Evolution of Prokaryote-Animal Symbiosis from a Genomics Perspective

Rosario Gil, Amparo Latorre, and Andrés Moya

Contents

Abstract Symbioses involving prokaryotes living in close relationship with eukaryotic cells have been widely studied from a genomic perspective, especially in the case of insects. In the process toward host accommodation, symbionts experience major genetic and phenotypic changes that can be detected in comparison with free-living relatives. But, as expected, several scenarios allowed the evolution of symbiotic associations, from the first stages of free-living bacteria,

R. Gil and A. Latorre
Institut Cavanilles de Biodiversitat i Biologia Evolutiva and Departament de Genètica, Universitat de València, Apartado Postal 22085, 46071 València, Spain
CIBER en Epidemiología y Salud Pública (CIBERESP), Spain

A. Moya (✉)
Institut Cavanilles de Biodiversitat i Biologia Evolutiva and Departament de Genètica, Universitat de València, Apartado Postal 22085, 46071 València, Spain
CIBER en Epidemiología y Salud Pública (CIBERESP), Spain
Centro Superior de Investigación en Salud Pública (CSISP), Avenida de Cataluña 21, 46020 València, Spain
e-mail: andres.moya@uv.es

through secondary and facultative symbiosis, towards the final point of obligate primary endosymbiosis. Particular relevance has the association formed by the coexistence of several symbionts into a given host. A summary of findings in this field, as well as an evolutionary scenario to explain these changes, is presented in this chapter.

1 Introduction

The term symbiosis refers to the close ecological relation between two (or more) species, able to report benefits to all (mutualism) or some of the implied organisms, with or without harm of one of the involved species (parasitism or commensalism, respectively). Symbiosis is an important source of evolutionary innovation, with examples in the entire Biosphere, being even at the origin of the eukaryotic cell (Margulis 1993). Since then, stable symbioses have evolved independently many times in diverse groups of eukaryotes (Moya et al. 2008). Most symbioses have a demonstrated biochemical basis: in some cases one of the partners benefits from organic compounds produced by the other; in others, its waste products (mainly nitrogen compounds) are recycled by the other. In mutualistic symbioses, matter and energy flow in both directions, so that both partners benefit from the association.

Numerous eukaryotic groups maintain a mutualistic relationship with prokaryotic cells, especially because many eukaryotic lineages present limited metabolic capabilities. Animal metabolism, in particular, is relatively narrow, and essential molecules (such as amino acids, vitamins, or fatty acids) must be retrieved from the environment for survival. Animals with specialized feeding behaviors tend to establish symbiotic associations with microorganisms, which provide the nutrients that are deficient in their diets. In fact, most intracellular mutualistic symbioses between bacteria and animals that have been analyzed at the genomic level (involving insects, nematodes and deep-sea animals, Table 1) are related to nutrient provision. Regarding insects, the most studied and diverse invertebrate group on earth, the presence of such associations throughout most of their evolutionary history suggests that symbiosis has been a driving force in the diversification of the group.

A high proportion of mutualistic symbiotic relationships established by insects imply the participation of bacteria. Frequently, the association is so tight that it is called endosymbiosis, when the bacteria (endosymbiont) obligatorily live inside specialized eukaryotic cells (bacterocytes), which can even form a specialized organ (the bacteriome), located inside the abdominal cavity of the insect. It has been estimated that up to a 15% of all insects species carry bacterial endosymbionts (Baumann 2005), attributing to them the great adaptive success of the Insecta class, by making possible the colonization of new ecological niches and allowing them to feed on restricted diets (such as plant sap, cereals or blood), poor in some essential nutrients that are provided by the endosymbionts. The elimination of these bacteria,

Table 1 Remarkable data of the mutualistic animal endosymbiont genomes sequenced as of November 2009

Stage of symbiosis	Organism	Phylum/class	Host	Symbiotic function	Genome size (Kb)	G + C content (%)	Repetitive DNA (%)	Accession number
S-symbionts	Sodalis glossinidius str. morsitans	Proteobacteria/ γ-Proteobacteria	Glossina morsitans (tsetse fly)	Related with trypanosome infections	4,171	54.7	Not indicated[b]	AP008232 AP008233 AP008234 AP008235 CP001278
	Hamiltonella defensa[a] 5AT	Proteobacteria/ γ-Proteobacteria	Acyrthosiphon pisum (aphid)	Resistance to parasitoid wasps	2,110	40.0	21	
	Regiella insecticola	Proteobacteria/ γ-Proteobacteria	Acyrthosiphon pisum (aphid)	Resistance to parasitoid wasps and fungal pathogens	>2,035	42.2	>14	ACYF01000000
	Wolbachia pipientis wPip	Proteobacteria/ α-Proteobacteria	Culex pipiens (mosquito)	Reproductive parasite	1,482	34.2	Not indicated[b]	AM999887
	Wolbachia pipientis wRi	Proteobacteria/ α-Proteobacteria	Drosopila simulans (fruit fly)	Reproductive parasite	1,446	35.0	22.1	CP001391
	Wolbachia pipientis wMel	Proteobacteria/ α-Proteobacteria	Drosophila melanogaster (fruit fly)	Reproductive parasite	1,268	35.2	8.9	NC_002978
P-endosymbionts	Ruthia magnifica[a]	Proteobacteria/ γ-Proteobacteria	Calyptogena magnifica (deep-sea-clam)	Nutrient provision (amino acids, vitamins and cofactors), Nitrogen recycling	1,200	35.0	0	CP000488
	Vesicomyosocius okutanii[a]	Proteobacteria/ γ-Proteobacteria	Calyptogena okutanii (deep-sea-clam)	Nutrient provision (amino acids, vitamins and cofactors), Inorganic Carbon fixation, Nitrogen recycling	1,022	31.6	0	AP009247

Table 1 (continued)

Stage of symbiosis	Organism	Phylum/class	Host	Symbiotic function	Genome size (Kb)	G + C content (%)	Repetitive DNA (%)	Accession number
	Wolbachia pipientis wBm	Proteobacteria/ α-Proteobacteria	Brugia malayi (nematode)	Nutrient provision (nucleotides, vitamins and cofactors)	1,080	34	0	AE017321
	Blochmannia pennsylvanicus[a]	Proteobacteria/ γ-Proteobacteria	Camponotus pennsylvanicus (carpenter ant)	Nutrient provision (amino acids), Nitrogen storage	792	29.6	0	CP000016
	Blochmannia floridanus[a]	Proteobacteria/ γ-Proteobacteria	Camponotus floridanus (carpenter ant)	Nutrient provision (amino acids), Nitrogen storage	706	27,4	0	BX248583
	Wigglesworthia glossinidia	Proteobacteria/ γ-Proteobacteria	Glossina brevipalpis (tsetse fly)	Nutrient provision (vitamins)	698	22.5	0	BA000021 AB063523
	Buchnera aphidicola BSg	Proteobacteria/ γ-Proteobacteria	Schizaphis graminum (aphid)	Nutrient provision (amino acids, vitamins, fatty acids)	653	25.3	0	AE013218 AF041836 Z21938
	Buchnera aphidicola BAp	Proteobacteria/ γ-Proteobacteria	Acyrthosiphon pisum (aphid)	Nutrient provision (amino acids, vitamins, fatty acids)	652	26.3	0	BA000003 AP001070 AP001070
	Buchnera aphidicola BBp	Proteobacteria/ γ-Proteobacteria	Baizongia pistaciae (aphid)	Nutrient provision (amino acids, vitamins, fatty acids)	618	26.0	0	AE016826 AF492591
	Carsonella ruddi[a]	Proteobacteria/ γ-Proteobacteria	Pachypsylla venusta (psyllid)	Nutrient provision?	160	16.0	0	AP009180
	Blattabacterium strain Bge	Bacteroidetes/ Flavobacteria	Blattella germanica (cockroach)	Nutrient provision (amino acids) and Nitrogen excretion	637	27.1	0	CP001487

Consortia	*Blattabacterium* strain BPLAN	Bacteroidetes/ Flavobacteria	*Periplaneta americana* (cockroach)	Nutrient provision (amino acids) and Nitrogen excretion[c]	637	28.2	0	CP001429 CP001430
	Buchnera aphidicola BCc	Proteobacteria/γ-Proteobacteria	*Cinara cedri* (aphid)	Nutrient provision (amino acids, vitamins, fatty acids)	422	20.1	0	CP000263 AY438025
	Sulcia muelleri GWSS	Bacteroidetes/ Flavobacteria	*Homalodisca vitripennis* (sharpshooter)	Nutrient provision (amino acids)	246	22.0	0	CP000770
	Baumannia cicadellinicola Hc	Proteobacteria/γ-Proteobacteria	*Homalodisca vitripennis* (sharpshooter)	Nutrient provision (amino acids and vitamins)	686	33.2	0	CP000238
	Sulcia muelleri SMDSEM	Bacteroidetes/ Flavobacteria	*Dieroprocta semicincta* (cicada)	Nutrient provision (amino acids)	276	22.0	0	CP001605
	Hodgkinia cicadicola[a] Dsem	Proteobacteria/α-Proteobacteria	*Dieroprocta semicincta* (cicada)	Nutrient provision (amino acids and vitamins)	144	58.0	0	CP001226

[a]These bacteria are called *Candidatus*

[b]These genomes contain repetitive DNA, but the relative amounts are not indicated on the bibliography

[c]Although the authors postulate a different interpretation of the metabolic capabilities (Sabree et al. 2009), a detailed metabolic analysis reflects that both *Blattabacterium* strains have the same mutualistic role, as described by López-Sánchez and coworkers for the Bge strain (López-Sánchez et al. 2009)

consequently, critically diminishes the biological fitness of the host, affecting to its growth, fertility or longevity.

The first to notice the link between a restricted diet and the presence of endosymbiotic bacteria in insects was Paul Buchner (1965), who coined the terms of primary (P-) endosymbiont and facultative or secondary (S-) symbiont, based on its morphologic characterization and its presence among the individuals of a certain taxonomic group. This classification has later been validated by means of molecular genetics techniques and the complete sequencing of genomes of an increasing number of endosymbiotic bacteria (Baumann 2005). Buchner classified as P-endo-symbionts those bacteria of a unique morphological type that are present in all the insects of a defined taxonomic group, confined inside specialized insect cells located in the abdominal cavity. Such P-endosymbionts are essential for its host fitness and survival. On the other side, the S-symbionts were identified as morpho-logically diverse bacteria, without a defined spatial distribution in the host body, and whose sporadic presence in some individuals of a defined taxon suggested that they were not essential for host survival. In fact, S-symbionts vary in number and distribution among species and among individuals of the same species, and can live outside of the eukaryotic cells. The congruence between the phylogenetic trees based on host and their corresponding P-endosymbionts sequences, respectively, indicate that each endosymbiont derives from a single infection of the ancestor of the host by the ancestors of its P-endosymbiont, and follows a path of vertical evolution, promoted by their exclusively maternal transmission between insect generations (Munson et al. 1992). On the contrary, the topological incongruence between the phylogenetic trees based on sequences of the S-symbionts and their hosts and the polyphyletic character of such bacteria suggest the existence of multiple events of infection and/or the horizontal transfer of these bacteria among insects (Russell et al. 2003).

Buchner and the early researchers of prokaryote–eukaryote symbioses did not differentiate between the two prokaryotic domains, since the existence of archaea was not recognized until the nineties of the past century (Woese et al. 1990). Therefore, certain early symbioses described as involving "bacteria" were, in fact, involving archaea (Hackstein et al. 2006). This was the case for many methanogenic symbionts hosted by protists that were described in termite and cockroach guts in the 1980s. However, symbioses between arthropods and metha-nogenic archaea do not seem to have a nutritional foundation. The archaea are always restricted to the hindgut, where they can appear free in the gut lumen, attached to digesta or to the hindgut wall, or as endosymbionts of anaerobic ciliated protozoa that occupy the same gut compartment. Little is known about the function of methanogenic archaea in the guts of arthropods, besides their role in lowering H_2 partial pressure by producing methane, while the archaea uses the hydrogen as a source for methane formation, which indicates that the relationship is mutualistic. Phylogenetic studies have been performed on anaerobic heterotrichous ciliates that keep an endosymbiotic association with methanogenic bacteria (van Hoek et al. 2000). This is an interesting study group because they live in the most divergent

niches, such as marine and freshwater sediments and the intestinal tract of animals. The topology of the phylogenetic trees indicates that the coevolution of host and endosymbiont can only be demonstrated in a few analyzed cases, which most likely means that, although probably hydrogenosome-bearing ciliates acquired methanogenic endosymbionts at the very beginning of their evolution towards anaerobiosis, prior to the anaerobic heterotrichous ciliates radiation, endosymbiont replacements must have accompanied the evolution of these protists. In addition to its role in hydrogen transfer, it has also been proposed that artropode intestinal methanogens can contribute to nitrogen–carbon balance in the hindgut by the fixation of atmospheric nitrogen, since these archaea posses a complete gene repertoire needed for nitrogen fixation (Raymond et al. 2004). Whole genome studies will help to identify other possible benefits of these methanogenic archaea to their hosts.

Even though new data are accumulating on archaeal symbionts of animals, most analyses concentrate on nutritional and physiological aspects. At the beginning of the genomics era, research on prokaryote endosymbionts of eukaryotic cells focused on a limited group of arthropods, mostly sap-sucking insects (Hemiptera: Sternorrhyncha), and those have been for quite a while the main models used to define the evolutionary and molecular aspects of prokaryote-animal symbioses. Therefore, we will focus mostly on bacterial endosymbionts of insects to detangle the molecular aspects of these symbioses from a genomics perspective, paying special attention to the genomic changes experienced by the bacterium in their adaptation to an endosymbiotic lifestyle.

The advent of genomics allowed the complete sequencing of genomes and the development of metagenomic methods, making possible the study of environmental samples and non-cultivable microorganisms, thus offering new opportunities for symbiosis research. The availability of many genomes of bacterial endosymbionts, opens the door to comparative analyses among them, unveiling common molecular aspects regarding the establishment and maintenance of symbiotic associations. In order to completely understand the different stages of genomic evolution of bacterial endosymbionts, it became necessary to analyze and compare genome sequences from endosymbionts in different stages of their symbiotic integration. These comparative analyses allowed researchers in the field to define a plausible scenario for the process of symbiotic integration, from a free-living bacterium to an obligate mutualistic lifestyle (Moya et al. 2009) (Fig. 1). The first step towards the establishment of an obligate intracellular mutualistic symbiosis takes place when a free-living bacterium infects an eukaryotic host. From this point, both organisms will co-evolve to adapt to the new situation. The host develops specialized cells to harbor the bacterium, which in turn provides benefits to the host that end up being essential. From an evolutionary point of view, this new stable situation triggers a cascade of changes that model the shape and content of the bacterial genome. In the course of this chapter, we will see how genomics and metagenomics studies helped researchers on the field to detangle the physiological and evolutionary changes that bacteria experience in their way towards an obligatory mutualistic intracellular symbiosis with eukaryotic hosts.
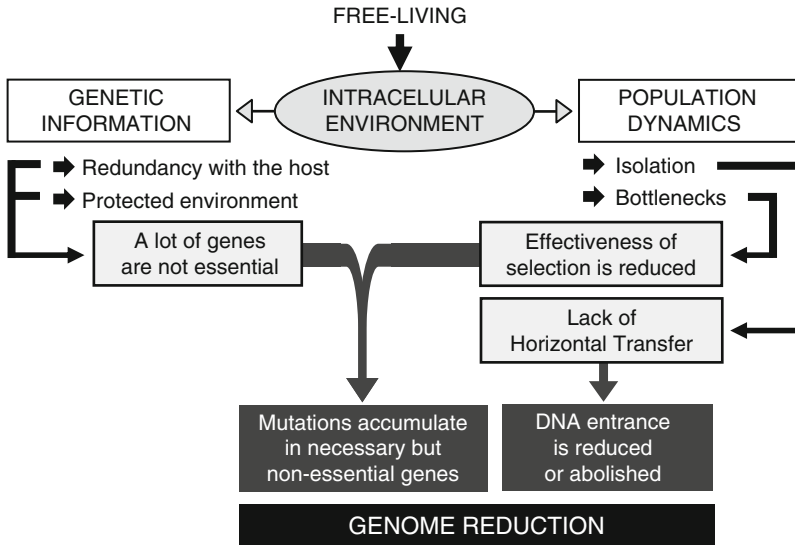
FREE-LIVING

```
GENETIC              INTRACELULAR          POPULATION
INFORMATION    ◁     ENVIRONMENT    ▷      DYNAMICS
```

➡ Redundancy with the host   ➡ Isolation
➡ Protected environment   ➡ Bottlenecks

| A lot of genes are not essential | | Effectiveness of selection is reduced |

| Lack of Horizontal Transfer |

| Mutations accumulate in necessary but non-essential genes | | DNA entrance is reduced or abolished |

GENOME REDUCTION

Fig. 1 Genetic and population factors involved in the genome reduction syndrome experienced by mutualistic endosymbionts. At the beginning of endosymbiosis, the new rich, protected and stable intracellular niche provided by the host makes superfluous some gene functions, that become redundant (since they can be contributed by the host) or unnecessary in a stable and protected environment, but forces the preservation of genes required for the maintenance and viability of the partnership. The decreased efficiency of the purifying selection causes a fast accumulation of slightly deleterious mutations on non-essential genes, increasing the rates of genomic evolution. In addition, the drastic reduction of the bacterial effective population size between successive insect generations increases the relative influence of random genetic drift. Furthermore, the obligatory intracellular life-style prevents the entrance of genetic material by horizontal gene transfer, making the losses irreversible

2 Survival, Replication and Transmission, the Three Biological Processes Involved In the Establishment of a Permanent Symbiotic Association

Mutualism and parasitism are two sides of the same coin. At the very beginning, it is not possible to determine if the relationship that would be established will be parasitic or mutualistic, since this distinction is based on the effect of the bacterium in the eukaryotic host but, from the bacterial point of view, the biological processes needed to successfully infect hosts are largely the same for both types of micro-organisms (Gil et al. 2004a). In both cases, it will be necessary to overcome the physical, cellular, and molecular barriers presented by the host, to achieve internalization, survival, and proficient replication of the prokaryote inside the eukaryotic host cell. No matter if the interaction is harmful, neutral, or beneficial to the host,

natural selection will favor the bacteria that achieve this goal (Ochman and Moran 2001). Most evolutionary transitions leading towards symbiotic lifestyles involve gene loss and horizontal gene transfer (HGT) of virulence genes within bacterial lineages. Genomic analyses indicate that, in many cases, the same molecular factors are involved both in pathogenic and mutualistic relationships although, in the case of mutualism, traditionally considered parasitic traits, at some point, became beneficial for both partners. In facultative symbionts, toxins that are known or suspected to target eukaryotic cells are involved in protecting the host against natural enemies (Oliver et al. 2009). Such toxins are encoded by genes present in lysogenic bacteriophages that participate in mutualistic functions but also act as hot spots for non-homologous recombination events that allow gene exchange of virulence cassettes among heritable symbionts (Degnan and Moran 2008). But even endosymbiotic bacteria with a long-time established relationship with their hosts, which have suffered a dramatic genome size reduction (as will be discussed below), maintain genes that encode essential endosymbiotic factors that are proposed to be virulence associated in bacterial pathogens, such as type III secretion systems and urease (Gil et al. 2003; Goebel and Gross 2001; Shigenobu et al. 2000). In many free-living bacteria, genes encoding the type III secretion system are located within pathogenicity islands that have been acquired by HGT. This system is present in many insect endosymbiotic bacteria where it has been proposed to be essential to invade the host cells, thus playing an essential role in the establishment of the symbiosis (Dale et al. 2001, 2002).

The establishment of a permanent intracellular association necessarily implies the development of efficient mechanisms for bacterial survival and replication inside the host cell. The bacteria must adapt their replication, so that their growth rates are coordinated with the development of their hosts in a way that depends on their location inside the host cell. In *Buchnera aphidicola*, which lives confined in vacuole-like organelles inside the aphid bacteriocytes, there is a tight coupling of bacterial cell number and aphid growth, with the bacteria showing a doubling time of approximately 2 days, much longer than the maximum exhibited by many free-living bacteria (Baumann and Baumann 1994). *Blochmannia floridanus* and *Wigglesworthia glossinidia*, which live free in the cytosol of bacteriocytes of their hosts (carpenter ants and tsetse flies, repectively), lack *dnaA*, the gene that encodes the essential DNA replication initiation protein in bacteria. Other alternative mechanisms reported so far for DNA replication initiation are also absent in *B. floridanus*. It has been suggested that this could imply the existence of a more direct control of DNA replication of the symbionts by the host (Gil et al. 2003).

An efficient transmission of the bacteria to the offspring must also be guaranteed. The acquisition of mechanisms ensuring maternal transmission to the host progeny allows the association to be heritable, resulting in the emergence of a new composite organism host-endosymbiont. The fine-tuning of this process detected in long-established obligate mutualistic symbioses suggests a long history of selection favoring host adaptations that help to maintain the association (Moran and Telang 1998).

3 Early Stages In the Symbiotic Relationship

The genomic era has allowed the sequencing of whole genomes of many bacteria living in symbiosis with eukaryotic hosts, allowing the comparison among the different evolutionary innovations carried out by these bacteria on their way from free-living to varied stages of integration with their respective hosts. To detangle the changes involved in each stage, over the next paragraphs we will follow the path from facultative symbiosis to early obligate endosymbiosis, as it has been revealed by molecular studies and comparative genomics over the past years.

3.1 Facultative Symbionts

Many different types of facultative or S-symbionts have been described in arthropods, and have been extensively studied in several lineages of aphid, psyllids, whiteflies, leafhoppers, tsetse flies, fruit flies and mosquitoes (Table 1). They can be maternally transmitted between host generations but, unlike P-endosymbionts, they can also be horizontally transferred among host individuals and species and, therefore they do not share long evolutionary histories with their hosts. S-symbionts do not reside exclusively in specialized cells and organs, and can also be found in gut tissues, glands or body fluids, and when a P-endosymbiont is also present, they can occupy cells surrounding the P-bacteriocytes, or even invade them. Phylogenetic studies indicate that facultative symbionts have established relatively recent associations with their hosts (Dale and Moran 2006). Thus, their genomes may resemble those in the early stages of a transition from a free-living lifestyle to an obligate mutualism.

Their uneven presence among species and individuals of the same species indicates that S-symbionts are not necessary for host survival, but their influence on host fitness is variable. A range of effects, from negative to beneficial, have been described. Some described S-symbionts have negative effects on growth and reproduction to the host or may establish neutral or parasitic associations. Heritable S-symbionts can spread among lineages by manipulating host reproduction to enhance matrilineal transmission through parthenogenesis, male killing and feminization of genetic males or cytoplasmic incompatibility. This is the case of *Wolbachia* infecting arthropods, where it undergoes transfer among host lineages (McGraw and O'Neill 2004). Remarkably, *Wolbachia* appears as a typical P-endosymbiont in filarial nematodes, where it is required for normal development. The complete genomes of four different *Wolbachia* strains are already available, allowing unraveling the molecular basis of their interaction with their respective hosts by comparative genomics. Three of them are reproductive parasites of arthropods, *Wolbachia pipientis* *w*Mel strain, found in *Drosophila melanogaster* (Wu et al. 2004); *w*Ri strain, from *Drosophila simulans* (Klasson et al. 2009), and *w*Pip strain, from the mosquitoes of the *Culex pipiens* group (Klasson et al. 2008); the last one, belongs to the *w*Bm strain, the obligate mutualist of the nematode

Brugia malayi (Foster et al. 2005). When the genomes of the parasitic strains were compared, a high degree of rearrangements was observed, revealing the most highly recombining obligate intracellular bacterial community examined to date (Klasson et al. 2009). The presence of abundant copies of transposable elements and prophages, that provide numerous sites for homologous recombination, can explain that. Most of the genome size differences are due to the presence of repeated elements, especially to the amplification of the WO prophage. Furthermore, the WO elements can experience intragenic recombination (Bordenstein and Wernegreen 2004). They present a conserved core of structural genes plus a variable fraction of genes that encode for ankyrin repeats, which correlate with the effects of the bacterial strain as reproductive parasite.

The first completely sequenced genome of a S-symbiont with no clear negative or positive effect corresponded to *Sodalis glossinidius* (Toh et al. 2006), the S-symbiot of the tsetse fly. It has been proposed to play a role in the acquisition of trypanosome infections (Welburn and Maudlin 1999). Its genome size (4.2 Mb) is close to that of free-living bacteria, but its coding capacity is highly diminished by the presence of a big amount of pseudogenes, only similar to what has been observed in some parasites such as *Mycobacterium leprae* (Cole et al. 2001; Gomez-Valero et al. 2007). The genome also contains certain amounts of repetitive and mobile DNA, such as transposable elements and bacteriophages, which could promote recombination. Therefore, it appears that this bacterium is at the early stages in the reductive process affecting symbiont genomes. *S. glossinidius* coexists in the gut lumen of tsetse flies with the P-endosymbiont, *W. glossinidia*, but occupying different portions of the insect gut, and it can be found both intra- and extracellularly (Toh et al. 2006). Moreover, it can be cultured in vitro (Dale and Maudlin 1999), an indication that the association with its host is not yet irreversible.

Many other S-symbionts described in aphids confer beneficial effects on the survival and reproduction rates of their hosts. They can rescue the host from heat damage (Chen et al. 2000; Montllor et al. 2002), provide resistance to natural enemies (Ferrari et al. 2004; Guay et al. 2009; Oliver et al. 2005; Scarborough et al. 2005) or stress (Russell and Moran 2006), are involved in host plant specialization and reproduction (Simon et al. 2003; Ferrari et al. 2004; Tsuchida et al. 2004), and even compensate the loss of the essential endosymbiont, as it was experimentally proven (Koga et al. 2003). Recently, the genome of one strain of *Candidatus* Hamiltonella defensa (from now on *H. defensa*), S-symbiont of the pea aphid *Acyrtosiphon pisum*, also became available (Degnan et al. 2009b). *H. defensa* can be found in aphids and other sap-feeding insects, where it has been proposed to play a beneficial role by protecting its host from attack by parasitoid wasps. Genes that encode for toxins, effector proteins, and two type-III secretion systems have been identified in the sequenced genome and seem to be involved in this function. The 2.1-Mb sequenced genome has undergone significant reduction in size relative to its closest free-living relatives, and important gene losses have been detected (it relies on the the P-endosymbiont *B*. *aphidicola* for the synthesis of 8 of the 10 essential amino acids), which indicates that the reductive process affecting endosymbiont genomes is already advanced. Nevertheless, the genome contains

considerable amounts of genes devoted to regulatory functions involved in regulation of virulence factors and quorum-sensing genes, which indicates that it still retains at least a partial ability to deal with changing environments and invasion of new host species. This genome also contains important amounts of repetitive DNA (21% of the genome), including insertion sequences, group II introns, prophages and plasmids. APSE, a lysogenic phage that infects many *H. defensa* populations, has been involved in the protective role of this bacterium against parasitic wasps, since the different variants of APSE identified all encode toxins that target eukaryotic tissues (Oliver et al. 2009). Therefore, the beneficial role of the phage toxins for the insect host fitness is contributing to the spread and maintenance of *H. defensa* in host populations. This is another evidence of the direct implication of virulence factors on the basis of a mutualistic symbiosis. Furthermore, the APSE lysis region is a hot spot for non-homologous recombination of novel virulence cassettes, allowing gene exchange among S-symbionts by horizontal transmission (Degnan and Moran 2008).

Candidatus Regiella insecticola (from now, *R. insecticola*) is another common facultative symbiont in aphids. Similar to *H. defensa*, it is not only involved in resistance to parasitoid wasps but also to fungal pathogens (Scarborough et al. 2005). Most of its genome (about 2.07 Mb) has been sequenced and compared with the close relative *H. defensa* (Degnan et al. 2009a). The complete genome assembly was not performed, because it was hampered by the presence of high amounts of repetitive DNA, mostly insertions sequence (IS) elements, representing up to 14% of the genes and pseudogenes. Similar to what has been found in the parasitic *W. pipientis* strains, the genomic architecture of these two genomes is highly divergent, as a consequence of recombination and gene inactivation facilitated by the presence of mobile DNA. In contrast, core genes reveal clonal evolution in *H. defensa* and *R. insecticola*, and the nucleotide divergence in this case is similar to what has been found in obligate mutualists. No intact prophages have been found in the already sequenced part of the genome of *R. insecticola*.

The genomes of two *Serratia symbiotica* strains, from the cedar and tuja aphids, are also being sequenced. Although some strains of *S. symbiotica* appear as typical facultative symbionts, this is not the case of the SCc strain, which has become essential for its host, the cedar aphid *Cinara cedri* (Gosalbes et al. 2008). Preliminary results of its genome project indicate that it has established a permanent and stable cooperative consortium with the host and the P-endosymbiont, *B. aphidicola* BCc, thus becoming essential for the maintenance of the fitness of all three partners (see Sect. 6).

3.2 Insertion Sequences, Shaping the First Steps Towards an Obligate Endosymbiosis

It has been postulated that soon after the establishment of obligate symbiosis, a massive gene loss must occur, probably by means of large deletion events that cause

the elimination of series of contiguous genes (Moran and Mira 2001). Later on, as it has been shown by comparative genomics, genome shrinkage proceeds through a process of gradual pseudogenization and gene loss scattered throughout the genome (Gomez-Valero et al. 2004; Silva et al. 2001). However, the mechanism involved in the large deletion events was unknown at that time. The identification of the genome changes that occur in these initial stages of the adaptation towards endo-symbiosis requires the genome analysis of clades of bacteria that have recently established such associations. For this purpose, our group selected SOPE, the P-endosymbiont of the rice weevil *Sitophilus oryzae*. With an estimated 3.0-Mb genome (Charles et al. 1997), within the range of many free-living bacteria, this γ-proteobacterium maintains a typical obligate mutualistic endosymbiosis with its host. The bacteria live inside bacteriocytes organized in an organ called bacteriome surrounding the midgut of the insect and near the female ovaries. The bacterium cannot be cultured outside the host, and it provides at least amino acids and vitamins to the insect, which has recognizable effects on fertility, development and the flying ability of adult insects (Heddi et al. 1999). SOPE is closely related with *S. glossinidius* (Dale and Welburn 2001; Heddi et al. 1998), which is still able to grow in laboratory culture conditions. Although SOPE and *S. glossinidius* are respectively P- and S-symbionts of hosts belonging to different insect orders (Coleoptera and Diptera), which feed on very different diets (storage grain and blood, respectively), their close phylogenetical position indicates a relative recent divergence. Therefore, the analysis of the similarities and differences between these two genomes will help to achieve a better understanding of the differences between the primary and secondary forms of endosymbiosis and what molecular events are implied in the establishment of an obligatory endosymbiosis.
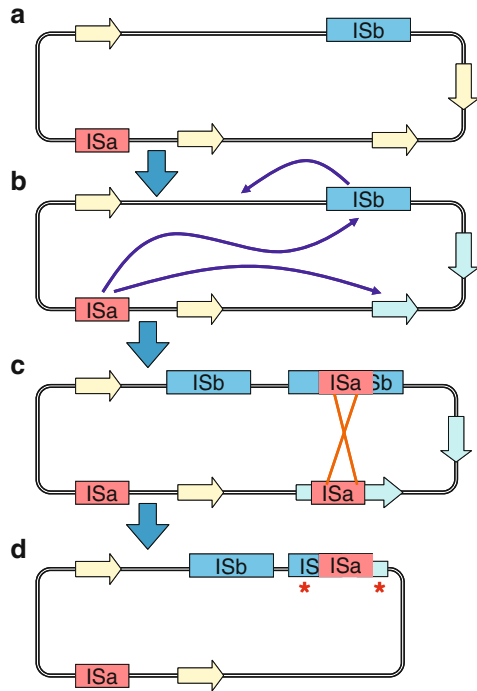
The association of insects of the genus *Sitophilus* and their present endosym-bionts is not antique. Some data indicate a recent endosymbiont replacement of an ancestral endosymbiont in the family Dryophtoridae to which the rice and maize weevils belong (Lefevre et al. 2004). During the first stages of the SOPE genome sequencing project (in progress), big amounts of repetitive DNA, mainly IS elements, were identified (Gil et al. 2008). It has been estimated that IS elements occupy about one third of its genome, and a similar situation has been observed in its close relative SZPE, the P-endosymbiont of the maize weevil (Plague et al. 2008). This impressive amount of repetitive DNA was not expected in an obligate mutualistic endosymbiont. Repetitive DNA is common in free-living bacteria, and its presence increases in bacteria that have recently evolved as specialized patho-gens (e.g., the enteric bacteria *Shigella* and *Salmonella enterica* Typhi) (Jin et al. 2002; Wei et al. 2003), intracellular parasites (e.g. *W. pipientis* strains, reproductive parasite of arthropods) (Klasson et al. 2008, 2009; Wu et al. 2004), or facultative insect symbionts (e.g. *H. defensa*, *Candidatus* Arsenophonus arthropodicus and *R. insecticola*) (Dale and Moran 2006; Degnan et al. 2009a,b). Thus, the increase in transposable elements is a common trait among bacteria that have recently estab-lished mutualistic relationships with their hosts, and must have subsequent effects on the outcome of the symbiotic process (Bordenstein and Reznikoff 2005; Moran and Plague 2004). However, it was assumed that after the establishment of an

obligate endosymbiont lifestyle, repetitive DNA tends to diminish until its total disappearance. Several observations support this conclusion, from total absence of phages or transposable elements in bacterial endosymbionts with a long-established obligatory relationship with their hosts, to the identification of only 5.4% of repetitive DNA, mostly composed of inactivated IS, in the mutualistic *W. pipientis* *w*Bm (Foster et al. 2005).

The IS are the most abundant and simplest transposable elements in nature (Touchon and Rocha 2007). Habitually they only include the elements needed in its own mobilization: short terminal inverted repetitive sequences (IR) define the ends of the IS and flank the ORF(s) that encodes the transposase activity that mediates the transposition events after the recognition and processing of the IR sequences. The IS are able to move between replicons of a certain genome and can also be transferred between genomes of different organisms by horizontal gene transfer. Its persistence is usually explained by an intense ability for intergenomic mobilization and to its more or less efficient infecting capacity. Four IS types have been identified in SOPE (Gil et al. 2008). At least two of them (IS*sope*1 and IS*sope*2) are present in large copy numbers in SZPE (Plague et al. 2008), and IS*sope*1 has also been identified in *S. glossinidius*, but representing just 2.5% of the total genome (Toh et al. 2006), is an indication that this element must have been present in a common ancestor of these bacteria.

The massive presence of IS must be related with some of the syndromes that appear at the beginning of the intracellular life (Fig. 2). IS elements are widespread in free-living bacteria, but their transposition is tightly controlled, so that only a few copies of a limited number of categories appear in each genome. The dramatic increase of these elements in intracellular bacteria must reflect an enhanced replicative transposition of elements that were already present at the onset of symbiosis, and can then act as a source of gene inactivation and chromosomal rearrangements. After the establishment of the symbiosis, the decrease in the selective pressure caused by functional conditions and population dynamics in the new environment, can favor the uncontrolled proliferation of such elements, which could be involved in the inactivation of non-essential genes. The high abundance of very similar (or even identical) repetitive elements in direct orientation can then serve as a substrate for unequal recombination, which would lead to a loss of the region between two elements, thus promoting genome size reduction in early stages. Additionally, the presence of these elements in opposite orientation, will lead to genome rearrangements. Comparative genomics analyses between several *B. aphidicola* strains form different aphids, *B. floridanus*, and close free-living relatives indicate that the massive gene loss that took place in the process towards the last common symbiotic ancestor (LCSA) of both species was accompanied by many chromosomal rearrangements. The former presence of repetitive elements, already disappeared in the present genomes, might explain such genome reorganizations, while the current lack of repetitive sequences, with a great potential as recombination sites, as well as the loss of *loci* needed for the catalysis of such recombination events in later stages of the symbiosis (see next section), appears to be in the origin of the high genomic-architecture stability levels in old endosymbionts, quite unusual among the

prokaryotes (Silva et al. 2003). This is an indication that most of genomic modeling, including chromosomal rearrangements and the loss of many functionally dispensable genes, must take place at an early stage of the process of genomic adaptation to intracellular life (Dougherty and Plague 2008; Touchon and Rocha 2007). Genes needed for DNA repair and recombination are also among the first losses detected, thus contributing to genomic stasis in further steps in the endosymbiotic evolutionary path. The loss of the genes coding for the enzymes RecA and RecF in SOPE, SZPE, and *S. glossinidius* (Dale et al. 2003) supports this idea.

4 Long-Established P-Endosymbioses

Most bacterial insect P-endosymbionts that have been analyzed belong to the γ-proteobacteria (Table 1). However, more recently, the genomes of several endosymbionts belonging to other groups of proteobacteria and to the phylum Bacteroidetes have also been analyzed (McCutcheon and Moran 2007; Lopez-Sanchez et al. 2008; Tokuda et al. 2008; Sabree et al. 2009), revealing convergent evolution among endosymbionts belonging to different phyla (López-Sánchez et al. 2009). In general, endosymbionts with a long-established relationship with their hosts have

genomes eight to ten times smaller than those of their free-living relatives. In bacteria, whose genomes are highly compact, gene content correlates quite well with genome size (Casjens 1998). Therefore, the reduced size of endosymbiont genomes reflects the presence of a smaller number of genes than those of free-living bacteria. Several additional characteristic genome features have traditionally been associated with the degenerative syndrome affecting endosymbiotic bacteria. These include almost total absence of recombination, increased rate of nucleotide substitution, high A + T content (although as it will be discussed later, this no longer can be considered a general trait), accumulation of deleterious mutations by random genetic drift, loss of codon bias towards A or T, and accelerated sequence evolution (Andersson and Kurland 1998; Clark et al. 1999; Moya et al. 2002; Wernegreen 2005). Most of these characteristics are linked with the above mentioned informational and demographic factors affecting bacteria that live in close association with eukaryotic cells, although the accommodation to symbiotic life varies according to the age of the association, the host lifestyle, and the way of living within the host.

The analysis of gene order in the first completely sequenced endosymbiont genomes lead to interesting observations regarding the evolution of these genomes. The availability of complete genome sequences from four different strains of *B. aphidicola* clonally evolving in their aphid hosts revealed that, after a short period of large genome rearrangements at the beginning of the symbiotic process, there were large periods of evolutionary stasis. All these strains present a nearly perfect gene-order conservation (Perez-Brocal et al. 2006; Shigenobu et al. 2000; Tamas et al. 2002; van Ham et al. 2003), which suggests that *B. aphidicola* can be considered as a "gene-order fossil", and that the onset of genomic stasis coincided with the establishment of the obligate symbiosis with aphids, 80–150 MY ago (von Dohlen and Moran 2000). As mentioned in the previous section, this astonishing genome stasis can be explained by the total absence of repetitive DNA in these genomes, as well as the loss of genes involved in DNA repair and recombination in early stages of the symbiotic integration. Repetitive DNA is quite abundant at the beginning of the obligate endosymbiosis, but these elements tend to disappear in the later stages of the relationship and are absolutely absent in endosymbionts that share long evolutionary histories with their hosts (Fig. 2). The progressive loss of transposable elements might have been favored by the energetic benefit of decreasing transposase activity and avoiding the increase in genome size derived from the proliferation of these elements or by the need to control the mutagenizing effect of its mobilization. It is presumable that, at some point, IS elements expansion will be deleterious and these elements would be also affected by the process of genome degradation that these genomes suffer. The sexual isolation of P-endosymbionts and the loss of recombination genes must also have participated in the process, since horizontal gene transfer is the way of entrance of these elements in prokaryotic genomes (Touchon and Rocha 2007). The reduced genomes of endosymbiotic bacteria and some pathogens have lost most (if not all) genes involved in recombination processes and, consequently, the genome size cannot be increased by acquisition of foreign DNA (Silva et al. 2003). Nevertheless, some recombination events can still take place in these reduced genomes, probably involving the RecBCD

system, which in the absence of RecA might serve as a general exonuclease repair enzyme (Sabater-Munoz et al. 2004), as revealed by the great plasticity of the plasmids involved in the biosynthesis of leucine in different lineages of *B. aphidicola*, showing that several events of insertion from a plasmid to the main chromosome have occurred since the divergence of these strains (Latorre et al. 2005).

In general, smaller genomes correlate with longer obligate associations. The differences in host lifestyle are also introducing changes in this degenerative process among strains of the same endosymbiont species. The small genomes of *B. aphidicola* are still suffering this reductive process, as evidenced by the fact that *B. aphidicola* strains from several aphid subfamilies showed differences up to 200 Kb (Gil et al. 2002), and the presence of pseudogenes in the *B. aphidicola* genomes that have been sequenced (Perez-Brocal et al. 2006; Shigenobu et al. 2000; Tamas et al. 2002; van Ham et al. 2003). In addition, the degenerative process is randomly affecting different genes in each genome, conditioning the essentiality of the rest of the genes that are present in these reduced genomes. Therefore, although we can hypothesize that the LCSA of *B. aphidicola* suffered a drastic genome reduction at the beginning of the symbiotic integration, since then, the different strains of the bacteria have undergone a reductive process in a way that correlates with their hosts.

In addition to changes in genome size, obligate and facultative endosymbionts of different insect hosts also differ in nucleotide composition. P-endosymbionts with an old association with their hosts have in general small genomes, and an A + T content higher than 70%, while P-endosymbionts with a younger association and S-symbionts have a genome size and an A + T percentage intermediate with respect to older P-endosymbionts and free-living relatives (Dale and Maudlin 1999; Heddi et al. 1998; McCutcheon and Moran 2007; Moya et al. 2002; Nakabachi et al. 2006). The loss of the bias in codon usage in these obligate intracellular bacteria, highly mitigated in P-endosymbionts with larger genomes and in S-symbionts and almost absent in *B. aphidicola*, is considered to be a consequence of this base composition bias (Moya et al. 2002; Rispe et al. 2004). This notable enrichment in A + T has been related to the loss of DNA repair enzymes, since the most common chemical changes in DNA (cytosine deamination and guanosine oxidation) led to changes in GC pairs leading to AT. However, several cases that do not follow this nucleotide composition rule have been described. The partial genome sequences available from *Candidatus* Tremblaya princeps, the P-endosymbiont of the mealybug *Planococcus citri*, indicated that this genome has a 57% G + C content, much higher than expected for an endosymbiont (Baumann et al. 2002). Recently, a remarkable small genome with a high G + C content has also been reported (McCutcheon et al. 2009) (see next section). *Candidatus* Hogkinia cicadicola (from now on *H. cicadicola*), P-endosymbiont of the cicada *Dieroprocta semicincta*, presents a 144-Kb genome with a 58.4% G + C content. Therefore, it has been proposed that, while gene loss associated with genome reduction is a critical step in endosymbiont genome evolution, mutational pressure favoring A + T is not.

There is only one case of advanced symbiosis described in archaea: *Nanoarchaeum equitans*, a tiny coccus living attached to the outside of the cells of its host, the

Crenarcheote *Ignicoccus hospitalis*. The study of this association, including the sequencing of the genome of both species (Waters et al. 2003; Podar et al. 2008), shows a highly specialized relationship, which so far cannot be assigned to any classical symbiosis type (mutualism, commensalism or parasitism). With a highly reduced genome (491 Kb), *N. equitans* was initially suggested to be a representative of a novel phylum within the domain archaea, Nanoarchaeota. However, further genomic analyses indicate that it is likely to be a highly derived Euryarchaeon, possibly related to the Thermococcales that has evolved through a unique pathway of genome degradation (Brochier et al. 2005; Makarova and Koonin 2005). Features such as the extreme *N. equitans* genome reduction, bias in codon usage, and evolutionary acceleration, are common to those observed in bacterial endosymbionts, probably an indication of the generality of the reductive mechanisms among prokaryotes. Interestingly, this reductive process has affected simultaneously both genomes, since the *I. hospitalis* genome is only 1.3 Mb in length, one of the smallest among free-living organisms. Further analyses will be necessary to understand the implication of this dual reductive genome process (Forterre et al. 2009).

5   Final Stages in Endosymbiotic Relationships

As the endosymbiotic integration progresses, genes that are rendered unnecessary experience a random process of gradual pseudogenization and gene loss scattered throughout the genome (Gomez-Valero et al. 2004; Silva et al. 2001). The final step of this minimization process might, in theory, lead to the loss of all genes except those that are essential for keeping the host-bacterial interaction reproducing. Therefore, even the most reduced genome must retain those genes involved in the symbiotic relationship, as well as a reduced repertoire of genes necessary to maintain the three essential functions that define a living cell: maintenance, reproduction and evolution (Luisi et al. 2002). One of the most comprehensive efforts to define the minimal core of essential genes was that presented by Gil et al. (2004b). This study can be a good starting point to identify essential genes involved in informational processes that must be present in any living cell, while the essential genes devoted to the symbiotic association can be deduced by the knowledge of the host needs for survival and reproduction. However, most extremely reduced genomes that have been described have lost part of such essential functions. In most cases, as it will be discussed in the next section, genome degradation can proceed over the expected limit because of the implication of a second endosymbiont on the relationship. But there is an intriguing case: *Candidatus* Carsonella ruddii (from now on, *C. ruddii*), considered the P-endosymbiont of the psyllid *Pachpsylla venusta*. Although a second bacterial symbiont has not been found in the psyllid, *C. ruddii* does not fulfil the conditions to be considered as a mutualistic endosymbiont, not even as a living organism. Its genome consists of a circular chromosome of 160 Kb, averaging 83.5% A + T content (Nakabachi et al. 2006). It also presents a high coding density (97%), and 182 described open reading frames, many of

which overlap and present a reduced gene length. A detailed analysis of the coding capacities of *C. ruddii*, revealed that the extensive degradation of the genome is affecting vital and symbiotic functions (Tamames et al. 2007). Most genes for DNA replication, transcription and translation are completely absent, and gene shortening causes, in some cases, the loss of essential domains and functional residues needed to fulfil these and other vital functions. In addition to the essential functions that define life, as a mutualistic endosymbiont, *C. ruddii* should provide its host all essential complements to its nutritionally deficient diet, limited to phloem sap, rich in sugars but relatively poor in nitrogenated compounds, especially essential amino acids. However, the genomic analysis revealed that the pathways for the synthesis of three essential amino acids (i.e. histidine, phenylalanine and tryptophan) are lost. Since this strain of *C. ruddii* is not able to sustain the requirements of its host, neither to sustain its own vital functions, it can be viewed as a further step towards the degeneration of the former P-endosymbiont, and its transformation in a subcellular new entity between living cells and organelles, which might be taking advantage of mitochondrial functions encoded by the nucleus, especially for basic informational processes needed for maintenance and multiplication. It might even be possible that some *C. ruddii* genes have been transferred to the host nuclear DNA, as it has been proved for present organelles. If confirmed, this would be the first example of such a scenario in animal cells.

6 Replacement or Complementation, and the Establishment of Microbial Consortia

Eventually, after the establishment of a permanent symbiotic association between a bacterium and an animal host, a second bacterial species can join the association. Although initially this new association can be facultative (as seen in Sect. 3.1), if the second bacterium provides benefits to the organization, with time, it can become essential for host fitness. The involvement of two bacteria in the fitness of an insect host adds one extra element to the evolutionary scenario that explains the reductive evolution of endosymbiont genomes, but there is no need to invoke any supplementary reductive factor in addition to the informational and population dynamis factors already indicated. Subsequently, all three components of the association will co-evolve, and the evolutionary process of genome shrinkage will now affect both bacteria. New genes will become unnecessary due to redundancy, but which one of the two bacterial genomes loses them will be a matter of chance. Depending on which genome is affected by the loss of genes needed for the synthesis of essential molecules, either both bacteria will become indispensable to keep a healthy consortium (complementation) or one of them can enter an extreme degenerative process, which may end with its extinction (replacement), and the retained bacteria will continue the degenerative process alone (Moya et al. 2009). Replacement has already been reported, for example, in the Family Dryophthoridae, where a former

endosymbiont *Candidatus* Nardonella was replaced by the ancestor of the *Sitophilus* P-endosymbionts (Lefevre et al. 2004). However, there are many more described cases in which both bacteria loose part of the gene complement necessary for their host fitness, so that both of them become indispensable and a stable consortium is established. Several symbiotic consortia have already been reported and sequenced, using metagenomics approaches, and many more will surely be available in the near future thanks to the use of new massive sequencing technologies.

One of the first described consortia involves strains of *B. aphidicola* and *S. symbiotica* living inside the cedar aphid. *S. symbiotica* appears as a facultative symbiont in many aphid species. However, it was always found in cedar aphids, coexisting with *B. aphidicola* BCc in the insect bacteriome, so that *S. symbiotica* strain SCc cannot be considered as a facultative symbiont. Comparative, functional and evolutionary genomic analysis, plus microscopic observations, led Perez-Brocal et al. (2006) to conclude that *S. symbiotica* SCc might be replacing *B. aphidicola* BCc. Contrary to other sequenced *B. aphidicola* strains, BCc has partially lost its symbiotic role, as it cannot synthesize tryptophan. Genes involved in the biosynthesis of this essential molecule were found in the genome of *S. symbiotica* SCc (Gosalbes et al. 2008), but included an additional surprise: the pathway to synthesize tryptophan is distributed between both genomes: *B. aphidicola* BCc produces a metabolic intermediate that is then provided to *S. symbiotica* SCc to synthesize the final product. Therefore, coexistence of both bacteria is needed to keep a healthy consortium due to metabolic complementation, and both of them keep an intracellular obligatory mutualistic association with their host.

The establishment of an endosymbiotic bacterium consortium can be in the origin of big evolutionary changes in host lifestyle. This is the case of the consortium formed by *Baumannia cicadellinicola* and *Sulcia muelleri*, co-resident P-endosymbionts of the xylem-feeding sharpshooter *Homalodisca vitripennis*. Their whole genome analysis revealed that they have complementary sets of biosynthetic capabilities needed to provide to their host the nutrients that are lacking in the xylem sap (Wu et al. 2006). While *B. cicadellinicola* contains a large number of pathways for biosynthesis of vitamins, *S. muelleri* encodes the enzymes involved in the biosynthesis of most essential amino acids. Phylogenetic studies indicate that *S. muelleri* was ancestrally present in a host lineage that acquired *B. cicadellinicola* at the same approximate time that the host ancestor switched to a xylem-feeding lifestyle, consistent with the view that *Baumannia's* nutrient-provisioning capabilities were a requirement for the acquisition of this new feeding behavior.

A newly described consortium also involving *S. muelleri* is on the basis of the dramatic genome reduction experienced by *H. cicadicola*, the P-endosymbiont of the cicada *D. semicincta* (McCutcheon et al. 2009). *H. cicadicola* is an α-proteobacteria with the smallest described genome to date (144 Kb), an unusually high $G + C$ content (58.4%), and a coding reassignement of UGA stop codon to Trp. It has been found in other cidadas, thus suggesting that this symbiont infected an ancestor of the cidadas and, since then, has been maternally transmitted.

Our group is also involved in the metagenomic study of another exceptional symbiotic consortium: the one established among the mealybug *P. citri* and their

two endosymbiotic bacteria: the P-endosymbiont *T. princeps*, a β-proteobacterium (Thao et al. 2002), which contains inside a γ-proteobacterium (von Dohlen et al. 2001), considered as an S-symbiont based on its polyphyletic origin (Thao et al. 2002). This is the first described case of a double-endosymbiosis, although the symbiotic relationship between the two bacteria (parasitic, commensal or mutualistic) has not been elucidated (Kono et al. 2008). As it has been mentioned in Sect. 4, *T. princeps* was the first endosimbiotic genome in which it was detected that the A + T bias rule does not apply (Baumann et al. 2002).

Some other consortia can involve more than two microorganisms. The marine oligochaete *Olavius algarvensis*, which lacks a digestive and excretory system, harbors four co-occurring symbionts essential for host survival (Woyke et al. 2006). The symbionts are located just below the worm cuticle, and they are essential to manage energy and waste needs of the host. The symbionts, γ1- and γ3-(sulphur-oxidizing chemolithoautotrophs), and δ1- and δ4-(sulphate reducer) proteobacteria, are engaged in an endosymbiotic sulphur cycle, fix CO_2, provide nutrients to the host, and are also involved in host waste recycling. They can heterotrophycally feed the host by taking up dissolved organic carbon compounds from the environment, and can synthesize almost all amino acids and several vitamins. The host probably takes these nutrients by digesting the bacteria (Fiala-Médioni et al. 1994). This is another case in which, contrary to most cases of obligate host-associated bacteria, the available genomic sequences do not show A + T bias.

7 Concluding Remarks

Symbiosis between prokaryotes and eukaryotes is an expanding field, thanks to the advent of the metagenomics and high-throughput sequencing technologies. Systems biology approaches are also allowing the exploration of metabolic interdependences among the members of the symbiotic consortium. Now that endosymbiont genomes are accumulating, comparative analyses allow making predictions on the evolutionary paths followed by endosymbiotic bacteria in their adaptation to the intracellular environment provided by the host. Now, more clearly than ever before, the association and functional interaction of genomes from different species observed during symbiosis can be viewed as a power, like mutation, recombination and other genome rearrangements, able to generate genetic variation, acting as a fuel for evolution. The action of forces such as natural selection and/or random drift will be the responsible of transforming this variation in evolutionary novelties. However, as the number of the available genomes increases, new features are appearing and open new questions that need to be experimentally solved. We do not know what drives symbiotic associations to mutualism versus parasitism, since both types of associations derive from common mechanisms for symbiont-host interaction. We cannot anticipate when a facultative association will become essential for host fitness and, when two or more prokaryotes are involved, we cannot determine which forces will lead it towards the establishment of a consortium or,

alternatively, will end up in a replacement. More recently, an additional question was opened about the nucleotide content bias, most of the times towards an increase of A + T, but also possible towards an increase in G + C content... For sure, we will learn a lot more about the molecular mechanisms and evolutionary forces acting on these systems once eukaryotic host genomes become available.

References

Andersson SG, Kurland CG (1998) Reductive evolution of resident genomes. Trends Microbiol 6:263–268

Baumann P (2005) Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol 59:155–189

Baumann L, Baumann P (1994) Growth kinetics of the endosymbiont *Buchnera aphidicola* in the aphid *Schizaphis graminum*. Appl Environ Microbiol 60:3440–3443

Baumann L, Thao ML, Hess JM, Johnson MW, Baumann P (2002) The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects. Appl Environ Microbiol 68:3198–3205

Bordenstein SR, Reznikoff WS (2005) Mobile DNA in obligate intracellular bacteria. Nat Rev Microbiol 3:688–699

Bordenstein SR, Wernegreen JJ (2004) Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. Mol Biol Evol 21:1981–1991

Brochier C, Gribaldo S, Zivanovic Y, Confalonieri F, Forterre P (2005) Nanoarchaea: representatives of a novel archaeal phylum or a fast-evolving euryarchaeal lineage related to Thermococcales? Genome Biol 6:R42

Buchner P (1965) Endosymbiosis of animals with plant microorganisms. Interscience Publishers, New York

Casjens S (1998) The diverse and dynamic structure of bacterial genomes. Annu Rev Genet 32:339–377

Charles H, Heddi A, Guillaud J, Nardon C, Nardon P (1997) A molecular aspect of symbiotic interactions between the weevil *Sitophilus oryzae* and its endosymbiotic bacteria: over-expression of a chaperonin. Biochem Biophys Res Commun 239:769–774

Chen D-Q, Montllor CB, Purcell AH (2000) Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid *A. kondoi*. Entomol Exp Appl 95:315–323

Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol Biol Evol 16:1586–1598

Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honore N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream MA, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K, Taylor K, Whitehead S, Woodward JR, Barrell BG (2001) Massive gene decay in the leprosy bacillus. Nature 409:1007–1011

Dale C, Maudlin I (1999) *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. Int J Syst Bacteriol 49 (Pt 1):267–275

Dale C, Moran NA (2006) Molecular interactions between bacterial symbionts and their hosts. Cell 126:453–465

Dale C, Welburn SC (2001) The endosymbionts of tsetse flies: manipulating host-parasite interactions. Int J Parasitol 31:628–631

Dale C, Young SA, Haydon DT, Welburn SC (2001) The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc Natl Acad Sci USA 98:1883–1888

Dale C, Plague GR, Wang B, Ochman H, Moran NA (2002) Type III secretion systems and the evolution of mutualistic endosymbiosis. Proc Natl Acad Sci USA 99:12397–12402

Dale C, Wang B, Moran N, Ochman H (2003) Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. Mol Biol Evol 20:1188–1194

Degnan PH, Moran NA (2008) Diverse phage-encoded toxins in a protective insect endosymbiont. Appl Environ Microbiol 74:6782–6791

Degnan PH, Leonardo TE, Cass BN, Hurwitz B, Stern D, Gibbs RA, Richards S, Moran NA (2009a) Dynamics of genome evolution in facultative symbionts of aphids. Environ Microbiol. October 16, Epub ahead of print

Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA (2009b) *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. Proc Natl Acad Sci USA 106:9063–9068

Dougherty KM, Plague GR (2008) Transposable element loads in a bacterial symbiont of weevils are extremely variable. Appl Environ Microbiol 74:7832–7834

Ferrari J, Darby AC, Daniell HCJG, Douglas AE (2004) Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. Ecol Entomol 29:60–65

Fiala-Médioni A, Michalski JC, Jollès J, Alonso C, Montreuil J (1994) Lysosomic and lysosome activities in gills of bivalves from deep hydrothermal vents. C R Acad Sci Paris 317:239–244

Forterre P, Gribaldo S, Brochier-Armanet C (2009) Happy together: genomic insights into the unique *Nanoarchaeum*/*Ignicoccus* association. J Biol 8:7

Foster J, Ganatra M, Kamal I, Ware J, Makarova K, Ivanova N, Bhattacharyya A, Kapatral V, Kumar S, Posfai J, Vincze T, Ingram J, Moran L, Lapidus A, Omelchenko M, Kyrpides N, Ghedin E, Wang S, Goltsman E, Joukov V, Ostrovskaya O, Tsukerman K, Mazur M, Comb D, Koonin E, Slatko B (2005) The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. PLoS Biol 3:e121

Gil R, Sabater-Munoz B, Latorre A, Silva FJ, Moya A (2002) Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. Proc Natl Acad Sci USA 99:4454–4458

Gil R, Silva FJ, Zientz E, Delmotte F, Gonzalez-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Hölldobler B, van Ham RCHJ, Gross R, Moya A (2003) The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. Proc Natl Acad Sci USA 100:9388–9393

Gil R, Latorre A, Moya A (2004a) Bacterial endosymbionts of insects: insights from comparative genomics. Environ Microbiol 6:1109–1122

Gil R, Silva FJ, Pereto J, Moya A (2004b) Determination of the core of a minimal bacterial gene set. Microbiol Mol Biol Rev 68:518–537

Gil R, Belda E, Gosalbes MJ, Delaye L, Vallier A, Vincent-Monegat C, Heddi A, Silva FJ, Moya A, Latorre A (2008) Massive presence of insertion sequences in the genome of SOPE, the primary endosymbiont of the rice weevil *Sitophilus oryzae*. Int Microbiol 11:41–48

Goebel W, Gross R (2001) Intracellular survival strategies of mutualistic and parasitic prokaryotes. Trends Microbiol 9:267–273

Gomez-Valero L, Latorre A, Silva FJ (2004) The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. Mol Biol Evol 21:2172–2181

Gomez-Valero L, Rocha EP, Latorre A, Silva FJ (2007) Reconstructing the ancestor of *Mycobacterium leprae*: the dynamics of gene loss and genome reduction. Genome Res 17:1178–1185

Gosalbes MJ, Lamelas A, Moya A, Latorre A (2008) The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. J Bacteriol 190:6026–6029

Guay J-F, Boudreault S, Michaud D, Cloutier C (2009) Impact of environmental stress on aphid clonal resistance to parasitoids: role of *Hamiltonella defensa* bacterial symbiosis in association with a new facultative symbiont of the pea aphid. J Insect Physiol 55:919–926

Hackstein JHP, van Alen TA, Rosenberg J (2006) Methane production by terrestrial arthropods. In: König H, Varma A (eds) Intestinal microorganisms of termites and other invertebrates. Springer, Heidelberg, pp 155–182

Heddi A, Charles H, Khatchadourian C, Bonnot G, Nardon P (1998) Molecular characterization of the principal symbiotic bacteria of the weevil *Sitophilus oryzae*: a peculiar G + C content of an endocytobiotic DNA. J Mol Evol 47:52–61

Heddi A, Grenier AM, Khatchadourian C, Charles H, Nardon P (1999) Four intracellular genomes direct weevil biology: nuclear, mitochondrial, principal endosymbiont, and *Wolbachia*. Proc Natl Acad Sci USA 96:6814–6819

Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, Wang J, Liu H, Yang J, Yang F, Zhang X, Zhang J, Yang G, Wu H, Qu D, Dong J, Sun L, Xue Y, Zhao A, Gao Y, Zhu J, Kan B, Ding K, Chen S, Cheng H, Yao Z, He B, Chen R, Ma D, Qiang B, Wen Y, Hou Y, Yu J (2002) Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. Nucleic Acids Res 30:4432–4441

Klasson L, Walker T, Sebaihia M, Sanders MJ, Quail MA, Lord A, Sanders S, Earl J, O'Neill SL, Thomson N, Sinkins SP, Parkhill J (2008) Genome evolution of *Wolbachia* strain wPip from the *Culex pipiens* group. Mol Biol Evol 25:1877–1887

Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SG (2009) The mosaic genome structure of the *Wolbachia* wRi strain infecting *Drosophila simulans*. Proc Natl Acad Sci USA 106:5725–5730

Koga R, Tsuchida T, Fukatsu T (2003) Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont *Buchnera* in an aphid. Proc Biol Sci 270:2543–2550

Kono M, Koga R, Shimada M, Fukatsu T (2008) Infection dynamics of coexisting beta- and gammaproteobacteria in the nested endosymbiotic system of mealybugs. Appl Environ Microbiol 74:4175–4184

Latorre A, Gil R, Silva FJ, Moya A (2005) Chromosomal stasis versus plasmid plasticity in aphid endosymbiont *Buchnera aphidicola*. Heredity 95:339–347

Lefevre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A (2004) Endosymbiont phylogenesis in the Dryophthoridae weevils: evidence for bacterial Replacement. Mol Biol Evol 21:965–973

Lopez-Sanchez MJ, Neef A, Patino-Navarrete R, Navarro L, Jimenez R, Latorre A, Moya A (2008) Blattabacteria, the endosymbionts of cockroaches, have small genome sizes and high genome copy numbers. Environ Microbiol 10:3417–3422

López-Sánchez MJ, Neef A, Peretó J, Patiño-Navarrete R, Pignatelli M, Latorre A, Moya A (2009) Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. PLoS Genet 5:e1000721

Luisi PL, Oberholzer T, Lazcano A (2002) The notion of a DNA minimal cell: a general discourse and some gidelines for an experimental approach. Helv Chim Acta 85:1759–1777

Makarova KS, Koonin EV (2005) Evolutionary and functional genomics of the Archaea. Curr Opin Microbiol 8:586–594

Margulis L (1993) Symbiosis in cell evolution. Microbial communities in the archaean and proterozoic eons. W. H. Freeman and Co., New York

McCutcheon JP, Moran NA (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc Natl Acad Sci USA 104:19392–19397

McCutcheon JP, McDonald BR, Moran NA (2009) Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. PLoS Genet 5:e1000565

McGraw EA, O'Neill SL (2004) *Wolbachia pipientis*: intracellular infection and pathogenesis in Drosophila. Curr Opin Microbiol 7:67–70

Montllor CB, Maxmen A, Purcell AH (2002) Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. Ecol Entomol 27:189–195

Moran NA, Mira A (2001) The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. Genome Biol 2:RESEARCH0054

Moran NA, Plague GR (2004) Genomic changes following host restriction in bacteria. Curr Opin Genet Dev 14:627–633

Moran NA, Telang A (1998) The evolution of bacteriocyte-associated endosymbionts in insects. Bioscience 48:295–304

Moya A, Latorre A, Sabater-Munoz B, Silva FJ (2002) Comparative molecular evolution of primary (*Buchnera*) and secondary symbionts of aphids based on two protein-coding genes. J Mol Evol 55:127–137

Moya A, Pereto J, Gil R, Latorre A (2008) Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9:218–229

Moya A, Gil R, Latorre A (2009) The evolutionary history of symbiotic associations among bacteria and their animal hosts: a model. Clin Microbiol Infect 15:11–13

Munson MA, Baumann P, Moran NA (1992) Phylogenetic relationships of the endosymbionts of mealybugs (Homoptera: Pseudococcidae) based on 16S rDNA sequences. Mol Phylogenet Evol 1:26–30

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

Ochman H, Moran NA (2001) Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292:1096–1099

Oliver KM, Moran NA, Hunter MS (2005) Variation in resistance to parasitism in aphids is due to symbionts not host genotype. Proc Natl Acad Sci USA 102:12795–12800

Oliver KM, Degnan PH, Hunter MS, Moran NA (2009) Bacteriophages encode factors required for protection in a symbiotic mutualism. Science 325:992–994

Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A (2006) A small microbial genome: the end of a long symbiotic relationship? Science 314:312–313

Plague GR, Dunbar HE, Tran PL, Moran NA (2008) Extensive proliferation of transposable elements in heritable bacterial symbionts. J Bacteriol 190:777–779

Podar M, Anderson I, Makarova KS, Elkins JG, Ivanova N, Wall MA, Lykidis A, Mavromatis K, Sun H, Hudson ME, Chen W, Deciu C, Hutchison D, Eads JR, Anderson A, Fernandes F, Szeto E, Lapidus A, Kyrpides NC, Saier MH Jr, Richardson PM, Rachel R, Huber H, Eisen JA, Koonin EV, Keller M, Stetter KO (2008) A genomic analysis of the archaeal system *Ignicoccus hospitalis–Nanoarchaeum equitans*. Genome Biol 9:R158

Raymond J, Siefert JL, Staples CR, Blankenship RE (2004) The natural history of nitrogen fixation. Mol Biol Evol 21:541–554

Rispe C, Delmotte F, van Ham RCHJ, Moya A (2004) Mutational and selective pressures on codon and amino acid usage in *Buchnera*, endosymbiotic bacteria of aphids. Genome Res 14:44–53

Russell JA, Moran NA (2006) Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. Proc Biol Sci 273:603–610

Russell JA, Latorre A, Sabater-Munoz B, Moya A, Moran NA (2003) Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea. Mol Ecol 12:1061–1075

Sabater-Munoz B, van Ham RC, Moya A, Silva FJ, Latorre A (2004) Evolution of the leucine gene cluster in *Buchnera aphidicola*: insights from chromosomal versions of the cluster. J Bacteriol 186:2646–2654

Sabree ZL, Kambhampati S, Moran NA (2009) Nitrogen recycling and nutritional provisioning by
    Blattabacterium, the cockroach endosymbiont. Proc Natl Acad Sci USA 106:19521–19526

Scarborough CL, Ferrari J, Godfray HC (2005) Aphid protected from pathogen by endosymbiont.
    Science 310:1781

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H (2000) Genome sequence of the
    endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature 407:81–86

Silva FJ, Latorre A, Moya A (2001) Genome size reduction through multiple events of gene
    disintegration in *Buchnera* APS. Trends Genet 17:615–618

Silva FJ, Latorre A, Moya A (2003) Why are the genomes of endosymbiotic bacteria so stable?
    Trends Genet 19:176–180

Simon JC, Carre S, Boutin M, Prunier-Leterme N, Sabater-Mun B, Latorre A, Bournoville R
    (2003) Host-based divergence in populations of the pea aphid: insights from nuclear markers
    and the prevalence of facultative symbionts. Proc Biol Sci 270:1703–1712

Tamames J, Gil R, Latorre A, Pereto J, Silva FJ, Moya A (2007) The frontier between cell and
    organelle: genome analysis of Candidatus *Carsonella ruddii*. BMC Evol Biol 7:181

Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran
    NA, Andersson SG (2002) 50 million years of genomic stasis in endosymbiotic bacteria.
    Science 296:2376–2379

Thao ML, Gullan PJ, Baumann P (2002) Secondary (gamma-Proteobacteria) endosymbionts infect
    the primary (beta-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve
    with their hosts. Appl Environ Microbiol 68:3190–3197

Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S (2006) Massive
    genome erosion and functional adaptations provide insights into the symbiotic lifestyle of
    Sodalis glossinidius in the tsetse host. Genome Res 16:149–156

Tokuda G, Lo N, Takase A, Yamada A, Hayashi Y, Watanabe H (2008) Purification and partial
    genome characterization of the bacterial endosymbiont *Blattabacterium cuenoti* from the fat
    bodies of cockroaches. BMC Res Notes 1:118

Touchon M, Rocha EP (2007) Causes of insertion sequences abundance in prokaryotic genomes.
    Mol Biol Evol 24:969–981

Tsuchida T, Koga R, Fukatsu T (2004) Host plant specialization governed by facultative symbiont.
    Science 303:1989

van Ham RC, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U, Fernandez JM, Jimenez
    L, Postigo M, Silva FJ, Tamames J, Viguera E, Latorre A, Valencia A, Moran F, Moya A
    (2003) Reductive genome evolution in *Buchnera aphidicola*. Proc Natl Acad Sci USA
    100:581–586

van Hoek AH, van Alen TA, Sprakel VS, Leunissen JA, Brigge T, Vogels GD, Hackstein JH
    (2000) Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates. Mol
    Biol Evol 17:251–258

von Dohlen CD, Moran NA (2000) Molecular data support a rapid radiation of aphids (Aphididae)
    in the Cretaceous and multiple origins of host alternation. Biol J Linn Soc 71:689–717

von Dohlen CD, Kohler S, Alsop ST, McManus WR (2001) Mealybug beta-proteobacterial
    endosymbionts contain gamma-proteobacterial symbionts. Nature 412:433–436

Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs L, Bolanos R,
    Keller M, Kretz K, Lin X, Mathur E, Ni J, Podar M, Richardson T, Sutton GG, Simon M, Soll D,
    Stetter KO, Short JM, Noordewier M (2003) The genome of *Nanoarchaeum equitans*: insights
    into early archaeal evolution and derived parasitism. Proc Natl Acad Sci USA 100:12984–12988

Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, Mayhew GF, Plunkett G
    3rd, Rose DJ, Darling A, Mau B, Perna NT, Payne SM, Runyen-Janecky LJ, Zhou S, Schwartz DC,
    Blattner FR (2003) Complete genome sequence and comparative genomics of *Shigella flexneri*
    serotype 2a strain 2457T. Infect Immun 71:2775–2786

Welburn SC, Maudlin I (1999) Tsetse-trypanosome interactions: rites of passage. Parasitol Today
    15:399–403

Wernegreen JJ (2005) For better or worse: genomic consequences of intracellular mutualism and parasitism. Curr Opin Genet Dev 15:572–583

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci USA 87:4576–4579

Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO, Boffelli D, Anderson IJ, Barry KW, Shapiro HJ, Szeto E, Kyrpides NC, Mussmann M, Amann R, Bergin C, Ruehland C, Rubin EM, Dubilier N (2006) Symbiosis insights through metagenomic analysis of a microbial consortium. Nature 443:950–955

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC, McGraw EA, Martin W, Esser C, Ahmadinejad N, Wiegand C, Madupu R, Beanan MJ, Brinkac LM, Daugherty SC, Durkin AS, Kolonay JF, Nelson WC, Mohamoud Y, Lee P, Berry K, Young MB, Utterback T, Weidman J, Nierman WC, Paulsen IT, Nelson KE, Tettelin H, O'Neill SL, Eisen JA (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. PLoS Biol 2:E69

Wu D, Daugherty SC, Van Aken SE, Pai GH, Watkins KL, Khouri H, Tallon LJ, Zaborsky JM, Dunbar HE, Tran PL, Moran NA, Eisen JA (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol 4:e188