Nature Reviews Genetics | AOP, published online 12 February 2008; doi:10.1038/nrg2319

Learning how to live together: genomic insights into prokaryote–animal symbioses

Andrés Moya‡, Juli Peretó*§, Rosario Gil*‡ and Amparo Latorre*‡*

Abstract | Our understanding of prokaryote–eukaryote symbioses as a source of evolutionary innovation has been rapidly increased by the advent of genomics, which has made possible the biological study of uncultivable endosymbionts. Genomics is allowing the dissection of the evolutionary process that starts with host invasion then progresses from facultative to obligate symbiosis and ends with replacement by, or coexistence with, new symbionts. Moreover, genomics has provided important clues on the mechanisms driving the genome-reduction process, the functions that are retained by the endosymbionts, the role of the host, and the factors that might determine whether the association will become parasitic or mutualistic.

Symbiosis
From the Greek, *sym* 'with' and *biosis* 'living'. A long-term association between two or more organisms of different species that is integrated at the behavioural, metabolic or genetic level. According to the level of dependence on the host, symbiosis can be obligate or facultative. The term was introduced by Anton de Bary and Albert Bernard Frank when discussing lichens and mycorrhizae, respectively, at the end of the 1870s.

**Institut Cavanilles de Biodiversitat i Biologia Evolutiva, Universitat de València, Apartado de correos 22085. 46071 València and CIBER de Epidemiología y Salud Pública, Spain.*
‡Departament de Genètica and §Departament de Bioquímica i Biologia Molecular, Universitat de València, Dr. Moliner, 50. 46100 Burjassot (València), Spain.
Correspondence to A.M.
e-mail: andres.moya@uv.es
doi:10.1038/nrg2319
Published online
12 February 2008

Prokaryotic microorganisms are widespread in all the environments on Earth. Given their ecological ubiquity, it is not surprising to find many prokaryotic species in close relationships with members of many eukaryotic taxa, often establishing a persistent association, which is known as symbiosis (BOX 1). According to the fitness effects on the members of the symbiotic relationship, associations can be referred to as parasitism, mutualism or commensalism and, depending on the location of the symbiont with respect to host cells, as ectosymbiosis or endosymbiosis. 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 dichotomic classification between obligate primary endosymbionts (P-endosymbionts) and facultative secondary symbionts (S-symbionts). P-endosymbionts generally have long evolutionary histories with their hosts, whereas S-symbionts seem to have established more recent associations, retaining the ability to return to a free-living condition[1–4]. In some cases, an S-symbiont can evolve to become an obligate partner and, if a P-endosymbiont is already present, microbial consortia can be established.

For several decades, the idea that microbial associations might be central to eukaryote evolution remained controversial. However, the work of Lynn Margulis since the late 1960s not only contributed to the establishment of a symbiotic theory for cell evolution[5], but also revived earlier proposals that were forgotten by biologists[6]. Today, there is a wide consensus on the essential role of symbioses during the origin and evolution of eukaryotic cells, although controversies about the details persist[7]. In addition, on the basis of its wide distribution across major phylogenetic taxa (BOX 1), symbiosis could have an important role in the evolution of species that are involved in such partnerships. In all the major biological phenomena that are classified as symbiotic, new cellular structures and/or metabolic capabilities emerge as a result of evolutionary forces, favouring the maintenance of the association[8]. The interplay of both partners or, in some cases, between a single host and more than one symbiont, forms an evolving biological community that changes throughout time.

In most cases of symbioses between prokaryotes and eukaryotes, the relationship between host and symbiont is so close that the microorganisms cannot be cultured, making them difficult to study. However, in the past few years, genome sequencing, transcriptomics and the recently emerged field of metagenomics — which avoid having to culture microbial cells — have offered new opportunities for symbiosis research. Whole genomes of several intracellular bacterial symbionts have been sequenced, allowing comparisons among the evolutionary innovations of these bacteria, from being free living to various stages of integration with their respective hosts[9–23] (TABLE 1). Such genomics approaches to the study of symbiosis will allow several questions to be tackled, contributing to an understanding of the evolutionary relevance of this phenomenon. In particular, these questions relate to the nature of the

Box 1 | **Spread of symbiosis in the biosphere**

Stable symbioses have independently evolved many times in diverse groups of eukaryotes. Although the scope of this Review is restricted to prokaryotic partners that are associated with animal hosts, there are many other remarkable examples of symbiosis between eukaryotic organisms, as exemplified by fungal associations with other fungi, protists, plants (that is, mycorrhizae), algae (that is, lichens) and animals (for an overview, see REF. 90). Most symbioses have a proven biochemical foundation. In some cases, one of the partners benefits from compounds that are produced by the other; in other cases, the waste products of one partner (especially nitrogen compounds) are recycled by the other.

The eukaryotic nucleocytoplasm has limited metabolic capabilities. The mitochondrial respiratory chain and oxygenic photosynthesis in the chloroplast are two examples of metabolic functions that have been acquired through symbiotic associations with prokaryotic partners during early eukaryotic evolution[8]. Countless symbiotic relationships have emerged more recently, or are in the process of developing, and they complement the limited metabolic networks of most eukaryotes with several prokaryotic metabolic capabilities, such as dinitrogen fixation[91], methanogenesis[92], chemolithoautotrophy[93], nitrogen assimilation[94] and essential-nutrient anabolism[95,96]. In particular, animal heterotrophic metabolism is relatively narrow and some amino acids, vitamins and fatty acids must be obtained from an external source. Thus, symbiotic associations with microorganisms have allowed animals to adapt to specialized feeding behaviours by providing the nutrients that are deficient in their restricted diets. In fact, all intracellular mutualistic symbioses between bacteria and animals that have been analysed at the genomic level[9–23] (TABLE 1) are related to nutrient provision and waste recycling. In the best-studied cases — those of symbioses involving insects[97] — the presence of such associations throughout most of insect evolutionary history suggests that symbiosis has been a driving force in the diversification of this group.

The figure shows the phylogenetic distribution of symbioses, indicating the bacterial and archaeal classes within which there are associations with eukaryotic hosts. Data were collected from the literature and are the result of a long tradition of studies that have used ecological, developmental, morphological, biochemical and genomic approaches to investigate symbiosis (a list of examples is given in the Supplementary information S2 (table)). The uneven distribution of symbiosis highlights the diverse interests and motivations of the scientific community. Advances in genomics have had a large impact on the research into microbial symbioses, which has implications for biotechnology (for example, sponge-associated prokaryotes[98]), agriculture (for example, nitrogen fixation in plant-associated bacteria[91]) and biomedicine (for example, the human intestine microbiome[99]). Metagenomic methods have notably increased our knowledge of the biodiversity of non-cultivable bacteria in symbiotic consortia, including the description of completely new candidate phyla — Endomicrobia[100] and Poribacteria[101]. Postulated symbiotic events leading to the evolutionary origin of organelles (mitochondria and chloroplasts) are indicated. The branching order in the eukaryotic phylogenetic tree has been adapted from REFS 102,103. The phylogenetic tree of animals is an adaptation from REF. 104, modified by recent molecular phylogenies[105,106]. The lengths of branches are not to scale. An asterisk indicates that complete genomes are available (see TABLE 1).

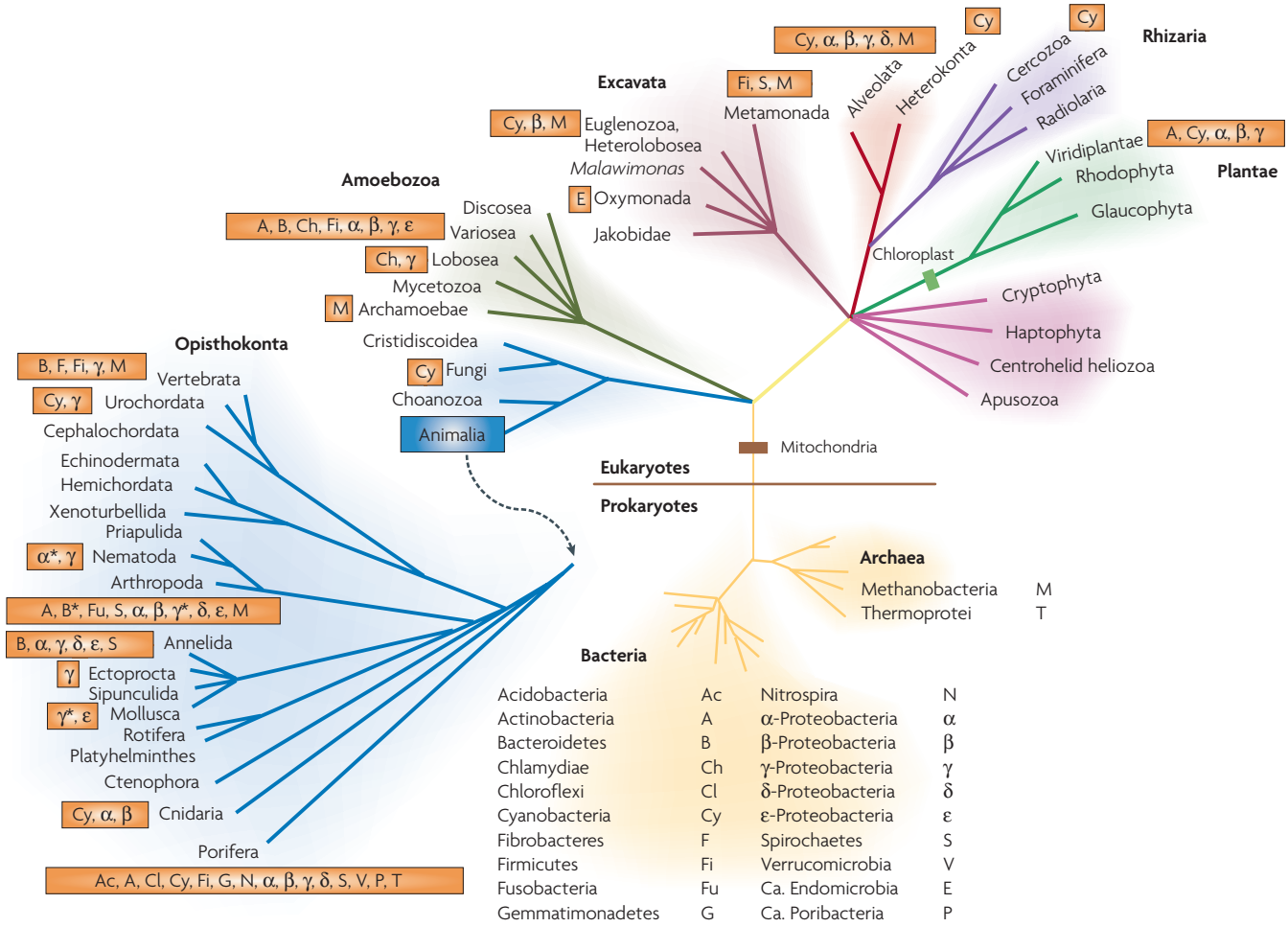

Table 1 | **Genomic data for mutualistic symbionts of animals**

Organism	Host	Metabolic mode	Genome size (kb)	GC content (%)	CDS	rRNAs	tRNAs	Pseudogenes	Accession number
Buchnera aphidicola BAp*	*Acyrthosiphon pisum* (aphid)‡	Heterotroph	652	26.24	574	3	32	12	BA000003, AP001070, AP001070
Buchnera aphidicola BSg*	*Schizaphis graminum* (aphid)	Heterotroph	653	26.3	556	3	32	33	AE013218, AF041836, Z21938
Buchnera aphidicola BBp*	*Baizongia pistaciae* (aphid)	Heterotroph	618	25.3	507	3	32	9	AE016826, AF492591
Buchnera aphidicola BCc*	*Cinara cedri* (aphid)	Heterotroph	422	20.2	362	3	31	3	CP000263, AY438025
*Blochmannia floridanus**§	*Camponotus floridanus* (carpenter ant)	Heterotroph	706	27.4	583	3	37	4	BX248583
*Blochmannia pennsylvanicus**§	*Camponotus pennsylvanicus* (carpenter ant)	Heterotroph	792	29.6	610	3	39	4	CP000016
*Wigglesworthia glossinidia**	*Glossina brevipalpis* (tsetse fly)	Heterotroph	698	22.5	617	6	34	14	BA000021, AB063523
*Sodalis glossinidius**	*Glossina morsitans* (tsetse fly)‡	Heterotroph (Secondary)	4,171	54.7	2,516	7	69	972	AP008232, AP008233, AP008234, AP008235
*Baumannia cicadellinicola**§	*Homalodisca coagulata* (sharpshooter)	Heterotroph	686	33.2	595	6	39	9	CP000238
Sulcia muelleri§‖	*Homalodisca coagulata* (sharpshooter)	Heterotroph	245	22.4	227	3	31	–	CP000770
Carsonella ruddii§	*Pachypsylla venusta* (psyllid)	Heterotroph	160	16.6	182	3	28	–	AP009180
Wolbachia wBm¶	*Brugia malayi* (nematode)#	Heterotroph	1,080	34	805	3	34	98	AE017321
Ruthia magnifica§	*Calyptogena magnifica* (Deep-sea clam)	Autotroph	1,200	34.0	976	3	36	–	CP000488
Vesicomyosocius okutanii§	*Calyptogena okutanii* (Deep-sea clam)	Autotroph	1,000	31.6	937	3	35	–	AP009247
Nitratiruptor sp**	Deep-sea-vent animals	Autotroph	1,878	39.7	~1,118	9	45	~739‡‡	AP009178
Sulfurovum sp**	Deep-sea-vent animals	Autotroph	2,563	43.8	~1,218	9	44	~1,248‡‡	AP009179

The data shown are for those symbionts that have been sequenced as of January 2008. Data were retrieved from the National Center for Biotechnology Information (NCBI). *γ-Proteobacteria. ‡Genome sequence in progress. §These bacteria are called *Candidatus*. ‖Bacteroidetes. ¶α-Proteobacteria. #Complete genome available, see REF. 78. **ε-Proteobacteria; although it is unclear whether these two isolates are epibiotic symbionts or another variation of symbionts, many genome features strongly support their symbiotic lifestyle. ‡‡Estimated from REF. 20. CDS, coding sequences; rRNA, ribosomal RNA.

genome-reduction process, the type of genes that are retained by the endosymbiont, the molecular pathways that are used by the host to control the endosymbiont population and the complex set of factors that determine whether the final outcome of the association is parasitic or mutualistic.

Here we focus on the study of mutualistic endosymbioses of animals for which genomic studies have been carried out, providing new biological and evolutionary insights. Although only one host genome has been sequenced, complete genomic sequences for several bacterial symbionts are available. Comparisons with free-living relatives have revealed dramatic changes in genome size and content, and have begun to reveal the mechanisms by which these changes take place and their functional consequences. Furthermore, the mechanisms that are involved in the establishment, maintenance and evolution of mutualistic associations are being scrutinized; this has led to the emergence of new insights into aspects such as metabolic interdependencies between partners and the survival of the prokaryote in the face of the host immune response.

Parasitism
Symbioses in which one species is increasing its fitness while the fitness of the other species is adversely affected.

Mutualism
Symbioses in which both species increase their fitness.

Commensalism
Symbioses in which one partner is increasing its fitness without affecting the other species.

Genomics coverage of prokaryotic endosymbionts

Obligate endosymbionts of animals. In recent years, the genomes of several bacterial P-endosymbionts of animals have been completely sequenced (TABLE 1). P-endosymbionts are inherited by strict vertical transmission and, in most cases, reside in specialized host cells called bacteriocytes. Functional analyses have corroborated the nutritional role of the associations; each bacterium provides the nutrients that are deficient in the diet of its animal-host, whereas the bacterium gains a permanent supply of a wide range of metabolites that are provided by the host.

Most of the endosymbionts that have been studied are γ-proteobacteria that live in obligate association with insects, and they provide various functions to

their host, including: the provision of essential amino acids[9,10,12–14,16,21,23], vitamins and cofactors[9–12,15] that are absent in the host diet; nitrogen recycling and storage[13,14]; and the provision of metabolic factors that are required for survival and fertility[19]. Whereas most of these species are heterotrophic endosymbionts, two recently sequenced endosymbionts that are harboured by deep-sea clams of the genus *Calyptogena*[17,18] are chemolithoautotrophic, and provide almost all nutrients to the host by fixing CO_2 and using H_2S as a source of energy and reducing power (FIG. 1 and Supplementary information S1 (figure)).

Facultative symbionts. Occasionally, hosts tolerate S-symbionts that coexist with a P-endosymbiont and that can be either deleterious or beneficial to the host[24]. In aphids, a number of facultative symbionts that reside in multiple host tissues, cells surrounding primary bacteriocytes or in their own bacteriocytes have been described[25]. Although they are normally vertically transmitted, their distribution patterns suggest that sporadic horizontal transmission among host individuals and host species must have occurred[26]. Among the possible beneficial roles that have been identified, some of these S-symbionts can rescue the host from heat damage[27–29], provide defence against natural enemies (parasitoids and pathogens)[30–32] and participate in host specialization[33–36]. In other cases, facultative symbionts can spread among lineages without conferring a benefit, by manipulating host reproduction (reviewed in REF. 37). For example, in *Wolbachia* infections of arthropods, the symbiont undergoes transfer among host lineages. Remarkably, *Wolbachia* can also be a P-endosymbiont in filarial nematodes[19], indicating that the same prokaryotic lineage can take part in more than one type of symbiotic association.

At present, the only completely sequenced genome of an S-symbiont is that of *Sodalis glossinidius* (TABLE 1). This bacterium coexists in the gut lumen of tsetse flies with the P-endosymbiont *Wigglesworthia glossinidia*, although they occupy different areas of the gut and can be found both intra- and extracellularly[22]. It has been suggested that *S. glossinidius* has an important role in the acquisition of trypanosome infections[38].

Endosymbiotic consortia. Although the importance of syntrophy between unrelated organisms has been recognized in several cases[39], the new field of metagenomics provides the ability to study bacterial endosymbionts that coexist in the same host, enabling the discovery of complex and stable associations, and analysis of the contribution of each partner to the relationship.

One endosymbiotic consortium that has recently been reported involves *Buchnera aphidicola* BCc and *Serratia symbiotica* SCc, which coexist in the bacteriome of the aphid *Cinara cedri*[40]. *S. symbiotica* is a facultative symbiont in other aphid species, but has been found in all the cedar-aphid populations that have been analysed, casting doubts over its facultative status in this species. Functional and evolutionary comparative analyses of the sequenced *B. aphidicola* genomes, as well as microscopic analysis of the two cedar-aphid endosymbionts (FIG. 2), led Pérez-Brocal and co-workers[21] to conclude that *S. symbiotica* SCc might have the potential to replace *B. aphidicola* BCc.

In a second example, genomic sequencing has been carried out for *Baumannia cicadellinicola* and the Bacteroidetes species *Sulcia muelleri*, the co-resident P-endosymbionts of the xylem-feeding sharpshooter *Homalodisca coagulata*[15,23]. Genomic analysis has revealed that the two symbionts have complementary biosynthetic capabilities, which are needed to provide their host with nutrients that are lacking in xylem sap. Phylogenetic analysis revealed that *Sulcia* was ancestrally present in a host lineage that acquired *Baumannia* at the same approximate time as the switch to xylem feeding, consistent with the view that its nutrient-provisioning capabilities were a requirement for the host to evolve towards this lifestyle[41]. As in the case of *Buchnera–Serratia*, the two symbionts live in close proximity within the host bacteriome[15].

Metagenomic approaches have also been used to analyse more complex consortia, including the four co-occurring symbionts from the segmented worm *Olavius algarvensis*[42]. In this instance, the host belongs to a group of oligochaetes that lack a mouth, gut and anus, and that are unique among annelids in having reduced their nephridial excretory system. *O. algarvensis* harbours at least four symbionts, two γ-proteobacteria (sulphur oxidizing) and two δ-proteobacteria (sulphate reducing), which fix CO_2 and provide the host with nutrients. Almost all amino acids and several vitamins can be synthesized by the symbionts and then provided to their host, probably through controlled digestion of the bacteria as suggested by the vacuole localization of the host[43]. The symbionts are also likely to be involved in host-waste recycling.

Genomic changes during endosymbiont evolution

Genomic changes experienced by endosymbionts. A general feature of intracellular pathogenic and mutualistic bacteria is that they have smaller genomes with a higher AT content than their free-living relatives[44]. The causes of reductive-genome evolution are related to both the genetic information that is needed in the new environment and to population dynamics, although the relative importance of these factors has not been determined[45,46]. The change from a free-living environment to one that is intracellular and protected implies that many genes are rendered unnecessary, whereas others become redundant because their functions can be supplied by the host. Therefore, some genetic material can be lost without a detrimental effect, and can accumulate mutations owing to the lack of effective natural selection to purge them. In addition, owing to the strict vertical transmission of obligate endosymbionts, only a few bacteria are transmitted, which generates continuous bottlenecks that favour the action of random genetic drift. The combination of these factors allows the accumulation of mutations in genes that can be beneficial but that are not essential — such as genes involved in DNA repair and recombination (see below) — further increasing the repertoire of genes that can be lost, and reducing the possibility of genetic exchange by homologous recombination. Furthermore,

Ectosymbiosis
A symbiosis in which the symbiont lives on the body surface of the host, including internal surfaces such as the lining of the digestive tube and the ducts of glands.

Endosymbiosis
Symbioses in which a prokaryote symbiont lives inside a eukaryotic cell.

Primary endosymbiont
(P-endosymbiont). Obligate bacterial endosymbionts that live inside specialized animal host cells called bacteriocytes. The association is obligate for both partners.

Secondary symbiont
(S-symbiont). Facultative bacterial endosymbiont that coexists with a P-endosymbiont. Often located in syncitial cells near the bacteriocyte and in various other insect tissue types. Secondary symbionts are not essential for host survival and are transferred horizontally among individuals of both the host species and other species.

Metagenomics
The application of genomic analyses to uncultured microorganisms. Also referred to as environmental genomics.

Vertical transmission
The endosymbionts are maternally transferred, that is, directly from a host to its offspring.

Bacteriocytes
Specialized cells of the host species in which symbiotic bacteria live.

Heterotrophy
The metabolic mode in which the carbon source is organic matter. By extension, this is a metabolic mode in which organic matter is the source of carbon, electrons and energy (chemoorganoheterotrophy).

Chemolithoautotrophy
The metabolic mode in which CO_2 is the carbon source and an inorganic chemical reaction is both the electron and energy source.

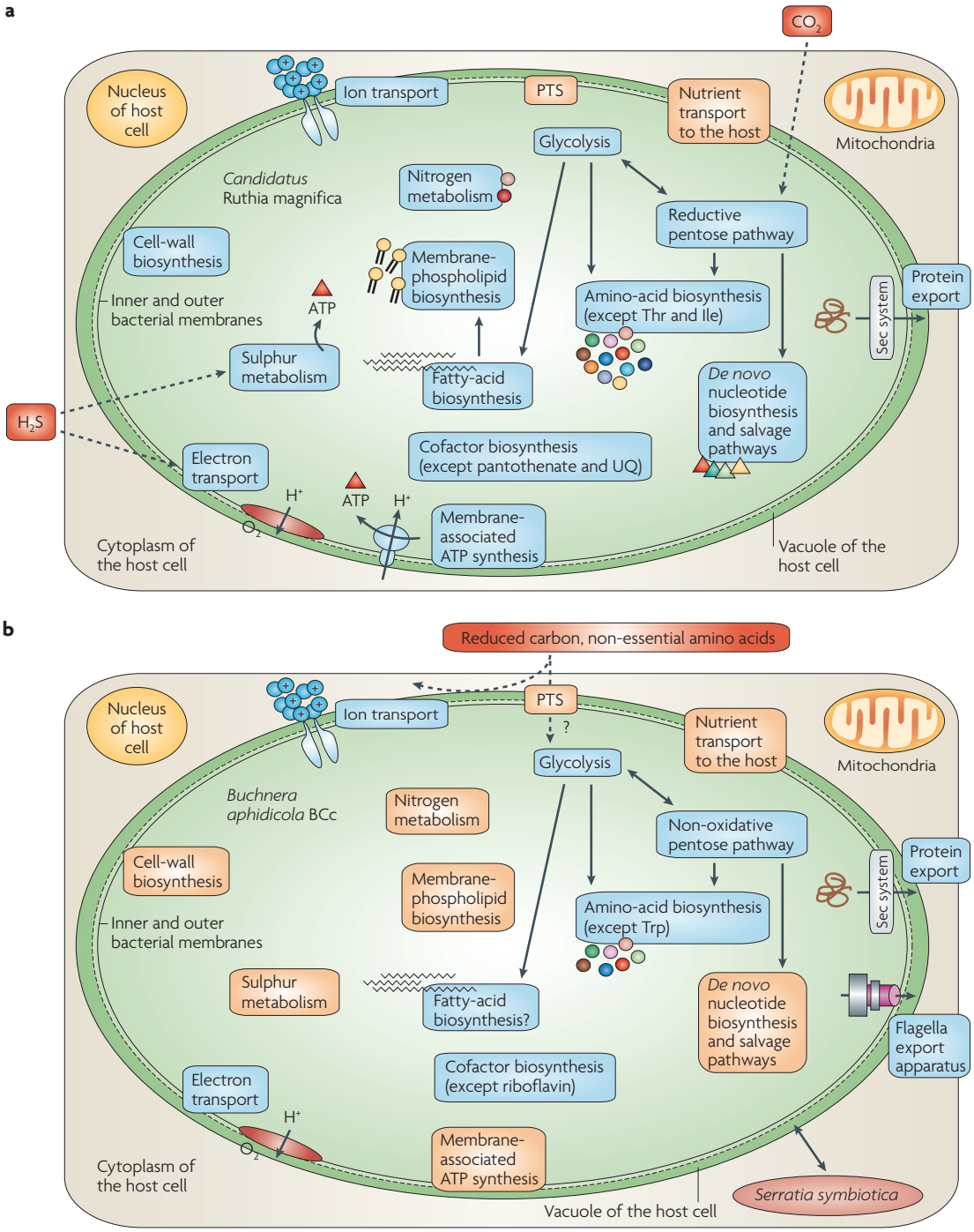

Figure 1 | **Comparative biochemistry of symbiont–host interdependence.** As a study case, we use information from complete genome sequences to infer and compare the metabolisms that are encoded by the reduced genomes of an autotrophic and a heterotrophic endosymbiont. Blue indicates the presence of a metabolic process, orange indicates its absence. **a** | *Candidatus* Ruthia magnifica[17] uses electrons from H_2S to drive an autotrophic metabolism based on the reductive pentose phosphate pathway (Calvin–Benson cycle), supplies essential nutrients (including amino acids and cofactors) to its host, and recycles nitrogen. As is the case in the *Calyptogena okutanii* endosymbiont[18], it has the potential for anabolism of all essential biomolecules except threonine (Thr), isoleucine (Ile) and ubiquinone (UQ). The missing steps in those pathways could be explained by the function of an unidentified enzyme, or by complementation with the host metabolism. **b** | *Buchnera aphidicola* BCc uses reduced-carbon molecules and non-essential amino acids to fuel its heterotrophic metabolism and synthesize essential nutrients, with the exception of tryptophan (Trp) and riboflavin[21]. These molecules could be provided by a second co-existing endosymbiont, *Serratia symbiotica*. The self-sufficiency of *R. magnifica* sharply contrasts with the incomplete metabolic abilities of *B. aphidicola* BCc. Among the few similarities between both metabolisms, note the absence of nutrient-transport systems to the host. A more detailed version can be found in the Supplementary information S1 (figure). PTS, phosphotransferase system; Sec system, secretory system.

Horizontal transmission
Some endosymbionts retain a generalized ability to colonize and persist in multiple hosts, that is, their transmission is between individuals of the same or different host species, rather than from parent to offspring.

Syntrophy
Emergence of new metabolic capabilities as a result of symbiosis, it is often essential for the survival of the consortium.

Bacteriome
An organ-like structure formed by bacteriocytes.

the intracellular environment isolates the symbiont from other bacterial species and thus reduces, and can even eliminate, the possibility of gaining new genetic material by horizontal gene transfer. Thus, gene losses can become irreversible.

In addition to having reduced genomes, an increased AT content is another general feature of the endosymbiont genomes that have been sequenced so far. The bias towards AT is probably due to the loss of genes for DNA repair, leading to a mutational GC to AT pressure[47] (for a metabolic hypothesis see REF. 48). Such a bias has a notable effect on many genes by altering the structure and function of the corresponding proteins. In fact, it has been proposed that GroEL, a chaperonin that is constitutively overexpressed in *B. aphidicola,* helps to post-translationally correct the altered structure of many proteins that is caused by this bias[49]. A by-product of the base-composition bias of these genomes is the loss of the codon-usage bias that is typical of free-living bacteria: it is almost absent from *B. aphidicola*, and is highly reduced in P-endosymbionts with larger genomes and in S-symbionts[50,51].

The degree of both genome-size reduction and increase in AT content vary among endosymbionts, and correlate with the age of the association. P-endosymbionts that are partners in old associations have generally smaller genomes and an AT content higher than 70%, whereas S-symbionts and P-endosymbionts that are part of younger associations have genome sizes and AT contents intermediate with respect to older P-endosymbionts and free-living relatives[2,25,50].

The genome-reduction process examined. To understand the different stages of the genome-reduction process it is necessary to analyse a range of genome sequences, from endosymbionts that still have genomes similar in size to those of their free-living relatives, to those that have undergone the most dramatic reductions. Such comparative analyses suggest that the process of genome shrinkage might have taken place in two separate stages.

A massive gene loss must have occurred soon after the establishment of the obligate symbiosis, probably by means of large deletions that would cause the elimination of a series of contiguous genes[52]. The accumulation of mobile elements, representing a source for chromosomal rearrangements and gene inactivation, seems to have an important role at this first stage. This is suggested by the fact that mobile elements are relatively abundant in bacteria that have recently acquired an obligate, intracellular way of life[53,54]. Mobile elements, such as bacterial insertion sequences (IS), are fairly abundant in the S-symbiont *S. glossinidius*[22], but even more so (estimated at more than 25% of genome content) in its close relatives SOPE (*Sitophilus oryzae* primary endosymbiont) and SZPE (*Sitophilus zeamais* primary endosymbiont). These species are P-endosymbionts of the rice and maize weevils, respectively, and both have recently established obligate associations with their hosts[55,56] (sequencing of the SOPE genome is in progress).

These data indicate that a common symbiotic ancestor of these two lineages must already have possessed these elements. The increase in frequency of these elements in the newly established P-endosymbiont must be due to an increase in the replicative transposition of elements that were resident at the onset of symbiosis[24]. All analysed endosymbiotic bacteria that are involved in older associations possess genomes that are free of mobile elements. Presumably, therefore, the expansion of mobile elements must at some point have become deleterious and they must have been removed as part of the process of genome degradation. The difference in the abundance of transposable elements in two strains of *Wolbachia* with different lifestyles and genome sizes supports this statement. In *Wolbachia pipientis* *w*Mel[57], a reproductive parasite of *Drosophila melanogaster*, 14% of the 1.27-Mb symbiont genome is occupied by repetitive DNA and mobile elements, whereas in *W. pipientis* *w*Bm[19], the obligate endosymbiont of the nematode *Brugia malayi*, these elements account for just 5.4% of the 1.08-Mb genome.

During the second stage of the genome-reduction process, genome shrinkage seems to have mostly occurred through a process of gradual gene loss, scattered along the genome. Such loss seems to follow a pattern that starts with the inactivation of a gene (pseudogenization) by single-nucleotide mutations,

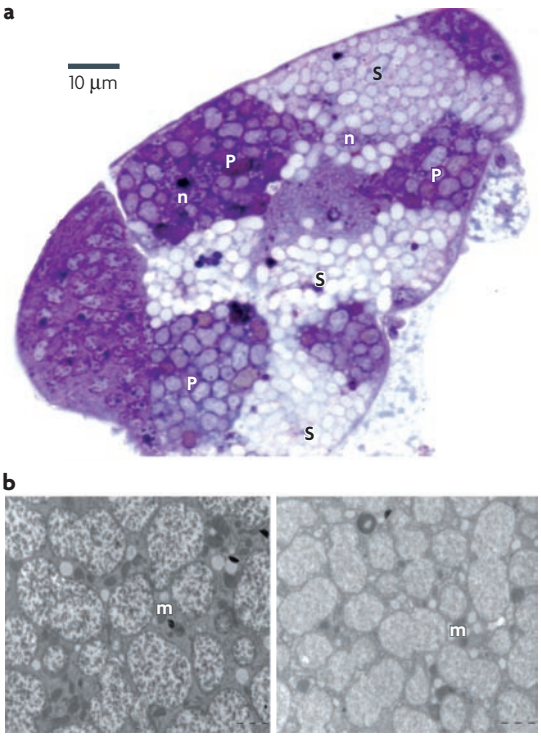

Figure 2 | **Bacteriocytes of the aphid *Cinara cedri*.** **a** | A 1.5 mm semi-thin section that shows both primary and secondary bacteriocytes, easily identifiable by their different tones under toluidine blue staining. **b** | Electron micrographs of the same individual that show the round shape of *Buchnera aphidicola* and *Serratia symbiotica*, respectively, in their bacteriocytes. P, primary symbiont (*B. aphidicola*); S, secondary symbiont (*S. symbiotica*); n, nuclei of both bacteriocytes; m, mitochondria. Electron micrographs courtesy of A. Lamelas, Universitat de València, Spain.

and continues with a rapid reduction in length until the original gene is completely eroded[58,59]. Even in advanced stages of the reductive process, genome-length reduction is mainly due to the loss of protein-coding genes, not to their shortening[21]. Regarding intergenic regions, there is a slight size reduction in the four sequenced *B. aphidicola* strains when compared with their close relative *Escherichia coli*[60]. The shortening of intergenic regions is more evident in extremely reduced genomes, as observed in *Carsonella ruddii*[16], in which this has been so extreme as to have led to overlapping genes. In *C. ruddii,* the annotated ORFs are also considerably shorter than their orthologues in other bacteria, although the functionality of these shortened ORFs is questionable[61].

Functional changes in hosts and symbionts

The first step towards the establishment of an obligate endosymbiosis takes place when a free-living bacterium infects the host. Then, both organisms co-evolve to adapt to the new situation. On the bacterial side, genomics 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; however, a symbiotic bacterium must still retain a number of functions to allow its continued survival in the host. By contrast, our understanding of how the animal host is affected at the genetic level, and how it has solved the problem of having a bacterium (or more than one) inside its cells, is poorly understood. However, host specialization implies that it is now dependent on the bacteria to survive and reproduce. Adaptive changes in the host include the development of specialized cells where bacteria reside, adequate systems to control bacterial populations, plus the modification of its immune response against the intruder cells.

General changes in endosymbiont gene content. The catalogue of individual genes that are shared by all analysed endosymbionts suggests that the molecular mechanisms that are necessary for survival in an intracellular environment share broad commonalities across all endosymbiotic associations[13]. Only about one-third of the coding capacity of each endosymbiont genome that has been sequenced so far is devoted to the requirements of the specific symbiosis, with these genes mainly reflecting differences in host lifestyle, nutritional needs and location within host cells.

Across the sequenced endosymbiont genomes, genes that are involved in essential functions — such as those involved in DNA replication, transcription and translation — are more likely to be retained than are genes of other functional classes, and sometimes account for more than one-third of the genome[21]. Chaperone systems and all essential components of the protein translocation machinery are also retained in these genomes, ensuring the correct folding and localization of protein components[9,49].

Although gene losses affect all functional categories, they are nonrandom. The most dramatic losses affect genes that are involved in metabolism but are not required for host survival, as we discuss later. Another general feature of all sequenced endosymbiont genomes is the loss of most DNA repair and transcriptional regulatory mechanisms. Furthermore, losses affect most genes of the latter category, indicating that there is no need for transcriptional regulation in a stable environment. This hypothesis is supported by the results of microarray experiments in *B. aphidicola*[62,63], which show constitutive gene expression of metabolic genes, regardless of changes in environmental conditions.

Other features of the functional content of endosymbiont genomes are more dependent on their specific lifestyles. For example, the bacterial cell envelope seems to be highly simplified, being less structured in bacteria living inside host-derived vesicles (for example, *B. aphidicola*) than in endosymbionts that live free in the cytosol of bacteriocytes (for example, *Blochmannia floridanus* and *W. glossinidia*)[13]. The simplification can be extreme; for example, *B. aphidicola* BCc has lost all the genes that are involved in the biosynthesis of the bacterial cell wall[21].

Metabolic changes and nutrient transport. The bigger endosymbiont genomes, belonging to more recently established partnerships, have retained many genes that are involved in several intermediary metabolic pathways, but most of these have been lost in the smaller genomes of well-established endosymbionts (FIG. 1; TABLE 1). The loss of metabolic genes is also affected to a certain extent by the presence of other bacterial species in the intracellular environment of the symbiont, in that the simultaneous presence of S-symbionts can compensate for the metabolic deficiencies of the P-endosymbiont, as well as those of the host[64].

Even though each endosymbiont (or the combination of symbionts within a consortium) has specifically retained the pathways that are necessary to synthesize the nutrients that respective hosts are unable to provide for themselves, endosymbionts exhibit limited transport capabilities[65] (FIG. 1). This situation has been observed in *B. aphidicola*, the two deep-sea-clam endosymbionts that were recently analysed[17,18] and the endosymbionts of *O. algarvensis*[42]. The subcellular localization of endosymbionts inside vacuole-like compartments[66,67], and the presence of lysozyme in the host cells[68,69], strongly suggest that, in these systems, the host nutritional needs could be satisfied by controlled weakening or killing of symbiont cells — a case of a necrotrophic host–symbiont relationship.

Invasion strategies: have pathogens been domesticated? Recent studies of the mechanisms that are used by endosymbionts to establish and maintain their infection of host tissues indicate that invasion strategies are based on the same molecular tools that are used by well-studied pathogenic bacteria (reviewed in REFS 24,47). These mechanisms include the use of various secretion systems for the attachment and invasion of host cells[64] and the utilization of the same quorum-sensing mechanisms for the regulation of virulence or mutualistic traits, depending on the type of association[17,18].

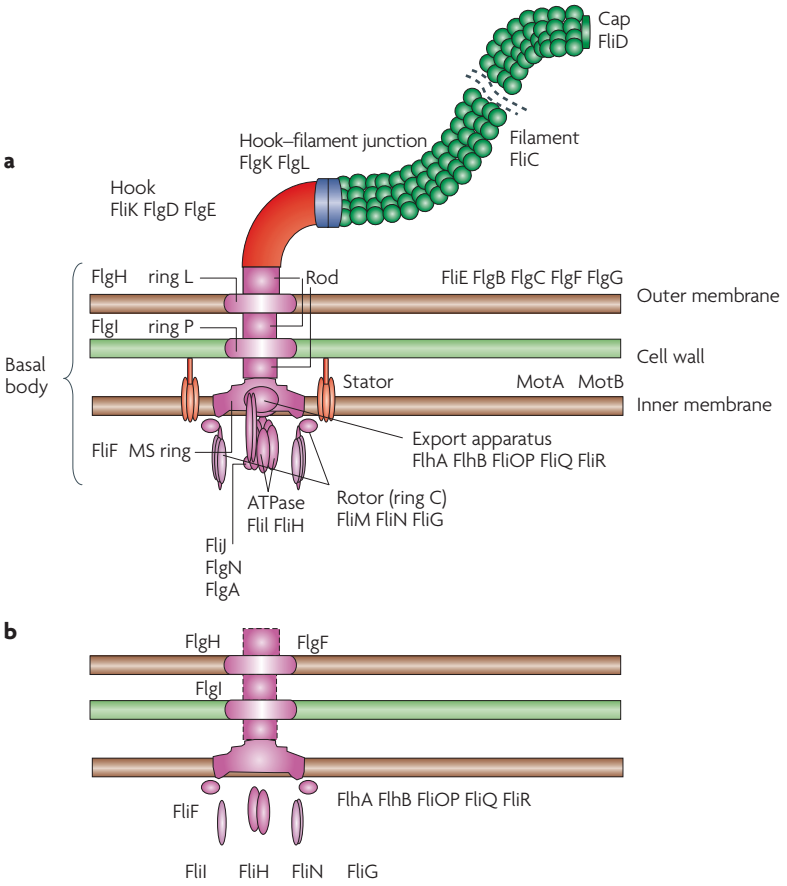

Figure 3 | **The flagellum, an example of genome reduction at the structural level.** **a** | The *Escherichia coli* flagellum. Protein and structural components are indicated: cap, filament (formed by ~20,000 copies of flagellin), hook–filament junction, hook and basal body. The basal body is formed by a set of rings for anchorage to the inner and outer membranes and to the cell wall, the rod, the stator, the rotor and the export apparatus (including one ATPase). **b** | Vestiges of the flagellar apparatus in *Buchnera aphidicola* BCc. Elements that are involved in the formation of the flagellar filament and hook, the stator/motor and the transcriptional regulators are lost. None of them is necessary for a non-motile bacterium. Many elements that are involved in the anchorage of the export apparatus to the cell envelope have also been lost, consistent with the absence of a well-structured outer membrane and cell wall. However, the MS-ring protein (FliF), two out of three components of the C-ring (FliN and FliG), the P-ring protein (FlgI) and L-ring protein (FlgH) are present. Only one protein for the rod formation (FlgF) is preserved. All the retained proteins are homologous to the components of the type III secretion system, supporting the hypothesis that this structure is used by the bacterium as a protein export system to help in the invasion of new host cells[71].

Sitophilus spp. P-endosymbionts, and their close relative the S-symbiont *S. glossindius*, encode presumably functional genes of the type III secretion system (T3SS)[70], an indication that the presence of the secretion system pre-dates the origin of the symbiotic association. The T3SS is a crucial virulence factor of many Gram-negative pathogens of plants and animals, including humans. However, it also seems to be a beneficial factor, enabling the establishment of a mutualistic, intracellular insect–bacteria association. In the *S. glossinidius* genome there are three regions that contain distinct T3SS systems. The best characterized are SSR-1 and SSR-2, which contain genes that are related to genes found in *Salmonella enterica* and *Chromobacterium violaceum*, respectively. Interestingly, both regions have a complete set of genes encoding a functional T3SS export apparatus, but they lack many genes encoding effector proteins in the orthologous pathogenic islands. Moreover, the only putative effector proteins that are preserved are specifically involved in the cytoskeletal rearrangements necessary for bacterial cell invasion[24].

Both parasitic and mutualistic *Wolbachia* lack a T3SS, but they do possess a complete set of genes for another bacterial secretion system, T4SS, which is used by many pathogens to secrete proteins in order to invade host cells[19,57]. *B. aphidicola* and *W. glossinidia* — well-established P-endosymbionts in an advanced stage of genome reduction and with highly simplified cell envelopes — also lack a canonical T3SS, but still retain surface mechanisms for exporting proteins, such as a simplified flagellar apparatus[71] (FIG. 3). Because these structures are quite abundant at the *B. aphidicola* surface — even though these are non-motile bacteria[72] — and only those elements homologous to the T3SS have been retained in all of these reduced genomes[21] it has been proposed that these structures must be involved in invading new bacteriocytes, ovaries and embryos to ensure transmission to host offspring[9].

The immune response and control of bacterial invasion. The molecular mechanisms that are involved in the host immune response to bacterial invasion remain poorly understood[73] but, clearly, hosts have adapted to maintain rather than eliminate endosymbionts. But how does the host immune system perceive these bacteria and control their growth and invasion without complete bacterial clearance?

Hosts have developed molecular systems to recognize conserved microbial cell-envelope motifs (for example, peptidoglycan) through receptors such as those that were identified following the whole-genome sequencing of *D. melanogaster*[74]. Endosymbionts that have an ancient relationship with their hosts have highly simplified cell envelopes and, therefore, these old associations are not good models to study this type of immune control. However, analyses of younger endosymbiotic associations provide insights into the early stages of the interplay between host immunity and bacterial virulence, for example, endosymbionts of weevils from the genus *Sitophilus*. Heddi *et al.* carried out an *in vivo* broad characterization of the transcriptional response of the bacteriocyte to intracellular bacteria, using SZPE as a model organism[75]. They found that the bacteriome expresses a peptidoglycan recognition protein (PGRP) that is homologous to the PGRP-LB protein of *D. melanogaster*, a catalytic member of the PGRP family. The amidase activity of PGRP-LB reduces the biological activity of peptidoglycan, and therefore downregulates the host immune response against Gram-negative bacteria[76]. The specific role of the weevil PRPG is still under investigation, but the fact that it contains the amino-acid residues that are responsible for amidase activity, and that PRPG transcripts only accumulate when endosymbiotic bacteria are present in the bacteriome[77], indicate that it might work in the same way, thus — reducing the host's defence against the endosymbiont to allow the long-term interaction.

Pathogenic island
A part of a genome, for which there is evidence of its acquisition by horizontal transfer, that encodes genes that contribute to the virulence of a pathogen.

Only one sequenced genome is available for a eukaryotic host of an endosymbiont that also has a sequenced genome — the genome of *B. malayi*[78], the nematode host of *Wolbachia w*Bm. Consistent with a mitigated immune response to control rather than eliminate the endosymbiont, the immune system of *B. malayi* has lost the ability to encode the small antibacterial peptides described in other sequenced nematodes[79]. The genome sequence of the aphid *Acyrthosiphon pisum* will soon be available and will provide us with additional clues into how hosts control bacteria in the later stages of the symbiotic association.

One additional insight into the control of endosymbionts by their hosts comes from the finding that endosymbionts that do not reside in host-derived vesicles, but that live free inside the cytosol, have lost the gene *dnaA*[11,13]. This gene encodes a protein that is involved in the initiation of DNA replication, which might indicate that there is a direct control of bacterial growth by the host. In the case of *B. aphidicola*, which does reside in host-derived vesicles, the high level of lysozyme expression detected in bacteriocytes[68,69] might be an important factor involved in the host control of bacterial abundance, as this enzyme catalyses the hydrolysis of peptidoglycan — a system normally used by animals as a constitutive defence against bacterial pathogens.

The evolution of specialized host cells. Using transcriptomic analysis, a pioneering study was carried out into the development and evolution of aphid cells that are destined to become bacteriocytes. Several transcription factors were found to be expressed in these cells early on in development — before colonization by the symbiont population acquired from the mother — following a pattern that has not been described in any other insect cells[80]. Even when bacteria are experimentally removed by antibiotic treatment, bacteriocytes are still specified and maintained. The main conclusion of this work is that bacteriocytes represent an evolutionarily novel cell fate, which is developmentally determined independently of the presence of bacteria. Regarding the role of the bacteriocyte itself in the symbiotic relationship, the identification of a number of bacteriocyte-specific, highly expressed genes in the aphid *A. pisum* by transcriptome analysis[69] revealed that most of them were involved in several functions: non-essential amino-acid metabolism, consistent with this process being among the most important in symbiotic systems; transport, indicating that the bacteriocyte mediates exchange of metabolites and substrates; and production of lysozyme, possibly involved in nutrient uptake by the host, as already mentioned. A similar study that was carried out in *S. zeamais* revealed that the bacteriocyte also allows increased sugar uptake and metabolism, and various anti-stress systems that control the resultant oxidative stress[75].

Evolutionary outcomes of symbioses

Once an association has been established and the genome-reduction process has started, the loss of some metabolic abilities, and other functions, irreversibly ties the intracellular bacteria to their host. This scenario was initially proposed in the 1930s by Lwoff[81] and has been confirmed by genomic studies (see previous section). However, the bacterial genome, no matter how small, must retain the genes that are essential for the maintenance of the symbiotic relationship (for example, essential nutrient provision), and a reduced repertoire of genes for self-maintenance, self-reproduction and evolution. Otherwise, genome reduction might eventually lead these bacteria towards extinction.

In time, a second symbiont can join the consortium. Although initially this new association is likely to be facultative, a new stable relationship can be established, in which case the two bacterial species will co-evolve. As the genome-reductive process continues, and genes that are shared with the second endosymbiont become redundant, two possible outcomes can occur. The actual outcome is a matter of chance, depending on which genome is affected by the loss of genes that are needed for the synthesis of molecules that are essential for fitness.

One possibility is that the presence of a second symbiont might accelerate the degenerative process of the first, leading to its extinction and replacement by the formerly facultative symbiont, which will then become obligatory. This replacement has been reported in symbioses of some weevils[56] and might also be taking place in the case of the *Buchnera–Serratia* consortium that has been established in the aphid *C. cedri*[21]. Unlike other sequenced *B. aphidicola* strains, BCc has partially lost its symbiotic role, as it cannot synthesize tryptophan and riboflavin (see Supplementary information S1(figure)), which need to come from another source, not only for the survival of the host, but also for that of *B. aphidicola*. Genes from *S. symbiotica* SCc that are involved in the biosynthesis of tryptophan have been identified, revealing that this species might be the source of this essential amino acid. The complete sequence of *S. symbiotica* SCc (currently in progress) will tell us whether this bacterium is able to perform all the metabolic functions needed to maintain its host's fitness, or whether it has also lost some pathways. In the first case, extinction of the P-endosymbiont (*Buchnera*) and replacement by *Serratia* would be the most plausible hypothesis.

Alternatively — highlighting the second possibility for the evolutionary outcome of a second symbiont establishing itself in a host — the obligate biochemical interdependence between two endosymbionts can evolutionarily seal the bacterial metabolic complementation, and the establishment of a stable consortium would be the expected evolutionary outcome. This appears to be the case in the sharpshooter *H. coagulata*, with two complementary endosymbionts, *B. cicadellinicola* and *S. muelleri*[15,23]. Whereas *B. cicadellinicola* is mainly devoted to the biosynthesis of vitamins, *S. muelleri* encodes the enzymes that are involved in the biosynthesis of most essential amino acids. The complementarity is striking. For example, *S. muelleri* has lost the pathway for the biosynthesis of histidine, which is the only biosynthetic pathway for essential amino acids retained in *B. cicadellinicola*.

Box 2 | **Minimal cells and synthetic biology**

The phenomenon of genome downsizing that has been observed in endosymbionts, intracellular parasites and organelles has inspired a research programme on minimal life, based on the hypothesis that genomes must retain essential genes that are involved in housekeeping functions, and a minimum number of metabolic transactions for cellular survival and replication. Apart from shedding light on this fundamental topic, the search for minimal genomes is of much value in the context of synthetic biology. In fact, one of the aims of this emerging field is the definition and chemical synthesis of a minimal genome, and its incorporation and expression in a suitable chassis, either derived from a cell (top-down strategy) or starting from a simple chemical system, such as a liposome (bottom-up approach). The top-down approach, also called genome-driven cell engineering[107] or the minimal cell project[108], is expected to provide methods for fabricating engineered cells, which will have a big impact both on biotechnology and on our basic understanding of living systems[109].

The first attempt to define a minimal genome based on comparative genomics used the genomes of two human parasitic bacteria: *Haemophilus influenzae* and *Mycoplasma genitalium*[110]. Owing to their parasitic lifestyles, these two bacteria have reduced genomes when compared with their closest free-living relatives. The analysis led to the proposal of a minimal gene set composed of just 256 genes, most of them involved in genetic-information storage and processing, protein chaperoning and a limited metabolic capability. Later, a combined study of all published research using computational or experimental methods, including the comparison of reduced genomes from insect endosymbionts, was used to define the minimal core of essential genes for a free-living bacterium thriving in a chemically rich environment[111]. The main difference between this study and previous efforts (see REF. 112 and references therein) was the emphasis on the functional completeness of the minimal metabolism encoded by the proposed gene repertoire (involving 62 protein-coding genes out of a minimal set of 208 genes). This aspect has been explored further, demonstrating the stoichiometric consistency of the minimal metabolic network, as well as providing insights into some of its architectural properties, such as its size, clustering and robustness[113].

It must be stressed that these latter studies present just one possible form of a minimal metabolism. Metabolic complexity is ecologically dependent; it is a function of the chemical richness of the environment and the primary energy source(s) available to the living system (FIG. 1). Different versions of minimal gene sets exist, that is, different combinations of metabolic pathways that can perform the essential biological functions of self-maintenance, self-reproduction and evolution[114] under the same external conditions. However, it is possible to define which functions should be performed in any living cell in a specified niche, and list the genes that would be necessary to maintain such functions. In this context, comparative and evolutionary genomics of endosymbionts provide insights into the different routes that endosymbionts have taken to solve the challenges of their own maintenance under the diverse selective pressures and environmental constraints that are the result of different host lifestyles.

Minimal genome
The smallest set of genes that is necessary and sufficient to sustain a living cell in the most favourable conditions; that is, in the presence of adequate nutrients and in the absence of stress factors.

The genome degeneration process seen in endosymbionts can also be accompanied by massive gene transfer to the host cell nucleus[82]. In order for a true cell organelle to be established, several processes must be acquired: new gene-regulatory mechanisms; specific target sequences for proteins that are encoded by the host nucleus but function in the symbiont; and a protein import apparatus that functions at the symbiont surface. These processes occurred in the α-proteobacterial ancestor of mitochondria and in the cyanobacterial ancestor that gave rise to chloroplast lineages. Is there any system that is currently *en route* from endosymbiont to organelle? At present, the smallest sequenced endosymbiont genome belongs to *C. ruddii*, which only codes for approximately 180 proteins. However, a careful analysis revealed that many genes that are involved in essential functions are absent. It remains to be elucidated whether *C. ruddii* is on its way to becoming an organelle, or if it is nearing extinction and replacement by an unidentified symbiont[61]. The transfer of genes from an organelle to the nucleus has been demonstrated under laboratory conditions (see REF. 83 and references therein). There is also evidence of gene transfer from *Wolbachia* to the nuclear genome of several insect and nematode hosts[84]. These remarkable cases of lateral gene transfer should be a caution for those carrying out eukaryotic genome sequencing, and should add new insights of relevance to the debate about the differences between cell organelles and endosymbionts[85].

Conclusions and future directions

The advent of genomics has significantly contributed to reinforcing the view of symbiosis as a widespread biological phenomenon, especially through the genome sequencing of uncultivable organisms. Symbiosis is a dynamic process in which the prokaryotic symbiont, while providing the host with new metabolic capabilities, experiences many genotypic and phenotypic changes compared with its free-living relatives in order to adapt to its new lifestyle, as revealed by the specific differences observed in the completely sequenced genomes of several endosymbiotic bacteria.

Many questions remain open for future research. On the host side, a better understanding of how these organisms control and/or domesticate prokaryotic symbionts is needed. Genomic and transcriptomic studies of model eukaryotic hosts that have well known and stable symbiotic associations will be valuable. Until now, only the genome of *B. malayi*, the nematode host of *W. pipiensis w*Bm, has been fully sequenced[78], but the sequence of the aphid *A. pisum* is on its way. Comparative analysis with model organisms that have already been sequenced, such as the nematode *Caenorhabditis elegans*[86] and the arthropod *D. melanogaster*[87], will give new insights into the changes that are experienced by the host after the establishment of the symbiosis. On the bacterial side, comparative studies of the molecular processes governing the nature of symbiotic associations, either mutualistic or parasitic, will provide important clues to understand why microorganisms evolve toward one lifestyle or the other.

Symbiosis is not always a matter of two organisms: we are gaining increasing evidence that well-established prokaryotic endosymbionts also maintain some metabolic interplay with new facultative symbionts. Such relationships could finish with the extinction of the P-endosymbiont (as described in insects from the family Dryopthoridae, to which *Sitophilus* spp. belongs[56], or the *Buchnera–Serratia* consortium in cedar aphids[21]), or the symbiotic consortia could continue over time (as described in the *Baumannia–Sulcia* consortium in sharpshooters[15]). Metagenomics and systems biology are new approaches that can be used in the analysis of complex symbiotic consortia. For example, the metagenomes of sponge microbial communities have been shown to contain genes and gene clusters for the biosynthesis of biologically active natural products. There is evidence to suggest that a mutually beneficial

Synthetic biology
The design and fabrication of artificial biological systems, with the aim of either optimizing their performance in the context of their technological utility or of deepening our understanding of the naturally occurring organisms.

relationship exists, at least between some of the bacteria and the sponges themselves[88]. Another interesting example is the symbiotic gut microbiome of mammals, in which we are starting to have a clearer picture of the role of the symbiotic microbial consortia in the metabolic phenotype of the mammalian host[89]. In this case, an extensive microbial–mammalian co-metabolism exists. As new complete genomes of eukaryote hosts become available, we will be able to make new metabolic inferences to understand the interplay between the host and its associated microbiota. Genomics and metagenomics are becoming essential tools to provide new insights into symbiotic associations, including those that are highly relevant to biotechnology and biomedicine.

Finally, synthetic biology will benefit from the new insights gained by comparative genomic analyses of extremely reduced genomes (BOX 2). This type of research makes important contributions to the conceptually intriguing and biologically fundamental question of the minimal genomic basis for sustaining cellular life. Many efforts in this area have attracted a great deal of public attention regarding both biosecurity risks and potential commercialization.

1. Hypsa, V. & Dale, C. *In vitro* culture and phylogenetic analysis of '*Candidatus Arsenophonus triatominarum*', an intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int. J. Syst. Bacteriol.* **47**, 1140–1144 (1997).
2. Dale, C. & Maudlin, I. *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *Int. J. Syst. Bacteriol.* **49**, 267–275 (1999).
3. Dale, C., Beeton, M., Harbison, C., Jones, T. & Pontes, M. Isolation, pure culture, and characterization of '*Candidatus Arsenophonus arthropodicus*', an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. *Appl. Environ. Microbiol.* **72**, 2997–3004 (2006).
4. Darby, A. C., Chandler, S. M., Welburn, S. C. & Douglas, A. E. Aphid-symbiotic bacteria cultured in insect cell lines. *Appl. Environ. Microbiol.* **71**, 4833–4839 (2005).
5. Sagan, L. On the origin of mitosing cells. *J. Theor. Biol.* **14**, 255–274 (1967).
6. Sapp, J. Evolution by Association. A History of Symbiosis (Oxford University Press, New York, 1994). **This book is the first historical report on the symbiosis concept and its central role in shaping contemporary evolutionary ideas about the origin and diversification of life.**
7. de Duve, C. The origin of eukaryotes: a reappraisal. *Nature Rev. Genet.* **8**, 395–403 (2007).
8. Margulis, L. Symbiosis in Cell Evolution. Microbial Communities in the Archaean and Proterozoic Eons (W. H. Freeman and Co., New York, 1993).
9. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407**, 81–86 (2000).
10. Tamas, I. *et al.* 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**, 2376–2379 (2002).
11. Akman, L. *et al.* Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature Genet.* **32**, 402–407 (2002).
12. van Ham, R. C. *et al.* Reductive genome evolution in *Buchnera aphidicola*. *Proc. Natl Acad. Sci. USA* **100**, 581–586 (2003).
13. Gil, R. *et al.* The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proc. Natl Acad. Sci. USA* **100**, 9388–9393 (2003).
14. Degnan, P. H., Lazarus, A. B. & Wernegreen, J. J. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. *Genome Res.* **15**, 1023–1033 (2005).
15. Wu, D. *et al.* Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* **4**, e188 (2006).
16. Nakabachi, A. *et al.* The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314**, 267 (2006).
17. Newton, I. L. *et al.* The *Calyptogena magnifica* chemoautotrophic symbiont genome. *Science* **315**, 998–1000 (2007).
18. Kuwahara, H. *et al.* Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, *Calyptogena okutanii*. *Curr. Biol.* **17**, 881–886 (2007).
19. Foster, J. *et al.* The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* **3**, e121 (2005).
20. Nakagawa, S. *et al.* Deep-sea vent ε-proteobacterial genomes provide insights into emergence of pathogens. *Proc. Natl Acad. Sci. USA* **104**, 12146–12150 (2007).
21. Pérez-Brocal, V. *et al.* A small microbial genome: the end of a long symbiotic relationship? *Science* **314**, 312–313 (2006). **This paper, together with reference 16, shows that the genomes of *B. aphidicola* BCc and *C. ruddii* represent the lower limits of genome size in insect endosymbionts. Further research in these systems might shed light on the evolutionary fate of endosymbionts.**
22. Toh, H. *et al.* Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res.* **16**, 149–156 (2006).
23. McCutcheon, J. P. & Moran, N. A. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl Acad. Sci. USA* **104**, 19392–19397 (2007).
24. Dale, C. & Moran, N. A. Molecular interactions between bacterial symbionts and their hosts. *Cell* **126**, 453–465 (2006).
25. Baumann, P. Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.* **59**, 155–189 (2005).
26. Russell, J. A., Latorre, A., Sabater-Munoz, B., Moya, A. & Moran, N. A. Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the *Aphidoidea*. *Mol. Ecol.* **12**, 1061–1075 (2003).
27. Chen, D.-Q., Montllor, C. B. & Purcell, A. H. 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 (2000).
28. Montllor, C. B., Maxmen, A. & Purcell, A. H. Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress. *Ecol. Entomol.* **27**, 189–195 (2002).
29. Russell, J. A. & Moran, N. A. Costs and benefits of symbiont infection in aphids: variation among symbionts and across temperatures. *Proc. Biol. Sci.* **273**, 603–610 (2006).
30. Oliver, K. M., Moran, N. A. & Hunter, M. S. Variation in resistance to parasitism in aphids is due to symbionts not host genotype. *Proc. Natl Acad. Sci. USA* **102**, 12795–12800 (2005).
31. Oliver, K. M., Russell, J. A., Moran, N. A. & Hunter, M. S. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proc. Natl Acad. Sci. USA* **100**, 1803–1807 (2003).
32. Scarborough, C. L., Ferrari, J. & Godfray, H. C. Aphid protected from pathogen by endosymbiont. *Science* **310**, 1781 (2005).
33. Tsuchida, T., Koga, R. & Fukatsu, T. Host plant specialization governed by facultative symbiont. *Science* **303**, 1989 (2004).
34. Simon, J. C. *et al.* 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 (2003).
35. Leonardo, T. E. & Muiru, G. T. Facultative symbionts are associated with host plant specialization in pea aphid populations. *Proc. Biol. Sci.* **270** (Suppl 2), S209–S212 (2003).
36. Leonardo, T. E. & Mondor, E. B. Symbiont modifies host life-history traits that affect gene flow. *Proc. Biol. Sci.* **273**, 1079–1084 (2006).
37. McGraw, E. A. & O'Neill, S. L. *Wolbachia pipientis*: intracellular infection and pathogenesis in *Drosophila*. *Curr. Opin. Microbiol.* **7**, 67–70 (2004).
38. Welburn, S. C. & Maudlin, I. Tsetse–trypanosome interactions: rites of passage. *Parasitol. Today* **15**, 399–403 (1999).
39. Moran, N. A. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc. Natl Acad. Sci. USA* **104** (Suppl 1), 8627–8633 (2007).
40. Gomez-Valero, L. *et al.* Coexistence of *Wolbachia* with *Buchnera aphidicola* and a secondary symbiont in the aphid *Cinara cedri*. *J. Bacteriol.* **186**, 6626–6633 (2004).
41. Takiya, D. M., Tran, P. L., Dietrich, C. H. & Moran, N. A. Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. *Mol. Ecol.* **15**, 4175–4191 (2006).
42. Woyke, T. *et al.* Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**, 950–955 (2006). **This paper and reference 15 are two pioneering works on the application of genomic analysis to uncultivable microbial consortia, which opened up a new field of research on symbiosis.**
43. Giere, O. & Erséus, C. Taxonomy and new bacterial symbioses of gutless marine Tubificidae (Annelida, Oligochaeta) from the island of Elba (Italy). *Org. Divers. Evol.* **2**, 289–297 (2002).
44. Wernegreen, J. J. Genome evolution in bacterial endosymbionts of insects. *Nature Rev. Genet.* **3**, 850–861 (2002).
45. Moran, N. A. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA* **93**, 2873–2878 (1996).
46. Itoh, T., Martin, W. & Nei, M. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc. Natl Acad. Sci. USA* **99**, 12944–12948 (2002).
47. Wernegreen, J. J. For better or worse: genomic consequences of intracellular mutualism and parasitism. *Curr. Opin. Genet. Dev.* **15**, 572–583 (2005).
48. Rocha, E. P. & Danchin, A. Base composition bias might result from competition for metabolic resources. *Trends Genet.* **18**, 291–294 (2002).
49. Fares, M. A., Ruiz-Gonzalez, M. X., Moya, A., Elena, S. F. & Barrio, E. Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* **417**, 398 (2002).
50. Moya, A., Latorre, A., Sabater-Munoz, B. & Silva, F. J. Comparative molecular evolution of primary (*Buchnera*) and secondary symbionts of aphids based on two protein-coding genes. *J. Mol. Evol.* **55**, 127–137 (2002).
51. Rispe, C., Delmotte, F., van Ham, R. C. & Moya, A. Mutational and selective pressures on codon and amino acid usage in *Buchnera*, endosymbiotic bacteria of aphids. *Genome Res.* **14**, 44–53 (2004).
52. Moran, N. A. & Mira, A. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* **2**, RESEARCH0054 (2001).
53. Moran, N. A. & Plague, G. R. Genomic changes following host restriction in bacteria. *Curr. Opin. Genet. Dev.* **14**, 627–633 (2004).
54. Plague, G. R., Dunbar, H. E., Tran, P. L. & Moran, N. A. Extensive proliferation of transposable elements in heritable bacterial symbionts. *J. Bacteriol.* **190**, 777–779 (2007).
55. Heddi, A., Charles, H., Khatchadourian, C., Bonnot, G. & Nardon, P. 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 (1998).

56. Lefevre, C. *et al.* Endosymbiont phylogenesis in the Dryophthoridae weevils: evidence for bacterial replacement. *Mol. Biol. Evol.* **21**, 965–973 (2004).
57. Wu, M. *et al.* Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**, e69 (2004).
58. Silva, F. J., Latorre, A. & Moya, A. Genome size reduction through multiple events of gene disintegration in *Buchnera* APS. *Trends Genet.* **17**, 615–618 (2001).
59. Gomez-Valero, L., Silva, F. J., Simon, J. C. & Latorre, A. Genome reduction of the aphid endosymbiont *Buchnera aphidicola* in a recent evolutionary time scale. *Gene* **389**, 87–95 (2007).
60. Gomez-Valero, L., Latorre, A. & Silva, F. J. The evolutionary fate of nonfunctional DNA in the bacterial endosymbiont *Buchnera aphidicola*. *Mol. Biol. Evol.* **21**, 2172–2181 (2004).
61. Tamames, J. *et al.* The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. *BMC Evol. Biol.* **7**, 181 (2007).
62. *Wilson*, A. C. *et al.* A dual-genome microarray for the pea aphid, *Acyrthosiphon pisum*, and its obligate bacterial symbiont, *Buchnera aphidicola*. *BMC Genomics* **7**, 50 (2006).
63. Wilcox, J. L., Dunbar, H. E., Wolfinger, R. D. & Moran, N. A. Consequences of reductive evolution for gene expression in an obligate endosymbiont. *Mol. Microbiol.* **48**, 1491–1500 (2003).
64. Koga, R., Tsuchida, T. & Fukatsu, T. 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 (2003).
65. Ren, Q. & Paulsen, I. T. Comparative analyses of fundamental differences in membrane transport capabilities in prokaryotes and eukaryotes. *PLoS Comput. Biol.* **1**, e27 (2005).
66. Fiala-Médioni, A. & Métivier, C. Ultrastructure of the gill of the hydrothermal vent bivalve *Calyptogena magnifica*, with a discussion of its nutrition. *Mar. Biol.* **90**, 215–222 (1986).
67. Baumann, P., Moran, N. A. & Baumann, L. in *The Prokaryotes* (ed. Dworkin, M.) 1–67 (Springer, New York, 2000).
68. Fiala-Médioni, A., Michalski, J. C., Jollès, J., Alonso, C. & Montreuil, J. Lysosomic and lysozyme activities in gills of bivalves from deep hydrothermal vents. *C. R. Acad. Sci. Paris* **317**, 239–244 (1994).
69. Nakabachi, A. *et al.* Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera*. *Proc. Natl Acad. Sci. USA* **102**, 5477–5482 (2005).
70. Dale, C., Young, S. A., Haydon, D. T. & Welburn, S. C. The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. *Proc. Natl Acad. Sci. USA* **98**, 1883–1888 (2001).
71. Young, G. M., Schmiel, D. H. & Miller, V. L. A new pathway for the secretion of virulence factors by bacteria: The flagellar export apparatus functions as a protein-secretion system. *Proc. Natl Acad. Sci. USA* **96**, 6456–6461 (1999).
72. Maezawa, K. *et al.* Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium *Buchnera aphidicola* sp. strain APS. *J. Bacteriol.* **188**, 6539–6543 (2006).
73. McFall-Ngai, M. Adaptive immunity: care for the community. *Nature* **445**, 153 (2007).
74. Hoffmann, J. A. The immune response of *Drosophila*. *Nature* **426**, 33–38 (2003).
75. Heddi, A. *et al.* Molecular and cellular profiles of insect bacteriocytes: mutualism and harm at the initial evolutionary step of symbiogenesis. *Cell. Microbiol.* **7**, 293–305 (2005).
76. Zaidman-Remy, A. *et al.* The *Drosophila* amidase PGRP-LB modulates the immune response to bacterial infection. *Immunity* **24**, 463–473 (2006).
77. Anselme, C., Vallier, A., Balmand, S., Fauvarque, M.-O. & Heddi, A. Host PGRP gene expression and bacterial release in endosymbiosis of the weevil *Sitophilus zeamais*. *Appl. Environ. Microbiol.* **72**, 6766–6772 (2006).
78. Ghedin, E. *et al.* Draft genome of the filarial nematode parasite *Brugia malayi*. *Science* **317**, 1756–1760 (2007).
The complete genome sequence of the nematode *B. malayi*, together with the previously reported genome sequence of its *Wolbachia* endosymbiont, will enable a systems approach for the study of host–symbiont interactions.
79. Kato, Y. & Komatsu, S. ASABF, a novel cysteine-rich antibacterial peptide isolated from the nematode *Ascaris suum*. Purification, primary structure, and molecular cloning of cDNA. *J. Biol. Chem.* **271**, 30493–30498 (1996).
80. Braendle, C. *et al.* Developmental origin and evolution of bacteriocytes in the aphid–*Buchnera* symbiosis. *PLoS Biol.* **1**, e21 (2003).
This paper was the first study of the development and evolution of a bacteriocyte containing endosymbiotic bacteria.
81. Lwoff, A. L'évolution physiologique. Étude des pertes des fonctions chez les microorganismes (Hermann, Paris, 1944).
82. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nature Rev. Genet.* **5**, 123–135 (2004).
83. Martin, W. Gene transfer from organelles to the nucleus: frequent and in big chunks. *Proc. Natl Acad. Sci. USA* **100**, 8612–8614 (2003).
84. Hotopp, J. C. *et al.* Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **317**, 1753–1756 (2007).
85. Bhattacharya, D., Archibald, J. M., Weber, A. P. M. & Reyes-Prieto, A. How do endosymbionts become organelles? Understanding early events in plastid evolution. *Bioessays* **29**, 1239–1246 (2007).
86. *C. elegans* Sequencing Consortium. Genome sequence of the nematode *Caenorhabditis elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
87. Myers, E. W. *et al.* A whole-genome assembly of *Drosophila*. *Science* **287**, 2196–2204 (2000).
88. Kennedy, J., Marchesi, J. R. & Dobson, A. D. Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Appl. Microbiol. Biotechnol.* **75**, 11–20 (2007).
89. Martin, F. P. *et al.* A top-down systems biology view of microbiome–mammalian metabolic interactions in a mouse model. *Mol. Syst. Biol.* **3**, 112 (2007).
90. Paracer, S. & Ahmadjian, V. Symbiosis. An Introduction to Biological Associations (Oxford University Press, New York, 2000).
91. Kneip, C., Lockhart, P., Voss, C. & Maier, U. G. Nitrogen fixation in eukaryotes — new models for symbiosis. *BMC Evol. Biol.* **7**, 55 (2007).
92. Schink, B. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* **61**, 262–280 (1997).
93. Stewart, F. J., Newton, I. L. & Cavanaugh, C. M. Chemosynthetic endosymbioses: adaptations to oxic–anoxic interfaces. *Trends Microbiol.* **13**, 439–448 (2005).
94. Minic, Z. & Herve, G. Biochemical and enzymological aspects of the symbiosis between the deep-sea tubeworm *Riftia pachyptila* and its bacterial endosymbiont. *Eur. J. Biochem.* **271**, 3093–3102 (2004).
95. Zientz, E., Dandekar, T. & Gross, R. Metabolic interdependence of obligate intracellular bacteria and their insect hosts. *Microbiol. Mol. Biol. Rev.* **68**, 745–770 (2004).
96. Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J. & Smith, A. G. Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* **438**, 90–93 (2005).
97. Buchner, P. Endosymbiosis of Animals with Plant Microorganisms (Interscience, New York, 1965).
98. Taylor, M. W., Radax, R., Steger, D. & Wagner, M. Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.* **71**, 295–347 (2007).
99. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host–bacterial mutualism in the human intestine. *Science* **307**, 1915–1920 (2005).
100. Stingl, U., Radek, R., Yang, H. & Brune, A. 'Endomicrobia': cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* **71**, 1473–1479 (2005).
101. Fieseler, L., Quaiser, A., Schleper, C. & Hentschel, U. Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum Poribacteria by environmental genomics. *Environ. Microbiol.* **8**, 612–624 (2006).
102. Moreira, D. *et al.* Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Mol. Phylogenet. Evol.* **44**, 255–266 (2007).
103. Rodriguez-Ezpeleta, N. *et al.* Toward resolving the eukaryotic tree: the phylogenetic positions of jakobids and cercozoans. *Curr. Biol.* **17**, 1420–1425 (2007).
104. Lecointre, G. & Le Guyader, H. The Tree of Life. A Phylogenetic Classification (Harvard University Press Reference Library) (Belknap Press, Cambridge, 2006).
105. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
106. Bourlat, S. J. *et al.* Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006).
107. O'Malley, M., Powell, A., Davies, J. & Calvert, J. Knowledge-making distinctions in synthetic biology. *Bioessays* **30**, 57–65 (2007).
108. Luisi, P. L. Chemical aspects of synthetic biology. *Chem. Biodivers.* **4**, 603–621 (2007).
109. Peretó, J. & Català, J. The renaissance of synthetic biology. *Biol. Theor.* **2**, 128–130 (2007).
110. Mushegian, A. R. & Koonin, E. V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl Acad. Sci. USA* **93**, 10268–10273 (1996).
111. Gil, R., Silva, F. J., Pereto, J. & Moya, A. Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**, 518–537 (2004).
This Review proposes a minimal genetic repertoire for a hypothetical heterotrophic bacterium, based on comparative genomics of endosymbionts, experimentally determined gene essentiality and the functions of a coherent metabolism.
112. Klasson, L. & Andersson, S. G. Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* **12**, 37–43 (2004).
113. Gabaldon, T. *et al.* Structural analyses of a hypothetical minimal metabolism. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **362**, 1751–1762 (2007).
114. Ruiz-Mirazo, K., Pereto, J. & Moreno, A. A universal definition of life: autonomy and open-ended evolution. *Orig. Life Evol. Biosph.* **34**, 323–346 (2004).

Acknowledgements
Financial support was provided by grants GV/2007/050 from Generalitat Valenciana, Spain, to R.G and BFU2006-06003 from Ministerio de Educación y Ciencia (MEC), Spain, to A.L. R.G is a recipient of a contract in the 'Ramón y Cajal' programme from the MEC, Spain. Our thanks to H. Escrivá (CNRS, Banyuls sur Mer), P. López (CNRS, Orsay) and D. Moreira (CNRS, Orsay) for their advice in the design of the eukaryotic phylogenetic tree in BOX 1.

DATABASES
Entrez Genome Project:
http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj
Acyrthosiphon pisum | *Baumannia cicadellinicola* | *Blochmannia floridanus* | *Brugia malayi* | *Buchnera aphidicola* | *Caenorhabditis elegans* | *Carsonella ruddii* | *Chromobacterium violaceum* | *Drosophila melanogaster* | *Sodalis glossinidius* | *Sulcia muelleri* | *Wigglesworthia glossinidia* | *Wolbachia pipientis*

FURTHER INFORMATION
Andrés Moya's homepage: www.uv.es/cavanilles/personal/Andres_Moya
Juli Peretó's homepage: http://www.uv.es/cavanilles/personal/Juli_Pereto
Rosario Gil's homepage: www.uv.es/cavanilles/personal/Rosario_Gil
Amparo Latorre's homepage: http://www.uv.es/cavanilles/personal/Amparo_Latorre/amparo_latorre.htm
Cavanilles Institute web page: www.uv.es/cavanilles
Genomes OnLine Database: http://www.genomesonline.org
Microbial Genome Database for Comparative Analysis: http://mbgd.genome.ad.jp
The Sanger Institute *Glossina* Genome Project: http://www.sanger.ac.uk/Projects/G_morsitans

SUPPLEMENTARY INFORMATION
See online article: S1 (figure) | S2 (table)
ALL LINKS ARE ACTIVE IN THE ONLINE PDF