection of this pathway to split up in the same fashion, when splitting actually occurs. However, this pathway is not found split in the same fashion, or even at all, in other di-symbiotic consortia such as the *Sulcia*-secondary (from sharpshooters, leafhopers, and cicadas) (McCutcheon and Moran, 2010), the *Carsonella-Profftella* (from a psyllid) (Nakabachi *et al.*, 2013), and the *Tremblaya*-secondary (from mealybugs) ones (Husnik and McCutcheon, 2016). This would point towards this feature being putatively somewhat specific to the *Buchnera*-*S. symbiotica* (and the two *Carsonella*-secondary) endosymbiotic consortia. In the case of most *Buchnera* strains, the *trpE* and *trpG* genes are present in a multi-copy plasmid, and in some cases presenting multiple copies of the two genes within the same replicon (Lai *et al.*, 1994, 1996; Rouhbakhsh *et al.*, 1996; van Ham *et al.*, 1999). This amplification phenomenon has been explained as a mechanism for *Buchnera* to overproduce anthranilate (Lai *et al.*, 1994). Within *Buchnera* strains with a *trpEG* plasmid, BCt is unique, in that it has undergone a fusion of this into the

leucine plasmid, putatively losing this amplification (Gil *et al.*, 2006). This would be consistent with SCt still preserving a full gene repertoire for the biosynthesis of tryptophan. Therefore, it could be speculated that when the secondary symbiont adapts to a certain point to the presence of *Buchnera* (a "better" anthranilate producer), a strong relaxation of selection in the *trpEG* genes would occur given the high availability of this compound. However, the fact of *Buchnera* being a "better" anthranilate producer remains to be experimentally tested.

2.2.4 Metabolic capabilities of BTs and STs endosymbionts

In order to gain insight into the metabolic capabilities of the BTs-STs endosymbiotic consortium, we conducted a metabolic reconstruction using Pathway tools (Karp *et al.*, 2015) (Figure 2.7 and 2.8). Regarding the biosynthesis of EAA, the main role undertaken by *Buchnera* in aphids, BTs preserves the capability of synthesizing the same EAAs (and precursors) as BAp (Hansen and Moran, 2011). However, as it is the case for *C.* (*Ci.*) *cedri*'s endosymbiotic system, the capacity of *Buchnera* to *de novo* synthesise tryptophan has been partially lost, and this role would now be shared by both members of *Tu. salignus*' endosymbiotic consortium (BTs and STs). Also, differently from BAp, but similarly to both BCc and BCt, the pathway leading to the biosynthesis of arginine has been truncated, and would depend solely on the import of ornithine and glutamine from the host. Regarding glycolysis, we observed an interesting convergence of BTs with BCt, which unlike BCc, have lost the *pgi* gene. This gene codes for a phosphoglucose isomerase, therefore, BTs would either use glucose-6P (and force glycolytic flux through the pentose phosphate pathway, as has been proposed for an *Escherichia coli pgi* knock-out mutant (Canonaco *et al.*, 2001; Charusanti *et al.*, 2010; Hua *et al.*, 2003) or fructose-6P as substrate for glycolysis. This last is probably the case of BCt (Lamelas

et al., 2011a), which has also lost the *pgl* and *zwf* genes of the oxidative pentose pathway. Finally, as BCc and BCt, BTs has lost the capacity to *de novo* synthesise the essential vitamin riboflavin (role now undertaken by STs) and has lost the genes *bioB*, *bioD*, and *bioA*, involved in the biosynthesis of biotin. This last vitamin could be putatively synthesised by the BTs-STs consortium, where BTs would synthesise pimeloyl-[acp] from an unknown precursor, this would in turn be converted to KAPA by the action of an unknown enzyme, and finally, KAPA would be imported into STs to synthesise biotin.

With regard to STs (Figure 2.8), it preserves a glucose PTS permease and it is able to start glycolysis from glucose-6P as well as to go through the pentose phosphate pathway to produce PRPP, ribulose-5P, and erythrose-4P. Regarding the biosynthesis of vitamins, as SCc and SCt, it would be able to synthesize riboflavin (*de novo*) and tetrahydrofolate (from dihydropterate). STs has completely lost the ability to synthesise thiamin and also the associated ABC transporter (coded by the genes *thiB*, *thiP*, and *thiQ*). However, it keeps the genes *thiK* and *thiL*, which catalyse the conversion of thiamin into TPP. STs preserves the capacity to synthesise peptidoglycan from UDP-N-acetyl-D-glucosamine and an almost complete pathway to lipid A-core. This last pathway is missing some steps, whose responsible enzymes are not yet identified in *Serratia*. Concerning the biosynthesis of nucleotides, it could import inosine, uridine, and cytidine; and synthesize all but dTTP, CTP, and dCTP. These last three would require the action of non-specific kinases yet to be identified, since STs has lost the gene *ndk* (dTTP from dTDP, CTP from CDP, and dTTP from dTDP) and *pyrG* (CTP from UTP). *ndk* has been shown to be non-essential for growth in *E. coli*, but mutants generated high levels of substitutions and frameshifts (Lu *et al.*, 1995; Miller *et al.*, 2002). In addition, we propose that the biosynthesis of coenzyme A might be occurring through the uptake of pantethine. This compound has recently been described to be the most advanced precursor that an *E. coli coaBC*-deletion mutant strain can utilize to

Figure 2.7 Metabolic reconstruction of BTs. Intact pathways are represented with solid black lines, and unclear ones (missing a specific gene or having it pseudogenised by a frameshift) with solid grey lines. Importers are displayed using green ovals, while exporters and exporters/importers are displayed using blue ovals. The name inside each oval states the family/superfamily they belong to (following TCDB's classification, otherwise the protein name is used. Essential and non-essential amino acids are shown in red and purple lettering, respectively. Cofactors and vitamins are shown in blue. Blurred compounds represent those for which biosynthesis or import cannot be accounted for based on genomic data. Blurred transporters represent those for which a part of the transporter is missing, therefore recently pseudogenised. Modified from (Manzano-Marín *et al.*, 2016).

bypass a CoaBC deficiency, in the absence of the pantothenate permease-encoding gene *panF* (Balibar *et al.*, 2011). This explanation would be consistent with the pseudogenisation of the *coaBC* gene in both STs and SCc, as well as the complete absence of the *panF* gene in both of these strains. Regarding the biosynthesis of amino acids, similarly to SCc, it has lost the capability of synthesising almost all EAAs, with the notable retention of the genes *trpA*, *trpB*, *trpC*, and *trpD*, which identically to SCc, would enable STs to perform the last steps in the

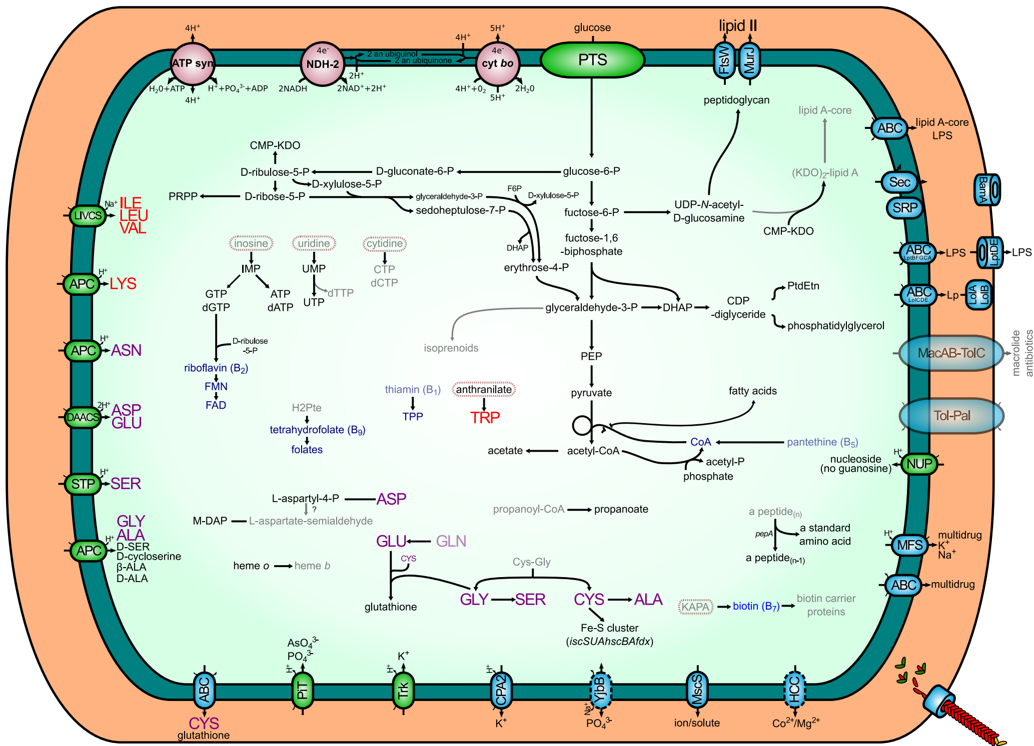


Figure 2.8 Metabolic reconstruction of STs. Intact pathways are represented with solid black lines, and unclear ones (missing a specific gene or having it pseudogenised by a frameshift) with solid grey lines. Importers are displayed using green ovals, while exporters and exporters/importers are displayed using blue ovals. The name inside each oval states the family/superfamily they belong to (following TCDB's classification, otherwise the protein name is used. Essential and non-essential amino acids are shown in red and purple lettering, respectively. Cofactors and vitamins are shown in blue. Blurred compounds represent those for which biosynthesis or import cannot be accounted for based on genomic data. Blurred transporters represent those for which a part of the transporter is missing, therefore recently pseudogenised. Modified from (Manzano-Marín *et al.*, 2016).

biosynthesis of tryptophan following the import of anthranilate, synthesised by BTs. It preserves specific importers for isoleucine, leucine, valine, lysine, asparagine, aspartate, glutamate, serine, glycine, and alanine, and could synthesise various NEAAs given the required inputs.

Summarising, we have corroborated the complete loss of the riboflavin biosynthetic pathway in *Buchnera* from all currently-sequenced

Lachninae, role now overtaken by *S. symbiotica*. We have also identified a second Lachninae-specific loss in *Buchnera* of the biotin biosynthetic genes *bioA*, *bioD*, and *bioB*, which have been interestingly retained in their *S. symbiotica* co-obligate partners. This pathway is not entirely complete in currently sequenced *Buchnera* strains from the Aphidinae subfamily, since most have lost the *bioC*, *bioH*, and *bioF* genes, but have preserved, over evolutionary time, all other genes in the pathway (Figure 2.9). However, in BBp, the sole *Buchnera* from the Fordini sequenced to date, the biotin pathway remains intact. Therefore, it can be inferred that this pathway could function in a yet-unknown way in most *Buchnera* and would represent an additional metabolic loss in the *Buchnera* common ancestor of the Lachninae. For the pea aphid *Ac. pisum*, it has been proven that riboflavin production by Buchnera is indeed essential for correct development of the aphid (Nakabachi and Ishikawa, 1999), but in the case of biotin, to our knowledge, no experimental evidence has yet

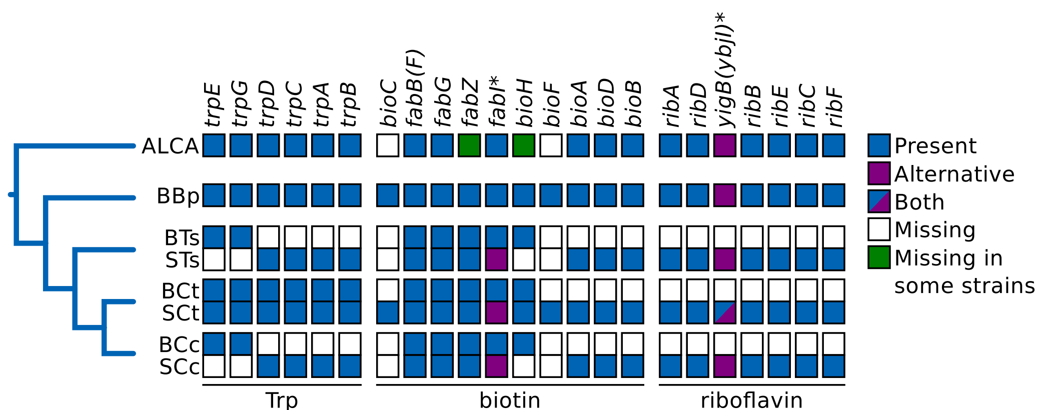


Figure 2.9 Proposed metabolic complementation found in the obligate di-symbiotic systems of different aphids. Diagram representing the proposed metabolic complementations in the biosynthesis of tryptophan and biotin in the currently available endosymbiotic systems of aphids as well as the riboflavin biosynthesis takeover/rescue by *S. symbiotica* in the endosymbiotic systems of the Lachninae aphids. The gene names are used as column names, while the abbreviations for the different endosymbiotic bacteria are used as row names, except for ALCA= Aphididae last common ancestor. Asterisks after gene names indicate a putative alternative enzyme could be performing the enzymatic function, a trans-2-enoyl-CoA reductase for the biotin and *yigL* for the riboflavin biosynthetic routes. Reproduced from (Manzano-Marín *et al.*, 2016).

been put forward to prove the supply or production of this vitamin by the symbiont. Regarding the host need for this last vitamin, to our knowledge, only experiments on artificial diets have been performed in two Macrosiphini aphids, proving these aphids can develop, albeit showing growth-defects (see I.3).

2.2.5 Metabolic and transporter-function convergence between *Tu. salignus* and *C.* (*Ci.*) *cedri*'s co-obligate endosymbiotic systems

As shown in Figure 2.5B, the genetic content of STs is mainly a subset of SCc's, although it has retained 34 genes that are not present in SCc, neither as CDSs nor as pseudogenes. It is important to remark that the genetic repertoire of both STs and SCc is merely a subset of that of *S. marcescens* and other *S. symbiotica* strains (Figure S.7), indicating that any retained gene could be inferred to be in the last common ancestor of *S. symbiotica*. Interestingly, STs's and SCc's corresponding *Buchnera*

Genes			
STs	**SCc**	**Function**	**EC number**
lysC	*thrA*	Asparate kinase	2.7.2.4
rdgB	*mazG*[a]	dITP/XTP pyrophosphatase	3.6.1.19
mdfA,mdlAB	*mdtK,emrAB-tolC*	Multidrug efflux	-
kefBG	*cvrA*	Potassium/H$^+$ antiporter	-
mltC	*mltA,mltD*	Membrane-bound lytic murein transglycosylase	-
mutY	*mutM*	Rescue 8-oxo-guanine:cytosine pre- (MutM) or post-replication (MutY)	-
iscUAhscBAfdx	*sufABCDSE*	iron-sulfur cluster biosynthesis	-
pgpB	*pgpA*	Phosphatidylglycerophosphatase	3.1.3.27
STsPAZIEG_0413	SCc_0167	fimbrial adhesin	-

[a] Pseudogene.

Table 2.3 Metabolic and transporter-functional convergence between STs and SCc. Table displaying the genes differentially retained between STs and SCc but that confer similar functions to their corresponding bacterium. Modified from Manzano-Marín *et al.* (2016)

partners are also very similar in terms of genetic content (Figure S.8). In terms of differential genome reduction and preservation of equivalent enzymes or transporters, convergent patterns can be noticed between STs and SCc. For example, within the lysine biosynthetic pathway, which is intact in SCc but apparently recently impaired in STs given the loss of the *asd* gene, there exists a most interesting differential degradation pattern. While STs has only retained the *lysC* gene, which codes for one of three isozymes in *E. coli* able to catalyse the aspartate kinase reaction, SCc has retained a truncated version of another one of these three isozymes, coded by the bifunctional *thrA* gene. This truncated gene could preserve its aspartate kinase activity, as has been tested *in vivo* with a similarly truncated version of the ThrA protein, carried in a plasmid using *E. coli* as a host (Omori and Komatsubara, 1993). Additionally, we analysed the 34 genes that are present in STs but not in SCc, and found that 16 of them actually have equivalent functions to genes present in SCc but that are missing in STs (Table 2.3). This indicates that common functions have been maintained, despite the two independent events of genome reduction. Lastly, we found a convergent retention of an apparently complete type-1 fimbriae-like system between STs and SCc. These fimbriae-like systems would be composed of four subunits: i) A major fimbrin (STSPAZIEG_0410 and SCc_0576), ii) a fimbriae assembly chaperone (STSPAZIEG_0411 and SCc_0575), iii) an outer membrane usher protein (STSPAZIEG_0412 and SCc_0574), and iv) a fimbrial adhesin (STSPAZIEG_0413 and SCc_0167). However, while the first three components seem to be orthologous, the fimbrial adhesins might be of different evolutionary origin, or have already diverged from each other beyond the limits for orthology detection. These types of fimbriae are used by some pathogenic bacteria for infection and surface attachment (reviewed in Korea *et al.*, 2011; Scheller and Cotter, 2015), and could thus reflect their need for this mechanism in order to maintain their obligate intracellular lifestyles inside bacteriocytes. In this line, it is important to mention that neither the facultative SAp nor the “early” co-obligate SCt, which are not found obligatorily inside

bacteriocytes, preserve intact fimbriae-related systems. Therefore, not only have SCc and STs evolved to preserve a large common genetic repertoire, but also part of the observed differential genome reduction undergone by this endosymbiont in both aphids can be attributed to the retention of alternative enzymes/transporters to perform similar or equivalent functions.

2.3 Conclusion

As has been shown, through comparative genomics and metabolic inference, we have found a high level of similarity between the endosymbiotic systems of *C.* (*Ci.*) *cedri* and *Tu. salignus*. We have found functional and genome size reduction convergence, as well as putative metabolic convergence in the complementation established between the endosymbiotic partners of both systems. However, STs holds a greatly reduced genome, even when compared to SCc, and would thus represent a later stage of genome reduction, where the leftover "junk" DNA and pseudogenes are purged from the genome. We have been able to determine that STs and SCc indeed diverged early in their establishment as obligate intracellular endosymbionts. Thus, two equally parsimonious scenarios (in terms of number of steps) can be proposed in light of the results: the "convergent" (Figure 2.10A) and the "non-convergent" one (Figure 2.10B), in terms of the biotin and tryptophan gene-losses. Although we cannot entirely discard replacement by a different *S. symbiotica* strain in *C.* (*Cu.*) *tujafilina*, taking into account the high incidence of obligate-like *S. symbiotica* strains within the Lachninae (across all 5 tribes) (Burke *et al.*, 2009; Chen *et al.*, 2015b; Jousselin *et al.*, 2016; Lamelas *et al.*, 2008), the "fragile" phylogenetic sister relation of STs and SCc (Figure 2.5A), the *S. symbiotica* rearrangement phylogeny showing SCc being more closely related to the facultative strains (Figure 2.5C), and the existence of *S. symbiotica* strains with both intact and pseudogenised/absent copies of

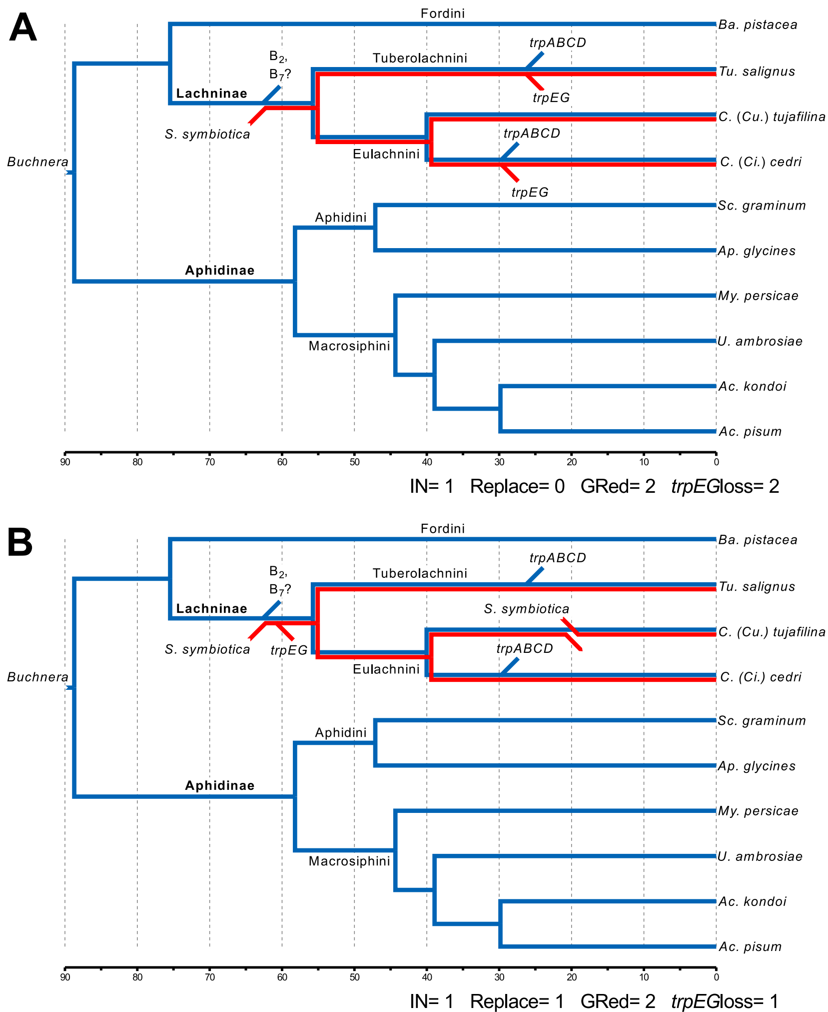


Figure 2.10 Proposed evolutionary scenarios for the establishment of *S. symbiotica* as a co-obligate endosymbiont in currently sequenced Lachninae aphids. Dendrogram representing the evolutionary history of *Buchnera* within aphids and the proposed **(A)** convergent and **(B)** non-convergent scenarios for the establishment of *S. symbiotica* as a co-obligate endosymbiont in the Lachninae. Blue and red lines represent *Buchnera* and *S. symbiotica*, respectively. The blue diagonal lines represent losses in *Buchnera*, while the red diagonal lines represent those in *S. symbiotica*. Names on top and below of branches indicate the aphid familiar name (bold) and tribal name. Scale bar at the bottom represents timing in Mya for the branching of the different aphids. IN= infection events, Replace= symbiont replacement events, GRed= Genome reduction events, *trpEG*loss= *trpEG* gene-loss events. Modified from (Manzano-Marín *et al.*, 2016).

the *trpEG* genes (Figure 2.6), we believe the most likely explanation to be a single infection some 55.74(SD=21.60) (replicate: 52.62[SD=16.50]) Mya in the *Tu. salignus*/Eulachnini common ancestor, followed by divergence (Figure 2.10A). The alternative non-convergent scenario (Figure 2.10B) would require at least two independent infections of *S. symbiotica* strains: one in the branch leading to *Tu. salignus*/Eulachnini, and one in the branch leading to the *C.* (*Cu.*) *tujafilina* one. Nevertheless, with no extra information from more *S. symbiotica* genomes from Lachninae aphids, the non-convergent scenario for the loss of biotin or tryptophan genes cannot be ruled out.

Through the comparison of the di-symbiotic consortia from Lachninae aphids *vs.* mono-symbiotic ones from other aphids, we have corroborated the riboflavin biosynthesis takover/rescue in at least the common ancestor of *Tu. salignus* and *Cinara*. Finally, we speculate that the secondary loss of the tryptophan biosynthetic capability by *Buchnera* and its putative complementation with *S. symbiotica* could have occurred independently in the lineages leading to *Tu. salignus* and *C.* (*Ci.*) *cedri*. This loss could be behind the establishment (obligatorily inside bacteriocytes) of *S. symbiotica* in, at least, these two Lachninae aphids. Since that moment, the genome reduction could have been accelerated due to drift, as a consequence of the bottlenecks undergone by these symbionts in each generation.

Chapter 3

Identity and localisation of secondary endosymbionts in aphids from the Lachninae subfamily

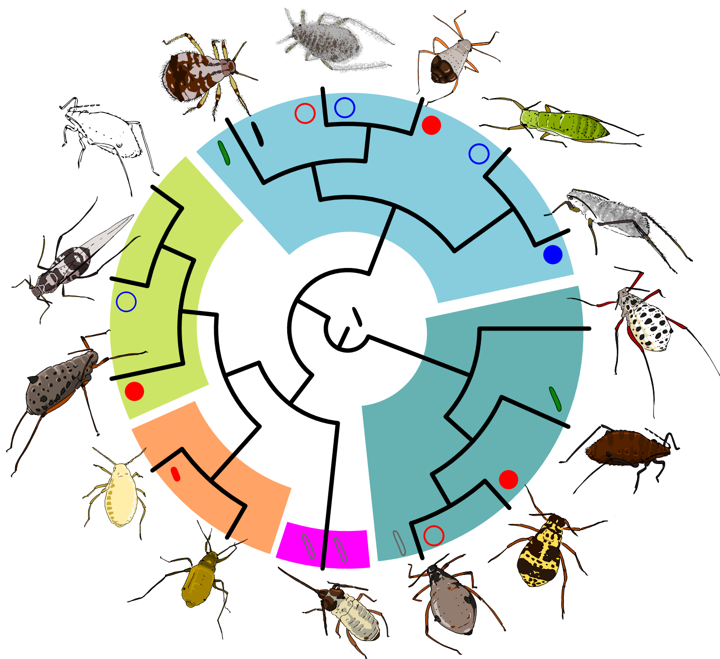


Figure 3.1 The Lachninae subfamily and its secondary endosymbionts. Dendrogram showing the phylogenetic relatioships of the different tribes of the Lachninae subfamily based on Chen *et al.* (2015a); Meseguer *et al.* (2015). Cartoons representing the different secondary endosymbionts harboured by aphids within this subfamily are shown in different colouring. Aphid cartoons by Jorge Mariano Collantes Alegre.

3.1 Introduction

As mentioned earlier, the *Buchnera* symbiosis seems to be universal in aphids (Buchner, 1953), with the notable exception of members belonging to the monophyletic Cerataphidini Baker, 1920 tribe, in which *Buchnera* has been replaced by an extracellular yeast-like symbiont (Fukatsu and Ishikawa, 1992; Fukatsu *et al.*, 1994). *Buchnera* cells have a spherical shape (Michalik *et al.*, 2014) (Figure 3.2A), and inhabit the cytoplasm of bacteriocytes (specialised cells evolved to house the endosymbiont), which make up a distinct organ-like structure called the bacteriome (Buchner, 1953; Fukatsu *et al.*, 1998) (Figure 3.2B). The onset of *Buchnera*'s symbiotic relationship with aphids dates back to 80-150 Mya (von Dohlen and Moran, 2000). *Buchnera*, as other "ancient"

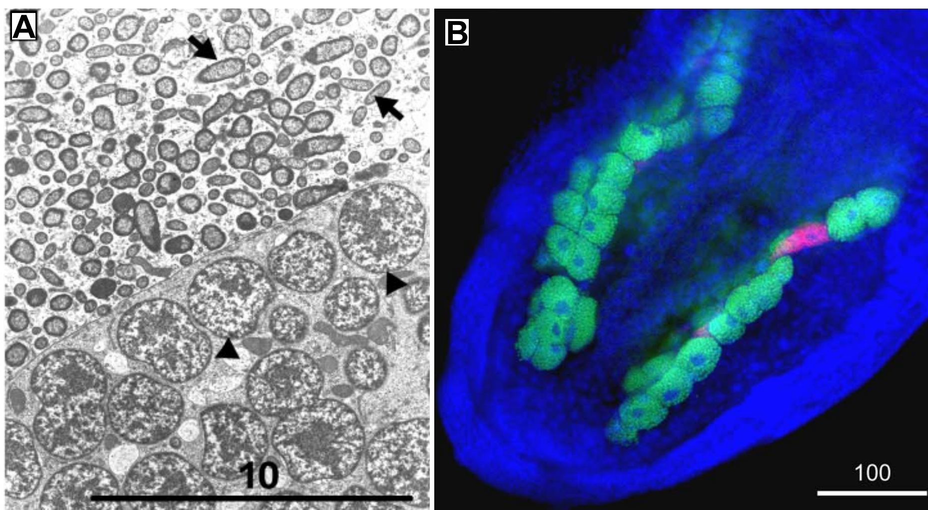


Figure 3.2 Morphology and localisation of *Buchnera* in *Ac. pisum* bacteriomes. (A) Electron micrograph of representatives of *Ca.* Hamiltonella defensa. Bar= 10 µm. Arrows and arrowheads indicate secondary symbionts and *Buchnera*, respectively. Reproduced from Moran *et al.* (2005). **(B)** Position of symbiont-containing bacteriocytes within the abdomens revealed by fluorescent *in situ* hybridisation using diagnostic probes. Blue is a general DNA stain, highlighting aphid nuclei, red indicates *Regiella*, and green indicates *Buchnera*. Bar= 100 µm. Reproduced from The International pea aphid genomics consortium (2010).

obligate endosymbionts, underwent a rapid genome erosion before the diversification of aphids, and since then both partners have been co-evolving. This has been evidenced through phylogenetic reconstructions using *Buchnera* DNA or amino acid sequences, which parallel their aphid hosts' evolutionary relations (Jousselin *et al.*, 2009; Liu *et al.*, 2013; Munson *et al.*, 1991b). Besides *Buchnera*, aphids can also harbour secondary endosymbionts (additional to the primary), these being of facultative or obligate nature in some lineages. To date, various secondary facultative bacterial endosymbionts have been identified, primarily in pea aphids (Degnan *et al.*, 2009a; Fukatsu *et al.*, 2000; Sakurai *et al.*, 2005; Tsuchida *et al.*, 2014). As mentioned earlier, these secondary symbionts have a very different tissue tropism than *Buchnera*, as they can be present in separate bacteriocytes (called secondary bacteriocytes), co-infecting the primary endosymbiont's bacteriocytes, located in sheath cells (at the periphery of the bacteriome and found closely associated to bacteriocytes), and/or free in the haemocoel (Figure S.2).

One early example of secondary co-obligate symbionts from aphids comes from the genus *Yamatocallis* (Drepanosiphinae Herrich-Schaeffer, 1857 subfamily), where 100% of the analysed individuals belonging to both *Yamatocallis hirayamae* and *Yamatocallis tokyoensis* species were found to harbour a secondary bacterium in distinct bacteriocytes (designated **YSMS** for ***Y**amatocallis* **s**econdary **m**ycetocyte **s**ymbiont), pointing towards its obligate status (Fukatsu, 2001) (Figure 3.3). In this case, the establishment of YSMS seems to have occurred in the common ancestor of *Yamatocallis* aphids. Evidence for this comes from both 16S rRNA gene sequences forming a distinct and isolated phylogenetic clade, and from negative diagnostic PCR results (using primers designed to specifically detect YSMS) from other Drepanosiphinae aphids.

In this respect, the subfamily Lachninae Herrich-Schaeffer, 1854 of aphids is peculiar, in that all members analysed thus far by microscopy techniques have been found to house secondary endosymbionts.

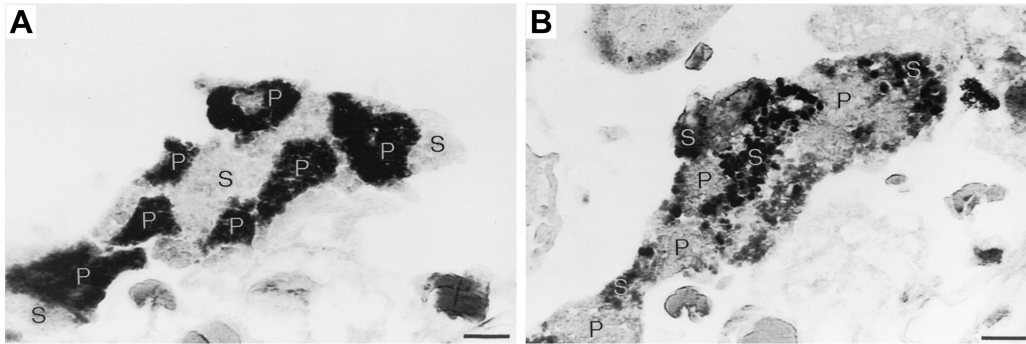


Figure 3.3 *in situ* hybridisation of the endosymbiotic bacteria in *Y. tokyoensis*. **(A)** Maternal bacteriome probed with DIG-KMGTP, which specifically targets the *Buchnera*. **(B)** Maternal bacteriome probed with DIG-KMGTS, which specifically targets the YSMS. YSMS cells in the cytoplasm between primary bacteriocytes are visualised. The sections shown in panels A and B were adjacent tissue sections. Bars = 20 µm. P= *Buchnera*, S= YSMS. Reproduced from Fukatsu (2001).

Secondary symbionts of Lachninae aphids differ in tissue tropism and cell shape (Buchner, 1953; Fukatsu and Ishikawa, 1998; Fukatsu *et al.*, 1998; Michalik *et al.*, 2014; Pyka-Fościak and Szklarzewicz, 2008) (Figure S.9), as well as in phylogenetic origin (Burke *et al.*, 2009; Lamelas *et al.*, 2008; Russell *et al.*, 2003). Although different bacterial taxa have been found associated to Lachninae aphids, many species of this subfamily have been systematically found associated with members of the bacterial genus *Serratia*, mainly *S. symbiotica* (Burke *et al.*, 2009; Fukatsu *et al.*, 1998; Jousselin *et al.*, 2016; Lamelas *et al.*, 2008). Particularly, most *Cinara* species have been consistently found to house *S. symbiotica* strains, which form two phylogenetically distinct clusters (FL and OL) (Burke *et al.*, 2009; Lamelas *et al.*, 2008) (see I.3.3). Whole genome sequencing and metabolic reconstruction of the *Buchnera*-*S. symbiotica* bacterial consortia of three Lachninae aphids has revealed that *S. symbiotica* strains, belonging to both FL and OL, have indeed established co-obligate associations along with *Buchnera* in these hosts (Gosalbes *et al.*, 2008; Lamelas *et al.*, 2011b; Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2016). Through comparative genomics of these di- (*Buchnera*-*S. symbiotica*) vs. mono-endosymbiotic

(*Buchnera*-only) systems, we have previously postulated that the establishment of *S. symbiotica* as a co-obligate endosymbiont in the LLCA was facilitated by a putative ancient pseudogenisation of the riboflavin biosynthetic pathway in *Buchnera*, and the complementation of this loss-of-function by *S. symbiotica* (Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2016).

Given the overwhelming evidence pointing towards a dependency on secondary endosymbionts within the Lachninae, we sought to further understand the evolutionary succession of establishments, replacements, and internalisations into distinct bacteriocytes of secondary endosymbionts within aphids of this subfamily. For this purpose, we have identified the secondary endosymbionts of distantly related Lachninae aphids belonging to 19 different species (8 different genera collected in three different countries [Austria, France, and Spain]) through 16S rRNA gene sequencing and phylogenetic analysis. In selected species, we have determined the location of the secondary endosymbionts using FISH with 16S-targeted specific oligonucleotide probes. We propose an evolutionary scenario for the establishment of an original secondary co-obligate endosymbiont in the LLCA, followed by symbiont replacements, internalisations of these into distinct bacteriocytes, and/or the putative establishment of tertiary obligate symbionts in different aphid lineages from this symbiont-diverse subfamily.

3.2 Results and Discussion

3.2.1 Phylogenetic analyses and fluorescence *in situ* hybridisation of Lachninae aphids' secondary endosymbionts

S. symbiotica and *S. marcescens*-like secondary symbionts

Most *Cinara* spp. investigated so far have been found to be associated with different *S. symbiotica* strains (Burke *et al.*, 2009; Jousselin *et al.*, 2016; Lamelas *et al.*, 2008). We have collected 19 representatives (comprising 11 species) of *Cinara* clades A (n=5), B (n=7), and C (n=7) (see Figure I.5), and have identified their endosymbionts through PCR, cloning, and sequencing of their 16S rRNA genes. We found that the secondary symbionts of all of the collected species – except for *Cinara* (*Cinara*) *confinis* (clade C) and *Cinara* (*Schizolachnus*) *obscurus* (clade B) – were indeed associated to *S. symbiotica* (details can be found on-line).

To re-evaluate the co-speciation previously observed for *Buchnera*-*Serratia* symbiont pairs within the Lachninae in the light of our new data, we performed a Bayesian phylogenetic reconstruction using currently available 16S rDNA sequences of both *Buchnera* (Figure 3.4A) and *Serratia* (Figure 3.4B) from Lachninae aphids. Contrary to earlier studies (Burke *et al.*, 2009; Lamelas *et al.*, 2008), we failed to recover the previously described *S. symbiotica* FL and OL clusters (see Figure I.7). We found that all Lachninae's *S. symbiotica* strains, but the one from *Trama caudata*, form a well-supported monophyletic clade nested within an unresolved group composed mainly of facultative strains of *S. symbiotica* from Aphidinae aphids, a strain from *Adelges tsugae* (Hemiptera: Adelgidae), and one from the Lachninae aphid *Trama troglodytes* (Figure 3.4B). This unresolved clade also contains the "early"

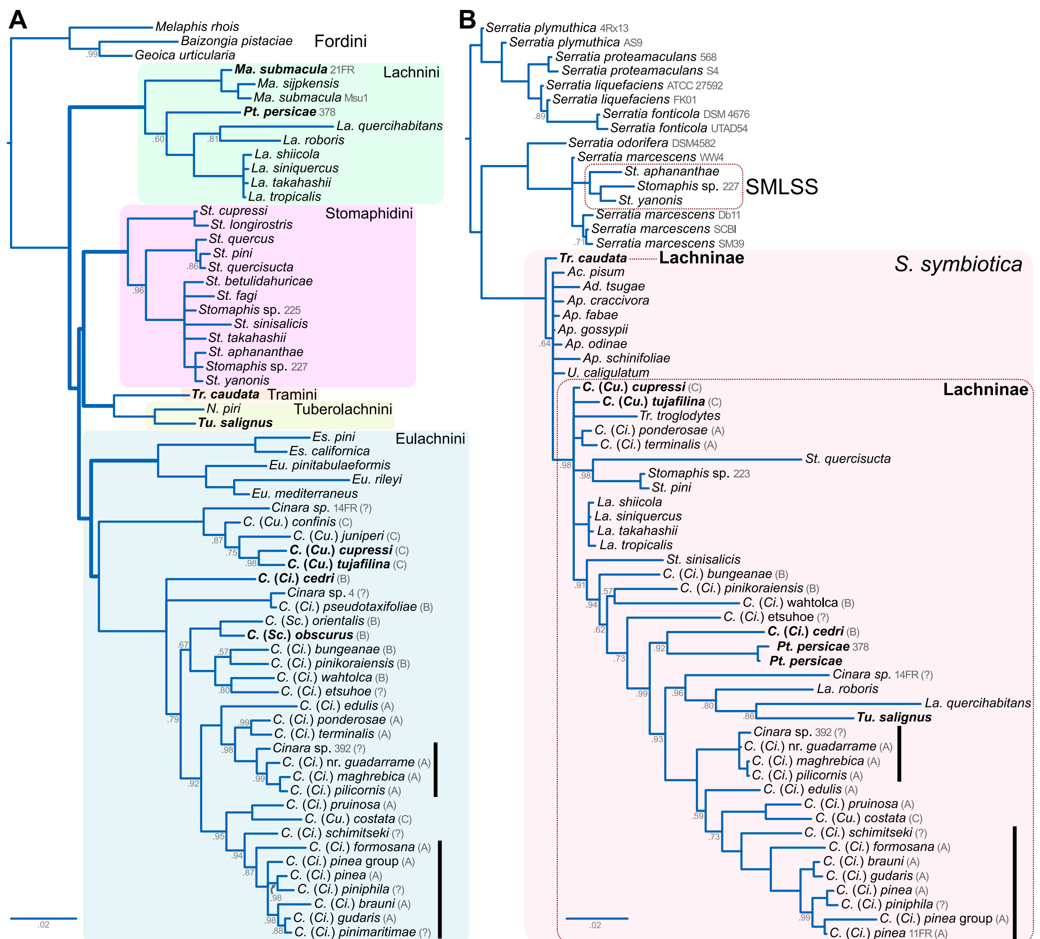


Figure 3.4 16S rRNA-based phylogenetic relationships of *Buchnera* and *Serratia* strains from Lachninae aphids. Bayesian phylogram of **(A)** *Buchnera* and **(B)** *Serratia* symbionts from selected aphids. *Buchnera* from the Fordini tribe and free-living *Serratia* strains were used for rooting the respective trees. Posterior probabilities equal to 1 are omitted from the trees to enhance clarity. For the *Buchnera* tree, the thicker branches represent constrained relationships within Lachninae tribes according to Chen *et al.* (2015a). Aphid tribe names in (A) are indicated at the top-right of the coloured boxes. The coloured box in B delimits the *S. symbiotica* clade, while dotted boxes delimit the SMLSS (from *Stomaphis* spp.) and the *S. symbiotica* strains from Lachninae aphids, respectively. In both (A) and (B), bold-lettered species names indicate the selected species for FISH microscopy. Reproduced from Manzano-Marín *et al.* (2016).

co-obligate *S. symbiotica* strain from *C.* (*Cu.*) *tujafilina* (Manzano-Marín and Latorre, 2014) and a strain from the closely related *Cinara* (*Cupressobium*) *cupressi*. Within the Lachninae clade, we recovered

three well-supported monophyletic clades made up from: **(i)** *Cinara* (*Cinara*) *ponderosae* and *Cinara* (*Cinara*) *terminalis* (both cluster A), **(ii)** some *Stomaphis* spp., and **(iii)** most *Lachnus* species. The latter belong to various closely related species, some suspected to be synonyms of *Lachnus tropicalis* (Blackman and Eastop, 2006), which would be consistent with the high sequence identity (>99%) of their *S. symbiotica* endosymbionts' 16S rRNA gene. Interestingly, most *S. symbiotica* strains from *Cinara* clade A form a well-supported monophyletic clade, and within this, there is high congruency with the phylogenetic relationships of the *Buchnera* strains found in the respective hosts, particularly within two subclades (Figure 3.4A and B, vertical black lines). In contrast, *S. symbiotica* strains from *Cinara* clade B are polyphyletic, and their phylogenetic relationships do not seem to mirror those of the respective *Buchnera* symbiont. Curiously, both the *Buchnera* and *S. symbiotica* from *Cinara* (*Cupressobium*) *costata* are recovered nested within strains form *Cinara* clade A. This contrasts *Cinara* phylogenetic relationships performed in Chen *et al.* (2015a) and Meseguer *et al.* (2015), which show this species to be part of a basal group of *Cinara* clade C. On the other hand, *S. symbiotica* strains from *Lachnus roboris*, *Lachnus quercihabitans*, *Tu. salignus*, and *Pterochloroides persicae* are all recovered within a clade encompassing most *S. symbiotica* strains from *Cinara* spp., reflecting no congruency with neither their hosts' nor their corresponding *Buchnera* phylogenetic relationships.

Serratia strains from *Stomaphis* spp. are recovered nested within both the free-living *S. marcescens* strains and the Lachninae *S. symbiotica* clade. The former constitutes what we denominate the ***S. m**arcescens*-**l**ike **s**econdary **s**ymbionts (**SMLSS**), all of which have been identified from aphids belonging to a single clade of *Stomaphis* spp. (Figure 3.4A and B). The latter are recovered as a monophyletic clade which is congruent with the *Buchnera* phylogeny, and as basal to the clade comprising most *S. symbiotica* strains from *Cinara* species.

Next, we investigated the diversity in localisation of *S. symbiotica*

within distantly related Lachninae by whole-mount FISH using aphid embryos. We found that 100% of the FISH-analysed specimens for *S. symbiotica* indeed housed this symbiont, meaning they were fixed in the population and providing clues towards their obligate status. Interestingly, we observed a great diversity in both cell shape and tissue tropism of *S. symbiotica* among the selected Lachninae (Figure 3.5). In *C.* (*Cu.*) *tujafilina* and *C.* (*Cu.*) *cupressi*, *S. symbiotica* is present in the periphery of the *Buchnera* bacteriocytes, co-infecting them, and also occupying its own bacteriocytes (Figure 3.5A and B). On the other hand,

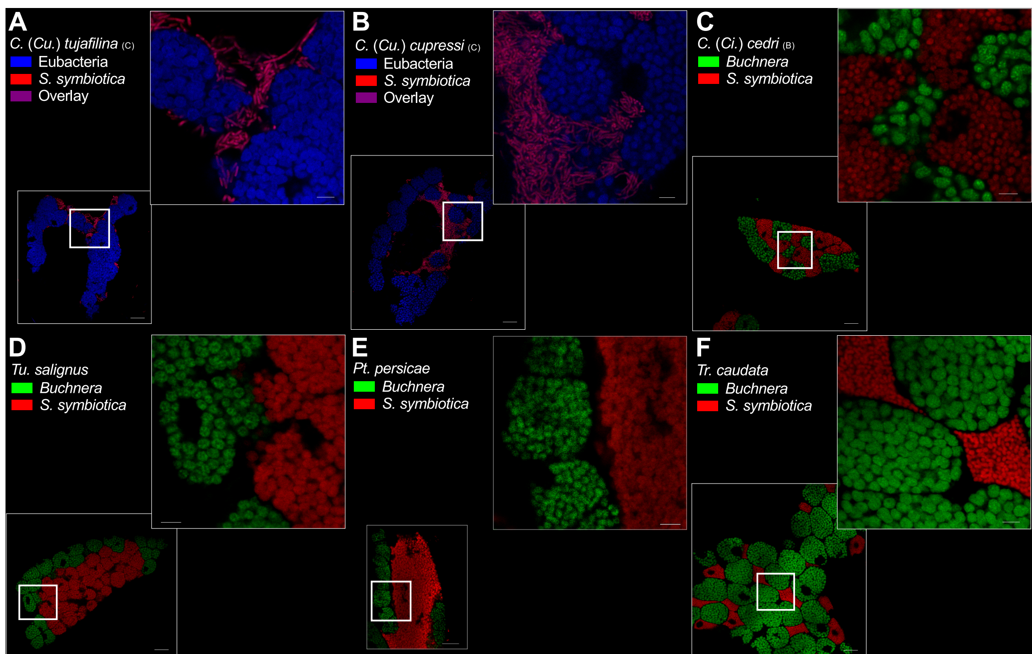


Figure 3.5 Location and morphology of *S. symbiotica* in selected Lachninae aphids. FISH microscopic images of aphid embryos from selected Lachninae aphids. **(A)** Ventral view of a *C.* (*Cu.*) *tujafilina* bacteriome. **(B)** Lateral-ventral view of a *C.* (*Cu.*) *cupressi* bacteriome. **(C)** Lateral view of a *C.* (*Ci.*) *cedri* bacteriome. **(D)** Lateral view of a *Tu. salignus* bacteriome. **(E)** Lateral-ventral view of a *Pt. persicae* bacteriome. **(F)** Ventral view of a *Tr. caudata* bacteriome. Thick white boxes indicate the magnification region, depicted in the top-right of each panel. The scientific name for each species along with the false colour code for each fluorescent probe and its target group are shown at the top-left of each panel. Scale bars from the unmagnified and magnified FISH images represent 20 and 5µm, respectively. Reproduced from Manzano-Marín *et al.* (2016).

in *C.* (*Ci.*) *cedri* (clade B), *Tu. salignus*, *Pt. persicae*, and *Tr. caudata*, *S. symbiotica* is housed exclusively inside distinct bacteriocytes (Figure 3.5C-F). The distribution of these secondary bacteriocytes is different among the aphid species. In *C.* (*Ci.*) *cedri* they are interspersed among *Buchnera* bacteriocytes (Figure 3.5C), in *Tu. salignus* they are found forming a "bacteriome core" surrounded by primary bacteriocytes (Figure 3.5D), and in *Pt. persicae* they form a "layer" along the bacteriome (Figure 3.5E). *S. symbiotica* cells appear round-shaped in *C.* (*Ci.*) *cedri*, *Tu. salignus*, and *Pt. Persicae*; while in *Tr. caudata* the secondary symbiont retains a rod shape and cell size similar to free-living *Serratia* strains (Figure 3.5F). Moreover, the *S. symbiotica* strains of *C.* (*Cu.*) *tujafilina* (Figure 3.5A), *C.* (*Cu.*) *cupressi* (Figure 3.5B) show an elongated filamentous cell shape, similarly to the facultative *S. symbiotica* symbiont of *Ac. pisum* (Moran *et al.*, 2005) (Figure S.2H).

***Sodalis*-like secondary symbionts**

Sodalis-like 16S rRNA gene sequences have been previously amplified from some Lachninae aphids, including *Eulachnus* spp., *Nippolachnus piri*, and *Cinara* (*Cinara*) *glabra* (cluster C) (Burke *et al.*, 2009). Here, we have confirmed, by 16S rRNA gene sequencing, the presence of a ***S**odalis*-**l**ike **s**econdary **s**ymbiont (**SLSS**) in different populations of *Eulachnus mediterraneus* (Spain) and *Eulachnus rileyi* (Austria, France, and Spain) (see collection table on-line). Additionally, using a specific primer designed to target the 16S rRNA gene of the SLSS of *Eulachnus* spp. (see M.3: Fluorescence *in situ* hybridisation), we detected and sequenced SLSS 16S amplicons from all the collected *C.* (*Sc.*) *obscurus* populations from Austria, France, and Spain, which were almost identical to each other, pointing towards this symbiont being fixed in this aphid species. By means of a Bayesian phylogenetic reconstruction, we determined the SLSS of Lachninae aphids to constitute at least 4 different lineages, nested within a clade made up of *Sodalis*-like symbionts from different insects (Figure 3.6A). Interestingly, the SLSS form *Eulachnus* spp. form a well-supported monophyletic clade,

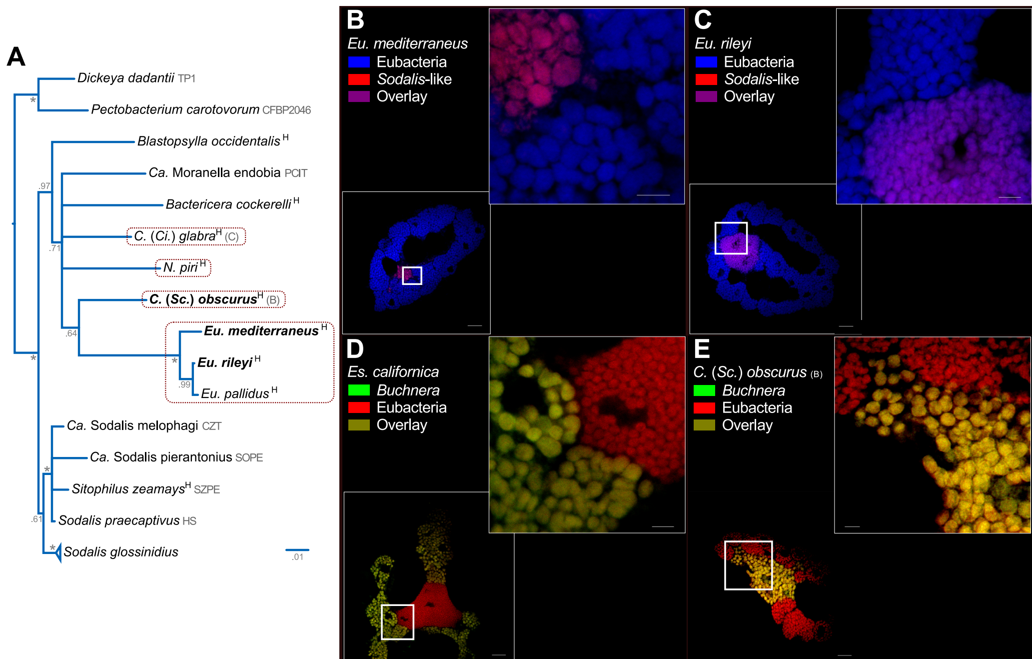


Figure 3.6 Location and 16S rRNA phylogenetic relationships of SLSS of Lachninae aphids. **(A)** Bayesian phylogram depicting the relationships and placement of known SLSS from aphids. The superscript "H" at the end of the full species name indicates the symbiont's host name was used. Bold-lettered species names indicate the species selected for FISH microscopy. **(B-E)** FISH microscopic images of aphid embryos of selected Lachninae aphids. **(B)** Lateral view of an *Eu. mediterraneus* bacteriome. **(C)** Ventral view of an *Eu. rileyi* bacteriome. **(D)** Lateral-ventral view of an *Es. californica* bacteriome. **(E)** Lateral view of a *C.* (*Sc.*) *obscurus* bacteriome. Thick white boxes indicate the magnification region, depicted in the top-right of each panel. The scientific name for each species along with the false colour code for each fluorescent probe and its target group is indicated at the top-left of each panel. Scale bars from the unmagnified and magnified FISH images represent 20 and 5 µm, respectively. Reproduced from Manzano-Marín *et al.* (2016).

reinforcing previous results (Burke *et al.*, 2009) and pointing towards a common origin. Given that members of *Essigella* have not been found associated to neither *S. symbiotica*, *H. defensa*, nor *Re. insecticola* endosymbionts (Russell *et al.*, 2003), and the close taxonomic and phylogenetic relationship of *Eulachnus* and *Essigella* (Figure I.5), we hypothesised that the SLSS detected in *Eulachnus* spp. could have been either fixed in the common ancestor of these two genera or right before

the diversification of *Eulachnus*. Regretfully, we were unable to recover any secondary symbiont's 16S rRNA from *Essigella californica* (France), neither by specific PCR nor by molecular cloning (50 colonies analysed).

To localise the SLSS within the bacteriome, we used the *Eulachnus* SLSS specific reverse PCR primer as a probe for FISH on dissected aphid embryos. We found that all individuals from both *Eu. rileyi* and *Eu. mediterraneus* harbour the SLSS inside specific bacteriocytes within the bacteriome (Figure 3.6B and C). Additionally, since we were unable to determine the 16S rRNA gene sequence of the putative secondary endosymbiont of *Es. californica*, we used a combination of a general bacterial and a *Buchnera* specific probe for FISH in embryos of this aphid species. We observed that there was indeed a distinct secondary bacterial symbiont with a very similar morphology and localisation as that of *Eulachnus* spp. (Figure 3.6D). In both *Eulachnus* spp. and *Es. californica*, we found that the symbiont seems to be under-represented when compared to *Buchnera*, similarly to what is observed for the SLSS of *N. piri* (Fukatsu *et al.*, 1998). This could be the reason why we failed to find it in *Es. californica*, even when 50 clones were analysed. Regarding *C.* (*Sc.*) *obscurus*, we did not observed a staining when the SLSS probe was used for FISH in this aphid species, and the probe lost its specificity in this host when formamide concentration was lowered to 25%, possibly also recognising other RNAs from *Buchnera*. Therefore, we used the same approach as for *Es. californica*. Using a general bacterial probe in combination with a *Buchnera* specific probe, we found that *C.* (*Sc.*) *obscurus* harbours two distinct spherical endosymbionts in separate bacteriocytes (Figure 3.6E). In contrast to *Eulachnus*, the SLSS bacteriocytes from *C.* (*Sc.*) *obscurus* are more abundant and located along the bacteriome surrounding *Buchnera* bacteriocytes.

"X-type" secondary symbionts

For the current study, we were able to collect two populations of *Maculolachnus submacula* aphids from France. Since an "X-type" symbiont was suspected to be the secondary symbiont of this aphid

species (Burke *et al.*, 2009; Lamelas *et al.*, 2008), we used a specific PCR assay to confirm the presence of this endosymbiont in both populations. Though this assay, we recovered 16S rRNA gene fragments sharing 100% sequence identity to each other. To facilitate phylogenetic analysis, we additionally performed molecular cloning of the 16S rRNA gene using universal primers 16SA1 and 16SB1 (Fukatsu and Nikoh, 1998) (see probe table on-line). Additionally, through the same method, we found an "X-type" symbiont associated with the aphid *C.* (*Cu.*) *confinis*. A Bayesian phylogenetic analysis of the different Aphididae "X-type" symbionts revealed that these form a monophyletic cluster closely related to *H. defensa* and *Re. insecticola*, facultative endosymbionts from *Ac. pisum* (Figure 3.7A). Particularly, the sequences obtained from *Ma. submacula* populations from three different countries form a well-supported monophyletic clade, separate

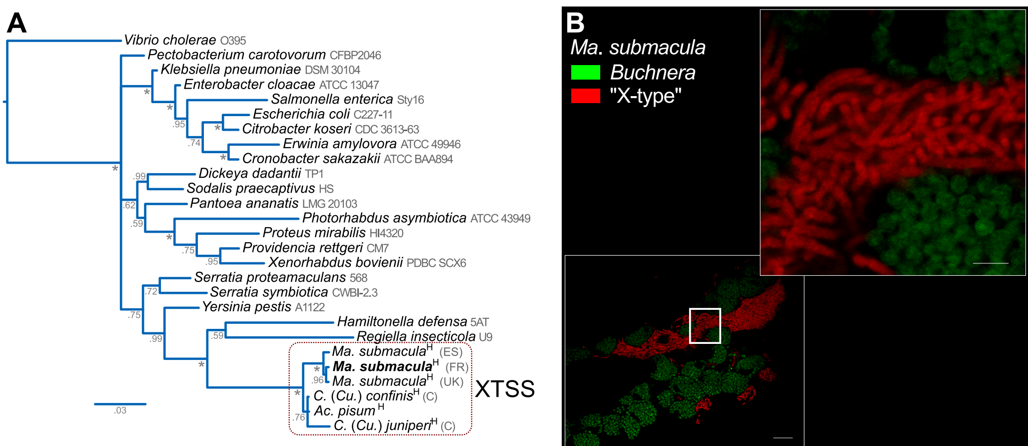


Figure 3.7 16S rRNA gene-based phylogenetic relationships of XTSSs and localisation of XTSS in *Ma. submacula*. (A) Bayesian phylogram depicting the relationships and placement of the currently available XTSSs from aphids and selected Enterobacteriaceae, using *Vibrio cholerae* as an outgroup. The superscript "H" indicates that the symbiont's host name was used. **(B)** FISH microscopic images of a lateral view of a *Ma. submacula* bacteriome. Thick white box indicates the magnification region, depicted in the top-right of each panel. The scientific name for the species along with the false colour code for each fluorescent probe and its target group is shown at the top-left of the panel. Scale bars from the unmagnified and magnified FISH images represent 20 and 5 µm, respectively. Modified from Manzano-Marín *et al.* (2016).

from that of *C.* (*Cu.*) *confinis*, *Cinara* (*Cupressobium*) *juniperi* (clade C), and *Ac. pisum*, and show a high sequence identity among each other (>99%). We found that 100% of all individuals analysed from both *Ma. submacula* populations contained XTSSs distributed along the bacteriome, both surrounding *Buchnera* bacteriocytes and in its own distinct ones (Figure 3.7B). Regarding *C.* (*Cu.*) *confinis*, we lacked enough individuals to perform FISH analysis, and therefore its localisation within the bacteriome remains undetermined.

3.2.2 Evolutionary succession of secondary endosymbionts within the Lachninae

Based on the aforementioned observations and previous microscopic and molecular work on symbionts from the Lachninae subfamily, we propose an evolutionary scenario for the settlement, internalisations, and replacements of the original secondary co-obligate endosymbiont from the LLCA (Figure 3.8). Firstly, with a combination of specific PCR assays, 16S rRNA gene sequencing, and FISH microscopy, we determined that all analysed specimens indeed harbour fixed secondary endosymbionts at the population/species level. This fact, in combination with previously published microscopic (Buchner, 1953; Fukatsu and Ishikawa, 1998; Fukatsu *et al.*, 1998; Lamelas *et al.*, 2008; Michalik *et al.*, 2014; Pyka-Fościak and Szklarzewicz, 2008) and molecular data from Lachninae aphids (Burke *et al.*, 2009; Lamelas *et al.*, 2011a,b; Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2016; Pérez-Brocal *et al.*, 2006), provide strong evidence for the dependence of members of this subfamily on co-obligate secondary endosymbionts, putatively due to the previously-mentioned ancient pseudogenisation of the riboflavin biosynthetic genes in the *Buchnera* harboured by the LLCA. The detection of *S. symbiotica* in different aphids from at least 6 Lachninae genera, along with the genomic data from three strains of this symbiont in the Lachninae found at different stages of the genome

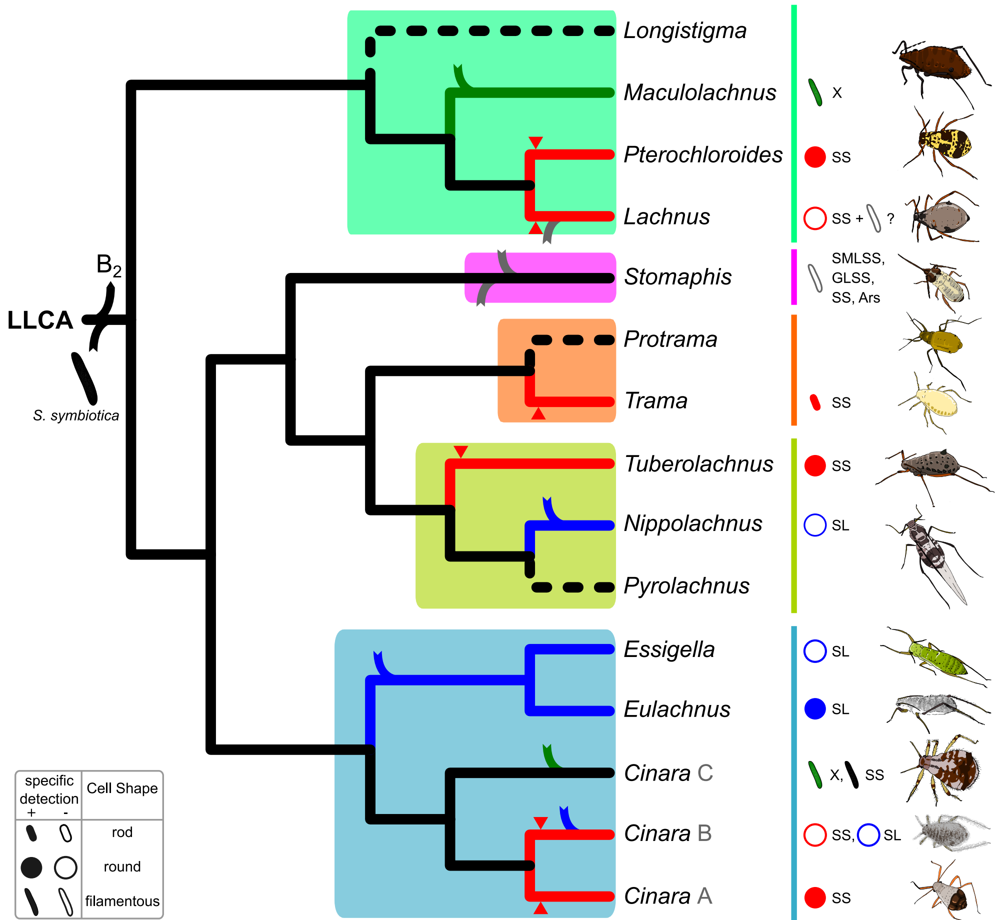


Figure 3.8 Evolutionary succession of secondary endosymbionts within the Lachninae. Cladogram displaying the relationships of Lachninae lineages by genera. Coloured boxes shading monophyletic clades as well as vertical lines on the right side delimit the five tribal clades according to Chen *et al.* (2015a); Meseguer *et al.* (2015). Incoming lines on branches symbolise the acquisition of co-obligate secondary symbionts. The outgoing line at the root of the tree stands for the loss of the riboflavin biosynthetic genes in the *Buchnera* from the LLCA. Green and blue branches represent lineages where XTSS or an SLSS have replaced the original *S. symbiotica* symbiont, respectively. Red branches with an arrowhead pointing to them reflect the internalisation of *S. symbiotica* into distinct bacteriocytes. At the leaves, shapes symbolising the bacterial endosymbionts' cell shapes according to the key (bottom-left) and cartoons of selected aphids form the different Lachninae genera are indicated. B_2=Riboflavin, SS=*S. symbiotica*, X=XTSS, Ars=*Arsenophonus*, SMLSS= *S. marcescens*-like secondary symbiont, GLSS= *Gilliamella*-like secondary symbiont, SL= SLSS, ?=unknown. Modified from Manzano-Marín *et al.* (2016).

reduction process (Lamelas *et al.*, 2011b; Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2016), point towards an early establishment of *S. symbiotica* as co-obligate in the LCCA. While spherical and found obligatorily inside bacteriocytes (with a highly reduced genome) in *Tu. salignus* and *C.* (*Ci.*) *cedri*, the filamentous and broadly distributed (with a mildly-reduced genome) *S. symbiotica* from *C.* (*Cu.*) *tujafilina* would preserve the traits of the putative "ancient" *S. symbiotica* from the LLCA. This would be consistent with the genomic and phenotypic traits of the facultative *S. symbiotica* from *Ac. pisum* (Burke and Moran, 2011; Moran *et al.*, 2005). We find this scenario to be most parsimonious, as it would require one single event of infection with a shared *S. symbiotica* ancestor in the LLCA, followed by at least four internalisations of *S. symbiotica* into bacteriocytes. This ancient secondary associate would have then undergone at least six independent events of symbiont replacement. An alternative scenario would require additional events of symbiont replacement with distinct *S. symbiotica* strains in specific Lachninae lineages.

Within the Lachnini tribe, there could have been either one or two independent events of internalisation and confinement of *S. symbiotica* into distinct bacteriocytes. The latter would be supported by lack of congruency between *Buchnera* and *S. symbiotica* lineages from *Pt. persicae* and *Lachnus* spp., suggesting separate events of genome reduction of *S. symbiotica* in these two aphid lineages. Furthermore, microscopic observations of *La. roboris* have revealed that it keeps a vertically transmitted association with a third filamentous-shaped bacteria, whose identity remains unknown (Buchner, 1953) (Figure S.9A). If this tertiary symbiotic bacteria was established in the common ancestor of *La. roboris* and *La. quercihabitans*, it could explain the longer branches, relative to *La. tropicalis*, leading to both of their *S. symbiotica* strains (Figure 3.4B), given that the presence of an additional symbiont could drive the rapid sequence evolution of the *S. symbiotica* symbiont. Further genome sequencing of the endosymbiotic bacteria

from these aphids will help clarify the specifics of these associations. Also, we have confirmed that *Ma. submacula* aphids indeed harbour XTSSs and determined its location within aphid embryos. We found that XTSS was present in 100% of the individuals analysed, which in combination with previous analyses detecting the presence of this symbiont in a Spanish (Lamelas *et al.*, 2008) and a UK (Burke *et al.*, 2009) population, points towards its obligate status. The morphology and localisation of XTSS (Figure 3.7B), resembles that observed for facultative endosymbionts of other aphids (Moran *et al.*, 2005) (Figure S.1), similarly to what is observed for the co-obligate *S. symbiotica* from *C.* (*Cu.*) *tujafilina* (Lamelas *et al.*, 2008; Manzano-Marín and Latorre, 2014) (Figure 1.2). This suggests XTSS has not yet undergone a massive genome reduction, contrary to what is observed in the pleomorphic *S. symbiotica* of *C.* (*Ci.*) *cedri* (Lamelas *et al.*, 2011b) and *Tu. salignus* (Manzano-Marín *et al.*, 2016). This, in combination with the lack of a *S. symbiotica* endosymbiont, points toward a replacement of the "ancient" secondary co-obligate endosymbiont, which must have occurred on the branch leading to *Ma. submacula*. It is important to note that *Ma. submacula* strains harbour at least two phylogenetically distinct *Buchnera* lineages (Figure 3.4A): one sister to *Maculolachnus sijpkensis* and one sister to this clade. In the current work, we refer to the latter as the one associated to XTSS.

Regarding the Stomaphidini, we postulate that at least 2 events of symbiont replacement have occurred. *Stomaphis pini* and *Stomaphis quercisucta* have been found to be associated with *S. symbiotica* (Burke *et al.*, 2009; Chen *et al.*, 2015b), and microscopic investigations into *St. quercus* have revealed that this species houses three vertically transmitted endosymbiotic bacteria: *Buchnera* plus two secondary symbionts which apparently share niche inside the same secondary bacteriocytes (Buchner, 1953; Pyka-Fościak and Szklarzewicz, 2008). These are likely to be an *Arsenophonus* and a *Gilliamella*-like secondary symbiont (Figure S.10), both of which have been found associated to

different Polish populations of this aphid species (Burke *et al.*, 2009). In addition, both *Stomaphis aphananthae* and *Stomaphis yanonis* are associated with SMLSSs (Burke *et al.*, 2009), suggesting an establishment of this symbiont in the branch leading to the clade comprising these two species. Additionally, microscopic analyses of *St. yanonis* bacteriomes have revealed this species indeed house a tubular secondary symbiont, putatively SMLSS, in separate bacteriocytes located on the surface of the bacteriome "core" formed by the *Buchnera* bacteriocytes (Fukatsu and Ishikawa, 1993, 1998). Consequently, we propose at least two events of acquisition of a new endosymbiont: one before the diversification of the clade comprising *St. cupressi* and another one before the expansion of the large unresolved clade including *St. yanonis* (Figure 3.4A).

With respect to the Tramini/Tuberolachnini clade, all currently analysed members have been found to be affiliated to *S. symbiotica* (Burke *et al.*, 2009), except for *N. piri*, which is associated with a putatively pleomorphic SLSS housed inside separate bacteriocytes (Burke *et al.*, 2009; Fukatsu and Ishikawa, 1998; Fukatsu *et al.*, 1998; Pérez-Brocal *et al.*, 2006). In the case of both *Tu. salignus* and *Tr. caudata*, the *S. symbiotica* symbiont is found exclusively within bacteriocytes (Figure 3.5D and F). However, the cell shape of the endosymbiotic bacterium is strikingly different. While in *Tu. salignus* it presents a large spherical-shaped cell, in *Tr. caudata* the symbiont keeps a small rod-shaped cell, resembling free-living *Serratia* strains. This could be indicative of this bacterium finding itself in the very first stages of intracellularisation into bacteriocytes, still preserving its rod shape and having a genome resembling that of *C.* (*Cu.*) *tujafilina*.

In regards to the Eulachnini, most *Cinara* spp. have been consistently found associated to *S. symbiotica* strains. Microscopic investigations of *Cinara* (*Cinara*) *pini* (clade A) and *C.* (*Ci.*) *cedri* have revealed they indeed harbour a pleomorphic secondary endosymbiotic bacterium obligatorily inside bacteriocytes (Fukatsu *et al.*, 1998) (Figure 3.5C and S.9B), and in the case of the latter, genomic-based metabolic inference

has established that both endosymbiotic partners are required for the biosynthesis of various essential nutrients (Gosalbes *et al.*, 2008; Lamelas *et al.*, 2011b). Additionally, a high level of congruency between the phylogenetic relationships of *Buchnera* and *S. symbiotica* strains from clade A's *Cinara* (Figure 3.4A and B) suggests a single event of drastic genome reduction followed by divergence, similar to what is observed for *Buchnera*. On the contrary, *S. symbiotica* from clade B's *Cinara* do not show this congruent pattern, pointing possibly to independent events of drastic genome reduction. Within *Cinara* clade B, *C.* (*Sc.*) *obscurus* would represent a case of symbiont replacement by a SLSS, which is present obligatorily inside bacteriocytes (Figure 3.6E). Whether this symbiont is widespread within the *Cinara* (*Schizolachnus*) subgenus remains to be explored. Nonetheless, microscopic analysis of *Cinara* (*Schizolachnus*) *pineti* have revealed it indeed harbours a spherical secondary endosymbiotic bacterium which is vertically transmitted in both oviparous and viviparous generations (Michalik *et al.*, 2014). As most *Cinara*, *C.* (*Cu.*) *tujafilina* and *C.* (*Cu.*) *cupressi*, are associated to *S. symbiotica* strains, however, both the location and 16S rRNA gene sequence of these more closely resemble the facultative strains of Aphidinae aphids (Figure 3.4B; 3.5A and B; Moran *et al.*, 2005). Genomic-based metabolic inference has provided evidence towards the obligate status of *S. symbiotica* in *C.* (*Cu.*) *tujafilina*, given the loss of the riboflavin biosynthetic capability of *Buchnera*, an essential co-factor now synthesised by *S. symbiotica* (Manzano-Marín and Latorre, 2014). This, in addition to the consistent association of these two *Cinara* (*Cupressobium*) species with *S. symbiotica*, led us to infer that these aphids do indeed keep an obligate association with closely related secondary endosymbiotic strains. Within *Cinara* clade C, evidence of at least one species (*C.* [*Ci.*] *glabra*) being affiliated to a SLSS (Figure 3.6A) and two (*C.* [*Cu.*] *confinis* and *C.* [*Cu.*] *juniperi*) to XTSS (Figure 3.7A), rather than *S. symbiotica*, suggests some events of symbiont replacement have occurred in this group of species. This could have been facilitated due to the niche occupied by *S. symbiotica*, being similar

to that of facultative endosymbionts of *Ac. pisum* (Moran *et al.*, 2005; Sakurai *et al.*, 2005; Sandström *et al.*, 2001; Tsuchida *et al.*, 2010) (Figure S.2). Finally, we propose a symbiont replacement event by a SLSS in the branch leading to the monophyletic clade formed by *Eulachnus* and *Essigella* species. Consistent with previous observations in *Eu. rileyi* (Michalik *et al.*, 2014), we found that both this species and *Eu. mediterraneus* harbour spherical SLSSs in separate bacteriocytes, spatially-arranged in a very similar fashion (Figure 3.6B and C). Also, even though we were unable to recover a 16S rRNA gene sequence belonging to a bacterial taxa other than *Buchnera* in *Es. californica*, we were able to detect by FISH microscopy the presence of secondary spherical bacterial endosymbionts residing in distinct bacteriocytes, localised similarly to those inhabited by SLSSs in *Eulachnus* spp. (Figure 3.6D). Therefore, pending further studies, we propose that the secondary symbiont found in *Es. californica* belongs to the same lineage as the SLSSs of *Eulachnus* species.

3.3 Conclusion

In summary, we propose an evolutionary framework which should assist in future studies on these symbiont-diverse subfamily. We observed that some putative co-obligate endosymbionts seem to have evolved from facultative ones, such as XTSS and *S. symbiotica*, while the phylogenetic relationships of others, namely SLSSs, remain to be determined. Additionally, the location within the bacteriome of some secondary endosymbionts, reveal that the distribution of these secondary bacteriocytes within the bacteriome does not follow a definite rule, having *S. symbiotica* bacteriocytes differently distributed among Lachninae aphids. Our findings reveal a dynamic pattern for the evolutionary history of "recently"-acquired endosymbionts, thus contributing to a better understanding of how mutualism in endosymbiotic associations can evolve. The role these bacteria have played in the adaptation of their

aphid hosts to different niches/feeding sites/plants and their role in speciation in this peculiar subfamily remains to be explored.

Chapter 4

Genome reduction in *Serratia symbiotica*

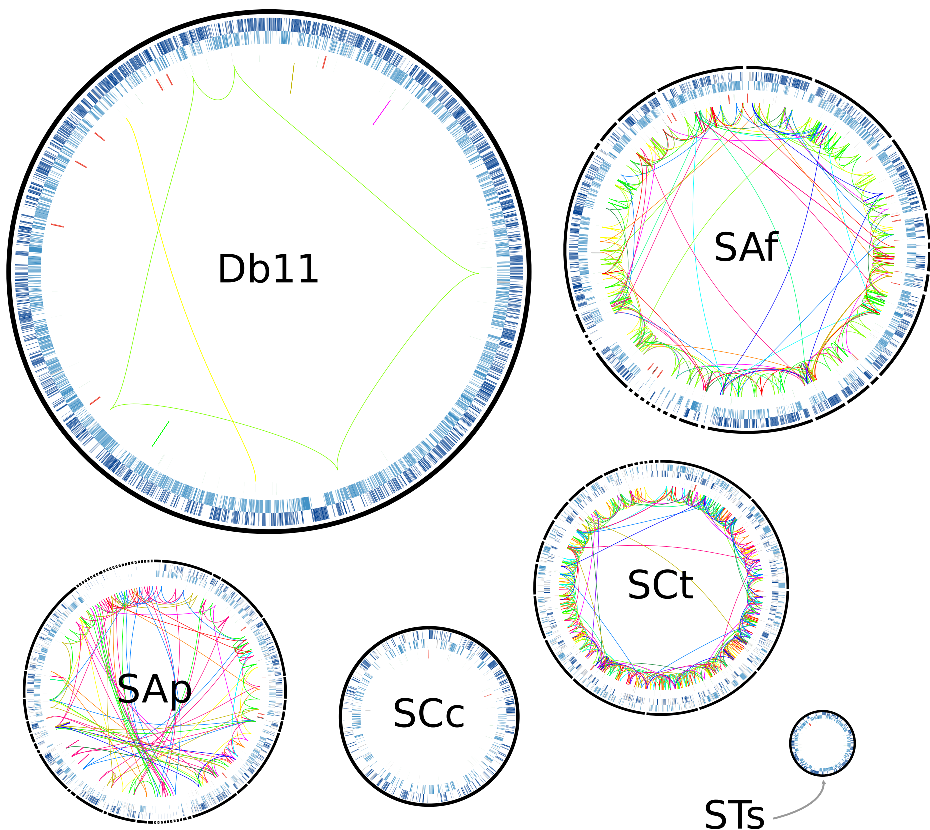


Figure 4.1 ***S. symbiotica* and its dwindling genomes.** Circular representations of *S. symbiotica* and *S. marcescens* genomes in scale. Light and dark blue highlights display genes in the forward and reverse strand, respectively. Red highlights mark the position of rRNA genes. Coloured links join mobile elements of the same type along the genome.

4.1 Introduction

Whichever a symbiont's function, taxonomic position, or status (facultative or obligate), a common feature from maternally-inherited endosymbiotic organisms is the possession of a reduced genome, when compared to their free-living counterparts (Burke and Moran, 2011; Clayton *et al.*, 2012; Foray *et al.*, 2014; Gil *et al.*, 2003; Gottlieb *et al.*, 2015; Kenyon *et al.*, 2015; Manzano-Marín *et al.*, 2015; Nakabachi *et al.*, 2013; Nikoh *et al.*, 2011; Smith *et al.*, 2015; van Ham *et al.*, 2003). The sequencing of these genomes has undoubtedly provided important clues into the distinct features these display along the erosion process. While mildly reduced genomes such as the one from *Sodalis glossinidius*, (facultative endosymbiont of the tsetse fly *Glossina morsitans morsitans*), *Sodalis pierantonius* (primary obligate endosymbiont from the rice weevil *Sitohpilus oryzae*), and *H. defensa* (facultative endosymbiont from the aphid *Ac. pisum*) show intermediate guanine-cytosine contents and a massive presence of both mobile elements (**MEs**) and pseudogenes (Degnan *et al.*, 2009b; Oakeson *et al.*, 2014; Toh *et al.*, 2005), highly reduced genomes such as the ones from *Buchnera* and *Blochmannia* (primary obligate endosymbiont from carpenter ants) are highly compact with few pseudogenes and show no traces of MEs (van Ham *et al.*, 2003; Williams and Wernegreen, 2015). However valuable the study of these genomes is, very few examples are available for different bacterial strains, belonging to a single genus, holding differentially reduced genomes. These examples include *Arsenohponus* symbionts of a parasitic wasp (*Nasonia vitripennis*) (INSDC:AUCC00000000.1), the brown planthopper (*Nilaparvata lugens*) (Xue *et al.*, 2014), and louse flies (Diptera: Hippoboscidae) (Nováková *et al.*, 2015) (INSDC:CP013920.1); *Coxiella* symbionts of ticks (Gottlieb *et al.*, 2015; Smith *et al.*, 2015); and *Sodalis* symbionts from the tsetse fly (Toh *et al.*, 2005) and rice weevil (Oakeson *et al.*, 2014).

S. symbiotica, is particular among currently sequenced bacterial

symbionts. First, while strains harboured by *Ap. fabae* (SAf) and *Ac. pisum* (SAp) (Aphidinae subfamily) are of facultative nature (Chandler *et al.*, 2008; Chen and Purcell, 1997; Chen *et al.*, 2000), strains from the aphids *C.* (*Cu.*) *tujafilina* (SCt), *C.* (*Ci.*) *cedri* (SCc), and *Tu. salignus* (STs) (Lachninae subfamily) have established co-obligate associations with both the aphids and its primary obligate endosymbiont, *Buchnera* (Lamelas *et al.*, 2011b; Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2016). In addition, depending on the strain, its cell shape is either rod-like (SAf) (Sabri *et al.*, 2011), filamentous (SAp and SCt) (Manzano-Marín and Latorre, 2014; Moran *et al.*, 2005), or spherical shape (SCc and STs) (Lamelas *et al.*, 2011b; Manzano-Marín *et al.*, 2016). Not surprisingly, these strains hold genomes of contrasting sizes, ranging from 3.58 to 0.65 Mbp (Burke and Moran, 2011; Foray *et al.*, 2014; Lamelas *et al.*, 2011b; Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2016). Furthermore, although most strains have not yet (to our knowledge) been axenically cultured, similarly to many insect obligate endosymbionts (reviewed in Kikuchi, 2009), SAf is able to grow freely in anaerobic conditions on a rich medium (Sabri *et al.*, 2011). Finally, while the phylogenetic relations of *S. symbiotica* are not fully resolved (Manzano-Marín *et al.*, 2016), they show a clear sister relationship to *S. marcescens*, a taxon comprised of various free-living bacteria for which complete genomes are available.

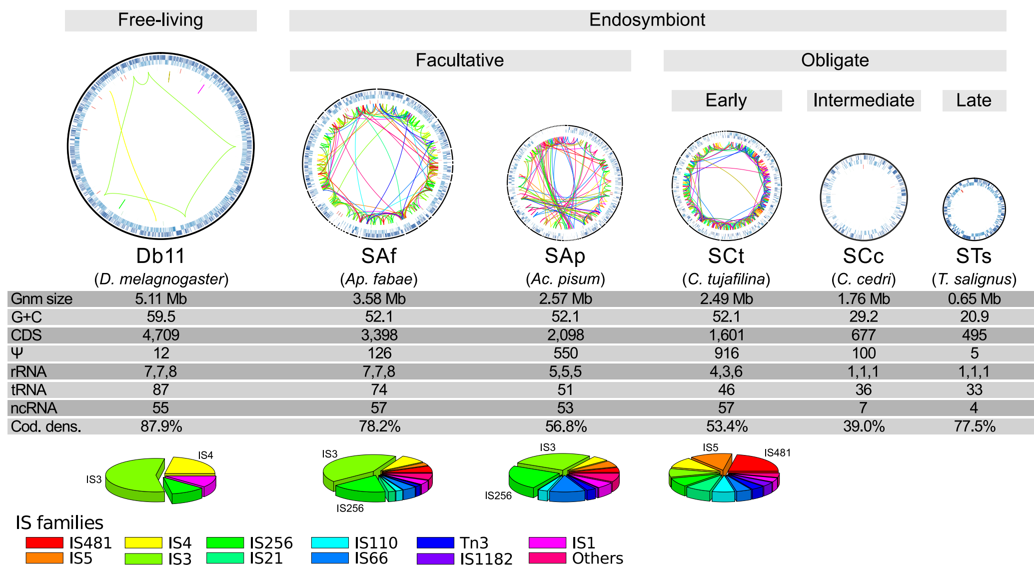
Given the observations for *S. symbiotica*, we analysed the genomes of currently available *S. symbiotica* strains and compared them to the free-living insect pathogen *S. marcescens* strain Db11 (**Db11**) (Flyg *et al.*, 1980; Iguchi *et al.*, 2014). Through comparative genomics we investigated genome rearrangement, the enrichment, and loss, of MEs, and the erosion of RNA features and the informational machinery undergone by *S. symbiotica* strains. Additionally, we describe the diminution of certain genes and the putative functional consequences of these reductions.

4.2 Results and Discussion

4.2.1 *S. symbiotica* strains and their shrinking genomes

Generally, "ancient" obligate endosymbionts hold highly-reduced genomes as small as 112 kbp (Bennett and Moran, 2013), conversely, more "recently" derived endosymbionts (including facultative ones) tend to display larger genomes, all the way up to the 4.5 Mbp genome of *S. glossinidius* (reviewed in Moran and Bennett, 2014). Accordingly, the different genomes of *S. symbiotica* strains land within and along this spectrum, from the large 3.58 Mbp genome of the facultative SAf to the small 0.65 Mbp genome of the co-obligate STs (Figure 4.2). Similarly to the other large endosymbiotic genomes (Degnan *et al.*, 2009a,b; Hansen *et al.*, 2012; Oakeson *et al.*, 2014; Toh *et al.*, 2005), SAf, SAp, and SCt's display a large enrichment of MEs, both in terms of diversity and number of them. Interestingly, the composition of the insertion sequence families (the most common type of MEs found within these genomes) seems to be lineage-specific. While IS3 and IS256 are the most prevalent in SAf and SAp (both facultative endosymbionts from Aphidinae aphids), IS481 and IS5 are the most common in co-obligate SCt. Conversely, the smaller genomes of SCc and STs lack any traces of MEs, congruent with similar-sized endosymbiotic genomes (see Moran *et al.*, 2008). Following the trend of many other dwindling genomes (Moran *et al.*, 2008), all *S. symbiotica* have a G+C content lower than that of their free-living counterpart, Db11. While this G+C content is very similar among SAf, SAp, and SCt (52.1%), there is a marked decrease in SCc (29.2%) and even more so in STs (20.9%). Additionally, while there is a great enrichment of pseudogenes in SAf (126), SAp (550), SCt (916), and SCc (100), the small STs is almost deprived of these gene remnants. This genetic erosion comes along with a decrease in coding density. Accordingly, while SAf shows only a small decrease when compared to Db11 (87.9% to 78.2%), SAp, SCt, and SCc exhibit a marked drop down

(56.8%, 53,4%, and 39.0%, respectively), mainly due to the increased pseudogenisation and "junk" DNA. On the other hand, the highly-reduced STs shows a high coding density (77.5%). This difference between SCc and STs is mainly due to high amount "junk" DNA that is present in SCc's genome, amounting to almost half of it (Lamelas *et al.*, 2011b). Finally, we also found a gradual loss (from free-living Db11 to co-obligate intracellular STs) of RNA features (rRNAs, tRNAs, and other non-coding RNAs [**ncRNAs**]), revealing their different levels of genomic erosion.



	Db11	SAf	SAp	SCt	SCc	STs
Gnm size	5.11 Mb	3.58 Mb	2.57 Mb	2.49 Mb	1.76 Mb	0.65 Mb
G+C	59.5	52.1	52.1	52.1	29.2	20.9
CDS	4,709	3,398	2,098	1,601	677	495
Ψ	12	126	550	916	100	5
rRNA	7,7,8	7,7,8	5,5,5	4,3,6	1,1,1	1,1,1
tRNA	87	74	51	46	36	33
ncRNA	55	57	53	57	7	4
Cod. dens.	87.9%	78.2%	56.8%	53.4%	39.0%	77.5%

Figure 4.2 Genome reduction in *S. symbiotica*. *Serratia* genomes are depicted as circular plots and are arranged from largest (leftmost) to smallest (rightmost). From outermost to innermost, the rings within the genome plots display features on the direct strand, the reverse one, and RNA genes. Inside the circles, coloured lines connect the same-family IS elements scattered throughout the genome, following the colour code at the very bottom of the image. The grey bars on top of the genome plots describe the lifestyle and genome reduction stage. Underneath the genome plots, the strain alias and the host, between parenthesis are shown. Below, a table showing the genomic features of each strain and pie charts displaying the relative abundance of IS-family elements, with the two most abundant highlighted by name. Underneath, the colour code for the different IS elements. Reproduced from Manzano-Marín and Latorre (2016).

As has been previously observed in other endosymbionts, genome erosion comes with a "disturbance" of the functional profile of the

organism, when compared to their free-living relatives (Clayton *et al.*, 2012; Gil *et al.*, 2003). Previous analyses have described that while the functional profiles of free-living *Serratia* strains were very stable, a displacement of it was evident in SCc and SAp (Manzano-Marín *et al.*, 2012). Accordingly, through a similar analysis using all five currently available *S. symbiotica* strains, we have found that while the recently derived SAf, SAp, and SCt strains cluster together forming a sister group to Db11, the highly reduced SCc and STs form a divergent cluster from the rest of *Serratia* strains (Figure 4.3A). These two *S. symbiotica* clusters differ mainly in the relative presence of MEs (category **X**) and translation-related genes (category **J**). While the former reflects the enrichment of SAf, SAp, and SCt in MEs, the latter evidences the common trend in highly reduced endosymbionts to retain housekeeping genes (e.g. category J includes all ribosomal proteins) (see Moran *et al.*, 2008).

In the early stages of an endosymbiont's genomic reduction, the genome's enrichment in MEs can lead to chromosomal rearrangement

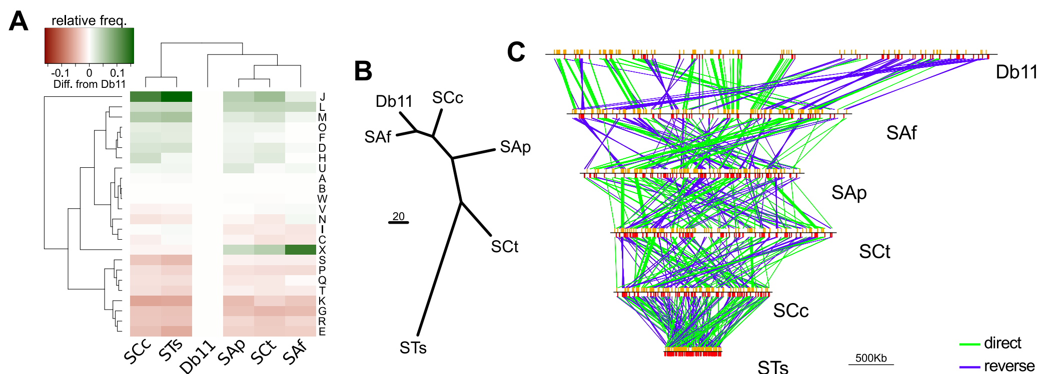


Figure 4.3 ***S. symbiotica* functional profile displacement and genome rearrangement.** **(A).** Heat map showing the two-way clustering of the *S. symbiotica* COG profile's differences, relative to the free-living *S. symbiotica*. **(B)** Unrooted tree as calculated by MGR for the minimum number of rearrangements undergone by *S. symbiotica*. **(C)** Graphic linear representation of the rearrangements undergone in *S. symbiotica*. Contigs that do not have single-copy conserved genes are not displayed. Reproduced from Manzano-Marín and Latorre (2016).

(Clayton *et al.*, 2012; Manzano-Marín and Latorre, 2014). This rearrangements get fixed in the endosymbiotic lineage once the MEs have been lost, as is observed by the general genome-wide synteny displayed in the "ancient" *Buchnera* (Tamas *et al.*, 2002; van Ham *et al.*, 2003), *Blochmannia* (Williams and Wernegreen, 2015), or *Blattabacterium* (Patiño-Navarrete *et al.*, 2013). Free-living *Serratia* strains display general genome-wide synteny (Manzano-Marín *et al.*, 2012), on the contrary, *S. symbiotica* genomes display various rearrangements when compared to free-living Db11's, and even among each other's (Figure 4.3B and C). Interestingly, while the mildly reduced genome of SAf displays the closest relationship (in terms of rearrangements) to Db11, the drastically reduced genome of STs has accumulated the highest number of rearrangements. Also, SCc and STs' genomes, which both lack MEs, display no synteny between them. These observations suggest that all *S. symbiotica* lineages have diverged before the loss of MEs, allowing a great number of lineage-specific reorganisation.

4.2.2 Erosion of essential amino acid biosynthetic routes

A general feature of endosymbiotic genomes is the loss of non-essential genes, leading to a highly reduced genome with a genetic repertoire specialised in the symbiotic function (reviewed in Moran *et al.*, 2008). In aphids, *Buchnera*, the primary obligate endosymibont, is mainly in charge of producing EAAs for its host. Therefore, it is expected that co-existing symbionts show degraded biosynthetic routes involved in the production of these compounds. By analysing these routes in *S. symbiotica*, it becomes immediately evident the gradual degradation (from Db11 to STs) of genes and operon attenuators implicated in synthesising EAAs (Figure 4.4). The recently derived SAf shows intact routes for most EAAs, with the notable expections of lysine and

methionine. As already described in a previous study (Manzano-Marín and Latorre, 2014), there is a marked difference in the retention of leucine, arginine, and histidine biosynthetic-related genes even between the closely related facultative SAp and co-obligate SCt. Finally, by comparing SCc and STs against the other *S. symbiotica* strains and each other, it becomes evident that both have become highly dependent on *Buchnera*'s supply of EAAs, with the main difference between SCc and STs being the purging of the remaining pseudogenes in the latter.

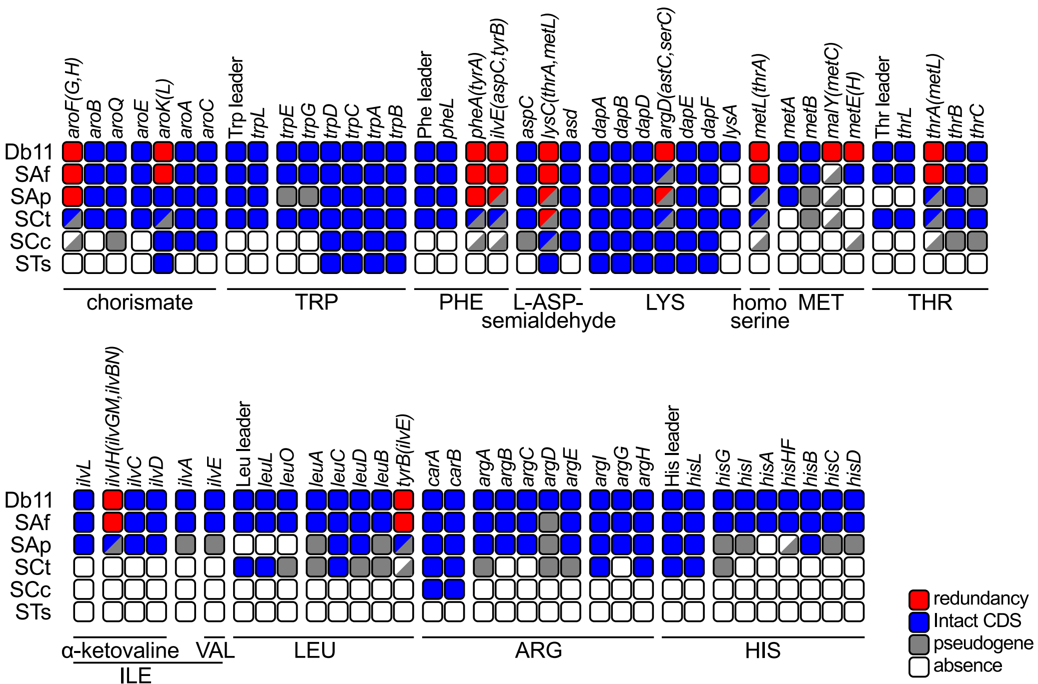


Figure 4.4 Erosion of essential amino acid biosynthetic genes in *S. symbiotica*. Inactivation tables showing the genes and leader sequences involved in the essential amino acid biosynthetic routes in *S. symbiotica* genomes compared to free-living Db11. At the top and left of the table, gene names for each enzymatic step and abbreviation for each *Serratia* strain, respectively. At the bottom of the table, black lines encompass the enzymatic steps required for the biosynthesis of each compound. Amino acid names are displayed using standard three-letter abbreviations. At the bottom-right, colour code for squares. Half-coloured boxes mean the genes catalysing the enzymatic step are present in different states. Reproduced from Manzano-Marín and Latorre (2016).

4.2.3 Decay of RNA features and the loss of regulation

Typically, highly reduced endosymbionts retain only a small number of ncRNAs and other RNA features (Figures S.11 and S.12). Through an annotation of these in the genomes of *S. symbiotica* and Db11, we have explored the erosion of RNA features (Figure 4.5: top panel). We found that in the recently derived endosymbionts SAf, SAp, and SCt, many of these features are still retained, although differentially. This points towards drift acting behind the loss of these features at the early stages of genome reduction. In the intermediate and advanced stages of drastic

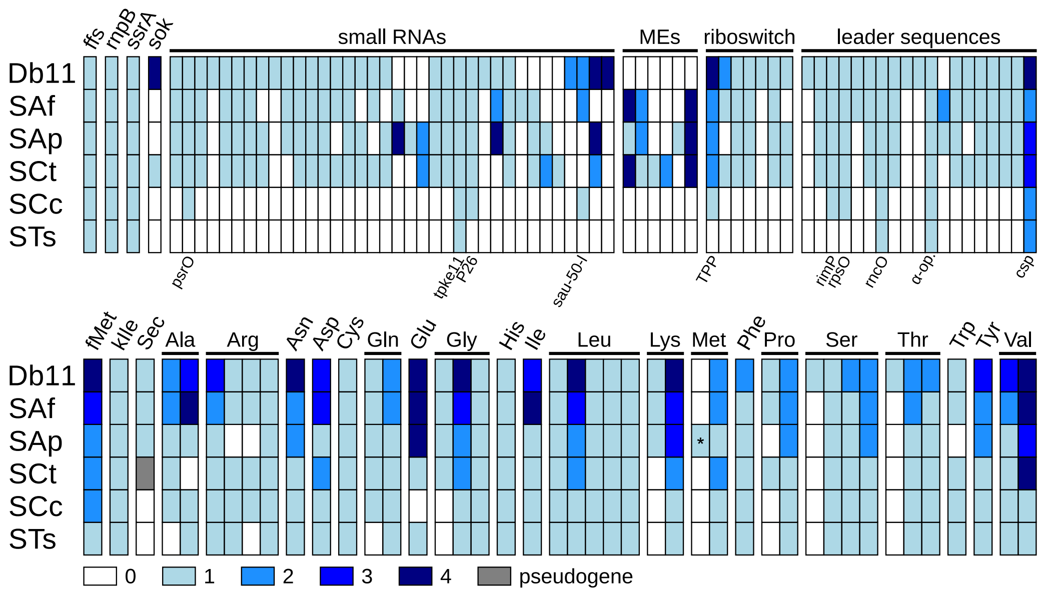


Figure 4.5 Decay of tRNA and other RNA features in *Serratia* genomes. (Top) Colour-coded diagram showing the decay of RNA features in different *S. symbiotica* genomes. On top of the matrix, gene names (for the first four columns) and RNA categories (for the rest) are indicated. On the bottom of the matrix, feature names are indicated for those features retained in SCc and STs. **(Bottom)** Colour-coded diagram showing the decay of tRNA features in different *S. symbiotica* genomes. On the top of the matrix, aminoacyl charging potential for each tRNA species (as inferred by TFAM). Each column represents a different anticodon. Asterisks indicate putative codon reassignments. fMet= *N*-Formylmethionine, kIle= lysylated isoleucine. Reproduced from Manzano-Marín and Latorre (2016).

genome reduction SCc and STs find themselves in, respectively, most of the RNA features have been lost. Conserved features across *S. symbiotica* are the the 4.5S RNA component of the signal recognition particle (SRP) (*ffs*), the RNase P M1 RNA component (*rnpB*), the tmRNA (*ssrA*), the tpke11 small RNA (of unknown function),the leader sequence from the *rnc-era* transcription unit (coding for the ribonuclease 3 and the GTPase Era), and the alpha operon leader (coding for the 30S ribosomal subunits S13, S11, and S4; the 50S ribosomal subunit L17; and the DNA-directed RNA polymerase subunit alpha). The first three are interestingly also retained in other small genomes, but unidentifiable in tiny genomes (Figures S.11), hinting at these being essential functions retained until the last stages of genome reduction. Since most of these RNA features are related to the regulation of gene expression (small antisense RNAs, riboswitches, and leader sequences [including amino acid operon attenuators]), these losses would reflect a general trend of gene-regulation-loss in endosymbiotic genomes through the erosion of RNA features.

Regarding tRNAs, we observed a drastic reduction in tRNA-gene number, particularly marked in SCc and STs (Figure 4.5: bottom panel). These losses, as in other reduced endosymbionts (Figures S.12), mainly affect redundancy rather than variety. Contrasting the other *S. symbiotica* genomes, we were unable to detect a tRNA with aminoacyl charging potential for glutamate in SCc. This is similar to what is observed in other tiny genomes, where some tRNAs with certain aminoacyl charging potential are absent (Figures S.12). However, the presence of a $tRNA^{Glu}$ in a yet-unidentified plasmid cannot be discarded. Also, a loss of the selenocysteine tRNA is already present in the early co-obligate SCt, consistent with the loss of other selenocysteine-related genes, and completely absent in the smaller SCc and STs. It is important to remark that, in SAp, one of the $tRNA^{Met}$ copies has undergone a mutation in its anticodon ($\underline{C}AT \rightarrow \underline{A}AT$), which could theoretically lead to the ATT codon to be recognised as coding for methionine. Finally, the tRNAs for

N-Formylmethionine (in charge of aminoacylation of the starting methionine) and lysylated isoleucine (tRNA^{fMet} and tRNA^{kIle}, respectively) are conserved even in the two smallest *S. symbiotica*. This follows the trend observed in other reduced genomes (Figures S.12) and points towards the essential nature of these.

4.2.4 Informational machinery

By analysing and comparing the informational machinery (ribosome-, transcription-, translation-, and DNA replication/repair-related genes) in *S. symbiotica* strains, both high preservation as well as gradual patterns of deterioration become evident in different categories. The ribosome, as well as the tRNA aminoacylation genes are mostly perfectly preserved (Figure 4.6). Marked differences include the presence of multiple copies

Figure 4.6 Informational machinery in *S. symbiotica* genomes, ribosome and tRNA aminoacylation. Circular plots displaying the different genes involved in ribosome and tRNA aminoacylation. Outer lines in the plot delimit the subcategory. Form outer to inner, the rings in the plot stand for the gene name, the colour-coded lines delimiting categories/complexes, and boxes standing for the presence or absence of the genes in Db11, SAf, SAp, SCt, SCc and STs. Reproduced from Manzano-Marín and Latorre (2016).

of the three rRNA genes in SAf, SAp, and SCt, and the absence of two ribosomal proteins (*rpsI* and *rplM*) and the prolyl tRNA–synthetase gene (*proS*). While the retention of only one copy of the three rRNA genes reflects the tendency of endosymbiotic, and other reduced genomes, to eliminate redundancy (Mendonça *et al.*, 2011), the loss of the *rpsI* gene (coding for the 30S ribosomal subunit S9) reflects the loss of a non-essential gene. In *E. coli*, it has been experimentally proven that a null mutant of the *rplI* gene is able to grow, albeit showing a slow growth phenotype (Bubunenko *et al.*, 2007; Shoji *et al.*, 2011). Most intriguing are the losses of the *rplM* (coding for the 50S ribosomal subunit L13) and *proS* genes. The former has been described as essential in *E. coli* (Shoji *et al.*, 2011), and its loss could be related to the loss of the *rpsI*, that together with *rplM* forms an operon. The latter, could reflect a putative functional replacement of the ProS protein activity by another non-specific aminoacyl tRNA–synthetase, phoenomenon that has been observed for other aminoacyl tRNA–synthetases (Zhang and Hou, 2004) (reviewed in Jacquin-Becker *et al.*, 2002 for Archea; see also Giegé and Springer, 2012).

Both rRNAs and tRNAs undergo a series of modifications that are required to produce the mature version of these ncRNAs (reviewed in Decatur and Fournier, 2002; Hagervall and Björk, 2013). By analysing the genes involved in both rRNA and tRNA modifications, we observed that while the recently derived SAf, SAp, and SCt hold a rather complete set (with particularly marked losses of 23S rRNA methyltransferases), the highly reduced SCc and STs retain only a small fraction of these genes (Figure 4.7). With the notable exceptions of the *fmt*, *tilS*, *trmD*, *tsaB*, *tsaC*, *tsaD*, and *tadA* genes, individual knockout mutants all of these genes (except *miaE*, which is not present in this organism) in *E. coli* have dimmed them non-essential (Baba *et al.*, 2006; Gerdes *et al.*, 2003; Joyce *et al.*, 2006; Purta *et al.*, 2008). The *fmt* and *tilS* genes code for the genes responsible for the attachment of a formyl group to the free amino group of methionyl-tRNA^{fMet} (for initiator methionine) (Kahn *et al.*,

1980) and the modification of the wobble base of the CAU anticodon of the tRNA^{kIle} (Grosjean and Björk, 2004; Muramatsu *et al.*, 1988), respectively. The retention of these two genes thereby insure both the correct charging of the initiator methionine in proteins (which is posttranslationally-removed) and the accurate recognition of the AUA codon as coding for Isoleucine. On the other hand, the genes *tsaB*, *tsaC*, and *tsaD* (along with the *tsaE* gene) are responsible for the biosynthesis of the threonylcarbamoyladenosine (t6A) residue at position 37 of ANN-decoding tRNAs (Deutsch *et al.*, 2012). Interestignly, the *tsaE* gene (an ATPase), which has been found to be non-essential in *E. coli* under anaerobic conditions (Mangat and Brown, 2008), is missing from STs, thus the biosynthesis of t6A would be either putatively impaired or working in an unknown way. Finally, the *tadA* gene codes for a tRNA-specific adenosine deaminase which is essential for viability in *E. coli* (Wolf, 2002).

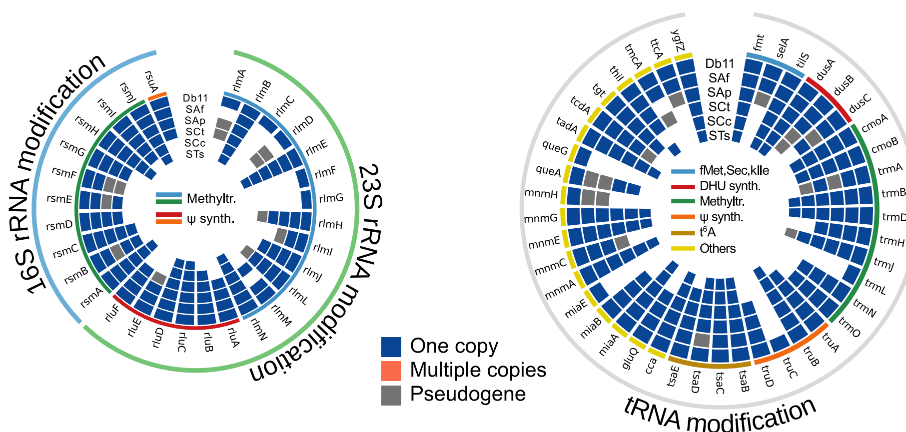


Figure 4.7 Informational machinery in *S. symbiotica* genomes, rRNA and tRNA modification. Circular plots displaying the different genes involved in rRNA and tRNA modification. Outer lines in the plot delimit the subcategory. Form outer to inner, the rings in the plot stand for the gene name, the colour-coded lines delimiting categories/complexes, and boxes standing for the presence or absence of the genes in Db11, SAf, SAp, SCt, SCc and STs. Reproduced from Manzano-Marín and Latorre (2016).

In regards to DNA replication and repair, the gene losses are particularly marked in the most genomically reduced symbionts, SCc and STS, affecting mostly DNA repair-related genes (Figure 4.8: left). This is also been observed in other reduced endosymbionts (see Moran and Bennett, 2014) and is possibly related to the triggering of a more drastic genome erosion (reviewed in Moran *et al.*, 2008). DNA replication-related losses affect the non-essential *holE* gene of the DNA polymerase, the priA-dependent primosome (retaining an elementary DNA-dependent one [missing the auxiliary Hup proteins]), and the *gyrA* subunit of the DNA gyrase. These latter, although identified as essential in *E. coli* (Baba *et al.*, 2006), has also been found to be missing from tiny genomes (Moran and Bennett, 2014), thereby suggesting its function could be taken over by an alternative enzyme or it being non-essential in some endosymbiotic organisms.

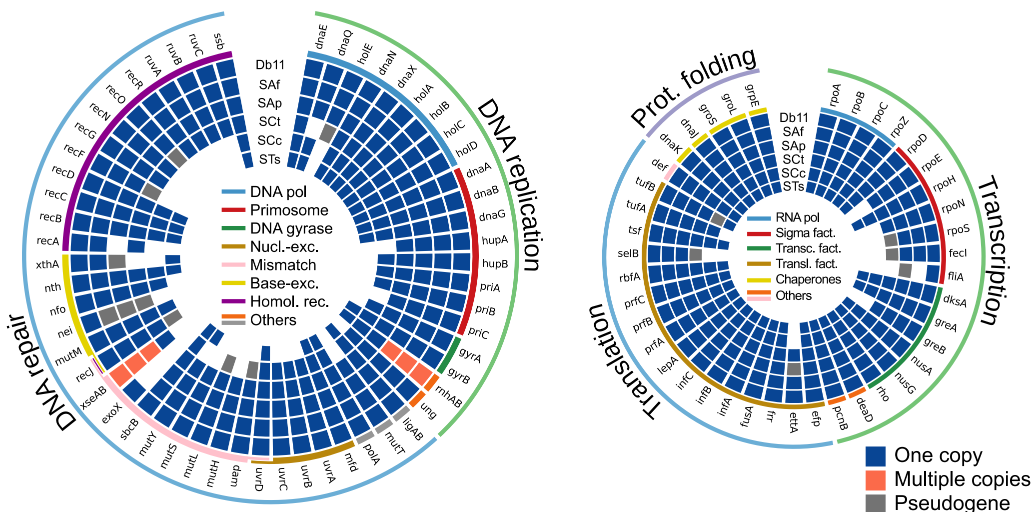


Figure 4.8 Informational machinery in *S. symbiotica* genomes, DNA replication/repair, transcription and translation. Circular plots displaying the different genes involved in DNA replication/repair, transcription and translation. Outer lines in the plot delimit the subcategory. Form outer to inner, the rings in the plot stand for the gene name, the colour-coded lines delimiting categories/complexes, and boxes standing for the presence or absence of the genes in Db11, SAf, SAp, SCt, SCc and STs. Modified from Manzano-Marín and Latorre (2016).

In terms of transcription- and translation-related genes, a high degree of retention in all *S. symbiotica* genomes can be observed (Figure **??**: right). Gene losses mainly affect the sigma factors, with STs retaining only the *rpoD* and *rpoH* genes, coding for σ^{70} and σ^{32}, respectively. While the former is preserved in endosymbionts (Moran and Bennett, 2014), the latter is missing from endosymbionts such as *Blattabacterium* and *Nasuia*. σ^{32} is required for the normal expression of heat shock genes and for the heat shock response through the regulation of the synthesis of heat shock proteins (Grossman *et al.*, 1987), and thus its retention/loss could be specific of certain endosymbiotic systems.

4.2.5 Dwindling genes: stripping proteins down to the bones

Through the manual curation of the annotation of SCt, SCc, and STs endosymbionts (Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.*, 2016), we noted that some genes (*atpC*, *cysJ*, *deaD*, *dnaX*, *ftsN*, *hscA*, *metG*, *pcnB*, *rnr*, and *tolC*) seemed to be shorter in STs, and sometimes consistently shrunken across *S. symbiotica*, compared to those of free-living *E. coli* and even Db11. However, while these genes showed truncated or missing domains, they displayed a high degree of sequence conservation when compared to Db11. Thorough examination of these shrunken genes revealed that experimental evidence, mainly from *E. coli*, have proven that truncated versions of these proteins were able to function with few to none obvious phenotypic consequences (details recorded in the annotation files available from the INSDC). Particularly evident is the loss of non-essential domains in six proteins: AceF, DnaX, FtsK, FtsN, and Rnr (Figure 4.9). The AceF protein (E2 component of pyruvate dehydrogenase complex) has undergone the loss of one or two biotin/lypoyl domains (PF00364) in all *S. symbiotica*, namely STs retains only one. In *E. coli*, it has been shown, through the *in vitro* deletion biotin/lipoyl domains, that one single domain suffices with

respect to enzyme activity and protein function (Guest *et al.*, 1985). The tau subunit of the DNA polymerase III is coded by the *dnaX* gene, however an alternative isoform, denominated gamma subunit, is produced due to a programmed ribosomal frameshifting leading to a premature stop codon in the -1 frame at codon 430 (Tsuchihashi and

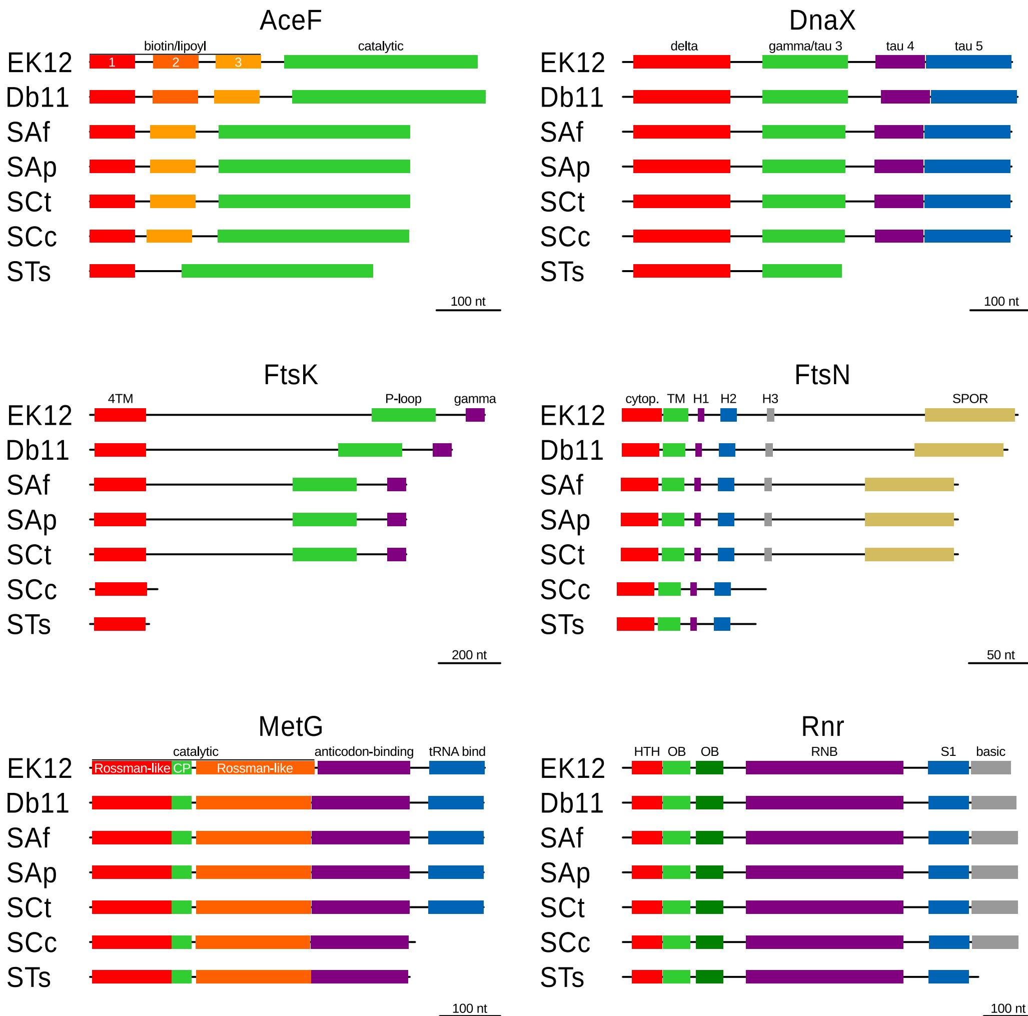


Figure 4.9 Diminution of genes in *S. symbiotica*. Graphic representation of the genes that show evident loss of non-essential domains (as judged by experimental evidence in *E. coli*) in *S. symbiotica*. Domains in each frame are represented by coloured boxes, with similar colours used for repeated domains in each protein. on top of each box, the domain's is provided. Above each frame, the protein's name is stated.On the bottom-right of every frame, a scale bar is provided. Reproduced from Manzano-Marín and Latorre (2016).

Kornberg, 1990). *in vitro* experiments with truncated his shorter isoform, which lacks the tau 4 and 5 domains (PF12168 and PF12170), indicate that gamma is sufficient for replication (Walker *et al.*, 2000). The most drastic gene diminutions are observed in the *ftsK* and *ftsN* genes (whose products are involved in cell division), where SCc and STs preserve only a very small portion of the original gene. Independent *in vivo* experiments in *E. coli* coding only for truncated FtsK (amino acids 1-200) (Draper *et al.*, 1998) or FtsN (amino acids 1-119) (Gerding *et al.*, 2009) proteins, have corroborated that these tiny versions are sufficient for cell division, although short to long filaments were observed to occur. Regarding MetG (Methionyl tRNA–synthetase), both SCc and STs are lacking the C-terminal putative tRNA binding domain (PF01588). Genetic complementation studies and characterization of C-terminally truncated enzymes in *E. coli*, established that MetG can be reduced to 547 residues without significant effect on either the activity or stability of the enzyme (Mellot *et al.*, 1989). Finally, a deletion of the C-terminal basic domain of the Rnr protein (Ribonuclease R) can be observed only in STs. This could lead to an increase in activity of this enzyme, since assays using purified truncated Rnr proteins from mutant *E. coli*, lacking the 83 residues from the C-terminus, were shown to display higher affinity and *circa* 2-fold higher activity than full length wild-type Rnr (on poly[A], A[17] and A[4] substrates) (Vincent and Deutscher, 2009). Through the alignment of the aforementioned putatively-functional proteins against other small and tiny genomes, we corroborated most of these gene diminutions are common among these organisms (Multiple alignment file available on-line). This suggests that selection might favour gene diminution (the retention of only essential domains of a protein), relaxing selective constraints in non-essential gene regions, thus contributing further to genome reduction.

4.3 Conclusion

S. symbiotica strains have been evolving under a similar environment, the aphid-*Buchnera* symbiotic system. We have established that *S. symbiotica* strains can be considered to be along the genome reduction spectrum from a free-living bacterium to a drastically reduced endosymbiont, thus providing "snapshots" of the genome reduction process. SAf would thus represent the very first stages of genome reduction, having not yet lost its ability to grow in axenic culture and having undergone a mild genome shrinkage when compared to the free-living Db11. SAp and SCt would be a stage further down the path, having a more reduced genome than Db11 and showing a massive enrichment in both pseudogenes and MEs. However, SCt has already done the transition to being a co-obligate endosymbiont, and thus shows more drastic gene losses in the EAAs' biosynthetic pathways. SCc and STs find themselves in more advanced stages of genome reduction and integration to their symbiotic systems, having established metabolic complementation with *Buchnera* for the synthesis of several essential compounds. SCc differs greatly from STs in genome size, which is explained by the former being in a very recent stage of an advanced genome erosion, thus retaining several pseudogenes and "junk" DNA. They also show a drastic genome-wide gene loss, and particularly in their ncRNA repertoire and informational machinery. Through the comparison of these *S. symbiotica* strains, we were able to hint at essential retained functions, which not surprisingly are shared with other highly-reduced endosymbionts. Finally, the detailed study of protein diminution revealed a common tendency of endosymbionts to loose non-essential protein domains, and thus constitute an additional route towards genome reduction. We expect the further study of this particular endosymbiont in aphids will continue to provide important clues into the intriguing process of genome reduction.

General conclusions

One crucial question within the field is how do mutualistic associations evolve. One way of approaching this question is through the study of "recently" acquired endosymbionts, particularly via the comparison of endosymbiotic systems from closely-related hosts. In this thesis, we have analysed both molecularly and microscopically the different endosymbiotic systems from several species of aphids belonging to the Lachninae subfamily.

First, we have corroborated that the presence of secondary symbionts across the Lachninae appears to be universal, given that all species analysed both in this study and in previous works have been found to house additional symbionts (other than *Buchnera*). Even though the great majority of these symbionts belong to the *S. symbiotica* species, some of the aphid lineages seem to have undergone symbiont replacement by distantly-related bacteria. Furhtermore, even within *S. symbiotica* endosymbionts, there is a great variety regarding cell shape, tissue tropism, and location within the bacteriome. The presence of *S. symbiotica* across all Lachninae tribes, led us to propose a single ancestral infection by a *S. symbiotica* in the LLCA. This bacterium would have then been fixed within this aphid subfamily due to the pseudogenisation of the genes implicated in the biosynthesis of riboflavin, an essential vitamin. After this event, the *S. symbiotica* lineages would have diverged and even been replaced by other symbiotic species.

Second, the sequencing of the genome from the "early" co-obligate SCt, and its comparison against the facultative SAp, revealed that the transition of *S. symbiotica* from a facultative endosymbiont to a co-obligate one was putatively triggered by a loss-of-function in *Buchnera*, rather than one in *S. symbiotica*. Through the thorough comparison of SCt

and SAp, we identified several pseudogenisations in the genes related to the biosynthesis of EAAs in SCt. This reinforces the hypothesis that this endosymbiont has already started to accomodate to the aphid-*Buchnera* symbiotic system.

Third, the genomic and microscopic analysis of the endosymbiotic system of *Tu. salignus*, revealed a high level of convergence between this system and that of *C.* (*Ci.*) *cedri*. We found that SCc and STs have undergone independent events of genome reduction, and putatively of gene-losses that would have led to metabolic complementation in the biosynthesis of both tryptophan and biotin.

Fourth, through the genomic comparison of the different strains of currently-available *S. symbiotica*, we "dissected" the different stages of genomic reduction each symbiont finds itself in. The genetic losses, the genomic erosion, the gene diminution, and the weathering of the informational machinery observed in *S. symbiotica* is shared with other dwindling endosymbiotic bacteria. This fact suggests that *S. symbiotica* is indeed a good model to study the genome reduction in endosymbionts.

Finally, this work has revealed the dynamic nature of the "recently" established symbiotic relationships within aphids of the Lachninae subfamily, and has revealed key processes involved in the transition from a free-living bacterium to a mutualistic endosymbiotic organism.

Resumen en Español

Introducción

El término "simbiosis" (del griego σύν 'juntos' y βίωσις 'vivir'), en su contexto biológico, fue acuñado por Heinrich Anton de Bary en su obra titulado "Erscheinung der Symbiose". El definió "simbiosis" como "la vida en conjunción de dos organismos disímiles", sin restringirlo a los efectos de la interacción sobre los miemros de ésta. Aunque el término se ha utilizado de manera más restrictiva, la definición de de Bary mantiene como la más inclusiva, teniendo en cuenta todo el espectro desde neutral, a las asociaciones perjudiciales. Dentro de este espectro, se han reconocido 6 grandes categorías de interacciones simbióticas, basadas en las posibles combinaciones de efectos neutrales (0), beneficiosos (+) y perjudiciales (-). El presente trabajo nos centraremos principalmente en un tipo de interacción: el mutualismo (+/+). el término mutualismo está reservado para las relaciones simbióticas donde cada miembro se beneficia de la actividad del otro. Al socio de mayor tamaño se le conce como el hospedero, mientras que el término simbionte se reserva comúnmente para el siocio de menor tamaño.

Las relaciones simbióticas están presentes en diversos organismos a lo largo del árbol evolutivo. Este tipo de relaciones han jugado un papel importante, por ejemplo, en el desarrollo de la célula eucariota. En 1967, la bióloga Lynn Margulis sintetizó la teoría endosimbiótica, en la cual integraba y exponía el conocimiento actual acerca de los organelos celulares y proponía que éstos se habían originado a partir de una simbiosis inicial de la célula protoeucariota con una bacteria aeróbica, en el caso de la mitocondria, y una Cyanobacteria, en el caso del cloroplasto, lo cuál se comprobó a través de análisis filogenéticos.

Como muchos animales, una importante cantidad de insectos mantiene asociaciones simbióticas obligadas con bacterias endosimbiontes. Éstas generalmente viven dentro de bacteriocitos (células especializadas en albergarlas), los cuales forman un órgano llamado el bacterioma. Una característica común de estos organismos endosimbióticos es la de tener un genoma reducido, lo cuál deriva de varios factores: **i)** Su modo vertical de transmisión, en el cual sólo una pequeña cantidad de éstos será heredado en la siguiente generación lo que acentúa el efecto de la deriva genética y el trinquete de Muller asociado; **ii)** La pérdida de genes de recombinación y reparación de ADN; **iii)** La relajación de la selección sobre genes redundantes e innecesarios para su nuevo estilo de vida. Durante este proceso de reducción genómica, los genomas endosimbióticos sufren una serie de cambios drásticos en contenido genético y arquitectura genómica. En los primeros estadíos, los genomas se encuentran altamente enriquecidos en pseudogenes y elementos móviles (**EMs**), principalmente elementos de inserción (**ISs**). En estadíos avanzados de esta reducción genómica, existe una pérdida total de la redundancia genética y de los EMs, acompañada de una pérdida casi total de los pseudogenes. En los casos más extremos, la pérdida de genes en tan drástica que se cuestiona su identidad como organismo celular.

Los áfidos (Hemiptera: Aphididae) mantienen una relación simbiótica obligada con bacterias endosimbiontes del género *Buchnera* (denominada endosimbionte primario), que habita en el citoplasma de bacteriocitos y está contenida dentro de una membrana simbiosomal (derivada de la membrana del hospedero). *Buchnera* provee al áfido de aminoácidos esenciales (**EAAs**), los cuáles son carentes en su estricta dieta basada en floema de planta. Además de *Buchnera*, los áfidos pueden contener endosimbiontes secundarios, los cuales pueden ser obligados o facultativos. Los endosimbiontes facultativos, al contrario de los obligados, no son necesarios para el correcto desarrollo y supervivencia del hospedero, sin embargo, pueden dotarlo de

características ventajosas bajo ciertos estreses ambientales. Por ejemplo, el endosimbionte *Serratia symbiotica* aumenta la tasa de supervivencia de áfidos de la especie *Acyrthosiphon pisum* después de sufrir un golpe de calor. Dentro de la subfamilia Lachninae de áfidos, se han encontrado endosimbiontes secundarios en todos los miembros que han sido analizados por medio de microscopía. Para la mayoría de las especies, este endosimbionte secundario pertenece a la especie *S. symbiotica*. El análisis genómico del endosimbionte *Buchnera* en dos especies de *Cinara* (*C.* (*Ci.*) *cedri* y *C.* (*Cu.*) *tujafilina*) pertenecientes a la subfamilia Lachninae (**BCc** y **BCt**, respectivamente), ha revelado que existe una pérdida común de los genes involucrados en la biosíntesis de riboflavina (vitamina B_2), la cual es esencial para el correcto desarrollo del hospedero. La secuenciación del endosimbionte secundario *S. symbiotica* de *C.* (*Ci.*) *cedri* (**SCc**), reveló que éste era requerido para la biosíntesis, en colaboración con *Buchnera*, de varios compuestos esenciales para el áfido, además de que retenía los genes necesarios para sintetizar riboflavina, subsanando así la pérdida de esta capacidad biosintética en *Buchnera*. Por ello se considera un endosimbionte obligado, junto con *Buchnera*. Asimismo, SCc presentaba un genoma mucho más reducido que el de la cepa facultativa de *S. symbiotica* de *Ac. pisum* (**SAp**), sin embargo, al contrario de otros endosimbiontes altamente reducidos, como *Buchnera*, mantenía una gran cantidad de genoma "basura" y pseudogenes. Por lo tanto, SAp y SCc representan dos estadíos distintos del proceso de reducción genómica.

Objetivos

Esta tesis trata de entender el proceso de adquisición de *S. symbiotica* en diversos miembros de la subfamilia Lachninae de áfidos, así como el proceso de reducción genómica sufrido por *S. symbiotica*. También, se explora la diversidad de endosimbiontes secundarios en Lachninae y se trata de reconstruír la historia evolutiva de estas asociaciones. Para esto

se han definido cuatro objetivos principales:

1. Explorar los cambios genómicos que se producen en las primeras etapas del proceso de acomodamiento, sufrido por un endosimbionte coobligado, al nuevo sistema endosimbiótico. Esto se abordó a través de la secuenciación y ensamblaje del genoma de *S. symbiotica* de *C.* (*Cu.*) *tujafilina* (**SCt**). Elegimos este organismo dado las diferencias histológicas observadas cuando se compara con SCc, las cuáles muestran a SCt más cercana a la facultativa SAp.

2. Contrastar la hipótesis de una adquisición ancestral de *S. symbiotica* como endosimbionte coobligado en los miembros de Lachninae (seguida de divergencia), frente a la sustitución de *S. symbiotica* en la rama que conduce al subgénero *Cinara* (*Cupressobium*). Esto se abordó a través de la comparación de los sistemas endosimbiótico coobligados de *C.* (*Ci.*) *cedri* (BCc y SCc) y *Tuberolachnus salignus* (**BTs** y **STs**), cuyas cepas de *S. symbiotica* pertenecen al mismo clado filogenético con características similares a endosimbiontes obligados.

3. Entender el establecimiento y posteriores putativos reemplazos de *S. symbiotica* dentro de la subfamilia Lachninae. Esto fue abordado a través de la caracterización molecular y análisis microscópico de endosimbiontes secundarios de especies representativas de diferentes clados de Lachninae.

4. Diseccionar el proceso de reducción del genoma de *S. symbiotica* y contrastar los resultados con otros endosimbiontes de genoma reducido, evaluando de esta manera la generalidad de las observaciones en *S. symbiotica*. Esto se abordó a través de la comparación de *Serratia marcescens* Db11, cepa de vida libre, contra las diferentes cepas de *S. symbiotica* aisladas de diferentes áfidos de las subfamilias Aphidinae y Lachninae.

Metodología y Resultados

Capítulo 1. El genoma de *Serratia symbiotica*, endosymbionte coobligado "reciente" del áfido *Cinara* (*Cupressobium*) *tujafilina*

Son particularmente interesantes los casos de endosimbiosis mutualista en la cual interviene más de un simbionte. Durante el proceso de adaptación que sufre una bacteria de vida libre en su camino a convertirse en un endosimbionte intracellular obligado, el simbionte experimenta pérdidas genómicas importantes y ajustes fenotípicos. A partir de las observaciones microscópicas y filogenéticas realizadas con anterioridad, se propuso que el endosimbionte secundario *S. symbiotica* de *C.* (*Cu.*) *tujafilina* podría haberse adaptado recientemente al sistema áfido-*Buchnera*, debido a su cercanía molecular e histológica a SAp. Por medio de una combinación de técnicas de secuenciación masiva (454 FLX titanium e Illumina HiSeq2000), hemos secuenciado y ensamblado el genoma de *S. symbiotica* de *C.* (*Cu.*) *tujafilina*, el cuál se esperaba se encontrara en un estadío muy temprano de adaptación como endosimbionte coobligado junto con el consorcio áfido-*Buchnera*. A continuación, se escrutaron los cambios en la coevolución de SCt y BCt, prestando especial atención a las transformaciones experimentadas por SCt, en comparación con SAp, para convertirse en un simbionte obligado. A pesar de que SCt está filogenética y genómicamente muy estrechamente relacionado con el endosimbionte facultativo SAp, muestra una variedad de alteraciones metabólicas, genéticas y de arquitectura genómica, que sugieren que este endosimbionte está un paso más cerca de convertirse en un endosimbionte intracelular coobligado que SAp. Utilizando técnicas de genómica comparativa, hemos descrito en profundidad el proceso de reorganización genómica sufrido por SCt y SAp, así como el papel que han jugado los MEs en

éste. Mediante la reconstrucción metabólica de SCt, hemos determinado que la pérdida de la capacidad biosintética de la riboflavina en BCt, postula a SCt como el provisor de este compuesto esencial. Finalmente, postulamos que la pérdida de la capacidad biosintética en la producción de riboflavina en *Buchnera*, ha sido la clave para la el establecimiento de *S. symbiotica* como endosimbionte coobligado en los áfidos que pertenecientes a la subfamilia Lachninane.

Capítulo 2. Convergencia evolutiva entre los sistemas endosimbióticos coobligados de los áfidos *Tuberolachnus salignus* y *Cinara* (*Cinara*) *cedri*

Como mencionado anteriormente, muchas especies de la subfamilia Lachninae parecen estar asociadas consistentemente con el endosimbionte secundario *S. symbiotica*. Hemos demostrado anteriormente que tanto *C.* (*Ci.*) *cedri* y *C.* (*Cu.*) *tujafilina* (Lachninae de la tribu Eulachnini) han establecido asociaciones de carácter coobligado tanto con *Buchnera* como con *S. symbiotica*. Sin embargo, mientras que los genomas de *Buchnera* de ambas especies de *Cinara* son genómica y funcionalmente muy similares, existe una degradación diferencial importante entre los genomas de sus correspondientes cepas de *S. symbiotica*. Para entender mejor la esencialidad y el grado de integración de *S. symbiotica* dentro de Lachninae, hemos secuenciado el genoma de ambos endosimbiontes, *Buchnera* y *S. symbiotica* del áfido *Tu. salignus* (Lachninae de la tribu Tuberolachnini) mediante técnicas de secuenciación masiva (Illumina HiSeq2000). A través de la genómica comparativa, hemos encontrado un sorprendente nivel de similitud entre el sistema endosimbiótico de este áfido y el de *C.* (*Ci.*) *cedri*. En ambos hospederos, *S. symbiotica* ha evolucionado un genoma muy reducido y, a través de microscopía de hibridación fluorescente *in situ* (**FISH**), hemos determinado que en ambas especies de áfidos el endosimbionte secundario se encuentra exclusivamente en el interior del

citoplasma de bacteriocitos. Mediante la reconstrucción metabólica, hemos encontrado que, curiosamente, los endosimbiontes de *Tu. salignus* presentan la misma complementación metabólica observada en *C.* (*Ci.*) *cedri* para la producción de triptófano, la cual no se observa en el sistema endosimbiótico coobligado de *C.* (*Cu.*) *tujafilina*. Por otra parte, hemos corroborado que en cuanto a la biosíntesis de riboflavina, BTs, al igual que BCc y BCt, ha perdido la capacidad de sintetizar esta vitamina y STs ahora realiza esta función, de este modo, proporcionando más apoyo a la hipótesis del establecimiento de un endosimbionte coobligado en el ancestro común de los áfidos de la subfamilia Lachninae (**LLCA**). Por último, hemos propuesto que la putativa división convergente de la biosíntesis de triptófano entre *Buchnera* y *S. symbiotica*, podría estar detrás de la internalización obligada de *S. symbiotica* al citoplasma de bacteriocitos propios y el desencadenamiento de una mayor degradación genómica.

Capítulo 3. Identidad y localización de endosimbiontes secundarios de la subfamilia Lachninae de áfidos

Como ya se ha indicado, la mayoría de los áfidos de la subfamilia Lachninae se han encontrado consistentemente albergando endosimbiontes secundarios, principalmente *S. symbiotica*. Esta aparente dependencia de endosimbiontes secundarios parece haber sido provocada por la pérdida de la capacidad de biosíntesis de riboflavina por *Buchnera* en el LLCA. Sin embargo, ningún análisis integral a gran escala de endosimbiontes secundarios en los Lachninae se ha realizado hasta la fecha, lo que dificulta la interpretación de los análisis evolutivos y genómicas de estos endosimbiontes. Por esta razón, hemos analizado los endosimbiontes de diferentes especies lejanas pertenecientes a siete diferentes géneros de Lachninae, que abarcan cuatro tribus, tanto por FISH (exploración de la morfología de los simbiontes y el tropismo celular) y la secuenciación del gen del 16S

rRNA. Hemos corroborado que todos los áfidos analizados poseen sistemas endosimbióticos duales, y aunque la mayoría alberga *S. symbiotica*, algunos han sufrido reemplazamiento del simbionte secundario por otros taxones bacterianos filogenéticamente distintos. Hemos determinamos que estos endosimbiontes secundarios muestran formas de células y el tropismo celulares contrastantes, y algunos parecen ser dependientes de linaje. Ello nos ha permitido proponer un escenario evolutivo para el establecimiento de un endosimbionte secundario ancestral en los Lachninae, seguido de internalizaciones independientes a bacteriocitos así como de reemplazamientos.

Capítulo 4. Reducción genómica en *Serratia symbiotica*

La reducción genómica es un fenómeno generalizado entre los organismos endosimbióticos heredados verticalmente, desde los intracelulares asociados a bacteriocitos hasta los extracelulares asociados al intestino. Esta erosión genómica es un proceso gradual en el cual los organismos de vida libre evolucionan para convertirse en endosimbiontes obligados, perdiendo genes y/o funciones no esenciales o redundantes. *S. symbiotica*, muestra varias características que lo hacen un excelente organismo modelo para el estudio de la reducción genómica. Mientras que algunas cepas son de carácter facultativo para el hospedero, otras han establecido asociaciones coobligado con sus respectivas especies de áfidos y su correspondiente endosimbionte primario (*Buchnera*). Además, las diferentes cepas tienen genomas de tamaños y características muy contrastantes, y además presentan un tropismo celular, formas y tamaños celulares sorprendentemente dispares. Por último, genomas pertenecientes a *S. marcescens*, bacterias de vida libre estrechamente relacionadas a *S. symbiotica*, también están disponibles para realizar análisis de genómica comparativa. A través de la reanotación genómica de genes involucrados en procesos de información (replicación y reparación de

ADN, ribosoma, modificación de ribosomas y tRNAs) y de genes de RNA, hemos diseccionado el proceso de reducción genómica en *S. symbiotica*. Asimismo, la anotación de estos mismos genes en otros organismos de genoma reducido ha corroborado que los patrones de reducción observados en *S. symbiotica* son compartidos con otros genomas menguantes. Esto demuestra que *S. symbiotica* es un buen modelo para el estudio del proceso de reducción genómica de un solo taxón bacteriano (*S. symbiotica*) que evoluciona en un nicho biológico similar (áfido-*Buchnera*).

Conclusiones Generales

Una cuestión clave en la biología evolutiva es la de cómo las asociaciones mutualista evolucionan. Una forma de abordar este problema es investigar las asociaciones mutualistas de reciente creación, en particular mediante la comparación de diferentes sistemas simbióticos en huéspedes estrechamente relacionados. En el presente trabajo hemos analizado tanto molecular como microscópicamente los sistemas endosimbióticos de varias especies de la subfamilia Lachninae.

Primero, hemos corroborado que la presencia de endosimbiontes secundarios en los Lachninae es aparentemente universal, ya que todos los miembros analizados por microscopía tanto en este trabajo como en previos estudios han revelado que contienen endosimbiontes adicionales a *Buchnera*. A pesar de que la mayoría de estos endosimbiontes pertenecen a la especie *S. symbiotica*, algunos linajes han sufrido reemplazamientos genómicos por bacterias pertenecientes a diferentes taxones. Asimismo, incluso dentro de las *S. symbiotica*, existe una gran variedad en tipo de relación, morfología celular y localización dentro del bacterioma. La presencia de *S. symbiotica* en especies de Lachninae pertenecientes a diferentes tribus lejanamente relacionadas, nos lleva a proponer una infección ancestral por *S. symbiotica* en el LLCA, la cual

fue fijada por la pseudogenización de los genes implicados en la biosíntesis de riboflavina, seguida de divergencia y reemplazamiento de simbionte.

Segundo, la secuenciación del genoma de SCt y su comparación contra la cepa facultativa SAp, ha revelado la transición de *S. symbiotica* de endosimbionte facultativo a obligado, desencadenada no por un cambio en *S. symbiotica*, sino por una pérdida en la capacidad biosintética de *Buchnera* para producir la riboflavina, una coenzyma esencial. Mediante la comparación minuciosa de los genomas de SCt y SAp, hemos encontrado varias pseudogenizaciones de genes relacionados a la síntesis de EAAs en SCt, lo cuál refuerza la hipótesis de que este endosimbionte ya ha comenzado el proceso de adaptación al sistema endosimbiótico del áfido y *Buchnera* en *C.* (*Cu.*) *tujafilina*.

Tercero, el análisis genómico y microscópico del sistema endosimbiótico de *Tu. salignus* reveló un gran nivel de convergencia evolutiva entre este sistema y el de *C.* (*Ci.*) *cedri*. Hemos encontrado que SCc y STs han sufrido eventos independientes de reducción genómica, y putativamente del establecimiento de complementación metabólica para la síntesis del triptófano y la biotina.

Cuarto, mediante la comparación genómica de las diferentes cepas de *S. symbiotica*, hemos diseccionado los distintos estadíos de reducción genómica en los que se encuentran las distintas cepas de *S. symbiotica*. Las pérdidas génicas, la erosión genómica y la reducción de genes y maquinaria informacional, son compartidas por otras bacterias endosimbióticas reducidas, lo cual apunta a la generalidad de estas observaciones y valida a *S. symbiotica* como un buen modelo para el estudio de la reducción genómica en endosimbiontes.

Finalmente, este trabajo ha revelado la naturaleza dinámica de las relaciones establecidas "recientemente" en áfidos de la subfamilia Lachninae, y ha revelado procesos claves de la transición de un

organismo de vida libre a uno endosimbiótico.

Bibliography

Altschul S. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**(17):3389–3402. URL http://www.ncbi.nlm.nih.gov/pubmed/9254694http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/25.17.3389. M.2.4

Ansorge J. 1996. *Insekten aus dem oberen Lias von Grimmen (Vorpommern, Norddeutschland)*, volume 2 of *Neue paläontologische Abhandlungen*. CPress, Dresden. ISBN 9783931689001. I.2.1

Ardell DH. 2006. TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. *Nucleic Acids Research*, **34**(3):893–904. URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkj449. M.2.4

Arneodo JD and Ortego J. 2014. Exploring the Bacterial Microbiota Associated With Native South American Species of *Aphis* (Hemiptera: Aphididae). *Environmental Entomology*, **43**(3):589–594. URL http://ee.oxfordjournals.org/lookup/doi/10.1603/EN13324. I.3.3

Attardo GM, Lohs C, Heddi A, Alam UH, Yildirim S, and Aksoy S. 2008. Analysis of milk gland structure and function in *Glossina morsitans*: Milk protein production, symbiont populations and fecundity. *Journal of Insect Physiology*, **54**(8):1236–1242. URL http://linkinghub.elsevier.com/retrieve/pii/S0022191008001212. M.4

Augustinos AA, *et al.* 2011. Detection and characterization of *Wolbachia* infections in natural populations of aphids: is the hidden diversity fully unraveled? *PloS one*, **6**(12):e28695. URL http://dx.plos.org/10.1371/journal.pone.0028695. I.3.2

Baba T, *et al.* 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, **2**(1):2006.0008. URL http://msb.embopress.org/cgi/doi/10.1038/msb4100050. 4.2.4, 4.2.4

Baca OG and Paretsky D. 1983. Q fever and *Coxiella burnetii*: a model for host-parasite interactions. *Microbiol. Mol. Biol. Rev.*, **47**(2):127–149. URL http://mmbr.asm.org/content/47/2/127.citation. I.1.4

Balibar CJ, Hollis-Symynkywicz MF, and Tao J. 2011. Pantethine Rescues Phosphopantothenoylcysteine Synthetase and Phosphopantothenoylcysteine Decarboxylase Deficiency in *Escherichia coli* but Not in *Pseudomonas aeruginosa*.

Journal of Bacteriology, **193**(13):3304–3312. URL http://jb.asm.org/content/193/13/3304.abstracthttp://jb.asm.org/cgi/doi/10.1128/JB.00334-11. 2.2.4

Baumann P, Baumann L, Lai CY, Rouhbakhsh D, Moran NA, and Clark MA. 1995. Genetics, Physiology, and Evolutionary Relationships of the Genus *Buchnera*: Intracellular Symbionts of Aphids. *Annual Review of Microbiology*, **49**(1):55–94. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.mi.49.100195.000415. 2.2.3

Belda E, Moya A, Bentley S, and Silva FJ. 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**(1):449. URL http://www.biomedcentral.com/1471-2164/11/449. 1.2.4

Bennett GM and Moran NA. 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. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt118. 4.2.1

Blackman RL and Eastop VF. 1994. *Aphids on the World's Trees: An Identification and Information Guide*. CAB International, Wallingford, UK, 1st edition. ISBN 978-0-851-98877-1. URL http://www.cabi.org/bookshop/book/9780851988771. I.2.1, I.2.3, 2.1

Blackman RL and Eastop VF. 2006. *Aphids on the World's Herbaceous Plants and Shrubs*. John Wiley & Sons, Ltd; Natural History Museum (London), Chichester, UK, 1st edition. ISBN 978-0-471-48973-3. URL http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471489735.html. 3.2.1

Blackman RL, De Boise E, and Czylok A. 2001. Occurrence of sexual morphs in *Trama troglodytes* von Heyden, 1837 (Hemiptera, Aphididae). *Journal of Natural History*, **35**(5):779–785. URL http://www.tandfonline.com/doi/abs/10.1080/00222930152023108. I.2.3

Boetzer M and Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biology*, **13**(6):R56. URL http://genomebiology.biomedcentral.com/articles/10.1186/gb-2012-13-6-r56. M.2.2

Boetzer M, Henkel CV, Jansen HJ, Butler D, and Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, **27**(4):578–579. URL http://bioinformatics.oxfordjournals.org/cgi/content/abstract/27/4/578http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btq683. M.2.2

Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, and Madden TL. 2012. Domain enhanced lookup time accelerated BLAST. *Biology Direct*, **7**(1):12. URL http://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-7-12. M.2.4

Bourque G and Pevzner PA. 2002. Genome-scale evolution: Reconstructing gene orders in the ancestral species. *Genome Research*, **12**(1):26–36. URL http://genome.cshlp.org/content/12/1/26.abstract. M.5

Brady A and Salzberg S. 2011. PhymmBL expanded: confidence scores, custom databases, parallelization and more. *Nature Methods*, **8**(5):367–367. URL http://www.nature.com/doifinder/10.1038/nmeth0511-367. M.2.2

Bubunenko M, Baker T, and Court DL. 2007. Essentiality of ribosomal and transcription antitermination proteins analyzed by systematic gene replacement in *Escherichia coli*. *Journal of Bacteriology*, **189**(7):2844–2853. URL http://jb.asm.org/cgi/doi/10.1128/JB.01713-06. 4.2.4

Buchner P. 1953. *Endosymbiose der Tiere mit Pflanzlichen Mikroorganismen*. Birkhäuser Basel, Basel. ISBN 978-3-0348-6959-1. URL http://link.springer.com/10.1007/978-3-0348-6958-4. I.1.3, I.3, 3.1, 3.1, 3.2.2, S.9

Burke GR and Moran NA. 2011. Massive Genomic Decay in *Serratia symbiotica*, a Recently Evolved Symbiont of Aphids. *Genome Biology and Evolution*, **3**(0):195–208. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evr002. I.3.2, I.3.3, 1.2.2, 1.2.3, 1.2.5, 3.2.2, 4.1

Burke GR, Normark BB, Favret C, and Moran NA. 2009. Evolution and diversity of facultative symbionts from the aphid subfamily Lachninae. *Applied and Environmental Microbiology*, **75**(16):5328–5335. URL http://aem.asm.org/cgi/doi/10.1128/AEM.00717-09. I.3.3, I.3.3, 1.2.4, 1.3, 2.3, 3.1, 3.2.1, 3.2.1, 3.2.1, 3.2.1, 3.2.2

Canonaco F, Hess TA, Heri S, Wang T, Szyperski T, and Sauer U. 2001. Metabolic flux response to phosphoglucose isomerase knock-out in *Escherichia coli* and impact of overexpression of the soluble transhydrogenase UdhA. *FEMS Microbiology Letters*, **204**(2):247–252. URL http://femsle.oxfordjournals.org/cgi/doi/10.1111/j.1574-6968.2001.tb10892.x. 2.2.4

Caspi R, *et al.* 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, **42**(D1):D459–D471. URL http://nar.oxfordjournals.org/content/42/D1/D459.longhttp://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkt1103. M.2.4

Chandler SM, Wilkinson TL, and Douglas AE. 2008. Impact of plant nutrients on the relationship between a herbivorous insect and its symbiotic bacteria. *Proceedings. Biological sciences / The Royal Society*, **275**(1634):565–70. URL http://rspb.royalsocietypublishing.org/content/275/1634/565. 4.1

Chari A, Oakeson KF, Enomoto S, Jackson DG, Fisher MA, and Dale C. 2015. Phenotypic characterization 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**(Pt 5):1400–1405. URL http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.000091. 1.2.4

Charusanti P, *et al.* 2010. Genetic basis of growth adaptation of *Escherichia coli* after deletion of *pgi*, a major metabolic gene. *PLoS genetics*, **6**(11):e1001186. URL http://dx.plos.org/10.1371/journal.pgen.1001186. 2.2.4

Chen DQ and Purcell AH. 1997. Occurrence and Transmission of Facultative Endosymbionts in Aphids. *Current Microbiology*, **34**(4):220–225. URL http://link.springer.com/10.1007/s002849900172. I.3.3, 4.1

Chen DQ, Campbell BC, and Purcell AH. 1996. A New *Rickettsia* from a Herbivorous Insect, the Pea Aphid *Acyrthosiphon pisum* (Harris). *Current Microbiology*, **33**(2):123–128. URL http://link.springer.com/10.1007/s002849900086. I.3.2

Chen DQ, Montllor CB, and Purcell AH. 2000. Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid, *A. kondoi*. *Entomologia Experimentalis et Applicata*, **95**(3):315–323. URL http://doi.wiley.com/10.1046/j.1570-7458.2000.00670.x. I.3.2, I.3.3, 4.1

Chen F, Mackey AJ, Vermunt JK, and Roos DS. 2007. Assessing Performance of Orthology Detection Strategies Applied to Eukaryotic Genomes. *PLoS ONE*, **2**(4):e383. URL http://dx.plos.org/10.1371/journal.pone.0000383. M.5

Chen R, Favret C, Jiang L, Wang Z, and Qiao G. 2015a. An aphid lineage maintains a bark-feeding niche while switching to and diversifying on conifers. *Cladistics*, page In Press. URL http://doi.wiley.com/10.1111/cla.12141. I.2.3, I.2.3, I.5, 3.1, 3.4, 3.2.1, 3.8

Chen R, Wang Z, Chen J, and Qiao GX. 2015b. Avoidance and Potential Remedy Solutions of Chimeras in Reconstructing the Phylogeny of Aphids Using the 16S rRNA Gene of *Buchnera*: A Case in Lachninae (Hemiptera). *International Journal of Molecular Sciences*, **16**(9):20152–20167. URL http://www.mdpi.com/1422-0067/16/9/20152/. I.3.3, I.3.3, 1.3, 2.3, 3.2.2

Cheng D and Hou R. 2001. Histological observations on transovarial transmission of a yeast-like symbiote in *Nilaparvata lugens* Stal (Homoptera, Delphacidae). *Tissue and Cell*, **33**(3):273–279. URL http://www.sciencedirect.com/science/article/pii/S0040816601901739. I.1.3

Chevreux B, Wetter T, and Suhai S. 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. In *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, 99, pages 45–46. Heidelberg, Germany. URL http://www.bioinfo.de/isb/gcb99/talks/chevreux/. M.2.2

Clark MA, Moran NA, and Baumann P. 1999. Sequence evolution in bacterial endosymbionts having extreme base compositions. *Molecular Biology and Evolution*, **16**(11):1586–1598. URL http://mbe.oxfordjournals.org/cgi/doi/10.1093/oxfordjournals.molbev.a026071. M.6, 2.2.2

Clayton AL, *et al.* 2012. A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS genetics*, **8**(11):e1002990. URL http://dx.plos.org/10.1371/journal.pgen.1002990. 1.2.4, 4.1, 4.2.1, 4.2.1

Collins CM and Leather SR. 2001. Effect of temperature on fecundity and development of the Giant Willow Aphid, *Tuberolachnus salignus* (Sternorrhyncha: Aphididae). *European Journal of Entomology*, **98**(2):177–182. URL http://www.eje.cz/doi/10.14411/eje.2001.033.html. 2.1

Dadd R and Mittler T. 1965. Studies on the artificial feeding of the aphid *Myzus persicae* (Sulzer)—III. Some major nutritional requirements. *Journal of Insect Physiology*, **11**(6):717–743. URL http://linkinghub.elsevier.com/retrieve/pii/002219106590154X. I.3

Dadd R, Krieger D, and Mittler T. 1967. Studies on the artificial feeding of the aphid *Myzus persicae* (Sulzer)—IV. Requirements for water-soluble vitamins and ascorbic acid. *Journal of Insect Physiology*, **13**(2):249–272. URL http://linkinghub.elsevier.com/retrieve/pii/0022191067901527. I.3

de Bary HA. 1879. *Erscheinung der Symbiose*. Karl J. Trubner, Strassburg. URL http://catalog.hathitrust.org/Record/009086346. I.1.1

Decatur WA and Fournier MJ. 2002. rRNA modifications and ribosome function. *Trends in Biochemical Sciences*, **27**(7):344–351. URL http://linkinghub.elsevier.com/retrieve/pii/S0968000402021096. 4.2.4

Degnan P, Lazarus A, Brock C, and Wernegreen JJ. 2004. Host–Symbiont Stability and Fast Evolutionary Rates in an Ant–Bacterium Association: Cospeciation of *Camponotus* Species and Their Endosymbionts, *Candidatus* Blochmannia. *Systematic Biology*, **53**(1):95–110. URL http://sysbio.oxfordjournals.org/cgi/doi/10.1080/10635150490264842. 2.2.3

Degnan PH, *et al.* 2009a. Dynamics of genome evolution in facultative symbionts of aphids. *Environmental Microbiology*, **12**(8):2060–2069. URL http://doi.wiley.com/10.1111/j.1462-2920.2009.02085.x. I.3.2, 1.2.2, 3.1, 4.2.1

Degnan PH, Yu Y, Sisneros N, Wing RA, and Moran NA. 2009b. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proceedings of the National Academy of Sciences*, **106**(22):9063–9068. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0900194106. I.3.2, 1.2.2, 4.1, 4.2.1

Degnan PH, Bittleston LS, Hansen AK, Sabree ZL, Moran NA, and Almeida RPP. 2011. Origin and Examination of a Leafhopper Facultative Endosymbiont. *Current Microbiology*, **62**(5):1565–1572. URL http://link.springer.com/10.1007/s00284-011-9893-5. I.3.2

Deutsch C, El Yacoubi B, de Crécy-Lagard V, and Iwata-Reuyl D. 2012. Biosynthesis of Threonylcarbamoyl Adenosine (t 6 A), a Universal tRNA Nucleoside. *Journal of Biological Chemistry*, **287**(17):13666–13673. URL http://www.jbc.org/lookup/doi/10.1074/jbc.M112.344028. 4.2.4

Dixon AFG. 1985. *Aphid Ecology An optimization approach*. Springer Netherlands, Dordrecht. ISBN 978-94-010-6480-4. URL http://link.springer.com/10.1007/978-94-011-5868-8. I.2.1

Douglas AE. 1989. Mycetocyte Symbiosis in Insects. *Biological Reviews*, **64**(4):409–434. URL http://doi.wiley.com/10.1111/j.1469-185X.1989.tb00682.x. I.3.1

Douglas AE. 1990. Nutritional interactions between *Myzus persicae* and its symbionts. In Campbell RK and Eikenbary RD, editors, *International Congress of Entomology*, pages 319–327. Elsevier, Amsterdam, The Netherlands. I.3.1

Douglas AE. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annu Rev Entomol*, **43**(1):17–37. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.ento.43.1.17. I.1.2

Douglas AE. 2006. Phloem-sap feeding by animals: problems and solutions. *Journal of Experimental Botany*, **57**(4):747–754. URL http://jxb.oxfordjournals.org/lookup/doi/10.1093/jxb/erj067. I.2.1

Douglas AE and Prosser WA. 1992. Synthesis of the essential amino acid tryptophan in the pea aphid (*Acyrthosiphon pisum*) symbiosis. *Journal of Insect Physiology*, **38**(8):565–568. URL http://linkinghub.elsevier.com/retrieve/pii/0022191092901070. 2.2.3

Draper GC, McLennan N, Begg K, Masters M, and Donachie WD. 1998. Only the N-terminal domain of FtsK functions in cell division. *Journal of bacteriology*, **180**(17):4621–7. URL http://jb.asm.org/content/180/17/4621.abstract. 4.2.5

Durak R and Borowiak-Sobkowiak B. 2013. Influence of temperature on the biological parameters of the anholocyclic species *Cinara tujafilina* (Hemiptera: Aphidoidea). *Central European Journal of Biology*, **8**(6):570–577. URL http://link.springer.com/10.2478/s11535-013-0161-x. 1.1

Durak R and Durak T. 2015. Redescription of males of the aphid species *Cinara* (*Cupressobium*) *tujafilina* and *Cinara* (*Cupressobium*) *cupressi* (Hemiptera, Lachninae). *Zootaxa*, **4032**(2):209. URL http://biotaxa.org/Zootaxa/article/view/zootaxa.4032.2.7. 1.1

Ehrhardt P. 1968. Der Vitaminbedarf einer siebröhrensaugenden Aphide, *Neomyzus circumflexus* Buckt. (Homoptera, Insecta). *Zeitschrift für Vergleichende Physiologie*, **60**(4):416–426. URL http://link.springer.com/10.1007/BF00297936. I.3

Engel P and Moran NA. 2013. The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, **37**(5):699–735. URL http://femsre.oxfordjournals.org/lookup/doi/10.1111/1574-6976.12025. I.1.3

Favret C. 2016. Aphid Species File. URL http://aphid.speciesfile.org. I.2.1, I.2.2, I.2.3

Ferrari J, Scarborough CL, and Godfray HCJ. 2007. Genetic Variation in the Effect of a Facultative Symbiont on Host-Plant Use by Pea Aphids. *Oecologia*, **153**(2):323–329. URL http://www.jstor.org/stable/40210868. I.3.2

Flyg C, Kenne K, and Boman HG. 1980. Insect Pathogenic Properties of *Serratia marcescens*: Phage-resistant Mutants with a Decreased Resistance to Cecropia Immunity and a Decreased Virulence to *Drosophila*. *Microbiology*, **120**(1):173–181. URL http://mic.sgmjournals.org/cgi/content/abstract/120/1/173http://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-120-1-173. 4.1

Foray V, *et al.* 2014. Whole-Genome Sequence of *Serratia symbiotica* Strain CWBI-2.3T, a Free-Living Symbiont of the Black Bean Aphid *Aphis fabae*. *Genome*

Announcements, **2**(4):e00767–14. URL http://genomea.asm.org/cgi/doi/10.1128/genomeA.00767-14. I.3.3, 4.1

Fukatsu T. 2001. Secondary intracellular symbiotic bacteria in aphids of the genus *Yamatocallis* (Homoptera: Aphididae: Drepanosiphinae). *Applied and Environmental Microbiology*, **67**(11):5315–20. URL http://aem.asm.org/cgi/doi/10.1128/AEM.67.11.5315-5320.2001. 3.1, 3.3

Fukatsu T and Ishikawa H. 1992. A novel eukaryotic extracellular symbiont in an aphid, *Astegopteryx styraci* (Homoptera, Aphididae, Hormaphidinae). *Journal of Insect Physiology*, **38**(10):765–773. URL http://linkinghub.elsevier.com/retrieve/pii/002219109290029D. 3.1

Fukatsu T and Ishikawa H. 1993. Occurrence of Chaperonin 60 and Chaperonin 10 in primary and secondary bacterial symbionts of aphids: Implications for the evolution of an endosymbiotic system in aphids. *Journal of Molecular Evolution*, **36**(6):568–577. URL http://link.springer.com/10.1007/BF00556361. 3.2.2, S.9

Fukatsu T and Ishikawa H. 1998. Differential immunohistochemical visualization of the primary and secondary intracellular symbiotic bacteria of aphids. *Applied Entomology and Zoology*, **33**(2):321–326. URL https://www.jstage.jst.go.jp/article/aez1966/33/2/33{_}2{_}321/{_}article. 3.1, 3.2.2

Fukatsu T and Nikoh N. 1998. Two intracellular symbiotic bacteria from the mulberry psyllid *Anomoneura mori* (Insecta, Homoptera). *Applied and environmental microbiology*, **64**(10):3599–606. URL http://aem.asm.org/content/64/10/3599. M.4, 3.2.1

Fukatsu T, Aoki S, Kurosu U, and Ishikawa H. 1994. Phylogeny of *Cerataphidini* Aphids Revealed by Their Symbiotic Microorganisms and Basic Structure of Their Galls - Implications for Host-Symbiont Coevolution and Evolution of Sterile Soldier Castes. *Zoological Science*, **11**(4):613–623. URL <GotoISI>://A1994PL57400016. I.1.3, 3.1

Fukatsu T, Watanabe K, and Sekiguchi Y. 1998. Specific detection of intracellular symbiotic bacteria of aphids by oligonucleotide-probed *in situ* hybridization. *Applied Entomology and Zoology*, **33**(3):461–472. URL https://www.jstage.jst.go.jp/article/aez1966/33/3/33{_}3{_}461/{_}article. 3.1, 3.1, 3.2.1, 3.2.2, S.9

Fukatsu T, Nikoh N, Kawai R, and Koga R. 2000. The Secondary Endosymbiotic Bacterium of the Pea Aphid *Acyrthosiphon pisum* (Insecta: Homoptera). *Applied and Environmental Microbiology*, **66**(7):2748–2758. URL http://aem.asm.org/cgi/doi/10.1128/AEM.66.7.2748-2758.2000. I.3.2, I.3.3, 1.2.1, 2.2.3, 3.1, S.2

Fukatsu T, Tsuchida T, Nikoh N, and Koga R. 2001. *Spiroplasma* Symbiont of the Pea Aphid, *Acyrthosiphon pisum* (Insecta: Homoptera). *Applied and Environmental Microbiology*, **67**(3):1284–1291. URL http://aem.asm.org/cgi/doi/10.1128/AEM.67.3.1284-1291.2001. I.3.2

Funk DJ, Helbling L, Wernegreen JJ, and Moran NA. 2000. Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proceedings of the Royal Society B: Biological Sciences*, **267**(1461):2517–2521. URL http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2000.1314. 2.2.3

Gerdes SY, *et al.* 2003. Experimental Determination and System Level Analysis of Essential Genes in *Escherichia coli* MG1655. *Journal of Bacteriology*, **185**(19):5673–5684. URL http://jb.asm.org/cgi/doi/10.1128/JB.185.19.5673-5684.2003. 4.2.4

Gerding MA, Liu B, Bendezu FO, Hale CA, Bernhardt TG, and de Boer PAJ. 2009. Self-Enhanced Accumulation of FtsN at Division Sites and Roles for Other Proteins with a SPOR Domain (DamX, DedD, and RlpA) in *Escherichia coli* Cell Constriction. *Journal of Bacteriology*, **191**(24):7383–7401. URL http://jb.asm.org/cgi/doi/10.1128/JB.00811-09. 4.2.5

Giegé R and Springer M. 2012. Aminoacyl-tRNA Synthetases in the Bacterial World. *EcoSal Plus*, **5**(1). URL http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.4.2.1. 4.2.4

Gil R, *et al.* 2003. The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences*, **100**(16):9388–9393. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1533499100. I.3.2, M.1, 4.1, 4.2.1

Gil R, Sabater-Muñoz B, Perez-Brocal V, Silva FJ, and Latorre A. 2006. Plasmids in the aphid endosymbiont *Buchnera aphidicola* with the smallest genomes. A puzzling evolutionary story. *Gene*, **370**(1):17–25. URL http://linkinghub.elsevier.com/retrieve/pii/S0378111905006943. 1.3, 2.2.1, 2.2.3

Gomez-Valero L, Soriano-Navarro M, Perez-Brocal V, Heddi A, Moya A, Garcia-Verdugo JM, and Latorre A. 2004. Coexistence of *Wolbachia* with *Buchnera aphidicola* and a Secondary Symbiont in the Aphid *Cinara cedri*. *Journal of Bacteriology*, **186**(19):6626–6633. URL http://jb.asm.org/cgi/doi/10.1128/JB.186.19.6626-6633.2004. I.3.2

Gosalbes MJ, Lamelas A, Moya A, and Latorre A. 2008. The Striking Case of Tryptophan Provision in the Cedar Aphid *Cinara cedri*. *Journal of Bacteriology*, **190**(17):6026–6029. URL http://jb.asm.org/cgi/doi/10.1128/JB.00525-08. I.3.3, I.9, 1.3, 2.2.1, 2.2.3, 3.1, 3.2.2

Gottlieb Y, Lalzar I, and Klasson L. 2015. Distinctive Genome Reduction Rates Revealed by Genomic Analyses of Two *Coxiella*-like Endosymbionts in Ticks. *Genome biology and evolution*, **7**(6):1779–1796. URL http://gbe.oxfordjournals.org/lookup/doi/10.1093/gbe/evv108. 4.1

Griffiths GW and Beck SD. 1975. Ultrastructure of pea aphid mycetocytes: Evidence for symbiote secretion. *Cell and Tissue Research*, **159**(3). URL http://link.springer.com/10.1007/BF00221782. I.3.1, I.6

Grosjean H and Björk GR. 2004. Enzymatic conversion of cytidine to lysidine in anticodon of bacterial tRNAIle – an alternative way of RNA editing. *Trends in Biochemical Sciences*, **29**(4):165–168. URL http://linkinghub.elsevier.com/retrieve/pii/S0968000404000532. 4.2.4

Grossman AD, Straus DB, Walter WA, and Gross CA. 1987. Sigma 32 synthesis can regulate the synthesis of heat shock proteins in *Escherichia coli*. *Genes & Development*, **1**(2):179–184. URL http://www.genesdev.org/cgi/doi/10.1101/gad.1.2.179. 4.2.4

Guay JF, Boudreault S, Michaud D, and Cloutier C. 2009. Impact of environmental stress on aphid clonal resistance to parasitoids: Role of *Hamiltonella defensa* bacterial symbiosis in association with a new facultative symbiont of the pea aphid. *Journal of Insect Physiology*, **55**(10):919–926. URL http://linkinghub.elsevier.com/retrieve/pii/S0022191009002364. I.3.2, M.4

Guest JR, Lewis HM, Graham LD, Packman LC, and Perham RN. 1985. Genetic reconstruction and functional analysis of the repeating lipoyl domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. *Journal of Molecular Biology*, **185**(4):743–754. URL http://linkinghub.elsevier.com/retrieve/pii/0022283685900592. 4.2.5

Hagervall TG and Björk GR. 2013. Transfer RNA Modification. *EcoSal Plus*, **1**(4). URL http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.4.6.2. 4.2.4

Hansen AK and Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. *Proceedings of the National Academy of*

Sciences, **108**(7):2849–2854. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1013465108. I.3.1, 1.2.5, 2.2.4

Hansen AK and Moran NA. 2014. The impact of microbial symbionts on host plant utilization by herbivorous insects. *Molecular Ecology*, **23**:1473–1496. URL http://doi.wiley.com/10.1111/mec.12421. I.1.2

Hansen AK, Vorburger C, and Moran NA. 2012. Genomic basis of endosymbiont-conferred protection against an insect parasitoid. *Genome Research*, **22**(1):106–114. URL http://genome.cshlp.org/cgi/doi/10.1101/gr.125351.111. I.3.2, 4.2.1

Heie OE. 1967. Studies on fossil aphids (Homoptera: Aphidoidea), especially in the Copenhagen collection of fossils in Baltic amber. In *Spolia zoologica Musei Hauniensis*, volume 26, page 274. Munksgaard, København. I.2.1

Heie OE. 1980. General Part. The Families Mindaridae, Hormaphididae, Thelaxidae, Anoeciidae, and Pemphigidae. In *The Aphidoidea (Hemiptera) of Fennoscandia and Denmark*, volume 1, page 236. Brill. ISBN 9788787491341. I.2.2

Heie OE. 1987. Paleontology and Phylogeny. In Minks AK and Harrewijn P, editors, *Aphids, Their Biology, Natural Enemies, and Control*, volume 2A, pages 367–391. Elsevier, Amsterdam, The Netherlands. S.1

Heie OE. 2015. A theory about the evolutionary history of Lachnidae and comments on the results of some molecular phylogenetic studies of aphids (Hemiptera: Aphidoidea)*. *Polish Journal of Entomology*, **84**(4). URL http://www.degruyter.com/view/j/pjen.2015.84.issue-4/pjen-2015-0024/pjen-2015-0024.xml. I.2.3

Heie OE and Wegierek P. 2011. *A List of Fossil Aphids (Hemiptera, Sternorryncha, Aphidomorpha)*, volume 6 of *Monographs of the Upper Silesian Museum*. Instytut Tarnogorski i Museum. ISBN 9788388880247. URL http://www.nhbs.com/title/184724/a-list-of-fossil-aphids-hemiptera-sternorryncha-aphidomorpha. I.2.3

Henry LM, Peccoud J, Simon JC, Hadfield JD, Maiden MJ, Ferrari J, and Godfray HCJ. 2013. Horizontally Transmitted Symbionts and Host Colonization of Ecological Niches. *Current Biology*, **23**(17):1713–1717. URL http://linkinghub.elsevier.com/retrieve/pii/S096098221300852X. I.3.2, I.3.3

Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, and Fukatsu T. 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. *PLoS biology*, **4**(10):e337. URL http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.0040337. I.1.3, I.1.4

Hosokawa T, Ishii Y, Nikoh N, Fujie M, Satoh N, and Fukatsu T. 2016. Obligate bacterial mutualists evolving from environmental bacteria in natural insect populations. *Nature Microbiology*, **1**(1):15011. URL http://www.nature.com/articles/nmicrobiol201511. (document), I.1.3

Houk EJ and Griffiths GW. 1980. Intracellular Symbiotes of the Homoptera. *Annual Review of Entomology*, **25**(1):161–187. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.en.25.010180.001113. I.3.1

Hua Q, Yang C, Baba T, Mori H, and Shimizu K. 2003. Responses of the Central Metabolism in *Escherichia coli* to Phosphoglucose Isomerase and Glucose-6-Phosphate Dehydrogenase Knockouts. *Journal of Bacteriology*, **185**(24):7053–7067. URL http://jb.asm.org/cgi/doi/10.1128/JB.185.24.7053-7067.2003. 2.2.4

Husnik F and McCutcheon JP. 2016. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *bioRxiv*, page Preprint. URL http://biorxiv.org/lookup/doi/10.1101/042267. 2.2.3

Husník F, Chrudimský T, and Hypša V. 2011. Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. *BMC Biology*, **9**(1):87. URL http://dx.doi.org/10.1016/j.cell.2013.05.040. I.3.2, M.6, 2.2.3

Husnik F, *et al.* 2013. Horizontal Gene Transfer from Diverse Bacteria to an Insect Genome Enables a Tripartite Nested Mealybug Symbiosis. *Cell*, **153**(7):1567–1578. URL http://dx.doi.org/10.1016/j.cell.2013.05.040. I.1.4

Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, and Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**(1):119. URL http://www.biomedcentral.com/1471-2105/11/119. M.2.4

Iguchi A, *et al.* 2014. Genome Evolution and Plasticity of *Serratia marcescens*, an Important Multidrug-Resistant Nosocomial Pathogen. *Genome Biology and Evolution*, **6**(8):2096–2110. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evu160. 4.1

Ishikawa H. 1989. Biochemical and Molecular Aspects of Endosymbiosis in Insects. In *International Review of Cytology*, volume 116, pages 1–45. URL http://linkinghub.elsevier.com/retrieve/pii/S0074769608606373. I.3.1

Jacquin-Becker C, Ahel I, Ambrogelly A, Ruan B, Söll D, and Stathopoulos C. 2002. Cysteinyl-tRNA formation and prolyl-tRNA synthetase. *FEBS Letters*, **514**(1):34–36. URL http://doi.wiley.com/10.1016/S0014-5793(02)02331-1. 4.2.4

Jaquiéry J, *et al.* 2013. Masculinization of the X Chromosome in the Pea Aphid. *PLoS Genetics*, **9**(8):e1003690. URL http://dx.plos.org/10.1371/journal.pgen.1003690. I.3

Jones P, *et al.* 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**(9):1236–1240. URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btu031. M.2.4

Jousselin E, Desdevises Y, and Coeur d'acier A. 2009. Fine-scale cospeciation between *Brachycaudus* and *Buchnera aphidicola*: bacterial genome helps define species and evolutionary relationships in aphids. *Proceedings of the Royal Society B: Biological Sciences*, **276**(1654):187–196. URL http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2008.0679. 2.2.3, 3.1

Jousselin E, *et al.* 2016. Assessment of a 16S rRNA amplicon Illumina sequencing procedure for studying the microbiome of a symbiont-rich aphid genus. *Molecular Ecology Resources*, **16**(3):628–640. URL http://doi.wiley.com/10.1111/1755-0998.12478. I.3.3, I.3.3, 2.3, 3.1, 3.2.1

Joyce AR, *et al.* 2006. Experimental and Computational Assessment of Conditionally Essential Genes in *Escherichia coli*. *Journal of Bacteriology*, **188**(23):8259–8271. URL http://jb.asm.org/cgi/doi/10.1128/JB.00740-06. 4.2.4

Kahn D, Fromant M, Fayat G, Dessen P, and Blanquet S. 1980. Methionyl-Transfer-RNA Transformylase from *Escherichia coli*. Purification and Characterisation. *European Journal of Biochemistry*, **105**(3):489–497. URL http://doi.wiley.com/10.1111/j.1432-1033.1980.tb04524.x. 4.2.4

Karp PD, *et al.* 2015. Pathway Tools version 19.0 update: software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, page bbv079. URL http://bib.oxfordjournals.org/lookup/doi/10.1093/bib/bbv079. M.2.4, 2.2.4

Katoh K and Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**(4):772–780. URL http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/mst010. M.6

Keeling PJ, McCutcheon JP, and Doolittle WF. 2015. Symbiosis becoming permanent: Survival of the luckiest. *Proceedings of the National Academy of Sciences*, **112**(33):10101–10103. URL http://www.pnas.org/lookup/doi/10.1073/pnas.1513346112. I.1.4

Kenyon LJ, Meulia T, and Sabree ZL. 2015. Habitat Visualization and Genomic Analysis of "*Candidatus* Pantoea carbekii," the Primary Symbiont of the Brown Marmorated Stink Bug. *Genome Biology and Evolution*, **7**(2):620–635. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evv006. 4.1

Keseler IM, *et al.* 2013. EcoCyc: fusing model organism databases with systems biology. *Nucleic acids research*, **41**(Database issue):D605–12. URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gks1027. M.2.4

Kikuchi Y. 2009. Endosymbiotic Bacteria in Insects: Their Diversity and Culturability. *Microbes and Environments*, **24**(3):195–204. URL http://joi.jlc.jst.go.jp/JST.JSTAGE/jsme2/ME09140S?from=CrossRef. 4.1

Kikuchi Y, Hosokawa T, Nikoh N, Meng XY, Kamagata Y, and Fukatsu T. 2009. Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC biology*, **7**(1):2. URL http://www.biomedcentral.com/1741-7007/7/2. I.1.3

Klasson L. 2006. Strong Asymmetric Mutation Bias in Endosymbiont Genomes Coincide with Loss of Genes for Replication Restart Pathways. *Molecular Biology and Evolution*, **23**(5):1031–1039. URL http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msj107. I.1.4

Kneip C, Lockhart P, Voß C, and Maier UG. 2007. Nitrogen fixation in eukaryotes – New models for symbiosis. *BMC Evolutionary Biology*, **7**(1):55. URL http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-7-55. I.1.2

Koga R, Tsuchida T, and Fukatsu T. 2003. Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont *Buchnera* in an aphid. *Proceedings. Biological sciences / The Royal Society*, **270**(1533):2543–2550. URL http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rspb.2003.2537. I.1.3

Koga R, Tsuchida T, and Fukatsu T. 2009. Quenching autofluorescence of insect tissues for *in situ* detection of endosymbionts. *Applied Entomology and Zoology*, **44**(2):281–291. URL http://joi.jlc.jst.go.jp/JST.JSTAGE/aez/2009.281?from=CrossRef. M.3

Koga R, Meng XY, Tsuchida T, and Fukatsu T. 2012. Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proceedings of the National Academy of Sciences*, **109**(20):E1230–E1237. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1119212109. I.3.3

Korea CG, Ghigo JM, and Beloin C. 2011. The sweet connection: Solving the riddle of multiple sugar-binding fimbrial adhesins in *Escherichia coli*. *BioEssays*, **33**(4):300–311. URL http://doi.wiley.com/10.1002/bies.201000121. 2.2.5

Lai CY, Baumann L, and Baumann P. 1994. Amplification of *trpEG*: adaptation of *Buchnera aphidicola* to an endosymbiotic association with aphids. *Proceedings of the National Academy of Sciences*, **91**(9):3819–3823. URL http://www.pnas.org/cgi/doi/10.1073/pnas.91.9.3819. 1.3, 2.2.1, 2.2.3

Lai CY, Baumann P, and Moran NA. 1995. Genetics of the tryptophan biosynthetic pathway of the prokaryotic endosymbiont (*Buchnera*) of the aphid *Schlechtendalia chinensis*. *Insect Molecular Biology*, **4**(1):47–59. URL http://doi.wiley.com/10.1111/j.1365-2583.1995.tb00007.x. 2.2.1

Lai CY, Baumann P, and Moran NA. 1996. The endosymbiont (*Buchnera* sp.) of the aphid *Diuraphis noxia* contains plasmids consisting of *trpEG* and tandem repeats of *trpEG* pseudogenes. *Applied and environmental microbiology*, **62**(2):332–339. URL http://aem.asm.org/content/62/2/332.abstract. 2.2.1, 2.2.3

Lamelas A, Perez-Brocal V, Gomez-Valero L, Gosalbes MJ, Moya A, and Latorre A. 2008. Evolution of the Secondary Symbiont "*Candidatus* Serratia symbiotica" in Aphid Species of the Subfamily Lachninae. *Applied and Environmental Microbiology*, **74**(13):4236–4240. URL http://aem.asm.org/cgi/doi/10.1128/AEM.00022-08. I.3.3, I.7, I.8, I.3.3, 1.1, 1.3, 2.1, 2.2.3, 2.3, 3.1, 3.2.1, 3.2.1, 3.2.1, 3.2.2

Lamelas A, Gosalbes MJ, Moya A, and Latorre A. 2011a. New Clues about the Evolutionary History of Metabolic Losses in Bacterial Endosymbionts, Provided by the Genome of *Buchnera aphidicola* from the Aphid *Cinara tujafilina*. *Applied and Environmental Microbiology*, **77**(13):4446–4454. URL http://aem.asm.org/cgi/doi/10.1128/AEM.00141-11. 1.1, 1.3, 2.2.2, 2.2.4, 3.2.2

Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, and Latorre A. 2011b. *Serratia symbiotica* from the Aphid *Cinara cedri*: A Missing Link from Facultative to Obligate Insect Endosymbiont. *PLoS Genetics*, **7**(11):e1002357. URL http://dx.plos.org/10.1371/journal.pgen.1002357. I.3.3, 1.2.3, 1.2.5, 2.1, 2.2.3, 3.1, 3.2.2, 4.1, 4.2.1

Langmead B and Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4):357–359. URL http://dx.doi.org/10.1038/nmeth.1923. M.2.2

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, **30**(22):3276–3278. URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btu531. M.6

Lartillot N, Lepage T, and Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics*, **25**(17):2286–2288. URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btp368. M.6

Laslett D. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, **32**(1):11–16. URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gkh152. M.2.4

Le SQ and Gascuel O. 2008. An improved general amino acid replacement matrix. *Molecular biology and evolution*, **25**(7):1307–20. URL http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/msn067. M.6

Lehane MJ. 2005. Managing the blood meal. In *The Biology of Blood-Sucking in Insects*, chapter 6, pages 84–115. Cambridge University Press, Cambridge, second edi edition. ISBN 9780521543958. URL http://dx.doi.org/10.1017/CBO9780511610493.007. I.1.3

Leonardo TE and Muiru GT. 2003. Facultative symbionts are associated with host plant specialization in pea aphid populations. *Proceedings of the Royal Society B: Biological Sciences*, **270**(Suppl_2):S209–S212. URL http://rspb.royalsocietypublishing.org/cgi/doi/10.1098/rsbl.2003.0064. I.3.2

Li L. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*, **13**(9):2178–2189. URL http://www.genome.org/cgi/doi/10.1101/gr.1224503. M.5

Li XY, Jiang LY, and Qiao GX. 2014. Is the subfamily Eriosomatinae (Hemiptera: Aphididae) monophyletic? *Turkish Journal of Zoology*, **38**(3):285–297. URL http://dergipark.ulakbim.gov.tr/tbtkzoology/article/view/5000022742. **??**

Liu L, Huang X, Zhang R, Jiang L, and Qiao G. 2013. Phylogenetic congruence between *Mollitrichosiphum* (Aphididae: Greenideinae) and *Buchnera* indicates insect-bacteria parallel evolution. *Systematic Entomology*, **38**(1):81–92. URL http://doi.wiley.com/10.1111/j.1365-3113.2012.00647.x. 3.1

López-Madrigal S, Latorre A, Porcar M, Moya A, and 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):5587–8. URL http://jb.asm.org/content/193/19/5587.long. I.1.4

Lowe TM and Eddy SR. 1997. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*, **25**(5):955–964. URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/25.5.0955. M.2.4

Lu Q, Zhang X, Almaula N, Mathews CK, and Inouye M. 1995. The Gene for Nucleoside Diphosphate Kinase Functions as a Mutator Gene in *Escherichia coli*. *Journal of Molecular Biology*, **254**(3):337–341. URL http://www.sciencedirect.com/science/article/pii/S0022283685706201http://linkinghub.elsevier.com/retrieve/pii/S0022283685706201. 2.2.4

Luo R, *et al.* 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, **1**(1):18. URL http://www.gigasciencejournal.com/content/1/1/18http://gigascience.biomedcentral.com/articles/10.1186/2047-217X-1-18. M.2.3

Mangat CS and Brown ED. 2008. Known Bioactive Small Molecules Probe the Function of a Widely Conserved but Enigmatic Bacterial ATPase, YjeE. *Chemistry & Biology*, **15**(12):1287–1295. URL http://linkinghub.elsevier.com/retrieve/pii/S1074552108004109. 4.2.4

Manzano-Marín A and Latorre A. 2014. Settling Down: The Genome of *Serratia symbiotica* from the Aphid *Cinara tujafilina* Zooms in on the Process of Accommodation to a Cooperative Intracellular Life. *Genome Biology and Evolution*, **6**(7):1683–1698. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evu133. 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 1.10, 2.1, 2.2.3, 2.2.3, 2.2.3, 3.1, 3.2.1, 3.2.2, 4.1, 4.2.1, 4.2.2, 4.2.5, S.3, S.4, S.5

Manzano-Marín A and Latorre A. 2016. Snapshots of a shrinking partner: Genome reduction in *Serratia symbiotica*. *bioRxiv*. URL http://biorxiv.org/lookup/doi/10.1101/057653. 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, S.11, S.12

Manzano-Marín A, Lamelas A, Moya A, and Latorre A. 2012. Comparative Genomics of *Serratia* spp.: Two Paths towards Endosymbiotic Life. *PLoS ONE*, **7**(10):e47274. URL http://dx.plos.org/10.1371/journal.pone.0047274. I.3.3, 2.2.3, 2.2.3, 4.2.1, 4.2.1

Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jiménez-García LF, and 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. URL http://gbe.oxfordjournals.org/lookup/doi/10.1093/gbe/evv188. 4.1

Manzano-Marín A, Szabo G, Simon JC, Horn M, and Latorre A. 2016. Happens in the best of subfamilies: Replacements and internalisations of Lachninae aphids' co-obligate *Serratia* endosymbiont. *bioRxiv*. URL http://biorxiv.org/lookup/doi/10.1101/059816. I.5, 3.4, 3.5, 3.6, 3.7, 3.8

Manzano-Marín A, Simon JC, and Latorre A. 2016. Reinventing the Wheel and Making It Round Again: Evolutionary Convergence in *Buchnera-Serratia* Symbiotic Consortia between the Distantly Related Lachninae Aphids *Tuberolachnus salignus* and *Cinara cedri*. *Genome Biology and Evolution*, **8**(5):1440–1458. URL http://gbe.oxfordjournals.org/lookup/doi/10.1093/gbe/evw085. 2.2, 2.3, 2.2, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 2.3, 2.10, 3.1, 3.2.2, 4.1, 4.2.5, S.6, S.7, S.8, S.10, S.1

Martin BD and Schwab E. 2012. Current Usage of Symbiosis and Associated Terminology. *International Journal of Biology*, **5**(1):32. URL http://www.ccsenet.org/journal/index.php/ijb/article/view/21139. I.1.1, I.1.1, I.1

McCutcheon JP. 2010. The bacterial essence of tiny symbiont genomes. *Current opinion in microbiology*, **13**(1):73–8. URL http://dx.doi.org/10.1016/j.mib.2009.12.002. I.1.4

McCutcheon JP and Moran NA. 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 million years of evolution. *Genome Biology and Evolution*, **2**:708–18. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evq055. 2.2.3

McCutcheon JP and Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. *Nature reviews. Microbiology*, **10**(1):13–26. URL http://dx.doi.org/10.1038/nrmicro2670. I.1.4

McCutcheon JP and von Dohlen CD. 2011. An Interdependent Metabolic Patchwork in the Nested Symbiosis of Mealybugs. *Current biology*, **21**(16):1366–1372. URL http://linkinghub.elsevier.com/retrieve/pii/S096098221100724X. I.1.4

McCutcheon JP, McDonald BR, and Moran NA. 2009. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet*, **5**(7):e1000565. URL http://dx.plos.org/10.1371/journal.pgen.1000565http://www.ncbi.nlm.nih.gov/pubmed/19609354. I.1.4

Mellot P, Mechulam Y, Le Corre D, Blanquet S, and Fayat G. 1989. Identification of an amino acid region supporting specific methionyl-tRNA synthetase: tRNA recognition. *Journal of Molecular Biology*, **208**(3):429–443. URL http://linkinghub.elsevier.com/retrieve/pii/002228368990507X. 4.2.5

Mendonça AG, Alves RJ, and Pereira-Leal JB. 2011. Loss of Genetic Redundancy in Reductive Genome Evolution. *PLoS Computational Biology*, **7**(2):e1001082. URL http://dx.plos.org/10.1371/journal.pcbi.1001082. 4.2.4

Mereschkowsky C. 1905. Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. In *Biologisches Centralblatt.*, volume 25, pages 593–604. Georg Thieme, Leipzig. URL http://www.biodiversitylibrary.org/item/17816. I.1.2

Mereschkowsky C. 1910. Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. In Rosenthal J, Goebel K, and Hertwig R, editors, *Biologisches Centralblatt.*, volume 30, pages 278–288,289–303,321–347,353–367. Georg Thieme, Leipzig. URL http://www.biodiversitylibrary.org/item/17797. I.1.2

Meseguer AS, Coeur d'acier A, Genson G, and Jousselin E. 2015. Unravelling the historical biogeography and diversification dynamics of a highly diverse conifer-feeding aphid genus. *Journal of Biogeography*, **88**(8):1482–1492. URL http://doi.wiley.com/10.1111/jbi.12531. I.2.3, I.5, 2.2.2, 3.1, 3.2.1, 3.8

Michalik A, Szklarzewicz T, Janakowska W, and Wieczorek K. 2014. Endosymbiotic microorganisms of aphids (Hemiptera: Sternorrhyncha: Aphidoidea): Ultrastructure, distribution and transovarial transmission. *European Journal of Entomology*, **111**(1):91–104. URL http://www.eje.cz/doi/10.14411/eje.2014.011.html. 3.1, 3.1, 3.2.2, S.9

Miller JH, *et al.* 2002. *Escherichia coli* Strains (*ndk*) Lacking Nucleoside Diphosphate Kinase Are Powerful Mutators for Base Substitutions and Frameshifts in Mismatch-Repair-Deficient Strains. *Genetics*, **162**(1):5–13. URL http://www.genetics.org/content/162/1/5. 2.2.4

Milne I, *et al.* 2013. Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, **14**(2):193–202. URL http://bib.oxfordjournals.org/cgi/doi/10.1093/bib/bbs012. M.2.2

Mittler T and Dadd R. 1963. Studies on the artificial feeding of the aphid *Myzus persicae* (Sulzer)—II. Relative survival, development, and larviposition on different diets. *Journal of Insect Physiology*, **9**(5):741–757. URL http://linkinghub.elsevier.com/retrieve/pii/0022191063900179. I.3

Mittler TE. 1957. Studies on the Feeding and Nutrition of *Tuberolachnus Salignus* (Gmelin) (Homoptera, Aphididae):: 1. the Uptake of Phloem Sap. *J. Exp. Biol.*, **34**(3):334–341. URL http://jeb.biologists.org/content/34/3/334.short. 2.1

Mittler TE. 1958a. Studies on the Feeding and Nutrition of *Tuberolachnus Salignus* (Gmelin) (Homoptera, Aphididae): II. The Nitrogen and Sugar Composition of Ingested Phloem Sap and Excreted Honeydew. *J. Exp. Biol.*, **35**(1):74–84. URL http://jeb.biologists.org/content/35/1/74.short. 2.1

Mittler TE. 1958b. Studies on the Feeding and Nutrition of *Tuberolachnus Salignus* (Gmelin) (Homoptera, Aphididae): III. The Nitrogen Economy. *J. Exp. Biol.*, **35**(3):626–638. URL http://jeb.biologists.org/content/35/3/626.abstract. 2.1

Moran NA, Russell JA, Koga R, and Fukatsu T. 2005. Evolutionary Relationships of Three New Species of Enterobacteriaceae Living as Symbionts of Aphids and Other Insects. *Applied and Environmental Microbiology*, **71**(6):3302–3310. URL http://aem.asm.org/cgi/doi/10.1128/AEM.71.6.3302-3310.2005. I.3.2, 1.2.1, 2.2.3, 3.2, 3.2.1, 3.2.2, 4.1, S.2

Montllor CB, Maxmen A, and Purcell AH. 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. *Ecological Entomology*, **27**:189–195. URL http://dx.doi.org/10.1046/j.1365-2311.2002.00393.x. I.3.2, I.3.3

Moran NA. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A*, **93**(7):2873–2878. URL http://www.pnas.org/content/93/7/2873. I.1.4

Moran NA and Bennett GM. 2014. The Tiniest Tiny Genomes. *Annual Review of Microbiology*, **68**(1):195–215. URL http://www.annualreviews.org/doi/abs/10.1146/annurev-micro-091213-112901. I.1.4, 4.2.1, 4.2.4, 4.2.4

Moran NA and Wernegreen JJ. 2000. Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends in Ecology & Evolution*, **15**(8):321–326. URL http://linkinghub.elsevier.com/retrieve/pii/S0169534700019029. I.1.4

Moran NA, McCutcheon JP, and Nakabachi A. 2008. Genomics and Evolution of Heritable Bacterial Symbionts. *Annual Review of Genetics*, **42**(1):165–190. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.genet.41.110306.130119. I.1.3, I.1.4, 2.2.1, 4.2.1, 4.2.1, 4.2.2, 4.2.4

Moya A, Peretó J, Gil R, and Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nature reviews. Genetics*, **9**(3):218–29. URL http://www.nature.com/doifinder/10.1038/nrg2319. I.1.2, I.1.4

Munson MA, Baumann P, and Kinsey MG. 1991a. *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. URL http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-41-4-566. I.3.1

Munson MA, Baumann P, Clark MA, Baumann L, Moran NA, Voegtlin DJ, and Campbell BC. 1991b. Evidence for the establishment of aphid-eubacterium endosymbiosis in an ancestor of four aphid families. *J. Bacteriol.*, **173**(20):6321–6324. URL http://jb.asm.org/content/173/20/6321.abstract. I.3.1, 3.1

Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T, and Yokoyama S. 1988. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature*, **336**(6195):179–181. URL http://www.nature.com/doifinder/10.1038/336179a0. 4.2.4

Myers EW, *et al.* 2000. A Whole-Genome Assembly of *Drosophila*. *Science (New York, N.Y.)*, **287**(5461):2196–2204. URL http://www.sciencemag.org/cgi/doi/10.1126/science.287.5461.2196. M.2.2

Nakabachi A and Ishikawa H. 1999. Provision of riboflavin to the host aphid, *Acyrthosiphon pisum*, by endosymbiotic bacteria, *Buchnera*. *Journal of Insect Physiology*, **45**(1):1–6. URL http://linkinghub.elsevier.com/retrieve/pii/S0022191098001048. 1.1, 1.2.3, 2.2.4

Nakabachi A, *et al.* 2013. Defensive Bacteriome Symbiont with a Drastically Reduced Genome. *Current Biology*, **23**(15):1478–1484. URL http://linkinghub.elsevier.com/retrieve/pii/S0960982213007525. 2.2.3, 4.1

Nakagawa S, *et al.* 2014. Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. *The ISME Journal*, **8**(1):40–51. URL http://dx.doi.org/10.1038/ismej.2013.131. I.1.2

Nawrocki EP and Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*, **29**(22):2933–2935. URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btt509. M.2.4

Nawrocki EP, *et al.* 2015. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*, **43**(D1):D130–D137. URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gku1063. M.2.4

Nikoh N, Hosokawa T, Oshima K, Hattori M, and Fukatsu T. 2011. Reductive Evolution of Bacterial Genome in Insect Gut Environment. *Genome Biology and Evolution*, **3**:702–714. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evr064. I.1.4, 4.1

Nikoh N, Hosokawa T, Moriyama M, Oshima K, Hattori M, and Fukatsu T. 2014. Evolutionary origin of insect-*Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences*, **111**(28):10257–10262. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1409284111. I.3.2

Nogge G. 1976. Sterility in tsetse flies (*Glossina morsitan*s Westwood) caused by loss of symbionts. *Experientia*, **32**(8):995–996. URL http://link.springer.com/10.1007/BF01933932. I.1.3

Normark BB. 2000. Molecular Systematics and Evolution of the Aphid Family Lachnidae. *Molecular Phylogenetics and Evolution*, **14**(1):131–140. URL http://linkinghub.elsevier.com/retrieve/pii/S1055790399906991. I.2.3

Nováková E, Hypša V, Klein J, Foottit RG, von Dohlen CD, and Moran NA. 2013. Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Molecular phylogenetics and evolution*, **68**(1):42–54. URL http://dx.doi.org/10.1016/j.ympev.2013.03.016. I.2.2, I.4, I.2.3, M.6, 2.2.2

Nováková E, Husník F, Šochová E, and Hypša V. 2015. *Arsenophonus* and *Sodalis* Symbionts in Louse Flies: an Analogy to the *Wigglesworthia* and *Sodalis* System in Tsetse Flies. *Applied and Environmental Microbiology*, **81**(18):6189–6199. URL http://aem.asm.org/lookup/doi/10.1128/AEM.01487-15. 4.1

Oakeson KF, *et al.* 2014. Genome Degeneration and Adaptation in a Nascent Stage of Symbiosis. *Genome Biology and Evolution*, **6**(1):76–93. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt210. I.3.2, 1.2.2, 1.2.4, 1.2.4, 4.1, 4.2.1

Ohtaka C and Ishikawa H. 1991. Effects of heat treatment on the symbiotic system of an aphid mycetocyte. *Symbiosis*, **11**:19–30. URL http://agris.fao.org/agris-search/search.do?recordID=US201301738764. I.1.3

Okonechnikov K, Golosova O, and Fursov M. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, **28**(8):1166–1167. URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/bts091. M.2.4

Oliver KM, Russell JA, Moran NA, and Hunter MS. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, **100**(4):1803–1807. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0335320100. I.3.2

Oliver KM, Moran NA, and Hunter MS. 2005. Variation in resistance to parasitism in aphids is due to symbionts not host genotype. *Proceedings of the National Academy of Sciences*, **102**(36):12795–12800. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0506131102. I.3.2

Oliver KM, Degnan PH, Burke GR, and Moran NA. 2010. Facultative Symbionts in Aphids and the Horizontal Transfer of Ecologically Important Traits. *Annual Review of Entomology*, **55**(1):247–266. URL http://www.annualreviews.org/doi/abs/10.1146/annurev-ento-112408-085305. I.3.2

Oliver KM, Smith AH, and Russell JA. 2014. Defensive symbiosis in the real world - advancing ecological studies of heritable, protective bacteria in aphids and beyond. *Functional Ecology*, **28**(2):341–355. URL http://doi.wiley.com/10.1111/1365-2435.12133. I.3.2

Omori K and Komatsubara S. 1993. Role of serine 352 in the allosteric response of *Serratia marcescens* aspartokinase I-homoserine dehydrogenase I analyzed by using site-directed mutagenesis. *Journal of Bacteriology*, **175**(4):959–965. URL http://jb.asm.org/content/175/4/959.abstract. 2.2.5

Ondov BD, Bergman NH, and Phillippy AM. 2011. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, **12**(1):385. URL http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-385. M.2.2

Ortiz-Rivas B and Martínez-Torres D. 2010. Combination of molecular data support the existence of three main lineages in the phylogeny of aphids (Hemiptera: Aphididae) and the basal position of the subfamily Lachninae. *Molecular Phylogenetics and Evolution*, **55**(1):305–317. URL http://linkinghub.elsevier.com/retrieve/pii/S1055790309005119. S.1

Özder N and Sağlam Ö. 2008. Effect of Temperature on the Biology of *Tuberolachnus salignus* (Gmelin) (Sternorrhyncha: Aphididae) on (*Salix alba*). *Journal of Central European Agriculture*, **9**(1):155–160. URL http://hrcak.srce.hr/24911. 2.1

Özder N, Sağlam Ö, Güven A, and Aydın V. 2007. Effect of temperature on development, mortality, fecundity and reproduction of *Tuberolachnus salignus* on weeping willow (*Salix babylonica*) and corkscrew willow (*Salix matsudana*). *Phytoparasitica*, **35**(4):386–391. URL http://link.springer.com/10.1007/BF02980702. 2.1

Patiño-Navarrete R, Moya A, Latorre A, and Peretó J. 2013. Comparative Genomics of *Blattabacterium cuenoti*: The Frozen Legacy of an Ancient Endosymbiont Genome. *Genome Biology and Evolution*, **5**(2):351–361. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evt011. 2.2.3, 4.2.1

Penz T, *et al.* 2012. Comparative Genomics Suggests an Independent Origin of Cytoplasmic Incompatibility in *Cardinium hertigii*. *PLoS Genetics*, **8**(10):e1003012. URL http://dx.plos.org/10.1371/journal.pgen.1003012. I.3.2

Pérez-Brocal V, *et al.* 2006. A Small Microbial Genome: The End of a Long Symbiotic Relationship? *Science*, **314**(5797):312–313. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1130441. I.3.3, M.6, 1.3, 2.2.2, 3.2.2

Philippe H and Roure B. 2011. Difficult phylogenetic questions: more data, maybe; better methods, certainly. *BMC Biology*, **9**(1):91. URL http://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-9-91. M.6

Pine L, George JR, Reeves MW, and Harrell WK. 1979. Development of a chemically defined liquid medium for growth of *Legionella pneumophila*. *J. Clin. Microbiol.*, **9**(5):615–626. URL http://jcm.asm.org/content/9/5/615.abstract?sid=21c72ab1-ed4d-407b-890a-dc627d47f422. I.1.4

Purta E, Kaminska KH, Kasprzak JM, Bujnicki JM, and Douthwaite S. 2008. YbeA is the m3 methyltransferase RlmH that targets nucleotide 1915 in 23S rRNA. *RNA*, **14**(10):2234–2244. URL http://www.rnajournal.org/cgi/doi/10.1261/rna.1198108. 4.2.4

Pyka-Fościak G and Szklarzewicz T. 2008. Germ cell cluster formation and ovariole structure in viviparous and oviparous generations of the aphid *Stomaphis quercus*. *The International Journal of Developmental Biology*, **52**(2-3):259–265. URL http://www.ijdb.ehu.es/web/paper.php?doi=10.1387/ijdb.072338gp. 3.1, 3.2.2, S.9

Remaudière G and Binazzi A. 2003. Les *Cinara* du Pakistan II. Le sous-genre *Cupressobium* (Hemiptera, Aphididae, Lachninae). *Revue française d'entomologie*, **25**(2):85–96. URL http://cat.inist.fr/?aModele=afficheN{&}cpsidt=14848754. 1.1

Rio RVM, *et al.* 2012. Insight into the Transmission Biology and Species-Specific Functional Capabilities of Tsetse (Diptera: Glossinidae) Obligate Symbiont *Wigglesworthia*. *mBio*, **3**(1):e00240–11–e00240–11. URL http://mbio.asm.org/cgi/doi/10.1128/mBio.00240-11. I.1.4

Rocha EP and Danchin A. 2002. Base composition bias might result from competition for metabolic resources. *Trends in Genetics*, **18**(6):291–294. URL http://linkinghub.elsevier.com/retrieve/pii/S0168952502026902. I.1.4

Ronquist F, *et al.* 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, **61**(3):539–542. URL http://sysbio.oxfordjournals.org/cgi/doi/10.1093/sysbio/sys029. M.6

Rouhbakhsh D, *et al.* 1996. The tryptophan biosynthetic pathway of aphid endosymbionts (*Buchnera*): Genetics and evolution of plasmid-associated anthranilate synthase (*trpEG*) within the Aphididae. *Journal of Molecular Evolution*, **42**(4):414–421. URL http://link.springer.com/10.1007/BF02498635. 2.2.1, 2.2.3

Russell JA, Latorre A, Sabater-Munoz B, Moya A, and Moran NA. 2003. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. *Molecular Ecology*, **12**(4):1061–1075. URL http://doi.wiley.com/10.1046/j.1365-294X.2003.01780.x. 3.1, 3.2.1

Sabree ZL and Moran NA. 2014. Host-specific assemblages typify gut microbial communities of related insect species. *SpringerPlus*, **3**(1):138. URL http://www.springerplus.com/content/3/1/138. I.1.3

Sabri A, Leroy P, Haubruge E, Hance T, Frère I, Destain J, and Thonart P. 2011. Isolation, pure culture and characterization of *Serratia symbiotica* sp. nov., the R-type of secondary endosymbiont of the black bean aphid *Aphis fabae*. *Iinternational Journal of Systematic and Evolutionary Microbiology*, **61**(9):2081–2088. URL http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijs.0.024133-0. I.3.3, 4.1

Sacchi L, Grigolo A, Biscaldi G, and Laudani U. 1993. Effects of heat treatment on the symbiotic system of Blattoidea: Morphofunctional alterations of bacteriocytes. *Bolletino di zoologia*, **60**(3):271–279. URL http://www.tandfonline.com/doi/abs/10.1080/11250009309355823. I.1.3

Saffo MB. 1992. Coming to terms with a field: Words and concepts in symbiosis. *Symbiosis*, **14**:17–31. I.1.1

Sagan L. 1967. On the origin of mitosing cells. *Journal of Theoretical Biology*, **14**(3):225–274. URL http://linkinghub.elsevier.com/retrieve/pii/0022519367900793. I.1.2

Sakurai M, Koga R, Tsuchida T, Meng XY, and Fukatsu T. 2005. *Rickettsia* Symbiont in the Pea Aphid *Acyrthosiphon pisum*: Novel Cellular Tropism, Effect on Host Fitness, and Interaction with the Essential Symbiont *Buchnera*. *Applied and Environmental Microbiology*, **71**(7):4069–4075. URL http://aem.asm.org/cgi/doi/10.1128/AEM.71.7.4069-4075.2005. I.3.2, 1.2.1, 3.1, 3.2.2

Sandstrom J and Moran NA. 1999. How nutritionally imbalanced is phloem sap for aphids? *Entomologia Experimentalis et Applicata*, **91**(1):203–210. URL http://doi.wiley.com/10.1046/j.1570-7458.1999.00485.x. I.2.1

Sandström JP, Russell JA, White JP, and Moran NA. 2001. Independent origins and horizontal transfer of bacterial symbionts of aphids. *Molecular Ecology*, **10**(1):217–228. URL http://doi.wiley.com/10.1046/j.1365-294X.2001.01189.x. I.3.2, 1.2.1, 3.2.2

Scarborough CL. 2005. Aphid Protected from Pathogen by Endosymbiont. *Science*, **310**(5755):1781–1781. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1120180. I.3.2

Scheller EV and Cotter PA. 2015. *Bordetella* filamentous hemagglutinin and fimbriae: critical adhesins with unrealized vaccine potential. *Pathogens and Disease*, **73**(8):ftv079. URL http://femspd.oxfordjournals.org/lookup/doi/10.1093/femspd/ftv079. 2.2.5

Schink B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.*, **61**(2):262–280. URL http://mmbr.asm.org/content/61/2/262.abstract. I.1.2

Schmieder R and Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, **27**(6):863–864. URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btr026. M.2.1

Schwartz R and Dayhoff M. 1978. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. *Science*, **199**(4327):395–403. URL http://www.sciencemag.org/cgi/doi/10.1126/science.202030. I.1.2

Shigenobu S and Wilson ACC. 2011. Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cellular and molecular life sciences : CMLS*, **68**(8):1297–309. URL http://dx.doi.org/10.1007{%}2Fs00018-011-0645-2. I.3.1

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, and Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, **407**(6800):81–6. URL http://dx.doi.org/10.1038/35024074. I.3.1

Shoji S, Dambacher CM, Shajani Z, Williamson JR, and Schultz PG. 2011. Systematic Chromosomal Deletion of Bacterial Ribosomal Protein Genes. *Journal of Molecular Biology*, **413**(4):751–761. URL http://linkinghub.elsevier.com/retrieve/pii/S0022283611009934. 4.2.4

Sloan DB and Moran NA. 2012. Genome Reduction and Co-evolution between the Primary and Secondary Bacterial Symbionts of Psyllids. *Molecular Biology and Evolution*, **29**(12):3781–3792. URL http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/mss180. 2.2.3

Smith TA, Driscoll T, Gillespie JJ, and Raghavan R. 2015. A *Coxiella*-like Endosymbiont Is a Potential Vitamin Source for the Lone Star Tick. *Genome Biology and Evolution*, **7**(3):831–838. URL http://gbe.oxfordjournals.org/cgi/doi/10.1093/gbe/evv016. 4.1

Staden R, Beal KF, and Bonfield JK. 1999. The Staden Package, 1998. In Misener S and Krawetz SA, editors, *Bioinformatics Methods and Protocols*, volume 132, pages 115–130. Springer, Totowa, New jersey. URL http://www.springerprotocols.com/Abstract/doi/10.1385/1-59259-192-2:115. M.2.2

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**(21):2688–2690. URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/btl446. M.6

Suzek BE, Ermolaeva MD, Schreiber M, and Salzberg SL. 2001. A probabilistic method for identifying start codons in bacterial genomes. *Bioinformatics*, **17**(12):1123–1130. URL http://bioinformatics.oxfordjournals.org/cgi/doi/10.1093/bioinformatics/17.12.1123. M.2.4

Talavera G and Castresana J. 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Systematic Biology*, **56**(4):564–577. URL http://sysbio.oxfordjournals.org/cgi/doi/10.1080/10635150701472164. M.6

Tamas I, *et al.* 2002. 50 Million Years of Genomic Stasis in Endosymbiotic Bacteria. *Science*, **296**(5577):2376–2379. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1071278. I.3.1, 2.2.3, 4.2.1

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*, **30**(12):2725–2729. URL http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/mst197. M.6

The International pea aphid genomics consortium. 2010. Genome Sequence of the Pea Aphid *Acyrthosiphon pisum*. *PLoS Biology*, **8**(2):e1000313. URL http://dx.plos.org/10.1371/journal.pbio.1000313. I.3.1, 3.2

Toft C and Andersson SGE. 2010. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nature reviews. Genetics*, **11**(7):465–75. URL http://www.nature.com/doifinder/10.1038/nrg2798. I.1.4, I.2

Toh H, Weiss BL, Perkin SAH, Yamashita A, Oshima K, Hattori M, and Aksoy S. 2005. 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. URL http://www.genome.org/cgi/doi/10.1101/gr.4106106. I.3.2, 4.1, 4.2.1

Tsuchida T, Koga R, and Fukatsu T. 2004. Host Plant Specialization Governed by Facultative Symbiont. *Science*, **303**(5666):1989–1989. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1094611. I.3.2

Tsuchida T, Koga R, Meng X, Matsumoto T, and Fukatsu T. 2005. Characterization of a Facultative Endosymbiotic Bacterium of the Pea Aphid *Acyrthosiphon pisum*. *Microbial Ecology*, **49**(1):126–133. URL http://link.springer.com/10.1007/s00248-004-0216-2. I.3.2, 1.2.1, S.2

Tsuchida T, *et al.* 2010. Symbiotic Bacterium Modifies Aphid Body Color. *Science*, **330**(6007):1102–1104. URL http://www.sciencemag.org/cgi/doi/10.1126/science.1195463. I.3.2, 1.2.1, 3.2.2, S.2

Tsuchida T, Koga R, Fujiwara A, and Fukatsu T. 2014. Phenotypic Effect of "*Candidatus* Rickettsiella viridis", a Facultative Symbiont of the Pea Aphid (*Acyrthosiphon pisum*), and Its Interaction with a Coexisting Symbiont. *Applied and Environmental Microbiology*, **80**(2):525–533. URL http://aem.asm.org/cgi/doi/10.1128/AEM.03049-13. I.3.2, 3.1

Tsuchihashi Z and Kornberg A. 1990. Translational frameshifting generates the gamma subunit of DNA polymerase III holoenzyme. *Proceedings of the National Academy of Sciences*, **87**(7):2516–2520. URL http://www.pnas.org/cgi/doi/10.1073/pnas.87.7.2516. 4.2.5

Unterman BM, Baumann P, and McLean DL. 1989. Pea aphid symbiont relationships established by analysis of 16S rRNAs. *Journal of Bacteriology*, **171**(6):2970–2974. URL http://jb.asm.org/cgi/content/abstract/171/6/2970. I.6, I.3.1

Van Domselaar GH, *et al.* 2005. BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research*, **33**(Web Server):W455–W459. URL http://nar.oxfordjournals.org/lookup/doi/10.1093/nar/gki593. M.2.4

van Ham RCHJ, Moya A, and Latorre A. 1997. Putative evolutionary origin of plasmids carrying the genes involved in leucine biosynthesis in *Buchnera aphidicola* (endosymbiont of aphids). *Journal of bacteriology*, **179**(15):4768–4777. URL http://jb.asm.org/content/179/15/4768.abstract. 2.2.1

van Ham RCHJ, Martínez-Torres D, Moya A, and Latorre A. 1999. Plasmid-encoded anthranilate synthase (*TrpEG*) in *Buchnera aphidicola* from aphids of the family Pemphigidae. *Applied and environmental microbiology*, **65**(1):117–125. URL http://aem.asm.org/cgi/content/abstract/65/1/117. 2.2.1, 2.2.3

van Ham RCHJ, *et al.* 2003. Reductive genome evolution in *Buchnera aphidicola*. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(2):581–586. URL http://www.pnas.org/cgi/doi/10.1073/pnas.0235981100. I.1.4, I.3.1, 2.2.3, 2.2.3, 4.1, 4.2.1

Vincent HA and Deutscher MP. 2009. The Roles of Individual Domains of RNase R in Substrate Binding and Exoribonuclease Activity: The Nuclease Domain is Sufficinet for Digestion of Structured RNA. *Journal of Biological Chemistry*, **284**(1):486–494. URL http://www.jbc.org/cgi/doi/10.1074/jbc.M806468200. 4.2.5

von Dohlen CD and Moran NA. 2000. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biological Journal of the Linnean Society*, **71**(4):689–717. URL http://doi.wiley.com/10.1111/j.1095-8312.2000.tb01286.x. I.2.1, M.6, 3.1

Walker JR, *et al.* 2000. *Escherichia coli* DNA polymerase III tau- and gamma-subunit conserved residues required for activity *in vivo* and *in vitro*. *Journal of bacteriology*, **182**(21):6106–13. URL http://jb.asm.org/cgi/doi/10.1128/JB.182.21.6106-6113.2000. 4.2.5

Wei JR, *et al.* 2006a. A mobile quorum-sensing system in *Serratia marcescens*. *Journal of Bacteriology*, **188**(4):1518–1525. URL http://jb.asm.org/cgi/doi/10.1128/JB.188.4.1518-1525.2006. 1.2.4

Wei JR, *et al.* 2006b. Regulatory roles of *spnT*, a novel gene located within transposon Tn*TIR*. *Biochemical and Biophysical Research Communications*, **348**(3):1038–1046. URL http://dx.doi.org/10.1016/j.bbrc.2006.07.152. 1.2.4

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. *Nature reviews. Genetics*, **3**(11):850–61. URL http://www.nature.com/nrg/journal/v3/n11/glossary/nrg931.html. I.1.4

Williams LE and Wernegreen JJ. 2015. Genome evolution in an ancient bacteria-ant symbiosis: parallel gene loss among *Blochmannia* spanning the origin of the ant tribe Camponotini. *PeerJ*, **3**:e881. URL https://peerj.com/articles/881. 4.1, 4.2.1

Wilson K. 2002. Preparation of Genomic DNA from Bacteria. In Ausbel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, and Struhl K, editors, *Short Protocols*

in Molecular Biology, chapter 2, pages Section I, Unit 2.4. John Wiley & Sons Inc, New York, NY, 5th edition. ISBN 978-0-471-25092-0. URL http://www.wiley.com/WileyCDA/WileyTitle/productCd-0471250929.html. M.1

Winkler HH. 1990. *Rickettsia* Species (As Organisms). *Annual Review of Microbiology*, **44**(1):131–153. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.mi.44.100190.001023. I.1.4

Wojciechowski W. 1992. Studies on the Systematic System of Aphids (Homoptera, Aphidinea). In *Prace Naukowe Uniwersytetu Śląskiego w Katowicach*, 1269, page 75. Uniwersytet Śąlski, Katowice. ISBN 9788322604489. S.1

Wolf J. 2002. tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *The EMBO Journal*, **21**(14):3841–3851. URL http://emboj.embopress.org/cgi/doi/10.1093/emboj/cdf362. 4.2.4

Wong M, Thornton D, Kennedy R, and Dolan M. 1995. A chemically defined liquid medium that supports primary isolation of *Rochalimaea* (*Bartonella*) *henselae* from blood and tissue specimens. *J. Clin. Microbiol.*, **33**(3):742–744. URL http://jcm.asm.org/content/33/3/742.abstract. I.1.4

Worning P, Jensen LJ, Hallin PF, Stærfeldt HH, and Ussery DW. 2006. Origin of replication in circular prokaryotic chromosomes. *Environmental Microbiology*, **8**(2):353–361. URL http://doi.wiley.com/10.1111/j.1462-2920.2005.00917.x. M.2.4

Wu M, *et al.* 2004. Phylogenomics of the Reproductive Parasite *Wolbachia pipientis* wMel: A Streamlined Genome Overrun by Mobile Genetic Elements. *PLoS Biology*, **2**(3):e69. URL http://dx.plos.org/10.1371/journal.pbio.0020069. I.3.2

Xue J, *et al.* 2014. Genomes of the rice pest brown planthopper and its endosymbionts reveal complex complementary contributions for host adaptation. *Genome Biology*, **15**(12):521. URL http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0521-0. 4.1

Zhang CM and Hou YM. 2004. Synthesis of Cysteinyl-tRNACys by A Prolyl-tRNA Synthetase. *RNA Biology*, **1**(1):34–40. URL http://www.tandfonline.com/doi/abs/10.4161/rna.1.1.910. 4.2.4

Zhang G and Chen X. 1999. Study on the phylogeny of Pemphigidae (Homoptera:Aphidinea). *Acta Entomologica Sinica*, **42**(2):176–183. URL http://www.insect.org.cn/EN/Y1999/V42/I2/176. **??**

Zhuravlev VV. 2003. The aphids of the genus *Cupressobium* (Homoptera, Aphidoidea, Lachnidae) of Eastern Europe, Caucasus and Middle Asia. *Vestnik Zoologii*, **37**(5):13–30. URL http://www.v-zool.kiev.ua/pdfs/2003/5/02.pdf. 1.1

Ziegler H. 1975. Nature of Transported Substances. In Zimmermann MH and Milburn JA, editors, *Encyclopedia of plant physiology. Volume 1. Transport in Plants I*, chapter 3, pages 59–100. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-642-66163-1. URL http://link.springer.com/10.1007/978-3-642-66161-7. I.1.3, I.2.1

Zientz E, Dandekar T, and Gross R. 2004. Metabolic Interdependence of Obligate Intracellular Bacteria and Their Insect Hosts. *Microbiology and Molecular Biology Reviews*, **68**(4):745–770. URL http://mmbr.asm.org/cgi/doi/10.1128/MMBR.68.4.745-770.2004. I.1.3

Łukasik P, van Asch M, Guo H, Ferrari J, and Godfray HCJ. 2013. Unrelated facultative endosymbionts protect aphids against a fungal pathogen. *Ecology Letters*, **16**(2):214–218. URL http://doi.wiley.com/10.1111/ele.12031. I.3.2

Appendix

S.1 Additional figures

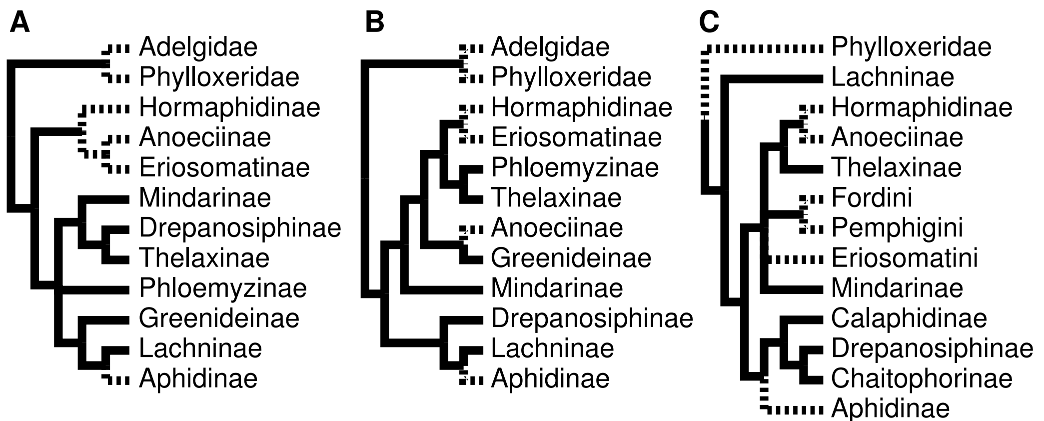


Figure S.1 Previous hypothesis of aphid phylogeny. Hypotheses of aphid phylogeny based on: cladistic analyses of morphological characters by **A)** Heie (1987) and **B)** Wojciechowski (1992); and **C)** molecular-based phylogentic analysis based on three genes (two nuclear and one mitochondrial) Ortiz-Rivas and Martínez-Torres (2010) (where Drepanosiphiniae are shown as three monophyletic families). Dotted branches represent host alternating lineages.

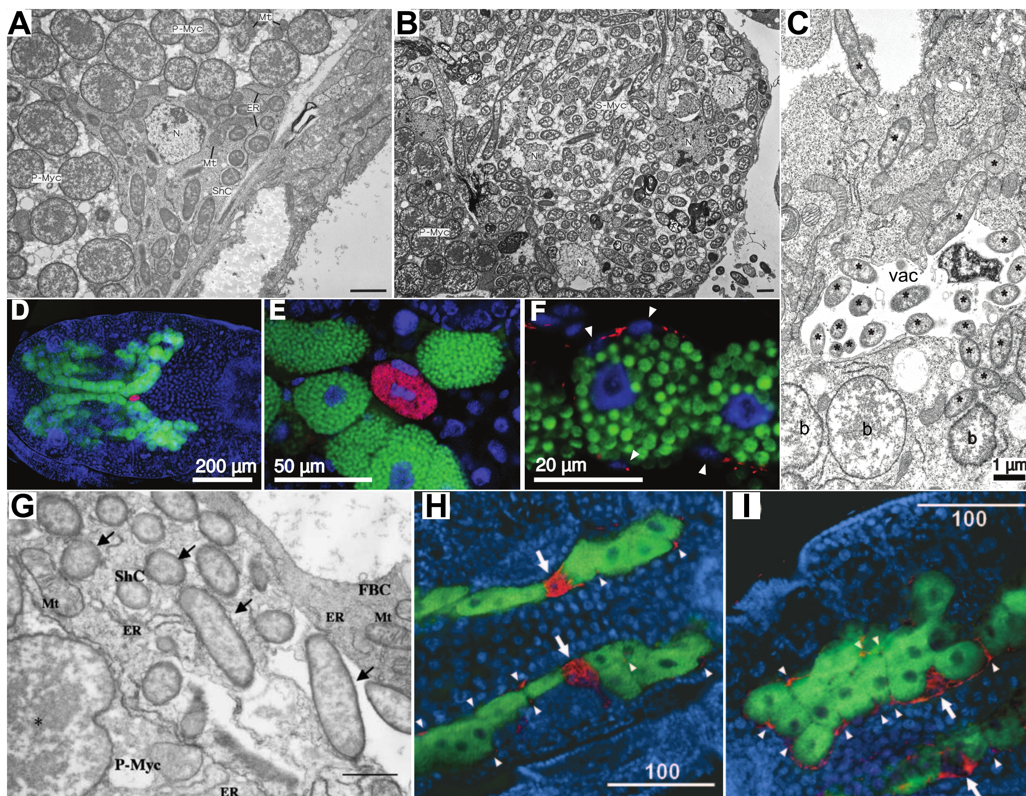


Figure S.2 Morphology and localisation of secondary endosymbionts of *Ac. pisum*. **(A** and **B)** Magnified image of a sheath cell and a secondary bacteriocyte harbouring *S. symbiotica*, respectively. Bars=2µm. **(C)** Image of a secondary bacteriocyte harbouring *Rickettsiella* and a primary bacteriocyte harboring *Buchnera*. v= vacuole, b= *Buchnera*, *= *Rickettsiella*. **(D)** Mature embryo (blue) containing many primary bacteriocytes harbouring *Buchnera* (green) and a secondary bacteriocyte harboring *Rickettsiella* (red). **(E)** Enlarged image of the secondary bacteriocyte. **(F)** Cells harbouring *Rickettsiella* adhering to the periphery of primary bacteriocytes (white arrowheads). **(G)** An enlarged image of *Regiella* cells inside sheath cells. **(H** and **I)** *Buchnera* (green), *S. symbiotica* (red), and *Regiella* (red) in aphid bacteriomes. ER= endoplasmic reticulum; Mt=mitochondrion, N=nucleus; P-Myc and S-myc=primary and secondary bacteriome, respectively; ShC=sheath cell. Reproduced from (A and B) Fukatsu *et al.* (2000), (C-F) Tsuchida *et al.* (2010), (G) Tsuchida *et al.* (2005), and (H and I) Moran *et al.* (2005).

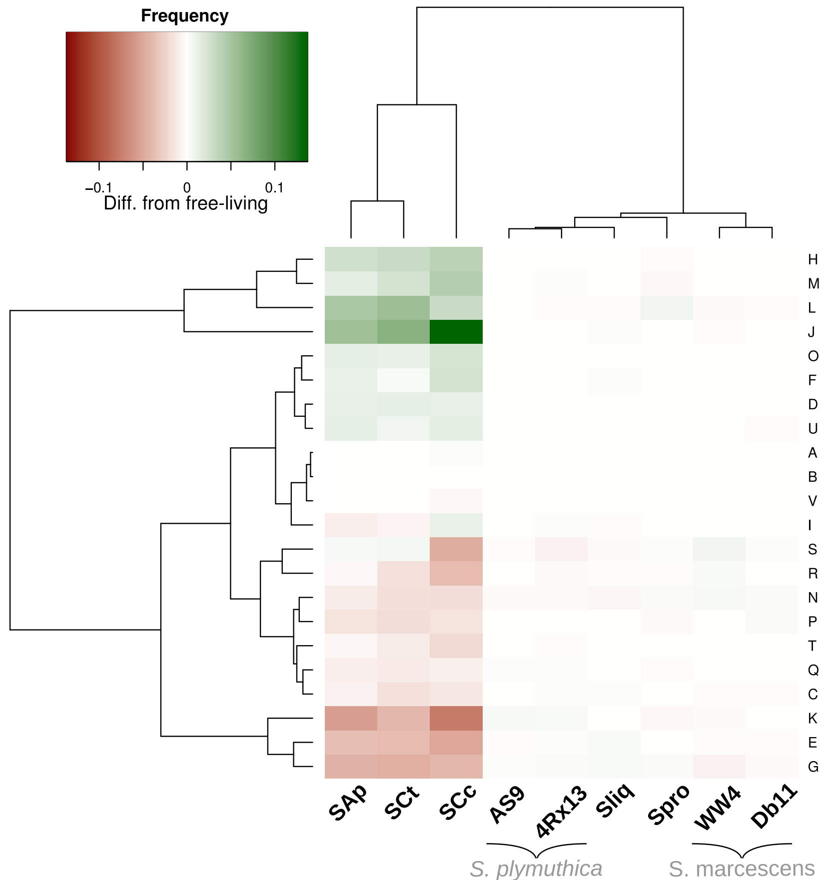


Figure S.3 Functional profile disruption in *S. symbiotica*. Heat map showing the two-way clustering of the COG profiles frequency divergence of *S. symbiotica* strains from the free-living *Serratia*. On the right-hand side of the heatmap, one-letter COG assignments for each row are displayed. On the bottom of each column, abbreviations for each if the *Serratia* strains are as follows: AS9= *S. plymuhtica* strain AS9; 4Rx13= *S. plymuthica* strain 4Rx13; Sliq= *S. liquefasciens* strain ATCC 27592; Spro= *S. protemaculans* strain 568; WW4= *S. plymuhtica* strain WW4; Db11= *S. marcescens* strain Db11. Reproduced from Manzano-Marín and Latorre (2014).

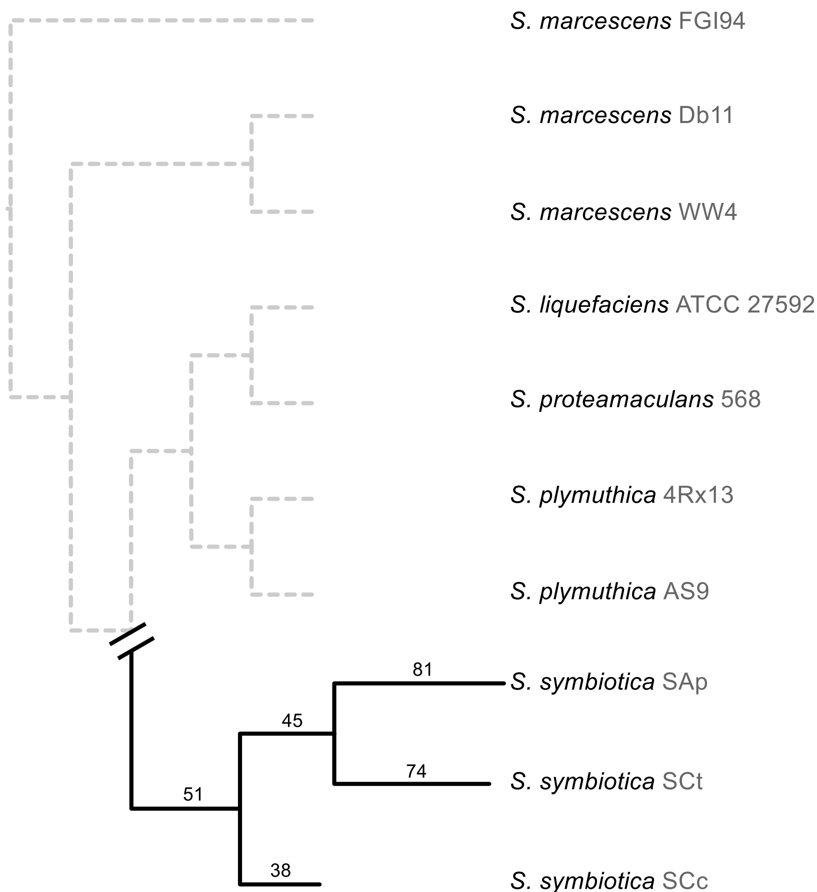


Figure S.4 *S. symbiotica* minimum number of rearrangements phylogeny. Rooted minimum number of rearrangements phylogeny using the 354 single-copy shared genes found among selected *Serratia* strains. SAp's scaffolds have been arranged taking SCt's as reference. Reproduced from Manzano-Marín and Latorre (2014).

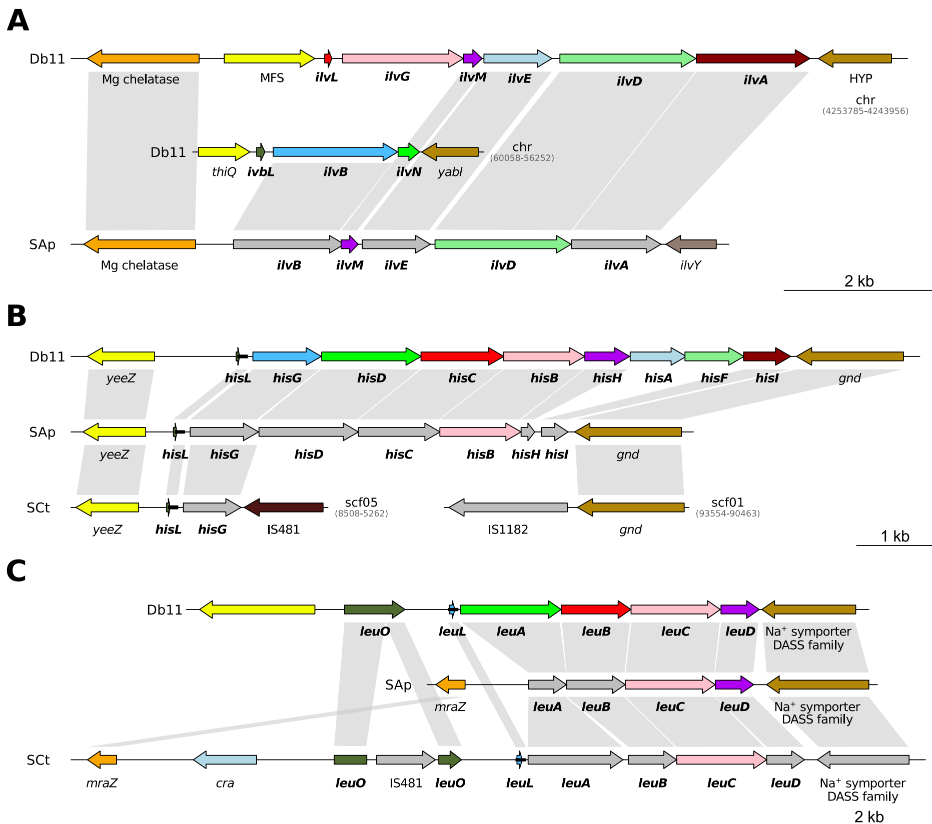


Figure S.5 Ile, His, and Leu biosynthetic operons and transcription units' degradation in *S. symbiotica*. Diagrams displaying the erosion of biosynthetic operons and transcription units in the facultative SAp and SCt, relative to Db11. **(A)** *ilvBN* an *ilvGMEDA* operons. **(B)** *hisGDCBHAFI* operon. **(C)** *leuABCD* operon. Modified from Manzano-Marín and Latorre (2014).

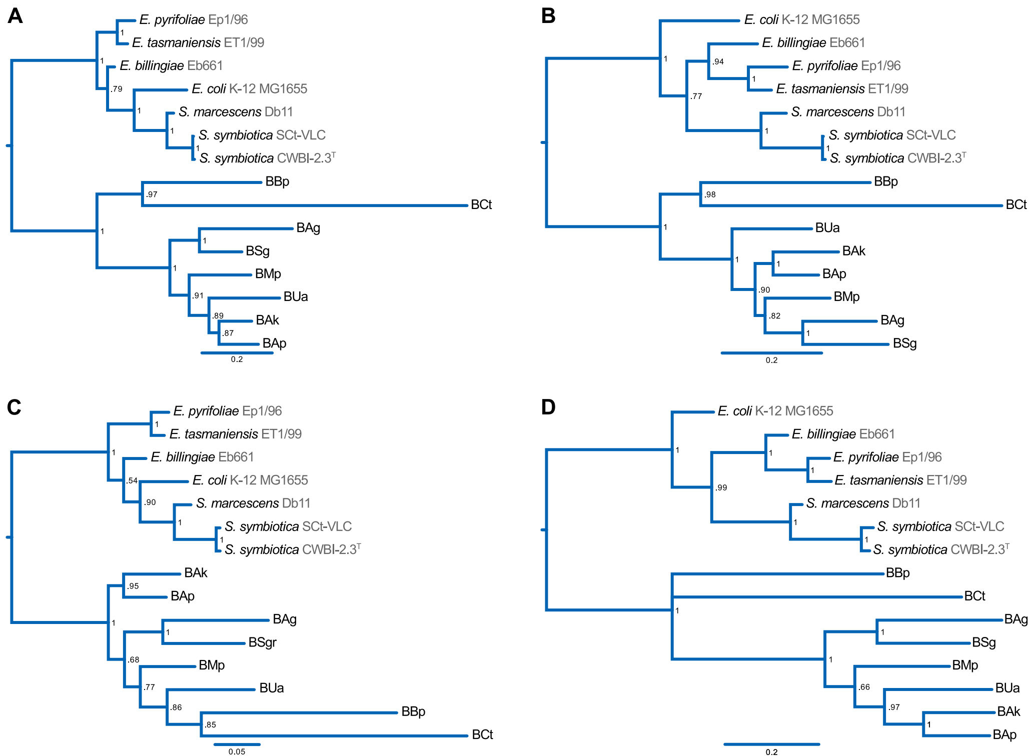


Figure S.6 Phylograms of the *trpD*, *trpC*, *trpB*, and *trpA* genes in *Buchnera*. Bayesian phylograms as calculated by MrBayes. The amino acid substitution models for each gene alignment were chosen using Prottest. **(A)** *trpD* (JTT+I+G+F). **(B)** *trpC* (CpRev+I+G+F). **(C)** *trpB* (CpRev+I+G+F). **(D)** *trpA* (mtRev+G+F). BAg= Buchnera from *Aphis glycines*, BAk= *Buchnera* from *Acyrthosiphon kondoi*, BAp= *Buchnera* from *Ac. pisum* strain APS, BBp= *Buchnera* from *B. pistacea*, BCt= *Buchnera* from *C.* (*Cu.*) *tujafilina*, BMp= *Buchnera* from *My. persicae* strain USDA, BSg= *Buchnera* from *Sc. graminum*, BUa= *Buchnera* from *Uroleucon ambrosiae*. Reproduced from Manzano-Marín *et al.* (2016).

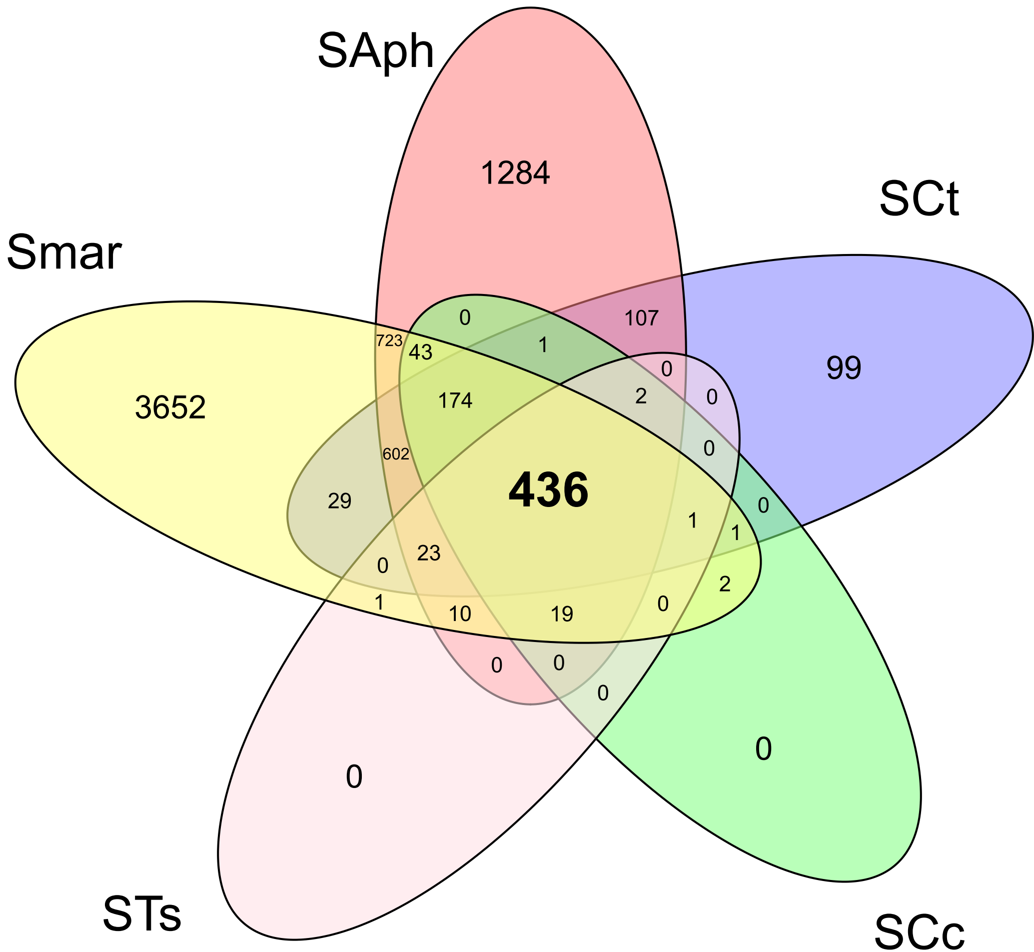


Figure S.7 Shared and unshared protein-coding genes among selected *S. marcescens* and *S. symbiotica* strains. Venn-like diagram displaying the shared (core) and unshared protein-coding genes among selected *S. marcescens* and *S. symbiotica* strains. Smar= *S. marcescens* pan-genome (strains Db11 [PRJEB4201], SM39 [PRJDB1121], WW4 [PRJNA88659]), SAph= *S. symbiotica* Aphidinae strains pan-genome (SAp and SAf). Reproduced from Manzano-Marín *et al.* (2016).

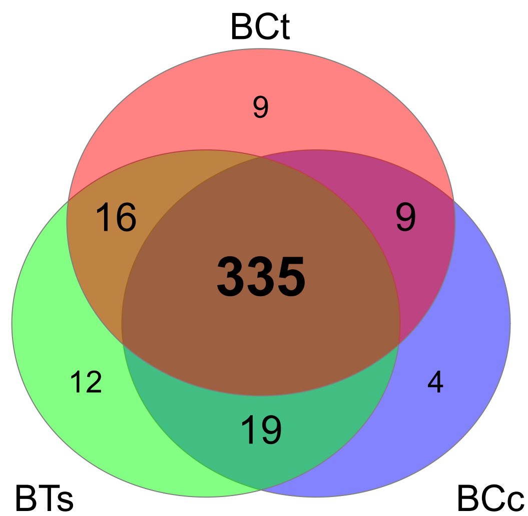


Figure S.8 Shared and unshared protein-coding genes in *Buchnera* from the Lachninae. Venn-like diagram displaying the shared (core) and unshared protein-coding genes among the currently available *Buchnera* strains from the Lachninae. Reproduced from Manzano-Marín *et al.* (2016).

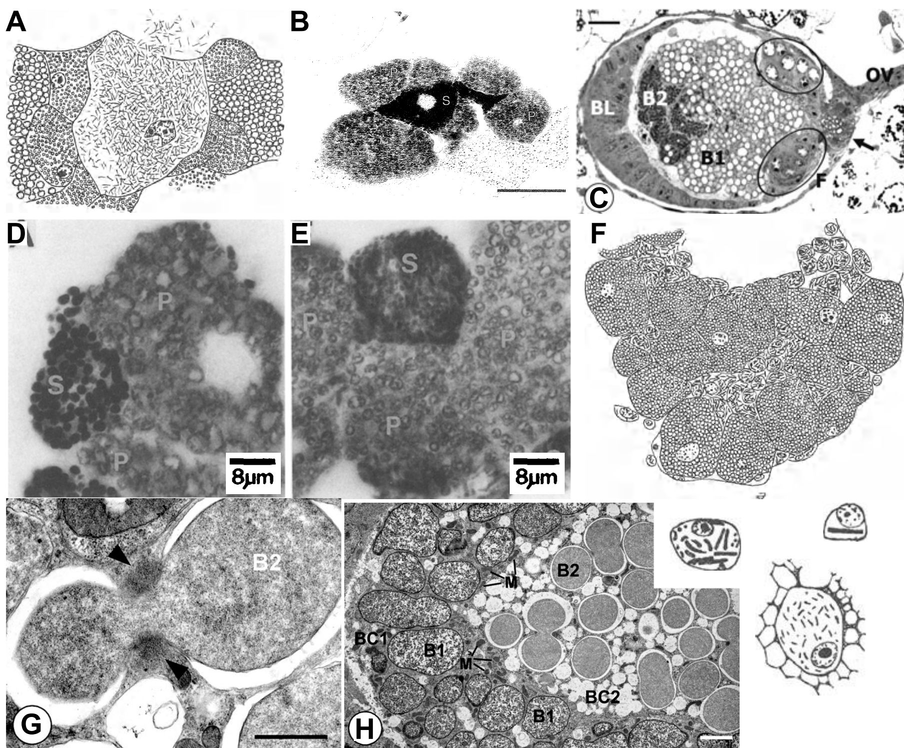


Figure S.9 Morphology and localisation of secondary endosymbionts of Lachninae (A) Detail of the three symbionts of *La. roboris*, 750x. **(B)** *in situ* hybridisation of the endosymbiotic system of *C.* (*Ci.*) *pini* with BIO-EUB338 showing differential staining of symbionts. P=primary symbiont, S=secondary symbiont. **(C)** Longitudinal section through a young emryo of *St. quercus* showing the posterior pole of it invaded by endosymbiotic microorganisms (arrow). Bar=20µm. B1 and B2= bacteriocytes containing different types of endosymbiotic bacteria. **(D** and **E)** Immunohistochemistry of Cpn60 with anti-GroL serum in embryos of *N. piri* (D) and *St. yanonis* (E) showing differential staining of primary and secondary symbionts. P=primary symbiont, S=secondary symbiont. **(F)** *St. quercus* adult bacteriome showing infection with three symbionts and detail of infected haemocytes, 385x and 720x respectively. **(G)** *Eu. rileyi* (viviparous generation) showing a dividing secondary bacterium (B2). **(H)** Fragment of the bacteriome of *C.* (*Sc.*) *pineti* (viviparous generation) with primary (B1) and secondary (B2) bacteria. M=mitochondria. (A and F) Reproduced from Buchner (1953) with permission of Springer. (B) Reproduced from Fukatsu *et al.* (1998) with permission from the author (Takema Fukatsu). (C) Reproduced from Pyka-Fościak and Szklarzewicz (2008). (D and E) Reproduced form Fukatsu and Ishikawa (1993) with permission of Springer. (G and H) Reproduced from Michalik *et al.* (2014).

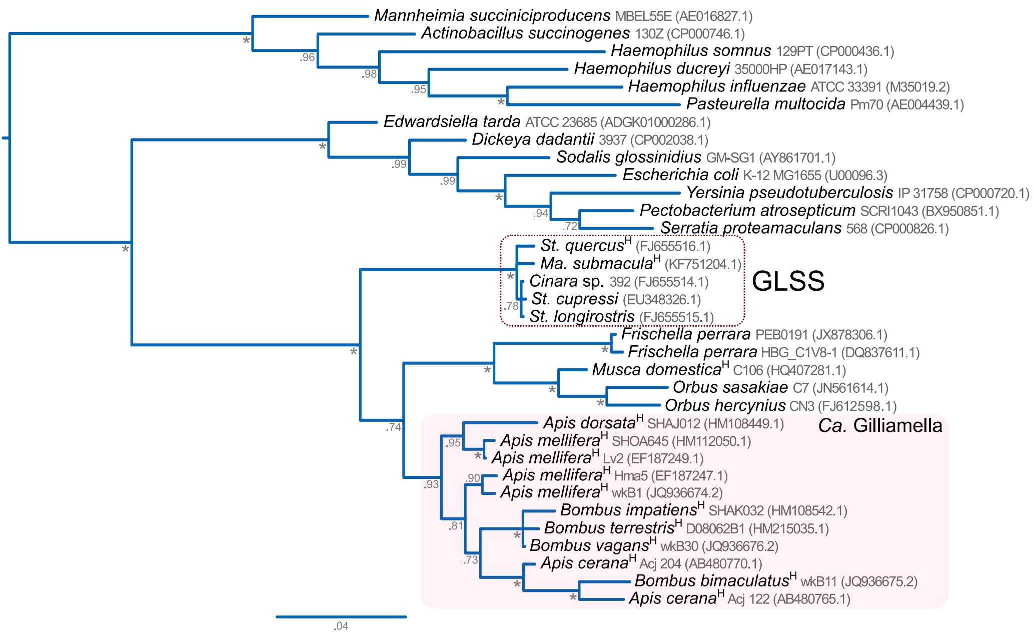


Figure S.10 16S rRNA phylogenetic relationships of GLSS of Lachninae aphids. Bayesian phylogram depicting the relationships and placement of known GLSS from Lachninae aphids and selected Enterobacteriaceae, Pasteurellaceae, and Orbaceae. The superscript "H" at the end of the full species name indicates the symbiont's host name was used. The accession number for each sequence used is indicated within parenthesis after the strain name. Reproduced from Manzano-Marín *et al.* (2016).

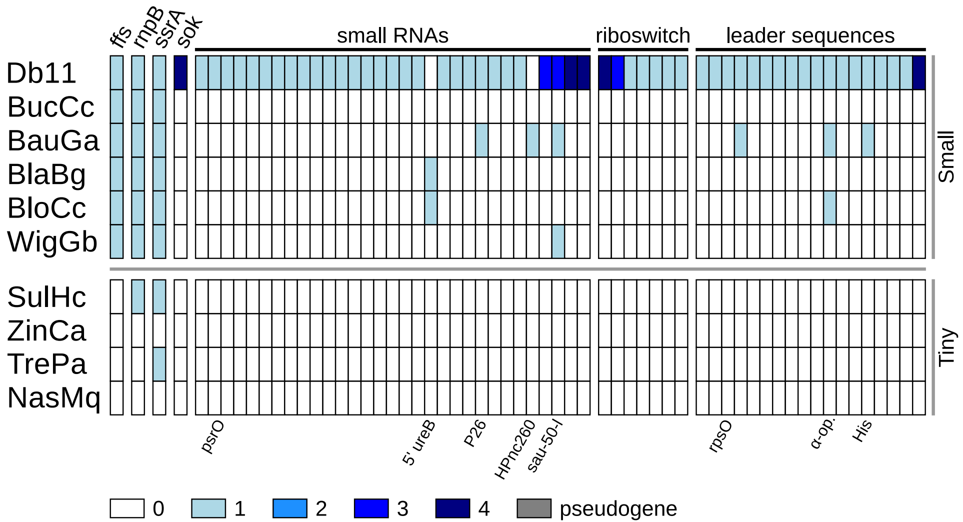


Figure S.11 Decay of RNA features in highly-reduced genomes. Colour-coded diagram showing the decay of RNA features in different highly-reduced genomes comparing against free-living Db11. On top of the matrix, gene names (for the first four columns) and RNA categories (for the rest) are indicated. On the bottom of the matrix, feature names are indicated for those features retained in highly-reduced genomes. Small and tiny genomes are separated by a grey bar. BucCc=*Buchnera* from *C.* (*Ci.*) *cedri*, BauGa=*Baumannia* from *Graphocephala atropunctata*, BlaBg=*Blattabacterium* from *Blattella germanica*, BloCc=*Blochmannia* from *Blochmannia chromaiodes*, WigGb=*Wigglesworthia* from *Glossina brevipalpis*, SulHc=*Sulcia* from *Homalodisca vitripennis*, ZinCa=*Zinderia* from *Clastoptera arizonana*, TrePa=*Tremblaya phenacola* from *Phenacoccus avenae*, NasMq=*Nasuia* from *Macrosteles quadripunctulatus*. Reproduced from Manzano-Marín and Latorre (2016).

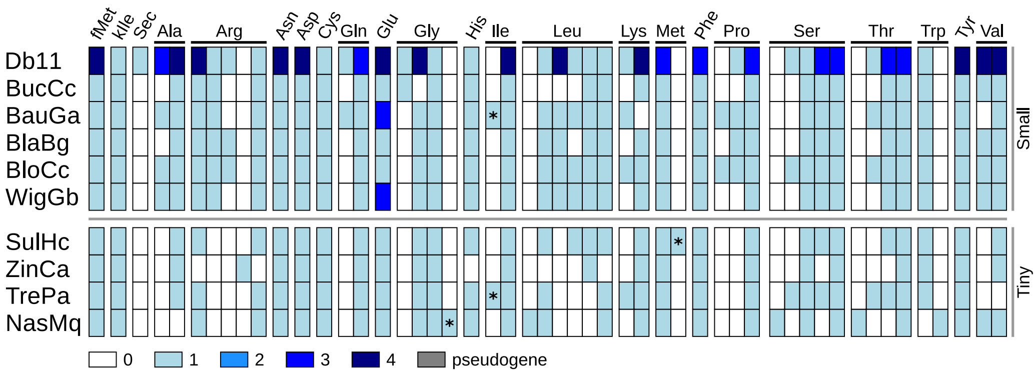


Figure S.12 Decay of tRNA features in highly-reduced genomes. Colour-coded diagram showing the decay of tRNA features in different highly-reduced genomes. On the top of the matrix, aminoacyl charging potential for each tRNA species (as inferred by TFAM). Each column represents a different anticodon. fMet= *N*-Formylmethionine, kIle= lysylated isoleucine. Asterisks indicate putative codon reassignments. Small and tiny genomes are separated by a grey bar. BucCc=*Buchnera* from *C.* (*Ci.*) *cedri*, BauGa=*Baumannia* from *Graphocephala atropunctata*, BlaBg=*Blattabacterium* from *Blattella germanica*, BloCc=*Blochmannia* from *Blochmannia chromaiodes*, WigGb=*Wigglesworthia* from *Glossina brevipalpis*, SulHc=*Sulcia* from *Homalodisca vitripennis*, ZinCa=*Zinderia* from *Clastoptera arizonana*, TrePa=*Tremblaya phenacola* from *Phenacoccus avenae*, NasMq=*Nasuia* from *Macrosteles quadripunctulatus*. Reproduced from Manzano-Marín and Latorre (2016). Reproduced from Manzano-Marín and Latorre (2016).

S.2 Additional tables

Node	Run	Mean date	Std. Err.	Inf. 95%	Sup. 95%
Aphidinae-Lachninae(Fordini)	A	88.7148	16.8568	60.2381	124.325
	B	86.0445	16.7205	59.1453	122.566
Fordini-Lachninae	A	75.4606	21.5947	39.1240	118.544
	B	71.4787	21.5785	37.1950	116.778
Tu. salignus-Cinara	A	55.7354	16.4463	27.9676	88.3250
	B	52.6158	16.5026	26.7994	87.7439
C. (*Cu.*) *tujafilina-C.* (*Ci.*) *cedri*	A	40.0253	12.2196	19.5770	64.8776
	B	37.9805	12.1630	18.6813	64.1443
Aphidini-Macrosiphini	A	58.2083	5.64011	50.3109	69.1486
	B	58.2244	5.61906	50.3005	69.1681
Ap. glycines-Sc. graminum	A	47.1148	5.30120	38.3409	57.9349
	B	47.0618	5.30809	38.2583	58.1197
My. persicae-U. ambrosiae(*Acyrthosiphon*)	A	44.3384	5.07670	35.8392	54.7956
	B	44.4842	5.08954	35.9161	55.0494
Ac. pisum-Ac. kondoi	A	29.8224	4.17634	22.6997	38.8299
	B	30.0263	4.21485	22.7702	38.9895

Table S.1 Statistics of the two independent runs of divergence time estimations for the *Buchnera* lineages. Reproduced from Manzano-Marín *et al.* (2016).

Abbreviations

Abbreviations that appear commonly and are not standardised names for molecules are indicated below.

bps	Base pairs
CDS	Protein-coding sequences
EAA	Esential amino acid
FISH	Fluorescence *in situ* hybridisation
IS	Insertion sequence
kbp	kilobase pairs
LLCA	Lachninae last common ancestor
Mbp	Megabase pairs
MEs	Mobile elements
Mya	Million years ago
ncRNA	Non-coding RNA
NEAA	Non-essential amino acid
Db11	*S. marcescens* from strain Db11
SMLSS	*S. marcescens*-like secondary symbiont
SLSS	*Sodalis*-like secondary symbiont
XTSS	"X-type" secondary symbiont
SAf	*S. symbiotica* strain CWBI-2.3^T from *Ap. fabae*
SAp	*S. symbiotica* strain Tucson from *Ac. pisum*
SCt	*S. symbiotica* strain SCt-VLC from *C.* (*Cu.*) *tujafilina*
SCc	*S. symbiotica* strain SCc from *C.* (*Ci.*) *cedri*
STs	*S. symbiotica* strain STs-Pazieg from *Tu. salignus*
BAp	*Buchnera* strain APS from *Ac. pisum*
BAg	*Buchnera* strain Ag from *Ap. glycines*
BAk	*Buchnera* strain AK from *Ac. kondoi*
BBp	*Buchnera* strain BBpi from *Ba. pistaciae*
BCt	*Buchnera* strain BCt from *C.* (*Cu.*) *tujafilina*

BCc *Buchnera* strain BCc from *C.* (*Ci.*) *cedri*
BMp *Buchnera* strain G002 from *My. persicae*
BTs *Buchnera* strain BTs-Pazieg from *Tu. salignus*
BUa *Buchnera* strain Ua from *U. ambrosiae*