Genomics of intracellular symbionts in insects

María José Gosalbes a,b, Amparo Latorre a,b, Araceli Lamelas a,b, Andrés Moya a,b,*

a Unidad Mixta de Investigación en Genómica y Salud Centro Superior de Investigación en Salud Pública (CSISP), Institut Cavanilles de Biodiversitat i Biologia Evolutiva (Universitat de València), Apartado Postal 22085, 46071 Valencia, Spain
b CIBER en Epidemiología y Salud Pública (CIBEResp), Spain

ARTICLE INFO

Article history:
Received 24 October 2009
Received in revised form 3 December 2009
Accepted 20 December 2009

Keywords:
Endosymbiosis
Insects
Genome sequence
Functional genomics

ABSTRACT

Endosymbiotic bacteria play a vital role in the evolution of many insect species. For instance, endosymbionts have evolved metabolically to complement their host’s natural diet, thereby enabling them to explore new habitats. In this paper, we will review and give some examples of the nature of the metabolic coupling of different primary and secondary endosymbionts that have evolved in hosts with different nutritional diets (i.e., phloem, xylem, blood, omnivores, and grain). Particular emphasis is given to the evolutionary functional convergence of phylogenetically distant endosymbionts, which are evolving in hosts with similar diets.

© 2010 Elsevier GmbH. All rights reserved.

Introduction

In recent years, genomics and metagenomics have opened up new avenues of research into symbiotic associations. Metagenomic studies of different habitats, such as sea water, ice cores, new avenues of research into symbiotic associations. Metagenomic studies of different habitats, such as sea water, ice cores, marine sponges, or human gut (Venter et al., 2004; Biddle et al., 2007; Schmitt et al., 2007; Bouguen et al., 2007) have led to the discovery of the presence of previously unknown and uncultivable bacteria, protozoa, and viruses. In particular, there are increasing numbers of genome sequences for many symbionts, revealing their complete set of genes as well as their functional contribution to their host’s metabolism.

Symbiotic associations are particularly well studied in insects. Insects are the most species-rich group of organisms, and it has been estimated that at least 15–20% of all insects live in symbiotic relationships with bacteria (Buchner, 1965; Douglas, 1998). According to the location of the symbiont with respect to the cells, associations can be referred to as ectosymbiotic or endosymbiotic. Among intracellular symbioses, there are differences regarding the extent of dependence between the animal host and the symbiont and the age of the association leading to the distinction between obligate primary endosymbiont (P-endosymbiont) and facultative secondary endosymbiont (S-endosymbiont). S-symbionts are considered facultative since they are not essential to host survival and reproduction even though, in some cases, they confer advantages to their hosts (Oliver et al., 2009). Although they are normally transmitted vertically through host generations, their distribution patterns suggest that sporadic horizontal transmission can occur (Russell et al., 2003; Russell and Moran, 2005). Insects that have established endosymbiotic associations with bacteria are characterized, in general, by feeding upon unbalanced diets, poor in essential nutrients such as amino acids, sterols, or vitamins, which are provided by the symbionts. Other symbionts provide different functions to their host, such as nitrogen recycling and storage or the provision of metabolic factors that are required for survival and fertility (Douglas, 1998; Baumann and Moran, 2000). On the other hand, the bacterium gains a permanent supply of a wide range of metabolites that are provided by the host as well as a safe environment. The first step towards obligate endosymbiosis becoming established is when a free-living bacterium infects the host. Then, both organisms coevolve to adapt to the new situation. On the bacterial side, genomic studies have revealed that the endosymbiont genome gets smaller during this adaptive process, owing to the loss of genes that are rendered unnecessary in the new environment (Gil et al., 2004; Zientz et al., 2004; Moya et al., 2008; Moran et al., 2008; Feldhaar and Gross, 2009).

To date, the following complete genomes of bacterial endosymbionts of insects have been fully sequenced: the endosymbionts of 4 aphid species (Buchnera spp.) (Shigenobu et al., 2000; Tamas et al., 2002; van Ham et al., 2003; Pérez-Brocal et al., 2006), 2 S-symbionts of the pea aphid (Hamiltonella defensa and Regiella insecticola) (Degnan et al., 2009a, 2009b), the endosymbiont of 2 carpenter ant...
Table 1
Functional features of intracellular symbionts.

<table>
<thead>
<tr>
<th>Endosymbiont</th>
<th>Host</th>
<th>Association age</th>
<th>Host nutrition</th>
<th>Genome size (kb)</th>
<th>Metabolic capacity</th>
<th>Insertion elements</th>
<th>T3SS</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Buchnera aphidicola</em> Bap (γ-Proteobacteria)</td>
<td><em>Acyrthosiphon pisum</em> (aphid)</td>
<td>150 my</td>
<td>Phloem (primary)</td>
<td>652</td>
<td>Amino acids, vitamins</td>
<td>No</td>
<td>Incomplete T3SS</td>
<td>BA000003</td>
</tr>
<tr>
<td><em>Buchnera aphidicola</em> Bsg (γ-Proteobacteria)</td>
<td><em>Schizaphis graminum</em> (aphid)</td>
<td>150 my</td>
<td>Phloem (primary)</td>
<td>653</td>
<td>Amino acids, vitamins</td>
<td>No</td>
<td>Incomplete T3SS</td>
<td>AE013218</td>
</tr>
<tr>
<td><em>Buchnera aphidicola</em> BBp (γ-Proteobacteria)</td>
<td><em>Baizongia pistaciae</em> (aphid)</td>
<td>150 my</td>
<td>Phloem (primary)</td>
<td>618</td>
<td>Amino acids, vitamins</td>
<td>No</td>
<td>Incomplete T3SS</td>
<td>AE016826</td>
</tr>
<tr>
<td><em>Buchnera aphidicola</em> Bct (γ-Proteobacteria)</td>
<td><em>Cinara tujafilina</em> (aphid)</td>
<td>150 my</td>
<td>Phloem (primary)</td>
<td>445</td>
<td>Amino acids, vitamins</td>
<td>No</td>
<td>Incomplete T3SS</td>
<td>AE014836</td>
</tr>
<tr>
<td><em>Buchnera aphidicola</em> Bcc (γ-Proteobacteria)</td>
<td><em>Cinara cedri</em> (aphid)</td>
<td>–</td>
<td>Phloem (primary)</td>
<td>422</td>
<td>Amino acids, Trp</td>
<td>No</td>
<td>Incomplete T3SS</td>
<td>CP0000263</td>
</tr>
<tr>
<td><em>Serratia symbiontica</em> (γ-Proteobacteria)</td>
<td><em>Cinara cedri</em> (aphid)</td>
<td>–</td>
<td>Phloem (primary)</td>
<td>2100</td>
<td>Parasitoid wasp protection</td>
<td>IS elements</td>
<td>2 complete T3SSs</td>
<td>CP001277</td>
</tr>
<tr>
<td><em>Coronella rudii</em> (γ-Proteobacteria)</td>
<td><em>Pachypsylla venusta</em> (psyllid)</td>
<td>120 my</td>
<td>Phloem (primary)</td>
<td>160</td>
<td>Symiotic role lost</td>
<td>No</td>
<td>No</td>
<td>CP001278</td>
</tr>
<tr>
<td><em>Buchnera cicadellinicola</em> (γ-Proteobacteria)</td>
<td><em>Homalodisca vitripennis</em> (sharpshooter)</td>
<td>170 my</td>
<td>Xylem (primary)</td>
<td>686</td>
<td>Vitamins, cofactors</td>
<td>No</td>
<td>No</td>
<td>CP002238</td>
</tr>
<tr>
<td><em>Sulcia muelleri</em> (Flavobacteria)</td>
<td><em>Homalodisca vitripennis</em> (sharpshooter)</td>
<td>&gt;270 my</td>
<td>Xylem (primary)</td>
<td>246</td>
<td>Amino acids</td>
<td>No</td>
<td>No</td>
<td>CP001277</td>
</tr>
<tr>
<td><em>Hodgkinia cicadellincola</em> (γ-Proteobacteria)</td>
<td><em>Diceroprocta semicincta</em> (singing cicada)</td>
<td>&gt;270 my</td>
<td>Xylem (primary)</td>
<td>277</td>
<td>Amino acids</td>
<td>No</td>
<td>No</td>
<td>CP001605</td>
</tr>
<tr>
<td><em>Wigglesworthia glossinidius</em> (γ-Proteobacteria)</td>
<td><em>Glossina brevipalpis</em> (tsetse fly)</td>
<td>190 my</td>
<td>Xylem (primary)</td>
<td>144</td>
<td>Amino acids</td>
<td>No</td>
<td>No</td>
<td>NC012960</td>
</tr>
<tr>
<td><em>Sodalis glossinidius</em> (γ-Proteobacteria)</td>
<td><em>Glossina morsitans</em> (tsetse fly)</td>
<td>–</td>
<td>Blood (secondary)</td>
<td>4171</td>
<td>Immunity</td>
<td>IS elements</td>
<td>3 complete T3SSs</td>
<td>AP008232</td>
</tr>
<tr>
<td><em>Blochmannia pennsvanicus</em> (γ-Proteobacteria)</td>
<td><em>Camponotus pennsylvanicus</em> (carpenter ant)</td>
<td>50 my</td>
<td>Omnivore (primary)</td>
<td>792</td>
<td>Nitrogen metabolism</td>
<td>No</td>
<td>No</td>
<td>CP000016</td>
</tr>
<tr>
<td><em>Blochmannia floridanus</em> (γ-Proteobacteria)</td>
<td><em>Camponotus floridanus</em> (carpenter ant)</td>
<td>50 my</td>
<td>Omnivore (primary)</td>
<td>706</td>
<td>Nitrogen metabolism</td>
<td>No</td>
<td>No</td>
<td>BX248583</td>
</tr>
<tr>
<td><em>Blattabacterium</em> (Flavobacteria)</td>
<td><em>Blattella germanica</em> (cockroach)</td>
<td>150 my</td>
<td>Omnivore (primary)</td>
<td>637</td>
<td>Nitrogen metabolism</td>
<td>No</td>
<td>No</td>
<td>CP001487</td>
</tr>
<tr>
<td><em>Blattabacterium</em> (Flavobacteria)</td>
<td><em>Periplaneta americana</em> (cockroach)</td>
<td>150 my</td>
<td>Omnivore (primary)</td>
<td>637</td>
<td>Nitrogen metabolism</td>
<td>No</td>
<td>No</td>
<td>CP001429</td>
</tr>
<tr>
<td><em>SOPE</em> (γ-Proteobacteria)</td>
<td><em>Staphylinus oryae</em> (weevil)</td>
<td>Less than 25 my</td>
<td>Grain (primary)</td>
<td>3000^</td>
<td>Amino acids, vitamins</td>
<td>IS elements</td>
<td>Complete T3SS</td>
<td>–</td>
</tr>
</tbody>
</table>

my, Millions of years.

^ Candidatus.
^ Secondary endosymbiont in other aphids.
^ Size estimated by pulse-field gel.
species (Blochmannia spp.)(Gil et al., 2003; Degnan et al., 2005), one symbiont of psyllids ("Candidatus Caronella ruddii") (Nakabachi et al., 2006), the P-endosymbiont (Wigglesworthia glossinidia) and the S-endosymbiont (Sodalis glossinidius) of tsetse flies (Akman et al., 2002; Toth et al., 2006), 2 different endosymbionts coexisting in a xylem-feeding sharpshooter and singing cicadas ("Candidatus Sulcia muelleri", Bauzmannia cicadellinicola, and "Candidatus Hodkginia cicadica") (McCutcheon and Moran, 2007; McCutcheon et al., 2009a, 2009b), and a Blattabacterium strain from the cockroach Blattella germanica (López-Sánchez et al., 2009) as well as from Periplaneta americana (Sabree et al., 2009). Furthermore, 5 genomes of Wolbachia, a widespread intracellular α-proteobacterium, have been sequenced (Wu et al., 2004; Foster et al., 2005; Klasson et al., 2008, 2009; Salzberg et al., 2009). In this review, we will focus on the functional interdependencies between partners from the data provided by genomic analyses (Table 1).

Does insect diet lead the evolutionary pathway towards minimal metabolism?

Genomic and metagenomic analyses have corroborated the nutritional role of such associations, as in each case the bacteria provide the nutrients lacking in the respective host diets. The host has, so to speak, domesticated the endosymbiont, so it devotes part of its gene repertoire to processes that are more beneficial to its host than to itself.

Phloem diet

Several insects are plant phloem-feeding insects, a diet that is rich in sugars but poor in nitrogenous compounds such as amino acids, vitamins, or cofactors. This is the case of aphids, psyllids, whiteflies, and mealybugs. They have established endosymbiotic associations with different bacteria (Buchnera aphidicola, "Candidatus Caronella ruddii", "Candidatus Portiera aleyrodicida" and "Candidatus Temblaya princeps", respectively) (Thao et al., 2000; Munson et al., 1991; Tremblay, 1989; Clark et al., 1992), which provide the nutrients lacking in each case.

Buchnera aphidicola and its aphid hosts are the best-studied insect–symbiont associations. Nowadays, 4 complete genomes of Buchnera aphidicola from 4 different aphids (Acrystosiphon pisum: B. aphidicola BAp, Schizaphis graminum: B. aphidicola BSp, Baisozzia pistaicae: B. aphidicola BBb, and Cnara cedri: B. aphidicola BCc) have been fully sequenced (Shigenobu et al., 2000; Tamas et al., 2002; van Ham et al., 2003; Perez-Brocal et al., 2006; Moran et al., 2009). Furthermore, sequencing of the genome of B. aphidicola BCc from Cnara tujaflina is currently underway in our laboratory. As a consequence of its association with the aphid, the bacterial genome has undergone a profound size reduction. Buchnera genome sequences revealed that some of the genes lost were those required for the synthesis of the nonessential amino acids that the aphid can make themselves. Aphids require 10 essential amino acids that are lacking in their phloem-sap diet and must be provided by their endosymbionts. In fact, nearly 10% of B. aphidicola coding capacity is devoted to amino acid biosynthesis. In 4 out of the 5 of the aphid lineages studied, the structural genes for leucine biosynthesis, and 2 regulatory genes (trpEC) of the tryptophan pathway are located in plasmids, while the rest of the genes of this pathway remain in the chromosome (Lai et al., 1994; Van Ham et al., 1999; Gil et al., 2006; Perez-Brocal et al., 2006). B. aphidicola from C. cedri has retained the biosynthetic capacity for all essential amino acids except tryptophan, whose pathway is incomplete (Gosalbes et al., 2008). The case of tryptophan in C. cedri will be discussed below.

Another characteristic feature of Buchnera genomes is the presence of a simplified flagellar apparatus, homologous to the type III secretion system (T3SS). Since Buchnera species are not motile, it has been proposed that these structures might be involved in invading bacteriocyes to ensure transmission to host offspring, and also a possible protein export function has been described (Shigenobu et al., 2000; Perez-Brocal et al., 2006; Toft and Fares, 2008). Comparison of the available genome sequences reveals deletions and inactivation of ancestral genes in each lineage.

B. aphidicola BSp possesses degenerate copies of genes underlying incorporation of inorganic sulphur, because the incorporation of this compound is unnecessary since the aphid’s grass diet provides the sulphur-containing organic compounds required to produce methionine and cysteine (Tamas et al., 2002; Moran, 2007). This loss implies that S. graminum is destined to remain on grasses or other hosts with sufficient sources of fixed sulphur (Tamas et al., 2002). Similar losses have occurred in B. aphidicola BBp, which lacks the pathway for arginine and panthothenate biosynthesis but retains genes underlying biotin biosynthesis, which have been lost in genomes of B. aphidicola from S. graminum, A. pisum, C. tujaflina, and C. cedri (van Ham et al., 2003; Perez-Brocal et al., 2006; Lamelas et al., unpublished). Hence, some of the observed gene losses are expected to affect host ecology.

In addition to B. aphidicola, some aphid populations harbour facultatively symbiotic intracellular bacteria. In the aphid C. cedri, the S-endosymbiont “Candidatus Serratia symbiotica” (hereafter, S. symbiotica) is reported to have become an obligate symbiont (Lamelas et al., 2008). Indeed, there are recent reports of a close endosymbiotic consortium that involves B. aphidicola and S. symbiotica in C. cedri (Gosalbes et al., 2008) (Fig. 1). Applying a metagenomic approach, we discovered a plasmid in B. aphidicola from C. cedri containing the trpEC genes, coding for anthranilate synthase, the first enzyme of the tryptophan biosynthesis pathway. This plasmid showed homology with all the tryptophan plasmids described previously for several B. aphidicola strains (Lai et al., 1994; Van Ham et al., 1999; Gil et al., 2006). The remaining genes for the pathway (trpDCBA) are located on the chromosome of S. symbiotica. These data show that both endosymbionts, B. aphidicola and S. symbiotica, are involved in the tryptophan biosynthesis that supplies this essential amino acid to both their host and themselves. Therefore, the obligate biochemical interdependence between 2 endosymbionts can represent an evolutionary seal of the bacterial metabolic complementation, and the establishment of a stable consortium would be the expected evolutionary outcome.

Another facultative symbiont, H. defensa, occurs in aphids and others sap-feeding insects such as psyllids or whiteflies. This S-endosymbiont protects its host by killing the parasitoid wasp larvae and allowing the insect to survive and reproduce. In the pea aphid (A. pisum), the degree of protection varies among Hamiltonellina strains and depends on toxins that are produced by the bacteria and by a temperate lambda-like bacteriophage, which infects H. defensa. The complete genome sequence (2.1 Mb) of H. defensa from A. pium has a complete inventory of pathogenicity factors such as 2 T3SS or diverse putative toxins (Moran et al., 2005; Degnan et al., 2009a). The genome also presents mobile DNA and insertion-sequence elements, indicating a dynamic nature and horizontal gene transfer. Related to the metabolism, this endosymbiont is host-dependent, synthesizing only 2 of the essential amino acids and most vitamins except thiamine and pantothenate. The diet requirements of both H. defensa and the aphid must rely on another endosymbiont, mainly Buchnera.
Psyllids are the other phloem sap-feeding insects, and the genome sequence of their primary endosymbiont, “Candidatus Carsonella ruddii” (hereafter C. ruddii), has been obtained (Thao et al., 2000; Nakabachi et al., 2006). The genome sequencing of C. ruddii from the psyllid Pachyprosypylla venusta revealed that it possess a 160-kb genome with only 182 predicted open reading frames, with many genes been lost (Nakabachi et al., 2006). In fact, about half of the biosynthesis pathways for essential amino acids are either partially or entirely missing, indicating a severe impairment of its symbiotic function. Since no secondary endosymbionts appear to be associated with these psyllids, it is difficult to understand how the hosts can thrive on their diet without additional symbiotic help. Furthermore, the ability of the cell to perform the most essential life-related functions is heavily impaired by the lack of genes involved in DNA replication, transcription, and translation. Therefore, it has yet to be clarified whether C. ruddii is on its way to becoming an organelle, or if it is nearing extinction and replacement by an unidentified symbiont (Tamames et al., 2007).

**Xylem diet**

Xylem is the component of the plant vascular system that is used to transport water and salts from the roots to the rest of the plant. Xylem sap is low in nutrients, containing mostly inorganic compounds and minerals, as well as small amounts of non-essential amino acids (glutamate, aspartate), sugars (primarily glucose), and organic acids (primarily malate) (Redak et al., 2000; Nakabachi et al., 2006). The genome sequencing of C. ruddii from the psyllid Pachyprosypylla venusta revealed that it possesses a 160-kb genome with only 182 predicted open reading frames, with many genes been lost (Nakabachi et al., 2006). In fact, about half of the biosynthesis pathways for essential amino acids are either partially or entirely missing, indicating a severe impairment of its symbiotic function. Since no secondary endosymbionts appear to be associated with these psyllids, it is difficult to understand how the hosts can thrive on their diet without additional symbiotic help. Furthermore, the ability of the cell to perform the most essential life-related functions is heavily impaired by the lack of genes involved in DNA replication, transcription, and translation. Therefore, it has yet to be clarified whether C. ruddii is on its way to becoming an organelle, or if it is nearing extinction and replacement by an unidentified symbiont (Tamames et al., 2007).

However, these γ-proteobacteria present a very limited set of essential amino acid biosynthesis pathways. The few capabilities that are present include histidine biosynthesis and synthesis of methionine if external homoserine is provided. This loss implies that both the host and *Baumannia* must obtain the amino acids from another external source. In this respect, *S. muelleri* plays an important role, producing 8 (lysine, valine, threonine, leucine, isoleucine, phenylalanine, arginine, tryptophan) of the 10 essential amino acids, despite having a small genome. The lack of vitamin and cofactor synthesis pathways in Sulcia indicated that both endosymbionts play complementary and not overlapping roles in the symbiotic system. Not only do they appear to provide different resources for the host, but each does not present the biosynthesis pathways corresponding to the resources synthesized by the other. A similar metabolic consortium has been observed, as described above, in phloem-feeding *C. cedri* with *Buchnera aphidicola* and *S. symbiotica* as endosymbionts. The complementarity between the host and each symbiont extends to mutual dependence between the symbionts that appear to depend on each other for the required compounds and for intermediaries in the metabolic processes.

Recently, a similar metabolic system has been described in singing cicadas, related to the glassy-winged sharpshooter (*McCutcheon et al., 2009a*). They feed exclusively on xylem sap from plant roots during their long underground juvenile stage, and they also harbour 2 endosymbionts: *S. muelleri* (272 kb) and an α-proteobacterium “Candidatus Hodgkinia cicadicola” (hereafter *H. cicadicola*) with the smallest genome described (144 kb). The *H. cicadicola* genome displays some other atypical features such as a high GC content (58.4%) and a recoding of UGA Stop-Trp (McCutcheon et al., 2009b). The *Sulcia* genome has nearly the same gene content and gene order as *Sulcia* from sharpshooter. In cicadas, this microorganism produces 8 of the essential amino acids while the remaining 2 are provided by *Hodgkinia*. Although they belong to different bacterial groups, *Baumannia* (γ-Proteobacteria) and *Hodgkinia* (α-Proteobacteria) play the same functional role for amino acid synthesis, except for the methionine biosynthetic pathway. *Hodgkinia* has lost 2 genes (metA and metB) essential to cystathione synthesis, an intermediary of the methionine pathway, as cicada feed on plant root exudate rich in this compound. Another striking feature is the use of cobalamin-dependent methionine synthase (MetH) by *Hodgkinia* instead of MetE (cobalamin-independent methionine synthase) for the last step in methionine biosynthesis like the other insect symbionts. The preference for MetH implies the

**Fig. 1.** Ultrathin section from *Cinara cedri* showing the bacteriocytes of the 2 symbionts that form the functional consortium. B. Buchnera aphidicola; S. Serratia symbiotica.
presence of cobalamin biosynthesis genes that have never been found in other symbiont genomes before. Unlike *Baumannia, Hodgskinia*, with a drastically reduced genome, has lost all vitamin and cofactor biosynthetic abilities, implying that the cicada and its symbionts must have access to external sources of these compounds, possibly from plant root xylem since cicadas spend most of their lives underground. On the other hand, the translational apparatus is incomplete in *Sulcia* and *Hodgskinia*, and it is still unknown whether this process occurs in these microorganisms as it does in *C. ruddii*.

**Blood diet**

Tsetse flies (*Glossina* spp.) are the vector for the trypanosome parasite, agents of sleeping sickness in humans and nagana in animals. These insects feed exclusively on vertebrate blood, which is vitamin deficient, mainly in B-complex vitamins. Tsetse flies harbour 2 symbiotic bacteria in gut tissue that are maternally transmitted to intrauterine larvae via milk secretions of the mother. *W. glossinidia* (γ-Proteobacteria) is the obligate primary endosymbiont and *S. glossinidius* (γ-Proteobacteria) is the communal secondary symbiont (*Aksoy*, 1995). Some populations of tsetse flies also harbour a third organism that belongs to the genus *Wolbachia* (γ-Proteobacteria) (*Cheng* et al., 2000). The complete genome sequence (698 kb) of *W. glossinidia* shows that the endosymbiont has retained 62 genes (10% of coding sequences) devoted to the biosynthesis of cofactors, prosthetic groups, and carriers, including B vitamins (*Akman* et al., 2002). As in the previous cases, this fact coupled with data from dietary supplementation experiments of antibiotic-fed, symbiont-free flies, show that the ability to reproduce was partially restored (Nogge, 1976). The small genome of *Wigglesworthia* presents the machinery for the synthesis of a complete flagellar apparatus. However, neither flagellum nor motility has been observed in this endosymbiont. Since the genome does not seem to encode a secretion system to penetrate the larval cells, the flagellar structure may work like a T3SS, as described in the well-established P-endosymbiont *Buchnera* (*Pérez-Brocal* et al., 2006; *Aksoy* and *Rio*, 2005).

*S. glossinidius* is the S-endosymbiont residing intra- and extracellularly in tsetse flies, principally in the midgut tissue, and it is the only insect symbiont to have been cultured (*Matthew* et al., 2005). The *Sodalis* genome (4 Mb) has reduced in size compared to its free-living relatives, but it is significantly larger than those of P-symbionts. In addition, it has 3 extrachromosomal plasmids and in some strains a bacteriophage-like element (*Darby* et al., 2005). The chromosome also contains an unusually high number of pseudogenes, suggesting a degradation of the genes whose function is no longer needed by this organism (*Akman* et al., 2002; *Toh* et al., 2006). All these features indicate its recent symbiotic affiliation. This symbiont has apparently retained many of the abilities of free-living bacteria, including vitamin and cofactor biosynthesis pathways. Thus, *Sodalis* benefits its tsetse host via the synthesis of these compounds, which are deficient in a blood diet, as well as *W. glossinidia*, the primary endosymbiont. Meanwhile, these bacteria have lost genes for carbon compound catabolism, central intermediary metabolism and, fatty acid phospholipid metabolism. In general, *Sodalis* seems to have retained synthetic rather than degradative, showing this bacterium has adapted to the energy-rich environment and nutritional ecology of the host insect, i.e. blood. The *Sodalis* chromosome encodes 3 putative T3SSs and, as occurring in other symbionts, they have been related to progeny transmission and the establishment of symbiosis (*Toh* et al., 2006). In experiments where *Sodalis* were selectively eliminated from tsetse, the trypanosome infection rate was significantly reduced. Thus, *Sodalis* plays a role in the efficacy of the host’s immune system. Since *Sodalis* can be cultivated in vitro, a new genetic transformation system, paratransgenesis, has recently been developed, by which foreign products are inserted and expressed in *Sodalis* and, in turn, expressed in the tsetse flies; furthermore, they are passed on to multiple generations. This technique can be used to diminish the vectorial capacity of flies by manipulating the relationship between host (tsetse fly), parasite (trypanosome), and endosymbiont (*Sodalis*) (*Dale* and *Welburn*, 2001; *Aksoy* and *Rio*, 2005; *Aksoy* et al., 2008).

Another interesting case of an insect with blood diet is the mosquito *Aedes aegypti*, the primary vector of dengue virus. *Wolbachia* species are obligatory endosymbionts of a wide range of invertebrates, including mosquitoes, fruit flies, and filarial nematodes. Recently, *W. aegypti* has been artificially infected with *Wolbachia pipiensis* that reduces the lifespan of insects (*McMeniman* et al., 2009). *Burley* et al. (2009) have shown that the old *Wolbachia*-infected mosquitoes present also a limited blood-feeding capability. Hence, *Wolbachia* infection could be considered as a biocontrol strategy given that only old mosquitoes transmit dengue.

**Omnivorous diet**

Ants, social insects, have motivated great interest due to their evolutionary success in terms of species richness and their extraordinary abundance in tropical regions. These insects have also managed to spread into extreme habitats, such as arctic regions or deserts. In the species-rich genus *Camponotus*, carpenter ants, the intracellular bacterium *Blochmannia* has been described, located in bacteriocytes intercalated between midgut cells and in ovaries of females (*Blochmann*, 1892). The diet of *Camponotus* ranges from nutrient-deficient plant secretions or homoptera exudate to complex sources like dead insects, bird excrement, or sweet food waste. Thus, given the ants' omnivorous diet, the nutritional role of *Blochmannia* was not clear. However, the genome sequences of “*Candidatus* Blochmannia floridanus” and “*Candidatus* Blochmannia pennsylvanicus” (hereafter *Blochmannia floridanus* and *Blochmannia pennsylvanicus*, respectively) showed that these endosymbionts contribute to nitrogen, sulphur, and lipid metabolism of the host (*Gil* et al., 2003; *Degnan* et al., 2005). A prominent feature of ant endosymbionts is the presence of a complete urease gene cluster (*Feldhaar* et al., 2007). This enzyme hydrolyzes urea to CO₂ and ammonia. This latter compound is a potent cell poison that is recycled by glutamine synthetase into amino acid before toxic concentrations are accumulated. *Blochmannia* has retained the biosynthetic pathways for essential amino acids (except arginine), while those for several non-essential amino acids have been lost. With respect to arginine amino acid, the endosymbiont presents the enzymes that catalyze citrulline synthesis from ornithine. Thus the endosymbiont helps the ants to recycle nitrogen compounds in developmental phases, such as metamorphosis when high anabolic activities support little or no uptake of substrates. When required, the stored nitrogen may be mobilized by the action of host cell-derived arginases and *Blochmannia* urease. In some pathogenic microorganisms, ureases have been identified as virulence factors, whereas in this symbiotic association, it has become beneficial. Therefore, *Blochmannia* confers a significant fitness advantage to its host by enhancing its competitiveness when compared with other ant species lacking this endosymbiont. Unlike other endosymbionts, *Blochmannia* has lost the flagellar apparatus that has been described as both a transport system and invasion mechanism of bacteriocytes, ovaries, or embryos.
Cockroaches, like ants, have an omnivorous diet. They harbour an endosymbiont, Blattabacterium, in bacteriocytes located in their abdominal fat body. Blattabacteria belong to the class Flavobacteria in the phylum Bacteroidetes. Studies into genome size have revealed the genome sizes of the endosymbionts of 3 cockroach species, Periplaneta americana, Blatta orientalis, and Blattella germanica, to be approximately 650 kb (López-Sánchez et al., 2008). Recently, the genome sequences of Blattabacterium strain Bge and Blattabacterium sp. BPLAN, the primary endosymbionts of B. germanica and P. americana, respectively, have been determined and shown to have both a genome size of 637 kb (López-Sánchez et al., 2009; Sabree et al., 2009). Blattabacterium strain Bge genome presents a striking trait, a complete urea cycle. This feature has been described in only one member of the Bacteroidetes phylum, a cellulolytic soil bacterium (Xie et al., 2007). The presence of closely related T3SS-encoding genes has been found in other weevil groups (Lefevre et al., 2004). The presence of closely related T3SS-encoding genes has been found in Sodalis and SZPE, and their expression is coincident with the timing of bacteriome infection in the developing weevil. Thus, T3SS was acquired by a common ancestor of SPZE and Sodalis and was retained as an essential element in the establishment and maintenance of infection and evolved in symbiotic relationships with different hosts that have distinct ecological niches and different specialized diets.

Grain diet

Grain is also an unbalanced diet, very rich in starch and carbohydrates but deficient in other nutrients. Grain weevils (Sitophilus species) harbour a primary endosymbiont that lives in bacteriocytes located surrounding the fore-midgut junction of the insect and in the apex of female ovaries. The best studied primary endosymbionts are SOPE and SZPE from Sitophilus oryzae and S. zeamais, respectively. SOPE, γ-Proteobacteria, maintains an obligate mutualistic endosymbiosis with the host and has not been cultured outside the weevil. Experiments based on heat treatments eliminated SOPE from the host, thus enabling its role within the association to be studied. These bacteria provide amino acids such as phenylalanine and proline as well as vitamins (Heddi et al., 1999). Comparative genomics using microarray analysis showed that in this endosymbiont the greatest gene losses have occurred in those involved in cellular processes (Río et al., 2003). The SOPE genome has been estimated by pulse-field gel electrophoresis at about 3 Mb, without observing drastic genome size reduction (Charles et al., 1997). Phylogenetic studies indicate that SOPE is closely related to S. glossinidius (Heddi et al., 1998). Partial sequencing of the SOPE and SZPE genomes shows a large number of insertion sequences, indicating that they have recently acquired an obligate intracellular way of life (Gil et al., 2008; Plague et al., 2007). Furthermore, it has been proposed that the grain weevil symbionts replaced a more ancient symbiont, which is still retained in other weevil groups (Lefevre et al., 2004). The presence of closely related T3SS-encoding genes has been found in Sodalis and SZPE, and their expression is coincident with the timing of bacteriome infection in the developing weevil. Thus, T3SS was acquired by a common ancestor of SPZE and Sodalis and was retained as an essential element in the establishment and maintenance of infection and evolved in symbiotic relationships with different hosts that have distinct ecological niches and different specialized diets.

Fig. 2. Comparison of nitrogen recycling metabolism deduced from the genome sequences of Blochmannia floridanus and Blattabacterium from Camponotus floridanus and Blattella germanica, respectively.
prokaryote–prokaryote endosymbiosis, found nowhere else in nature. *Tremblaya princeps*, a β-proteobacterium, harbours within its cell other Gram-negative bacteria (secondary endosymbiont) belonging to the β-Proteobacteria. Metagenomic analyses of this association will lead to an understanding of the nutritional, physiological, and ecological interactions between their members (Baumann et al., 2002; Kono et al., 2008). The whitefly *Bemisia tabaci* harbours *Portiera aleyrodicida*, an obligatory symbiotic bacterium, as well as several secondary symbionts including *Rickettsia*, *Hamiltonella*, *Wolbachia*, *Arsenophonus*, *Cardinium*, and *Fritschea*, the function of which has yet to be studied in depth. *B. tabaci* is a species complex composed of numerous biotypes, which may differ from each other both genetically and biologically, and where the different biotypes carry different S-endosymbionts (Baumann et al., 2004; Chiel et al., 2007). In summary, genomics and metagenomics constitute powerful tools that can be used in the analyses of complex symbiotic consortia, in order to understand their role in the dynamics and evolution of insect populations.

**Acknowledgements**

Financial support was provided by grants BFU2006/06003/BMC and BFU2009-12895-C02-01/BMC (Ministerio de Ciencia e Innovación, Spain) to A. Latorre and the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement number 212894 and Prometeo/2009/092 (Conselleria d’Educació, Generalitat Valenciana, Spain) to A. Moya. A. Lamelas is funded by a predoctoral fellowship from the Generalitat Valenciana (Spain).

**References**

Akman, L., Yamashita, A., Watanabe, H., Shiba, T., Hattori, M., Oshima, K., Shiba, T., Hattori, M., Aksoy, S., 2008. Paratransgenesis applied for control of tsetse flies, *Tremblaya princeps*, a β-proteobacterium, harbours within its cell other Gram-negative bacteria (secondary endosymbiont) belonging to the β-Proteobacteria. Metagenomic analyses of this association will lead to an understanding of the nutritional, physiological, and ecological interactions between their members (Baumann et al., 2002; Kono et al., 2008). The whitefly *Bemisia tabaci* harbours *Portiera aleyrodicida*, an obligatory symbiotic bacterium, as well as several secondary symbionts including *Rickettsia*, *Hamiltonella*, *Wolbachia*, *Arsenophonus*, *Cardinium*, and *Fritschea*, the function of which has yet to be studied in depth. *B. tabaci* is a species complex composed of numerous biotypes, which may differ from each other both genetically and biologically, and where the different biotypes carry different S-endosymbionts (Baumann et al., 2004; Chiel et al., 2007). In summary, genomics and metagenomics constitute powerful tools that can be used in the analyses of complex symbiotic consortia, in order to understand their role in the dynamics and evolution of insect populations.


