Facultad de Ciencias Biológicas

Instituto de Biología Integrativa de Sistemas

Programa de Doctorado en Biomedicina y Biotecnología

Comparative genomics between two insect symbiotic models: Innate immune system and amino acid biosynthetic pathways of the rice weevil *Sitophilus oryzae* and the cedar aphid *Cinara cedri*

Memoria presentada por Carlos Alberto Vargas Chávez para optar al grado de Doctor por la Universidad de Valencia.

Directores:
Dra. Amparo Latorre Castillo, Catedrática de Genética. Universidad de Valencia.
Dr. Abdelaziz Heddi, Catedrático en el Instituto Nacional de Ciencias Aplicadas de Lyon. Institut National des Sciences Appliquées de Lyon.

Mayo 2019

AMPARO LATORRE CASTILLO, Catedrática de Genética de la Universidad de Valencia y ABDELAZIZ HEDDI, Catedrático en el Institut National des Sciences Appliquées de Lyon,

CERTIFICAN que el trabajo para optar al grado de Doctor en Biomedicina y Biotecnología, y que lleva por título "Comparative genomics between two insect symbiotic models: Innate immune system and amino acid biosynthetic pathways of the rice weevil *Sitophilus oryzae* and the cedar aphid *Cinara cedri*", ha estado realizado bajo su dirección en el Instituto de Biología Integrativa de Sistemas por CARLOS ALBERTO VARGAS CHÁVEZ.

Y para que conste, en el cumplimiento de la legislación vigente, firmamos el presente certificado en Valencia a 27 de mayo de 2019.

Amparo Latorre Castillo　　　　Abdelaziz Heddi

TABLE OF CONTENTS

Table of Contents v

Acknowledgements vii

I General Introduction...9

I.1 Insect origin and evolution 9

I.1.1 The first insects 9

I.1.2 Pterygota: winged insects 11

I.1.3 Endopterygota: insects with complete metamorphosis 12

I.2 Symbiosis in insects 13

I.2.1 Symbiosis 14

I.2.2 Biosynthesis of amino acids in animals and symbiosis 17

I.2.3 Role of symbiosis in insect evolution 20

I.3 Innate immunity of insects 23

I.3.1 The origins of the arthropod immune system 24

I.3.2 Characteristics of the insects' innate immune system 25

I.4 Beetles (Coleoptera) 30

I.4.1 Evolutionary history of beetles 31

I.5.2 Coleoptera as pests 32

I.4.3 Endosymbionts in beetles 34

I.5 The Aphids (Hemiptera: Aphidoidea) 36

I.5.1 Evolutionary history of aphids 38

I.5.2 Aphids as pests 39

I.5.3 Endosymbionts in aphids 41

II Objectives 45

III Materials and Methods 47

III.1 Insects samples and DNA extraction 47

III.1.1 *S. oryzae* rearing and DNA extraction 47

III.1.2 *C. cedri* collection and DNA extraction 48

III.2 Genome assembly and annotation 48

III.2.1 *S. oryzae* genome sequencing and assembly 48

III.2.2 *S. oryzae* genome annotation 49

III.2.3 *C. cedri* genome sequencing and assembly 50

III.2.4 *C. cedri* genome annotation 51

III.2.5 *Wolbachia* Cced genome assembly and annotation 51

III.3 Orthology assignment 51
III.4 Identification of genes involved in the innate immune system pathways 52
III.5 Identification of genes involved in amino acid biosynthetic pathways 53
III.6 Identification of putative horizontally transferred genes 53
IV Chapter 1: The rice weevil *Sitophilus oryzae* 55
IV.1 Introduction 55
IV.2 Results and discussion 58
IV.2.1 Genome assembly and annotation 58
IV.2.2 Gene orthology and gene family evolution 60
IV.2.3 Identification of genes involved in amino acid biosynthesis 61
IV.2.4 Identification of genes involved in immune system pathways 65
IV.2.5 Manual analysis of other specific aspects 70
IV.3 Conclusions 76
V Chapter 2: The cedar aphid *Cinara cedri* 79
V.1 Introduction 79
V.2 Results and discussion 82
V.2.1 Genome assembly and annotation 82
V.2.2 Gene orthology 83
V.2.3 Identification of genes involved in amino acid biosynthesis 84
V.2.4 Identification of genes involved in immune system pathways 88
V.2.5 The third endosymbiont, *Wolbachia* 92
V.3 Conclusions 96
VI Chapter 3: Comparisons between both models..................99
VI.1 Introduction 99
VI.2 Results and discussion 101
VI.2.1 Gene orthology 101
VI.2.2 Identification of genes involved in amino acid biosynthesis 103
VI.2.3 Identification of genes involved in immune system pathways 107
VI.3 Conclusions 112
VII General Conclusions..115
Resumen en Español..119
Bibliography...135
Appendix 169
S.1 Additional Tables 169

ACKNOWLEDGEMENTS

Reaching this point has not been easy, and it would not have been possible at all for me without the help of numerous persons at different stages of my life. Of course, it will be impossible to name everyone, but I will do my best.

Primero que nada, a mi familia, a mis padres que desde mis primeros años se esforzaron por darme la mejor educación posible apoyándome siempre en mis decisiones que han llevado hasta la culminación de este proyecto. ¡Son los mejores! A mi hermana con quien he compartido infinitos momentos y siempre preguntándome sobre lo que hago y mostrándose muy orgullosa. Has sido un gran apoyo siempre. A Marcela, muchas gracias por todas las oportunidades para vivir nuevas experiencias y tener una perspectiva más amplia para ver el mundo. También a mi segunda madre en Valencia, Dioli, me has hecho sentir parte de una familia ¡muchas gracias por todo!

Del laboratorio en especial a Mariana que ha sido una gran amiga a lo largo de muchos años y que fue un gran apoyo a mi llegada en Valencia y con quien compartí tanto todo este tiempo. A Ana por todas las risas, gatos y vinos. A Sergio con quien coincidí también en Lyon por las buenas charlas sobre cualquier tema y cenas. A Maria (sin acento) la murciana que mejor valenciano habla y a Paolo, muchas gracias por los buenos momentos dentro y fuera del laboratorio y los viajes.

A Toni y a los miembros de su grupo, en especial a Damian y Irene. To Cristina, thank you for the motivation and for the wonderful meals and to Rita for the peace and wisdom you spread. To Patrice, thanks a lot for the help and the patience when I started using the servers at the BF2i. To Carole thanks a lot for all the help, both for the project and for helping me with many aspects of the French bureaucracy. To Séverine, thank you for the talks at the cantine and the recommendations of places to visit. To Justin (golden boy) thank you for showing me nice places to get a drink in Lyon. Nico, thanks a lot for everything, without you this project would simply have not been possible, thanks for your patience and your help.

To my two advisors: Amparo, muchas gracias por haber confiado en mí y haberme recibido hace ya bastantes años para empezar este proyecto, siempre me he sentido muy bien acogido en tu grupo. Gracias por la oportunidad de formarme y crecer. To Aziz, thank you for all the enthusiastic lab meetings from which great ideas emerged and for the opportunity of working and learning a lot from an amazing model such as *Sitophilus*. Also, for the chance of becoming a member of your group during both short stays, it was a wonderful experience.

Por último, Asier, no hay mucho que te pueda decir que no sepas ya. Tú has estado a mi lado desde que comencé este proyecto y has vivido a mi lado todo lo bueno y todo lo malo siempre motivándome, dándome ánimos y celebrando los éxitos. Me encanta llegar a casa y hablar contigo sobre cualquier cosa, incluido este proyecto. Has sido el compañero perfecto a lo largo de todo este camino. Ahora termina esta etapa y comienza una nueva y me alegro de que sea a tu lado.

¡Gracias a todos!

I GENERAL INTRODUCTION

I.1 Insect origin and evolution

Insects can be found in nearly every environment on Earth and they are capable of exploiting almost every available food source. They are the most diverse group of animals with an estimated number of extant species of 5 million, although estimates vary widely. All insects have broadly the same body plan consisting of three segments (head, thorax and abdomen), three pairs of legs, one pair of antennae and compounds eyes, however each species has specialized body parts fitting their lifestyle. They can move by crawling, jumping, running, flying, swimming and even striding on the surface of water. Their behaviours are extremely diverse, ranging from predators capable of outrunning, outswimming of outflying their prey to sessile parasitic forms. In addition, while most insects are solitary, some are social, and they live in highly organized colonies reaching populations of millions of inhabitants with castes specialized on a given task.

I.1.1 The first insects

Arthropods (from the Greek ἄρθρον arthron, joint and πούς pous, foot) comprise a phylum of invertebrate animals with a rigid exoskeleton and a segmented body (Budd and Telford, 2009). With over 1.5 million described species it is the most abundant group of animals encompassing more than 80% of the animal species (Zhang, 2013). The Arthropoda lineage originated in a marine environment, yet it has undergone at least three land colonizing events: the earliest for the subphylum Myriapoda 554 million years ago (MYA) and two independent events 495 MYA for the class Arachnida and for the subphylum of the six-legged arthropods, Hexapoda, which comprises insects (Lozano-Fernandez J. *et al.,* 2016; Misof *et al.,* 2016). Their body structure allowed them to migrate to land and withstand different kinds of stress including wider temperature ranges, desiccation and the lack of the support that their former aqueous environment provided. Thus, groups of arthropods living in the tidal zone started colonizing the land, a pristine niche without competitors and safe from water dwelling predators.

Phylogenetic studies propose that the Insecta class, a member of the Arthropoda phylum and the Hexapoda subphylum, emerged during the Ordovician period (443-485 MYA; Misof *et al.,* 2014), though more conservative studies place it during the Silurian (419-443 MYA; Grimaldi*,* 2010; see Figure I.1 for a phylogenetic tree of insects). Nonetheless, the oldest insect fossil is 396–407 million years (MY) old and it belongs to a now extinct genus (Engel and Grimaldi, 2004). Their emergence seems to be related to the appearance of terrestrial plants, which represented a potential novel food source and habitat. These primitive insects belonged to the Apterygota subclass and they were wingless. Additionally, they were hemimetabolous, meaning that they underwent little or even no metamorphosis with the young closely resembling the adults. Around 80 MY later, during the Devonian period, insects were the first animals to develop the spectacular ability of flight, which enabled them to colonize the entire globe (see Figure I.1).

During the Carboniferous (359 to 299 MYA) the Pterygotes, also known as winged insects, experienced a major radiation. This period is characterized by the highest atmospheric contents of oxygen in Earth’s geological history, thus allowing terrestrial invertebrates to reach their largest sizes ever (Engel, 2015). This expansion was followed by another radiation during the Permian (252-299 MYA) on which the Endopterygota, or holometabolous insects (with complete metamorphosis), first appeared. The end of the Permian is marked by the largest extinction event in the history of the Earth. During this incident the marine biodiversity took the heaviest blow losing up to 96% of its species, while on the land two thirds of the tetrapods families went extinct. In the case of insects, eight orders were lost in this event, thus being the only recorded incident in history where insects suffered a mass extinction (Sahney and Benton, 2008). The survivors of this event endured further extinction pulses during the Triassic period (201-252 MYA) and the survivors evolved into the insect orders that are still around today. Most of the modern families emerged during the Jurassic (201 to 145 MYA).

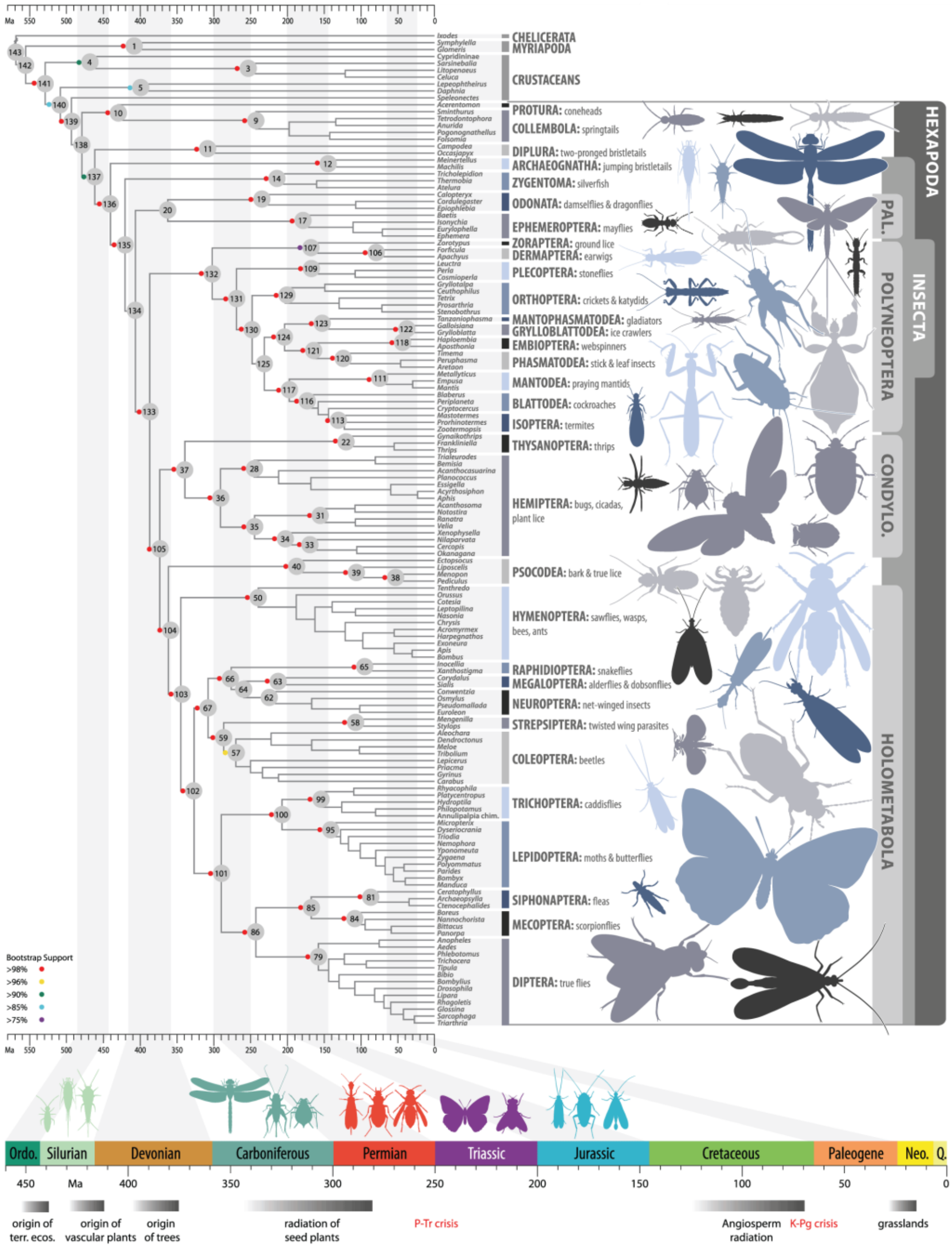

Figure I.1 Dated phylogenetic tree of insect relationships. The time line at the bottom of the tree relates the geological origin of insect clades to major geological and biological events. From Misof, Bernhard, et al. 2014. "Phylogenomics Resolves the Timing and Pattern of Insect Evolution." Science 346 (6210): 763–67. Reprinted with permission from AAAS.

I.1.2 Pterygota: winged insects

The development of insect's wings is probably the most important event in their evolutionary history allowing them to become the most diverse and abundant animal group (Engel, 2015). While the initial stages of the wing evolution are practically missing from the fossil record, it has been proposed that they emerged around 350 MYA. However, phylogenetical studies now set this event around 400 MYA, up to 170 MY before any other animal evolved this ability (Wootton and Kukalova-Peck, 2000; Misof *et al.,* 2014). The first winged insects are known as Paleoptera and surviving groups include dragonflies (Odonata) and mayflies (Ephemeroptera). Subsequent changes allowed a new group of insects, the Neoptera, to fold their wings flat against their body thus enabling major differentiation between the fore- and hindwings for different purposes including defence, communication or thermoregulation (Engel, 2015; Medved *et al.,* 2015).

The Neoptera are further divided into two linages, Polyneoptera and Eumetabola (comprising both Paraneoptera and Holometabola). Polyneoptera is a highly heterogenous group including grasshoppers (Orthoptera), stick and leaf insects (Phasmatodea), mantises (Mantodea) and roaches (Blattaria) among many others (Song *et al.,* 2016). On the other hand, Paraneoptera includes lice (Phthiraptera) and the true bugs (Hemiptera), a group that includes insect groups such as cicadas, aphids, planthoppers, leafhoppers, and shield bugs. Members of the Paraneoptera have a wide array of feeding habits, they can be phytophagous, predatory, or hematophages, among others (Grimaldi and Engel, 2005). A key trait of both Polyneoptera and Paraneoptera is that their members undergo an incomplete metamorphosis with only three stages: egg, nymph (with several stages of development called instars) and adult. While Paraneoptera is an obviously highly diverse group, the group that undergoes a complete metamorphosis, the Holometabola, is responsible for most of the insects' diversity highlighting the relevance of this trait for the success of insects (Grimaldi and Engel, 2005).

I.1.3 Endopterygota: insects with complete metamorphosis

Holometabola encompasses nearly 85% of all insect diversity and contains more species than all remaining animal and plant phyla combined (Grimaldi and Engel, 2005). Holometabolism (also known as complete metabolism) refers to a form of insect development with four stages: egg, larva, pupa and adult. The morphological and behavioural differences between larvae and adults allows them to thrive in different ecological niches without competing and allowing each stage to specialize on a different environment and food source. In general, larvae focus on feeding and growing, thus accumulating essential resources for the following stages while adults focus on mating and dispersal of the eggs.

The emergence of holometabolism is of great significance in the evolutionary history of insects, probably as relevant as the emergence of wings. While the origin of complete metamorphosis in insects has not been fully elucidated, one hypothesis is that the larval stage is analogous to a lengthy pronymph, a brief stage between hatching and the first larval instar, which often goes unnoticed given that sometimes it occurs inside the egg. Conversely, it seems that the pupa is a condensation of all nymphal stages before becoming an adult. While one could be tempted to think that after this trait emerged holometabolous insects immediately experienced a major radiation, in fact, it took millions of years before these adaptations led to the success of this group of insects (Nel *et al.,* 2007; Engel 2015). Additionally, these expansions were not simultaneous, with several radiations occurring through time each taking advantage of different conditions.

The Holometabola includes several well-known orders such as the Hymenoptera that comprises ants, bees and wasps, Lepidoptera (moths and butterflies), Diptera (true flies), and Coleoptera (beetles), the most successful group in terms of species number. Many members of these groups are examples of co-evolution, having reciprocally evolved with flowering plant during the Cretaceous (66-145 MYA; McKenna *et al.,* 2015; see Figure I.1). Some insects have developed a close association with flowering plants and while some damage them through their feeding habits and by being vectors of other diseases, others are pollinators essential for the completion of the life cycle of many flowering species.

I.2 Symbiosis in insects

Symbiosis has played and continues to play a crucial role in eukaryotic organisms' development and evolution. Microbes can biosynthesize metabolites and catalyse chemical processes that animals and plants are unable to, thus most symbiotic relationships are nutritional. While gene transfer between species, and thus transfer of metabolic functions, is widespread among prokaryotes, it does not appear to be common in multicellular eukaryotes. The strategy that has been implemented in several animals has been establishing symbiotic relationships with microbes to complement their metabolic capabilities, thus incorporating a large novel gene set in a single event. The new ecological unit created through these associations is known as holobiont. In addition, symbiosis affects the evolution of the genomes involved in the relationship (or hologenome) by facilitating the loss of genes that are present in all participating partners and by allowing the transfer of genes from one genome to the other.

As a result of these symbiotic relationships, insects have been able to successfully thrive in almost every environment. They can exploit every available food source, sometimes even nutritionally deficient or toxic food sources, given that the microbes they are associated with can provide the missing nutrients, mainly amino acids and vitamins, or detoxify their diet. These relationships have led to a diversification of many insect groups by allowing them to access a formerly unreachable niche. Additionally, these associations commonly lead to the emergence of specialized organs and mechanisms to keep the symbionts under control.

I.2.1 Symbiosis

The term symbiosis was defined by de Bary as “the living together of unlike organisms" (de Bary, 1879) considering the whole spectrum or partnerships, ranging from beneficial to harmful. More recently, Martin and Schwab (2012a) have reviewed the definition of symbiosis recognizing categories and referring to the larger partner as the host and the smaller as the symbiont (Figure I.2). Regarding the fitness of the members of the association, it is common to refer to the association as mutualism when both host and symbiont benefit of the association, parasitism when one member is negatively affected and commensalism when a member benefits without affecting the other.

Additionally, depending on the symbiont’s location and level of integration the interaction can be divided in ectosymbiosis, where the symbiont is located on the host's body surface, including internal surfaces such as the lining of the digestive tube, and endosymbiont where the symbiont is located inside the host's cells or tissues. Endosymbionts are further divided into intracellular, if they are located inside the host's cells either free in the cytoplasm or within membrane derived vacuoles, or extracellular if they are located within host's cavities.

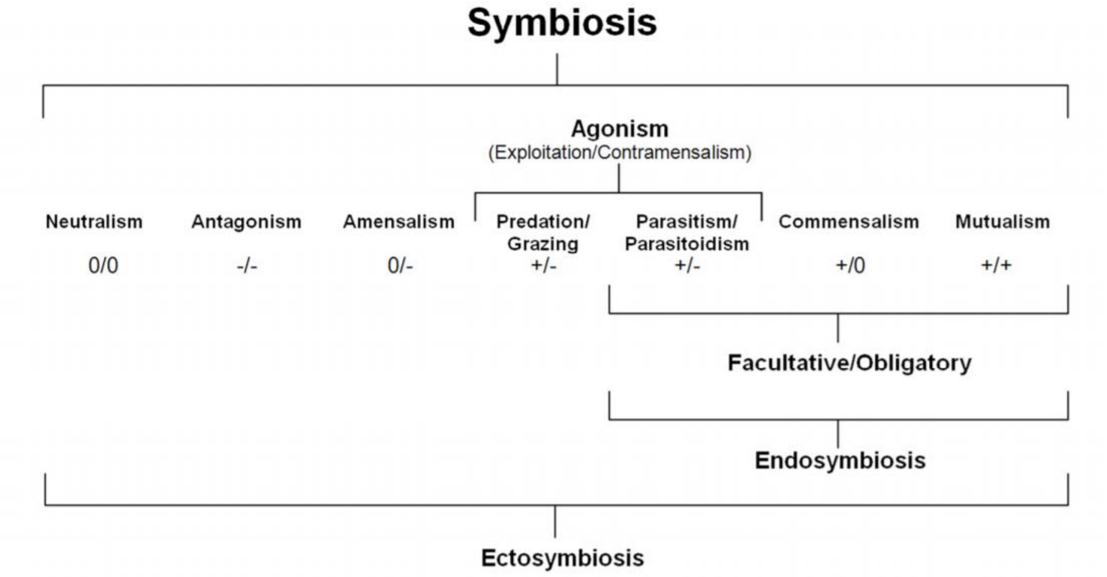

Figure I.2 Scheme of symbiosis. For every pair of interacting species: + represents a beneficial effect, - a harmful effect, and 0 a neutral effect. Endosymbiosis can be subdivided into intracellular and extracellular. From Martin, Bradford D., and Ernest Schwab. 2012b. "Current Usage of Symbiosis and Associated Terminology." International Journal of Biology 5 (1): 32–45.

Several insects harbour a rich gut microbiota comprised by bacteria, archaea, yeasts and protozoa that may permanently reside in the mucosa of the gut or may only pass through the digestive tract along with the food (Engel and Moran, 2013). These microorganisms may provide essential nutrients and aid in the food digestion while also providing protection from colonization by pathogenic microbes to their hosts. Additionally, a common feature of many obligate insect-associated symbionts is the intracellular localization within specialized cells known as bacteriocytes, which can form organs known as bacteriomes (Baumann *et al.,* 2006). Such structures occur in several insect orders including Hemiptera, Coleoptera, Diptera, and Hymenoptera, among others (Douglas, 1989). This project will focus on hosts that have established intracellular mutualistic relationships.

Lastly, the degree of dependency of both partners can be broadly classified as obligate or facultative. An obligate (or primary) symbiosis is required for the correct development, reproduction and survival of one or both partners whereas a facultative (or secondary) symbiont, while beneficial, is not essential and can be removed without much consequences for either member of the interaction.

Endosymbiosis did not arise from one single event. This is easily seen given the existence of bacterial endosymbionts in several branches of the tree of life, notably Gammaproteobacteria, Betaproteobacteria and Bacterioidetes (see Figure I.3). And while it is known that endosymbionts can be acquired from the environment and in some cases replace the original primary endosymbiont, the reasons that originate and allow this phenomenon on insect lineages are not yet well understood (Latorre and Manzano-Marín, 2017).

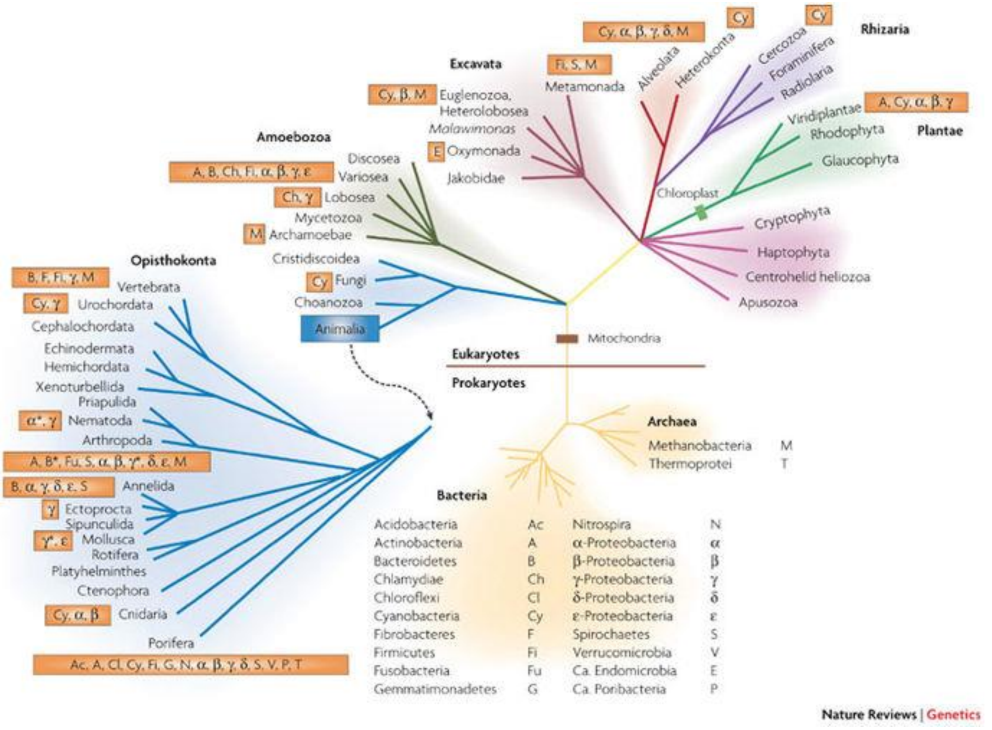

Figure I.3 Phylogenetic distribution of symbioses, indicating the bacterial and archaeal classes within which there are associations with eukaryotic hosts. From Moya, Andrés, et al. 2008. "Learning How to Live Together: Genomic Insights into Prokaryote-Animal Symbioses." Nature Reviews Genetics 9 (3): 218–29.

I.2.2 Biosynthesis of amino acids in animals and symbiosis

Amino acids are fundamental in cellular metabolism and while they have an obvious role in composing proteins, they can also act as chemical messengers or be used as energy metabolites. The origin of the amino acids biosynthetic pathways is central for cellular metabolism and it seems to predate even glycolysis and gluconeogenesis (Cunchillos and Lecointre, 2007), therefore it is highly conserved in all living organisms. However, there are also key differences in the pathways that have evolved to adapt to the unique metabolic needs of each organism.

All animals are heterotrophs meaning that they cannot fix carbon and therefore they must consume organic compounds produced by other organisms as their main energy source. When an organism becomes a consumer by eating other organisms the amino acids and vitamins become available in the diet and no longer need to be synthesized. Unless these biosynthetic pathways serve other essential functions, they are unnecessary and dispensable. The best-known example is the loss in mammals of the ability to synthesize the nine essential amino acids (EAAs), namely: histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine (Payne and Loomis, 2006).

A key requirement for the biosynthesis of amino acids is a nitrogen atom which can be acquired in the form of ammonia or from other amino acids. Additionally, some types of bacteria and archaea are capable of fixating atmospheric nitrogen, but this process is metabolically expensive. Besides nitrogen, the synthesis of all amino acids requires the carbon skeletons from intermediaries of the glycolysis, the citric acid cycle or the pentose phosphate pathway (see Figure I.4). The biosynthetic pathways can be organized in families according to the metabolic precursor of each amino acid as follows: α-ketoglutarate, 3-phosphoglycerate, oxaloacetate, pyruvate, phosphoenolpyruvate and erythrose-4-phosphate, and ribose-5-phosphate.

Glutamate, glutamine, proline and arginine are produced using α-ketoglutarate, an intermediate in the Citric Acid Cycle, as a precursor (see Figure I.4). Glutamate is the first amino acid produced requiring a transamination using another amino acid as the donor. Additionally, glutamate can be deaminated thus allowing the disposal of excess nitrogen as ammonia. Glutamine is produced through the condensation of glutamate and ammonia while proline and arginine require larger sets of reactions. All four non-essential amino acids are produced by humans, but arginine is produced at low rates and is thus considered semi essential; however, birds and most insects are incapable of synthesizing arginine given that they lack the urea cycle.

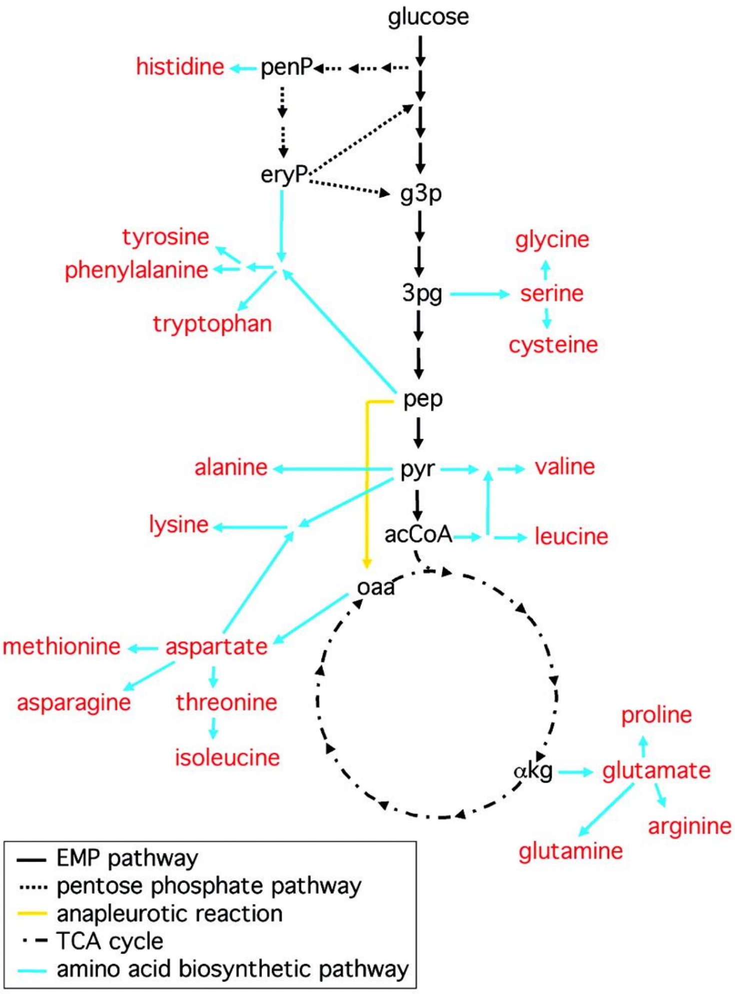

Figure I.4 Fuelling reactions and amino acid biosynthetic pathways in E. coli. Fuelling reactions and amino acid biosynthetic pathways are shown as black and blue arrows, respectively. Akashi, Hiroshi, and Takashi Gojobori. 2002. "Metabolic Efficiency and Amino Acid Composition in the Proteomes of Escherichia Coli and Bacillus Subtilis." PNAS 99 (6): 3695–3700.

Oxaloacetate is another intermediate in the Citric Acid Cycle and it is used to produce aspartate, asparagine, methionine, threonine, lysine and isoleucine (see Figure I.4). The first amino acid produced is aspartate, and similarly to the previous family, it requires a transamination of oxaloacetate using another amino acid as the donor. Animals are also capable of producing asparagine through the amination of aspartate using glutamine as the donor of the ammonium group. However, they are incapable of producing the other four amino acids which therefore must be acquired from their diet and are thus considered to be EAAs.

3-phosphoglycerate is an intermediate from glycolysis and it is used to produce serine, glycine and cysteine (see Figure I.4) and animals can produce all three. Serine is the first amino acid produced and it is further converted to glycine by a serine hydroxymethyltransferase which catalyses the reversible reaction. The formation of cysteine requires a sulphur atom which is transferred from homocysteine, a methionine derivate. Another derivate of glycolysis, pyruvate, can be used to produce alanine through its transamination. Alanine can be further transformed to valine and leucine, but animals are incapable of catalysing these reactions.

The synthesis of the aromatic amino acids phenylalanine, tyrosine and tryptophan requires the condensation of phosphoenolpyruvate and erythrose-4-phosphate and additional processing to generate chorismate. Animals do not have the set of tools required to generate it; however, they can produce tyrosine from phenylalanine through the addition of a hydroxyl group. Lastly, ribose-5-phosphate leads to the formation of histidine, but it is a complex pathway requiring ten reactions and animals are incapable of performing them.

In summary, most animals are incapable of producing ten amino acids, the aforementioned nine EAAs plus arginine, and therefore must acquire them from external sources, namely from their diet. Several animals have opted for different strategies to obtain these amino acids when they are missing in their diet such as establishing nutritional symbiosis with microbes. This has allowed them to conquer environments with low levels of readily available nutrients. In fact, long-term and obligate symbiosis, commonly related with the production of amino acids, has been documented many times in the eukaryotic evolution (Moya *et al.,* 2008, see Figure I.3 and Figure I.5).

I.2.3 Role of symbiosis in insect evolution

Insects can thrive on nutrient-deficient diets such as blood (including the body louse *Pediculus humanus*, Kirkness *et al.,* 2010; the tsetse fly *Glossina morsitans*, International Glossina Genome Initiative, 2014; and the bed bug *Cimex lectularius*, Rosenfeld *et al.,* 2016) and notably plant sap (including the pea aphid *Acyrthosiphon pisum*, The International Aphid Genomics Consortium, 2010; the rice pest brown planthopper *Nilaparvata lugens*, Xue *et al.,* 2014; and the whitefly *Bemisia tabaci*, Chen *et al.,* 2016) when aided by obligate bacterial endosymbionts. Metabolic assays and sequencing studies of several of these endosymbionts determined that they had the capacity to synthesise essential compounds for their insect hosts and in some cases even shared metabolic pathways requiring the transportation of certain metabolites from symbiont to host and vice versa (Zientz *et al.,* 2004; Lamelas *et al.,* 2011; Ponce-de-Leon *et al.,* 2017). While bacterial endosymbionts are predominant, there are also several examples of eukaryotic endosymbionts (Cheng and Hou, 2001; Fukatsu *et al.,* 1994).

Additionally, some insects have established relationships with antibiotic producing bacteria or fungi, which protect them from other pathogens (Currie *et al.,* 1999; Beemelmanns *et al.,* 2016). Furthermore, symbiotic bacteria are also capable of providing protection against pathogens or parasitoids acting as additional components of the immune system (Eleftherianos *et al.,* 2013; Nakabachi *et al.,* 2013). Also, bacteria such as *Wolbachia* and *Rickettsia* can influence host reproductive biology (Anderson and Karr, 2001). Many of these widespread symbiotic relationships between eukaryotes and fully sequenced bacteria are gathered and stored in the Symbiotic Genomes Database (SymbioGenomesDB; Reyes-Prieto *et al.,* 2015).

The first large catalogue of these associations was built by Paul Buchner mainly including hemipteran families with nutritionally challenging diets (Buchner, 1965). To confirm the obligate nature of these endosymbionts, they have been eliminated by means of antibiotic treatment, from several insects including the pea aphid (Ohtaka and Ishikawa, 1991; Koga *et al.,* 2003), the tsetse fly (Nogge, 1976) and some cockroaches (Sacchi *et al.,* 1993). Aposymbiotic (artificially symbiont-free) insects generally display an increased lethality, impaired development and either a decrease in fertility or complete sterility, demonstrating the obligate nature of these endosymbionts.

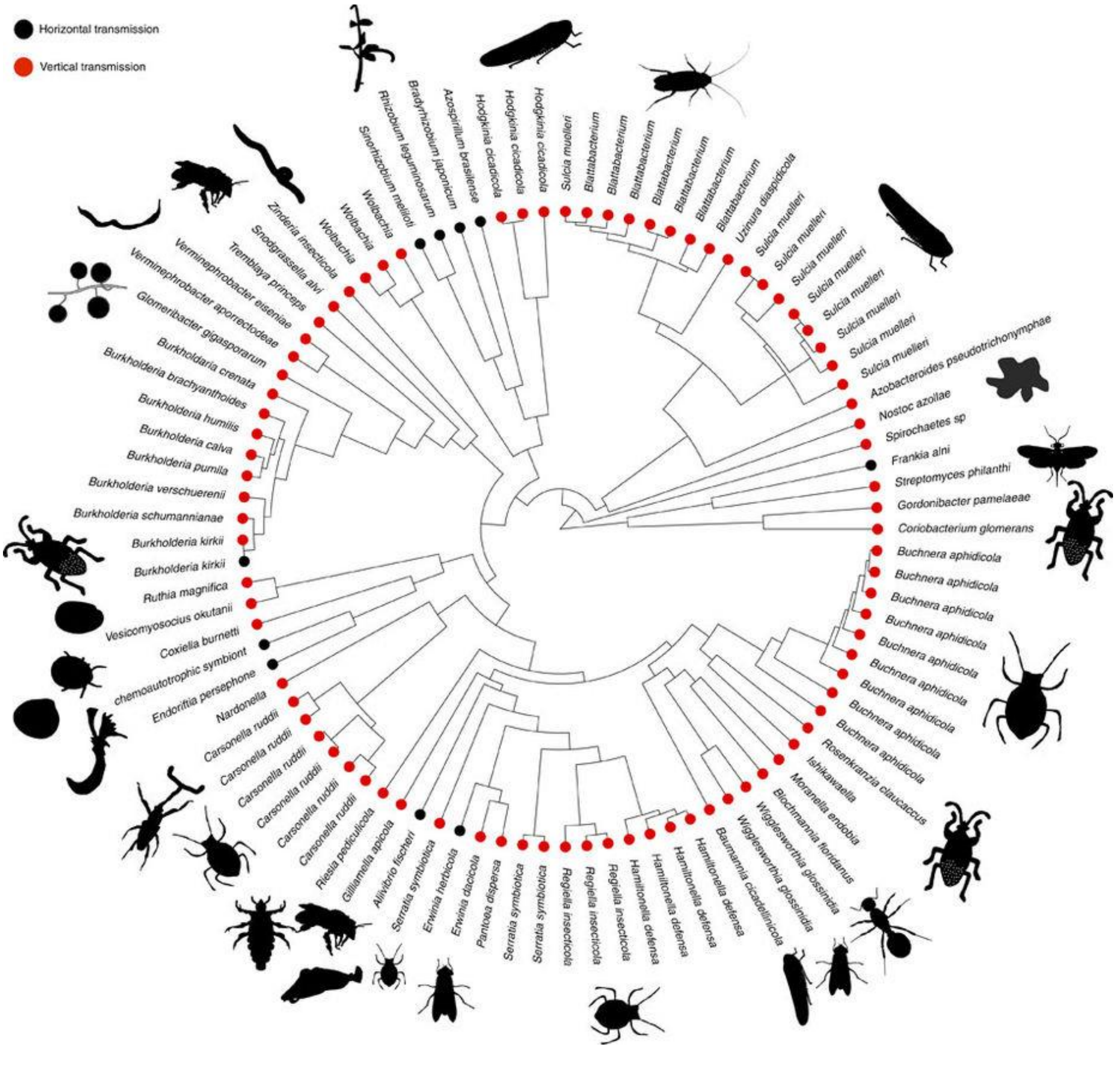

Figure I.5 Bacterial symbiosis across the tree of life. A phylogeny of bacterial symbionts with images of their host groups. Vertically transmitted symbionts are shown with red circles and horizontally transmitted symbionts are shown with black circles. From Fisher, Roberta M., et al. 2017. ''The Evolution of Host-Symbiont Dependence.'' Nature Communications 8: 1–8.

The constant and protected environment of the bacteriome allows a relaxation of natural selection. There, most of the endosymbionts cell functions become redundant and unnecessary and start to become gradually lost (Moran *et al.,* 2008; Moya *et al.,* 2008). Also, there are several examples of enzymes that lose their substrate specificity thus facilitating genome reduction (Price and Wilson, 2014). Furthermore, the endosymbiont is incapable of regaining lost functions through horizontal gene transfer (HGT) given that the bacteriomes seclude endosymbionts from free-living bacteria and the bacterial transmission across host generations occurs generally maternally through female germ cells. In addition, the transmission mode ensures that only a small subpopulation is inherited to the next generation therefore increasing the effect of genetic drift (Moran, 1996). These effects are accelerated after the DNA repair and recombination mechanisms are lost from the endosymbiont's genome (Moran *et al.,* 2008; McCutcheon and Moran, 2012; Wernegreen, 2015). This leads to the irreversible fixation of deleterious mutations in a process known as Muller's ratchet.

Another feature of obligate endosymbionts is a bias towards an Adenine and Thymine (AT)-rich genome, with some exceptions such as *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). While the causes behind this phenomenon have not been conclusively defined, there are several explanations including the higher energetic cost of the biosynthesis of the guanosine and cytosine triphosphate when compared to the adenine and thymidine triphosphate (Rocha and Danchin, 2002), the increase in cytosine deaminations leading to cysteine to thymidine transitions due to the loss of DNA repair genes (Klasson, 2006) and deletions due to polymerase slippage in G+C rich regions (Clayton *et al.,* 2016).

In many cases, the highly reduced primary endosymbiont genomes are static, such as in the case of *Buchnera aphidicola* (Roeland *et al.,* 2003), *Portiera aleyrodidarum* (Santos-Garcia *et al.,* 2015) and *Ca*. Tremblaya princeps (Husnik and McCutcheon, 2016) where an almost perfect synteny is observable among strains. On the other hand, on the genus *Serratia* it has been possible to define the main features on the process from free-living to intracellular endosymbiont, as well as the switch from secondary to coprimary (Manzano-Marín *et al.,* 2016). Furthermore, there are other complex cases such as *Ca.* Hodgkinia cicadicola, an endosymbiont of cicadas, which has split into two new species while sharing the same bacteriocytes and being metabolically interdependent (Van Leuven *et al.,* 2014; Campbell *et al.,* 2015). However, the most spectacular and complex case is that of *Ca.* Tremblaya phenacola str.

PPER, an endosymbiont of the bougainvillea mealybug, with a chimeric genome product of the integration of the genomes of an ancestral Betaproteobacteria and a Gammaproteobacteria (Gil *et al.,* 2018).

In summary, symbiosis between insects and bacteria has evolved as a strategy to confer chemical reactions, metabolic pathways of functions that have been lost in eukaryotes. In return, the primary bacterial endosymbionts benefit from a stable and protected environment where they can proliferate and be transmitted to the next generation. A consequence of this is a relaxation of natural selection and thus an accelerated loss of redundant and nonessential genes. This leads to the highly reduced genomes of endosymbionts. Nonetheless, the growth of the endosymbionts is kept under control through several mechanisms of their hosts like limiting the availability of key metabolites or the immune system as we will see in the next section.

I.3 Innate immunity of insects

Given the cosmopolitan nature of insects they are exposed to a great diversity of pathogens of all types, therefore they have developed a wide variety of strategies to face these threats. Their first barrier is their strong exoskeleton, the cuticle, which shields them from physical trauma and from the invasion of most pathogens (Balabanidou *et al.,* 2018). However, in case this tough barrier fails, their immune system is able to face all types of threats either bacterial, fungal or viral.

The immune system can be classified in two main categories: adaptive and innate. The innate immune defences are described to be non-specific, act immediately and do not confer long-lasting immunity. On the other hand, the adaptive immunity mounts a tailored response against a given antigen while also storing the necessary information to quickly mount a response against a repeatedly infecting pathogen. Like all invertebrates, insects only display innate immunity. However, while immune regulations have been well deciphered in relation to microbial pathogens, very little is known about host immune interaction with beneficial symbionts. In other words, how the immune system is regulated to both fight against pathogens while preserving and regulating mutualistic symbionts.

I.3.1 The origins of the arthropod immune system

From the beginning of life, organisms have been forced to defend themselves from attacks by microorganisms. Throughout life, all living cells have selected different strategies and molecular mechanisms to cope with invaders. Bacteria must defend themselves mostly from phages, and to do so they use several mechanisms including the well-known CRISPR and restriction-modification systems among many others (Doron *et al.,* 2018). After the evolution of eukaryotes new strategies emerged, notably the ability to differentiate self from non-self and phagocytosis. Both traits were initially used to allow the first eukaryotic cells to phagocytise foreign materials mainly as a food source.

The appearance of multicellularity one billion years ago allowed more complex molecular functions to emerge to distinguish self from non-self (Buchmann, 2014). During the transition to Metazoa, molecules with the role of recognizing other cells appeared and the functions of phagocytosis started being used to eliminate non-self particles. Even sponges, the most basal group of Metazoans, have the required mechanisms to recognize, encapsulate and eliminate pathogens (Müller and Müller, 2003). While plants have similar mechanisms to identify and eliminate pathogens, it has been suggested that these similarities are due to convergence rather than having a common ancestor (Haney *et al.,* 2014).

Given its high relevance in human health, the immune system has been widely studied predominantly in vertebrates, and more recently in invertebrates as well. While there are great differences between both groups of animals, most notably, as stated before, the presence of the adaptive immune system only in vertebrates, there are also important similarities. Palmer and Jiggins suggested an ancient origin for the innate immune system, with examples of conservation between vertebrates and arthropods, despite the split between this two groups having occurred around 600 MYA (Palmer and Jiggins, 2015).

I.3.2 Characteristics of the insects' innate immune system

Insects can mount a strong innate immune response to deal with the invasion of microbial pathogens. Several holometabolous insects, notably *Drosophila melanogaster* (Diptera, Drosophilidae; De Gregorio *et al.,* 2001), *Anopheles gambiae* (Diptera, Culicidae; Christophides *et al.,* 2002), and *Tribolium castaneum* (Coleoptera, Tenebrionidae; Zou *et al.,* 2007), are traditional models for genomic and functional investigations of insect innate immunity. The high level of similarity in the core components of the immune system among these insects is suggestive of a highly conserved innate immune system. Nevertheless, there are significant differences depending on the lifestyle of the insect. Some striking examples are some sap-feeding hemipterans including the pea aphid (*A. pisum*) and the Asian citrus psyllid (*Diaphorina citri*) that have lost key elements of signalling pathways (Gerardo *et al.,* 2010; Arp *et al.,* 2016).

The innate immune system includes cellular and humoral responses and is the second line of defence against microbial infections in insects after the cuticle has failed. These defence responses are triggered by pattern recognition receptors (PRRs), which detect and bind to conserved microbial surface structures known as microbe-associated molecular patterns (MAMPs). The humoral immune response leads to the production of antimicrobial peptides (AMPs) that protect against a broad array of microbial infectious agents and even eukaryotic parasites.

The molecular mechanisms involved in these defence reactions have been most thoroughly studied in *D. melanogaster* (De Gregorio *et al.,* 2002; Hoffmann, 2003; Leclerc and Reichhart, 2004; Lemaitre and Hoffmann, 2007; Valanne *et al.,* 2011; Myllymäki *et al.,* 2014). In this model organism, genes encoding AMPs are activated by NF-κB transcription factors in response to infection through the two main immune system signalling pathways: The Toll and the immune deficiency (IMD) signalling pathways. Other well-known immune pathways in insects are the JNK and JAK-STAT pathways, which participate in cell stress or wound response while the JAK-STAT pathway also participates in antiviral response (Dostert *et al.,* 2005). Efforts have been made to catalogue all these components among insects (Brucker *et al.,* 2012). (See Figure I.6 for a summary)

The IMD pathway (Figure I.6) is involved in the recognition of diaminopimelic acid (DAP)-type peptidoglycan (DAP-PG) of bacteria, a component of the cell wall in most Gram-negative bacteria, and some Gram-positive *Bacillus* and *Listeria* species, thus it is considered to be almost specific for Gram-negative bacteria (Le Bourhis *et al.,* 2007). The main receptors associated with this pathway are the peptidoglycan recognition proteins (PGRP) PGRP-LC, and PGRP-LE which interestingly can sense both extra- and intracellular MAMPs. These proteins bind to the peptidoglycan from the bacterial cell wall of Gram-negative bacteria and initiate the signalling cascade by activating the death domain protein IMD, which recruits the adaptor protein FADD that in turn recruits the caspase Dredd. Then, Dredd is activated through its ubiquitination, performed by IAP2 in a complex with Uev1A, Effete and Bend, and upon activation cleaves IMD, exposing a new site which can be ubiquitinated by the IAP2 complex. Active IMD is then capable of activating the protein kinase TAK1 that triggers both the JNK pathway (through Hep, Bsk, Jra and Kay) and phosphorylation of nuclear factor Relish (through IRD5 and Kenny) promoting its translocation into the nucleus and initiating the transcription of genes including AMPs. While the members of the signalling pathway are conserved among insects, usually with a one-to-one relationship (Brucker *et al.,* 2012; Xia *et al.,* 2015), the receptors and effectors are highly diverse.

The Toll pathway (Figure I.6) is involved in the recognition of LYS-type peptidoglycan (LYS-PG) mainly from Gram-positive bacteria, and fungi. This pathway activates the production of antimicrobial peptides, induces the propagation and differentiation of haemocytes and promotes phagocytosis and encapsulation of parasites (Valanne *et al.,* 2011). The pathway begins with the extracellular recognition of virulence factors followed by three different protease cascades depending on the type of recognized organism. The first involves the activation of protease Psh which recognizes both fungal and Gram-positive bacterial virulence factors. The second is involved in the recognition of β-glucan from the fungal cell wall through the Gram-negative binding protein (GNBP) 3, and the third in the recognition of LYS-PG through GNBP1, PGRP-SA and PGRP-SD converging at the modular serine protease ModSP. In turn, this protease activates the Gram-positive–specific serine protease Grass, which activates the proteases Sphinx1/2, Spirit and Spheroid. At this point all three cascades converge with the activation of the Späetzle processing enzyme (SPE) which cleaves späetzle allowing it to bind to the extracellular region of the Toll receptors, which also have an intracellular domain whose role is to initiate the intracellular signalling process (Cao *et al.,* 2015; see Figure I.6).

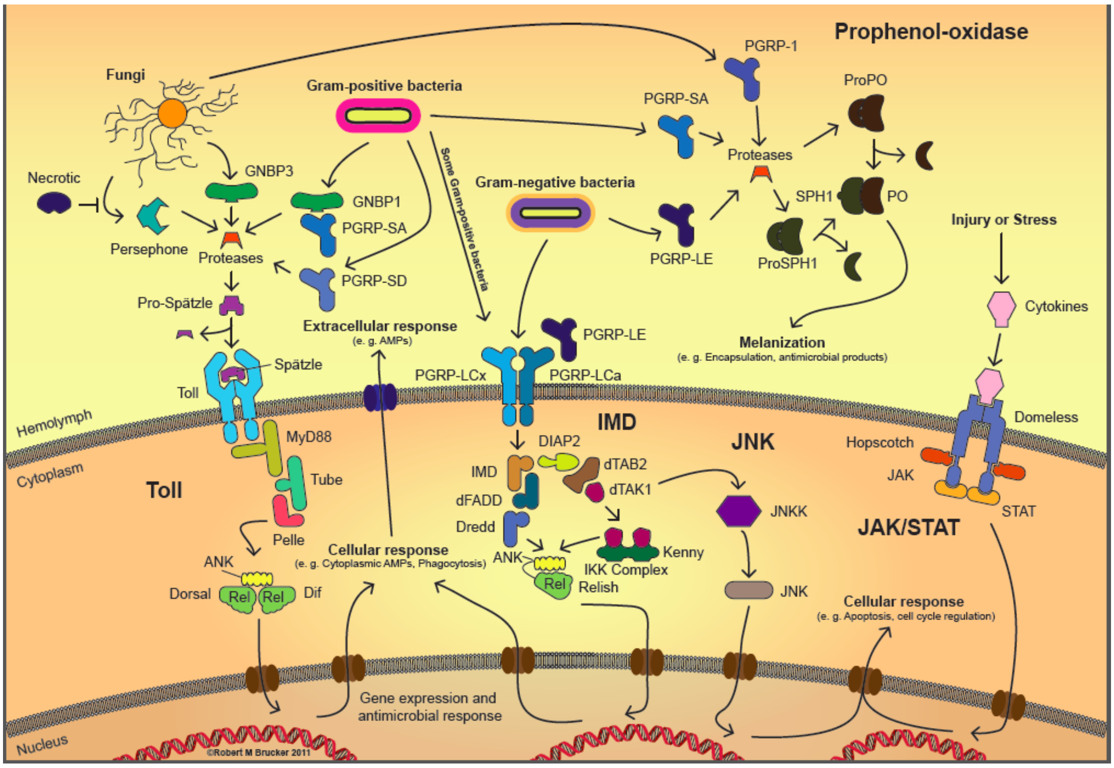

Figure I.6 Insect innate immune system pathways based on Drosophila literature. Only the most relevant proteins from each pathway are displayed. Please refer to the text for more details. From Brucker, Robert M, et al. 2012. "Insect Innate Immunity Database (IIID): An Annotation Tool for Identifying Immune Genes in Insect Genomes." PLoS ONE 7 (9): e45125.

Upon activation of the Toll receptor, the adaptor protein MyD88, Tube and the kinase Pelle are recruited. Additionally, it seems that Pellino interacts with Pelle ubiquitinating it, thus acting as a positive regulator. Next, Pelle phosphorylates Cactus leading to its ubiquitination and subsequent degradation. Cactus is generally bound to the NF-κB transcription factor(s) Dorsal and/or Dif inhibiting its translocation to the nucleus. The Atypical protein kinase C (aPKC), along with Ref(2)P and TRAF2, and the G protein-coupled receptor kinase-2

(GPRK2) also seem to be able to interact with Cactus (or directly with Dorsal) to promote Dorsal activation. Once in the nucleus, Dorsal and/or Dif promotes the transcription of effector genes, including AMPs.

The JAK/STAT pathway (Figure I.6) is activated through the binding of the ligand Upd by the Dome receptor. The Janus Kinase protein (JAK, or Hopscotch) is normally associated with the Dome receptor and upon binding of the ligand, the receptor stimulates JAK which in turn phosphorylates itself thus creating a binding site for the Signal Transducer and Activator of Transcription (STAT) and afterwards phosphorylating it. Upon phosphorylation STAT dimerizes and moves to the nucleus where it activates the transcription of target genes (Zeidler and Bausek, 2013).

The main effectors of the insects' immune system are AMPs. These are small peptides (<100 amino acids), highly diverse and in many cases species specific, though they might display broad-spectrum antimicrobial activity. Their modes of action include inhibition of microbial proliferation, tagging and promotion of aggregation and lysis of the invaders. Based on their structures or unique sequences, insect AMPs can be classified into four families: (i) the α-helical linear peptides without cysteine residues (cecropin and moricin), (ii) β-sheet globular cysteine-rich peptides (defensin and drosomycin), and peptides with an unusually high content of a given amino acid, (iii) apidaecin, drosocin, and lebocin are rich in proline, and (iv) coleoptericin, diptericin, attacin and gloverin in glycine (Yi *et al.,* 2014; Mylonakis *et al.,* 2016). Another group of antimicrobial enzymes are lysozymes, a protein family that defends against bacteria by cleaving peptidoglycans within cell walls, especially of Gram-positive bacteria. Additionally, thaumatins are described as anti-fungal proteins with a glucanase function first described in plants (Brandazza *et al.,* 2004; Altincicek *et al.,* 2008; Anselme *et al.,* 2008), although their precise mechanism of action is still unknown.

While some of the aforementioned pathways also participate in controlling the effects of viral infections, antiviral RNA interference (RNAi) has been described to silence viral RNA in a sequence specific manner. There are three main RNAi pathways (Karlikow *et al.,* 2014; Cao *et al.,* 2016): i) small interfering RNAs (siRNAs), which can be produced from exogenous or endogenous double stranded RNAs (dsRNAs) and depend on the Dicer (Dcr) -2, R2D2 and Argonaute (Ago) 2 proteins, ii) microRNAs (miRNAs), which are mostly encoded in intergenic regions and are dependent on Drosha and Pasha proteins in the nucleus and on the Dcr-1 and Ago1 proteins and iii) Piwi-interacting RNAs (piRNAs) which are encoded by clusters of genes and are involved in epigenetic and post-transcriptional gene silencing of transposons and possibly in the antiviral response and are dependent on Piwi, Aubergine (Aub) and Ago3 proteins and are Dcr independent.

The activity of the siRNA pathway begins with the recognition and cleaving of viral dsRNAs by Dcr-2 (Galiana-Arnoux *et al.,* 2006; see Figure I.7). This generates siRNAs that are loaded into Ago2, which is then capable of recognizing target RNA using the guide RNA, afterwards the targeted viral RNA is degraded. The miRNA pathway begins with the transcription of a miRNA gene that adopts a hairpin structure. This structure is recognized by the Drosha/Pasha complex, cleaved into a pre-miRNA and exported to the cytoplasm. There, Dcr-1 cleaves the pre-miRNA into its mature miRNA form. The miRNA is loaded into Ago1 and, as in the siRNA pathway, the target gene is silenced. Finally, the piRNA pathway involves the transcription of a precursor and its recognition and processing by Piwi or Aub. Afterwards, through a process known as ping-pong amplification loop it increases the number of copies by recognizing both the sense and antisense transcripts using Ago3 and Aub respectively and processing their targets.

As we have seen, the immune system is highly complex and capable of detecting and fending off pathogens of all kinds. However, it is presented with an extraordinary challenge. While it should be capable of detecting and eliminating pathogens, it must still allow beneficial microorganisms to thrive in the surface or in the gut of the insect. The interactions between the insect gut microbiota and the immune system have been studied in *Drosophila* (Broderick and Lemaitre, 2012) and it is known that both have a profound influence on each other. Regarding the interaction between endosymbionts and the immune system, one of the best studied models is *Sitophilus oryzae* (Coleoptera, Dryophthoridae), as we shall see in detail next.

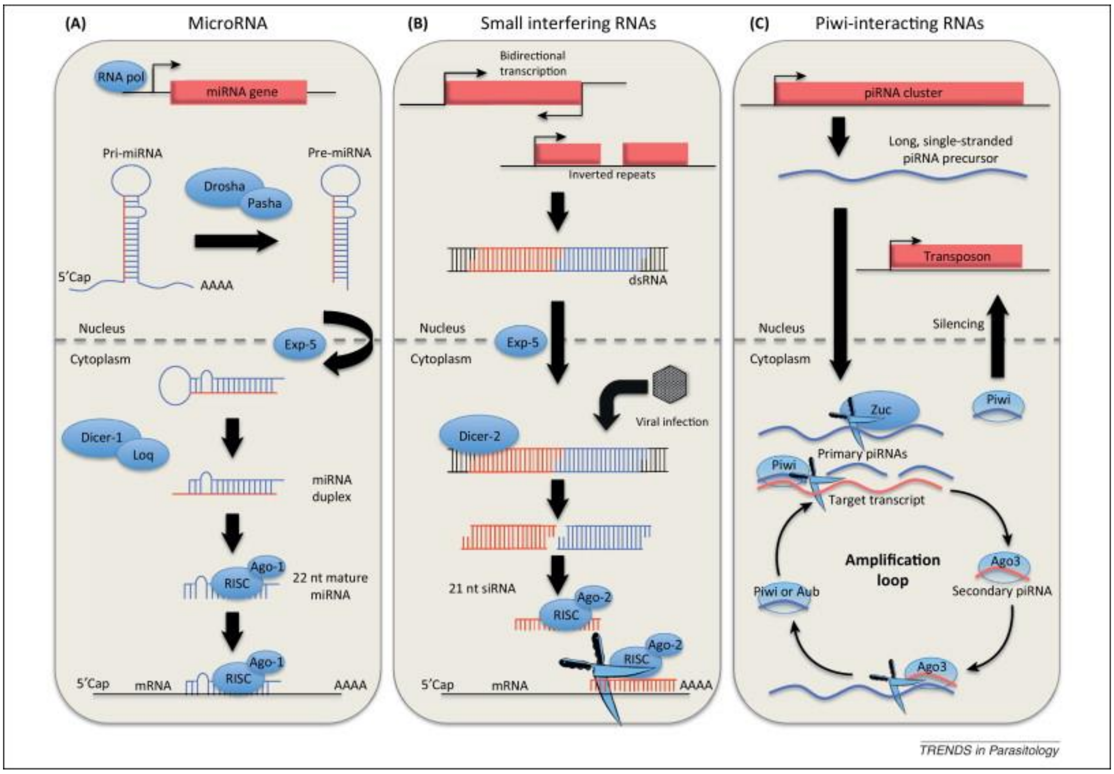

Figure I.7 Three main classes of regulatory small RNAs. Only the most relevant proteins from each pathway are displayed. Please refer to the text for more details. From Lucas, Keira J., et al. 2013. "Small RNAs: A New Frontier in Mosquito Biology." Trends in Parasitology 29 (6): 295–303.

I.4 Beetles (Coleoptera)

A possibly apocryphal, but brilliant reply attributed to the British geneticist J. B. S. Haldane when asked what could be inferred about God from a study of his works was "An inordinate fondness for beetles." Beetles (order Coleoptera) are the largest order of animals with an estimated number of described species of 400,000 (Hammond, 1992) and an extraordinary degree of diversity. Some examples of their diversity include the massive Hercules beetles which can reach 17.5 cm, being the longest beetle and among the largest insects in the world; the featherwing beetles, the smallest insects in the world, whose size ranges from 0.3 to 4 mm; the remarkable bombardier beetle which can spray its predators with near boiling liquids; and firefly beetles whose glow is so appealing to us. Remarkably, in this order the largest family comprises the true weevils, Curculionidae, with more than 51,000 species, arguably the most species-rich family of animals (Oberprieler *et al.,* 2007). This high degree of success can be attributed to several factors including short life cycles, high fertility, a highly protective exoskeleton and a low extinction rate (McKenna *et al.,* 2015).

I.4.1 Evolutionary history of beetles

The order Coleoptera is divided into four suborders: Adephaga, most of them are predators; Archostemata, whose larvae feed on decaying wood and adults on pollen; Myxophaga, which are usually aquatic and feed on algae; and by far the largest is Polyphaga, with a huge diversity on their feeding habits though most feed on plants, either healthy or dead and decaying. The earliest fossil from Archostemata is from the late Permian and fossils from Adephaga and Polyphaga are from the Triassic. While there are no fossil specimens from the Myxophaga some likely extinct relatives are known from the Permian (McKenna and Farrel, 2009). The exact date of the origin of beetles is not easy to determine given the scarcity of fossils of the early beetles. However, they are thought to have originated during the Early Permian, and a fossil from the Carboniferous suggests an even earlier origin (Bethoux, 2009).

Given the lack of information in the fossil record, molecular studies have been used to better define the date of origin. Nonetheless, this is also challenging given the enormous diversity of beetles and has mostly been insufficient for solving the macrostructure of their phylogeny (Hunt *et al.,* 2007; Bocack *et al.,* 2014). A recent study by Zhang and colleagues has attempted to establish a reliable phylogeny by using 95 nuclear protein-coding genes from 373 species comprising all suborders and 124 out of the 186 recognized families (Zhang *et al.,* 2018; see Figure I.8).

The work of Zhang and colleagues supports the hypothesis that Coleoptera originated during the early Permian and that they experienced intense radiation during the Cretaceous. While this radiation has traditionally been attributed to the emergence of angiosperms, Hunt and colleagues suggested that their diversity is not linked to the diversification of flowering plants but rather to their long evolutionary history and high lineage survival (Hunt *et al.,* 2007; Zhang *et* *al.,* 2018). Whichever is the case it ultimately led them to be the most abundant group of animals on the planet.

I.5.2 Coleoptera as pests

The wide variety of beetles, their cosmopolitan nature and their abilities to exploit almost every food source has put them into close contact with humans throughout all our history and while sometimes forgotten, it is important to keep in mind that beetles can have beneficial roles in our economy. A well-known example is lady beetles (Coccinellidae) that can control the populations of sap feeding pests including aphids, whiteflies and mealybugs by feeding on them. Another group of beetles that have a positive effect on agriculture are ground beetles (Carabidae), which are mostly predatory, actively hunting for invertebrates, thus sometimes used in crops as biological control agents. A perhaps not so evidently beneficial group of beetles is dung beetles (Scarabidae), which feed on dung after quickly rolling and burying it, thus rendering it unavailable for breeding of pestilent flies that target cattle. Additionally, the burying of the dung improves the soil fertility.

In contrast, given their predominantly phytophagous nature (around 75% of the beetles are herbivorous during both their larval and adult stages; Gillot, 2005), they are well known agricultural pests. While some of them are monophagous, specializing on a single plant host, other are highly polyphagous being able to devastate diverse crops. Additionally, besides feeding on the plants they are also vectors of numerous diseases and to make matters worse they can quickly adapt and become resistant to pesticides (Odeyemi *et al.,* 2010).

Pests include the Colorado potato beetle (*Leptinotarsa decemlineata*; Schoville *et al.,* 2018), which is a serious pest of members of the Solanaceae family that includes tomatoes, eggplants and peppers. Another example is the apple blossom weevil (*Anthonomus pomorum*), which, although not a problem in commercial apple orchards, is a serious pest in organic orchards. However, perhaps the most relevant coleoptera pests are those that target stored goods such as cereals, including the red flour beetle (*T. castaneum*; Tribolium

Genome Sequencing Consortium, 2008), the sawtoothed grain beetle (*Oryzaephilus surinamensis*) and cereal weevils of the genus *Sitophilus*.

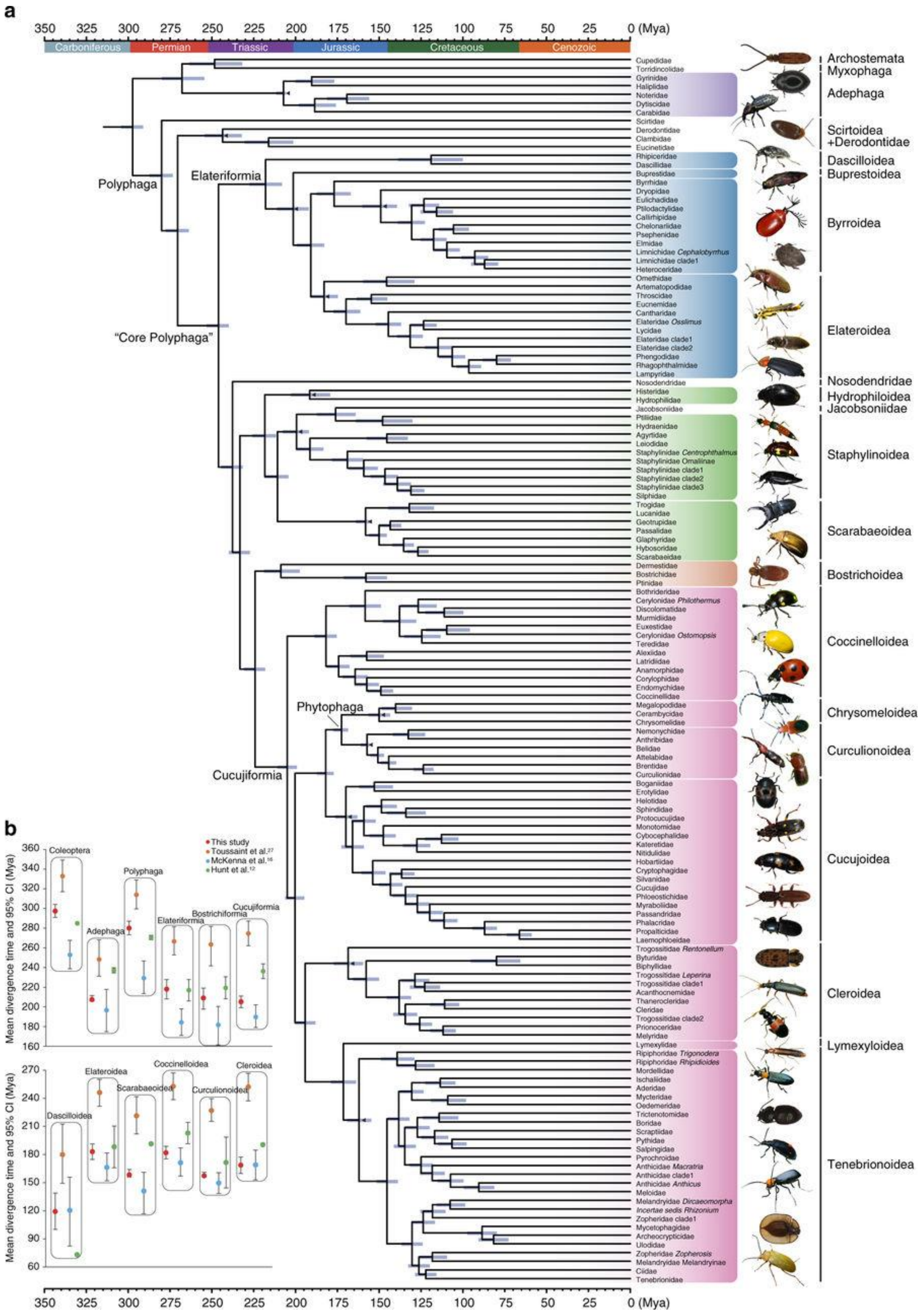

Figure I.8 Time-calibrated tree of beetle families. b. Comparison of divergence time estimates for twelve major nodes sharing across four beetle time trees. The circle represents the mean age, and the whiskers mark the 95% credibility internals. From Zhang, Shao Qian, et al. 2018. "Evolutionary History of Coleoptera Revealed by Extensive Sampling of Genes and Species." Nature Communications 9 (1): 1–11.

Additionally, xylophagous beetles also cause losses in other commodities such lumber through their feeding activities. Some examples are the Asian longhorned beetle (*Anoplophora glabripennis*; McKenna *et al.,* 2016), which feeds on a wide range of trees or the mountain pine beetle (*Dendroctonus ponderosae*; Keeling *et al.,* 2013), which is responsible for the destruction of large areas of lodgepole pine forests in Canada. Other examples are the coffee borer beetle (*Hypothenemus hampei*; Vega *et al.,* 2015), which feeds on the coffee berries thus causing loss of the economic value of the coffee grains, the small hive beetle (*Aethina tumida*), which causes damage to bee hives sometimes even forcing the bees to abandon their hives, and the boll weevil (*Anthonomus grandis*) which has caused devastating losses to the cotton producers all around the world by feeding on the cotton bugs and flowers.

I.4.3 Endosymbionts in beetles

Feeding exclusively on plants is no easy feat for animals. Besides producing defensive phytochemicals that repel or poison herbivores and hostile structures such as thorns and waxy tough cuticles, plant tissues are normally high in carbohydrates but low in proteins and essential vitamins. Plants are a highly recalcitrant food source and herbivorous animals require a specialized set of tools to overcome the plants defences and assimilate the intracellular plant metabolites. An important group of these enzymes are plant cell wall degrading enzymes (PCWDE), which help herbivores to digest the highly complex cell wall components (Wybouw *et al.,* 2016). PCWDE capable of targeting the polysaccharide network with cellulase, hemicellulase and pectinase activities have been acquired through horizontal gene transfer (HGT) by arthropods from microbes many times through the evolution (Kirsch *et al.,* 2014).

While beetles possess several PCWDE, the fact that they can thrive on a wide variety of diets so efficiently suggests that they might be receiving some assistance from symbionts including ecto- and/or endosymbionts and/or a more or less rich microbiota. In fact, there are well studied cases such as the symbiosis between the mountain pine beetle (*D. ponderosae*) where the beetle carries spores from the pathogenic blue stain ascomycete *Grosmannia clavigera* within a specialized on its mouthparts which are used to infect the host trees (Keeling *et al.,* 2013). As the fungus develops, it provides additional nutrients for the beetle, helps overcoming the trees defences and blocks the water and nutrient transport in the tree thus killing it in the end (Bracewell and Six, 2015). However, this might be a love-hate relationship, given that the symbiotic fungus may sometimes hamper the development of the larvae (Wang *et al.,* 2012). Notably, given the relevance of this pest, the genomes of the fungus and the beetle have been sequenced (DiGuistini *et al.,* 2009; Keeling *et al.,* 2013).

Besides studies on ectosymbionts, the endosymbionts in the Curculionoidea superfamily have been a subject of intense study and in fact, it is a great model for studying endosymbiont establishment, coevolution and replacement. Several studies propose that there was a single event around 125 MYA where *Candidatus* Nardonella (hereafter known as *Nardonella*) infected the ancestor of the Curculionidae and Brentidae families before their divergence and started coevolving along with its host (Lefevre *et al.,* 2004; Conord *et al.,* 2008; Toju *et al.,* 2013; Zhang *et al.,* 2017). To date *Nardonella* has only been found in association with these two lineages where it has been identified and molecularly characterized in several species (Kuriwada *et al.,* 2010; Huang *et al.,* 2016; Anbutsu *et al.,* 2017). However, it is far from being universally found in weevils and several cases of endosymbiont elimination or replacement have been identified. The best studied case is that of species of the genus *Sitophilus*, where an Enterobacterium of the genus *Sodalis* recently replaced the ancestral *Nardonella*. (Rio *et al.,* 2003; Lefèvre *et al.,* 2004; Clayton *et al.,* 2012). In the case of species of the *Curculio* genus, it was hypothesized that the endosymbiont replacement may have been due to a switch on their diet (Lefèvre *et al.,* 2004; Toju *et al.,* 2010; Toju *et al.,* 2013). The higher content of carbohydrates in cereal grains, and the lack of vitamins and AA may have enabled the competition between *Sodalis* with a full genome and *Nardonella* with reduced and exhausted genome.

Other examples of endosymbionts include *Candidatus* Dasytiphilus stammeri from pollen eating beetles of the genus *Dasytes* (Coleoptera, Dasytidae) and *Candidatus* Stammera capleta (hereafter known as Stammera) from the tortoise leaf beetle *Cassida rubiginosa*. The presence of intracellular endosymbionts in the Dasystes is perhaps most surprising given the fact that the larvae are predaceous and thus not dependent on a nutritionally unbalanced diet. Their presence might be better explained by the fact that the adults feed on pollen grains, suggesting that they might require enzymes from the endosymbionts to digest the walls from the pollen grains or amino acids and vitamins that might be present in low amounts in pollen grains (Weiss and Kaltenpoth, 2016). On the other hand, the extracellular endosymbiont *Stammera* remarkably has the smallest known genome of an extracellular bacterium (Salem *et al.,* 2017). Sequencing of its genome showed that it has been drastically streamlined to break down pectin having lost the capacity to produce all essential amino acids and B-vitamins.

As we have seen, beetles are the most diverse group of animals. This success in part can be attributed to their strong cuticle and their ability to establish symbiotic relationships with microbes that allow them to thrive on recalcitrant, toxic of nutritionally poor diets. A particularly efficient group at establishing these relationships is the Curculionoidea superfamily, where endosymbionts have been found in most of the studied species. A notable member of this group is *S. oryzae* which, as aforementioned, is a model for studying host-symbiont interactions besides being a serious pest of grains. In Chapter 1 we will see more unique and remarkable characteristics of this model.

I.5 The Aphids (Hemiptera: Aphidoidea)

Aphids (Hemiptera, Aphididae) belong to a group of insects with over 5,000 extant species (Grimaldi and Engel, 2005). Being hemipterans, aphids do not go through a complete metamorphosis but rather they only have three stages: egg, nymph and adult. Nonetheless, their life cycles are far from simple and involve a series of morphologically distinct morphs. Although aphids have life cycles with characteristic features, most of them present a holocyclic life cycle (see Figure I.9).

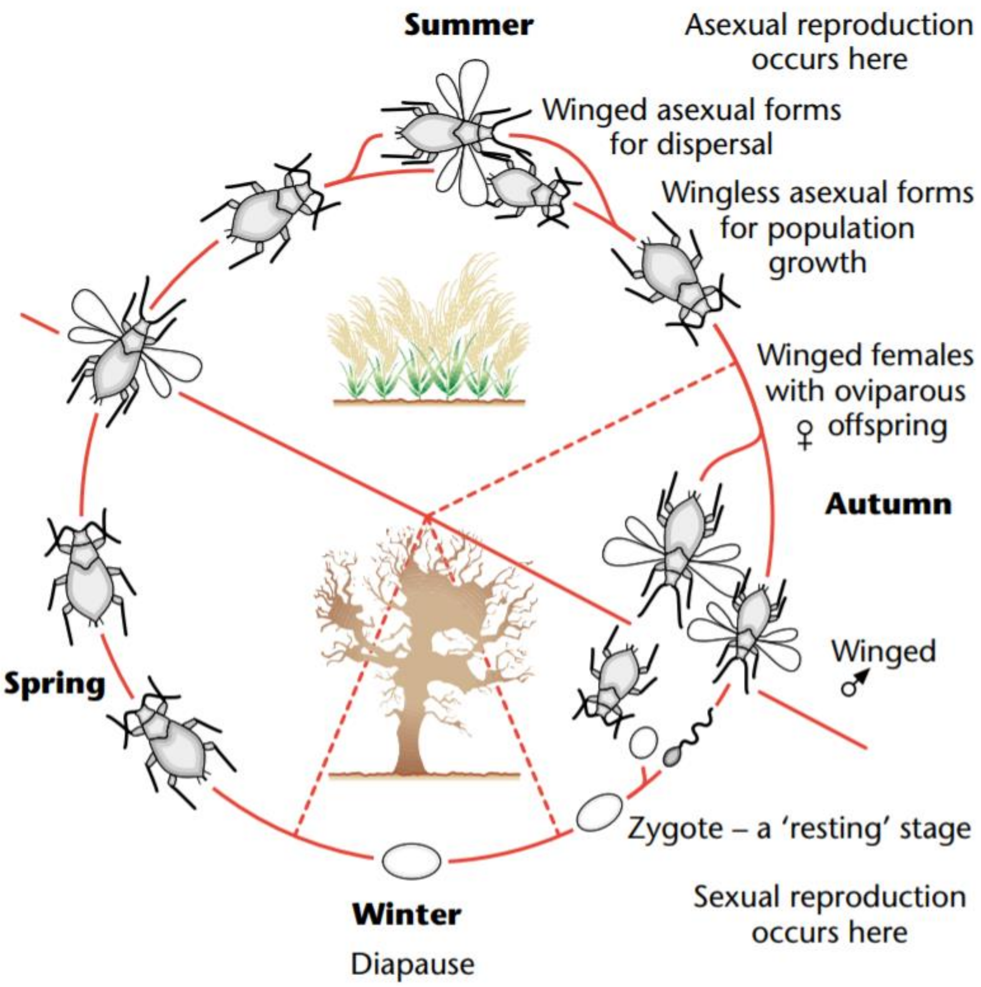

Figure I.9 Life cycle of an aphid with primary and secondary host plants. During the spring and summer asexual parthenogenetic females are produced in the secondary host plants. Afterwards, in the autumn, a generation of sexual morphs is produced, and adults might migrate to their primary hosts where they mate and lay eggs capable of withstanding the harsh conditions of the winter. The rising temperatures of the springs trigger their hatching and the cycle begins again. From Olive, Peter JW. 2002. "Reproduction and Life Cycles in Invertebrates." In Encyclopedia of Life Sciences. Chichester: John Wiley & Sons, Ltd.

This yearly life cycle involves several generations in which only parthenogenetic females are produced and one generation of sexual morphs, which lays eggs (Blackman and Eastop, 1994; see Figure I.9). The eggs resulting from the sexual reproduction are the means to survive the harsh conditions of the winter and they will hatch in spring, giving rise to viviparous females that will reproduce by parthenogenesis during spring and summer. Later, the dropping temperatures of the autumn trigger the appearance of the sexual morphs, which will lay fertilized eggs, thus starting the cycle over again. This feature is an important fact to explain the huge success of aphids and it was probably developed during the Triassic.

Additionally, there are more complex life cycles that also involve host alternation, where first the sexual morphs mate and lay fertilised eggs on a tree or shrub, their primary host. Afterwards, the aphids originating from the hatched eggs migrate to an unrelated herbaceous or woody plant also known as the secondary host where the parthenogenetic generations occur. Then, before the next sexual generations the aphids must move back to their primary host. Finally, some aphid species or populations are anholocyclic, meaning that they are parthenogenetic throughout their whole life cycle. This is the case for some populations of *Rhopalosiphon padi* that live in warm areas (Simon *et al.,* 1996).

I.5.1 Evolutionary history of aphids

The earliest aphid fossil so far dates to around 225 MYA, in the mid Triassic period (Hong *et al.,* 2009); however, the common ancestor of aphids is hypothesized to have arisen around 280 MYA (Heie, 1996), in the early Permian. It is thought that while several lineages from the Aphidoidea superfamily went extinct during the Triassic and Jurassic, the surviving lineages underwent a rapid radiation during the Cretaceous period possibly related to the expansion of angiosperms during this period (von Dohlen and Moran, 2000).

In order to catalogue the great diversity of aphids, members of this family are organised into subfamilies, tribes, and subtribes. Their taxonomic organization is an active area of study with multiple works attempting to correctly organize all this diversity (Martínez-Torres *et al.,* 2001; Ortiz-Rivas and Martínez-Torres, 2010; Nováková *et al.,* 2013). However, it has not yet been defined conclusively. In addition, given that the infection with *Buchnera* ancestor took place a single time in an aphid’s ancestor, both aphids and their endosymbiont have evolved strictly in parallel (Moran *et al.,* 1993). Therefore, molecular data from their endosymbionts has also been used to improve the information regarding the phylogenetic relationships between aphids’ lineages.

The Lachninae subfamily is presumably the sister taxa of all other aphid subfamilies and thus the most basal (Ortiz-Rivas and Martínez-Torres, 2010). Currently it groups 402 extant species (Favret, 2018) and almost all of them feed on tree trunks and branches and most of them are tended by ants (Heie, 2015). Species within this subfamily are classified into five tribes: the Eulachnini, the Lachnini, the Stomaphidini, the Tramini and the Tuberolachnini. The Eulachnini, which feed exclusively on conifers, are classified into four genera, *Cinara, Essigella*, *Eulachnus*, and *Pseudessigella*.

The largest within the Lachninae is *Cinara* (comprising 252 species) also known as giant conifer aphids. This genus has traditionally been taxonomically classified into two main subgenera: *Cinara (Cinara)* and *Cinara (Cupressobium)*. A third subgenus *Cinara (Schizolachnus)* was recently transferred to this genus (Chen *et al.,* 2016) and the fourth, *Cinara (Cedrobium)*, has not yet been molecularly analysed and thus, the phylogenetic placement of this subgenus remains uncertain. Extensive molecular work in *Cinara* divides the genus into three major phylogenetic clades, termed simply A, B, C (Meseguer *et al.,* 2015). While clade A includes only members of the *Cinara (Cinara)* subgenus, clade B groups some *Cinara (Cinara)* species and all *Cinara (Schizolachnus)* and clade C includes some members of the *Cinara (Cinara)* subgenus and all the *Cinara (Cupressobium)* species.

I.5.2 Aphids as pests

Being sap-feeding insects, aphids have inevitably started taking advantage of many crops, thus gaining their pest status. While feeding on the plant sap inevitably reduces crop yields and damages crops, the real impact is by being vectors of plant viruses (Guerrieri and Digilio, 2008). Additionally, not only host plants are infected given that winged adults select their hosts by using visual cues, followed by olfaction using the antennae. If everything seemed right, they proceed with the probing of the plant by inserting their stylus to test the sap. Non-host plants will be rejected; however, the transfer of viruses occurs at the beginning of the process thus allowing non-host plants to become infected (Will *et al.,* 2007). Additionally, the feeding activity of aphids produces large amounts of honeydew which is quickly colonized by sooty mold which by itself is harmless for the plant, however its growth can block sunlight from reaching the leaves, thus hampering photosynthesis (Dhami *et al.,* 2013).

The dispersal winged morphs are not particularly capable fliers and they lose their wings after a few days. However, given their small size it is easy for them to be dispersed by air currents or precipitation and additionally, movement of plant materials by humans also increases their dispersion (Fereres *et al.,* 2017). Additionally, given their successful reproductive strategies, once a plant has been colonized by a single individual it can quickly create a large population of clonal individuals.

While around 450 aphid species are associated with crops only around 100 are economically relevant. These aphids belong mainly to the Aphidinae subfamily and includes the pea aphid *Acyrthosiphon pisum*, which is the best studied aphid capable of feeding on several Fabaceae; the cowpea aphid *Aphis craccivora* that is even more polyphagous than *A. pisum* and a vector of up to 30 viruses; the green peach aphid *Myzus persicae*, which is extremely polyphagous and cosmopolitan found wherever peaches are available; and the Russian wheat aphid *Diuraphis noxia*, which feeds on grasses including wheat and is currently expanding its range among others (Blackman and Eastop, 2017). Members of the Lachninae subfamily, on the other hand, are not important agricultural pest.

Chemical strategies have been used to attempt to control the proliferation of aphids, however these strategies not always prove to be successful. One example are green peach aphids, which have shown an extraordinary ability to evolve resistance to a wide range of insecticides (80 different chemicals, though might be an underestimation; Whalon *et al.,* 2018). Resistances to organophosphates, pyrethroids and pirimicarbare have been described in *M. persicae* (Fenton *et al.,* 2010). Resistance to neonicotinoids, such as thiamethoxam, has been documented in the cotton aphid, *Aphis gossypii*, with Gore and colleagues highlighting the high reproductive rate of aphids as a source of resistance and thus recommending rotation strategies (Gore *et al.,* 2013).

In addition to chemical control strategies, several agents of biological control are used including parasitoids such as *Aphidius ervi*, fungi such as *Beauveria bassiana*, nematodes (to specifically target root dwelling aphids) and other arthropods such as lacewings (*Chrysoperla rufilabris*) and ladybirds (*Coccinella septempunctata*) that prey on aphids (Kundoo and Kahn, 2017). However, as mentioned before, aphids excrete honeydew which is a rich food source that's exploited by several hymenopterans including bees, which can use it to produce honey, and several species of ants which defend aphids from their predators to maintain their food source (Styrsky and Eubanks, 2007). Furthermore, as an additional defensive strategy, aphids can harbour beneficial facultative bacteria that protect them from parasitoids. For example, the facultative *Serratia symbiotica* has shown to confer resistance to the attack of parasitoids in *A. pisum* (see below).

I.5.3 Endosymbionts in aphids

As aforementioned, phloem, while rich in carbohydrates and some amino acids, is deficient in EAAs and B-vitamins. Early studies on *M. persicae* (Aphidinae: Macrosiphini) reared on an artificial diet lacking amino acids or B-vitamins demonstrated that these aphids have a need for these essential compounds, which are indeed absent from their diet (Mittler and Dadd, 1963; Dadd *et al.,* 1967). Given that animals are unable to synthesize EAAs these results suggested that the endosymbiotic bacteria found in the bacteriomes of aphids and first identified by Paul Buchner could be providing them with the missing nutrients (Buchner, 1965).

Several years later, with more advanced molecular techniques, it was demonstrated that their primary endosymbiont, which is hosted in their bacteriocytes, was a Gammaproteobacterium (Unterman *et al.,* 1989). Further studies using endosymbionts from different aphid subfamilies demonstrated that the sequences of their primary endosymbionts were arranged as a monophyletic clade, which mirrored the phylogenetic relationships of their hosts. With this considered, the genus *Buchnera* and the species *Buchnera aphidicola* were proposed to name the primary endosymbionts from the aphid *Schizaphis graminum* (Munson *et al.,* 1991). The name *B. aphidicola* was then used for all the primary endosymbionts of aphids, which are now considered strains. Thus, given that *B. aphidicola* is the only species from now on we will refer to it as *Buchnera*.

The next milestone was reached with the sequencing of the first *Buchnera* genome (Shigenobu *et al.,* 2000) revealing that it had a genome significantly smaller than free living bacteria, including its free-living relative *Escherichia coli*. Additionally, it was observed that *Buchnera* was devoting its genome mainly to the production of EAAs and that it lacked genes for both the production of non-essential amino acids (NEAAs) and cell surface components, thus emphasizing on its role as a primary endosymbiont and its inability to survive in niches other than inside its host's bacteriocytes. A couple of years later other *Buchnera* genomes were sequenced and it was shown that they displayed a high level of synteny among them, thus suggesting a rapid genome reduction before the diversification of aphids (Tamas *et al.,* 2002; van Ham *et al.,* 2003), and in 2006 the smallest *Buchnera* genome was published. Remarkably, it had lost the ability to synthesize some essential nutrients for the host, mainly the synthesis of the EAA tryptophan and the vitamin riboflavin (Pérez-Brocal *et al.,* 2006). Some years later, the genome of the pea aphid was published (The International Aphid Genomics Consortium, 2010), revealing that indeed the host could provide *Buchnera* with NEAAs and not only that, but it was also demonstrated that some metabolic pathways were shared between the two members of the consortium (Shigenobu and Wilson, 2011).

Regardless of the benefits that endosymbionts provide, keeping bacteria as permanent residents in their bodies would pose a problem for aphids given that their immune system would need to be kept under control to tolerate the endosymbiont. This is not a problem for *A. pisum* given that it has lost almost the entire IMD pathway (Gerardo *et al.,* 2010), which is involved in the recognition and eradication of Gram-negative bacteria. Therefore, *Buchnera* is presumably allowed to thrive "unrecognized" raising the question of how it is perceived and regulated by the host and more importantly how does it defend itself against pathogens. While this is an efficient strategy to allow conserving its endosymbiont, it also renders the aphid highly vulnerable to infections by

Gram negative bacteria (Altincicek *et al.,* 2011). Nonetheless, it is proposed that given that aphids feed on a relatively sterile food source, perhaps they don't need such an efficient immune system. Additionally, the energy that would otherwise be used on this pathway is available to be redirected to other functions such as reproduction (Barribeau *et al.,* 2010).

In addition to harbouring *Buchnera*, many aphids contain secondary endosymbionts, which are not essential for their host's nutrition but might provide some benefits under particular environmental circumstances (conditional mutualism). Given their non-essential status, secondary symbionts normally do not have nutritional roles and normally have one or more functions that provide their host with an added benefit. Some examples of these beneficial yet non-essential functions are: protection against parasitoid wasps by *S. symbiotica*, *Hamiltonella defensa*, and *Regiella insecticola* (Oliver *et al.,* 2003; Hansen *et al.,* 2014), protection against fungal parasites (*R. insecticola*, *Rickettsia* sp., *Rickettsiella* sp., and *Spiroplasma* sp.) and an increased resistance to heat stress (*Rickettsia* sp. and *S. symbiotica*) (Chen *et al.,* 2000; Montllor *et al.,* 2002). Another remarkable difference between primary and secondary endosymbionts is that while secondary endosymbionts are sometimes kept inside bacteriocytes, in their own bacteriocytes or sometimes even sharing bacteriocytes with *Buchnera* (Michalik *et al.,* 2014), they can also be found in cells surrounding the bacteriome or free in the haemolymph. On occasions, as stated before, the primary endosymbiont has lost essential functions that are supplemented by the facultative bacteria. If this happens, a symbiotic consortium is established as it has occurred in the Lachninae subfamily (Lamelas *et al.,* 2011; Manzano-Marin *et al.,* 2017).

Given their efficient and complex lifecycle, aphids are capable of rapidly infesting their food source. While some of them are highly polyphagous other target a single genus or even species. Aphids have been widely studied to better understand the host-symbiont relationship given the prevalence of their primary endosymbiont *Buchnera*. Additionally, as aforementioned, some aphids have lost parts of the signalling pathways of the immune system, thus making them attractive models to understand the relationship between the endosymbiont and the immune system. Furthermore, members of the Lachninae subfamily seem prone to acquire secondary endosymbionts. Therefore, they are good models for studying endosymbiont complementation and the process by which a secondary endosymbiont drifts into to a coprimary (Manzano-Marin *et al.,* 2017). Among them, perhaps the best studied are aphids from the *Cinara* genus, especially the cedar aphid *Cinara cedri*, whose genome has been sequenced as part of this thesis and is presented in the second Chapter.

II OBJECTIVES

This thesis manuscript focuses on the comparison of the innate immune system and amino acid biosynthetic pathways of two insects that harbour endosymbionts. Both sets of pathways were chosen given that they are the main mechanisms through which both members of the holobiont interact with each other. The selected organisms are the rice weevil *S. oryzae* and the cedar aphid *C. cedri*. These organisms were chosen because we plan to identify differences on genetic repertoire of insects with endosymbiotic relationships of different ages. While the relationship between *C. cedri* and *Buchnera* is ancient (established at least 150 Ma) that of *S. oryzae* and *S. pierantonius* is quite young (established around 30,000 years ago). Another relevant aspect for choosing both systems is that they have been profusely studied by the groups of Abdelaziz Heddi and Amparo Latorre respectively and that the genomic sequences for their endosymbionts had previously been obtained. In the case of *C. cedri* the sequence of *Serratia*, the coprimary endosymbiont is also available. While comparing both systems would allow us to identify differences between an ancient and a recently established endosymbiotic relationship it is important to keep in mind that there are many additional differences between both models, including their diet, life cycle and taxonomic position.

The first objective is to obtain the fully assembled genome of *S. oryzae* and annotate it to allow the identification of genes involved in the immune system pathways and in the amino acids metabolism. The gene sets will then be compared with that of *T. castaneum* and *D. melanogaster*, the models of beetles and insects respectively. This would allow us to observe if there are relevant differences between our system and the model of its order, a beetle which does not harbour a primary endosymbiont. Our hypothesis is that given that both *S. oryzae* and *T. castaneum* belong to the same order and share their ecological niche there should not be large differences in the gene sets we will assess. While *T. castaneum* does not harbour an endosymbiont, the fact that the relationship between *S. oryzae* and *S. pierantonius* was recently established should not yet be reflected in the genome. Additionally, comparing with *D. melanogaster* will allow us to identify features that could be specific to the Coleoptera order.

The second objective is to obtain the genes involved in the innate immune system and amino acid biosynthetic pathways of *C. cedri* through the assembly and annotation of its genome. Upon obtention, this set of genes will be compared with that of *A. pisum* and *D. melanogaster*, the models of aphids and insects respectively. In this case, *A. pisum* also harbours an endosymbiont which in fact shares a common ancestor with that of *C. cedri*, we therefore expect that both insects will have similar gene sets. Nonetheless, this comparison would also allow us to identify features particular to *C. cedri*. Through the comparison with *D. melanogaster* we will be able to identify differences between the genome of the model of the insects and that of an insect with an ancestral endosymbiotic relationship.

The third objective is to compare the genes involved in the innate immune system and the amino acid biosynthetic pathways of *S. oryzae* and *C. cedri*. This would allow us to accomplish the main goal of this project; however, as aforementioned, there are additional differences between the two models besides the age of the establishment of their symbiotic relationships. To try to exclude other factors that could be causing differences we observe, other insects will be incorporated to this comparison to set it into context. It has previously been suggested that endosymbiosis is a source of evolutionary innovation and its effect on the genomes of the endosymbiont has been greatly studied, however not much is known yet regarding the genomes of the hosts, thus this will be a great opportunity to assess its effects in the genomes of the hosts.

III MATERIALS AND METHODS

III.1 Insects samples and DNA extraction

III.1.1 *S. oryzae* rearing and DNA extraction

Individuals of both sexes of *S. oryzae* were reared on wheat grains at 27.5°C with 70% relative humidity. The aposymbiotic strain was obtained by treating the symbiotic strain during one month at 35°C and 90% relative humidity as previously described (Nardon, 1973). This strain is viable, fertile and was raised in the same conditions as the symbiotic strain. The aposymbiotic status is regularly checked by PCR and histology.

The genome size of five male and five female adults of *S. oryzae* were estimated through flow cytometry using the protocol described in Lopes *et al.,* (2009). Male and female aposymbiotic adults were used for DNA extraction and only the gonads were used to minimize DNA contamination from its diet, which could still be present in the gut. A DNA extraction protocol specific for *S. oryzae* using a STE buffer (100 mM NaCl, 1 mM Na_2EDTA pH 8, 10 mM Tris HCl pH 8) was performed. Tissues were homogenized in STE buffer, then treated successively with SDS 10%, proteinase K and RNase.

Genomic DNA was purified by two successive extractions with phenol:chloroform:isoamyl alcohol (25/24/1) followed by an extraction with 1 vol of chloroform:isoamyl alcohol (24/1). Genomic DNA was then precipitated with 0.7 vol isopropanol. After washing the pellet with 70% ethanol, genomic DNA was recovered in TE (1mM EDTA, 10 mM Tris HCl pH8) buffer.

Four different DNA samples were obtained: three from males and one from females with each sample made up of the genomic DNA from 20 individuals. The DNA concentration in each of these samples was quantified using a NanoDrop spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA).

III.1.2 *C. cedri* collection and DNA extraction

C. cedri specimens were collected from a permanent population reared on two cedar trees of the species *Cedrus atlantica* (Glauca group). Total insect DNA (tDNA) was extracted immediately after sampling using the method of Latorre *et al.,* (1986). 6.5 mg of insects were gently homogenized in 160 µL buffer I (10 mM Tris-HCl [pH 7.8]; 60 mM NaCl; 5% sucrose; 10 mM EDTA) at 4°C. After the addition of 200 µL of buffer II (300 mM Tris-HCl [pH 8.0]; 1.25% SDS; 5% sucrose; 10 mM EDTA), the sample was incubated at 65°C for 30 min, neutralized with 60 µL 3 M potassium acetate (pH 5.0), and kept at –20°C for 20 min. The tDNA was concentrated by precipitation using standard protocols, resuspended in ultrapure water, and stored at –20°C. The concentration and quality of the tDNA were measured using a PicoGreen dsDNA Quantification Assay (Invitrogen [Thermo Fisher Scientific], Waltham, MA, USA).

III.2 Genome assembly and annotation

III.2.1 *S. oryzae* genome sequencing and assembly

We used two sequencing platforms: Illumina and PacBio. For each sex, two Illumina libraries, with read size of 101 bp, were generated: one paired-end with an average fragment size of 500 bp and one mate pair with an average fragment size of 5 kbp. The libraries were sequenced using an Illumina HiSeq 2000 platform; the paired-end (PE) libraries were sequenced at ProfileXpert (Lyon, France) while the mate paired (MP) were sequenced at Macrogen (Seoul, South Korea). Two additional male samples were used to build an Illumina HiSeq 2500 library, with read size of 125 bp and average fragment size of 200 bp, and a PacBio library. Seven SMRT cells were sequenced using P6-C4 chemistry generating reads with an average size of 5,900 bp. These last two libraries were sequenced at KeyGene (Wageningen, The Netherlands).

The overlapping PE Illumina reads were error-corrected using SGA (Simpson *et al.,* 2012) and the PacBio data were filtered using PROOVREAD (Hackl *et al.,* 2014) with the overlapping PE Illumina reads. The mitochondrion was assembled using MITOBIM (Hahn *et al.,* 2013) followed by a hybrid strategy using PLATANUS (Kajitani *et al.,* 2014), SSPACE (Boetzer *et al.,* 2011) and GAPFILLER (Boetzer and Pirovano, 2012) with the Illumina non-overlapping PE and MP data and PBJELLY (English *et al.,* 2012) with the PacBio data. Next, reads mapping to the mitochondrion from the overlapping PE Illumina set were removed and the remaining reads were assembled using PLATANUS with an initial scaffolding and gap filling followed by three rounds of re-scaffolding with REDUNDANS (Pryszcz and Gabaldón, 2016) using SSPACE and GAPFILLER. These scaffolding rounds were also complemented using PBJELLY with the cleaned PacBio data. Next, REAPR (Hunt *et al.,* 2013) was used to assess the assembly. Regions not supported by the Illumina PE and MP data were removed from the assembly and probable misassembled contigs were broken. Contigs smaller than 1 kb were eliminated, and the assembly was rescaffolded and gap filled another round using the aforementioned tools. We ran the assembly through NCBI’s Contamination Screen and discarded four complete scaffolds, and regions matching adapters. If the matching region was in the middle of a scaffold, the scaffold was split into two new scaffolds. The completeness of the genome assembly was assessed by searching for the presence of conserved genes using CEGMA and BUSCO.

III.2.2 *S. oryzae* genome annotation

Repeated elements were identified de novo using REPEATMODELER (Smit and Hubley, 2008) and masked with REPEATMASKER (Smit *et al.,* 2013) using the custom library generated by REPEATMODELER. After the genome had been masked, it was annotated with the MAKER (Cantarel *et al.,* 2008) pipeline. First, GENEMARK (Besemer *et al.,* 2005) was self-trained and then, using the RNA-seq information from 12 libraries and a set of proteins from coleopterans (78,686 proteins), a first round of MAKER was performed. Two rounds of training of SNAP (Johnson *et al.,* 2008) were performed using MAKER’s output to define a high-quality training set. Afterwards, we selected several gene models from the output of CEGMA and manually curated them to train AUGUSTUS (Stanke *et al.,* 2005). To obtain the preliminary set of annotated genes (OGS v1.0), MAKER was run using GENEMARK, AUGUSTUS, SNAP and TRNASCAN (Lowe and Eddy, 1997). The prefix SORY followed by an underscore and an increasing eight-digit number was used to name every gene model. The full protein set was compared against the UniProt/Swiss-Prot database using BLASTP (Altschul *et al.,* 1990) to define a putative gene function. Additionally, INTERPROSCAN (Jones *et al.,* 2014) was used to add protein domain information.

The *S. oryzae* genome was curated to improve the OGS v1.0 in terms of both gene models and functional annotations of genes using WebApollo (Lee *et al.,* 2013). Several groups of experts participated on the manual curation of genes or gene families of interest including development, metabolism, immune pathways, olfactory receptors, epigenetics machinery and horizontally transferred genes. After the manual curation, the new official gene set OGS v1.1 was generated.

III.2.3 *C. cedri* genome sequencing and assembly

The DNA was used to build a PE library with a fragment size of approximately 410 bp and two MP libraries with insert sizes of 3,000 and 5,000 bp. All libraries were sequenced using the Illumina HiSeq2000 in paired end mode, which outputs 101 bp reads (2x101bp). The sequence reads where trimmed and filtered using the Trim Galore! wrapper script (Krueger, 2015), CUTADAPT (Martin, 2011) and gem-mapper (Marco-Sola *et al.,* 2012). The genome size and complexity were estimated using JELLYFISH v1.1 (Marçais and Kingsford, 2011) and GENOMESCOPE v1.0 (Vurture *et al.,* 2017) respectively.

The genome was assembled using SGA preqc analysis (Simpson, 2014), ASM (Frias and Ribeca, 2018) and ABYSS v1.5.2, GEM-MAPPER (Marco-Sola *et al.,* 2012) and SSPACEv3.0 (Boetzer *et al.,* 2011). The gaps were closed with GAPFILLER (Boetzer and Pirovano, 2012) and single nucleotide substitutions and short insertion–deletion errors were corrected as described in Cruz *et al.,* (2016). Finally, gene completeness was evaluated with CEGMAv2.4 (Parra *et al.,* 2007), and BUSCO v3.0.2 (Simão *et al.,* 2015) with the insecta_odb9 that included 42 species and 1,658 genes.

III.2.4 *C. cedri* genome annotation

A combination of the Program to Assemble Spliced Alignments (PASA v2.0.2) and Evidence Modeler (EVM v1.1.1) (Haas *et al.,* 2008) was used to obtain consensus coding sequence (CDS) models using three main sources of evidence: aligned transcripts, aligned proteins and gene predictions. Finally, ncRNAs were annotated employing CMSEARCH (Cui *et al.,* 2016) and TRNASCAN-SE (Lowe and Eddy, 1997), and lncRNAs were obtained from the PASA assemblies without protein-coding gene annotation that were longer than 200bp.

III.2.5 *Wolbachia* Cced genome assembly and annotation

Given that the tDNA from *C. cedri* was obtained from whole insects, we took the opportunity to assemble and annotate the genome of the third endosymbiont, *Wolbachia*. The annotation was performed using prokka (Seeman T, 2014). Using AMPHORA2 (Wu and Scott, 2012) the 31 markers used in Ramirez-Puebla ST *et al.,* (2015) were identified and concatenated using the EMBOSS union web tool (*http://emboss.bioinformatics.nl/cgi-bin/emboss/union*). This sequence was added to the alignment from the aforementioned work, gaps were removed, and sequences realigned using MUSCLE (Edgar, 2004). The alignment was filtered using GBLOCKS (Castresana, 2000) to remove poorly aligned positions and the obtained blocks were used to perform a maximum-likelihood analysis using the JTT substitution model in PHYML 3.0 (Guindon *et al.,* 2010).

III.3 Orthology assignment

EGGNOG-MAPPER BUILD (Huerta-Cepas *et al.,* 2016a) was used to identify orthologs among the selected species (see Table S.1) using the diamond mode and the Arthropoda (artNOG) dataset. All 1,345 genes with one-to-one orthologues in every studied species were selected and their alignments concatenated to generate the species tree using ETE-BUILD (Huerta-Cepas *et al.,* 2016b) following the phylomedb4 gene tree workflow and the sptree_raxml_85 species tree workflow. The tree was made ultrametric using the ape package (Paradis *et al.,* 2004) from R (R Core Team, 2018). This ultrametric tree was used along with the gene families' data to estimate gene losses and gains using CAFE after having estimated the error rate and corrected for it. Using the gene loss and gain rate the ancestral state gene counts are inferred and a p-value is calculated to evaluate the relevance of the gene family changes along each branch.

The genomes of both *S. oryzae* and *C. cedri* were compared to other arthropod genomes to characterize both orthology and paralogy relationships. The catalogue of gene phylogenies, called the phylome, was reconstructed using the PHYLOMEDB v4 pipeline (Huerta-Cepas *et al.,* 2014). A search was performed against the proteome database of other arthropods using all proteins from *S. oryzae* or *C. cedri* as seeds. Multiple alignments of homologous sequences were built and then used for phylogenetic tree reconstruction as described in Huerta-Cepas *et al.,* (2014).

III.4 Identification of genes involved in the innate immune system pathways

A database of genes involved in insect immune system was generated by updating and expanding the database from Insect Innate Immunity Database (Brucker *et al.,* 2012). The source references for the jewel wasp (*Nasonia vitripennis*; Werren *et al.,* 2010), the honey-bee (*Apis mellifera*; Evans *et al.,* 2006), the fruit fly (*D. melanogaster*; De Gregorio *et al.,* 2001), the African malaria mosquito (*A. gambiae*; Christophides *et al.,* 2002) and the pea aphid (*A. pisum*; Gerardo *et al.,* 2010) were revised and the genes they described were retrieved. To increase completeness of the list the immune system genes from the red flour beetle (*T. castaneum*; Zou *et al.,* 2007), the diamondback moth (*Plutella xylostella*; Xia *et al.,* 2015), the tobacco hornworm (*Manduca sexta*; Cao *et al.,* 2015), the head louse (*P. humanus*; Kang *et al.,* 2015), the Florida carpenter ant (*Camponotus floridanus*; Gupta *et al.,* 2015), the Asian citrus psyllid (*D. citri*; Arp *et al.,* 2016) and the silkworm (*Bombyx mori*; Tanaka *et al.,* 2008) were also added.

III.5 Identification of genes involved in amino acid biosynthetic pathways

All the genes from all the arthropods in our dataset were annotated using BLASTKOALA (Kanehisa *et al.,* 2016) to ensure all genes were given the same treatment and mapped into the pathway for the biosynthesis of amino acids (map01230) using KEGG mapper. Gaps in the pathways were filled manually searching for individual missing enzymes.

III.6 Identification of putative horizontally transferred genes

DARKHORSE (Podell and Gaasterland, 2007) and HGT-FINDER (Nguyen *et al.,* 2015a) were used to identify putative horizontally transferred genes. The candidates were manually assessed, and likely retroviral transfers were discarded. The remaining candidates were manually curated using WEBAPOLLO (Lee *et al.,* 2013). The putative donor was identified evaluating the phylogeny of the hits with highest identity using the webserver version of BLAST+ (Camacho *et al.,* 2009).

III.7 Disclaimer

While sampling and DNA extraction for both models are included in this manuscript they were not performed by the author of the manuscript. Additionally, in the case of *C. cedri* the genome was assembled and annotated at the CNAG following the previously described strategy.

IV CHAPTER 1: THE RICE WEEVIL *SITOPHILUS ORYZAE*

IV.1 Introduction

The Dryophthoridae family comprises some of the most destructive pests including the oligophagous grain feeders of the genus *Sitophilus* (Coleoptera, Dryophthoridae) which cause stored cereal losses ranging from 25 to 40% on average but reaching up to 80% when favourable conditions are met (Champ and Dyte, 1976; Ladang *et al.,* 2008; Tefera *et al.,* 2013). Among the cereal weevils, the rice weevil *S. oryzae* (see Figure IV.1) is possibly the most important storage pest of every top cereal of agronomic and economic relevance (wheat, maize, rice, sorghum and barley), causing extensive quantitative and qualitative losses not only in stored grains, but also in grain products, such as pasta, throughout the world (Zunjare *et al.,* 2016; Stejskal *et al.,* 2004; Grenier *et al.,* 1997). Additionally, the infestation of grains by weevils also releases dust which attracts secondary pests such as mites and the red flour beetle *T. castaneum*, among others (Hardman, 1977), as well as mycotoxin poisoning of grains given that insects can act as carriers of fungal contaminants (Tefera *et al.,* 2011). While *S. oryzae* is primarily a pest of stored products, it can also attack cereals on the field.

As in other holometabolous insects, the life cycle of *S. oryzae* is divided in four stages: egg, larva, pupa and adult. The first three stages are completed fully inside the grain. Females lay up to six eggs per day and around 300 eggs over their entire lifespan. The female chews a small hole in the grain, deposits a single egg within the hole and finally seals it with secretions from her ovipositor. After the egg hatches, the larva starts developing within the seed, consuming it from the inside as it feeds. It pupates within the grain kernel eventually leaving the grain as an adult. The whole process takes about 30 days in average (Koehler, 1994). Feeding on plant tissues is no easy feat, nonetheless, some herbivore insects such as true weevils, including *S. oryzae*, are recipients of genes coding for PCWDE from both fungi and bacteria which enables them to digest pectin (Shen *et al.,* 2003; Kirsch *et al.,* 2014; Kirsch *et al.,* 2016).

Figure IV.1 The rice weevil S. oryzae over rice grains. Reproduced from Bugwood.org, photo by Joseph Berger.

Another important feature of cereal weevils is their permanent association with endosymbionts that supply them with nutrients that are not readily available in the grains, hence increasing their fitness and invasive power. *S. oryzae* currently maintains a mutualistic intracellular relationship with the intracellular gamma-proteobacterium *Candidatus* Sodalis pierantonius (hereafter known as *S. pierantonius*; Heddi *et al.,* 1998; Oakeson *et al.,* 2014). These endosymbionts are transmitted maternally and early during embryogenesis they induce the differentiation of the bacteriocytes in the first instar larvae (Heddi *et al.,* 2001).

The symbiosis between *Sitophilus* weevils and *S. pierantonius* was established recently (less than 30,000 years ago) replacing the ancestral *Nardonella* and thus being an interesting model to study this phenomenon (Lefèvre *et al.,* 2004; Clayton *et al.,* 2012). Such replacements are hypothesized to take place during a change of the host’s habitat triggering a competition between the ancestral primary endosymbiont which had an already reduced and exhausted genome and secondary endosymbionts that had not yet undergone drastic genome reduction (Lefèvre *et al.,* 2004; Conord *et al.,* 2008; Toju *et al.,* 2010; Anbutsu *et al.,* 2017).

Several works have demonstrated that *S. pierantonius* improves the host abilities including fertility, developmental time and flight capacity (Heddi *et al.,* 1999; Rio *et al.,* 2003; Grenier *et al.,* 2011). Additionally, the genome of *S. pierantonius* was recently sequenced (Oakeson *et al.,* 2014) and the putative inferred metabolism suggested that it was capable of providing the weevil with every amino acid except for methionine, tryptophan and histidine. Contrary to long-lasting insect endosymbionts, the genome of *S. pierantonius* is only partially reduced and highly pseudogenized, with a GC-content of 56.06%, similar to free-living relatives, and it contains a large number of mobile elements (18% of the genome’s size), which are normally absent from long-lasting endosymbiotic bacterial genomes (Gil *et al.,* 2008; Oakeson *et al.,* 2014; Latorre and Manzano-Marín, 2017). Comparative analyses of the genome of *S. pierantonius* with a free-living relative, namely *Sodalis praecaptivus*, highlighted numerous genomic rearrangements, duplications, and deletions enabled by a recent expansion of insertion sequence elements (Plague *et al.,* 2008). It was shown that *S. pierantonius* has lost the ability to synthesize several essential amino acids and vitamins and almost half of its protein-coding genes have been pseudogenized. *S. pierantonius* was also demonstrated to be highly involved in the cuticle synthesis of emerging adults through the supply of tyrosine and phenylalanine aromatic amino acids, and that the host is capable of finely controlling the endosymbionts population throughout its metamorphosis (Vigneron *et al.,* 2012; Vigneron *et al.,* 2014).

Altogether, these traits underlie the extraordinary success of *S. oryzae*, which is thus far considered as one of the greatest threats to postharvest agricultural products. Moreover, this insect pest is of increasing concern due to its ability to rapidly evolve resistance to insecticides such as phosphine, the fumigant used to protect stored grain from insect pests (Champ and Dyte, 1977; Nguyen *et al.,* 2015b; Mills, 2001).

IV.2 Results and discussion

IV.2.1 Genome assembly and annotation

We sequenced and assembled the genome of *S. oryzae* at a depth of 101X using a pool of 20 adults resulting in an assembly with a size of 652 Mb in 17,786 scaffolds (see Table IV.1). This assembly size is consistent with the genome size measured through flow cytometry (641 Mb in females and 635 Mb in males). However, a recently published work (Silva *et al.,* 2018) estimated the genome size to be of 769 Mb. These different measurements are currently unexplained, nonetheless it is possible that differences in the genome size between the strain we sequenced and the sample they obtained are due to the number of transposable elements (see below).

At 652 Mb, *S. oryzae* has the third largest genome among the Coleoptera with data available in NCBI, its size is comparable to the fourth largest (*L. decemlineata*). We assessed the completeness of the genome using two pipelines to annotate core genes in eukaryotic genomes: CEGMA (82.26% complete and 11.29% partial) and BUSCO (89% complete and 7% partial).

Statistic	***S. oryzae***	***D. ponderosae***	***A. glabripennis***	***T. castaneum***
No. sequences (>= 1 kb)	17,786	8,188	10,462	1,544
Largest scaffold	818 Kb	4,163 Kb	5,511 Kb	31,381 Kb
Total length	652 Mb	253 Mb	708 Mb	210 Mb
Scaffold N50	106 Kb	629 Kb	659 Kb	975 Kb
GC%	32.62	38.40	32.70	33.9
Gap length	81 Mb	51 Mb	105 Mb	59 Mb
Median coverage	101×	443×	121×	NA
BUSCO (% complete/partial)	89/96	83/94	87/97	95/99
Gene count (protein coding)	17,159	13,021	13,894	12,862
Transcript mean size (bp)	1,369.31	2,162.37	1,987.62	1,984.73

Table IV.1 Assembly statistics of S. oryzae and a comparison with the statistics of D. ponderosae, A. glabripennis and T. castaneum.

All the aforementioned statistics fall within the range of previously sequenced genomes from the order Coleoptera (Richards *et al.*, 2008; Keeling *et al.*, 2013; McKenna *et al.,* 2016). However, with a scaffold N50 of 106 kb the degree of fragmentation of the assembly is higher than in other insect genomes assemblies and is most likely due to the fact that a pool of individuals was used as the DNA source, thus increasing the heterozygosity of the assembly. More importantly, a surprisingly large number of repeated elements (or repeatome) was found on the genome, thus further complicating the assembly.

The genome was repeat-masked using REPEATMODELER before it was annotated and at this stage it became evident that the genome of *S. oryzae* contained an overabundance of transposable elements (TEs). This was further explored by Clément Goubert using DNAPIPETE (Goubert *et al.,* 2015) and he demonstrated that remarkably, half (52.1%) of the genome of *S. oryzae* was composed of repetitive elements, thus explaining the large genome size and the degree of fragmentation of the assembly. The amount of TEs in the genome placed *S. oryzae* among the insects with the largest repeatomes, such as the Asian tiger mosquito *Aedes albopictus* (50%, Goubert *et al.,* 2015) and the housefly *Musca domestica* (52%, Scott *et al.,* 2014).

The three species of *Sitophilus* feeding on grains (*S. oryzae*, *S. zeamais* and *S. granarius*) have a similar TE content (52.1%, 45.4% and 42.4% respectively) while the repeatome of *S. linearis* spans 27.8% of the genome. Interestingly, this species is the only member of the genus that feeds on a richer food source (tamarind) and that doesn’t harbour any endosymbiont. This perhaps could highlight the relevance of transposable elements in the adaptation to stress, such as an endosymbiotic replacement, in this genus.

Using a customized MAKER pipeline (Cantarel *et al.,* 2008) and incorporating RNA-seq information from 12 different libraries, we annotated 17,026 gene models producing an official gene set (OGS v1.0). A total of 1,675 genes were manually curated in terms of structural and functional annotation by groups of experts focusing on metabolism, immunity, development, epigenetics, olfaction, and horizontally transferred genes. The automated annotations and manual curations were merged into a new official gene set (OGS v1.1) with 17,159 gene models.

The manual annotation certainly represented an improvement from the automated annotation. However, the fact that the genome is quite fragmented has complicated the process of annotation. This fragmentation seems to be reflected in the smaller transcript size of *S. oryzae* given that, while its mean transcript size is around 1.3 kb that of the other three used coleoptera is around 2 kb. Although gene structure studies would be complicated using this assembly, we can certainly use it to identify genes of interest. Also, given the fact that the completeness statistics are similar to those of the other genomes, we can perform reliable presence–absence studies.

IV.2.2 Gene orthology and gene family evolution

We used EGGNOG-MAPPER to assess the conservation degree of protein-coding genes between *S. oryzae* and 21 other arthropod species to determine orthology relationships. These species were selected to represent the four main orders of the holometabolous insects and a selection of hemipterans. All the chosen insects have a fully assembled and annotated genome and are well studied. Additionally, they have different diets and environments and several of them harbour endosymbionts (see Table S.1). To minimize differences in the number of isoforms due to the accuracy of the genome annotations only one isoform (the longest) per gene was used for the analysis.

Over 85% of *S. oryzae* genes have orthologues in at least one arthropod species. Of these, 1,345 were universal single copy orthologs (indicated as one2one in Figure IV.2) across at least 19 species, which were used to determine the maximum likelihood phylogeny. The phylogeny accurately reflects the known phylogenetic relationships between the species. Interestingly, *S. oryzae* has the most lineage specific genes and orthologs in the many2many category among Coleoptera (see Figure IV.2). This could be explained by the high expansion rates of many families of genes in this species.

The rapidly evolving families in Coleoptera were identified using the phylogeny and the orthology relationships among the gene sets of each species. With 174 rapidly expanding families, *S. oryzae* was the beetle with the fastest expansion rate (0.409 genes per million years) and highest number of genes gained. It seems likely that the increase in gene family sizes in *S. oryzae* is due to the high levels of TEs not unlike what was observed in termites where the expanded gene families had an increased amount of TEs in their flanking regions (Harrison *et al.,* 2018). While proteins involved with transposition and mobile elements compose a large proportion of the rapidly evolving gene families (47%), there are other families of interest such as carboxylases and ABC transporters which might be associated with higher insecticide resistance (Lü *et al.,* 2015) as well as cathepsins which are known to have a digestive role and defend against plant proteases on herbivorous insects (Bansal *et al.,* 2018). This will be further discussed.

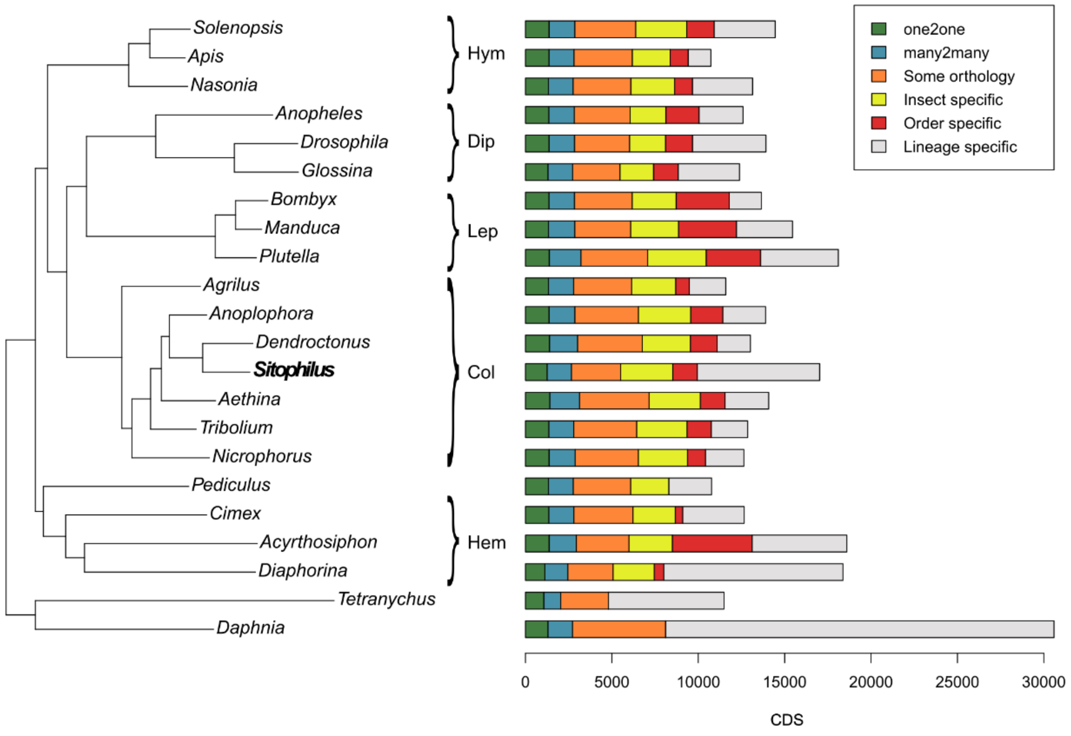

Figure IV.2 Phylogenetic relationships of the orthologs of S. oryzae and 21 other insect genomes. The one2one category refers to orthologs with only one copy in at least 19 of the species used. The many2many category includes orthologs also found in at least 19 of the species used but with more than one copy in at least one of the species. Some orthology refers to orthologs found in less than 19 of the species used. Insect, family and lineage specific refers to orthologs found in all the species belonging to that group. To group species included in each order curly brackets were used: Hym stands for Hymenoptera, Dip for Diptera, Lep for Lepidoptera, Col for Coleoptera and Hem to Hemiptera.

IV.2.3 Identification of genes involved in amino acid biosynthesis

As aforementioned, insects, and for that matter *S. oryzae*, are incapable of synthesizing ten amino acids: the nine mammalian EAAs (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine) plus arginine given that they lack the urea cycle (there are some exceptions, such as cockroaches). Most insects can obtain the necessary amino acids from their diets, and in that sense holometabolous insects have an additional advantage. This advantage is that they can exploit different niches, and thus food sources, during their larval and adult stages. However, most monophagous insects rely on a single food source during both their larval and their adult stages meaning that they must obtain all their amino acids from a single food source. *S. oryzae* feeds its entire life cycle wholly on cereal grains, and while they are a rich food source, the abundance of some amino acids is too low to sustain animals adequately. Additionally, the amino acid abundance on the grains depends on many factors, including the genotype, the weather and the soil the plants are growing on (Garcia del Moral, *et al.,* 2007), therefore an animal relying entirely on a single food source must be capable of coping with these aspects.

To evaluate the ability of *S. oryzae* to synthesize amino acids, we identified all genes involved in their biosynthesis. Additionally, the capabilities of *S. pierantonius* to synthesize amino acids were also considered (Oakeson *et al.,* 2014). Through a detailed reassessment of these pathways we found that besides being incapable of producing methionine, tryptophan and histidine, *S. pierantonius* was also incapable of producing leucine, valine, isoleucine, alanine and proline. This highlighted the important role of the host (and/or the diet) in complementing the supply of EAAs.

Regarding amino acids requiring α-ketoglutarate, we identified in the genome of *S. oryzae* the pathways involved in the biosynthesis of glutamate (the precursor of many other amino acids), glutamine and proline, thus the only missing amino acid from the family was arginine. Nonetheless, this gap could be filled by *S. pierantonius*, which can catalyse all the steps from glutamine to arginine (see Figure IV.3).

Another group of amino acids uses oxaloacetate as its precursor and while in *S. oryzae* we only identified the pathways to synthesize aspartate and asparagine, *S. pierantonius* can produce lysine and threonine. Neither the host nor the endosymbiont can produce isoleucine, *S. oryzae* lacks three out of the five steps needed to produce isoleucine from threonine and *S. pierantonius* has its dihydroxy-acid dehydratase (ilvD) pseudogenized, thus interrupting the pathway (see Figure IV.3) and rendering the holobiont dependent on its food source for obtaining isoleucine. Finally, while the holobiont might obtain methionine entirely from its food source it has also been suggested that insects might synthesize methionine using cystathionine as a precursor (Russell *et al.,* 2013). This would require CBS, a cystathionine-β-synthase to perform as a cystathionine-β-lyase to produce homocysteine and the last step would be catalysed by mmuM, a homocysteine methyltransferase.

The next group of amino acids uses 3-phosphoglycerate as its carbon backbone. Both *S. oryzae* and *S. pierantonius* are capable of synthesizing serine and using it to produce glycine. While both can also produce cysteine, they use different pathways, the host using cystathionine as its intermediary and the symbiont acetyl-serine. Pyruvate is used as their carbon backbone by another group of amino acids and interestingly only *S. oryzae* can produce a member of this group, alanine, through the transamination of pyruvate using alaA. This means that *S. pierantonius* is dependent on its host for this NEAA. Pyruvate can also be used to obtain valine and leucine, but neither member of the consortia can catalyse these reactions given that both ilvD and leuA appear to be pseudogenized in *S. pierantonius* and *S. oryzae* lacks all genes of the pathway except for an aminotransferase involved in the last step (see Figure IV.3).

Regarding the aromatic amino acids, *S. pierantonius* is capable of catalysing the reactions leading to chorismate through the condensation of phosphoenolpyruvate and erythrose-4-phosphate. The endosymbiont is also capable of producing both phenylalanine and tyrosine using the chorismate, and the host can produce tyrosine through the addition of a hydroxyl group to the phenylalanine. Conversely, neither can produce tryptophan, therefore the insect must obtain all the amount it needs entirely from its food. Finally, the holobiont is unable to produce histidine, the complete pathway is missing from both members and therefore it must also be obtained from its diet.

In summary, alanine and proline are provided by the host, threonine, lysine, phenylalanine and arginine by the endosymbiont and valine, leucine, isoleucine, tryptophan and histidine must be obtained from their diet. While methionine might need to be obtained from its diet, we should not discard the possibility that *S. oryzae* might be capable of producing methionine using cystathionine. It is also worth considering that while grains are a rich source of several amino acids, they are generally poor in lysine and threonine and in some cases tryptophan as well (Jiang *et al.,* 2016). Taking this into account it is remarkable that while *Nardonella* was only able to produce tyrosine (Anbutsu *et al.,* 2017) and depended on the host for the provision of the remaining amino acids, *S. pierantonius* has conserved the pathways for several amino acids, namely lysine, threonine, phenylalanine and arginine. This would have given a nutritional advantage to *S. pierantonius* over *Nardonella*.

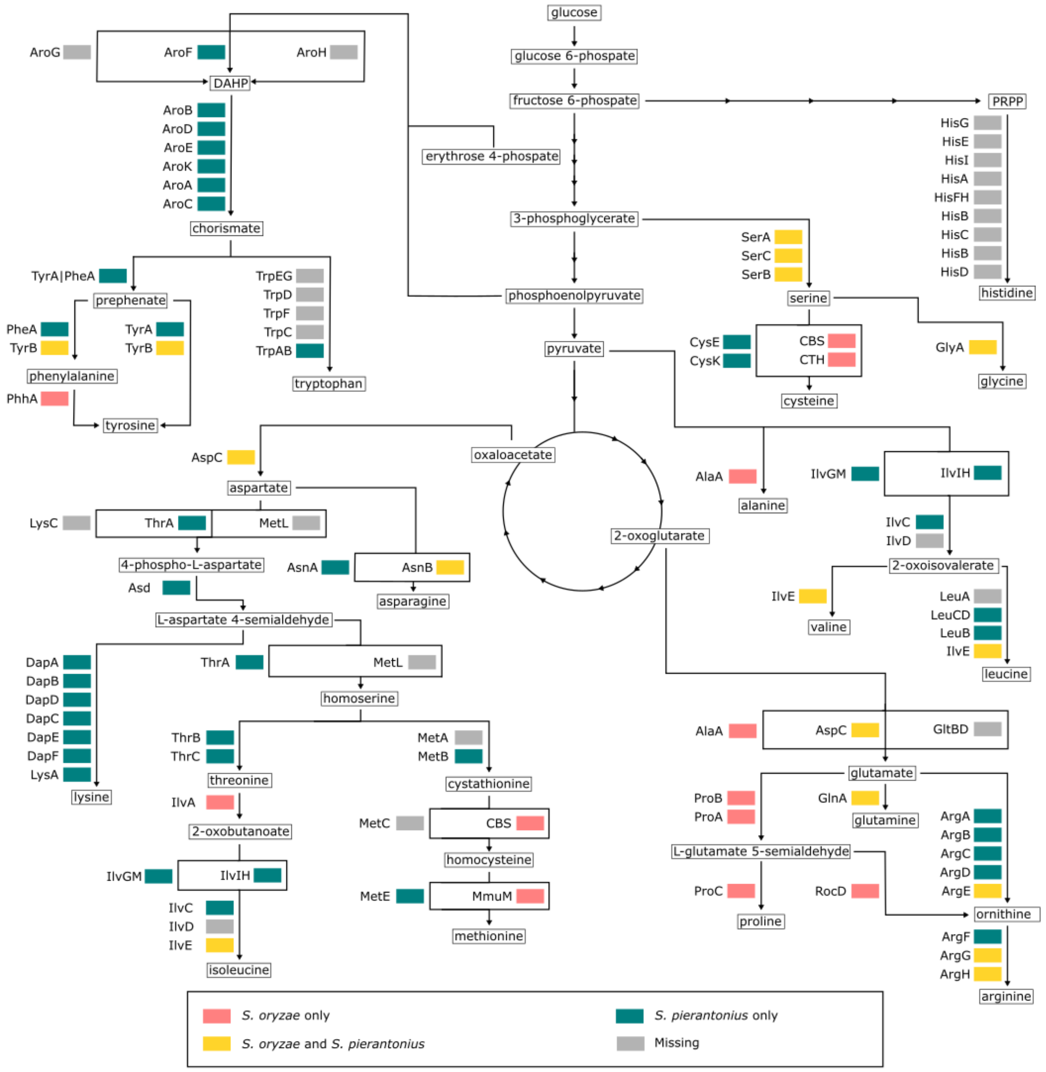

Figure IV.3 Amino acids biosynthesis pathway reconstruction in the S. oryzae and S. pierantonius holobiont.

Finally, perhaps the most interesting amino acid regarding *S. oryzae* is tyrosine given its role in the cuticle formation. The cuticle is made up by chitin lipids, hydrocarbons, proteins and polyphenols with the last two components in some cases representing up to 33% of the total cuticle weight (Kramer *et al.,* 1989). Given the fact that polyphenols are produced mainly via the oxidation of tyrosine, this amino acid must be readily available at the time of each moult and especially at the last one. To cope with this requirement, during their larval phase, insects store aromatic amino acids in proteins known as arylphorins which accumulate in the haemolymph and fat body of insects (Delobel *et al.,* 1993). Remarkably, besides accumulating these storage proteins, *S. oryzae* exquisitely controls the growth of *S. pierantonius* allowing it to increase in number before its last moult to allow a rapid increase in the amount of tyrosine right when it is needed (Vigneron *et al.,* 2014). Additionally, it has recently been demonstrated that *Nardonella* specializes solely in producing this amino acid (Anbutsu *et al.,* 2017), thus highlighting its importance. While *S. pierantonius* can produce tyrosine and most likely also supply it to its host, it is also important recalling that it is able to produce phenylalanine, which is the precursor of tyrosine in the weevil, as well. Therefore, *S. oryzae* could use phenylalanine as an additional source of tyrosine.

IV.2.4 Identification of genes involved in immune system pathways

As aforementioned, *S. oryzae* is an interesting model to study the insect's innate immune system for several reasons. One of them is the fact that it harbours a recently acquired intracellular endosymbiont, thus posing an interesting dilemma for the host, given that it needs to tolerate and control the proliferation of its endosymbiont while maintaining the ability to respond to other infections. In addition, the fact that it spends all its non-adult stages inside the cereal grains, in a relatively sterile environment, could relax the need for a strong immune system, at least during these stages.

Given that *S. pierantonius* is a Gram-negative bacterium, perhaps the most interesting immune system pathway to analyse is the IMD pathway considering its role in the recognition of DAP-type (diaminopimelate) peptidoglycan (DAP-PG) of bacteria and the fact that *S. pierantonius* has the pathway that generates peptidoglycan intact (Oakeson *et al.,* 2014; Maire *et al.,* 2018). Additionally, it is known that *S. pierantonius*, contrary to most long-lasting endosymbionts, can elicit an immune response when injected in the haemolymph of *S. oryzae* (Anselme *et al.,* 2008; Vigneron *et al.,* 2012). As aforementioned, the main receptors associated with this pathway are PGRP-LC and PGRP-LE. A PGRP receptor was identified in the weevil's genome and additional clustering analysis suggested that this receptor could not be included in the groups of either PGRP-LC, LE or LF and that it should rather be placed in a new group with members equally alike to all three groups (J. Orlans, personal communication, 2018). A similar case was previously observed in *C. floridanus* (Gupta *et al.,* 2015).

Remarkably, PGRP-LE was not identified in either *S. oryzae* or the tsetse fly *G. morsitans* (International Glossina Genome Initiative, 2014) and this is especially relevant given that both harbour intracellular endosymbionts. The fact that they are missing the main intracellular receptor for gram-negative bacteria suggests that they might be unable to detect intracellular bacteria, possibly as a result of the adaptation to endosymbiosis. Besides lacking this intracellular receptor, the remaining members of the IMD pathway are conserved with an almost perfect one-to-one ortholog relationship with *T. castaneum* (Zou *et al.,* 2007) and *D. melanogaster* (Myllymäki *et al.,* 2014), thus suggesting that this pathway is highly conserved among different insect orders (see Figure IV.4). This agrees with the conclusions of Maire *et al.,* (2018) where they demonstrated that both the internal and external bacteriome immune responses rely on IMD and Relish. The fact that the IMD immune response is tissue specific and dependent on the endosymbiont load suggests that there might be additional mechanisms for sensing intracellular bacteria besides PGRP-LE. Additionally, this so far unknown receptor might be essential for modulating the immune response and thus exhibiting different tissue specific responses.

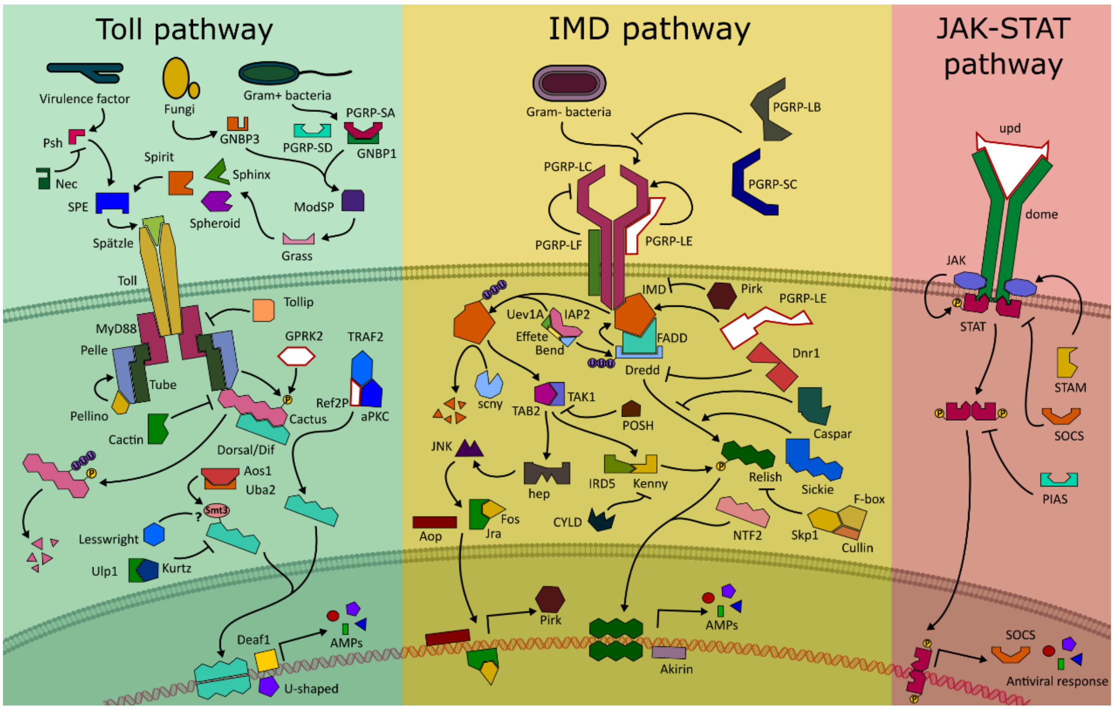

Figure IV.4 Reconstruction of the immune signalling pathways in S. oryzae. Missing elements are identified outlined in red and painted white.

The Toll pathway is involved in the recognition of fungi and Gram-positive bacteria. As mentioned before, the pathway starts with the activation of the Toll receptor by spätzle, however it must be processed before it can activate the receptor. This processing is the outcome of a cascade involving several receptors and proteases. While all of them are most likely conserved in *S. oryzae*, establishing a one-to-one relationship of orthology with the sequences from *D. melanogaster* or *T. castaneum* was impossible given the complexity of the serine proteases and serine protease inhibitor families. These families have a large and highly variable repertoire particular to each species. Regarding the receptors, 11 different Toll receptors were identified which is comparable to the 11 identified on *T. castaneum* and the 10 identified on *D. melanogaster*. Of the remaining members of the Toll pathway, all were identified in *S. oryzae* except for Gprk2 and ref(2)P (see Figure IV.4). Ref(2)P is a partner of the atypical protein kinase C which is capable of directly activating Dorsal, and while it has been shown to be important for the correct functioning of the Toll pathway in *D. melanogaster* (Avila *et al.*, 2002) it appears to be missing from *T. castaneum* as well.

On the other hand, Gprk2 is a serine/threonine kinase that modulates G-protein coupled receptors and it is thought to amplify the signal through its interaction with cactus but without being involved in the degradation of this protein (Valanne *et al.,* 2011). This gene was found in *T. castaneum* and therefore, it seems to be the only gene of the Toll pathway missing in *S. oryzae*. While losing this gene might attenuate the immune response, the fact that it is involved in the Toll pathway and thus most likely not involved in the response towards *S. pierantonius,* it seems unlikely that this gene was lost as an adaptation to tolerate the endosymbiont. Another difference is that while *D. melanogaster* has one gene for dif and another for dorsal, *T. castaneum* seems to have two copies for dorsal (Zou *et al.,* 2007; Valanne *et al.,* 2011) and *S. oryzae* a single copy. Both appear to lack dif.

While these pathways respond to different types of stress and react differently to Gram-positive and Gram-negative bacteria, they both produce a cocktail of AMPs with a different profile. Given their small size, AMPs can be challenging to identify and additionally they sometimes are species specific (Franzenburg *et al.,* 2013). In *S. oryzae* we were able to identify six attacins, two cecropins, two coleoptericins, one defensin, four thaumatins and five lysozymes. Several of these had already been identified on previous studies (Login *et al.,* 2011; Vigneron *et al.,* 2012; Masson *et al.,* 2015) and it is known that they have different tissue specific profiles (Anselme *et al.,* 2008, Maire *et al.,* 2018).

Among the identified AMPs, the most relevant is coleoptericin A, which is known to inhibit the cellular division of *S. pierantonius*, thus causing large filamentous polyploid cells which are unable to escape the bacteriocytes (Login *et al.,* 2011). The bacteriostatic activity of this AMP seems to have been selected to keep the endosymbiont confined without the need for a novel control mechanism. Interestingly, this AMP is highly expressed in the bacteriome while other immune genes are not in normal conditions (Anselme *et al.,* 2008). However, this tissue can respond to infection and activate the production of other AMPs which are exported to the surrounding tissues to contend with the invading pathogens (Masson *et al.,* 2016). While the exact mechanisms that involved in the recognition of the endosymbiont are unknown, it has been demonstrated that both the endosymbiont control and the immune response against pathogens are mediated by the IMD pathway (Maire *et al.,* 2018). This raises the question of how are both different responses mediated by the same pathway and if there is an external factor provided by the endosymbiont or if the attenuation of the immune response in the bacteriome is mediated solely by the host. If the latter is the case, it would have to involve non-canonical elements of the immune system pathways.

Regarding the defence against viruses, from the JAK-STAT signalling pathway we were able to identify all members of the pathway except for unpaired, the ligand of the domeless receptor (see Figure IV.4). This is not too surprising given that it is also missing in *T. castaneum*, therefore perhaps there is a different ligand in Coleoptera capable of activating the receptor. With respect to the RNA interference pathways, *S. oryzae* contains all members of the piRNA pathway. The main difference is that unlike *D. melanogaster*, which has different piwi and aubergine genes (Haase, 2016), both *S. oryzae* and *T. castaneum* have a single gene. Concerning the siRNA pathway, *S. oryzae* contains every member of the pathway with one-to-one orthology to *D. melanogaster* and *T. castaneum*. The only difference is that *T. castaneum* has two copies of the nuclease argonaute 2. Finally, regarding genes involved in the miRNA pathway, all genes are found in *S. oryzae* and all of them have a one-to-one orthology relationship with *T. castaneum* and *D. melanogaster*. The fact that practically every gene involved in the RNA interference pathways is conserved between *S. oryzae* and *T. castaneum* and even *D. melanogaster* highlights the importance of this mechanism in insects.

In summary, the immune system pathways of *S. oryzae* seem quite similar to those of *T. castaneum* and even *D. melanogaster*. Genes involved in the signalling cascades are conserved among these insects and the most notable differences are in the receptors and the effectors. One of the most striking differences is that *S. oryzae* lacks a PGRP-LE receptor, which might reduce its efficacy at identifying intracellular bacteria, including endosymbionts. Another missing gene is the kinase Gprk2, which is involved in the amplification of the signal in the Toll pathway and thus might not be essential. Additionally, establishing a clear orthology relationship to several of the proteases involved in the cascade upstream of the Toll receptor was not possible. Finally, AMPs already identified on previous studies were annotated in the genome assembly and new ones were identified as well.

IV.2.5 Manual analysis of other specific aspects

Horizontally transferred genes

Besides identifying the genes involved in the innate immune system pathways and in the metabolism of amino acids we were able to evaluate numerous other aspects of the genome of *S. oryzae*. We assessed horizontal gene transfer and notably we identified hundreds of lateral gene transfer candidates; however, after closer scrutiny most of them were discarded given that the majority were likely retroviral transfers. We kept 22 candidates, manually reannotated them and attempted to identify their putative donor. Perhaps unsurprisingly most of the LGT candidates had a digestive role with 20 of the candidates being putative plant cell wall degrading enzymes (PCWDEs). In fact, 17 of these enzymes had previously been described on *S. oryzae* using a midgut cDNA sequencing approach (Pauchet *et al.,* 2010). Out of the three novel candidates, one is virtually an identical copy of one of the previously reported endo-beta-1,4-glucanases, and the other two are putative glycosyl hydrolase family protein 32 with orthologs the closest orthologs in the sugar cane weevil *Sphenophorus levis*.

The role of the two remaining candidates seems harder to define. One of them is a putative uracil-DNA glycosylase, and while this enzyme is found in both prokaryotes and eukaryotes there are significant differences between the enzymes from each domain. *S. oryzae* contains a gene coding for the eukaryotic version of the protein while it also contains another copy which might be of bacterial origin. The fact that this gene is expressed and found in two exons suggests that this gene has been fully integrated into the genome of *S. oryzae* while is putative role is unclear.

The last candidate is a hypothetical protein which is expressed as well and with orthologs among other members of the order Coleoptera; however, there are proteins with a high degree of identity also in bacteria. The role of this gene is a complete mystery. Regarding the origin of the sequences, ten seem to be of fungal origin, most likely occurring in two different events. A transfer of an endo-beta-1,4-glucanase GH45 from an unknown fungal donor and its subsequent amplification into four copies. The other an endopolygalacturonase GH28 from a member of the Leotiomyceta which amplified into six copies. The remaining 12 genes are of bacterial origin occurring in five different events. The first is a glycosyl hydrolase family protein 32 from an Enterobacteriales donor that has duplicated, the second is a hypothetical protein from an unknown bacterial donor, the third is a glycoside hydrolase family protein 48 from a Streptomycetaceae donor an which has duplicated, the fourth is a pectin methylesterase from an Enterobacteriales donor which has expanded five times, and the fifth is the uracil-DNA glycosylase from an unknown bacterial donor. Currently we can’t discard the possibility that more than one gene was transferred in a single event.

S. pierantonius resides in the ovaries of *S. oryzae* and infects the oocytes very early during their development. Furthermore, this endosymbiont is not secluded within an M3 membrane (or symbiosomal membrane derived from the bacteriocyte’s membrane), contrasting with other endosymbionts such as *Buchnera*. Therefore, one would think that it is in the ideal environment to transfer genes to its host that would be incorporated in their genome and inherited by the offspring. Nonetheless, we did not find any evidence of host genes originating from *S. pierantonius*. This suggests that *S. oryzae* has an efficient system for controlling the incorporation of foreign DNA into the germline genome or perhaps *S. pierantonius* is not a good donor. On the other hand, the genes that have an Enterobacteriales donor might come from *Nardonella* but determining if this is the case would be complicated given the lack of a reliable free-living relative of *Nardonella*.

This is in line with what has been observed on aphids, where some ancestral transfers of carotenoid biosynthetic genes from fungi were detected (Moran and

Jarvik, 2010). Additionally, genes of bacterial origin were also detected but their suggested role is not a nutritional one, but rather involved in the maintenance of *Buchnera* (Nikoh and Nakabachi, 2009). The donor for these genes is unknown but they seem to come from a rickettsial bacterium closely related to *Wolbachia*. Additionally, the possibility that *Buchnera* has transferred genes to its host has been discarded (Nikoh *et al.,* 2010). These observations contrast with what has been observed in *G. morsitans* where large segments of several hundred kilobases of the genome of *Wolbachia* have been integrated into the genome of the host (Doudoumis *et al.,* 2012; Brelsfoard *et al.,* 2014). It is thought that these events can promote reproductive isolation in their hosts.

Global metabolic network

The team of Federica Calevro has reconstructed the metabolic network of *S. oryzae* using the CycADS pipeline. A total of 1,205 enzymes activities were predicted, demonstrating that it does not differ much from other beetles such as *D. ponderosae* with 1,170 predicted enzymatic activities and *T. castaneum* with 1,282. Afterwards, 40 pathway maps were analysed focusing on major metabolic pathways from the KEGG Pathway Database. Interestingly, the metabolic network of *S. pierantonius* was reconstructed as well and incorporated into the reconstruction of *S. oryzae* thus building a metabolic reconstruction focusing on the metabolic interplay and the pathway interconnections between the associated partners (see Figure IV.5).

When comparing the metabolic reconstruction of *S. pierantonius* with that of *S. praecaptivus* it became evident that *S. pierantonius* has lost the nitrate and nitrite reductase activities and is thus unable to assimilate nitrate and reduce it into ammonia. Therefore, the organic nitrogen must be obtained from the food. This is not an issue for *S. oryzae* given that it has midgut digestive proteinases (Liang *et al.,* 1991) which release the amino acids from food proteins. The main candidate is the diamino acid glutamine, which is very abundant in the prolamins, a group of seed storage proteins found in several monocotyledons (Shewry and Halford, 2002). The nitrogen from this amino acid can afterwards be used by the glutamine synthetase/glutamine oxoglutarate aminotransferase

(GS/GOGAT) in both *S. oryzae* and *S. pierantonius* for its incorporation in other molecules.

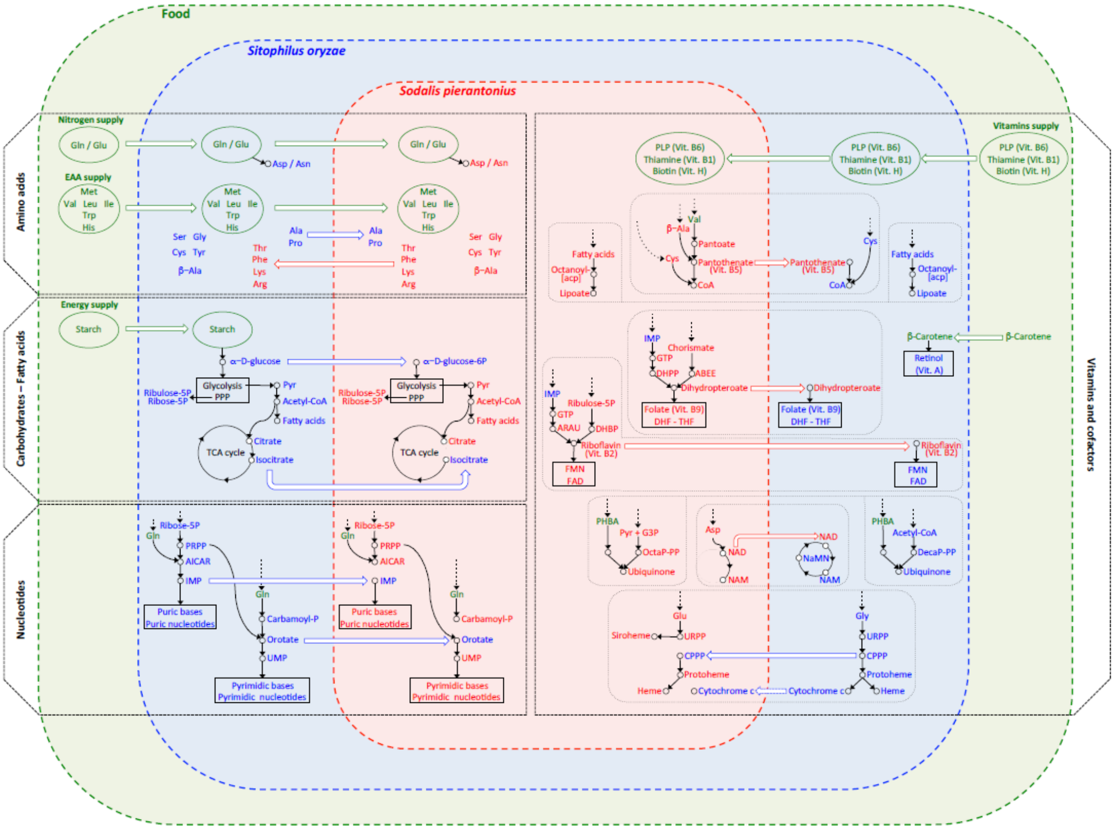

Figure IV.5 Metabolic reconstruction of the S. oryzae-S. pierantonius holobiont. The inner compartment in red exemplifies S. pierantonius, the blue one surrounding it represents S. oryzae and the outermost on green represents the food source. White arrows represent the fluxes between compartments and the colour surrounding the arrow represents the source of the metabolite being transported.

The starch is the main component of cereal seeds and it is the principal source of energy for the weevil holobiont. Glucose and/or glucose-1-phosphate obtained through the digestion of starch are the entry metabolites of the glycolysis pathway, which is completely operative in *S. oryzae* just like the citric acid cycle. *S. pierantonius* internalizes glucose from its host using a specific sugar transporting phosphotransferase system which simultaneously transports and phosphorylates glucose. Glycolysis is fully preserved in *S. pierantonius* while the citric acid cycle is interrupted by the pseudogenization of the gene encoding aconitate hydratase that interconverts citrate and isocitrate and therefore the host must provide it with isocitrate (see Figure IV.5). Additionally, citrate is probably a dead-end product that must shuttled to the host for salvage. The pentose phosphate pathway, which generates pentoses including the ribose-5-phosphate precursor of nucleotide synthesis, was identified in both host and endosymbiont.

S. oryzae, was also demonstrated to be capable of performing salvage and *de novo* biosynthesis of purines and pyrimidines thus being able to incorporate preformed nitrogenous bases obtained from their diet. Nonetheless, if required, it can also produce inosine monophosphate (IMP, the precursor of purines) from ribose-5-phosphate and amino acids and uridine monophosphate (UMP, the precursor of pyrimidines) from carbamoyl phosphate. Remarkably, *S. pierantonius* is only capable of executing the salvage pathways thus being dependent on the supply of IMP and UMP by *S. oryzae* (see Figure IV.5).

Vitamins are essential for the correct functioning of the metabolic processes of an organism and those that are incapable of synthesizing them must obtain them from an external source. While *S. oryzae* might obtain an adequate amount of some vitamins from its diet it also depends on *S. pierantonius* for the supplementation of those that are deficient in cereal grains. The biosynthesis pathways of vitamins B6, B1 and H (PLP, thiamine and biotin respectively) are disrupted in *S. pierantonius* and *S. oryzae* is incapable of producing them, therefore they must be obtained from their diet. These dietary requirement for thiamine contradicts previous reports (Oakeson *et al.,* 2014). *S. pierantonius* has kept full pathways for the biosynthesis of vitamin B5, B2 and B9 (pantothenate, riboflavin and folate respectively) given it is provided with the adequate precursors, namely valine for B5 and IMP for both B2 and B9 (see Figure IV.5). After *S. oryzae* is supplied with these vitamins by the bacterial symbiont, it can perform the final reactions to synthesize the active cofactors. Hence, for these three vitamins, host and bacterial pathways are highly interconnected and interdependent thus highlighting the coevolution of both symbiotic partners and rapid genomic reshaping.

Insecticide resistance genes

Resistance to insecticides is an increasing problem when trying to ensure a sustainable production of plant crops. Given their short generation times and metabolic arsenal to detoxify plant secondary metabolites insects are quite successful at developing resistance to the most commonly used insecticides. This is especially relevant in grain weevils, given that it has been demonstrated that they are the oldest pest through the analysis of pottery up to 10,500 years old (Obata *et al.,* 2011). Therefore, it is very likely that *S. oryzae* has been exposed to many xenobiotic agents throughout the history of its interaction with humans and has managed to withstand them.

While many strategies are employed by insects to counteract insecticides, it is known that several gene families are particularly relevant when developing resistance to xenobiotics (Panini *et al.,* 2016). 317 genes belonging to these families were identified: 62 ABC transporters (ABC), 52 carboxyl/cholinesterases (CCE), 90 cytochrome P450s (CYP), 35 glutathione s-transferases (GST), 34 ligand-gated ion channels (LGIC), 38 glucuronosyltransferases (UGT) and six voltage gated sodium channels (VGSC). Broadly, regarding the number of genes, *S. oryzae* is within the range of other beetles.

At a glance *S. oryzae* seems to have an average amount of CYPs and less than *T. castaneum*, which lives in a comparable environment and is exposed to similar types of insecticides. However, there is an expansion of the CYP6BQ gene from the CYP3 family reaching 18 copies. These members of the family are known to be involved on insecticide resistance and studies have shown that while there are differences in the gene copies the affinities are not significantly different, but rather the resistance is achieved varying expression levels (Zhu *et al.,* 2013). Additionally, the mitochondrial cytochromes CYP12 and the cytochrome CYP9 from the family CYP3 are also known to be implicated in detoxification but the number of genes is similar among beetles.

Regarding GSTs, *S. oryzae* is the second beetle with most members of the Epsilon class and these are known to be implicated in insecticide resistance while the Delta class is reduced, like in all other beetles (Shi *et al.,* 2012). The LGICs superfamily includes nicotinic acetylcholine receptors (nAChr) which might be implicated in sensitivity to neonicotinoids, including imidacloprid (Jones and Sattelle, 2007; Clements *et al.,* 2016). *S. oryzae* is the beetle with the most members in the nAChr class among those studied. The C class from the ABC superfamily has experienced a great expansion in *T. castaneum*, and this expansion was also observed in *S. oryzae*. Interestingly this class seems to be involved in insecticide resistance (Broehan *et al.,* 2013). The B and G classes might also be implicated, but in these cases an expansion has not taken place in *S. oryzae*.

The CCEs superfamily contains members that are strongly involved in insecticide resistance, particularly members of the A class (Lü *et al.,* 2015) which are greatly expanded in *S. oryzae* and *L. decemlineata*. The mechanism of action might be sequestering the insecticide rather than catabolizing it given the observable cross-resistance of this resistant strain to other hydrophobic insecticides, such as other SPs and DDT. Abamectin might be sequestered by members of the class D; however, there is only a single member of the D class in *S. oryzae*.

In summary, while *S. oryzae* does not have an exaggerated arsenal to cope with insecticides it is certainly prepared to deal with insecticides just as efficiently as Tribolium or Leptinotarsa. Perhaps some of the most interesting expansions are the class A of the CCEs superfamily and the CYP6BQ class of the CYP superfamily. Nonetheless, it seems that simple solutions such as increasing the expression of a gene or point mutations can confer resistance to insecticides. Additionally, expansions of these types of genes have been observed in other insects, and while they might be related with transposable elements it does not seem to be a feature particular to *S. oryzae*.

IV.3 Conclusions

S. oryzae has the third largest genome among the Coleoptera with data available in NCBI. Its genome size and number of genes are within the range of those of other beetles.

The size of the repeatome in *S. oryzae* is among the largest found in insects and the largest of any studied Coleoptera. This might be one of the reasons behind the large genome size of this beetle. Additionally, while highly interesting it complicated obtaining a genome with a larger N50.

S. oryzae had the fastest gene family expansion rate among the beetles evaluated and some of these families included those potentially involved on insecticide resistance. It is important to keep in mind that while the high number of genes might be due to fragmented genes being counted twice, it is undeniable that *S. oryzae* has numerous gene families that have rapidly expanded.

The interdependence between the host and the endosymbiont is easily observed when analysing not only amino acid biosynthesis, but the whole putative metabolism between both partners. Regarding EAAs, some are provided by the host, others by the endosymbiont and some must be obtained from the diet. Also, perhaps an advantage *S. pierantonius* had over the ancestral endosymbiont *Nardonella* was its ability to synthesize more amino acids, the most important might have been lysine and threonine given the low abundance of these amino acids in cereals. This is certainly a possibility given the fact that *S. pierantonius* has already lost the ability to produce several amino acids but has conserved these two among others, including tyrosine that is especially relevant for the production of the beetle’s cuticle. Thus, the replacement of *Nardonella* by *S. pierantonius* could have had a major role in the success of the *Sitophilus* lineage. The availability of genomic data of these two interacting organisms permitted a detailed analysis of potential interdependencies in the exchanges of vitamins and cofactors, and this has already helped to confirm and reject previous studies.

Regarding the immune system pathways, these pathways seem highly conserved when compared to the models *T. castaneum* and *D. melanogaster*. One of the most striking differences is that *S. oryzae* lacks a PGRP-LE receptor which might explain in part how weevils tolerate their endosymbiont. Additionally, both novel and already known AMPs were identified. The fact that the immune system of *S. oryzae* is so similar to that of other non-symbiotic insects suggests that small modifications perhaps are enough to allow harbouring an intracellular pathogen, such as the adaptation of colA to control the growth and localization of *S. pierantonius* and perhaps the loss of PGRP-LE.

We did not find any event of HGT in the genome of *S. oryzae* that seemed to have originated from *S. pierantonius*. This suggests that perhaps *S. oryzae* has an efficient system for controlling the incorporation of foreign DNA or perhaps there just hasn’t been enough time given the recent origin of the association.

While *S. oryzae* has an averagely sized set of genes to cope with insecticides it seems to be able to manage xenobiotics just as efficiently as *L. decemlineata*. Additionally, the expansions of these types of genes might be related with transposable elements.

These novel insights on the biology of *S. oryzae* and the availability of its genome sequence will provide hints as to novel ways to control this pest insect with better planned strategies directed to specific aspects of its metabolism ultimately decreasing the losses this insect causes to the agriculture. Additionally, through the evaluation of the repertoire of insecticide resistance genes better insecticide candidates could be selected. By eliminating or decreasing the efficiency of its endosymbiont at producing tyrosine the adults would experience a great decrease in their invasiveness. Conversely, genes known to keep the endosymbiont under control could be selected to develop RNAi-Mediated Crop Protection (developing plants which stably express double-stranded RNAs (dsRNAs) that target genes in their pests). One example would be targeting highly specific receptors involved in the recognition of entomopathogens to increase their efficiency.

V CHAPTER 2: THE CEDAR APHID *CINARA CEDRI*

V.1 Introduction

As it was mentioned above, aphids are known to harbour more than one endosymbiont and among them, the Lachninae subfamily seems to be especially prone to harbour other endosymbionts in addition to *Buchnera*, most commonly bacteria from the genus *Serratia*. Therefore, they are good models for studying endosymbiont complementation, and the process by which a secondary endosymbiont drifts into to a coprimary (Manzano-Marin *et al.,* 2017). The *Cinara* genus from the Lachninae subfamily, especially the cedar aphid *Cinara cedri* (see Figure V.1) is among the best studied models in addition to *A. pisum* from the Aphidinae subfamily.

Figure V.1 The cedar aphid C. cedri. With permission from Angel Umaran.

The sequencing of both endosymbionts from *C. cedri*: *Buchnera* BCc (Gomez-Valero *et al.,* 2004; Pérez-Brocal *et al.,* 2006) and *S. symbiotica* (Lamelas *et al.,* 2011) determined that these two endosymbionts are not mere cohabitants of the bacteriocyte but rather they have established a co-obligate association with the aphid host. It was observed that *Buchnera* BCc has lost the ability to synthesise riboflavin and tryptophan (Pérez-Brocal *et al.* 2006) and while the biosynthesis of riboflavin is now performed entirely by *S. symbiotica* (Lamelas *et al.,* 2011), the biosynthesis of tryptophan is shared between *Buchnera* BCc and *S. symbiotica*. This involves the biosynthesis of anthranilate by *Buchnera* BCc and its subsequent transfer into *S. symbiotica* which converts it to tryptophan thus making both endosymbionts essential for their host and for each other as well (Gosalbes *et al.,* 2008). Additionally, the sequencing of *Buchnera* BCc marked a milestone given that up to date it is the smallest sequenced *Buchnera* genome. Further sequencing of the co-obligate endosymbionts of two other members of the Lachninae subfamily, *Cinara tujafilina* and *Tuberolagnus salignus* confirmed that the establishment of the consortium *Buchnera-Serratia* predates the diversification of the linage (see Figure V.2), as all *Buchnera* have small genomes and have lost the complete pathway for the biosynthesis of riboflavin (Manzano-Marín and Latorre, 2014; Manzano-Marín *et al.,* 2016).

C. cedri lives in colonies on gymnosperms, and while it can feed on several species of the *Cedrus* genus, it is generally found on *Cedrus atlantica* and *Cedrus deodora*. Its original indigenous area was believed to be the Moroccan Medium Atlas Mountains, where the samples used to describe the species for the first time were collected (Remaudiere, 1954). However, recently a parasitoid, *Pauesia anatolica*, was found to be capable of targeting *C. cedri*. It is believed that when cedar trees were exported, they carried the aphids, but not the parasitoids, thus complicating the identification of the parasitoid (Michelena *et al.,* 2005). Given the geographical distribution of *P. anatolica*, *C. cedri*'s native range is now believed to be southern Turkey.

C. cedri was first observed in Europe in Italy on 1974 (Covassi and Binazzi, 1974; Binazzi, 1978) and since then its presence has been documented in almost all of Europe and in many other countries of the Near East. In Europe, *C. cedri* has only been reported to feed on *C. atlantica* and *C. deodara* while in Turkey it has been observed on the Cedar of Lebanon *Cedrus libani* (Tuatay and Remaudière, 1964; Covassi and Binazzi, 1974; Notario *et al.,* 1984). While the two formers have never been reported to be damaged by *C. cedri*, *C. libani* is known to suffer early needle falling, and chlorosis among other types of damage (Binazzi *et al.,* 2015). This is of particular concern given the vulnerable conservation status of the Cedar of Lebanon. When the climatic conditions are favourable, this aphid can develop very dense colonies on cedar's branches. These infestations are easy to spot given the large amounts of honeydew that are excreted which allow the sooty mold fungi to thrive (Binazzi and Scheurer, 2009) but it also provides an important resource that honeybees can use to produce honey (Ülgentürk *et al.,* 2013).

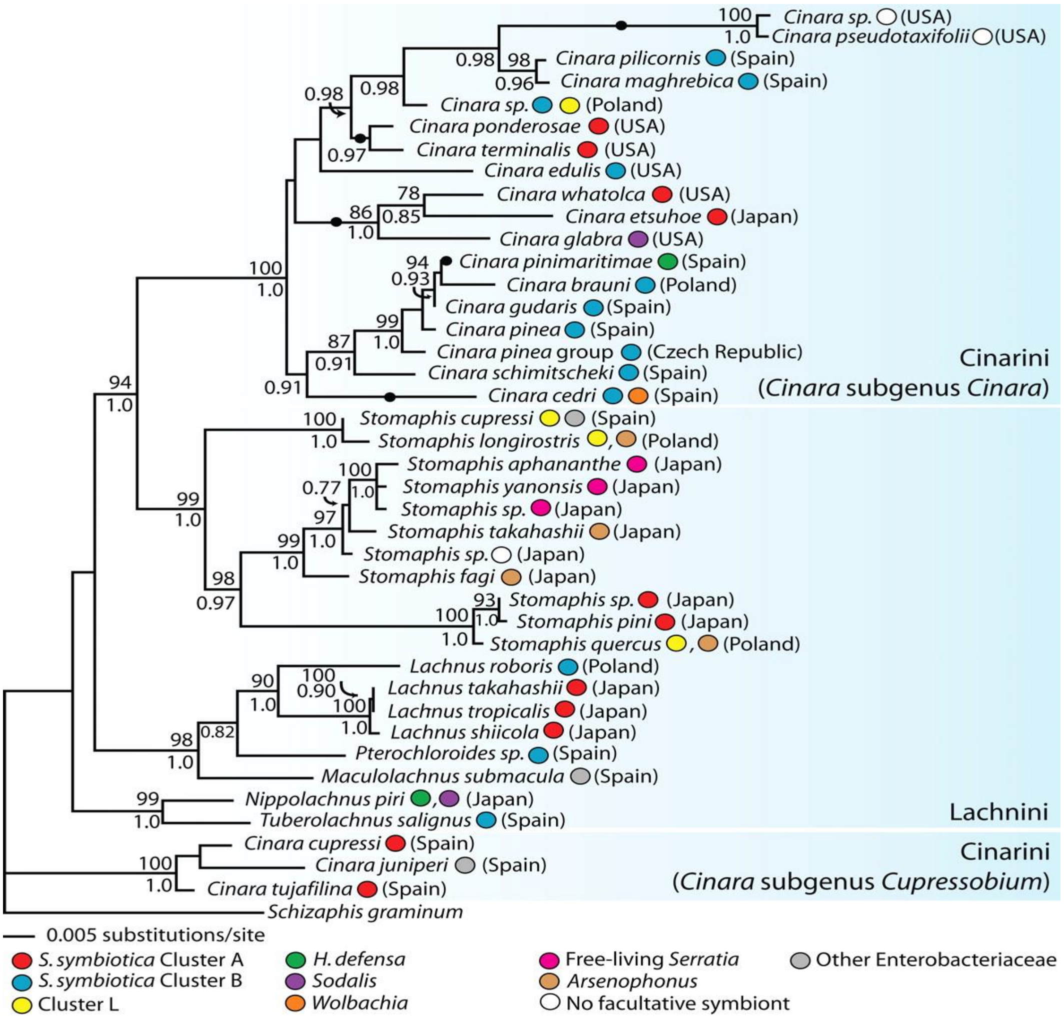

Figure V.2 ML tree using the 16S rRNA gene from Buchnera from members of the Lachninae, based on genes. Most species of Lachninae sampled were infected with a facultative symbiont. There are two main S. symbiotica clusters and they display a strong geographic pattern. Cluster A is found mainly in Asia and North America and cluster B in Europe. From Burke, Gaelen R, et al. 2009. "Evolution and Diversity of Facultative Symbionts from the Aphid Subfamily Lachninae." Applied and Environmental Microbiology 75 (16). American Society for Microbiology: 5328–35.

V.2 Results and discussion

V.2.1 Genome assembly and annotation

The genome of *C. cedri* was sequenced and assembled at the National Centre for Genomic Analysis (CNAG). The assembled genome comprised 1,740 scaffolds, had a size of 396 Mb and a GC content of 30.55 (see Table V.1). Currently there are seven publicly available sequenced aphids and *C. cedri* would rank on third place regarding its genome size. The completeness of the genome was assessed using BUSCO and 87% of the core genes were found to be complete and 3% fragmented, thus demonstrating that the degree of completeness of the genome was similar, or even higher, to that of other available high-quality aphid genomes (The International Aphid Genomics Consortium, 2010; Mathers *et al.,* 2017). Using a customized annotation pipeline 16,996 protein coding genes were predicted.

Statistic	***C. cedri***	***A. pisum***	***M. persicae***
No. sequences (>= 1 kb)	1,740	12,970	4,017
Largest scaffold	6,897 Kb	3,073 Kb	2,200 Kb
Total length	396 Mb	542 Mb	347 Mb
Scaffold N50	1,239 Kb	519 Kb	436 Kb
GC%	30.55	29.76	30.03
Gap length	1.7 Mb	41.78 Mb	1.84 Mb
BUSCO (% complete/partial)	87/90	82/90	83/91
Gene count (protein coding)	16,996	18,601	18,529
Transcript mean size (bp)	1,873	2,039	2,318

Table V.1 Assembly statistics of C. cedri and a comparison with the statistics of A. pisum and M. persicae.

The size of the genome of *C. cedri* was 50 Mb larger than that of *M. persicae*, conversely it is around 150 Mb smaller than that of *A. pisum*, suggesting that the sizes of aphid's genomes are within this range. Regarding the quality of the assembly, the N50 and the largest scaffold from the assembly of *C. cedri* are more than twice as big as that of *A. pisum* and thrice as big as that of *M. persicae* thus highlighting the high quality of the assembly. All three aphids have a similar GC content, however there are some differences on the gap length in their assemblies, while for *C. cedri* and *M. persicae* these values are small, in the case of *A. pisum* more than 40 Mb of the assembly are Ns, thus unknown.

There are important differences regarding the number of predicted genes and their sizes between aphids, with *A. pisum* and *M. persicae* having a similar number of predicted protein coding genes, and *C. cedri* having more than 1,500 less predicted genes. Given the high quality of the assembly, it seems likely that these genes are expansions specific to the Aphididae family to which both *A. pisum* and *M. persicae* belong. Finally, the transcripts in *C. cedri* are smaller than those in the other two aphids, this could mean that the gene prediction could require some improvement to identify missing gene features, or that the genes in this aphid are in fact smaller.

V.2.2 Gene orthology

We used EGGNOG-MAPPER to assess the conservation degree of protein-coding genes between *C. cedri* and 20 other arthropod species to determine orthology relationships. These species were selected to represent the four main orders of the holometabolous insects and include a variety of hemimetabolous insects. All the chosen insects have a fully assembled and annotated genome and are well studied. Additionally, they have different diets and environments and several of them harbour endosymbionts. (see Table S.1). To minimize differences in the number of isoforms due to the accuracy of the genome annotations only one isoform (the longest) per gene was used for the analysis.

Around 62% of genes of *C. cedri* have orthologues in at least another arthropod. Of these, 1,341 were universal single copy orthologs (indicated as one2one in Figure V.3) across at least 18 species, which were used to determine the maximum likelihood phylogeny. The phylogeny accurately reflects the known phylogenetic relationships between the species. Interestingly, while *C. cedri* has the smallest number of genes among aphids, it has the largest number of lineage specific genes (6,449, around 38% of its genes) (see Figure V.3). However, among hemipterans, with 10,368 (or 56% of its genes) *D. citri* has the largest set of lineage specific genes. Interestingly, aphids have the largest number of order specific genes among hemipterans, suggesting that they have undergone expansions specific to their superfamily.

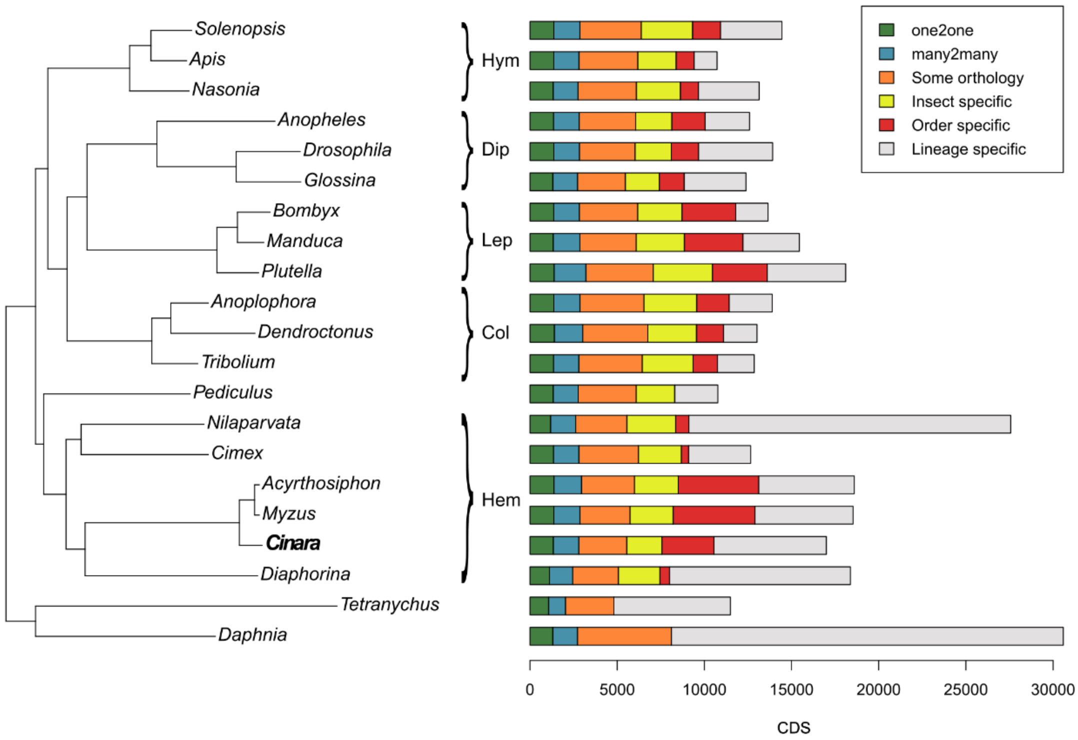

Figure V.3 Phylogenetic relationships of the orthologs of C. cedri and 20 other insect genomes. The one2one category refers to orthologs with only one copy in at least 18 of the species used. The many2many category includes orthologs also found in at least 18 of the species used but with more than one copy in at least one of the species. Some orthology refers to orthologs found in less than 19 of the species used. Insect, family and lineage specific refers to orthologs found in all the species belonging to that group. To group species included in each order curly brackets were used: Hym stands for Hymenoptera, Dip for Diptera, Lep for Lepidoptera, Col for Coleoptera and Hem to Hemiptera.

V.2.3 Identification of genes involved in amino acid biosynthesis

Much like *S. oryzae*, *C. cedri* is incapable of synthesizing the ten EAAs: histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, valine and arginine. While holometabolous insects can exploit several diet sources given the differences between their larval and adult stages, aphids have a somewhat monophagous diet feeding on plant phloem throughout all their lives. Phloem is an unbalanced food source, with a high content of carbon and energy and nitrogen in the form of sugars and free amino acids respectively, and generally free of toxins or feeding deterrents. However, animals feeding solely on phloem, such as aphids, must cope with the low nitrogen quality in sap, meaning a low abundance of EAAs, sometimes even reaching a 1:20 ratio in comparison to non-essential amino acids. This is known as the nitrogen barrier (Douglas, 2006), and to overcome it, phloem-feeding hemipterans rely on endosymbionts that provide them with the amino acids with low abundance in the phloem.

As aforementioned, members of the subfamily Lachninae, such as *C. cedri*, do not rely on a single endosymbiont to produce amino acids and other essential metabolites, but rather two coprimary endosymbionts, namely *Buchnera* BCc and *S. symbiotica* have established a metabolic consortium. Therefore, the metabolic pathways of both endosymbionts were also considered and incorporated into the reconstruction of *C. cedri*. When the genome of *Buchnera* BCc was published, it was revealed that it was capable of synthesizing 12 amino acids, namely nine out of the ten EAAs, lacking only tryptophan (Pérez-Brocal et al, 2006). On the other hand, *S. symbiotica* was only capable of synthesizing six amino acids, including tryptophan which was synthetized by a metabolic complementation between the two bacteria (Lamelas *et al.,* 2011). In most cases, the final steps of the synthesis of the amino acids was proposed to be carried out by the host, indicating not only metabolic complementation between the endosymbionts but also with the host. To evaluate the metabolic capabilities of *C. cedri* to synthesize amino acids, we identified all genes involved in their biosynthesis.

Among the amino acids requiring α-ketoglutarate, we identified the pathways to produce glutamate, glutamine and proline and all of them are missing in both endosymbionts. Regarding arginine, *C. cedri* can produce ornithine from a precursor of proline using rocD; afterwards, *Buchnera* BCc can catalyse the following three reactions to produce arginine (see Figure V.4).

Concerning the amino acids that use oxaloacetate as a precursor, *C. cedri* is only capable of producing aspartate and asparagine. While it has been stated that *S. symbiotica* is capable of catalysing the reaction from aspartate to asparagine (Lamelas *et al.,* 2011), in the current study we were unable to identify the required enzyme, conversely *S. symbiotica* has an L-asparaginase with which it is able to catalyse the reaction from asparagine to aspartate. Also, asparagine is the most abundant amino acid in the phloem of several plants, probably minimizing the relevance of the enzyme catalysing this reaction (Shigenobu and Wilson, 2011). *Buchnera* BCc can produce lysine using aspartate as its precursor, however it lacks the dapC aminotransferase which is one of the differences with *Buchnera* APS from *A. pisum* where this gene is not absent. *S. symbiotica* lacks these gene as well, and the first and the last genes leading from aspartate to lysine, so it seems that only *Buchnera* BCc is capable of synthesizing lysine. *Buchnera* BCc can also produce threonine on its own. For the biosynthesis of isoleucine *C. cedri* and *Buchnera* BCc must cooperate, with the former catalysing the first and last steps of the pathway leading from threonine to isoleucine and *Buchnera* BCc catalysing the remaining three.

Finally, how the cedar aphids obtain methionine remains somewhat obscure given that neither *Buchnera* BCc nor *S. symbiotica* have the complete pathways. While the former has the last step of the pathway, a methionine synthase, it would require an external source of homocysteine and it has been suggested that the host could provide this metabolite through the reversal of the transsulfuration pathway, producing homocysteine from the cysteine. (Wilson *et al.,* 2010). However, it has been demonstrated that the source metabolite to produce methionine is not cysteine but rather cystathionine (Russell *et al.,* 2013). Therefore, exploiting an unknown source of cystathionine, *C. cedri* could use CBS, its cystathionine-β-synthase, reversely to produce homocysteine and the last step would be catalysed by *Buchnera* BCc.

The three amino acids that use 3-phosphoglycerate for their carbon backbone, serine, glycine and cysteine can be synthesized by *C. cedri*. Regarding the endosymbionts, both can produce glycine from serine and *S. symbiotica* is also capable of using serine to produce cysteine via a pathway different from that of their host. This is another of the differences with *A. pisum*, given that *Buchnera* APS is capable of producing cysteine. Of the amino acids that use pyruvate as their precursor, alanine can be produced only by *C. cedri*. While it has been suggested that both *Buchnera* BCc and *S. symbiotica* can produce alanine (Lamelas *et al.,* 2011), on this reassessment of the genomes we were unable to identify the required enzymes and thus, the host seems to be the only source of alanine. Regarding valine and leucine, *Buchnera* BCc is capable of catalysing all the needed reactions except for the last one (see Figure V.4). This step, which requires an aminotransferase, is performed by the host, thus effectively conferring it the ability to control the production rate of both amino acids.

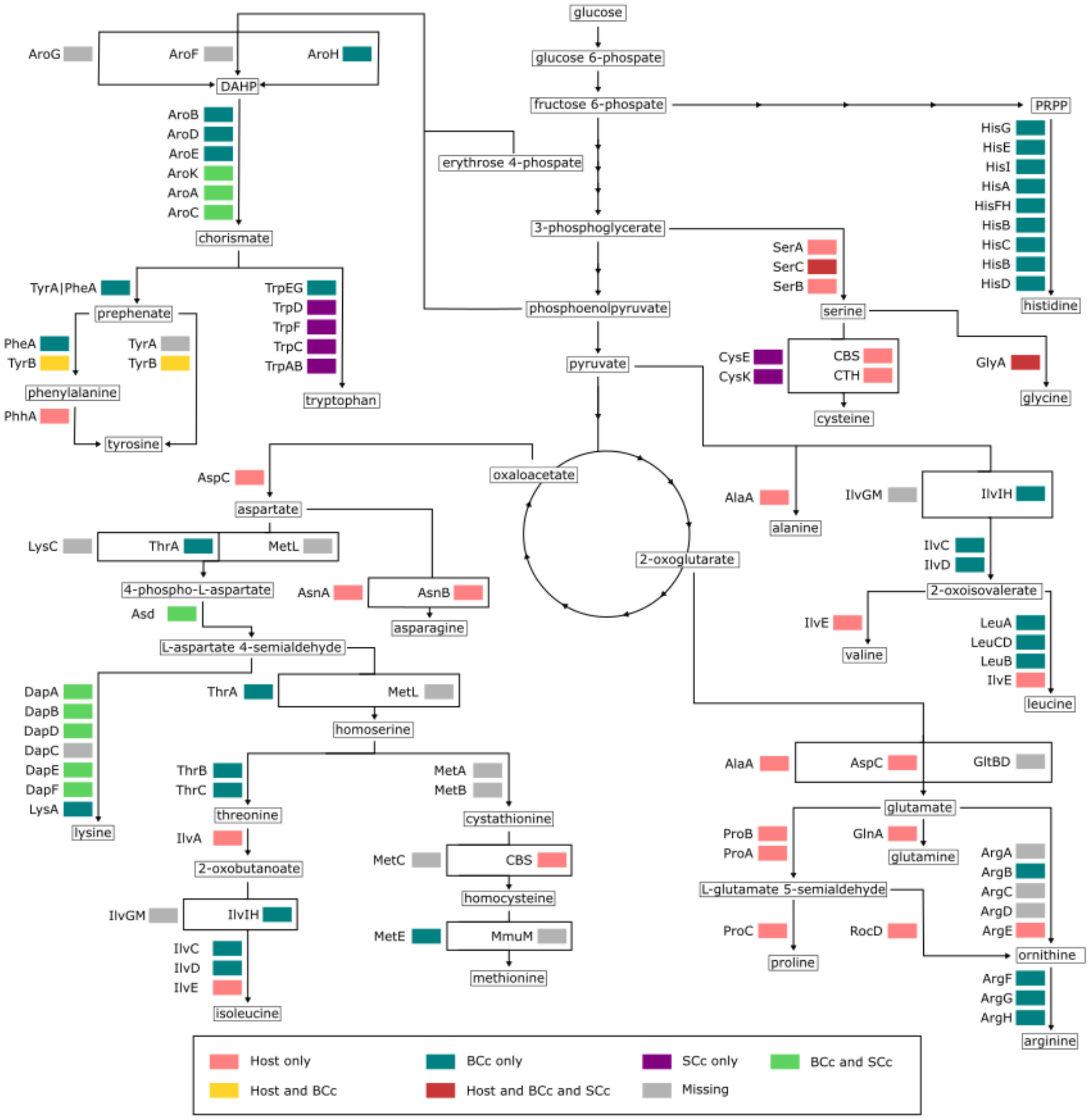

Figure V.4 Amino acids biosynthesis pathway reconstruction of the C. cedri, Buchnera and S. symbiotica holobiont.

Regarding histidine, only *Buchnera* BCc contains all the needed genes and is thus capable of producing it. Finally, in the case of the aromatic amino acids, *C. cedri* is only capable of producing tyrosine using phenylalanine as its substrate. Phenylalanine can be produced by *Buchnera* BCc, and while some steps can also be performed by either *S. symbiotica* or the host, it would appear that *Buchnera* BCc can produce it on its own. On the other hand, the synthesis of tryptophan is the paradigm of cooperation between the members of an endosymbiotic consortia. While *Buchnera* BCc is able to produce anthranilate, the remaining four reactions of the pathway are performed by *S. symbiotica* thus producing tryptophan from anthranilate for the whole holobiont (Lamelas *et al.,* 2011). This is the key difference between *A. pisum* and *C. cedri* given that *Buchnera* APS is capable of synthesizing tryptophan on its own, while on *C. cedri* the collaboration of both primary endosymbionts is needed.

In summary, histidine, phenylalanine, lysine and threonine are entirely provided by *Buchnera* BCc. Serine, alanine, glutamate, glutamine, proline, aspartate, asparagine and tyrosine can be synthesized by the host alone. Cysteine can be produced by the host and *S. symbiotica* independently. Valine, leucine, arginine and isoleucine are provided by a collaboration between *C. cedri* and *Buchnera* BCc. Glycine can be produced by all three components of the holobiont. Tryptophan is generated through a collaboration between *Buchnera* BCc and *S. symbiotica*. And finally, methionine is either obtained from the diet or produced through a collaboration between *C. cedri* and *Buchnera* BCc, however this would require an enzyme from *C. cedri* to work in the opposite direction and this has never been observed in any animal (Russell *et al.,* 2013).

V.2.4 Identification of genes involved in immune system pathways

Given that *C. cedri*, like most aphids, is capable of harbouring multiple endosymbionts one would expect its immune system to have adapted to tolerate and provide an adequate environment for its endosymbiont while still being able to cope with infections. It is known that the immune system of *A. pisum* is reduced when compared to other insect models such as *D. melanogaster* (Gerardo *et al.,* 2010). Several explanations have been postulated for this reduction, notably the fact that their diet is relatively sterile, their large investment in reproduction and the fact that secondary endosymbionts could also confer them a certain degree of protection against pathogens (Altincicek *et al.,* 2008).

Notably, *Buchnera* BCc and *S. symbiotica* are Gram-negative bacteria, thus the IMD pathway would be involved in their recognition. However, while *S. symbiotica* retains the ability to synthesize peptidoglycan and liposaccharides (Lamelas *et al.,* 2011), in this system *Buchnera* BCc has lost all the genes for amino sugar and peptidoglycan biosynthesis as a consequence of the genome reduction it has suffered (Pérez-Brocal *et al.,* 2006). In the case of this pathway, *C. cedri* has suffered important losses comparable to those in *A. pisum*. It lacks all PGRPs, imd, Fadd, Dredd, relish, sickie, CYLD, Fos and pirk (see Figure V.5). Not having PGRPs means that *C. cedri* lacks the main mechanisms to recognize Gram-negative bacteria. The main signalling proteins Fadd, Dredd and Imd are also missing, thus interrupting the signalling cascade from the start. Additionally, the transcription factor relish was not identified, thus all central genes involved in the pathway are missing. Regarding the other losses: Pirk regulates the interaction between Fadd, Dredd and Imd, and CYLD and sickie enhance the activation of relish; thus, unless they have another role they don't appear to be needed. While the IMD pathway seems to be non-functional, the JNK pathway is only missing Fos and therefore it seems functional. While the usual activation of the JNK pathway requires the IMD pathway, it has also been suggested that it could be activated via Eiger and its receptor Wengen (Igaki *et al.,* 2002).

On the other hand, the Toll pathway seems to be rather complete. As mentioned before, the activation of the pathway requires several steps of processing involving a cascade of proteases and establishing a clear one-to-one orthology between the serine proteases in *D. melanogaster* with those *in C. cedri* is complicated, thus it is not clear if the same genes are involved or if some of them might be missing. Additionally, this pathway is also affected by the losses of PGRP receptors and GNBP. While *D. melanogaster* has three GNBP and *A. pisum* two, *C. cedri* has retained only one. It is not clear if one copy is missing from the assembly or if *C. cedri* can activate the pathway using a single more versatile receptor. A functional study is needed to clarify this situation.

Regarding the receptors, nine different Toll receptors were identified, and this number seems comparable to the 13 identified on *A. pisum* and the 10 identified on *D. melanogaster*. Of the remaining canonical members of the Toll pathway, all were identified in *C. cedri* except for ref(2)P (see Figure V.5). While Ref(2)P has been shown to be important for the correct functioning of the Toll pathway in *D. melanogaster* (Avila *et al.,* 2002) it is missing in both *A. pisum* and *S. oryzae*, thus suggesting that this gene might be specific to *D. melanogaster*. Another difference shared between *C. cedri* and *A. pisum* (and *S. oryzae*) is that while *D. melanogaster* has one gene for dif and another for dorsal, the other insects seem to lack dif.

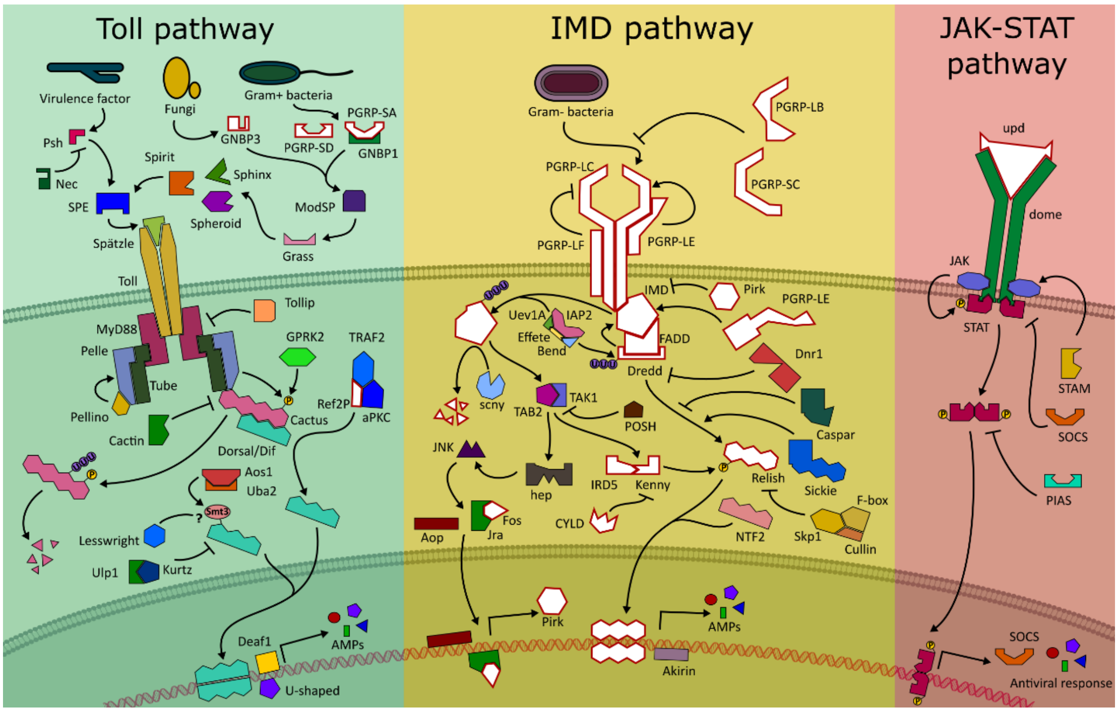

Figure V.5 Reconstruction of the immune signalling pathways in C. cedri. Missing elements are identified outlined in red and painted white.

As aforementioned, the main activation product of these pathways are AMPs, thus an attempt to identify them was performed. Given the fact that they are small and sometimes species specific (Franzenburg *et al.,* 2013), they are challenging to identify. In *C. cedri* the only identifiable AMP was thaumatin, for which four copies existed. Unsurprisingly, this was the only AMP group identified in *A pisum* as well (Gerardo *et al.,* 2010). Regarding its role, it is known that these short peptides have antifungal properties in plant tissue (Shatters *et al.,* 2006), thus they might be involved in controlling fungal infections in aphids as well.

The absence of AMPs in aphids is intriguing and raises the question of how aphids control the proliferation of their endosymbionts and how they cope with pathogen invasions as well. Although there might be additional previously unidentified AMPs which are specific to aphids, it seems more likely that their AMPs repertoire is indeed limited. While they might be employing still not fully elucidated immune responsive genes to contend with infection, it has been suggested that their strategy when exposed to a pathogen is to increase their fecundity in a phenomenon known as fecundity compensation (Barribeau *et al.,* 2010). This phenomenon implicates that the host increases its investment in reproduction in order to maintain fitness regardless of a decreased longevity due to infection. In the case of aphids, having lost an entire pathway would allow saving resources that could otherwise be redirected to fecundity, however the question of how aphids keep their endosymbionts under control remains. Given the lack of AMPs, and immune responsive genes in general, it seems that perhaps their strategy involves controlling the availability of essential metabolites instead of the immune system.

The JAK-STAT signalling pathway is involved in defence against viral infections and we were able to identify all members of the pathway except for unpaired (see Figure V.5). The lack of this gene is a common feature in the genomes of *A. pisum* and *S. oryzae* as well, thus suggesting that unpaired is perhaps a ligand specific of *D. melanogaster* or Diptera. The other set of pathways that could potentially contend with viral infection are the RNA interference pathways. All members of the piRNA pathway were identified on *C. cedri* as well as in *A. pisum*. The main difference is that while *D. melanogaster* has different piwi and aubergine genes (Haase, 2016), *C. cedri* has two copies which seem orthologous to aubergine and *A. pisum* has four. Regarding the siRNA pathway, *C. cedri* contains every member of the pathway with one-to-one orthology to *D. melanogaster* except for trsb and vig (which are also missing in *A. pisum*) and argonaute 2, where five copies were identified on *C. cedri* versus only one on *A. pisum*. Finally, regarding genes involved in the miRNA pathway, only two genes are missing: HPS4 and nibbler. HPS4 is involved in vesicle trafficking and this activity has been linked with efficiency in loading miRNA silencing with mutants of *D. melanogaster* exhibiting enhanced activity (Lee *et al.,* 2009). On the other hand, nibbler codes for an exoribonuclease involved in the 3' end processing of microRNAs enhancing miRNA function (Han *et al.,* 2011). Both genes have antagonistic roles, they are not essential, and they are also missing in *A. pisum*. The fact that virtually every gene involved in the RNA interference pathways is conserved between *C. cedri* and *A. pisum* and even *D. melanogaster* and *S. oryzae* suggest that all three pathways existed in the common ancestor of insects and that they are essential for their survival.

In summary, the immune system pathways of *C. cedri* are highly similar to those of *A. pisum* with only small differences regarding the copy number of some genes. This could implicate that the presence of *Serratia* as a co-obligate endosymbiont in *C. cedri* (and other members of the Lachninae subfamily) is not due to gene losses from the immune system of *C. cedri* given that it appears to be already unable to recognize any Gram-negative bacterium. Instead, the incorporation of *Serratia* seems to have happened to acquire a second healthier bacterium due to the loss of the riboflavin pathway, as already pointed out (Manzano-Marín and Latorre, 2014) not unlike the replacement of *Nardonella* by *S. pierantonius* in *S. oryzae*.

While establishing a clear orthology relationship to the proteases involved in the cascade upstream of the Toll receptor was not possible we suggest that the cascade is fully functional either by using orthologs to the known genes or different ones. As observed before, there are major differences between aphids and other holometabolous insects such as the insect model *D. melanogaster* or the rice weevil *S. oryzae* especially concerning losses in the IMD pathway. Additionally, similarly as observed in *S. oryzae*, there are differences in the number of receptors and the effectors. One of the most striking differences is the apparent lack of antimicrobial peptides besides Thaumatin suggesting that there might be some yet unidentified alternative AMPs specific of aphids.

V.2.5 The third endosymbiont, *Wolbachia*

While the two endosymbiotic partners of *C. cedri* are well studied they share their host with another endosymbiont, namely *Wolbachia* (Gomez-Valero *et al.,* 2004). Although having three endosymbiotic partners is uncommon, *C. cedri* is not the only insect where this happens. This has also been observed in other aphids (Augustinos *et al.,* 2011; Guo *et al.,* 2017), in tsetse flies (from the *Glossina* genus) which harbours *Wigglesworthia* as primary endosymbiont, *Sodalis* as a secondary endosymbiont with the third being *Wolbachia* as well (Beard *et al.,* 1993; Dale *et al.,* 2001).

Wolbachia is an extremely widespread α-proteobacterium carried by most insect species and some mites, crustaceans and nematodes as well. Its relationship with its hosts can range from parasitic to mutualistic and its effects in the host reproduction have been widely studied (Serbus *et al.,* 2008). It is very successful in transmitting itself both horizontally and vertically through the manipulation of the hosts' reproduction. It employs different strategies such as increasing the ratio of females in the offspring through mechanisms such as sperm-egg cytoplasmic incompatibility (Zabalou *et al.,* 2004) or feminization of the genetic males (Asgharian *et al.,* 2014). Despite its high prevalence in arthropods, originally it was thought that aphids did not harbour *Wolbachia*. However, it was demonstrated that *C. cedri* aphids were associated with this bacterium (Gomez-Valero *et al.,* 2004). Additionally, further studies have demonstrated that *Wolbachia* is indeed found in several other populations of aphids besides *C. cedri* (Augustinos *et al.,* 2011). Regarding the infection route, it has been suggested that *Wolbachia* might have been transferred from parasitoids that can also get infected with it (Vavre *et al.,* 1999). The role of *Wolbachia* in aphids, if any, is not known. Moreover, it is difficult to elucidate why a bacterium that usually produces sexual syndromes is infecting insects with a holocyclic and thus complex life cycle.

The genome of *Wolbachia* Cced was assembled into 21 contigs with a total size of 1.35 Mb and a GC content of 30.3%; both are within the range of other species that are in the process of adaptation to intracellular life. The annotation yielded 1,442 features, including 1,403 CDS, 3 rRNA and 36 tRNA. To correctly place *Wolbachia* in the phylogenetic tree of available *Wolbachia* strains the strategy implemented by Ramirez-Puebla *et al.,* (2015) was followed. This confirmed the original hypothesis, which states that *Wolbachia* from *C. cedri* belongs to the B supergroup along with *Wolbachia* from Diptera (*Culex pipiens* *molestus w*Pip, *Culex molestus w*Pip_Mol, *Culex quinquefasciatus w*Pel and *w*Pip_JHB, *Aedes albopictus w*AlbB, *Drosophila simulans w*No), Lepidoptera (*Hypolimnas bolina w*Bol1-b), Hymenoptera (*Nasonia vitripennis w*VitB), Hemiptera (*Dactylopius coccus w*DacB, *Diaphorina citri w*Dia) and Siphonaptera (*Ctenocephalides felis w*Cte) with the closest being that from *D. citri* which is also the closest insect to *C. cedri* (see Figure V.6).

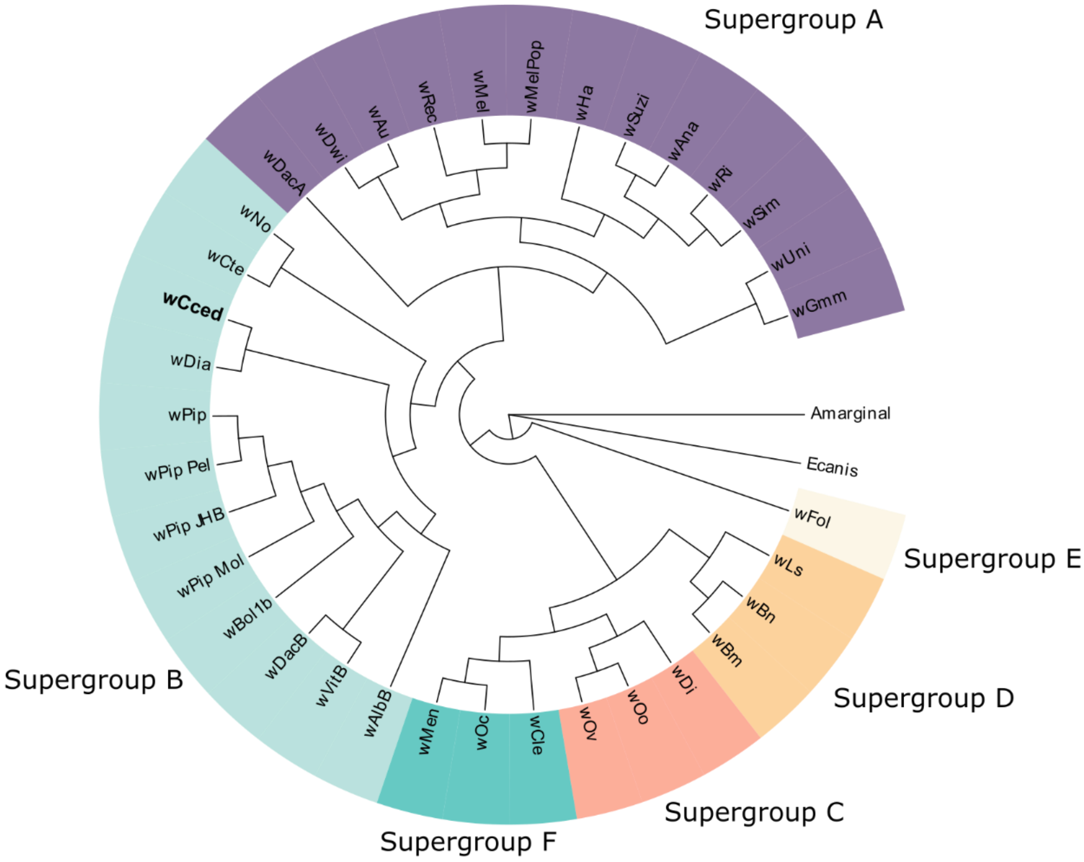

Figure V.6 ML tree using an alignment performed with concatenated sequences for 31 marker genes from Wolbachia as described in Ramirez-Puebla et al., (2015). Six supergroups are observed using this methodology. Hosts for the species used are: Aedes albopictus wAlbB; Brugia malayi wBm; Cimex lectularius wCle; Cinara cedri wCced; Ctenocephalides felis wCte; Culex molestus wPip_Mol; Culex pipiens molestus wPip; Culex quinquefasciatus wPel and wPip_JHB; Dactylopius coccus wDacA and wDacB; Diaphorina citri wDia; Dirofilaria immitis wDi; Drosophila ananassae wAna; Drosophila melanogaster wMel and wMelPop; Drosophila recens wRec; Drosophila simulans Riverside wRi; Drosophila simulans wHa, wSim, wNo and wAu; Drosophila suzukii wSuzi; Drosophila willistoni wDwi; Folsomia candida wFol; Glossina morsitans morsitans wGmm; Hypolimnas bolina wBol1-b; Litomosoides sigmodontis wLs; Mengenilla moldrzyki wMen; Muscidifurax uniraptor wUni; Nasonia vitripennis wVitB; Onchocerca ochengi wOo; Onchocerca volvulus strain Cameroon wOv; Osmia caerulescens wOc; and Wuchereria bancrofti wBn.

The genomes for *Wolbachia w*VitB, *w*Pip_Mol, *w*Pip_JHB, *w*Pip, *w*Dia, *w*Bol1-b and *w*AlbB were used as input to obtain the pangenome for supergroup B along with the newly assembled *w*Cced. The smallest genome was that of *w*VitB (1.1 Mb) and the largest that of *w*Pip_JHB (1.5 Mb). The total number of genes in the pangenome was of 10,602 of which *w*VitB contributed with 966 and *w*Pip_JHB with 1,495. The pangenome was divided in 2,836 groups of which 584 were found in all strains and 1,357 were strain-specific. With 367 strain-specific genes *w*Cced has the highest amount of strain specific genes (see Table V.2), 246 of them correspond to hypothetical proteins, 41 to transposases and 11 to ankyrin repeats. *Wolbachia* genomes from insects are known to harbour a large number of repeated elements multi-gene families such as ankyrin proteins; however, the nature of their role hasn't been elucidated yet. It has been suggested that ankyrin proteins play an important role in symbiotic interactions given that they are known to participate in protein-protein interactions (Siozios *et al.,* 2013).

*w*AlbB	*w*Bol1-b	*w*Cced	*w*Di	*w*Pip	*w*Pip_JHB	*w*Pip_Mol	*w*VitB
141	158	367	191	55	293	45	107

Table V.2 Number of strain-specific genes per strain. Host used are: Aedes albopictus wAlbB; Hypolimnas bolina wBol1-b; Cinara cedri wCced; Dirofilaria immitis wDi; Culex pipiens molestus wPip; Culex quinquefasciatus wPel and wPip_JHB; Culex molestus wPip_Mol; Nasonia vitripennis wVitB.

Regarding the metabolism of amino acids, *Wolbachia* can catalyse eight of the nine reactions necessary to produce lysine from aspartate lacking only the final step, a diaminopimelate decarboxylase, to metabolize meso-2,6-diaminoheptanedioate into lysine; therefore, it has argD, the gene missing in *Buchnera* BCc. While we could hypothesize that it participates in the synthesis of lysine and thus has been conserved for this reason, the fact that argD is missing in several other endosymbionts (Sloan *et al.,* 2014; Santos-*Garcia et al.,* 2017) suggests that other enzymes could replace its activity and that Wolbachia is not needed. Regarding the production of methionine, an enzyme that could participate in the synthesis of methionine is cystathionine beta-lyase from *Wolbachia* which catalyses the conversion of cystathionine into homocysteine. Using this enzyme would not require the other enzymes from the host to operate in the reverse sense but it would require a supply cystathionine. However, given that *Wolbachia* mostly relies on its host for the supply of amino acids (White *et al.,* 2017) it is more likely that the pathways it conserves are involved in other biological processes unrelated to the biosynthesis of amino acids including the synthesis of peptidoglycan in the case of meso-2,6-diaminoheptanedioate.

In summary, while *Wolbachia* is found in all *C. cedri* samples screened to date, its role (if any) is still not clear. It does not seem to participate or interfere with the biosynthesis of amino acids nor has any additional pathways that could prove relevant for the holobiont. It does have the largest number of strain specific genes, but many of them are either involved in transposition or hypothetical proteins. The analysis of its genome revealed that it is massively infected with mobile genetic elements not unlike what was observed in *w*Mel from the A supergroup (Wu *et al.,* 2004). It seems that these mobile elements might have originated from phages and that these phages can perhaps allow transfers between strains in cases where a host is infected by multiple strains (Jamnongluk *et al.,* 2002).

V.3 Conclusions

C. cedri has the third largest genome among the aphids with data available in NCBI; however, all available aphids have a genome size ranging from 300 to 400 Mb except for *A. pisum*, thus its remarkably larger genome seems like an anomaly among aphids.

When compared to *M. persicae* and *A. pisum*, *C. cedri* had the smallest number of genes yet the largest number of lineage specific genes. Additionally, the aphids' superfamily had the largest number of order specific genes among hemipterans, suggesting that they have suffered lineage specific expansions.

The main role of *Buchnera* BCc is the provision of EAA. However, the metabolism of amino acids is coupled between *C. cedri*, *Buchnera* BCc and *S. symbiotica*. While the amino acids produced by both endosymbionts had been previously studied in detail, the genome sequencing has allowed us to specifically define the metabolites provided by *C. cedri*. Serine, alanine, glutamate, glutamine, proline, aspartate, asparagine and tyrosine can be synthesized by the host alone. Cysteine was known to be produced by *S.*

symbiotica, but it can also be produced by the host. Valine, leucine, arginine and isoleucine are provided by a collaboration between *C. cedri* and *Buchnera* BCc. Finally, methionine is either obtained from the diet or produced through a collaboration between *C. cedri* and *Buchnera* BCc.

The immune system pathways of *C. cedri* are highly similar to those of *A. pisum* and *M. persicae* with only small differences regarding the copy number of some genes. As observed before, there are major differences in the IMD pathway between aphids and other holometabolous insects. Additionally, there is an apparent lack of antimicrobial peptides besides Thaumatin.

While *Wolbachia* is found in all *C. cedri* samples screened to date, the elucidation of its role through the sequencing of its genome was not possible. It does not seem to collaborate in the biosynthesis of amino acids nor has any additional pathways which could prove relevant for the holobiont. Many of its strain-specific genes are either involved in transposition or hypothetical proteins. It seems that these mobile elements might have originated from phages.

VI CHAPTER 3: COMPARISONS BETWEEN BOTH MODELS

VI.1 Introduction

While we are mainly interested in the differences between both models due to the different age of the relationship between each host and its endosymbiont, it is evident that there are many other factors that should be considered. To consider the environment, the diet, the taxonomic status, and whether they are holo- or hemimetabolous we included 19 other arthropods with fully sequenced genomes (see Table S.1). The selected outgroups were one crustacean, the water flea *Daphnia pulex* (Colbourne *et al.,* 2011), and one arachnid, the two-spotted spider mite *Tetranychus urticae* (Grbić *et al.,* 2011). While several arthropods harbour *Wolbachia* we will only take it into account when it is known to be essential.

Regarding hemimetabolous insects we included a selection of hemipterans: (i) the cedar aphid *C. cedri* which harbours *Buchnera* BCc and *S. symbiotica*; the pea aphid *A. pisum* (The International Aphid Genomics Consortium, 2010) and the green peach aphid *Myzus persicae* (Mathers *et al.,* 2017) which harbour *Buchnera* (APS and F009 respectively); and the Asian citrus psyllid *D. citri* (Hunter *et al.,* 2014), which harbours *Candidatus* Carsonella ruddii DC, all four belong to the Sternorrhyncha suborder and are phloem feeders; (ii) the brown planthopper *N. lugens* (Xue *et al.,* 2014) from the Auchenorrhyncha suborder, which harbours a yeast-like symbiont (YLS) and is also a phloem feeder; (iii) the bedbug *Cimex lectularius* (Rosenfeld *et al.,* 2016) which harbours the gram-negative *Wolbachia* wCle and feeds on blood; and (iv) the louse *Pediculus humanus* (Kirkness *et al.,* 2010) which harbours *Candidatus* Riesia pediculicola, a gram-negative bacterium, and feeds on blood as well.

Given the vast diversity of holometabolous insects and that they comprise most of the diversity of their class, we selected four species from the four largest classes. From the Hymenoptera: (i) the Florida carpenter ant *Camponotus* *floridanus* (Bonasio *et al.,* 2010); and the red imported fire ant *Solenopsis invicta* (Wurm *et al.,* 2011)*,* both of which are omnivores capable of exploiting several food sources ranging from seeds and honeydew to other animals; (ii) the western honeybee *Apis mellifera* (The Honeybee Genome Sequencing Consortium, 2006), which feeds solely on pollen and nectar; and (iii) the jewel wasp *Nasonia vitripennis* (Werren *et al.,* 2010), which is a parasitoid of the larvae of several parasitic carrion flies. Among them, only *C. floridanus* harbours an endosymbiont, the gram negative *Blochmania floridanus* (Gil *et al.,* 2003).

From the Diptera we chose: (i) two members of the Culicidae family, the yellow fever mosquito *Aedes aegypti* (Nene *et al.,* 2007) and *Anopheles gambiae* (Holt *et al.,* 2002), which during their larval stage spend most of their time feeding on algae, bacteria, and other microorganisms and during their adult stage they feed on nectar and other sugar sources and additionally, the females feed on blood to allow the development of their eggs; (ii) the fruit fly *D. melanogaster* (Adams *et al.,* 2000), which feeds on decaying plant matter; and (iii) the tsetse fly *G. morsitans* (International Glossina Genome Initiative, 2014), which solely feeds on the blood of vertebrates. Only the tsetse fly harbours endosymbionts, the obligate *Wigglesworthia glossinidia* and the facultative *Sodalis glossinidius*.

From the Lepidoptera four species were selected as well: the silkworm *Bombyx mori* (Mita *et al.,* 2004), the tobacco hornworm *Manduca sexta* (Kanost *et al.,* 2016), the monarch butterfly *Danaus plexippus* (Zhan *et al.,* 2011) and the diamondback moth *Plutella xylostella* (You *et al.,* 2013). All of them feed on the leaves of several plants during their larval stage and feed on nectar, if anything at all, during their adult stage. To our knowledge no endosymbiont has been found associated to lepidopterans.

Finally, from the large Coleoptera order we selected: (i) the emerald ash borer *Agrilus planipennis*, the Asian longhorned beetle *A. glabripennis* (McKenna *et al.,* 2016) and the mountain pine beetle *D. ponderosae* (Keeling *et al.,* 2013), which feed on the inner bark and phloem of several species of trees; (ii) the small hive beetle *Aethina tumida*, which feeds on honey, pollen and bee brood; (iii) the rice weevil *S. oryzae* and the red flour beetle *T. castaneum* (Tribolium

Genome Sequencing Consortium, 2008) a primary and secondary pest respectively of stored grains; and (iv) *Nicrophorus vespilloides* (Cunningham *et al.,* 2015) a burying beetle which feeds on carrion. The only beetle with an endosymbiont in our set is *S. oryzae*.

VI.2 Results and discussion

VI.2.1 Gene orthology

We used EGGNOG-MAPPER to assess the conservation degree of protein-coding genes between our two models and 26 other arthropod species to assess their orthology relationships. We categorized these set of genes in several groups according to their characteristics. The first group is the one2one set which includes single copy genes found in at least 25 species. The second group is many2many which encompasses genes found in at least 25 species but without having to be single copy genes in every species. The insect specific group, as its name implies, includes genes found in every insect species and not in the two outgroups. Order specific genes include genes that are particular to a given order and not found in any other species. Lineage specific genes include genes that apparently have no orthologs in any other species and finally the some orthology group encompasses all genes which did not fit in any of the aforementioned groups.

On average, there are 1,327 genes with a one2one orthologous relationship per species (see Figure VI.1); additionally, 1,478 genes have a many2many relationship which amounts to a total of 2,805 genes with orthologs in at least 25 of the other 27 species. Thus, both groups comprise the core genome of arthropods. Regarding insects, there is an additional set of 2,578 genes in average which are found in all insects and not in the two outgroups.

Concerning order specific orthologs, hemipterans have 2,340 genes with orthologs only found on members of this order, however this distribution is not homogenous, both *A. pisum* and *M. persicae* have more than 4,600 genes belonging to this group while *C. lectularius* has just over 420 genes in this group. On average, there are 1,356 genes per coleopteran which are found only in their order. The distribution on this group is more homogenous, however *A.*

planipennis only has 784 genes in this category. The fact that it is the most ancestral beetle in our study suggests that perhaps the ancestor of all other lineages experienced expansions after it diverged from the ancestor of *A. planipennis*. Dipterans have 1,822 genes in this group on average; however, both mosquitoes have more genes in this category than flies. Lepidopterans have 3,106 order specific genes, the largest amount of all holometabolous insects in our analysis and with a homogeneous distribution. Finally, hymenopterans have 1,243 order specific genes, with both ants having more genes in this group.

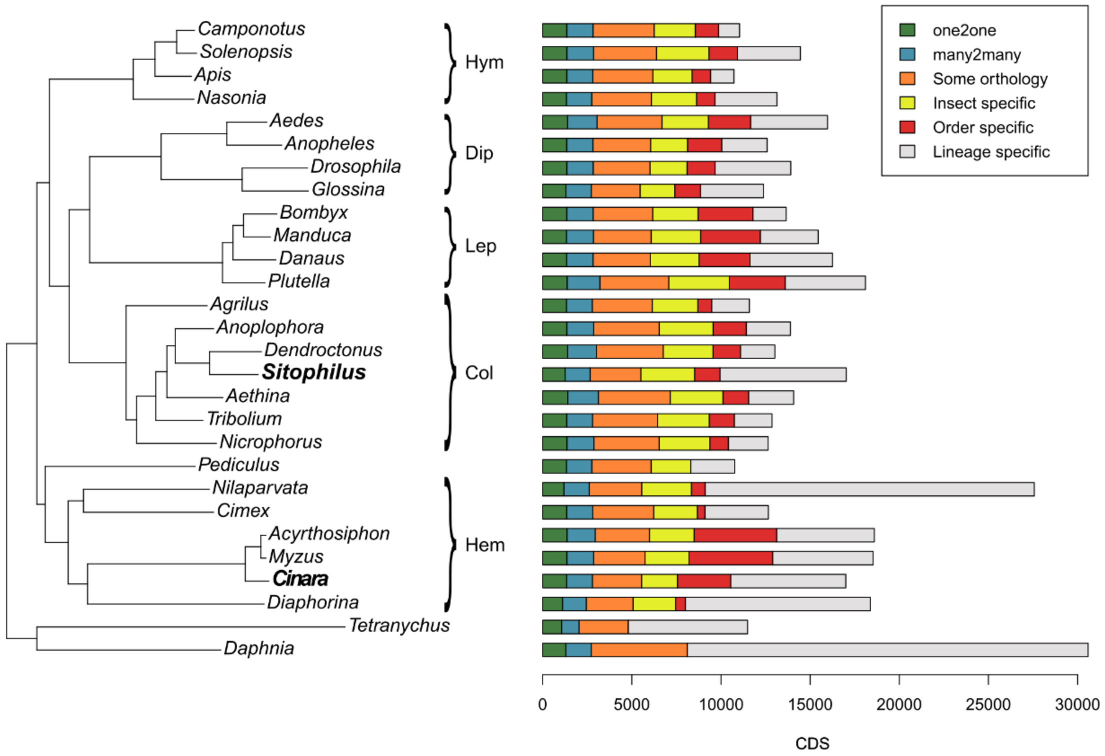

Figure VI.1 Phylogenetic relationships of the orthologs of 28 genomes. The one2one category refers to orthologs with only one copy in at least 25 of the species used. The many2many category includes orthologs also found in at least 25 of the species used but with more than one copy in at least one of the species. Some orthology refers to orthologs found in less than 25 of the species used. Insect, family and lineage specific refers to orthologs found in all the species belonging to that group. To group species included in each order curly brackets were used: Hym stands for Hymenoptera, Dip for Diptera, Lep for Lepidoptera, Col for Coleoptera and Hem to Hemiptera.

Regarding lineage specific genes, both *D. pulex* and *N. lugens* have a huge number of genes belonging to this category, 73% and 67% of their total number of genes respectively (see Figure VI.1). Given that we can't compare the water flea with other crustaceans we can't say whether this is a fact specific to *D. pulex* or if all crustaceans behave similarly. Conversely, *N. lugens* represents quite an anomaly given its large gene repertoire and the large number of genes which are specific to this species and the fact that other insects have a similar environment and diet as *N. lugens*. Finally, there is a group of 3,365 genes on average per specie which have an ortholog in another species, but which do not fit in any of the previous categories.

VI.2.2 Identification of genes involved in amino acid biosynthesis

The insects we selected for this analysis live in very different niches and must face numerous types of stress, including consuming all the nutrients they need. While we will focus on amino acid biosynthesis, the production of vitamins is another essential role of nutritional endosymbionts. The conservation of pathways for the biosynthesis of vitamins has a patchy distribution highly dependent on the needs of their host (Serbus *et al.,* 2017). This also holds true for the biosynthesis of amino acids given that while most insects are incapable of synthesizing the ten EAAs, there are some interesting differences.

Regarding the amino acids requiring α-ketoglutarate, we were able to identify the aminotransferases that are proposed to synthesize glutamate in all the evaluated arthropods and in *S. pierantonius*, *B. floridanus*, *C.* Carsonella ruddii DC and YLS. The prevalence of this enzyme highlights its relevance in the biosynthesis of amino acids given that the newly created amino group can then be used to produce other amino acids through other transamination reactions. The pathways to synthesize proline were also found in all arthropods and in YLS. Glutamine can also be produced by all arthropods and by *S. pierantonius* and *B. floridanus*, thus seeming to be dispensable in most endosymbionts. Arginine cannot be produced by arthropods; however, some endosymbionts can supply its host with the amino acid, namely *S. pierantonius* and *Buchnera* from the strains APS and F009 (see Figure VI.2 and Table S.2). *Buchnera* Cce and *C.* Carsonella ruddii DC can also produce arginine if supplied with ornithine. Interestingly, lepidopterans have an ornithine carbamoyltransferase which is missing in all other evaluated arthropods, thus potentially allowing them to produce arginine from ornithine instead of from citrulline as other insects. This enzyme is part of the urea cycle, which is absent from most insects, however the degree of distribution of this enzyme among arthropods is not known. While it is conserved in other animals this might represent a transfer given its similarity with the sequence of prokaryotes, nonetheless it is currently impossible to determine its origin.

On the amino acids that use oxaloacetate as a precursor, all arthropods can produce aspartate. Additionally, *S. pierantonius*, *W. glossinidia*, *Wolbachia* wCle and *B. floridanus* are also capable of producing it. All arthropods are also capable of producing asparagine; however, all *Buchnera*, *C.* Carsonella ruddii DC, *S. symbiotica* and *B. floridanus* are unable to produce it and thus rely on their hosts for the obtention of this amino acid. Notably, all but *B. floridanus* are endosymbionts of sap-sucking insects and thus this loss might be related to the abundance of asparagine in phloem. Also, although *C. floridanus* is omnivorous, an important source of its diet is the honeydew produced by sap-feeding insects (Sauer *et al.,* 2002) and thus, this might be another cause of nutritional stress.

Neither lysine, methionine, isoleucine or threonine can be synthesized by any of the arthropods evaluated; however, lysine can be produced by all *Buchnera*, *S. pierantonius*, *B. floridanus*, *C.* Carsonella ruddii DC and YLS. While many endosymbionts lack argD, an acetylornithine aminotransferase, it has been suggested that another aminotransferase can substitute this enzyme. Thus, this pathway is only missing from the genomes of endosymbionts from blood-feeders rendering them dependent on their food source for the obtention of this amino acid. The pathways for the biosynthesis of methionine are complete only in *B. floridanus* and YLS, nonetheless if the mechanisms that were suggested in the previous chapter for the biosynthesis of methionine in *Buchnera* indeed occur they could possibly also work in *S. pierantonius* and *C.* Carsonella ruddii DC. Isoleucine can only be synthesized by *B. floridanus*, *C.* Carsonella ruddii DC, YLS and *Buchnera*, but interestingly *Buchnera* lacks the last step which is an aminotransferase found in arthropods, thus allowing its host to have a higher degree of control in the production of this amino acid (see Figure VI.2).

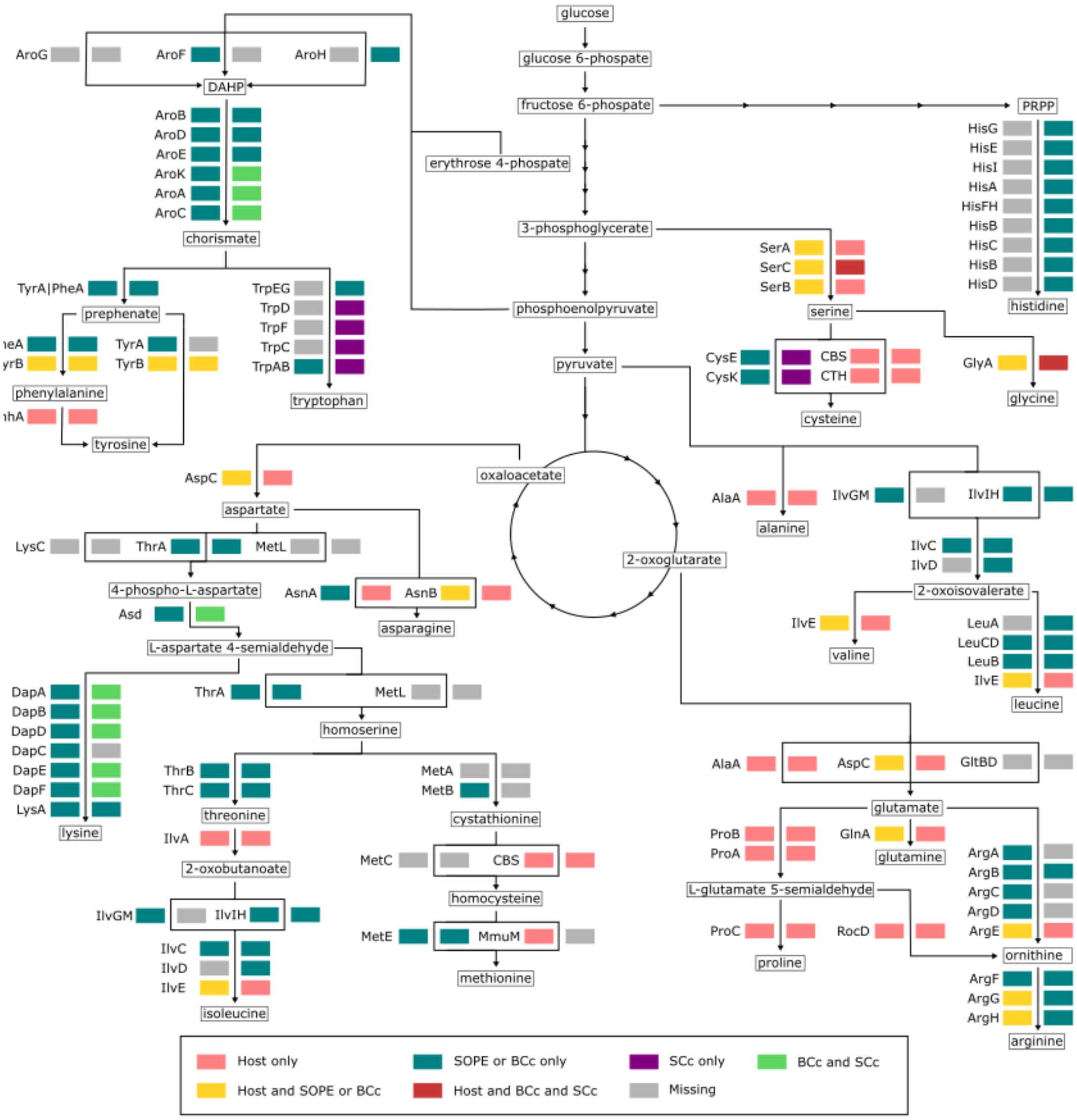

Figure VI.2 Amino acids biosynthesis pathway reconstruction for both the S. oryzae and S. pierantonius holobionts (on the left) and C. cedri, Buchnera and S. symbiotica (on the right).

Threonine can be produced by *Buchnera*, YLS, *S. pierantonius* and *B. floridanus*. *C.* Carsonella ruddii DC seems to lack thrB, homoserine kinase, which is needed to produce threonine seems to be absent such as in the *C*. Carsonella ruddi from *Pachpsylla venusta* (Tamames *et al.,* 2007) and therefore either another enzyme is substituting ThrB or there is another threonine source in the holobiont. Notably, lepidopterans have a threonine synthase (thrC) which is not found in any other arthropod. This enzyme catalyses the last step gene of the threonine biosynthesis pathway and it appears to have been transferred from an alpha protobacterium. Interestingly, to our knowledge there is no lepidopteran with an endosymbiont, thus either this transfer occurred thanks to a now lost endosymbiont or it has a non-endosymbiotic origin.

Serine, glycine and cysteine, the three amino acids that use 3-phosphoglycerate for their carbon backbone, can be synthesized by all arthropods. Regarding the endosymbionts, only *S. pierantonius* and YLS can produce serine. Glycine is produced by all endosymbionts, being the only amino acid produced by all the evaluated organisms. *B. floridanus*, YLS, *S. pierantonius*, *S. symbiotica* and *Buchnera* APS and F009 can produce cysteine. Of the amino acids that use pyruvate as their precursor, alanine can be produced by all evaluated arthropods. Conversely, the ability to produce it has been lost from many of the endosymbionts, being only conserved in YLS, which is the only eukaryotic endosymbiont. Therefore, the pressure to conserve the enzyme that produces alanine does not seem to have been removed from YLS. Regarding valine and leucine YLS, *B. floridanus* and *C.* Carsonella ruddii DC can produce them, while *Buchnera* is capable of catalysing all the needed reactions except for the last one (see Figure VI.2), comparable to what was described previously for isoleucine.

In the case of the aromatic amino acids, no arthropod is capable of producing them and therefore they must rely on their diet or on their endosymbionts to produce them at adequate rates. *W. glossinidia* is only capable of producing chorismate, the precursor of aromatic amino acids. *C.* Carsonella ruddii DC might be capable of producing tryptophan, although it seems to lack aroE a shikimate dehydrogenase, and the other two must be obtained elsewhere. *B. floridanus* is only missing the last step to produce both tyrosine and phenylalanine, but this enzyme is found in *C. floridanus* so all three amino acids can be produced by the holobiont. YLS and *Buchnera* APS and F009 can produce tryptophan and phenylalanine which can then be exported to their hosts to produce tyrosine while *S. pierantonius* can produce both phenylalanine and tyrosine. *Buchnera* Cce can only produce phenylalanine on its own and as was mentioned before it requires the cooperation of *S. symbiotica* for the biosynthesis of tryptophan. Finally, regarding histidine, *Buchnera*, YLS and *B. floridanus* contain all the needed genes and are thus capable of producing it. *C.* Carsonella ruddii DC lacks only one gene, hisB, and while it has been suggested that it is unable to produce histidine (Tamames *et al.,* 2007), the fact that the rest of the pathway is conserved could perhaps mean that the role of hisB is being replaced by another enzyme.

VI.2.3 Identification of genes involved in immune system pathways

As aforementioned, the insect's innate immune system is an ancestral mechanism involved with the defence of the host against a wide variety of pathogens. Therefore, we expect its main components to be shared between both *C. cedri* and *S. oryzae* with some differences reflecting their biological needs. Broadly we had observed that *C. cedri* lacked many components of the IMD pathway, possibly even rendering it non-functional. Through its comparison with *A. pisum* and *M. persicae* we observed that it seemed to be a common trend between aphids. To define whether this is a trend specific to aphids or more general we took advantage of the wide variety of insects included for the identification of the orthologous genes.

A matrix of the orthologous genes for members of the innate immune system pathways for the 28 arthropods was generated including 106 genes of interest belonging to miRNA, siRNA, piRNA, IMD, Toll, JNK and JAK/STAT pathways (see Table S.3). This matrix allowed us to easily observe differences and similarities between the insects included; nonetheless it is important to keep in mind that using a homology-based approach would only identify elements known in the reference genomes, thus missing species-specific novel genes or even pathways. Additionally, homologous genes might have diverged and acquired different functions according to the needs of the organism. Also, having a similar copy number of a given gene in different organisms does not necessarily mean that the immune response will be comparable given that there could be key differences in the sequences of the genes themselves or in their regulatory regions.

The first observable feature is that AMPs are highly order- or even lineage-specific. While most holometabolous insects have members of all the assessed families, in the case of hemimetabolous insects only defensins and thaumatins were identified. Furthermore, in non-insect arthropods none of the assessed AMPs were identified. However, extracting conclusions from these facts is complicated given that it is known that identifying AMPs by mere homology can be challenging because short sequences can rapidly lose signals of homology due to the high sequence divergence in this large peptide family. Additionally, it is known that genome annotation pipelines generally ignore small peptides (Plaza *et al.,* 2017), thus masking the still unexplored abundance of AMP. A study by Mylonakis and colleagues (2016), which elegantly described the phylogenetic origin of the most widespread AMP families demonstrated that most of the known AMPs belong to the best studied orders (see Figure VI.3). Thus, there is surely many AMPs waiting to be discovered, nonetheless, identifying them and experimentally validating them will certainly be challenging. Currently a great effort it being made to improve identification of AMPs given their possible usage as therapeutic agents as a novel strategy to cope with an increasing resistance toward conventional antibiotics (Mahlapuu *et al.,* 2016).

Regarding antiviral RNAi, we can see that all arthropods have seemingly fully functional pathways with only some order- or species-specific losses. From the miRNA pathway HPS4 and Nbr are missing from most non-holometabolous insects including *C. cedri*. As aforementioned, both genes have antagonistic roles involved with the loading of the miRNA in Ago1 and they are not essential. Given the fact that *P. humanus* has both genes it seems more likely that they have been lost from hemipterans rather than recruited in holometabolous insects.

Concerning the piRNA pathway, essential elements were identified in all explored arthropods with the major difference being the absence of Yb in some insects. This could be explained by the fact that Yb belongs to the class of the Tudor-domain-containing proteins (TDRDs) which are known to play crucial roles in the piRNA pathway (Ku and Lin, 2014). Therefore, given that there are several known TDRDs perhaps another set has taken the role of Yb in other arthropods. Another difference is that while in *D. melanogaster* there are two copies for Piwi/Aub, in most other insects there is a single copy. It has been suggested that two copies of Piwi/Aub and the Ago3 genes were present in the last common ancestor of insects (Dowling *et al.,* 2017).

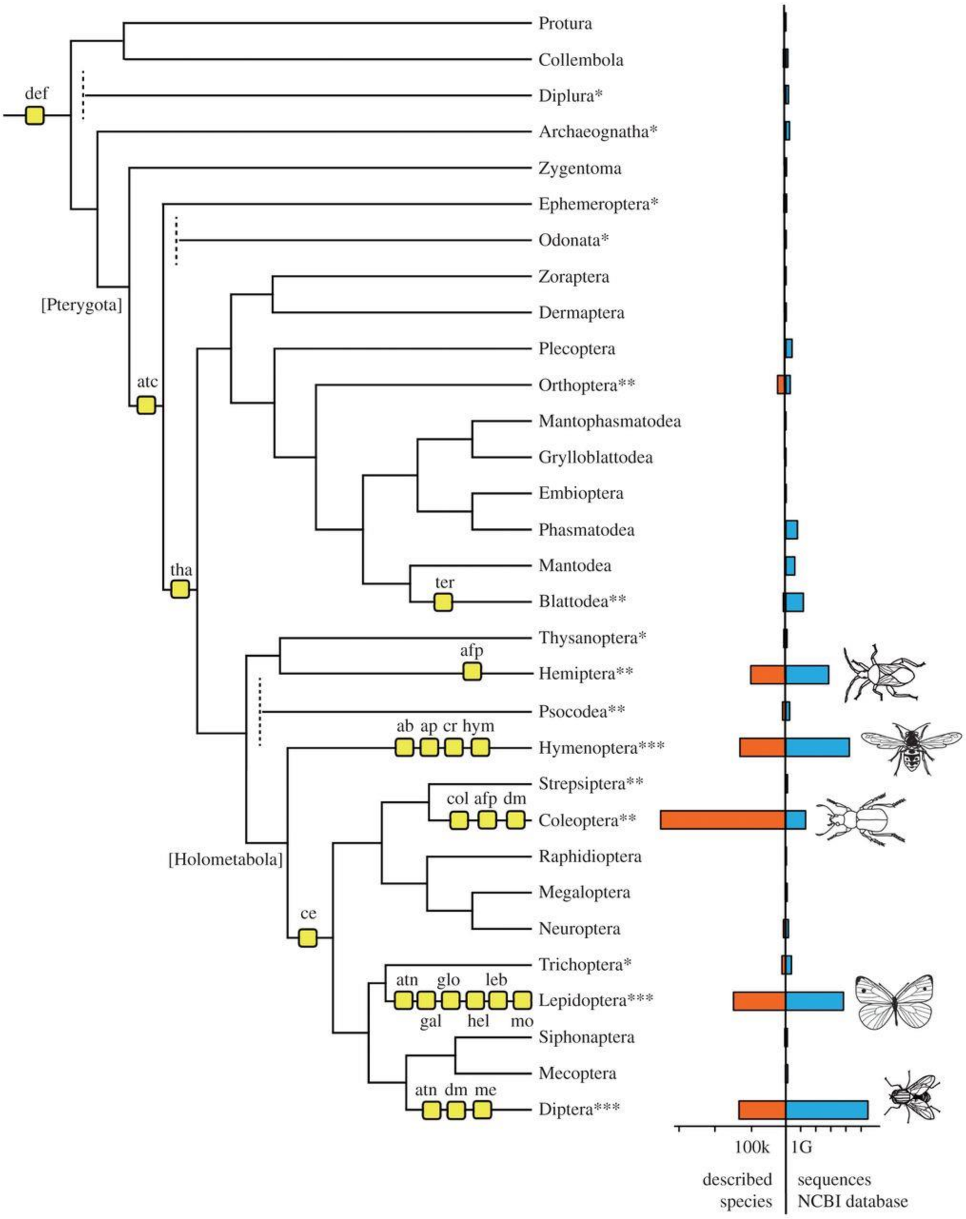

Figure VI.3 Phylogeny of insects depicting the major groups. Asterisks indicate taxa with available genome projects (one asterisk: data available in the database of the 15k pilot project, two asterisks: one to three published projects and three asterisks: more than three published genome projects). The yellow boxes indicate the hypothesized origin of each AMP family. The bar plot to the left shows the number of described species while the one to the right shows the amount of sequences in the NCBI database. Abbreviations: ab, abaecin; afp, antifungal protein; ap; apidaecin; atc, attacin C-terminal domain; atn, attacin N-terminal domain; ce, cecropin; col, coleoptericin; cr, crustin; def, defensin; dm, drosomycin; gal, gallerimycin; glov, gloverin; hel, heliomicin; hym, hymenoptaecin; leb, lebocin; mor, moricin; ter, termicin; tha, thaumatin. From Mylonakis, Eleftherios, Lars Podsiadlowski, Maged Muhammed, and Andreas Vilcinskas. 2016. "Diversity, Evolution and Medical Applications of Insect Antimicrobial Peptides." Philosophical Transactions of the Royal Society B: Biological Sciences 371 (1695).

All key members of the siRNA pathway were identified in all insects. However, all aphids are missing two genes: trsn and vig. Vig is known to be a component of the RISC complex, however its role has not been described. Trsn is an endonuclease also likely involved with siRNA, but its role has not been well described either; regardless, the siRNA pathway is known to be functional in aphids (Mutti *et al.,* 2006). Also, lepidopterans do not have a homolog for r2d2, but it has been suggested that r3d1 could replace it (Cao *et al.,* 2015).

The JAK/STAT pathway was highly conserved, with core components in all insects. The main difference is that while the ligand for dome has been identified in *D. melanogaster* it has not been identified on any other insect. This trend is shared by the Toll pathway, which is also highly conserved among all the compared insects. The main difficulty in both *S. oryzae* and *C. cedri* (and most other insects for that matter) was to identify a clear homologue to the upstream proteases involved in the processing of späetzle. Grass, nec, spheroide and spirit were not identified in most orders besides Diptera. Additionally, SPE, ModSP and psh were not identified in most hemimetabolous insects. However, this does not mean that those proteases do not exist in other insects, it just highlights the challenges for establishing a clear one to one relationship between genes of large families. The absence of ref(2)P seems to be a feature shared by most coleoptera and aphids. The fact that it was only found in *A. planipennis*, the sister group to all other Coleoptera, could suggest that it was lost after these taxa diverged.

Perhaps the most relevant feature is the high variability observed in the IMD pathway. As aforementioned, aphids are missing CYLD and we can observe that this is the only group lacking this gene; additionally, ird5, imd, relish and tab2 apparently are missing from Sternorrhyncha. Dredd, Fadd, key and pirk seems to be missing in all Hemiptera. Thus, while holometabolous insects have a conserved IMD pathway, there seems to be an absence of canonical components of the IMD pathway among hemimetabolous insects. This patchy distribution of the IMD pathway genes has been observed throughout the Arthropoda (see Figure VI.4) so it doesn’t seem to be a feature unique to hemimetabolous insects.

It has been suggested that elements from a gut remodelling network could have been recruited to participate in the immune system given that the IMD pathway is involved in the intestinal immune response of dipterans (Georgel *et al.,* 2001;

Zumaya-Estrada *et al.,* 2018). During metamorphosis there is an extensive intestinal remodelling requiring apoptosis and it is known that the IMD pathway can promote it through the activation of reaper (White *et al.,* 1996; Jiang *et al.,* 1997); in addition to this, at this stage potentially pathogenic bacteria could be released, thus controlling both processes at the same time could be advantageous. While Zumaya-Estrada and colleagues suggested that this recruitment might have occurred after the holometabolous lineage split, we consider this recruitment was earlier during evolution of insects given that the American cockroach *Periplaneta americana* (Li *et al.,* 2018) and the migratory locust *Locusta migratoria* (Wang *et al.,* 2014) have fully functional IMD pathways.

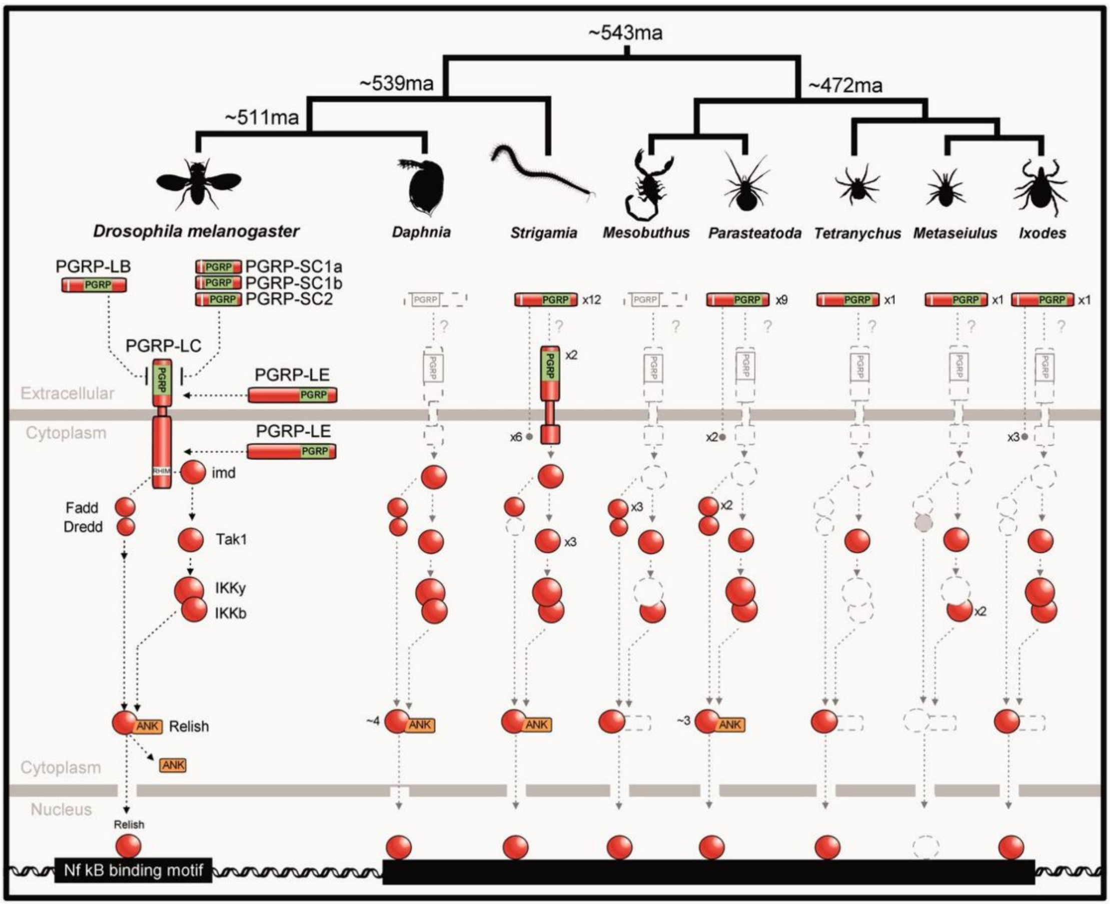

Figure VI.4 Presence or absence of IMD canonical component across representative arthropods obtained through comparative genomics. From Palmer, William J., and Francis M. Jiggins. 2015. "Comparative Genomics Reveals the Origins and Diversity of Arthropod Immune Systems." Molecular Biology and Evolution 32 (8): 2111–29.

In summary, the fact that most hemipteran insects thrive on largely sterile diets (phloem, xylem or blood) might have led to a reduced microbial load in their gut, thus reducing the pressure for conserving the costly IMD pathway. Additionally, as aforementioned these diets are poor, thus most hemipterans have established nutritional symbiotic relationships with microbes, most of them Gram-negative bacteria. Having a reduced IMD pathway made the hemipterans ideal hosts for Gram-negative bacteria given the ease to invade them. On the other hand, the Toll, JAK-STAT and RNAi pathways are highly conserved among all insects.

VI.3 Conclusions

On average, there are of 2,805 genes in each species with orthologs in at least 25 of the other 27 species, thus representing the core genome of arthropods. Taking only insects into consideration, there is an additional set of 2,578 genes thus setting number of genes in the insect core genome at around 5,382. Regarding order specific orthologs, hemipterans on average have 2,340 genes with orthologs only found on members of this order however the numbers in this other vary greatly; on the other hand, there are 1,356 coleopteran-specific genes with a more homogeneous distribution.

Regarding the metabolism of amino acids, all arthropods have a similar gene repertoire, thus supporting the hypothesis of a great-loss at the origin of metazoans; however, there are some relevant differences when considering the holobionts: (i) Arginine cannot be produced by arthropods; however, some endosymbionts can supply its host with the amino acid. Interestingly, lepidopterans have an ornithine carbamoyltransferase which potentially allows them to produce arginine from ornithine. (ii) Lysine can be produced by all evaluated endosymbionts except those from blood-feeders. (iii) Alanine can be produced by all evaluated arthropods; conversely, this ability has been lost from many of the endosymbionts, being only conserved in YLS, thus it seems that there is a selective pressure on eukaryotes for the production of this amino acid. (iv) *S. pierantonius* can produce both phenylalanine and tyrosine while *Buchnera* Cce can only produce phenylalanine on its own and requires the cooperation of *S. symbiotica* for the biosynthesis of tryptophan.

AMPs are highly order- or even lineage-specific. In the case of hemimetabolous insects only defensins and thaumatins were identified. Regarding antiviral RNAi, we can see that all arthropods have seemingly fully functional pathways with only some order- or species-specific losses. This is also the case for the JAK/STAT pathway, where the main difference is that the ligand for dome has not been identified on any other insect besides *D. melanogaster*. The Toll pathway is also highly conserved among all the compared insects; however, identifying the upstream proteases involved in the processing of spaetzle is not trivial. The absence of ref(2)P seems to be a feature shared by most coleoptera and aphids suggesting that in these species activation of Dif/Dorsal regulated genes would be dependent of cactus degradation.

Perhaps the most relevant feature is the high variability observed in the IMD pathway. While aphids are missing the largest number of elements of this pathway, all Hemiptera are missing members. Thus, while holometabolous insects have a conserved IMD pathway, there seems to be an absence of canonical components of the IMD pathway among hemimetabolous insects. This would impair the ability of hemipterans to recognize invading Gram-negative bacteria. While Zumaya-Estrada and colleagues suggested that the recruitment of the IMD pathway occurred after the holometabolous lineage split, we consider it happened earlier during evolution of insects given that the American cockroach, among other hemimetabolous insects, have fully functional IMD pathways.

VII GENERAL CONCLUSIONS

The repeatome in *S. oryzae* is among the largest found in insects and the largest of any studied Coleoptera. Additionally, *S. oryzae* had the fastest gene family expansion rate among the beetles evaluated and some of these families included those potentially involved on insecticide resistance. This phenomenon could potentially be linked to the large number of mobile elements.

An interdependence between the host and the endosymbiont is observed when analysing the joint metabolism of the rice weevil and its endosymbiont. Some of the EAAs are provided by the host, others by the endosymbiont and some must be obtained from the diet. A likely advantage *S. pierantonius* had over the ancestral endosymbiont *Nardonella* was its ability to synthesize more amino acids, the most important might have been lysine and threonine given the low abundance of these amino acids in cereals.

The immune system pathways of *S. oryzae* are highly conserved when compared to the models *T. castaneum* and *D. melanogaster*. One of the most striking differences is that *S. oryzae* lacks a PGRP-LE receptor which might partly explain how weevils tolerate their endosymbiont. Additionally, both novel and already known AMPs were identified. The fact that the immune system of *S. oryzae* is so similar to that of other non-symbiotic insects suggests that small modifications perhaps are enough to allow harbouring an intracellular pathogen, such as the adaptation of colA to control the growth and localization of *S. pierantonius* and perhaps the loss of PGRP-LE.

No events of HGT that seemed to have originated from *S. pierantonius* were identified in the genome of *S. oryzae*. This suggests that perhaps *S. oryzae* has an efficient system for controlling the incorporation of foreign DNA or there just hasn’t been enough time given the recent origin of the association.

The metabolism of amino acids is coupled between *C. cedri*, *Buchnera* BCc and *S. symbiotica*. Serine, alanine, glutamate, glutamine, proline, aspartate, asparagine and tyrosine can be synthesized by the host alone. Cysteine was known to be produced by *S. symbiotica*, but it can also be produced by the host. Valine, leucine, arginine and isoleucine are provided by a collaboration between *C. cedri* and *Buchnera* BCc. Finally, methionine is either obtained from the diet or produced through a collaboration between *C. cedri* and *Buchnera* BCc.

The immune system pathways of *C. cedri* are highly similar to those of *A. pisum* with only small differences regarding the copy number of some genes. As observed before, there are major differences in the IMD pathway between aphids and other holometabolous insects. Additionally, there is an apparent lack of antimicrobial peptides besides Thaumatin.

While *Wolbachia* is found in all *C. cedri* samples screened to date, the elucidation of its role through the sequencing of its genome was not possible. It does not seem to collaborate in the biosynthesis of amino acids nor has any additional pathways which could prove relevant for the holobiont.

The aphids' superfamily had the largest number of order specific genes among hemipterans, suggesting that they have suffered lineage specific expansions. On the other hand, coleopteran-specific genes have a more homogeneous distribution.

Regarding the metabolism of amino acids, all arthropods have a similar gene repertoire, thus supporting the hypothesis of a great-loss at the origin of metazoans. However, (i) Arginine cannot be produced by arthropods but some endosymbionts can supply its host with the amino acid. Interestingly, lepidopterans have an ornithine carbamoyltransferase which potentially allows them to produce arginine from ornithine; (ii) Lysine can be produced by all evaluated endosymbionts except those from blood-feeders; (iii) Alanine can be produced by all evaluated arthropods; conversely, this ability has been lost from many of the endosymbionts, being only conserved in YLS, thus it seems that there is a selective pressure on eukaryotes for the production of this amino acid. (iv) *S. pierantonius* can produce both phenylalanine and tyrosine while *Buchnera* Cce can only produce phenylalanine on its own and requires the cooperation of *S. symbiotica* for the biosynthesis of tryptophan.

AMPs are highly order- or even lineage-specific. In the case of hemimetabolous insects only defensins and thaumatins were identified. Regarding antiviral RNAi, we can see that all arthropods have seemingly fully functional pathways with only some order- or species-specific losses. This is also the case for the JAK/STAT pathway, where the main difference is that the ligand for dome has not been identified on any other insect besides *D. melanogaster*. The Toll pathway is also highly conserved among all the compared insects. The absence of ref(2)P seems to be a feature shared by most coleoptera and aphids suggesting that in these species activation of Dif/Dorsal regulated genes would be entirely dependent of cactus degradation.

While aphids are missing the largest number of elements of the IMD pathway, all Hemiptera are missing some members. Thus, while holometabolous insects have a conserved IMD pathway, there seems to be an absence of canonical components of the IMD pathway among hemimetabolous insects. This would impair the ability of hemipterans to recognize invading Gram-negative bacteria.

The differences between the immune system and the synthesis of amino acids in *C. cedri* and *S. oryzae* seem to be mostly due to their taxonomic position and not due to the differences in the age of their endosymbiotic relationships.

RESUMEN EN ESPAÑOL

Introducción

Los insectos se pueden encontrar en la mayoría de los ecosistemas del planeta y son capaces de aprovechar prácticamente cualquier fuente de alimento. Son el grupo de animales más diverso con un número de especies estimado en cinco millones. Todos los insectos comparten un mismo plan corporal que consiste en tres segmentos (cabeza, tórax y abdomen), tres pares de patas, un par de antenas y ojos compuestos; sin embargo, cada especie tiene partes especializadas acordes con su estilo de vida.

A pesar de que los insectos son capaces de aprovechar múltiples fuentes de alimento, en múltiples ocasiones dependen de compañeros microbianos para obtener nutrientes escasos en su dieta. Por ello, la simbiosis juega un papel crucial en el desarrollo y la evolución de este grupo de animales. La unidad ecológica producto de estas asociaciones se denomina holobionte y los genomas de los participantes en la relación se conocen como el hologenoma. Estos hologenomas son propensos a sufrir pérdidas de genes duplicados entre los participantes de la interacción y también pueden sufrir transferencias de genes de un genoma al otro.

En muchos casos estas interacciones también llevan al desarrollo de órganos especializados y mecanismos novedosos para regular el crecimiento de sus endosimbiontes. Además de mantener bajo control a sus endosimbiontes, los insectos deben de ser capaces de enfrentarse a numerosas infecciones. Para ello cuentan con una primera barrera que es su exoesqueleto que los protege de la invasión de la gran mayoría de los patógenos. Sin embargo, cuando esta protección falla los patógenos se enfrentan a otra barrera: el sistema inmune. A pesar de que los insectos únicamente cuentan con un sistema inmune innato, este es capaz de proteger a los insectos contra la mayoría de las infecciones. Sin embargo, un fenómeno poco estudiado es cómo se regula el sistema inmune de los insectos para protegerlos contra patógenos, pero tolerar a sus endosimbiontes.

Entre los insectos no cabe duda qué los coleópteros son el orden más exitoso con un número estimado de 400,000 especies descritas. El éxito de este grupo se debe a numerosos factores, entre ellos sus cortos ciclos de vida, su alta fertilidad, su exoesqueleto altamente resistente y unas bajas tasas de extinción. De todas las familias de este orden, la familia Curculionidae es la más grande, con más de 51,000 especies descritas, posiblemente siendo la familia con más especies de entre todos los animales. La familia Dryophthoridae contiene algunos de los insectos plaga más destructivos incluyendo a los granívoros *Sitophilus* (Coleoptera, Dryophthoridae) que pueden causar perdidas en los cereales almacenados de entre el 25 y el 40% del peso total. De las especies de este género, el gorgojo del arroz, *S. oryzae*, es la más destructiva. Además de las pérdidas causadas por estos gorgojos, el polvo liberado durante su alimentación atrae a pestes secundarias que pueden acarrear micotoxinas.

Al igual que otros insectos holometábolos, el ciclo de vida de *S. oryzae* se divide en cuatro estadios: huevo, larva, pupa y adulto. Los primeros tres estadios ocurren dentro de los granos. Las hembras hacen un pequeño orificio en el grano utilizando sus mandíbulas y depositan un huevo que posterior cubren con una secreción. Al eclosionar, la larva comienza a desarrollarse dentro del grano consumiéndolo desde el interior. Posteriormente pupa dentro del grano y emerge como un adulto.

Una característica importante de los gorgojos del género *Sitophilus* es su asociación permanente con endosimbiontes que les proporcionan nutrientes poco abundantes en los granos. Actualmente *S. oryzae* mantiene una relación mutualista con *Candidatus* Sodalis pierantonius (de ahora en delante *S. pierantonius*). Estos endosimbiontes se transmiten por vía maternal y en las larvas inducen la formación de los bacteriocitos. Esta interacción se estableció recientemente (hace menos de 30,000 años) reemplazando a *Nardonella*, el simbionte anterior.

Numerosos trabajos han demostrado que *S. pierantonius* aumenta las capacidades invasivas de *S. oryzae* al incrementar su fertilidad y capacidad de vuelo. Además, al contar con la secuencia del genoma de su genoma se ha podido inferir su metabolismo y se determinó que el endosimbionte es capaz de proveer a su hospedador con todos los aminoácidos excepto metionina, triptófano e histidina. El estudio de su genoma también permitió determinar que a diferencia de otros genomas de endosimbiontes más antiguos, este aun no experimenta una reducción tan drástica en su tamaño. Tiene un alto contenido de GC, numerosos genes pseudogenizados y un gran número de elementos móviles que abarcan el 18% de su genoma. Otro punto importante es que de determinó que *S. pierantonius* tiene un rol muy importante en la síntesis del exoesqueleto de los adultos a través de la tirosina y fenilalanina que proporciona al hospedador. Interesantemente, el hospedador es capaz de controlar con precisión las etapas de crecimiento de *S. pierantonius* permitiendo su proliferación cuando requiere mayores cantidades de tirosina.

Por otra parte, los áfidos (Hemiptera, Aphididae) forman parte de un grupo de insectos con más de 5,000 especies descritas. Al ser hemípteros estos no tienen una metamorfosis completa y únicamente tienen tres estadios: huevo, ninfa y adulto. No obstante, sus ciclos de vida distan de ser simples y cuentan con adultos con diferentes morfologías. La mayoría de ellos presentan un ciclo de vida holocíclico en el cual existen varias generaciones de hembras partenogenéticas. Con la llegada del frio producen una generación de formas sexuales las cuales depositan huevos que son capaces de resistir las temperaturas extremas del invierno. Con la llegada de la primavera estos huevos eclosionan y comienza nuevamente el ciclo. Esta característica indudablemente es una de las razones de su éxito.

Los áfidos también albergan endosimbiontes, la gran mayoría de ellos cuenta con *Buchnera* como su endosimbionte primario. Sin embargo, son capaces de albergar múltiples endosimbiontes primarios, y la subfamilia Lachninae parece ser especialmente propensa a esto. En su caso, además de albergar a *Buchnera* también cuentan con *Serratia* lo cual los vuelve excelentes modelos para estudiar la complementación endosimbiótica y el proceso por el cual un endosimbionte secundario puede pasar a ser co-primario. El género *Cinara*, y en especial el áfido del cedro *Cinara cedri* es uno de los mejor estudiados dentro de esta subfamilia.

La secuenciación de ambos endosimbiontes primarios de *C. cedri* determinó que estos habían establecido una relación coobligada con su hospedador. Se observó que *Buchnera* había perdido la habilidad de producir riboflavina y triptófano. Mientras que *Serratia* se encarga por completo de la producción de riboflavina se descubrió que la síntesis de triptófano estaba compartida entre ambos endosimbiontes. Esto implicaba la biosíntesis de antranilato por parte de *Buchnera* y su transferencia a *Serratia* la cual se encarga de convertirlo en triptófano. Posteriormente la secuenciación de otros endosimbiontes de miembros de la subfamilia Lachninae confirmó que el consorcio *Buchnera-Serratia* se estableció antes de la diversificación del linaje ya que en todos los casos *Buchnera* contaba con un genoma pequeño y había perdido la habilidad de sintetizar riboflavina.

A pesar de que estamos principalmente interesados en identificar las diferencias entre ambos modelos a causa de la edad de la relación entre el hospedador y su endosimbionte es obvio que hay múltiples diferencias que debemos tener en cuenta. Entre ellas el hábitat en el que se encuentran, su dieta, su posición taxonómica y si son holo- o hemimetábolos. Para ello incluimos otros 19 artrópodos con genomas completos disponibles. Como grupos externos se incluyeron un crustáceo, la pulga de agua *Daphnia pulex*, and un arácnido que a pesar de ser conocido como araña roja se trata de un ácaro, *Tetranychus urticae*.

En cuanto a los insectos hemimetábolos incluimos varios hemípteros: (i) el áfido del cedro *C. cedri* que como se mencionó anteriormente alberga *Buchnera* y *Serratia*; el áfido del guisante *Acyrthosiphon pisum* y el áfido verde del melocotonero *Myzus persicae* los cuales únicamente albergan *Buchnera*; y el psílido asiático de los cítricos *Diaphorina citri* el cual alberga *Candidatus* Carsonella ruddii DC, los cuatro pertenecen al suborden Sternorrhyncha y se alimentan de floema; (ii) el salta hojas marrón *Nilaparvata lugens* del suborden Auchenorrhyncha, el cual alberga una levadura como endosimbionte y también se alimenta de floema; (iii) la chinche *Cimex lectularius*, la cual alberga

Wolbachia y se alimenta de sangre; y (iv) el piojo *Pediculus humanus* el cual alberga *Candidatus* Riesia pediculicola y se alimenta de sangre también.

De la clase Himenóptera: (i) la hormiga carpintera de Florida *Camponotus floridanus*; y la hormiga roja de fuego *Solenopsis invicta,* ambas omnívoras; (ii) la abeja europea *Apis mellifera*, que se alimenta exclusivamente de polen y néctar; y (iii) la avispa parasitoide *Nasonia vitripennis*, la cual es parasitoide de varias moscas. Entre ellos, solo *C. floridanus* alberga un endosimbionte, *Blochmania floridanus*.

De los Diptera: (i) dos miembros de la familia Culicidae, el mosquito de la fiebre amarilla *Aedes aegypti* y *Anopheles gambiae*, los cuales durante su estadio larvario se alimentan de bacterias, algas y otros microorganismos y durante su etapa adulta se alimentan de néctar, mientras que las hembras también se alimentan de sangre para permitir el desarrollo de su huevos; (ii) la mosca del vinagre *Drosophila melanogaster,* que se alimenta de materia vegetal en descomposición; y (iii) la mosca tse-tsé *Glossina morsitans*, la cual se alimenta exclusivamente de sangre. Solo la mosca tse-tsé alberga un endosimbionte obligado *Wigglesworthia glossinidia*.

De Lepidóptera: el gusano de la seda *Bombyx mori*, el gusano del tabaco *Manduca sexta*, la mariposa monarca *Danaus plexippus* y la palomilla dorso de diamante *Plutella xylostella*. Todos ellos se alimentan de las hojas de distintas plantas durante su estadio larvario y de néctar durante cuando son adultos. Ninguna de ellas tiene endosimbiontes asociados.

Por último, del orden Coleóptera: (i) el barrenador esmeralda del fresno *Agrilus planipennis*, el escarabajo asiático de los cuernos *Anoplophora glabripennis* y el escarabajo del pino de montaña *Dendroctonus ponderosae*, los cuales se alimentan de la corteza interna y del floema de diversas especies de árboles; (ii) el escarabajo de la colmena *Aethina tumida*, el cual se alimenta de miel y polen; (iii) el gorgojo del arroz *S. oryzae* y el escarabajo rojo de la harina *Tribolium castaneum* que se alimentan de granos almacenados; y (iv) *Nicrophorus vespilloides* un escarabajo que se alimenta de cadáveres de otros animales. El único escarabajo con un endosimbionte del grupo que seleccionamos es *S. oryzae*.

Objetivos

El objetivo principal de esta tesis es comparar el sistema inmune innato y las rutas de biosíntesis de aminoácidos entre insectos que albergan endosimbiontes. Ambas rutas de señalización fueron elegidas ya que es a través de ellas que el insecto interactúa con sus endosimbiontes. Los organismos elegidos fueron el gorgojo del arroz *Sitophilus oryzae* y el pulgón del cedro *Cinara cedri*. Dichos organismos fueron elegidos ya que nos interesa identificar las diferencias en el repertorio genético de insectos con relaciones endosimbióticas con diferentes edades. Mientras que la relación entre *C. cedri* y *Buchnera* es bastante antigua (se estableció hace al menos 150 millones de años), la de *S. oryzae* y *Sodalis pierantonius* es mucho más joven, estableciéndose hace alrededor de 30,000 años. Además, contamos con la secuencia de los endosimbiontes de ambos sistemas mismos que han sido estudiados a profundidad por los grupos de Abdelaziz Heddi y Amparo Latorre.

1. El primer objetivo es obtener la secuencia del genoma de *S. oryzae* y anotarla para identificar los genes involucrados tanto en el sistema inmune como en el metabolismo de aminoácidos. Este grupo de genes se comparará con el repertorio de *T. castaneum* y *Drosophila melanogaster*, el modelo de los coleópteros y el de los insectos respectivamente. Esta comparación nos permitirá identificar diferencias entre un escarabajo que alberga un endosimbionte y otro que no tiene endosimbiontes. Es importante recordar que *S. oryzae* y *T. castaneum* pertenecen al mismo orden y comparten nicho ecológico. Además, la relación entre *S. oryzae* y *S. pierantonius* se estableció muy recientemente, por lo tanto, no sería sorprendente encontrar pocas diferencias entre ambos sistemas. La comparación con *D. melanogaster* nos permitirá identificar características que son específicas del orden Coleóptera.
2. Obtener la lista de genes involucrados en el sistema inmune innato y en la biosíntesis de aminoácidos en *C. cedri* utilizando la secuencia del

genoma ensamblado y anotado. Dicho listado de genes se comparará con los ortólogos obtenidos en *Acyrthosiphon pisum* y *D. melanogaster*. En este caso *A. pisum* también alberga un endosimbionte y la comparación entre ambos áfidos nos permitirá identificar las características únicas de *C. cedri* a pesar de su similitud con *A. pisum*. Al comparar con *D. melanogaster* podremos identificar las características únicas de los insectos con relaciones simbióticas ancestrales que no se encuentran en el modelo de los insectos.

3. Comparar los genes involucrados en el sistema inmune o en la biosíntesis de aminoácidos entre *S. oryzae* y *C. cedri*. Esto nos permitirá alcanzar el objetivo principal de este proyecto. No obstante, existen múltiples diferencias entre ambos modelos además de la fecha del establecimiento de sus relaciones endosimbióticas. Para tomar esto en cuenta e identificar las diferencias entre ambos ocasionadas por otros factores se incluirán otros insectos en la comparación.

Metodología y resultados

Capítulo 1. El gorgojo del arroz *Sitophilus oryzae*

El genoma de *S. oryzae* fue secuenciado con una cobertura 101X utilizando 20 adultos. El tamaño del genoma obtenido fue de 652 Mb en 17,786 scaffolds. La cantidad de elementos transponibles (48.6% del genoma) se encuentra entre las más elevadas descritas en insectos, entre ellos el mosquito tigre *Aedes albopictus* con 50% y la mosca *Musca domestica* con el 52%.

Incorporando información de RNA-seq de 12 librerías generadas en diferentes condiciones y proteínas anotadas en otros coleópteros se anotó el genoma de *S. oryzae*. Se identificaron 17,026 modelos de genes con la predicción automática y 1,675 genes se curaron manualmente enfocándose en metabolismo, inmunidad, desarrollo, epigenética, sistema olfativo y genes transferidos horizontalmente. El número final de genes anotados en *S. oryzae* fue de 17,159. De ellos más del 85% tiene un ortólogo en al menos una de las especies de artrópodos incluidas en nuestro estudio. Además, determinamos que S. oryzae tiene el mayor número de genes linaje específicos dentro de los coleópteros. Esto posiblemente se encuentra asociado a la alta tasa de expansión de familias de genes que calculamos en esta especie.

Utilizando una filogenia de los coleópteros utilizados en nuestro análisis identificamos las familias con evolución acelerada en cada una de las especies y con 174 familias, S. oryzae fue el escarabajo con la tasa de expansión más elevada (0.409 genes por millón de años). Este proceso podría estar ligado al alto número de elementos transponibles tal como fue descrito en las termitas.

En cuanto al metabolismo de aminoácidos determinamos que la alanina y la prolina son proporcionadas por el insecto en tanto que la treonina, lisina, fenilalanina y arginina son proporcionadas por el endosimbionte. Ya que el endosimbionte ancestral únicamente era capaz de producir tirosina esto podría haberle dado una ventaja a *S. pierantonius* sobre *Nardonella*. Finalmente, la valina, leucina, isoleucina, triptófano, metionina e histidina deben ser obtenidas de la dieta.

Se determine que el sistema inmune de S. oryzae es muy similar al de *T. castaneum* e incluso *D. melanogaster*. Los genes involucrados en las cascadas de señalización se encuentran conservados y las diferencias principales se encuentran a nivel de los receptores y efectores. Una de las diferencias más significativas es que S. oryzae carece del receptor PGRP-LE, lo cual podría afectar su capacidad para reconocer bacterias intracelulares, incluyendo a sus propios endosimbiontes. Otro gen no identificado en S. oryzae es la cinasa Gprk2 la cual está involucrada en la amplificación de la señal en la ruta Toll, y por ello posiblemente no es esencial. Las proteasas involucradas en el inicio de la cascada de señalización que lleva a la activación del receptor Toll no pudieron ser identificadas debido al gran número de proteasas y la dificultad para establecer relaciones de ortología entre ellas. Finalmente, se identificaron péptidos antimicrobianos, tanto aquellos descritos previamente como algunos nuevos.

Capítulo 2. El áfido del cedro *Cinara cedri*

El genoma de *C. cedri* fue secuenciado, ensamblado y anotado en el Centro Nacional de Análisis Genómico. El tamaño del genoma ensamblado fue de 396

Mb distribuidas en 1,740 scaffolds. A pesar de que no existen grandes diferencias entre los tamaños de los genomas de los áfidos cuyos genomas se encuentran disponibles si existen diferencias entre los números de genes identificados en cada especie. En tanto que *A. pisum* y *M. persicae* tienen un número similar de genes predichos, en *C. cedri* se encontraron 1,500 menos genes. Dada la alta calidad del ensamble, parece que no se trata de errores técnicos y que efectivamente C. cedri tiene un número menor de genes, sugiriendo perdidas especificas en su linaje o expansiones en la familia Aphididae a la cual pertenecen *A. pisum* y *M. persicae*.

Aproximadamente el 62% de los genes en *C. cedri* tienen ortólogos en al menos algún otro artrópodo. Al comparar el número de genes linaje específicos de cada especie observamos que a pesar de que C. cedri tiene el menor número de genes entre los áfidos que incluimos en nuestro estudio, también tiene el mayor número de genes linaje específicos (6,449, aproximadamente el 38% de sus genes).

Respecto al metabolismo de aminoácidos en el holobionte *C. cedri*, *Buchnera* proporciona la histidina, fenilalanina, lisina y treonina mientras que la serina, alanina, glutamato, glutamina, prolina, aspartato, asparagina y tirosina pueden ser sintetizadas por el insecto. La cisteína puede ser producida tanto por el insecto como por Serratia y la glicina por los tres miembros del consorcio. La valina, leucina, arginina e isoleucina son producidas por una colaboración entre *C. cedri* y *Buchnera*. El triptófano se produce por una colaboración entre *Buchnera* y *Serratia*, siendo esta una de las principales razones por las cuales ambos son endosimbiontes co-primarios. Finalmente, se ha sugerido que la metionina puede producirse a través de una colaboración entre *C. cedri* y *Buchnera*, sin embargo, esto requeriría que una enzima de *C. cedri* funcionara en la dirección opuesta a la que lo hace normalmente y esto nunca se ha observado en la naturaleza. La alternativa es que la metionina se obtiene de la dieta.

El sistema inmune de *C. cedri* es muy parecido al de *A. pisum* con pequeñas diferencias en el número de copias de algunos genes. Entre las características que tienen en común se encuentra la ausencia de la mayoría de los genes involucrados en la señalización de la ruta IMD. Ello implica que los áfidos son incapaces de reconocer a las bacterias Gramnegativas, al menos por la principal vía descrita. Esto posiblemente permitió no solo la adquisición de *Buchnera* el endosimbionte primario de los áfidos, si no de múltiples endosimbiontes secundarios o co-primarios, como el caso de *Serratia* en *C. cedri*, pues al tratarse también de una bacteria Gramnegativa, esta no tendría demasiados problemas para colonizar al insecto. En el caso de *C. cedri*, tampoco fue posible establecer una relación clara de ortología entre las proteasas involucradas en la activación de la ruta Toll del áfido y aquellas definidas en *D. melanogaster*, No obstante, se sabe que la ruta esta activa, ya sea utilizando los ortólogos de las proteasas en la mosca del vinagre u otras proteasas. Finalmente, una de las diferencias más llamativas entre los áfidos y los demás insectos es la ausencia de péptidos antimicrobianos a excepción de la taumatina. Sin embargo, dado el reto que representa identificar péptidos antimicrobianos es posible que los áfidos cuenten con otras de estas moléculas que aún no han sido identificadas entre su arsenal.

Otro factor interesante en *C. cedri* es que todos los individuos muestreados hasta la fecha cuentan con *Wolbachia* como endosimbionte, lo cual sugiere que tiene algún rol en la biología del insecto. Debido a que la secuencia de su genoma se obtuvo al secuenciar al insecto, esta también se analizó y no se identificaron rutas para la biosíntesis de aminoácidos ausentes en el holobionte ni rutas metabólicas que puedan ser de interés para *C. cedri*. El hecho de que la gran mayoría de sus genes cepa específicos están involucradas en la transposición o son proteínas hipotéticas dificulta hipotetizar sobre su rol.

Capítulo 3. Comparación entre ambos modelos

Para poner la comparación entre *C. cedri* y *S. oryzae* en contexto se incorporaron los otros 26 artrópodos previamente mencionados. Se identificaron 1,327 genes conservados con una ortología uno a uno entre todos los artrópodos. Además 1,478 tenían ortólogos en todos los artrópodos sin ser estrictamente uno a uno llevando a un total de 2,805 genes que formarían el genoma-núcleo de los artrópodos. Y si consideramos únicamente a los insectos se incluiría un grupo adicional de 2,578 genes que se encuentran en todos ellos, pero no en los dos artrópodos no insectos.

Analizando específicamente a los coleópteros se identificó un grupo de 1,356 genes presentes únicamente en ellos y se observa una distribución más menos homogénea, sin embargo, en *A. planipennis* se observan solo 784 en esta categoría. Esto es especialmente interesante ya que es el linaje más antiguo dentro de los escarabajos incluidos en el estudio, sugiriendo que pudo existir una serie de duplicaciones en el ancestro de los otros coleópteros. De interés especial también es el grupo de los lepidópteros ya que se encontraron 3,106 genes específicos de su orden, el numero más grande de entre los holometábolos analizados y además el más homogéneo. En el caso de los hemimetábolos *N. lugens* parece representar una anomalía. Este insecto tiene un repertorio de genes muy por encima de cualquier otro incluido en el análisis.

Los insectos que elegimos para este análisis viven en ambientes muy distintos y deben enfrentarse a diferentes tipos de estrés incluido obtener todos los nutrientes necesarios para su correcto desarrollo. Nosotros nos enfocamos únicamente en los aminoácidos, sin embargo, la obtención de otros factores como las vitaminas también es crucial y es sabido que los endosimbiontes participan en esa tarea.

En cuanto a la síntesis de glutamato y aspartato, todos los artrópodos evaluados, así como varios endosimbiontes son capaces de producirlo. Esto no es ninguna sorpresa ya que estos aminoácidos son esenciales para la síntesis de otros aminoácidos. La glutamina, serina, glicina y cisteína también pueden ser producidas por todos los artrópodos y algunos endosimbiontes. La síntesis de prolina y alanina también esta conservada entre todos los eucariotas, incluido el endosimbionte de *N. lugens*. Sin embargo, no se encuentra en ningún endosimbionte bacteriano sugiriendo que estas rutas tienen una mayor relevancia en los eucariotas y que son prescindibles en los endosimbiontes bacterianos. La asparagina puede ser producida por todos los artrópodos y los simbiontes de los insectos que se alimentan de floema son incapaces de producirla, esto puede estar relacionado con el alto contenido de este aminoácido en el floema.

La arginina no puede ser producida en su totalidad por ningún artrópodo, pero todos pueden producirla a partir de la citrulina. Algunos endosimbiontes pueden producirla a partir de la ornitina e interesantemente parece que los lepidópteros también tienen la habilidad de catalizar esta reacción. La lisina, metionina, valina, leucina, isoleucina y treonina no pueden producirse por ningún artrópodo, sin embargo, algunos endosimbiontes son capaces de producirla. En el caso de la lisina, aquellos que no son capaces de producirla son los endosimbiontes de insectos que se alimentan de sangre, sugiriendo que el contenido de este aminoácido en la dieta es suficiente para satisfacer los requerimientos. En el caso de los aminoácidos aromáticos, ningún artrópodo es capaz de producirlos, sin embargo, algunos endosimbiontes son capaces de producirlos. Tal es el caso del holobionte *C. floridanus* así como los áfidos que puede producir los tres o *S. oryzae* que puede producir fenilalanina y tirosina. Por último, en cuanto a la histidina *Buchnera*, *B. floridanus*, el endosimbionte de *N. lugens* y posiblemente *C.* Carsonella ruddii DC son capaces de producirla.

Al comparar los sistemas inmunes de los artrópodos que seleccionamos para nuestro estudio observamos que los péptidos antimicrobianos son altamente específicos para cada orden e incluso en cada linaje. A pesar de que sabemos que la mayoría de la diversidad de estos péptidos no ha sido explorada debido a las dificultades para identificarlos podemos concluir que mientras que los áfidos únicamente cuentan con una familia de péptidos antimicrobianos los holometábolos cuentan con una gran diversidad de estos efectores. En cuanto a respuesta antiviral RNAi, todos los artrópodos cuentan con rutas mayormente completas y únicamente se observan pequeñas perdidas como las pérdidas de HPS4 y Nbr que están involucradas en la carga del miRNA en la proteína Ago1, sin embargo, se sabe que estas no son esenciales. La ruta JAK/STAT también está altamente conservada, sin embargo, un aspecto importante de esta ruta es que el principal ligando, dome, no ha sido identificado en ninguna otra especie además de la mosca del vinagre.

La ruta de señalización Toll se encuentra conservada en la mayoría de los artrópodos, pero existen dificultades para identificar correctamente las proteasas involucradas en el procesamiento del ligando principal. Esto se debe sobre todo a que se trata de una familia de enzimas muy grande y establecer las relaciones de ortología entre ellas no es trivial. Un aspecto similar entre los áfidos y los coleópteros es la ausencia de ref(2)P. De entre todas las especies estudiadas en ambos grupos únicamente se identificó en *A. planipennis*, el grupo hermano de todos los demás coleópteros, sugiriendo que este gen se perdió tras la divergencia de estos linajes.

Posiblemente el aspecto más relevante de este estudio son las grandes diferencias en la vía de señalización IMD. Los áfidos son el único grupo en el que no se encuentra CYLD, y además en ninguna especie del suborden Sternorrhyncha se encuentran ird5, imd, relish ni tab2. Dredd, Fadd, key y pirk no se encuentran en ninguno de los miembros del orden Hemíptera, por lo tanto, sugiriendo que este grupo de organismos no son capaces de identificar y posiblemente responder ante infecciones de bacterias Gramnegativas. Ya que la mayoría de los hemípteros se alimentan de dietas mayormente estériles esto podría no representar un problema demasiado grande, además ahorrando los recursos que son necesarios para mantener dicha vía del sistema inmune. Además, el hecho de no contar con la ruta IMD permite que estos insectos sean colonizados con mayor facilidad por potenciales endosimbiontes.

Conclusiones generales

La cantidad de repeticiones en el genoma de *S. oryzae* se encuentra entre las mayores observadas en otros insectos y la mayor en cualquier coleóptero estudiado. El número de elementos móviles posiblemente ha propiciado que sea el escarabajo con la tasa de expansión de familias génicas más elevada entre aquellos que incluimos en nuestro estudio.

Se observa una clara dependencia entre el metabolismo de *S. oryzae* y su endosimbionte. En cuanto a los aminoácidos esenciales, algunos se producen por el hospedador, otros por el endosimbionte y otros deben obtenerse a partir de la dieta. El hecho de que *S. pierantonius* es capaz de producir más aminoácidos que *Nardonella* es una posible razón para explicar el reemplazo dada la baja abundancia de lisina y treonina en los cereales.

El sistema inmune de *S. oryzae* es muy similar al de *T. castaneum* e incluso *D. melanogaster*. Una de las pocas diferencias es la ausencia del receptor PGRP-LE, lo cual es una posible puede explicar que *S. oryzae* sea capaz de tolerar a su endosimbionte. También se identificaron varios péptidos antimicrobianos. El hecho de que su sistema inmune sea tan parecido al de otros insectos sugiere que pequeños cambios son suficientes para permitir albergar un endosimbionte obligado.

No encontramos eventos de transferencia horizontal en el genoma de *S. oryzae* que parezcan provenir de *S. pierantonius*. Esto sugiere que existen mecanismos para evitar este fenómeno o quizás no ha habido tiempo suficiente para observar esta transferencia.

El metabolismo de aminoácidos esta compartido entre *C. cedri*, *Buchnera* y *Serratia*. La cisteína puede ser producida tanto por el hospedador como por *Serratia*. La valina, leucina, arginina e isoleucina son producidas por una colaboración entre *C. cedri* and *Buchnera*. Por último, aún no está claro si la metionina es obtenida de la dieta o mediante una colaboración entre *C. cedri* and *Buchnera*.

El sistema inmune de *C. cedri* es muy similar al de *A. pisum* con pequeñas diferencias en el número de copias de algunos genes. En cambio, en comparación con otros insectos existen grandes diferencias en la ruta de señalización IMD. Además, no se identificaron otros péptidos antimicrobianos además de la taumatina.

Mientras que en todos los áfidos de la especie *C. cedri* se ha encontrado *Wolbachia* no ha sido posible definir su rol, si es que tiene alguno. No parece colaborar en la biosíntesis de aminoácidos ni se identificaron otras rutas metabólicas que puedan ser de importancia para el holobionte. La mayoría de los genes específicos de esta cepa están involucrados en la transposición o son proteínas hipotéticas.

En cuanto al metabolismo de aminoácidos, todos los artrópodos tienen un repertorio de genes parecido, apoyando la idea de una gran pérdida de genes en el origen de los metazoos. Sin embargo, (i) la arginina no puede ser producida por los artrópodos, sin embargo algunos endosimbiontes pueden proporcionar este aminoácido a sus hospedadores, además los lepidópteros cuentan con una enzima que les permite producirla a partir de la ornitina; (ii) la lisina puede ser producida por todos los endosimbiontes, excepto aquellos que habitan en insectos que se alimentan de sangre; (iii) la alanina puede ser producida por todos los artrópodos, sin embargo se ha perdido en todos los endosimbiontes bacterianos, sugiriendo que existe una presión selectiva para que los eucariotas la conserven; (iv) *S. pierantonius* puede producir fenilalanina y tirosina mientras que *Buchnera* únicamente puede producir fenilalanina por si solo y requiere a *Serratia* para producir triptófano.

Los péptidos antimicrobianos son orden-, o incluso linaje-específicos. En el caso de los insectos hemimetábolos únicamente fue posible identificar defensina y taumatina. En cuanto a la respuesta antiviral, todos los artrópodos cuentan con las vías de señalización intactas, lo cual habla de su importancia. Eso también ocurre en el caso de la ruta de señalización JAK/STAT, la cual se conserva en todos los artrópodos a pesar de que el ligando únicamente se ha identificado en *D. melanogaster*. En cuanto a la ruta Toll, es posible que existan diferencias en las proteasas involucradas en el procesamiento del ligando del receptor Toll, sin embargo, es difícil de afirmar por la dificultad para asignar ortología entre las proteasas.

Los áfidos carecen de un gran número de elementos de la ruta de señalización IMD; sin embargo, todos los hemípteros carecen de ciertos elementos de esta ruta. Esto implica que los hemípteros tienen una menor capacidad para reconocer y posiblemente responder ante patógenos Gramnegativos. Consideramos que se trata de perdidas en el grupo de los hemimetábolos y no de una adquisición tardía puesto que esta vía de señalización se encuentra con todos sus elementos en algunos insectos hemimetábolos como la cucaracha americana.

BIBLIOGRAPHY

Adams, M D, S E Celniker, R A Holt, C A Evans, J D Gocayne, P G Amanatides, S E Scherer, et al. 2000. “The Genome Sequence of Drosophila Melanogaster.” Science (New York, N.Y.) 287 (5461): 2185–95.

Akashi, Hiroshi, and Takashi Gojobori. 2002. “Metabolic Efficiency and Amino Acid Composition in the Proteomes of Escherichia Coli and Bacillus Subtilis.” Proceedings of the National Academy of Sciences 99 (6): 3695–3700. doi:10.1073/pnas.062526999.

Altincicek, Boran, Eileen Knorr, and Andreas Vilcinskas. 2008. “Beetle Immunity: Identification of Immune-Inducible Genes from the Model Insect Tribolium Castaneum.” Developmental and Comparative Immunology 32 (5): 585–95. doi:10.1016/j.dci.2007.09.005.

Altincicek, Boran, Bas ter Braak, Alice M. Laughton, Klas I. Udekwu, and Nicole M. Gerardo. 2011. “Escherichia Coli K-12 Pathogenicity in the Pea Aphid, Acyrthosiphon Pisum, Reveals Reduced Antibacterial Defense in Aphids.” Developmental and Comparative Immunology 35 (10). Elsevier Ltd: 1091–97. doi:10.1016/j.dci.2011.03.017.

Altschul, S F, W Gish, W Miller, E W Myers, and D J Lipman. 1990. “Basic Local Alignment Search Tool.” Journal of Molecular Biology 215 (3): 403–10. doi:10.1016/S0022-2836(05)80360-2.

Anbutsu, Hisashi, Minoru Moriyama, Naruo Nikoh, Takahiro Hosokawa, Ryo Futahashi, Masahiko Tanahashi, Xian-Ying Meng, et al. 2017. “Small Genome Symbiont Underlies Cuticle Hardness in Beetles.” Proceedings of the National Academy of Sciences, 201712857. doi:10.1073/pnas.1712857114.

Anderson, Cort L., and Timothy L. Karr. 2001. “Wolbachia: Evolutionary Novelty in a Rickettsial Bacteria.” BMC Evolutionary Biology 1. doi:10.1186/1471-2148-1-10.

Anselme, Caroline, Vicente Pérez-Brocal, Agnès Vallier, Carole Vincent-Monegat, Delphine Charif, Amparo Latorre, Andrés Moya, and Abdelaziz Heddi. 2008. “Identification of the Weevil Immune Genes and Their Expression in the Bacteriome Tissue.” BMC Biology 6: 1–13. doi:10.1186/1741-7007-6-43.

Arp, Alex P, Wayne B Hunter, Kirsten S Pelz-Stelinski, and Nicolas Durand. 2016. “Annotation of the Asian Citrus Psyllid Genome Reveals a Reduced Innate Immune System.” Frontiers in Physiology 7 (570). doi:10.3389/fphys.2016.00570.

Asgharian, Hosseinali, Peter L. Chang, Peter J. Mazzoglio, and Ilaria Negri. 2014. “Wolbachia Is Not All about Sex: Male-Feminizing Wolbachia Alters the Leafhopper Zyginidia Pullula Transcriptome in a Mainly Sex-Independent

Manner.” Frontiers in Microbiology 5 (SEP): 1–10. doi:10.3389/fmicb.2014.00430.

Augustinos, Antonis A., Diego Santos-Garcia, Eva Dionyssopoulou, Marta Moreira, Aristeidis Papapanagiotou, Marios Scarvelakis, Vangelis Doudoumis, et al. 2011. “Detection and Characterization of Wolbachia Infections in Natural Populations of Aphids: Is the Hidden Diversity Fully Unraveled?” Edited by Richard Cordaux. PLoS ONE 6 (12): e28695. doi:10.1371/journal.pone.0028695.

Avila, A., N. Silverman, M. T. Diaz-Meco, and J. Moscat. 2002. “The Drosophila Atypical Protein Kinase C-Ref(2)P Complex Constitutes a Conserved Module for Signaling in the Toll Pathway.” Molecular and Cellular Biology 22 (24): 8787–95. doi:10.1128/mcb.22.24.8787-8795.2002.

Balabanidou, Vasileia, Linda Grigoraki, and John Vontas. 2018. “Insect Cuticle: A Critical Determinant of Insecticide Resistance.” Current Opinion in Insect Science. Elsevier Inc. doi:10.1016/J.COIS.2018.03.001.

Bansal, Raman, and Andy Michel. 2018. “Expansion of Cytochrome P450 and Cathepsin Genes in the Generalist Herbivore Brown Marmorated Stink Bug.” BMC Genomics 19 (1). BMC Genomics: 1–14. doi:10.1186/s12864-017-4281-6.

Barribeau, Seth M., Daniel Sok, and Nicole M. Gerardo. 2010. “Aphid Reproductive Investment in Response to Mortality Risks.” BMC Evolutionary Biology 10 (1). doi:10.1186/1471-2148-10-251.

Baumann P. 2006: Diversity of prokaryote-insect associations within the Sternorrhyncha (psyllids, whiteflies, aphids, mealybugs). In Bourtzis K. & Miller T.A. (eds): Insect Symbiosis. Vol. 2. Contemporary Topics in Entomology Series. CRC Press, Boca Raton, FL, pp. 1-24

Beard, C. B., C. R. Woese, F. F. Richards, S. L. O’Neill, L. Mandelco, P. Mason, R. B. Tesh, and S. Aksoy. 2007. “Genetic Transformation and Phylogeny of Bacterial Symbionts from Tsetse.” Insect Molecular Biology 1 (3): 123–31. doi:10.1111/j.1365-2583.1993.tb00113.x.

Beemelmanns, Christine, Huijuan Guo, Maja Rischer, and Michael Poulsen. 2016. “Natural Products from Microbes Associated with Insects.” Beilstein Journal of Organic Chemistry 12: 314–27. doi:10.3762/bjoc.12.34.

Besemer, John, and Mark Borodovsky. 2005. “GeneMark: Web Software for Gene Finding in Prokaryotes, Eukaryotes and Viruses.” Nucleic Acids Research 33 (July). Oxford University Press: 451–54. doi:10.1093/nar/gki487.

Bethoux, O. 2009. “Gaps and Nodes between Fossil and Extant Insects.” Systematic Entomology 34 (4): 599–609. doi:10.1111/j.1365-3113.2009.00484.x.

Binazzi, Andrea, and Stephan Scheurer. 2009. Atlas of the Honeydew Producing Conifer Aphids of Europe. Aracne.

Binazzi, Francesco, Giuseppino Sabbatini Peverieri, and Pio Federico Roversi. 2015. “First Record in Cyprus of Cinara (Cinara) Cedri Mimeur (Aphididae Lachninae) on Cedrus Brevifolia (Hooker Fil.) Henry.” Redia 98 (December): 151–54.

Binazzi, A. 1978. Contributi alla conoscenza degli afididelle conifere. I. Le specie dei genn. Cinara Curt.,Schizolachnus Mordv.,Cedrobium Remaud.edEulachnus D.Gu. presenti in Italia (HomopteraAphidoidea Lachnidae).- Redia, LXI: 291-400.

Blackman, Roger L., and Victor Frank Eastop. 1994. Aphids on the World’s Trees : An Identification and Information Guide. CAB International in association with the Natural History Museum.

Blackman, RL, and VF Eastop. 2017. “Aphids on the World’s Plants: An Online Identification and Information Guide.” http://www.aphidsonworldsplants.info/.

Bocak, Ladislav, Christopher Barton, Alex Crampton-Platt, Douglas Chesters, Dirk Ahrens, and Alfried P. Vogler. 2014. “Building the Coleoptera Tree-of-Life for > 8000 Species : Composition of Public DNA Data and Fit with Linnaean Classification.” Systematic Entomology 39: 97–110. doi:10.1111/syen.12037.

Boetzer, Marten, and Walter Pirovano. 2012. “Toward Almost Closed Genomes with GapFiller.” Genome Biology 13 (6): R56. doi:10.1186/gb-2012-13-6-r56.

Boetzer, Marten, Christiaan V. Henkel, Hans J. Jansen, Derek Butler, and Walter Pirovano. 2011. “Scaffolding Pre-Assembled Contigs Using SSPACE.” Bioinformatics 27 (4): 578–79. doi:10.1093/bioinformatics/btq683.

Bonasio, Roberto, Qiye Li, Jinmin Lian, Navdeep S Mutti, Lijun Jin, Hongmei Zhao, Pei Zhang, et al. 2012. “Genome-Wide and Caste-Specific DNA Methylomes of the Ants Camponotus Floridanus and Harpegnathos Saltator.” Current Biology : CB 22 (19). Howard Hughes Medical Institute: 1755–64. doi:10.1016/j.cub.2012.07.042.

Bracewell, Ryan R, and Diana L Six. 2015. “Experimental Evidence of Bark Beetle Adaptation to a Fungal Symbiont.” Ecology and Evolution 5 (21). Wiley-Blackwell: 5109–19. doi:10.1002/ece3.1772.

Brandazza, Anna, Sergio Angeli, Mariella Tegoni, Christian Cambillau, and Paolo Pelosi. 2004. “Plant Stress Proteins of the Thaumatin-like Family Discovered in Animals.” FEBS Letters 572 (1–3): 3–7. doi:10.1016/j.febslet.2004.07.003.

Brelsfoard, Corey, George Tsiamis, Marco Falchetto, Ludvik M. Gomulski, Erich Telleria, Uzma Alam, Vangelis Doudoumis, et al. 2014. “Presence of Extensive Wolbachia Symbiont Insertions Discovered in the Genome of Its Host Glossina

Morsitans Morsitans.” Edited by Jesus G. Valenzuela. PLoS Neglected Tropical Diseases 8 (4). Public Library of Science: e2728. doi:10.1371/journal.pntd.0002728.

Broderick, Nichole A, and Bruno Lemaitre. 2012. “Gut-Associated Microbes of Drosophila Melanogaster.” Gut Microbes 3 (4). Taylor & Francis: 307–21. doi:10.4161/gmic.19896.

Broehan, Gunnar, Tobias Kroeger, Marcé Lorenzen, and Hans Merzendorfer. 2013. “Functional Analysis of the ATP-Binding Cassette (ABC) Transporter Gene Family of Tribolium Castaneum.” BMC Genomics 14 (6). doi:10.1186/1471-2164-14-6.

Brucker, Robert M, Lisa J Funkhouser, Shefali Setia, Rini Pauly, and Seth R Bordenstein. 2012. “Insect Innate Immunity Database (IIID): An Annotation Tool for Identifying Immune Genes in Insect Genomes.” PLoS ONE 7 (9): e45125. doi:10.1371/journal.pone.0045125.

Buchmann, Kurt. 2014. “Evolution of Innate Immunity: Clues from Invertebrates via Fish to Mammals.” Frontiers in Immunology 5 (SEP): 1–8. doi:10.3389/fimmu.2014.00459.

Buchner, Paul. 1965. Endosymbiosis of Animals with Plant Microorganisms. Interscience Publishers.

Budd, Graham E., and Maximilian J. Telford. 2009. “The Origin and Evolution of Arthropods.” Nature 457 (7231): 812–17. doi:10.1038/nature07890.

Burke, Gaelen R, Benjamin B Normark, Colin Favret, and Nancy A Moran. 2009. “Evolution and Diversity of Facultative Symbionts from the Aphid Subfamily Lachninae.” Applied and Environmental Microbiology 75 (16). American Society for Microbiology: 5328–35. doi:10.1128/AEM.00717-09.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. 2009. “BLAST+: Architecture and Applications.” BMC Bioinformatics 10 (1). BioMed Central: 421. doi:10.1186/1471-2105-10-421.

Campbell, Matthew A, James T Van Leuven, Russell C Meister, Kaitlin M Carey, Chris Simon, and John P McCutcheon. 2015. “Genome Expansion via Lineage Splitting and Genome Reduction in the Cicada Endosymbiont Hodgkinia.” Proceedings of the National Academy of Sciences of the United States of America 112 (33). National Academy of Sciences: 10192–99. doi:10.1073/pnas.1421386112.

Cantarel, Brandi L, Ian Korf, Sofia M C Robb, Genis Parra, Eric Ross, Barry Moore, Carson Holt, Alejandro Sánchez Alvarado, and Mark Yandell. 2008. “MAKER: An Easy-to-Use Annotation Pipeline Designed for Emerging Model Organism Genomes.” Genome Research 18 (1). Cold Spring Harbor Laboratory Press: 188–96. doi:10.1101/gr.6743907.

Cao, Xiaolong, Yan He, Yingxia Hu, Yang Wang, Yun Ru Chen, Bart Bryant, Rollie J. Clem, Lawrence M. Schwartz, Gary Blissard, and Haobo Jiang. 2015. “The Immune Signaling Pathways of Manduca Sexta.” Insect Biochemistry and Molecular Biology 62 (April). Elsevier Ltd: 64–74. doi:10.1016/j.ibmb.2015.03.006.

Castresana, J. 2000. “Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis.” Molecular Biology and Evolution 17 (4): 540–52. doi:10.1093/oxfordjournals.molbev.a026334.

Champ, BR, and CE Dyte. 1976. Report on the FAO global survey of pesticide susceptibility of stored grain pests. FAO, Rome, Italy.

Champ, B.R., and C.E Dyte. 1977. FAO global survey of pesticide susceptibility of stored grain pests. FAO Plant Protection Bulletin, Vol. 25, pp. 49-67, ISSN 0254-9727

Chen, De-Qiao, Clytia B. Montllor, and Alexander H. Purcell. 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). John Wiley & Sons, Ltd (10.1111): 315–23. doi:10.1046/j.1570-7458.2000.00670.x.

Chen, Wenbo, Daniel K. Hasegawa, Navneet Kaur, Adi Kliot, Patricia Valle Pinheiro, Junbo Luan, Marcus C. Stensmyr, et al. 2016. “The Draft Genome of Whitefly Bemisia Tabaci MEAM1, a Global Crop Pest, Provides Novel Insights into Virus Transmission, Host Adaptation, and Insecticide Resistance.” BMC Biology 14 (1). BMC Biology: 1–15. doi:10.1186/s12915-016-0321-y.

Cheng, D J, and R F Hou. 2001. “Histological Observations on Transovarial Transmission of a Yeast-like Symbiote in Nilaparvata Lugens Stal (Homoptera, Delphacidae).” Tissue & Cell 33 (3): 273–79.

Christophides, George K, Evgeny Zdobnov, Carolina Barillas-Mury, Ewan Birney, Stephanie Blandin, Claudia Blass, Paul T Brey, et al. 2002. “Immunity-Related Genes and Gene Families in Anopheles Gambiae.” Science 298 (159): 159–65. doi:10.1126/science.1077136.

Clayton, Adam L., Kelly F. Oakeson, Maria Gutin, Arthur Pontes, Diane M. Dunn, Andrew C. von Niederhausern, Robert B. Weiss, Mark Fisher, and Colin Dale. 2012. “A Novel Human-Infection-Derived Bacterium Provides Insights into the Evolutionary Origins of Mutualistic Insect-Bacterial Symbioses.” PLoS Genetics 8 (11). doi:10.1371/journal.pgen.1002990.

Clayton, Adam L., D. Grant Jackson, Robert B. Weiss, and Colin Dale. 2016. “Adaptation by Deletogenic Replication Slippage in a Nascent Symbiont.” Molecular Biology and Evolution 33 (8): 1957–66. doi:10.1093/molbev/msw071.

Clements, Justin, Sean Schoville, Nathan Peterson, Que Lan, and Russell L. Groves. 2016. “Characterizing Molecular Mechanisms of Imidacloprid Resistance in Select Populations of Leptinotarsa Decemlineata in the Central Sands Region of Wisconsin.” Edited by Raul Narciso Carvalho Guedes. PLOS ONE 11 (1). Public Library of Science: e0147844. doi:10.1371/journal.pone.0147844.

Colbourne, J. K., M. E. Pfrender, D. Gilbert, W. K. Thomas, A. Tucker, T. H. Oakley, S. Tokishita, et al. 2011. “The Ecoresponsive Genome of Daphnia Pulex.” Science 331 (6017): 555–61. doi:10.1126/science.1197761.

Conord, Cyrille, Laurence Despres, Agnès Vallier, Séverine Balmand, Christian Miquel, Stéphanie Zundel, Guy Lemperiere, and Abdelaziz Heddi. 2008. “Long-Term Evolutionary Stability of Bacterial Endosymbiosis in Curculionoidea: Additional Evidence of Symbiont Replacement in the Dryophthoridae Family.” Molecular Biology and Evolution 25 (5): 859–68. doi:10.1093/molbev/msn027.

Covassi, M., and A. Binazzi. 1974. Note corologiche emorfologiche sulla Cinara cedriMim. in Italia(Homoptera Aphidoidea Lachnidae).- Redia, LV: 331-341.

Cruz, Fernando, Irene Julca, Jèssica Gómez-Garrido, Damian Loska, Marina Marcet-Houben, Emilio Cano, Beatriz Galán, et al. 2016. “Genome Sequence of the Olive Tree, Olea Europaea.” GigaScience 5 (1). BioMed Central: 29. doi:10.1186/s13742-016-0134-5.

Cui, Xuefeng, Zhiwu Lu, Sheng Wang, Jim Jing-Yan Wang, and Xin Gao. 2016. “CMsearch: Simultaneous Exploration of Protein Sequence Space and Structure Space Improves Not Only Protein Homology Detection but Also Protein Structure Prediction.” Bioinformatics (Oxford, England) 32 (12). Oxford University Press: i332–40. doi:10.1093/bioinformatics/btw271.

Cunchillos, C, and G Lecointre. 2007. “Ordering Events of Biochemical Evolution.” Biochimie 89 (5): 555–73. doi:10.1016/j.biochi.2006.12.007.

Cunningham, Christopher B., Lexiang Ji, R. Axel W. Wiberg, Jennifer Shelton, Elizabeth C. McKinney, Darren J. Parker, Richard B. Meagher, et al. 2015. “The Genome and Methylome of a Beetle with Complex Social Behavior, Nicrophorus Vespilloides (Coleoptera: Silphidae).” Genome Biology and Evolution 7 (12): 3383–96. doi:10.1093/gbe/evv194.

Currie, C. R., R. C. Summerbell, J. a. Scott, and D. Malloch. 1999. “Fungus-Growing Ants Use Antibiotic-Producing Bacteria to Control Garden Parasites.” Nature 398: 701–4.

Dadd, R.H., D.L. Krieger, and T.E. Mittler. 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). Pergamon: 249–72. doi:10.1016/0022-1910(67)90152-7.

Dale, C., S A Young, D T Haydon, and S C Welburn. 2001. “The Insect Endosymbiont Sodalis Glossinidius Utilizes a Type III Secretion System for Cell Invasion.” Proceedings of the National Academy of Sciences 98 (4). National Academy of Sciences: 1883–88. doi:10.1073/pnas.021450998.

De Bary, A. (1879). Die Erscheinung der Symbiose. Strassburg: Verlag Trubner.

De Gregorio, E, P T Spellman, G M Rubin, and B Lemaitre. 2001. “Genome-Wide Analysis of the Drosophila Immune Response by Using Oligonucleotide Microarrays.” Proceedings of the National Academy of Sciences of the United States of America 98 (22): 12590–95. doi:10.1073/pnas.221458698.

De Gregorio, Ennio, Paul T Spellman, Phoebe Tzou, Gerald M Rubin, and Bruno Lemaitre. 2002. “The Toll and Imd Pathways Are the Major Regulators of the Immune Response in Drosophila.” The EMBO Journal 21 (11): 2568–79. doi:10.1093/emboj/21.11.2568.

Delobel, Alex., Maurice. Tran, and O.R.S.T.O.M. (Agency : France). 1993. Les Coléoptères Des Denrées Alimentaires Entreposées Dans Les Régions Chaudes. Orstom/CTA, Institut français de recherche pour le développement en coopération centre technique de coopération agricole et rurale.

Dhami, Manpreet K., Bevan S. Weir, Michael W. Taylor, and Jacqueline R. Beggs. 2013. “Diverse Honeydew-Consuming Fungal Communities Associated with Scale Insects.” Edited by Kurt O. Reinhart. PLoS ONE 8 (7). Public Library of Science: e70316. doi:10.1371/journal.pone.0070316.

DiGuistini, Scott, Nancy Y Liao, Darren Platt, Gordon Robertson, Michael Seidel, Simon K Chan, T Roderick Docking, et al. 2009. “De Novo Genome Sequence Assembly of a Filamentous Fungus Using Sanger, 454 and Illumina Sequence Data.” Genome Biology 10 (9). BioMed Central: R94. doi:10.1186/gb-2009-10-9-r94.

Doron, Shany, Sarah Melamed, Gal Ofir, Azita Leavitt, Anna Lopatina, Mai Keren, Gil Amitai, and Rotem Sorek. 2018. “Systematic Discovery of Antiphage Defense Systems in the Microbial Pangenome.” Science 359 (6379): eaar4120. doi:10.1126/science.aar4120.

Dostert, Catherine, Emmanuelle Jouanguy, Phil Irving, Laurent Troxler, Delphine Galiana-Arnoux, Charles Hetru, Jules A. Hoffmann, and Jean Luc Imler. 2005. “The Jak-STAT Signaling Pathway Is Required but Not Sufficient for the Antiviral Response of Drosophila.” Nature Immunology 6 (9): 946–53. doi:10.1038/ni1237.

Doudoumis, Vangelis, George Tsiamis, Florence Wamwiri, Corey Brelsfoard, Uzma Alam, Emre Aksoy, Stelios Dalaperas, et al. 2012. “Detection and Characterization of Wolbachia Infections in Laboratory and Natural Populations of Different Species of Tsetse Flies (Genus Glossina).” BMC Microbiology 12 (Suppl 1). BioMed Central: S3. doi:10.1186/1471-2180-12-S1-S3.

Douglas, A. E. 1989. “Mycetocyte Symbiosis in Insects.” Biological Reviews 64 (4). John Wiley & Sons, Ltd (10.1111): 409–34. doi:10.1111/j.1469-185X.1989.tb00682.x.

Douglas, A. E. 2006. “Phloem-Sap Feeding by Animals: Problems and Solutions.” Journal of Experimental Botany 57 (4). Oxford University Press: 747–54. doi:10.1093/jxb/erj067.

Dowling, Daniel, Thomas Pauli, Alexander Donath, Karen Meusemann, Lars Podsiadlowski, Malte Petersen, Ralph S. Peters, et al. 2017. “Phylogenetic Origin and Diversification of RNAi Pathway Genes in Insects.” Genome Biology and Evolution 8 (12). Oxford University Press: evw281. doi:10.1093/gbe/evw281.

Duan, Jun, Tim Ladd, Daniel Doucet, Michel Cusson, Kees vanFrankenhuyzen, Omprakash Mittapalli, Peter J. Krell, and Guoxing Quan. 2015. “Transcriptome Analysis of the Emerald Ash Borer (EAB), Agrilus Planipennis: De Novo Assembly, Functional Annotation and Comparative Analysis.” Edited by Kun Yan Zhu. PLOS ONE 10 (8). Public Library of Science: e0134824. doi:10.1371/journal.pone.0134824.

Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” Nucleic Acids Research 32 (5): 1792–97. doi:10.1093/nar/gkh340.

Eleftherianos, Loannis, Jaishri Atri, Julia Accetta, and Julio C. Castillo. 2013. “Endosymbiotic Bacteria in Insects: Guardians of the Immune System?” Frontiers in Physiology 4 MAR (March): 1–10. doi:10.3389/fphys.2013.00046.

Engel, Michael S. 2015. “Insect Evolution.” Current Biology 25 (19). Elsevier: R868–72. doi:10.1016/j.cub.2015.07.059.

Engel, Michael S, and David A Grimaldi. 2004. “New Light Shed on the Oldest Insect.” Nature 427 (February): 627–30. doi:10.1038/nature02334.1.

Engel, Philipp, and Nancy A. Moran. 2013. “The Gut Microbiota of Insects - Diversity in Structure and Function.” FEMS Microbiology Reviews 37 (5): 699–735. doi:10.1111/1574-6976.12025.

English, Adam C., Stephen Richards, Yi Han, Min Wang, Vanesa Vee, Jiaxin Qu, Xiang Qin, et al. 2012. “Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology.” Edited by Zhanjiang Liu. PLoS ONE 7 (11). Public Library of Science: 1–12. doi:10.1371/journal.pone.0047768.

Evans, J D, K Aronstein, Y P Chen, C Hetru, J-L Imler, H Jiang, M Kanost, G J Thompson, Z Zou, and D Hultmark. 2006. “Immune Pathways and Defence Mechanisms in Honey Bees Apis Mellifera.” Insect Molecular Biology 15 (5): 645–56. doi:10.1111/j.1365-2583.2006.00682.x.

Favret, C. 2018. Aphid Species File. Version 5.0/5.0. <http://Aphid.SpeciesFile.org>.

Fenton, Brian, John T. Margaritopoulos, Gaynor L. Malloch, and Stephen P. Foster. 2010. “Micro-Evolutionary Change in Relation to Insecticide Resistance in the Peach-Potato Aphid, Myzus Persicae.” Ecological Entomology 35 (January). John Wiley & Sons, Ltd (10.1111): 131–46. doi:10.1111/j.1365-2311.2009.01150.x.

Fereres, A., M. E. Irwin, and G. E. Kampmeier. 2017. “Aphid Movement: Process and Consequences.” In Aphids as Crop Pests, 196–224. Wallingford: CABI. doi:10.1079/9781780647098.0196.

Fisher, Roberta M., Lee M. Henry, Charlie K. Cornwallis, E. Toby Kiers, and Stuart A. West. 2017. “The Evolution of Host-Symbiont Dependence.” Nature Communications 8 (May). Nature Publishing Group: 1–8. doi:10.1038/ncomms15973.

Franzenburg, Sören, Jonas Walter, Sven Künzel, Jun Wang, John F Baines, Thomas C G Bosch, and Sebastian Fraune. 2013. “Distinct Antimicrobial Peptide Expression Determines Host Species-Specific Bacterial Associations.” Proceedings of the National Academy of Sciences of the United States of America 110 (39). National Academy of Sciences: E3730-8. doi:10.1073/pnas.1304960110.

Frias, L, and P Ribeca. 2018. ASM scripts. https://github.com/lfrias81/anchor-asm/tree/master/wrapper.

Fukatsu T, Aoki S, Kurosu U, Ishikawa H. 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. Zool Sci. 1994;11:613–623.

Galiana-Arnoux, Delphine, Catherine Dostert, Anette Schneemann, Jules A Hoffmann, and Jean-Luc Imler. 2006. “Essential Function in Vivo for Dicer-2 in Host Defense against RNA Viruses in Drosophila.” Nature Immunology 7 (6): 590–97. doi:10.1038/ni1335.

García del Moral, Luis F., Yahia Rharrabti, Vanessa Martos, and Conxita Royo. 2007. “Environmentally Induced Changes in Amino Acid Composition in the Grain of Durum Wheat Grown under Different Water and Temperature Regimes in a Mediterranean Environment.” Journal of Agricultural and Food Chemistry 55 (20): 8144–51. doi:10.1021/jf063094q.

Georgel, P, S Naitza, C Kappler, D Ferrandon, D Zachary, C Swimmer, C Kopczynski, G Duyk, J M Reichhart, and J A Hoffmann. 2001. “Drosophila Immune Deficiency (IMD) Is a Death Domain Protein That Activates Antibacterial Defense and Can Promote Apoptosis.” Developmental Cell 1 (4): 503–14.

Gerardo, Nicole M., Boran Altincicek, Caroline Anselme, Hagop Atamian, Seth M. Barribeau, Martin de Vos, Elizabeth J. Duncan, et al. 2010. “Immunity and Other Defenses in Pea Aphids, Acyrthosiphon Pisum.” Genome Biology 11 (2). doi:10.1186/gb-2010-11-2-r21.

Gil, R., F. J. Silva, E. Zientz, F. Delmotte, F. Gonzalez-Candelas, A. Latorre, C. Rausell, et al. 2003. “The Genome Sequence of Blochmannia Floridanus: Comparative Analysis of Reduced Genomes.” Proceedings of the National Academy of Sciences 100 (16): 9388–93. doi:10.1073/pnas.1533499100.

Gil, Rosario, Eugeni Belda, María J Gosalbes, Luis Delaye, Agnès Vallier, Carole Vincent-Monégat, Abdelaziz Heddi, Francisco J Silva, Andrés Moya, and Amparo Latorre. 2008. “Massive Presence of Insertion Sequences in the Genome of SOPE, the Primary Endosymbiont of the Rice Weevil Sitophilus Oryzae.” International Microbiology 11 (1): 41–48.

Gil, Rosario, Carlos Vargas-Chavez, Sergio López-Madrigal, Diego Santos-García, Amparo Latorre, and Andrés Moya. 2018. “Tremblaya Phenacola PPER: An Evolutionary Beta-Gammaproteobacterium Collage.” ISME Journal 12 (1). Nature Publishing Group: 124–35. doi:10.1038/ismej.2017.144.

Gómez-Valero, Laura, Mario Soriano-Navarro, Vicente Pérez-Brocal, Abdelaziz Heddi, Andrés Moya, José Manuel García-Verdugo, and Amparo Latorre. 2004. “Coexistence of Wolbachia with Buchnera Aphidicola and a Secondary Symbiont in the Aphid Cinara Cedri.” Journal of Bacteriology 186 (19). American Society for Microbiology Journals: 6626–33. doi:10.1128/JB.186.19.6626-6633.2004.

Gore, J, D Cook, A Catchot, B R Leonard, S D Stewart, G Lorenz, and D Kerns. 2013. “Cotton Aphid (Heteroptera: Aphididae) Susceptibility to Commercial and Experimental Insecticides in the Southern United States.” Journal of Economic Entomology 106 (3): 1430–39.

Gosalbes, M. J., A. Lamelas, A. Moya, and A. Latorre. 2008. “The Striking Case of Tryptophan Provision in the Cedar Aphid Cinara Cedri.” Journal of Bacteriology 190 (17): 6026–29. doi:10.1128/JB.00525-08.

Goubert, Clément, Laurent Modolo, Cristina Vieira, Claire Valiente Moro, Patrick Mavingui, and Matthieu Boulesteix. 2015. “De Novo Assembly and Annotation of the Asian Tiger Mosquito (Aedesalbopictus) Repeatome with DnaPipeTE from Raw Genomic Reads and Comparative Analysis with the Yellow Fever Mosquito (Aedes Aegypti).” Genome Biology and Evolution 7 (4): 1192–1205. doi:10.1093/gbe/evv050.

Grbić, Miodrag, Thomas Van Leeuwen, Richard M. Clark, Stephane Rombauts, Pierre Rouzé, Vojislava Grbić, Edward J. Osborne, et al. 2011. “The Genome of Tetranychus Urticae Reveals Herbivorous Pest Adaptations.” Nature 479 (7374). Nature Publishing Group: 487–92. doi:10.1038/nature10640.

Grenier, Anne-Marie, Mbailao Mbaiguinam, and Bernard Delobel. 1997. “Genetical Analysis of the Ability of the Rice Weevil Sitophilus Oryzae (Coleoptera, Curculionidae) to Breed on Split Peas.” Heredity 79: 15–23.

Grimaldi, David, and Michael S. Engel. 2005. Evolution of the Insects. Cambridge University Press.

Grimaldi, David A. 2010. “400 Million Years on Six Legs: On the Origin and Early Evolution of Hexapoda.” Arthropod Structure and Development 39 (2–3). Elsevier Ltd: 191–203. doi:10.1016/j.asd.2009.10.008.

Guerrieri, Emilio, and Maria Cristina Digilio. 2008. “Aphid-Plant Interactions: A Review.” Journal of Plant Interactions 3 (4): 223–32. doi:10.1080/17429140802567173.

Guindon, Stéphane, Jean Franois Dufayard, Vincent Lefort, Maria Anisimova, Wim Hordijk, and Olivier Gascuel. 2010. “New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.” Systematic Biology 59 (3): 307–21. doi:10.1093/sysbio/syq010.

Guo, Jianqing, Séverin Hatt, Kanglai He, Julian Chen, Frédéric Francis, and Zhenying Wang. 2017. “Nine Facultative Endosymbionts in Aphids. A Review.” Journal of Asia-Pacific Entomology. doi:10.1016/j.aspen.2017.03.025.

Gupta, Shishir K., Maria Kupper, Carolin Ratzka, Heike Feldhaar, Andreas Vilcinskas, Roy Gross, Thomas Dandekar, and Frank Förster. 2015. “Scrutinizing the Immune Defence Inventory of Camponotus Floridanus Applying Total Transcriptome Sequencing.” BMC Genomics 16 (1): 1–21. doi:10.1186/s12864-015-1748-1.

Haas, Brian J, Steven L Salzberg, Wei Zhu, Mihaela Pertea, Jonathan E Allen, Joshua Orvis, Owen White, C Robin Buell, and Jennifer R Wortman. 2008. “Automated Eukaryotic Gene Structure Annotation Using EVidenceModeler and the Program to Assemble Spliced Alignments.” Genome Biology 9 (1): R7. doi:10.1186/gb-2008-9-1-r7.

Haase, Astrid D. 2016. “A Small RNA-Based Immune System Defends Germ Cells against Mobile Genetic Elements.” Stem Cells International 2016. Hindawi Limited: 7595791. doi:10.1155/2016/7595791.

Hackl, Thomas, Rainer Hedrich, Jörg Schultz, and Frank Förster. 2014. “Proovread: Large-Scale High-Accuracy PacBio Correction through Iterative Short Read Consensus.” Bioinformatics 30 (21): 3004–11. doi:10.1093/bioinformatics/btu392.

Hahn, Christoph, Lutz Bachmann, and Bastien Chevreux. 2013. “Reconstructing Mitochondrial Genomes Directly from Genomic Next-Generation Sequencing Reads - A Baiting and Iterative Mapping Approach.” Nucleic Acids Research 41 (13). Oxford University Press: e129. doi:10.1093/nar/gkt371.

Hammond, Peter. 1992. “Species Inventory.” Global Biodiversity. Dordrecht: Springer Netherlands, 17–39. doi:10.1007/978-94-011-2282-5_4.

Han, Bo W., Jui-Hung Hung, Zhiping Weng, Phillip D. Zamore, and Stefan L. Ameres. 2011. “The 3′-to-5′ Exoribonuclease Nibbler Shapes the 3′ Ends of MicroRNAs Bound to Drosophila Argonaute1.” Current Biology 21 (22). Cell Press: 1878–87. doi:10.1016/J.CUB.2011.09.034.

Haney, Cara H, Jonathan M Urbach, and Frederick M Ausubel. 2014. “Differences and Similarities: Innate Immunity in Plants and Animals.” The Biochemist 36 (5): 40–44.

Hansen, Allison K., and Nancy A. Moran. 2014. “The Impact of Microbial Symbionts on Host Plant Utilization by Herbivorous Insects.” Molecular Ecology 23: 1473–96. doi:10.1111/mec.12421.

Hardman, J. M. 1977. “Environmental Changes Associated with the Growth of Populations of Sitophilus Oryzae (L.) Confined in Small Cells of Wheat.” Journal of Stored Products Research 13 (2): 45–52. doi:10.1016/0022-474X(77)90057-1.

Harrison, Mark C., Evelien Jongepier, Hugh M. Robertson, Nicolas Arning, Tristan Bitard-Feildel, Hsu Chao, Christopher P. Childers, et al. 2018. “Hemimetabolous Genomes Reveal Molecular Basis of Termite Eusociality.” Nature Ecology and Evolution 2 (3). Springer US: 557–66. doi:10.1038/s41559-017-0459-1.

Heddi, Abdelaziz, Hubert Charles, Chaqué Khatchadourian, Guy Bonnot, and Paul Nardon. 1998. “Molecular Characterization of the Principal Symbiotic Bacteria of the Weevil Sitophilus Oryzae: A Peculiar G + C Content of an Endocytobiotic DNA.” Journal of Molecular Evolution 47 (1): 52–61. doi:10.1007/PL00006362.

Heddi, Abdelaziz, Anne-Marie Grenier, Chaqué Khatchadourian, Hubert Charles, and Paul Nardon. 1999. “Four Intracellular Genomes Direct Weevil Biology: Nuclear, Mitochondrial, Principal Endosymbiont, and Wolbachia.” Evolution 96 (June): 6814–19. doi:10.1073/pnas.96.12.6814.

Heddi, A, H Charles, and C Khatchadourian. 2001. “Intracellular Bacterial Symbiosis in the Genus Sitophilus: The ‘biological Individual’ Concept Revisited.” Research in Microbiology 152 (5): 431–37.

Heie OE. 1996. The evolutionary history of aphids and a hypothesis on the coevolution of aphids and plants. Bollettino di Zoologia agraria e di Bachicoltura, Series II 28: 149–155.

Heie, Ole E. 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): 275–87. doi:10.1515/pjen-2015-0024.

Hoffmann, Jules a. 2003. “The Immune Response of Drosophila.” Nature 426 (6962): 33–38. doi:10.1038/nature02021.

Holt, Robert A., G. Mani Subramanian, Aaron Halpern, Granger G. Sutton, Rosane Charlab, Deborah R. Nusskern, Patrick Wincker, et al. 2002. “The Genome Sequence of the Malaria Mosquito Anopheles Gambiae.” Science 298 (5591): 129–49. doi:10.1126/science.1076181.

Hong, Youchong, Zhijun Zhang, Xinrong Guo, and O. E. Heie. 2009. “A New Species Representing the Oldest Aphid (Hemiptera, Aphidomorpha) from the Middle Triassic of China.” J. Paleont 83 (5): 826–31.

Huang, Xu, Yunshan Huang, Jingyu Zhang, Fang Lu, Jing Wei, and Mingxing Jiang. 2016. “The Symbiotic Bacteria Nardonella in Rice Water Weevil (Coleoptera: Curculionidae): Diversity, Density, and Associations With Host Reproduction.” Annals of the Entomological Society of America 109 (3): 415–23. doi:10.1093/aesa/saw015.

Huerta-Cepas, Jaime, Salvador Capella-Gutiérrez, Leszek P. Pryszcz, Marina Marcet-Houben, and Toni Gabaldón. 2014. “PhylomeDB v4: Zooming into the Plurality of Evolutionary Histories of a Genome.” Nucleic Acids Research 42 (D1): 897–902. doi:10.1093/nar/gkt1177.

Huerta-Cepas, Jaime, Damian Szklarczyk, Kristoffer Forslund, Helen Cook, Davide Heller, Mathias C. Walter, Thomas Rattei, et al. 2016a. “EGGNOG 4.5: A Hierarchical Orthology Framework with Improved Functional Annotations for Eukaryotic, Prokaryotic and Viral Sequences.” Nucleic Acids Research 44 (D1): D286–93. doi:10.1093/nar/gkv1248.

Huerta-Cepas, Jaime, François Serra, and Peer Bork. 2016b. “ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data.” Molecular Biology and Evolution 33 (6): 1635–38. doi:10.1093/molbev/msw046.

Hunt, Toby, Johannes Bergsten, Zuzana Levkanicova, Anna Papadopoulou, Oliver St John, Ruth Wild, Peter M Hammond, et al. 2007. “A Comprehensive Phylogeny of Beetles Reveals the Evolutionary Origins of a Superradiation.” Science 318 (5858): 1913–16. doi:10.1126/science.1146954.

Hunt, Martin, Taisei Kikuchi, Mandy Sanders, Chris Newbold, Matthew Berriman, and Thomas D Otto. 2013. “REAPR: A Universal Tool for Genome Assembly Evaluation.” Genome Biology 14 (5). BioMed Central: R47. doi:10.1186/gb-2013-14-5-r47.

Hunter, W. B, Reese, J., & International Psyllid Genome Consortium, T. 2014. The Asian Citrus Psyllid Genome (Diaphorina citri, Hemiptera). Journal of Citrus Pathology, 1(1).

Husnik, Filip, and John P. McCutcheon. 2016. “Repeated Replacement of an Intrabacterial Symbiont in the Tripartite Nested Mealybug Symbiosis.”

Proceedings of the National Academy of Sciences 113 (37): E5416–24. doi:10.1073/pnas.1603910113.

Husnik, Filip, Naruo Nikoh, Ryuichi Koga, Laura Ross, Rebecca P. Duncan, Manabu Fujie, Makiko Tanaka, et al. 2013. “Horizontal Gene Transfer from Diverse Bacteria to an Insect Genome Enables a Tripartite Nested Mealybug Symbiosis.” Cell 153 (7): 1567–78. doi:10.1016/j.cell.2013.05.040.

Igaki, Tatsushi, Hiroshi Kanda, Yuki Yamamoto-Goto, Hirotaka Kanuka, Erina Kuranaga, Toshiro Aigaki, and Masayuki Miura. 2002. “Eiger, a TNF Superfamily Ligand That Triggers the Drosophila JNK Pathway.” The EMBO Journal 21 (12). European Molecular Biology Organization: 3009–18. doi:10.1093/emboj/cdf306.

International Glossina Genome Initiative, International Glossina Genome. 2014. “Genome Sequence of the Tsetse Fly (Glossina Morsitans): Vector of African Trypanosomiasis.” Science (New York, N.Y.) 344 (6182). American Association for the Advancement of Science: 380–86. doi:10.1126/science.1249656.

Jamnongluk, Wanwisa, Pattamaporn Kittayapong, Visut Baimai, and Scott L. O’Neill. 2002. “Wolbachia Infections of Tephritid Fruit Flies: Molecular Evidence for Five Distinct Strains in a Single Host Species.” Current Microbiology 45 (4): 255–60. doi:10.1007/s00284-002-3746-1.

Jiang, C, E H Baehrecke, and C S Thummel. 1997. “Steroid Regulated Programmed Cell Death during Drosophila Metamorphosis.” Development (Cambridge, England) 124 (22): 4673–83.

Johnson, Andrew D., Robert E. Handsaker, Sara L. Pulit, Marcia M. Nizzari, Christopher J. O’Donnell, and Paul I W De Bakker. 2008. “SNAP: A Web-Based Tool for Identification and Annotation of Proxy SNPs Using HapMap.” Bioinformatics 24 (24): 2938–39. doi:10.1093/bioinformatics/btn564.

Jones, Andrew K., and David B. Sattelle. 2007. “The Cys-Loop Ligand-Gated Ion Channel Gene Superfamily of the Red Flour Beetle, Tribolium Castaneum.” BMC Genomics 8: 1–16. doi:10.1186/1471-2164-8-327.

Jones, Philip, David Binns, Hsin Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, et al. 2014. “InterProScan 5: Genome-Scale Protein Function Classification.” Bioinformatics 30 (9): 1236–40. doi:10.1093/bioinformatics/btu031.

Kajitani, Rei, Kouta Toshimoto, Hideki Noguchi, Atsushi Toyoda, Yoshitoshi Ogura, Miki Okuno, Mitsuru Yabana, et al. 2014. “Efficient de Novo Assembly of Highly Heterozygous Genomes from Whole-Genome Shotgun Short Reads.” Genome Research 24 (8). Cold Spring Harbor Laboratory Press: 1384–95. doi:10.1101/gr.170720.113.

Kanehisa, Minoru, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. 2016. “KEGG as a Reference Resource for Gene and Protein Annotation.” Nucleic Acids Research 44 (D1): D457–62. doi:10.1093/nar/gkv1070.

Kang, Jae Soon, Yong-Jun Cho, Ju Hyeon Kim, Sang Hyeon Kim, Seungil Yoo, Seung-Jae Noh, Junhyung Park, et al. 2015. “Comparison of the Genome Profiles between Head and Body Lice.” Journal of Asia-Pacific Entomology 18 (3). Elsevier: 377–82. doi:10.1016/J.ASPEN.2015.04.010.

Kanost, Michael R., Estela L. Arrese, Xiaolong Cao, Yun-Ru Chen, Sanjay Chellapilla, Marian R. Goldsmith, Ewald Grosse-Wilde, et al. 2016. “Multifaceted Biological Insights from a Draft Genome Sequence of the Tobacco Hornworm Moth, Manduca Sexta.” Insect Biochemistry and Molecular Biology 76 (September): 118–47. doi:10.1016/j.ibmb.2016.07.005.

Karlikow, Margot, Bertsy Goic, and Maria-Carla Saleh. 2014. “RNAi and Antiviral Defense in Drosophila: Setting up a Systemic Immune Response.” Developmental & Comparative Immunology 42 (1): 85–92. doi:10.1016/j.dci.2013.05.004.

Keeling, Christopher I, Macaire MS Yuen, Nancy Y Liao, T Roderick Docking, Simon K Chan, Greg A Taylor, Diana L Palmquist, et al. 2013. “Draft Genome of the Mountain Pine Beetle, Dendroctonus Ponderosae Hopkins, a Major Forest Pest.” Genome Biology 14 (3). BioMed Central: R27. doi:10.1186/gb-2013-14-3-r27.

Kirkness, E. F., B. J. Haas, W. Sun, H. R. Braig, M. A. Perotti, J. M. Clark, S. H. Lee, et al. 2010. “Genome Sequences of the Human Body Louse and Its Primary Endosymbiont Provide Insights into the Permanent Parasitic Lifestyle.” Proceedings of the National Academy of Sciences 107 (27): 12168–73. doi:10.1073/pnas.1003379107.

Kirsch, Roy, Lydia Gramzow, Günter Theißen, Blair D. Siegfried, Richard H. Ffrench-Constant, David G. Heckel, and Yannick Pauchet. 2014. “Horizontal Gene Transfer and Functional Diversification of Plant Cell Wall Degrading Polygalacturonases: Key Events in the Evolution of Herbivory in Beetles.” Insect Biochemistry and Molecular Biology 52 (1). Elsevier Ltd: 33–50. doi:10.1016/j.ibmb.2014.06.008.

Kirsch, Roy, David G. Heckel, and Yannick Pauchet. 2016. “How the Rice Weevil Breaks down the Pectin Network: Enzymatic Synergism and Sub-Functionalization.” Insect Biochemistry and Molecular Biology 71. Elsevier Ltd: 72–82. doi:10.1016/j.ibmb.2016.02.007.

Klasson, Lisa, and Siv G. E. Andersson. 2006. “Strong Asymmetric Mutation Bias in Endosymbiont Genomes Coincide with Loss of Genes for Replication Restart Pathways.” Molecular Biology and Evolution 23 (5): 1031–39. doi:10.1093/molbev/msj107.

Koehler, Philip Gene. 1994. “Rice Weevil, Sitophilus Oryzae (Coleoptera: Curculionidae).” Entomology and Nematology Department, Florida Cooperative Extension Services, Institute of Food and Agricultural Sciences Series No. ENY-261, University of Florida.

Koga, R., T. Tsuchida, and T. Fukatsu. 2003. “Changing Partners in an Obligate Symbiosis: A Facultative Endosymbiont Can Compensate for Loss of the Essential Endosymbiont Buchnera in an Aphid.” Proceedings of the Royal Society of London. Series B: Biological Sciences 270 (1533): 2543–50. doi:10.1098/rspb.2003.2537.

Kramer, K.J., Morgan, T.D., Hopkins, T.L., Christensen, A.M., Schaefer, J., 1989. Solidstate 13C-NMR and diphenol analyses of sclerotized cuticles from stored product Coleoptera. Insect Biochemistry 19, 753–757.

Krueger F. 2015. Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Ku, Hsueh-Yen, and Haifan Lin. 2014. “PIWI Proteins and Their Interactors in PiRNA Biogenesis, Germline Development and Gene Expression.” National Science Review 1 (2): 205–18. doi:10.1093/nsr/nwu014.

Kundoo, Ajaz Ahmad, and Akhtar Ali Khan. 2017. “Coccinellids as Biological Control Agents of Soft Bodied Insects: A Review.” Journal of Entomology and Zoology Studies 5 (5): 1362–73.

Kuriwada, Takashi, Takahiro Hosokawa, Norikuni Kumano, Keiko Shiromoto, Dai Haraguchi, and Takema Fukatsu. 2010. “Biological Role of Nardonella Endosymbiont in Its Weevil Host.” PloS ONE 5 (10). doi:10.1371/journal.pone.0013101.

Ladang, Y D, L T S Ngamo, M B Ngassoum, P M Mapongmestsem, and T Hance. 2008. “Effect of Sorghum Cultivars on Population Growth and Grain Damages by the Rice Weevil, Sitophilus Oryzae L. (Coleoptera: Curculionidae).” African Journal of Agricultural Research 3 (2): 255–58.

Lamelas, Araceli, María José Gosalbes, Alejandro Manzano-Marín, Juli Peretó, Andrés Moya, and Amparo Latorre. 2011. “Serratia Symbiotica from the Aphid Cinara Cedri: A Missing Link from Facultative to Obligate Insect Endosymbiont.” Edited by Paul M. Richardson. PLoS Genetics 7 (11). Public Library of Science: e1002357. doi:10.1371/journal.pgen.1002357.

Latorre, A., A. Moya, and F. J. Ayala. 1986. Evolution of mitochondrial DNA in Drosophila subobscura. Proc. Natl. Acad. Sci. U.S.A. 83 8649–8653. 10.1073/pnas.83.22.8649

Leclerc, Vincent, and Jean-Marc Reichhart. 2004. “The Immune Response of Drosophila Melanogaster.” Immunological Reviews 198 (April): 59–71.

Lee, Young Sik, Sigal Pressman, Arlise P Andress, Kevin Kim, Jamie L White, Justin J Cassidy, Xin Li, et al. 2009. “Silencing by Small RNAs Is Linked to Endosomal Trafficking.” Nature Cell Biology 11 (9). NIH Public Access: 1150–56. doi:10.1038/ncb1930.

Lee, Eduardo, Gregg A. Helt, Justin T. Reese, Monica C. Munoz-Torres, Chris P. Childers, Robert M. Buels, Lincoln Stein, Ian H. Holmes, Christine G. Elsik, and Suzanna E. Lewis. 2013. “Web Apollo: A Web-Based Genomic Annotation Editing Platform.” Genome Biology 14 (8). doi:10.1186/gb-2013-14-8-r93.

Lefèvre, Cédric, Hubert Charles, Agnès Vallier, Bernard Delobel, Brian Farrell, and Abdelaziz Heddi. 2004. “Endosymbiont Phylogenesis in the Dryophthoridae Weevils: Evidence for Bacterial Replacement.” Molecular Biology and Evolution 21 (6): 965–73. doi:10.1093/molbev/msh063.

Lemaitre, Bruno, and Jules Hoffmann. 2007. “The Host Defense of Drosophila Melanogaster.” Annual Review of Immunology 25 (1): 697–743. doi:10.1146/annurev.immunol.25.022106.141615.

Li, Sheng, Shiming Zhu, Qiangqiang Jia, Dongwei Yuan, Chonghua Ren, Kang Li, Suning Liu, et al. 2018. “The Genomic and Functional Landscapes of Developmental Plasticity in the American Cockroach.” Nature Communications 9 (1). Nature Publishing Group: 1008. doi:10.1038/s41467-018-03281-1.

Liang, C., G. Brookhart, G. H. Feng, G. R. Reeck, and K. J. Kramer. 1991. “Inhibition of Digestive Proteinase of Stored Grain Coleoptera by Oryzacystatin, a Cysteine Proteinase Inhibitor from Rice Seed.” FEBS Letters 278 (2): 139–42. doi:10.1016/0014-5793(91)80102-9.

Login, F. H., S. Balmand, A. Vallier, C. Vincent-Monegat, A. Vigneron, M. Weiss-Gayet, D. Rochat, and A. Heddi. 2011. “Antimicrobial Peptides Keep Insect Endosymbionts Under Control.” Science 334 (6054): 362–65. doi:10.1126/science.1209728.

Lopes, Denilce Meneses, C. R. de Carvalho, W. R. Clarindo, Milene Miranda Praça, and Mara Garcia Tavares. 2009. “Genome Size Estimation of Three Stingless Bee Species (Hymenoptera, Meliponinae) by Flow Cytometry.” Apidologie 40 (5): 517–23. doi:10.1051/apido/2009030.

López-Madrigal, Sergio, Amparo Latorre, Manuel Porcar, Andrés Moya, and Rosario Gil. 2011. “Complete Genome Sequence of "Candidatus Tremblaya Princeps" Strain PCVAL, an Intriguing Translational Machine below the Living-Cell Status.” Journal of Bacteriology 193 (19). American Society for Microbiology (ASM): 5587–88. doi:10.1128/JB.05749-11.

Lowe, Todd M., and Sean R. Eddy. 1996. “TRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence.” Nucleic Acids Research 25 (5): 955–64. doi:10.1093/nar/25.5.0955.

Lozano-Fernandez, Jesus, Robert Carton, Alastair R. Tanner, Mark N. Puttick, Mark Blaxter, Jakob Vinther, Jørgen Olesen, Gonzalo Giribet, Gregory D. Edgecombe, and Davide Pisani. 2016. “A Molecular Palaeobiological Exploration of Arthropod Terrestrialization.” Philosophical Transactions of the Royal Society B: Biological Sciences 371 (1699): 20150133. doi:10.1098/rstb.2015.0133.

Lü, Feng Gong, Kai Yun Fu, Qian Li, Wen Chao Guo, Tursun Ahmat, and Guo Qing Li. 2015. “Identification of Carboxylesterase Genes and Their Expression Profiles in the Colorado Potato Beetle Leptinotarsa Decemlineata Treated with Fipronil and Cyhalothrin.” Pesticide Biochemistry and Physiology 122: 86–95. doi:10.1016/j.pestbp.2014.12.015.

Lucas, Keira J., Kevin M. Myles, and Alexander S. Raikhel. 2013. “Small RNAs: A New Frontier in Mosquito Biology.” Trends in Parasitology 29 (6): 295–303. doi:10.1016/j.pt.2013.04.003.

Mahlapuu, Margit, Joakim Håkansson, Lovisa Ringstad, and Camilla Björn. 2016. “Antimicrobial Peptides: An Emerging Category of Therapeutic Agents.” Frontiers in Cellular and Infection Microbiology 6. Frontiers Media SA: 194. doi:10.3389/fcimb.2016.00194.

Maire, Justin, Carole Vincent-Monégat, Florent Masson, Anna Zaidman-Rémy, and Abdelaziz Heddi. 2018. “An IMD-like Pathway Mediates Both Endosymbiont Control and Host Immunity in the Cereal Weevil Sitophilus Spp.” Microbiome 6 (1). Microbiome: 1–10. doi:10.1186/s40168-017-0397-9.

Manzano-Marín, Alejandro, and Amparo Latorre. 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): 1683–98. doi:10.1093/gbe/evu133.

Manzano-Marín, Alejandro, Jean-Christophe Simon, and Amparo Latorre. 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): 1440–58. doi:10.1093/gbe/evw085.

Manzano-Marín, Alejandro, Gitta Szabó, Jean-Christophe Simon, Matthias Horn, and Amparo Latorre. 2017. “Happens in the Best of Subfamilies: Establishment and Repeated Replacements of Co-Obligate Secondary Endosymbionts within Lachninae Aphids.” Environmental Microbiology 19 (1): 393–408. doi:10.1111/1462-2920.13633.

Marçais, Guillaume, and Carl Kingsford. 2011. “A Fast, Lock-Free Approach for Efficient Parallel Counting of Occurrences of k-Mers.” Bioinformatics 27 (6): 764–70. doi:10.1093/bioinformatics/btr011.

Marco-Sola, Santiago, Michael Sammeth, Roderic Guigó, and Paolo Ribeca. 2012. “The GEM Mapper: Fast, Accurate and Versatile Alignment by Filtration.” Nature Methods 9 (12): 1185–88. doi:10.1038/nmeth.2221.

Martin, Bradford D.;, and Ernest Schwab. 2012a. “Symbiosis: ‘Living Together’ in Chaos.” Studies in the History of Biology 4 (4): 7–25.

Martin, Bradford D., and Ernest Schwab. 2012b. “Current Usage of Symbiosis and Associated Terminology.” International Journal of Biology 5 (1): 32–45. doi:10.5539/ijb.v5n1p32.

Martin, Marcel. 2011. “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads.” EMBnet.Journal 17 (1): 10. doi:10.14806/ej.17.1.200.

Martinez-Torres, David, Celia Buades, Amparo Latorre, and Andres Moya. 2001. “Molecular Systematics of Aphids and Their Primary Endosymbionts.” Molecular Phylogenetics and Evolution 20 (3): 437–49. doi:10.1006/mpev.2001.0983.

Masson, Florent, Yves Moné, Aurélien Vigneron, Agnès Vallier, Nicolas Parisot, Carole Vincent-Monégat, Séverine Balmand, Marie-Christine Carpentier, Anna Zaidman-Rémy, and Abdelaziz Heddi. 2015. “Weevil Endosymbiont Dynamics Is Associated with a Clamping of Immunity.” BMC Genomics 16 (1). BioMed Central: 819. doi:10.1186/s12864-015-2048-5.

Masson, Florent, Anna Zaidman-Rémy, and Abdelaziz Heddi. 2016. “Antimicrobial Peptides and Cell Processes Tracking Endosymbiont Dynamics.” Philosophical Transactions of the Royal Society B: Biological Sciences 371 (1695). doi:10.1098/rstb.2015.0298.

Mathers, Thomas C., Yazhou Chen, Gemy Kaithakottil, Fabrice Legeai, Sam T. Mugford, Patrice Baa-Puyoulet, Anthony Bretaudeau, et al. 2017. “Rapid Transcriptional Plasticity of Duplicated Gene Clusters Enables a Clonally Reproducing Aphid to Colonise Diverse Plant Species.” Genome Biology 18 (1). BioMed Central: 27. doi:10.1186/s13059-016-1145-3.

McCutcheon, John P., and Nancy A. Moran. 2012. “Extreme Genome Reduction in Symbiotic Bacteria.” Nature Reviews Microbiology 10 (1): 13–26. doi:10.1038/nrmicro2670.

McCutcheon, John P., and Carol D. von Dohlen. 2011. “An Interdependent Metabolic Patchwork in the Nested Symbiosis of Mealybugs.” Current Biology 21 (16): 1366–72. doi:10.1016/j.cub.2011.06.051.

McCutcheon, John P., Bradon R. McDonald, and Nancy A. Moran. 2009. “Origin of an Alternative Genetic Code in the Extremely Small and GC–Rich Genome of a Bacterial Symbiont.” Edited by Ivan Matic. PLoS Genetics 5 (7). Public Library of Science: e1000565. doi:10.1371/journal.pgen.1000565.

McKenna, Duane D, and Brian D Farrell. 2009. “Beetles (Coleoptera).” In The Timetree of Life, edited by S. Blair Hedges and Sudhir Kumar, 278–89. Oxford University Press.

Mckenna, Duane D., Alexander L. Wild, Kojun Kanda, Charles L. Bellamy, Rolf G. Beutel, Michael S. Caterino, Charles W. Farnum, et al. 2015. “The Beetle Tree of Life Reveals That Coleoptera Survived End-Permian Mass Extinction to Diversify during the Cretaceous Terrestrial Revolution.” Systematic Entomology 40 (4): 835–80. doi:10.1111/syen.12132.

McKenna, Duane D., Erin D. Scully, Yannick Pauchet, Kelli Hoover, Roy Kirsch, Scott M. Geib, Robert F. Mitchell, et al. 2016. “Genome of the Asian Longhorned Beetle (Anoplophora Glabripennis), a Globally Significant Invasive Species, Reveals Key Functional and Evolutionary Innovations at the Beetle–plant Interface.” Genome Biology 17 (1). BioMed Central: 227. doi:10.1186/s13059-016-1088-8.

Medved, Victor, James H. Marden, Howard W. Fescemyer, Joshua P. Der, Jin Liu, Najmus Mahfooz, and Aleksandar Popadić. 2015. “Origin and Diversification of Wings: Insights from a Neopteran Insect.” Proceedings of the National Academy of Sciences 112 (52): 15946–51. doi:10.1073/pnas.1509517112.

Meseguer, Andrea S., Armelle Coeur d’acier, Gwenaelle Genson, and Emmanuelle Jousselin. 2015. “Unravelling the Historical Biogeography and Diversification Dynamics of a Highly Diverse Conifer-Feeding Aphid Genus.” Journal of Biogeography 42 (8). John Wiley & Sons, Ltd (10.1111): 1482–92. doi:10.1111/jbi.12531.

Michelena, J. M., Fabienne Assael, and Z. Mendel. 2005. “Description of Pauesia (Pauesia) Anatolica (Hymenoptera: Braconidae, Aphidiinae) Sp. Nov., a Parasitoid of the Cedar Aphid Cinara Cedri.” Phytoparasitica 33 (5). Springer Netherlands: 499–505. doi:10.1007/BF02981399.

Mills, K. A. 2001. “Phosphine Resistance: Where To Now ?” In Proc. Int. Conf. Controlled Atmosphere and Fumigation in Stored Products, edited by E. J. Donahaye, S. Navarro, and J. G. Leesch, 583–91. Fresno, CA: Executive Printing Services.

Misof, Bernhard, Shanlin Liu, Karen Meusemann, Ralph S. Peters, Alexander Donath, Christoph Mayer, Paul B. Frandsen, et al. 2014. “Phylogenomics Resolves the Timing and Pattern of Insect Evolution.” Science 346 (6210): 763–67. doi:10.1126/science.1257570.

Mita, Kazuei, Masahiro Kasahara, Shin Sasaki, Yukinobu Nagayasu, Tomoyuki Yamada, Hiroyuki Kanamori, Nobukazu Namiki, et al. 2004. “The Genome Sequence of Silkworm, Bombyx Mori.” DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes 11 (1): 27–35.

Mittler, T.E., and R.H. Dadd. 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). Pergamon: 741–57. doi:10.1016/0022-1910(63)90017-9.

Montllor, Clytia B., Amy Maxmen, and Alexander H. Purcell. 2002. “Facultative Bacterial Endosymbionts Benefit Pea Aphids Acyrthosiphon Pisum under Heat Stress.” Ecological Entomology 27 (2). John Wiley & Sons, Ltd (10.1111): 189–95. doi:10.1046/j.1365-2311.2002.00393.x.

Moran, N. A., and T. Jarvik. 2010. “Lateral Transfer of Genes from Fungi Underlies Carotenoid Production in Aphids.” Science 328 (5978): 624–27. doi:10.1126/science.1187113.

Moran, Nancy A., Mark A. Munson, Paul Baumann, and Hajime Ishikawa. 1993. “A Molecular Clock in Endosymbiotic Bacteria Is Calibrated Using the Insect Hosts.” Proceedings of the Royal Society of London. Series B: Biological Sciences 253 (1337): 167–71. doi:10.1098/rspb.1993.0098.

Moran, Nancy A., John P. McCutcheon, and Atsushi Nakabachi. 2008. “Genomics and Evolution of Heritable Bacterial Symbionts.” Annual Review of Genetics 42 (1): 165–90. doi:10.1146/annurev.genet.41.110306.130119.

Moran, N A. 1996. “Accelerated Evolution and Muller’s Rachet in Endosymbiotic Bacteria.” Proceedings of the National Academy of Sciences of the United States of America 93 (7). National Academy of Sciences: 2873–78. doi:10.1073/PNAS.93.7.2873.

Moya, Andrés, Juli Peretó, Rosario Gil, and Amparo Latorre. 2008. “Learning How to Live Together: Genomic Insights into Prokaryote-Animal Symbioses.” Nature Reviews Genetics 9 (3): 218–29. doi:10.1038/nrg2319.

Müller, Werner E G, and Isabel M Müller. 2003. “Origin of the Metazoan Immune System: Identification of the Molecules and Their Functions in Sponges.” Integrative and Comparative Biology 43 (2): 281–92. doi:10.1093/icb/43.2.281.

Munson, M. A., P. Baumann, and M. G. Kinsey. 1991. “Buchnera Gen. Nov. and Buchnera Aphidicola Sp. Nov., a Taxon Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Aphids.” International Journal of Systematic Bacteriology 41 (4). Microbiology Society: 566–68. doi:10.1099/00207713-41-4-566.

Mutti, Navdeep S., Yoonseong Park, John C. Reese, and Gerald R. Reeck. 2006. “RNAi Knockdown of a Salivary Transcript Leading to Lethality in the Pea Aphid, Acyrthosiphon Pisum.” Journal of Insect Science 6 (38): 1–7. doi:10.1673/031.006.3801.

Myllymaki, H., S. Valanne, and M. Ramet. 2014. “The Drosophila Imd Signaling Pathway.” The Journal of Immunology 192 (8): 3455–62. doi:10.4049/jimmunol.1303309.

Mylonakis, Eleftherios, Lars Podsiadlowski, Maged Muhammed, and Andreas Vilcinskas. 2016. “Diversity, Evolution and Medical Applications of Insect Antimicrobial Peptides.” Philosophical Transactions of the Royal Society B: Biological Sciences 371 (1695). The Royal Society. doi:10.1098/RSTB.2015.0290.

Nakabachi, Atsushi, Reiko Ueoka, Kenshiro Oshima, Roberta Teta, Alfonso Mangoni, Mihaela Gurgui, Neil J. Oldham, et al. 2013. “Defensive Bacteriome Symbiont with a Drastically Reduced Genome.” Current Biology 23 (15): 1478–84. doi:10.1016/j.cub.2013.06.027.

Nardon P. 1973. Obtention d’une souche asymbiotique chez le charançon Sitophilus sasakii Tak: différentes méthodes d’obtention et comparaison avec la souche symbiotique d’origine C. R. Acad. Sci. Paris Ser; 277 : 981-984

Nel, André, Patrick Roques, Patricia Nel, Jakub Prokop, and J. Sébastien Steyer. 2007. “The Earliest Holometabolous Insect from the Carboniferous: A ‘Crucial’ Innovation with Delayed Success (Insecta Protomeropina Protomeropidae).” Annales de La Societe Entomologique de France 43 (3): 349–55. doi:10.1080/00379271.2007.10697531.

Nene, Vishvanath, Jennifer R. Wortman, Daniel Lawson, Brian Haas, Chinnappa Kodira, Zhijian Tu, Brendan Loftus, et al. 2007. “Genome Sequence of Aedes Aegypti, a Major Arbovirus Vector.” Science 316 (5832): 1718–23. doi:10.1126/science.1138878.

Nguyen, Marcus, Alex Ekstrom, Xueqiong Li, and Yanbin Yin. 2015a. “HGT-Finder: A New Tool for Horizontal Gene Transfer Finding and Application to Aspergillus Genomes.” Toxins 7 (10). Multidisciplinary Digital Publishing Institute: 4035–53. doi:10.3390/toxins7104035.

Nguyen, Tam T., Patrick J. Collins, and Paul R. Ebert. 2015b. “Inheritance and Characterization of Strong Resistance to Phosphine in Sitophilus Oryzae (L.).” PLoS ONE 10 (4): 1–14. doi:10.1371/journal.pone.0124335.

Nikoh, Naruo, John P. McCutcheon, Toshiaki Kudo, Shin-ya Miyagishima, Nancy A. Moran, and Atsushi Nakabachi. 2010. “Bacterial Genes in the Aphid Genome: Absence of Functional Gene Transfer from Buchnera to Its Host.” Edited by Gregory P. Copenhaver. PLoS Genetics 6 (2). Public Library of Science: e1000827. doi:10.1371/journal.pgen.1000827.

Nikoh, Naruo, and Atsushi Nakabachi. 2009. “Aphids Acquired Symbiotic Genes via Lateral Gene Transfer.” BMC Biology 7 (1). BioMed Central: 12. doi:10.1186/1741-7007-7-12.

Nogge, G. 1976. “Sterility in Tsetse Flies (Glossina Morsitans Westwood) Caused by Loss of Symbionts.” Experientia 32 (8): 995–96.

Notario A, Binazzi A, Castresana L, Baragaño J, Montoya R. 1984. Los pulgones del Cedro: Cinara cedri Mimeur y Cedrobium laportei Remaudiere.

Naturalia Hispanica Series Monografias / Instituto Nacional para la Conservacion de la Naturaleza, 33. Ministerio de Agricultura, Pesca y Alimentación.

Nováková, Eva, Václav Hypša, Joanne Klein, Robert G. Foottit, Carol D. von Dohlen, and Nancy A. Moran. 2013. “Reconstructing the Phylogeny of Aphids (Hemiptera: Aphididae) Using DNA of the Obligate Symbiont Buchnera Aphidicola.” Molecular Phylogenetics and Evolution 68 (1): 42–54. doi:10.1016/j.ympev.2013.03.016.

Oakeson, Kelly F, Rosario Gil, Adam L Clayton, Diane M Dunn, Andrew C von Niederhausern, Cindy Hamil, Alex Aoyagi, et al. 2014. “Genome Degeneration and Adaptation in a Nascent Stage of Symbiosis.” Genome Biology and Evolution 6 (1): 76–93. doi:10.1093/gbe/evt210.

Obata, Hiroki, Aya Manabe, Naoko Nakamura, Tomokazu Onishi, and Yasuko Senba. 2011. “A New Light on the Evolution and Propagation of Prehistoric Grain Pests: The World’s Oldest Maize Weevils Found in Jomon Potteries, Japan.” PLoS ONE 6 (3): e14785. doi:10.1371/journal.pone.0014785.

Oberprieler, Rolf G, Adriana E Marvaldi, and Robert S Anderson. 2007. “Weevils, Weevils, Weevils Everywhere.” Zootaxa 1668: 491–520.

Odeyemi, OO, MO Ashamo, RO Akinkurolere, and AA Olatunji. 2010. “Resistance of Strains of Rice Weevil , Sitophilus Oryzae (Coleoptera: Curculionidae) to Pirimiphos Methyl.” 10th International Working Conference on Stored Product Protection, no. November: 167–72. doi:10.5073/jka.2010.425.433.

Ohtaka C, and Ishikawa H., 1991 - Effects of heat treatment on the symbiotic system of an aphid mycetocyte. Symbiosis, 11: 19-30.

Olive, Peter JW. 2002. “Reproduction and Life Cycles in Invertebrates.” In Encyclopedia of Life Sciences. Chichester: John Wiley & Sons, Ltd. doi:10.1038/npg.els.0003649.

Oliver, K. M., J. A. Russell, N. A. Moran, and M. S. Hunter. 2003. “Facultative Bacterial Symbionts in Aphids Confer Resistance to Parasitic Wasps.” Proceedings of the National Academy of Sciences 100 (4): 1803–7. doi:10.1073/pnas.0335320100.

Ortiz-Rivas, Benjamín, and David Martínez-Torres. 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): 305–17. doi:10.1016/j.ympev.2009.12.005.

Palmer, William J., and Francis M. Jiggins. 2015. “Comparative Genomics Reveals the Origins and Diversity of Arthropod Immune Systems.” Molecular Biology and Evolution 32 (8): 2111–29. doi:10.1093/molbev/msv093.

Panini M, Manicardi G, Moores G, Mazzoni E, 2016. An overview of the main pathways of metabolic resistance in insects.- Invertebrate Survival Journal, 13: 326-335.

Paradis, Emmanuel, Julien Claude, and Korbinian Strimmer. 2004. “APE: Analyses of Phylogenetics and Evolution in R Language.” Bioinformatics 20 (2): 289–90. doi:10.1093/bioinformatics/btg412.

Parra, G., K. Bradnam, and I. Korf. 2007. “CEGMA: A Pipeline to Accurately Annotate Core Genes in Eukaryotic Genomes.” Bioinformatics 23 (9): 1061–67. doi:10.1093/bioinformatics/btm071.

Pauchet, Yannick, Paul Wilkinson, Ritika Chauhan, and Richard H. Ffrench-Constant. 2010. “Diversity of Beetle Genes Encoding Novel Plant Cell Wall Degrading Enzymes.” PLoS ONE 5 (12). doi:10.1371/journal.pone.0015635.

Payne, Samuel H., and William F. Loomis. 2006. “Retention and Loss of Amino Acid Biosynthetic Pathways Based on Analysis of Whole-Genome Sequences.” Eukaryotic Cell 5 (2): 272–76. doi:10.1128/EC.5.2.272-276.2006.

Perez-Brocal, V., R. Gil, S. Ramos, A. Lamelas, M. Postigo, J. M. Michelena, F. J. Silva, A. Moya, and A. Latorre. 2006. “A Small Microbial Genome: The End of a Long Symbiotic Relationship?” Science 314 (5797): 312–13. doi:10.1126/science.1130441.

Plague, Gordon R, Helen E Dunbar, Phat L Tran, and Nancy a Moran. 2008. “Extensive Proliferation of Transposable Elements in Heritable Bacterial Symbionts.” Journal of Bacteriology 190 (2): 777–79. doi:10.1128/JB.01082-07.

Podell, Sheila, and Terry Gaasterland. 2007. “DarkHorse: A Method for Genome-Wide Prediction of Horizontal Gene Transfer.” Genome Biology 8 (2). BioMed Central: R16. doi:10.1186/gb-2007-8-2-r16.

Ponce-de-Leon, Miguel, Daniel Tamarit, Jorge Calle-Espinosa, Matteo Mori, Amparo Latorre, Francisco Montero, and Juli Pereto. 2017. “Determinism and Contingency Shape Metabolic Complementation in an Endosymbiotic Consortium.” Frontiers in Microbiology 8 (NOV). doi:10.3389/fmicb.2017.02290.

Price, Daniel R G, and Alex C C Wilson. 2014. “A Substrate Ambiguous Enzyme Facilitates Genome Reduction in an Intracellular Symbiont.” BMC Biology 12 (110): 1–9. doi:10.1186/s12915-014-0110-4.

Pryszcz, Leszek P., and Toni Gabaldón. 2016. “Redundans: An Assembly Pipeline for Highly Heterozygous Genomes.” Nucleic Acids Research 44 (12): e113. doi:10.1093/nar/gkw294.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Ramírez-Puebla, Shamayim T., Luis E. Servín-Garcidueñas, Ernesto Ormeño-Orrillo, Arturo Vera-Ponce de León, Mónica Rosenblueth, Luis Delaye, Julio Martínez, and Esperanza Martínez-Romero. 2015. “Species in Wolbachia? Proposal for the Designation of ‘Candidatus Wolbachia Bourtzisii’, ‘Candidatus Wolbachia Onchocercicola’, ‘Candidatus Wolbachia Blaxteri’, ‘Candidatus Wolbachia Brugii’, ‘Candidatus Wolbachia Taylori’, ‘Candidatus Wolbachia Collembolicola’ and ‘Candidatus Wolbachia Multihospitum’ for the Different Species within Wolbachia Supergroups.” Systematic and Applied Microbiology 38 (6): 390–99. doi:10.1016/j.syapm.2015.05.005.

Remaudiere, G. 1954. Les Cinarini (Hom. Aphidoidea Lachnidae) du Cèdre en Afrique du Nord. Rev.Path.Vég.Ent.Agric.Fr., XXXIII (2): 115-122.

Reyes-Prieto, Mariana, Carlos Vargas-Chávez, Amparo Latorre, and Andrés Moya. 2015. “SymbioGenomesDB: A Database for the Integration and Access to Knowledge on Host–symbiont Relationships.” Database 2015 (November). Oxford University Press: bav109. doi:10.1093/database/bav109.

Richards, Stephen, Richard A Gibbs, George M Weinstock, Susan J Brown, Robin Denell, Richard W Beeman, Richard Gibbs, et al. 2008. “The Genome of the Model Beetle and Pest Tribolium Castaneum.” Nature 452 (7190): 949–55. doi:10.1038/nature06784.

Rio, Rita V M, Cedric Lefevre, Abdelaziz Heddi, and Serap Aksoy. 2003. “Comparative Genomics of Insect-Symbiotic Bacteria: Influence of Host Environment on Microbial Genome Composition.” Applied and Environmental Microbiology 69 (11): 6825–32.

Rocha, Eduardo P.C., and Antoine Danchin. 2002. “Base Composition Bias Might Result from Competition for Metabolic Resources.” Trends in Genetics 18 (6): 291–94. doi:10.1016/S0168-9525(02)02690-2.

Rosenfeld, Jeffrey A., Darryl Reeves, Mercer R. Brugler, Apurva Narechania, Sabrina Simon, Russell Durrett, Jonathan Foox, et al. 2016. “Genome Assembly and Geospatial Phylogenomics of the Bed Bug Cimex Lectularius.” Nature Communications 7 (1): 10164. doi:10.1038/ncomms10164.

Russell, Calum W., Sophie Bouvaine, Peter D. Newell, and Angela E. Douglas. 2013. “Shared Metabolic Pathways in a Coevolved Insect-Bacterial Symbiosis.” Applied and Environmental Microbiology 79 (19): 6117–23. doi:10.1128/AEM.01543-13.

Sacchi, Luciano, Aldo Grigolo, Gianfranca Biscaldi, and Ugo Laudani. 1993. “Effects of Heat Treatment on the Symbiotic System of Blattoidea: Morphofunctional Alterations of Bacteriocytes.” Bolletino Di Zoologia 60 (3). Taylor & Francis Group : 271–79. doi:10.1080/11250009309355823.

Saha, Surya, Prashant S Hosmani, Krystal Villalobos-Ayala, Sherry Miller, Teresa Shippy, Mirella Flores, Andrew Rosendale, et al. 2017. “Improved Annotation of the Insect Vector of Citrus Greening Disease: Biocuration by a

Diverse Genomics Community.” Database : The Journal of Biological Databases and Curation 2017. Oxford University Press. doi:10.1093/database/bax032.

Sahney, Sarda, and Michael J Benton. 2008. “Recovery from the Most Profound Mass Extinction of All Time.” Proceedings of the Royal Society B: Biological Sciences 275 (1636): 759–65. doi:10.1098/rspb.2007.1370.

Salem, Hassan, Eugen Bauer, Roy Kirsch, Aileen Berasategui, Michael Cripps, Benjamin Weiss, Ryuichi Koga, et al. 2017. “Drastic Genome Reduction in an Herbivore’s Pectinolytic Symbiont.” Cell 171 (7). Elsevier Inc.: 1520–1531.e13. doi:10.1016/j.cell.2017.10.029.

Santos-Garcia, Diego, Carlos Vargas-Chavez, Andrés Moya, Amparo Latorre, and Francisco J. Silva. 2015. “Genome Evolution in the Primary Endosymbiont of Whiteflies Sheds Light on Their Divergence.” Genome Biology and Evolution 7 (3). Oxford University Press: 873–88. doi:10.1093/gbe/evv038.

Santos-Garcia, Diego, Francisco J Silva, Shai Morin, Konrad Dettner, and Stefan Martin Kuechler. 2017. “The All-Rounder Sodalis: A New Bacteriome-Associated Endosymbiont of the Lygaeoid Bug Henestaris Halophilus (Heteroptera: Henestarinae) and a Critical Examination of Its Evolution.” doi:10.1093/gbe/evx202.

Sauer, Christina, Dieter Dudaczek, Bert Ho, and Roy Gross. 2002. “Tissue Localization of the Endosymbiotic Bacterium ‘Candidatus Blochmannia Floridanus’ in Adults and Larvae of the Carpenter Ant Camponotus Floridanu.” Applied and Environmental Microbiology 68 (9): 4187–93. doi:10.1128/AEM.68.9.4187.

Schoville, Sean D., Yolanda H. Chen, Martin N. Andersson, Joshua B. Benoit, Anita Bhandari, Julia H. Bowsher, Kristian Brevik, et al. 2018. “A Model Species for Agricultural Pest Genomics: The Genome of the Colorado Potato Beetle, Leptinotarsa Decemlineata (Coleoptera: Chrysomelidae).” Scientific Reports 8 (1): 1–18. doi:10.1038/s41598-018-20154-1.

Scott, Jeffrey G, Wesley C Warren, Leo W Beukeboom, Daniel Bopp, Andrew G Clark, Sarah D Giers, Monika Hediger, et al. 2014. “Genome of the House Fly, Musca Domestica L., a Global Vector of Diseases with Adaptations to a Septic Environment.” Genome Biology 15 (10): 466. doi:10.1186/s13059-014-0466-3.

Seemann, T. 2014. “Prokka: Rapid Prokaryotic Genome Annotation.” Bioinformatics 30 (14): 2068–69. doi:10.1093/bioinformatics/btu153.

Serbus, Laura R., Catharina Casper-Lindley, Frédéric Landmann, and William Sullivan. 2008. “The Genetics and Cell Biology of Wolbachia -Host Interactions.” Annual Review of Genetics 42 (1): 683–707. doi:10.1146/annurev.genet.41.110306.130354.

Serbus, Laura R., Brian Garcia Rodriguez, Zinat Sharmin, A. J. M. Zehadee Momtaz, and Steen Christensen. 2017. “Predictive Genomic Analyses Inform the Basis for Vitamin Metabolism and Provisioning in Bacteria-Arthropod Endosymbioses.” G3&#58; Genes|Genomes|Genetics 7 (6): 1887–98. doi:10.1534/g3.117.042184.

Shatters, Robert G., Laura M. Boykin, Stephen L. Lapointe, Wayne B. Hunter, and A.A. Weathersbee. 2006. “Phylogenetic and Structural Relationships of the PR5 Gene Family Reveal an Ancient Multigene Family Conserved in Plants and Select Animal Taxa.” Journal of Molecular Evolution 63 (1): 12–29. doi:10.1007/s00239-005-0053-z.

Shen, Zhicheng, Michael Denton, Navdeep Mutti, Kirk Pappan, Michael R Kanost, John C Reese, and Gerald R Reeck. 2003. “Polygalacturonase from Sitophilus Oryzae: Possible Horizontal Transfer of a Pectinase Gene from Fungi to Weevils.” Journal of Insect Science 3 (24): 1–9.

Shewry, Peter R., and Nigel G. Halford. 2002. “Cereal Seed Storage Proteins: Structures, Properties and Role in Grain Utilization.” Journal of Experimental Botany 53 (370). Oxford University Press: 947–58. doi:10.1093/jexbot/53.370.947.

Shi, Houxia, Lianghong Pei, Shasha Gu, Shicheng Zhu, Yanyun Wang, Yi Zhang, and Bin Li. 2012. “Glutathione S-Transferase (GST) Genes in the Red Flour Beetle, Tribolium Castaneum, and Comparative Analysis with Five Additional Insects.” Genomics 100 (5): 327–35. doi:10.1016/j.ygeno.2012.07.010.

Shigenobu, Shuji, and Alex C. C. Wilson. 2011. “Genomic Revelations of a Mutualism: The Pea Aphid and Its Obligate Bacterial Symbiont.” Cellular and Molecular Life Sciences 68 (8): 1297–1309. doi:10.1007/s00018-011-0645-2.

Shigenobu, Shuji, Hidemi Watanabe, Masahira Hattori, Yoshiyuki Sakaki, and Hajime Ishikawa. 2000. “Genome Sequence of the Endocellular Bacterial Symbiont of Aphids Buchnera Sp. APS.” Nature 407 (6800): 81–86. doi:10.1038/35024074.

Silva, Alexandra Avelar, Lucas Soares Braga, Alberto Soares Corrêa, Valerie Renee Holmes, John Spencer Johnston, Brenda Oppert, Raul Narciso Carvalho Guedes, and Mara Garcia Tavares. 2018. “Comparative Cytogenetics and Derived Phylogenic Relationship among Sitophilus Grain Weevils (Coleoptera, Curculionidae, Dryophthorinae).” Comparative Cytogenetics 12 (2). Pensoft Publishers: 223–45. doi:10.3897/CompCytogen.v12i2.26412.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2015. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” Bioinformatics 31 (19): 3210–12. doi:10.1093/bioinformatics/btv351.

Simon, Jean-Christophe, David Martinez-Torres, Amparo Latorre, Andres Moya, and Paul D N Hebert. 1996. “Molecular Characterization of Cyclic and Obligate Parthenogens in the Aphid Rhopalosiphum Padi (L.).”

Simpson, Jared T., Richard Durbin, Daniel R. Zerbino, Ewan Birney, Kim Wong, and Shaun D Jackman. 2012. “Efficient de Novo Assembly of Large Genomes Using Compressed Data Structures Sequence Data.” Genome Research 22 (3): 549–56. doi:10.1101/gr.126953.111.

Simpson, Jared T. 2014. “Exploring Genome Characteristics and Sequence Quality without a Reference.” Bioinformatics 30 (9). Oxford University Press: 1228–35. doi:10.1093/bioinformatics/btu023.

Siozios, Stefanos, Panagiotis Ioannidis, Lisa Klasson, Siv G. E. Andersson, Henk R. Braig, and Kostas Bourtzis. 2013. “The Diversity and Evolution of Wolbachia Ankyrin Repeat Domain Genes.” Edited by Richard Cordaux. PLoS ONE 8 (2). Public Library of Science: e55390. doi:10.1371/journal.pone.0055390.

Sloan, Daniel B, Atsushi Nakabachi, Stephen Richards, Jiaxin Qu, Shwetha Canchi Murali, Richard A Gibbs, and Nancy A Moran. 2014. “Parallel Histories of Horizontal Gene Transfer Facilitated Extreme Reduction of Endosymbiont Genomes in Sap-Feeding Insects.” Molecular Biology and Evolution 31 (4): 857–71. doi:10.1093/molbev/msu004.

Smit, AFA, and R Hubley. RepeatModeler Open-1.0. 2008-2015 <http://www.repeatmasker.org>.

Smit, AFA, R Hubley, and P Green. RepeatMasker Open-4.0. 2013-2015 <http://www.repeatmasker.org>.

Song, Nan, Hu Li, Fan Song, and Wanzhi Cai. 2016. “Molecular Phylogeny of Polyneoptera (Insecta) Inferred from Expanded Mitogenomic Data.” Scientific Reports 6 (October): 1–10. doi:10.1038/srep36175.

Stanke, Mario, and Burkhard Morgenstern. 2005. “AUGUSTUS: A Web Server for Gene Prediction in Eukaryotes That Allows User-Defined Constraints.” Nucleic Acids Research 33 (SUPPL. 2): 465–67. doi:10.1093/nar/gki458.

Stejskal, V., Z. Kucerova, and J. Lukas. 2004. “Evidence and Symptoms of Pasta Infestation by Sitophilus Oryzae (Curculionidae; Coleoptera) in the Czech Republic.” Plant Protection Science 40 (3): 107–11.

Styrsky, John D, and Micky D Eubanks. n.d. “Ecological Consequences of Interactions between Ants and Honeydew-Producing Insects.” doi:10.1098/rspb.2006.3701.

Tamames, Javier, Rosario Gil, Amparo Latorre, Juli Peretó, Francisco J Silva, and Andrés Moya. 2007. “The Frontier between Cell and Organelle: Genome

Analysis of Candidatus Carsonella Ruddii.” BMC Evolutionary Biology 7 (1). BioMed Central: 181. doi:10.1186/1471-2148-7-181.

Tamas, I., Lisa Klasson, Björn Canbäck, A Kristina Näslund, Ann-Sofie Eriksson, Jennifer J Wernegreen, Jonas P Sandström, Nancy A Moran, and Siv G E Andersson. 2002. “50 Million Years of Genomic Stasis in Endosymbiotic Bacteria.” Science 296 (5577): 2376–79. doi:10.1126/science.1071278.

Tanaka, Hiromitsu, Jun Ishibashi, Kosuke Fujita, Yoshiro Nakajima, Aki Sagisaka, Kazuya Tomimoto, Noriko Suzuki, et al. 2008. “A Genome-Wide Analysis of Genes and Gene Families Involved in Innate Immunity of Bombyx Mori.” Insect Biochemistry and Molecular Biology 38 (12): 1087–1110. doi:10.1016/j.ibmb.2008.09.001.

Tarver, Matthew R., Qiang Huang, Lilia de Guzman, Tom Rinderer, Beth Holloway, Justin Reese, Daniel Weaver, and Jay D. Evans. 2016. “Transcriptomic and Functional Resources for the Small Hive Beetle Aethina Tumida , a Worldwide Parasite of Honey Bees.” Genomics Data 9 (September): 97–99. doi:10.1016/j.gdata.2016.06.003.

Tefera, Tadele, Fred Kanampiu, Hugo De Groote, Jon Hellin, Stephen Mugo, Simon Kimenju, Yoseph Beyene, Prasanna M. Boddupalli, Bekele Shiferaw, and Marianne Banziger. 2011. “The Metal Silo: An Effective Grain Storage Technology for Reducing Post-Harvest Insect and Pathogen Losses in Maize While Improving Smallholder Farmers’ Food Security in Developing Countries.” Crop Protection 30 (3). Elsevier Ltd: 240–45. doi:10.1016/j.cropro.2010.11.015.

The Honeybee Genome Sequencing Consortium. 2006. “Insights into Social Insects from the Genome of the Honeybee Apis Mellifera.” Nature 443 (7114). Nature Publishing Group: 931–49. doi:10.1038/nature05260.

The International Aphid Genomics Consortium. 2010. “Genome Sequence of the Pea Aphid Acyrthosiphon Pisum.” Edited by Jonathan A. Eisen. PLoS Biology 8 (2). Public Library of Science: e1000313. doi:10.1371/journal.pbio.1000313.

The International Silkworm Genome. 2008. “The Genome of a Lepidopteran Model Insect, the Silkworm Bombyx Mori.” Insect Biochemistry and Molecular Biology 38 (12): 1036–45. doi:10.1016/j.ibmb.2008.11.004.

Toju, Hirokazu, Takahiro Hosokawa, Ryuichi Koga, Naruo Nikoh, Xian Ying Meng, Nobutada Kimura, and Takema Fukatsu. 2010. “‘Candidatus Curculioniphilus Buchneri,’ a Novel Clade of Bacterial Endocellular Symbionts from Weevils of the Genus Curculio.” Applied and Environmental Microbiology 76 (1): 275–82. doi:10.1128/AEM.02154-09.

Toju, Hirokazu, Hirotoshi Sato, Satoshi Yamamoto, Kohmei Kadowaki, Akifumi S Tanabe, Shigenobu Yazawa, Osamu Nishimura, and Kiyokazu Agata. 2013. “How Are Plant and Fungal Communities Linked to Each Other in Belowground Ecosystems? A Massively Parallel Pyrosequencing Analysis of the Association

Specificity of Root-Associated Fungi and Their Host Plants.” Ecology and Evolution 3 (9). Wiley-Blackwell: 3112–24. doi:10.1002/ece3.706.

Tuatay, N., Remaudiere, G. 1964. Premiere Contribution au Catalogue des Aphididae (Hom.) de la Turquie. Rev. de Path. Veg. Et Ent. Agr. de Fr. 43: 243-278.

Ülgentürk S, Szentkiralyi F, Uygun N, Fent M, Gimari SD, Civelek H, Ayha B. 2013. Predators of Marchalina hellenica (Hemiptera: Marchalinidae) on pine forests in Turkey. Phytoparasitica 41(5), 529-537.

Unterman, B M, P Baumann, and D L McLean. 1989. “Pea Aphid Symbiont Relationships Established by Analysis of 16S RRNAs.” Journal of Bacteriology 171 (6): 2970–74.

Valanne, S., J.-H. Wang, and M. Ramet. 2011. “The Drosophila Toll Signaling Pathway.” The Journal of Immunology 186 (2): 649–56. doi:10.4049/jimmunol.1002302.

van Ham, R. C. H. J., J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J. M. Fernandez, et al. 2003. “Reductive Genome Evolution in Buchnera Aphidicola.” Proceedings of the National Academy of Sciences 100 (2): 581–86. doi:10.1073/pnas.0235981100.

Vavre, F., F. Fleury, D. Lepetit, P. Fouillet, and M. Bouletreau. 1999. “Phylogenetic Evidence for Horizontal Transmission of Wolbachia in Host-Parasitoid Associations.” Molecular Biology and Evolution 16 (12): 1711–23. doi:10.1093/oxfordjournals.molbev.a026084.

Vega, Fernando E, Stuart M Brown, Hao Chen, Eric Shen, Mridul B Nair, Javier A Ceja-navarro, Eoin L Brodie, Francisco Infante, Patrick F Dowd, and Arnab Pain. 2015. “Draft Genome of the Most Devastating Insect Pest of Coffee Worldwide : The Coffee Berry Borer , Hypothenemus Hampei.” Nature Publishing Group, no. July. Nature Publishing Group: 1–17. doi:10.1038/srep12525.

Vigneron, Aurélien, Delphine Charif, Carole Vincent-Monégat, Agnès Vallier, Frédérick Gavory, Patrick Wincker, and Abdelaziz Heddi. 2012. “Host Gene Response to Endosymbiont and Pathogen in the Cereal Weevil Sitophilus Oryzae.” BMC Microbiology 12 Suppl 1 (January): 1–25. doi:10.1186/1471-2180-12-S1-S14.

Vigneron, Aurélien, Florent Masson, Agnès Vallier, Séverine Balmand, Marjolaine Rey, Carole Vincent-Monégat, Emre Aksoy, Etienne Aubailly-Giraud, Anna Zaidman-Rémy, and Abdelaziz Heddi. 2014. “Insects Recycle Endosymbionts When the Benefit Is Over.” Current Biology : CB 24 (19): 2267–73. doi:10.1016/j.cub.2014.07.065.

von Dohlen, Carol, and Nancy A. Moran. 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). Oxford University Press: 689–717. doi:10.1111/j.1095-8312.2000.tb01286.x.

Vurture, Gregory W, Fritz J Sedlazeck, Maria Nattestad, Charles J Underwood, Han Fang, James Gurtowski, and Michael C Schatz. 2017. “GenomeScope: Fast Reference-Free Genome Profiling from Short Reads.” Edited by Bonnie Berger. Bioinformatics 33 (14): 2202–4. doi:10.1093/bioinformatics/btx153.

Wang, B., M. Lu, C. Cheng, C. Salcedo, and J. Sun. 2012. “Saccharide-Mediated Antagonistic Effects of Bark Beetle Fungal Associates on Larvae.” Biology Letters 9 (1): 20120787–20120787. doi:10.1098/rsbl.2012.0787.

Wang, Xianhui, Xiaodong Fang, Pengcheng Yang, Xuanting Jiang, Feng Jiang, Dejian Zhao, Bolei Li, et al. 2014. “The Locust Genome Provides Insight into Swarm Formation and Long-Distance Flight.” Nature Communications 5 (1). Nature Publishing Group: 2957. doi:10.1038/ncomms3957.

Weiss, Benjamin, and Martin Kaltenpoth. 2016. “Bacteriome-Localized Intracellular Symbionts in Pollen-Feeding Beetles of the Genus Dasytes (Coleoptera, Dasytidae).” Frontiers in Microbiology 7 (SEP): 1–10. doi:10.3389/fmicb.2016.01486.

Wernegreen, Jennifer J. 2015. “Endosymbiont Evolution: Predictions from Theory and Surprises from Genomes.” Annals of the New York Academy of Sciences 1360 (1). John Wiley & Sons, Ltd (10.1111): 16–35. doi:10.1111/nyas.12740.

Werren, J. H., S. Richards, C. A. Desjardins, O. Niehuis, J. Gadau, J. K. Colbourne, L. W. Beukeboom, et al. 2010. “Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species.” Science 327 (5963): 343–48. doi:10.1126/science.1178028.

Whalon, M.E., D. Mota-Sanchez, R.M. Hollingworth, and L. Duynslager. 2018. Arthropod Pesticide Resistance Database. www.pesticideresistance.org, accessed on 2018.

White, K, E Tahaoglu, and H Steller. 1996. “Cell Killing by the Drosophila Gene Reaper.” Science (New York, N.Y.) 271 (5250): 805–7.

White, Pamela M, Laura R Serbus, Alain Debec, Adan Codina, Walter Bray, Antoine Guichet, R Scott Lokey, and William Sullivan. 2017. “Reliance of Wolbachia on High Rates of Host Proteolysis Revealed by a Genome-Wide RNAi Screen of Drosophila Cells.” Genetics 205 (4). Genetics Society of America: 1473–88. doi:10.1534/genetics.116.198903.

Will, Torsten, W Fred Tjallingii, Alexandra Thönnessen, and Aart J E van Bel. 2007. “Molecular Sabotage of Plant Defense by Aphid Saliva.” Proceedings of the National Academy of Sciences of the United States of America 104 (25). National Academy of Sciences: 10536–41. doi:10.1073/pnas.0703535104.

Wilson, A. C.C., P. D. Ashton, F. Calevro, H. Charles, S. Colella, G. Febvay, G. Jander, et al. 2010. “Genomic Insight into the Amino Acid Relations of the Pea Aphid, Acyrthosiphon Pisum, with Its Symbiotic Bacterium Buchnera Aphidicola.” Insect Molecular Biology 19 (SUPPL. 2): 249–58. doi:10.1111/j.1365-2583.2009.00942.x.

Wootton, Robin J., and Jarmila Kukalová-Peck. 2000. “Flight Adaptations in Palaeozoic Palaeoptera (Insecta).” Biological Reviews 75 (1): 129–67. doi:10.1111/j.1469-185X.1999.tb00043.x.

Wu, Martin, Ling V Sun, Jessica Vamathevan, Markus Riegler, Robert Deboy, Jeremy C Brownlie, Elizabeth A McGraw, et al. 2004. “Phylogenomics of the Reproductive Parasite Wolbachia Pipientis WMel: A Streamlined Genome Overrun by Mobile Genetic Elements.” Edited by Nancy A. Moran. PLoS Biology 2 (3). Public Library of Science: e69. doi:10.1371/journal.pbio.0020069.

Wu, Martin, and Alexandra J. Scott. 2012. “Phylogenomic Analysis of Bacterial and Archaeal Sequences with AMPHORA2.” Bioinformatics 28 (7). Oxford University Press: 1033–34. doi:10.1093/bioinformatics/bts079.

Wurm, Yannick, John Wang, Oksana Riba-Grognuz, Miguel Corona, Sanne Nygaard, Brendan G Hunt, Krista K Ingram, et al. 2011. “The Genome of the Fire Ant Solenopsis Invicta.” Proceedings of the National Academy of Sciences of the United States of America 108 (14). National Academy of Sciences: 5679–84. doi:10.1073/pnas.1009690108.

Wybouw, Nicky, Yannick Pauchet, David G Heckel, and Thomas Van Leeuwen. 2016. “Horizontal Gene Transfer Contributes to the Evolution of Arthropod Herbivory.” Genome Biology and Evolution 8 (6). Oxford University Press: 1785–1801. doi:10.1093/gbe/evw119.

Xia, Xiaofeng, Liying Yu, Minqian Xue, Xiaoqiang Yu, Liette Vasseur, Geoff M. Gurr, Simon W. Baxter, Hailan Lin, Junhan Lin, and Minsheng You. 2015. “Genome-Wide Characterization and Expression Profiling of Immune Genes in the Diamondback Moth, Plutella Xylostella (L.).” Scientific Reports 5 (March). Nature Publishing Group: 1–13. doi:10.1038/srep09877.

Xue, Jian, Xin Zhou, Chuan Xi Zhang, Li Li Yu, Hai Wei Fan, Zhuo Wang, Hai Jun Xu, et al. 2014. “Genomes of the Rice Pest Brown Planthopper and Its Endosymbionts Reveal Complex Complementary Contributions for Host Adaptation.” Genome Biology 15 (12). doi:10.1186/s13059-014-0521-0.

Yi, Hui-Yu, Munmun Chowdhury, Ya-Dong Huang, and Xiao-Qiang Yu. 2014. “Insect Antimicrobial Peptides and Their Applications.” Applied Microbiology and Biotechnology 98 (13): 5807–22. doi:10.1007/s00253-014-5792-6.Insect.

You, Minsheng, Zhen Yue, Weiyi He, Xinhua Yang, Guang Yang, Miao Xie, Dongliang Zhan, et al. 2013. “A Heterozygous Moth Genome Provides Insights into Herbivory and Detoxification.” Nature Genetics 45 (2). Nature Publishing Group: 220–25. doi:10.1038/ng.2524.

Zabalou, Sofia, Markus Riegler, Marianna Theodorakopoulou, Christian Stauffer, Charalambos Savakis, and Kostas Bourtzis. 2004. “Wolbachia-Induced Cytoplasmic Incompatibility as a Means for Insect Pest Population Control.” Vol. 101.

Zeidler, Martin P, and Nina Bausek. 2013. “The Drosophila JAK-STAT Pathway.” Jak-Stat 2 (3): e25353. doi:10.4161/jkst.25353.

Zhan, Shuai, Christine Merlin, Jeffrey L. Boore, and Steven M. Reppert. 2011. “The Monarch Butterfly Genome Yields Insights into Long-Distance Migration.” Cell 147 (5): 1171–85. doi:10.1016/j.cell.2011.09.052.

Zhang, Guanyang, Patrick Browne, Geng Zhen, Andrew Johnston, Hinsby Cadillo-Quiroz, and Nico Franz. 2017. “Endosymbiont Diversity and Evolution across the Weevil Tree of Life.” BioRxiv, no. Oberprieler 2007. doi:https://doi.org/10.1101/171181.

Zhang, Shao Qian, Li Heng Che, Yun Li, Liang Dan, Hong Pang, Adam Ślipiński, and Peng Zhang. 2018. “Evolutionary History of Coleoptera Revealed by Extensive Sampling of Genes and Species.” Nature Communications 9 (1): 1–11. doi:10.1038/s41467-017-02644-4.

Zhang, Zhi-Qiang. 2013. “Phylum Arthropoda.” Zootaxa 3703 (1): 17–26.

Zhu, Fang, Timothy W. Moural, Kapil Shah, and Subba Reddy Palli. 2013. “Integrated Analysis of Cytochrome P450 Gene Superfamily in the Red Flour Beetle, Tribolium Castaneum.” BMC Genomics 14 (1). doi:10.1186/1471-2164-14-174.

Zientz, Evelyn, Thomas Dandekar, and Roy Gross. 2004. “Metabolic Interdependence of Obligate Intracellular Bacteria and Their Insect Hosts.” Microbiology and Molecular Biology Reviews : MMBR 68 (4). American Society for Microbiology: 745–70. doi:10.1128/MMBR.68.4.745-770.2004.

Zou, Zhen, Jay D Evans, Zhiqiang Lu, Picheng Zhao, Michael Williams, Niranji Sumathipala, Charles Hetru, Dan Hultmark, and Haobo Jiang. 2007. “Comparative Genomic Analysis of the Tribolium Immune System.” Genome Biology 8 (8): R177. doi:10.1186/gb-2007-8-8-r177.

Zumaya-Estrada, Federico Alonso, Jesús Martínez-Barnetche, Andrés Lavore, Rolando Rivera-Pomar, and Mario Henry Rodríguez. 2018. “Comparative Genomics Analysis of Triatomines Reveals Common First Line and Inducible Immunity-Related Genes and the Absence of Imd Canonical Components among Hemimetabolous Arthropods.” Parasites & Vectors 11 (1). BioMed Central: 48. doi:10.1186/s13071-017-2561-2.

Zunjare, Rajkumar, Firoz Hossain, Vignesh Muthusamy, Sunil K. Jha, Pradyumn Kumar, Javaji C. Sekhar, Nepolean Thirunavukkarasu, and Hari S. Gupta. 2016. “Genetic Variability among Exotic and Indigenous Maize Inbreds for Resistance to Stored Grain Weevil (Sitophilus Oryzae L.) Infestation.”

Cogent Food & Agriculture 2 (1). Cogent: 1–10. doi:10.1080/23311932.2015.1137156.

APPENDIX

S.1 Additional Tables

Name	Order	Family	Holo/hemi metabolous	Primary Endosymbiont	Reference
Aethina tumida	Coleoptera	Nitidulidae	Holometabolous		Tarver *et al.,* 2016
Agrilus planipennis	Coleoptera	Buprestidae	Holometabolous		Duan *et al.,* 2015
Anoplophora glabripennis	Coleoptera	Cerambycidae	Holometabolous		McKenna *et al.,* 2016
Dendroctonus ponderosae	Coleoptera	Curculionidae	Holometabolous		Keeling *et al.,* 2013
Nicrophorus vespilloides	Coleoptera	Silphidae	Holometabolous		Cunningham *et al.,* 2015
Sitophilus oryzae	Coleoptera	Dryophthoridae	Holometabolous	*Candidatus Sodalis pierantonius*	
Tribolium castaneum	Coleoptera	Tenebrionidae	Holometabolous		Tribolium Genome Sequencing Consortium, 2008
Daphnia pulex	Diplostraca	Daphniidae			Colbourne *et al.,* 2011
Aedes aegypti	Diptera	Culicidae	Holometabolous		Nene et al., 2007
Anopheles gambiae	Diptera	Culicidae	Holometabolous		Holt*et al.,* 2002
Drosophila melanogaster	Diptera	Drosophilidae	Holometabolous		Adams *et al.,* 2000
Glossina morsitans	Diptera	Glossinidae	Holometabolous	*Wigglesworthia glossinidia*	International Glossina Genome Initiative, 2014
Acyrthosiphon pisum	Hemiptera	Aphididae	Hemimetabolous	*Buchnera aphidicola*	The International Aphid Genomics Consortium, 2010
Cimex lectularius	Hemiptera	Cimicidae	Hemimetabolous	*Wolbachia wCle*	Rosenfeld, *et al.,* 2015
Cinara cedri	Hemiptera	Lachnidae	Hemimetabolous	*Buchnera aphidicola* and *Serratia symbiotica*	
Diaphorina citri	Hemiptera	Liviidae	Hemimetabolous	*Candidatus Carsonella ruddii*	Saha *et al.,* 2017
Myzus persicae	Hemiptera	Aphididae	Hemimetabolous	*Buchnera aphidicola*	Mathers et al., 2017

Nilaparvata lugens	Hemiptera	Delphacidae	Hemimetabolous	Yeast like symbiont	Xue et al., 2014
Apis mellifera	Hymenoptera	Apidae	Holometabolous		Honeybee Genome Sequencing Consortium, 2006
Camponotus floridanus	Hymenoptera	Formicidae	Holometabolous	*Blochmania floridanus*	Bonasio *et al.,* 2010
Nasonia vitripennis	Hymenoptera	Pteromalidae	Holometabolous		Werren *et al.,* 2010
Solenopsis invicta	Hymenoptera	Formicidae	Holometabolous		Wurm *et al.,* 2011
Bombyx mori	Lepidoptera	Bombycidae	Holometabolous		The International Silkworm Genome Consortium, 2008
Danaus plexippus	Lepidoptera	Nymphalidae	Holometabolous		Zhan et al., 2011
Manduca sexta	Lepidoptera	Sphingidae	Holometabolous		Cao and Jiang, 2015
Plutella xylostella	Lepidoptera	Plutellidae	Holometabolous		You *et al.,* 2013
Pediculus humanus	Phthiraptera	Pediculidae	Hemimetabolous	*Candidatus Riesia pediculicola*	KirknesS *et al.,* 2010
Tetranychus urticae	Trombidiformes	Tetranychidae			Grbić *et al.,* 2011

Table S.1 Arthropod species used in the analysis.

name	dpu	tur	dci	cce	Bap	Ssy	Wcc	mpe	api	cle	nlu	phu	nve	tca	atu	sor	Spa	dpo	agl	apl	pxy	dpl	mse	bmo	gmo	dme	aga	aae	nvi	ame	sin	cfl	
K00615 (E2.2.1.1)	1	2	0	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	His
K01783 (rpe)	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	
K01807 (rpiA)	1	1	2	1	1	1	0	1	1	1	1	1	1	1	1	1	1	4	1	1	1	1	2	1	0	1	1	1	1	1	0	1	
K01808 (rpiB)	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
K00948 (PRPS)	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	
K00765 (hisG)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K11755 (hisIE)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K11755 (hisIE)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01814 (hisA)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K02500 (hisF)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K02501 (hisH)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01089 (hisB)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00817 (hisC)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01089 (hisB)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00013 (hisD)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00850 (pfkA)	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Glycer

K01623 (ALDO)	1	0	1	0	0	0	0	1	1	1	1	1	1	1	3	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	
K01624 (FBA)	0	0	0	0	1	1	0	0	3	0	0	0	0	0	0	0	2	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	
K11645 (fbaB)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00134 (GAPDH)	1	0	1	0	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	0	1	1	2	1	1	1	1	1	1	
K00927 (PGK)	1	1	1	1	1	1	1	1	1	0	2	0	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	
K00058 (serA)	1	2	1	1	0	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	Ser
K00831 (serC)	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	
K01079 (serB)	1	1	0	1	0	0	0	1	1	1	1	0	1	1	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	
K00640 (cysE)	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	Cys
K01738 (cysK)	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	3	0	0	0	1	2	1	1	0	0	0	0	0	0	0	0	
K01697 (CBS)	2	1	0	1	0	0	0	1	1	1	0	2	1	1	1	1	0	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	
K01758 (CTH)	1	1	1	1	0	0	0	2	2	1	1	1	2	1	1	1	0	1	1	2	1	0	1	1	1	1	1	1	1	1	1	2	
K17217 (mccB)	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00600 (glyA)	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Gly
K01834 (PGAM)	1	0	0	1	1	1	0	1	1	0	1	0	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	Phosphoenolpyruva
K15633 (gpmI)	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	
K01689 (ENO)	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	
K01626 (E2.5.1. 54)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	Chorismate
K01735 (aroB)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K03786 (aroQ)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00014 (aroE)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00891 (E2.7.1. 71)	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00800 (aroA)	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01736 (aroC)	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01657 (trpE)	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Trp
K01658 (trpG)	0	0	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00766 (trpD)	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K13498 (trpCF)	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K13498 (trpCF)	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01695 (trpA)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01696 (trpB)	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01694 (TRP)	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01850 (E5.4.9 9.5)	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Prephenate
K14187 (tyrA)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K04093 (pheA1)	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K14170 (pheA)	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Phe
K14170 (pheA)	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01713 (pheC)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K14454 (GOT1)	1	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	
K14455 (GOT2)	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	
K00812 (aspB)	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00813 (aspC)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K11358 (yhdR)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	

K00815 (TAT)	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
K00817 (hisC)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00832 (tyrB)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K00500 (phhA)	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	Tyr
K14187 (tyrA)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K14454 (GOT1)	1	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	
K14455 (GOT2)	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	
K00812 (aspB)	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00813 (aspC)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K11358 (yhdR)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	
K00815 (TAT)	1	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
K00817 (hisC)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00832 (tyrB)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K00873 (PK)	1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	Pyruvate
K00814 (GPT)	1	0	1	1	0	0	0	1	1	1	1	1	1	0	1	1	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	Ala
K01652 (E2.2.1.6L)	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Oxoisovalerate
K01653 (E2.2.1.6S)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K11258 (ilvM)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00053 (ilvC)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01687 (ilvD)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	1	0	0	0	
K00826 (E2.6.1.42)	1	1	0	1	0	0	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	2	1	1	Val
K01649 (leuA)	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	Leu
K01703 (leuC)	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K01704 (leuD)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00052 (leuB)	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
K00826 (E2.6.1.42)	1	1	0	1	0	0	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	2	1	1	
K01958 (PC)	1	1	0	1	0	0	0	1	1	1	0	1	1	1	2	1	0	1	1	1	2	1	0	1	0	1	1	1	1	1	1	1	Oxaloacetate
K14454 (GOT1)	1	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	Asp
K14455 (GOT2)	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	
K00812 (aspB)	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00813 (aspC)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K11358 (yhdR)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	
K01914 (asnA)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Asn
K01953 (asnB)	0	1	1	1	0	0	0	1	1	0	2	1	1	1	1	1	1	1	1	2	1	1	1	1	0	1	1	2	1	1	0	1	
K00928 (lysC)	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Lys
K12526 (lysAC)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K12524 (thrA)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00133 (asd)	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01714 (dapA)	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	
K00215 (dapB)	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00674 (dapD)	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

K00821 (argD)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
K01439 (dapE)	0	0	0	0	1	1	1	0	1	3	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K01778 (dapF)	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01586 (lysA)	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K12526 (lysAC)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K12524 (thrA)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Homoserine
K00872 (thrB1)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Thr
K02204 (thrB2)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01733 (thrC)	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	2	1	1	0	0	0	0	0	0	0	0	
K17989 (SDS)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Ile
K01754 (E4.3.1.19)	1	0	1	1	0	0	0	1	1	1	1	0	1	1	1	1	0	1	1	0	1	1	1	1	0	1	1	2	0	1	0	0	
K01652 (E2.2.1.6L)	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K01653 (E2.2.1.6S)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K11258 (ilvM)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00053 (ilvC)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01687 (ilvD)	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	1	0	0	0	
K00826 (E2.6.1.42)	1	1	0	1	0	0	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	2	1	1	
K00651 (metA)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	Met
K01739 (metB)	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01760 (metC)	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00549 (metE)	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	
K01758 (CTH)	1	1	1	1	0	0	0	2	2	1	1	1	2	1	1	1	0	1	1	2	1	0	1	1	1	1	1	1	1	1	1	2	
K17217 (mccB)	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K01697 (CBS)	2	1	0	1	0	0	0	1	1	1	0	2	1	1	1	1	0	2	1	1	1	1	1	2	1	1	1	1	1	1	1	1	
K00547 (mumM)	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	
K01647 (CS)	1	0	2	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	2	1	1	2	1	1	1	1	2-Oxoglutarate
K01681 (ACO)	1	1	2	1	0	0	1	3	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	2	1	2	1	1	1	1	
K00031 (IDH1)	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
K00030 (IDH3)	1	2	1	1	0	0	0	1	2	1	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	
K00814 (GPT)	1	0	1	1	0	0	0	1	1	1	1	1	1	0	1	1	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	Glu
K14454 (GOT1)	1	1	0	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	
K14455 (GOT2)	1	1	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	
K00812 (aspB)	0	0	0	0	0	0	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00813 (aspC)	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K11358 (yhdR)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	
K12657 (ALDH18A1)	1	2	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	Pro
K00147 (proA)	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	
K12657 (ALDH18A1)	1	2	0	1	0	0	0	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	
K00286 (proC)	1	0	1	1	0	0	0	1	1	1	0	1	2	1	2	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	
K01915 (glnA)	1	0	0	1	0	0	0	1	1	1	0	1	1	1	2	1	1	1	1	1	0	0	1	0	1	1	1	1	1	2	1	1	Gln

K14682 (argAB)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Arg
K00930 (argB)	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K00145 (argC)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
K00818 (E2.6.1.11)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
K00821 (argD)	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
K01438 (argE)	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
K14677 (ACY1)	0	0	0	1	0	0	0	1	1	1	0	1	2	1	2	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	2	
K00819 (rocD)	1	2	0	1	0	0	0	2	1	0	1	1	2	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	2	1	2	2	
K00611 (OTC)	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	
K01940 (argG)	1	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	1	1	
K01755 (argH)	1	0	6	0	1	0	0	0	1	1	0	0	1	1	1	1	1	1	1	0	2	1	1	1	1	1	1	1	1	1	3	1	

Table S.2 Matrix of counts of the number of genes identified which potentially code for enzymes involved in the biosynthesis of amino acids. The first column refers to the KEGG id and the last column to the right refers to the amino acid or metabolite produced through those reactions. The middle columns to a given arthropod. Note that they are taxonomically ordered. Abbreviations are as follows: Daphnia pulex (dpu), Tetranychus urticae (tur), Diaphorina citri (dci), Cinara cedri (cce), Buchnera aphidicola (Bap), Serratia symbiotica (Ssy), Wolbachia (Wcc), Myzus persicae (mpe), Acyrthosiphon pisum (api), Cimex lectularius (cle), Nilaparvata lugens (nlu), Pediculus humanus (phu), Nicrophorus vespilloides (nve), Tribolium castaneum (tca), Aethina tumida (atu), Sitophilus oryzae (sor), Candidatus Sodalis pierantonius (Spa), Dendroctonus ponderosae (dpo), Anoplophora glabripennis (agl), Agrilus planipennis (apl), Plutella xylostella (pxy), Danaus plexippus (dpl), Manduca sexta (mse), Bombyx mori (bmo), Glossina morsitans (gmo), Drosophila melanogaster (dme), Anopheles gambiae (aga), Aedes aegypti (aae), Nasonia vitripennis (nvi), Apis mellifera (ame), Solenopsis invicta (sin), Camponotus floridanus (cfl).

Pathway	Symbol	dpu	tur	dci	cce	mpe	api	cle	nlu	phu	nve	tca	atu	sor	dpo	agl	apl	pxy	dpl	mse	bmo	gmo	dme	aga	aae	nvi	ame	sin	cfl
AMP	Att	0	0	0	0	0	0	0	0	0	2	1	3	6	1	5	0	0	2	1	3	2	3	1	2	0	0	0	0
AMP	Cec	0	0	0	0	0	0	0	0	0	0	0	0	2	0	4	0	0	0	0	0	2	4	1	6	0	0	0	0
AMP	Col	0	0	0	0	0	0	0	0	0	2	2	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
AMP	Def	0	0	0	0	0	0	2	1	2	1	3	2	1	1	1	1	0	0	0	0	0	1	1	3	5	2	3	2
AMP	Thaumatin	0	0	0	0	4	4	0	0	0	2	3	6	4	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
AMP	Gloverin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	5	0	0	0	0	0	0	0	0
AMP	Lebocin	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0	0	0	0	0	0	0	0
AMP	Lys	3	4	3	1	2	2	6	2	2	8	3	8	5	2	7	3	3	7	6	6	3	12	9	8	1	1	2	2
Recognition	GNBP	0	0	0	1	2	2	1	1	0	1	2	6	3	2	3	3	5	2	3	2	2	2	2	2	2	2	2	2
Recognition	PGRP	0	0	0	0	0	0	0	0	0	12	7	8	6	6	10	4	10	7	9	10	4	6	6	8	8	4	5	4
mi/siRNA	HPS4	0	0	1	0	0	0	0	2	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
miRNA	AGO1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	0	1	1	1
miRNA	Dcr-1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	2	1	1	0	1	1	1	1	1	1	1	1
miRNA	drosha	1	1	0	6	1	1	1	1	1	1	1	1	1	0	1	1	2	0	1	1	1	1	1	1	1	1	1	1
miRNA	Ge-1	1	0	1	1	1	1	1	0	1	1	1	3	1	1	1	0	2	0	1	2	1	1	1	1	1	1	1	1
miRNA	gw	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	2	3	1	1	1	1
miRNA	me31B	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
miRNA	Nbr	0	0	0	0	0	0	0	0	1	1	1	1	1	4	1	1	1	1	1	1	0	1	1	1	1	1	1	1
miRNA	pasha	1	2	1	2	1	2	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
miRNA	r2d2	0	0	0	1	1	1	2	1	0	1	1	1	1	1	1	1	0	0	0	0	2	1	0	0	3	1	1	1
miRNA	Ranbp21	1	0	0	1	1	1	1	2	1	1	1	1	1	3	1	1	1	1	1	0	1	1	1	1	1	1	1	1
piRNA	AGO3	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
piRNA	armi	1	0	0	2	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
piRNA	aubergine	0	0	0	2	2	4	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	4	2	1	1	1
piRNA	fs(1)Yb	1	0	2	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0
piRNA	shu	1	1	1	2	1	2	2	1	1	1	2	1	1	1	1	1	2	1	2	2	0	1	1	1	1	0	1	1
piRNA	zuc	0	0	1	1	1	1	1	0	1	1	1	1	1	0	1	1	2	1	1	1	1	1	0	0	1	1	1	1
si/miRNA	loqs	1	0	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
siRNA	AGO2	0	0	0	5	1	1	2	1	0	1	2	3	1	1	1	1	2	1	1	1	5	1	1	1	3	1	1	1
siRNA	Ars2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
siRNA	bel	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1

siRNA	cbc	2	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
siRNA	Dcr-2	0	0	0	2	1	1	1	1	2	3	1	1	1	2	1	1	3	1	1	1	1	1	1	1	1	1	1	1
siRNA	trsn	1	0	0	0	0	0	1	3	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	2	1	1	1	1
siRNA	Tudor-SN	1	0	0	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
siRNA	vig	1	0	1	0	0	0	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IMD	akirin	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	2	0	1	1	1	1	1	1	1	1	1	1
IMD	ben	0	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
IMD	casp	1	0	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
IMD	Cul1	1	1	0	1	1	1	1	1	2	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
IMD	CYLD	1	1	1	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1	2	1	1	1	1	1	1	1	1	1
IMD	Diap2	1	0	1	2	1	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	0	1	1	1	1	1
IMD	dnr1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	1
IMD	Dredd	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IMD	eff	0	2	1	1	1	1	1	1	0	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	0	1
IMD	Fadd	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1
IMD	IKKβ	1	0	0	0	0	0	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2	1
IMD	imd	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1
IMD	key	1	0	0	0	0	0	0	0	1	1	1	1	1	1	2	1	1	2	3	1	1	1	1	1	1	1	1	1
IMD	Ntf-2	0	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1
IMD	pirk	0	0	0	0	0	0	0	0	0	1	1	1	1	0	1	0	1	1	1	1	0	1	0	0	1	1	1	1
IMD	POSH	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IMD	Rel	1	1	0	0	0	0	1	0	0	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1
IMD	scny	1	0	0	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
IMD	sick	1	1	0	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
IMD	Skp2	1	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1	3	1	0	1	1	1	1	1	1	1	1	1
IMD	SkpA	0	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
IMD	Tab2	0	0	0	0	0	0	0	1	0	1	1	1	1	1	1	1	3	1	1	1	0	1	0	0	1	1	1	1
IMD	Tak1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
IMD	UEV1a	1	1	0	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JAK/STAT	dome	1	0	1	2	2	2	1	0	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	5
JAK/STAT	hop	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1
JAK/STAT	Socs36E	1	0	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1
JAK/STAT	Stam	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	2	1	1	1	1	1	1	1	1	1	1
JAK/STAT	Stat92E	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	2	1	2	1	2	1	1	1	1	1	1	1
JAK/STAT	Su(var)2-10	1	1	2	2	1	4	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JAK/STAT	upd3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
JAK/STAT	zfh1	1	0	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JNK	aop	1	0	4	1	1	1	1	2	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JNK	bsk	2	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	0	1	1	1	1	1	1	1	1	1
JNK	hep	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	0	1	1	1	1	2	1	1	1	2
JNK	Jra	1	0	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
JNK	kayak	1	0	2	0	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Toll	Aos1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Toll	aPKC	1	1	0	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	2	1	0	1	0	1	1	1	1	1
Toll	cact	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	1	1	3	2	2	2
Toll	cactin	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Toll	Deaf1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1
Toll	dl	1	1	0	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1
Toll	Gprk2	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
Toll	grass	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	2	2	4	9	0	0	0	0
Toll	krz	2	2	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	0	1	1	1
Toll	lwr	1	1	0	1	1	1	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Toll	modSP	0	0	0	0	0	0	1	0	0	1	2	2	1	1	1	1	1	3	2	1	4	1	1	1	1	1	2	1
Toll	Myd88	1	1	0	2	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Toll	necrotic	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	1	1	1	1
Toll	Pli	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Toll	pll	1	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1	3	1	0	1	1	1	1	1	2	1	1	1
Toll	psh	0	0	0	0	0	0	0	0	0	1	1	2	1	1	2	1	1	2	1	1	1	2	1	1	1	1	1	1
Toll	ref(2)P	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	1	0	1	1	0	1	0	0	1	1	1	1
Toll	smt3	0	0	1	1	1	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	0	1	2	2	1	1	1	1
Toll	SPE	0	0	0	0	0	0	1	0	0	1	2	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1
Toll	spheroide	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
Toll	sphinx1	0	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	2	2	2	1	1	3	1	1	1	0	1	1
Toll	sphinx2	0	0	0	0	0	0	0	0	0	1	1	0	1	3	3	1	0	0	0	0	1	3	0	0	0	0	0	0
Toll	spirit	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0
Toll	spz	3	1	3	4	4	4	4	4	4	4	4	6	5	3	4	3	5	7	6	6	4	4	4	5	3	4	3	4
Toll	Toll	6	2	7	9	9	12	10	12	7	11	11	15	12	14	10	10	17	19	21	16	9	10	14	15	16	12	12	13
Toll	Tollip	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	1	1	1	1
Toll	Traf4	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
Toll	Traf6	0	0	0	1	1	1	1	0	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	0	0	1	2	1

Toll	tub	1	1	4	1	1	1	1	0	1	1	1	2	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
Toll	Uba2	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1
Toll	Ulp1	1	1	0	1	1	1	1	0	2	2	1	1	1	2	1	1	1	1	1	1	5	1	0	1	1	1	1	1
Toll	ush	1	0	1	1	1	1	1	2	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
Toll	wisp	1	1	0	1	1	1	1	0	1	1	1	1	1	2	1	1	3	1	1	1	2	2	1	1	1	1	1	1

Table S.3 Matrix of counts of the number of genes identified which potentially are members of the immune system signalling pathways. The first column refers to the pathway and the second column to the gene name. The remaining columns refer to a given arthropod. Note that they are taxonomically ordered. Abbreviations are as follows: Daphnia pulex (dpu), Tetranychus urticae (tur), Diaphorina citri (dci), Cinara cedri (cce), Myzus persicae (mpe), Acyrthosiphon pisum (api), Cimex lectularius (cle), Nilaparvata lugens (nlu), Pediculus humanus (phu), Nicrophorus vespilloides (nve), Tribolium castaneum (tca), Aethina tumida (atu), Sitophilus oryzae (sor), Dendroctonus ponderosae (dpo), Anoplophora glabripennis (agl), Agrilus planipennis (apl), Plutella xylostella (pxy), Danaus plexippus (dpl), Manduca sexta (mse), Bombyx mori (bmo), Glossina morsitans (gmo), Drosophila melanogaster (dme), Anopheles gambiae (aga), Aedes aegypti (aae), Nasonia vitripennis (nvi), Apis mellifera (ame), Solenopsis invicta (sin), Camponotus floridanus (cfl).