University of South Bohemia in České Budějovice

Faculty of Science

Evolutionary origins of intracellular symbionts in arthropods

Master thesis

Bc. Filip Husník

Supervisor: Prof. RNDr. Václav Hypša, Csc.

České Budějovice 2012

Husník, F., 2012: Evolutionary origins of intracellular symbionts in arthropods. Mgr. Thesis, in English. - 34 p. (+ 18 p. suppl. - research article), Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:
Intracellular symbionts are widespread among arthropods, particularly within insects. Obligate symbiotic associations are known to have originated multiple times between the arthropods feeding on nutrient-poor diets and bacteria from various groups. However, exact phylogenetic positions and relationships among these symbiotic lineages are mostly unclear or vague. This thesis consists of an exemplary case study on the most symbiont-rich bacterial group, Enterobacteriaceae, already published in BMC Biology. It uses advanced phylogenetic tools and extended taxonomic sample to establish phylogenetic relationships among individual symbiotic lineages and their phylogenetic affinity to free-living relatives. To provide it with broader background, the publication is accompanied by a review on general evolutionary forces influencing origin and maintenance of intracellular symbiosis in arthropods. Apart from overviewing the current known diversity of the symbiotic bacteria, it also points out specific drawbacks in inferring symbionts phylogeny and consequences that can phylogeny have on our understanding of intracellular symbiosis.

Prohlašuji, že svoji diplomovou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 26. dubna 2012

Filip Husník

Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches

Filip Husník*, Tomáš Chrudimský and Václav Hypša*
BMC Biology 2011, 9:87, doi:10.1186/1741-7007-9-87
http://www.biomedcentral.com/1741-7007/9/87

Additional data:
http://www.biomedcentral.com/1741-7007/9/87/additional
http://purl.org/phylo/treebase/phylows/study/TB2:S11451
http://users.prf.jcu.cz/husnif00

Authors' contributions:
FH: ~70%
TCH: ~5%
VH: ~25%

FH carried out the sequence alignments and phylogenetic analyses, and participated in the study design, evolutionary interpretation of the results and preparation of the manuscript. TCH compiled and analyzed the AT/GC reduced matrices. VH conceived of the study and participated in its design, evolutionary interpretation of the results and preparation of the manuscript. All authors read and approved the final manuscript.

Acknowledgements:
It is a pleasure to thank those who made this thesis possible. Firstly, I would like to thank my supervisor Václav Hypša for helpful suggestions and critical comments during my master's study. Secondly, I am indebted to my many of my friends and colleagues from Laboratory of Molecular Phylogeny and Evolution of Parasites (University of South Bohemia) and John McCutcheon's Microbial Genomics and Symbiosis lab (University of Montana). Last but not least, I thank my friends and family for supporting me.

Contents

1. Background

1.1 Evolutionary origins of intracellular symbionts in arthropods

1.1.1 Symbiosis as an evolutionary innovation

Symbiotic lifestyle is an important source of evolutionary innovations which gave rise for example to the origin of eukaryotic cell. In a human body, bacterial cells outnumber the host cells and taken together would make an organ larger than a liver [1, 2]. Therefore, it is no surprise that the most species-rich group on Earth, arthropods, has evolved numerous symbiotic associations with not only bacteria, but also fungi and various unicellular eukaryotes, and the symbiotic habit assisted this group in its extreme diversification.

The core of this MSc. thesis is a published study [3] on the richest source of insect bacterial symbionts, gammaproteobacterial family Enterobacteriaceae. In this introduction, I provide a broader context to the published results, focusing mainly on the evolution and origins of intracellular symbiosis between bacteria and arthropods and trying to highlight neglected or uncertain parts of its research. It will deal with the following questions of the evolutionary history of intracellular symbioses. How frequently have intracellular symbioses originated among different groups of bacteria and arthropods? Are some taxonomical or ecological groups predisposed to form intracellular symbiotic associations? How common is transition between pathogenic and symbiotic lifestyle or vice versa? Which changes affect symbiotic associations and how common are losses, replacements or complementations of established symbionts. How intracellular lifestyle adjusts genomes, transcriptomes or proteomes of both symbiotic partners? Are symbionts with extremely reduced genomes bacteria, organelles or something in between?

1.1.2 Multiplication of languages: obligate vs. facultative symbionts

Symbiotic bacteria of arthropods are usually assigned to two main ecological categories called primary/obligate (P) and secondary/facultative (S) symbionts. P-symbionts are obligate mutualists inherited maternally by vertical transmission. They are harbored in specialized cells called bacteriocytes that can form an organ called bacteriome (older terms: mycetocytes, mycetome) and provide their hosts with compounds unavailable from their unbalanced diet or

recycle waste products. Typical hosts of P-symbionts are thus phloem/xylem sap sucking or blood-sucking arthropods. Inevitable consequences of this relationship are that P-symbionts co-speciate with their hosts for millions of years and are highly adapted to the intracellular environment, so that they cannot survive outside their host and the host cannot survive or reproduce without them [4-9].

In comparison to P-symbionts, S-symbionts is a heterogeneous assemblage of arthropods-associated bacteria including facultative commensals, facultative mutualists and sometimes even bacteria with negative effects on its host, such as reproductive manipulators. Traits typical for these bacteria are that they are not necessary for the host survival [10] and that they are usually present in nonspecialized cells and tissues both intracellularly and extracellularly. Unlike P-symbionts, their characteristics allow them to be also horizontally transferred among different arthropod groups [11-17]. Although some S-symbionts are cultivable in axenic culture [18-23], there is currently no study that would confirm that S-symbionts have life phase outside of arthropod hosts, but several possible arthropod-to-arthropod transmission hypotheses have been suggested. These hypotheses include e.g. sexual transmission, transmission through parasites or parasitoids, co-feeding on an identical plant/host, feces contamination and hemolymph sucking during phoresis (e.g. chewing lice or mites). Mutualistic phenotype of S-symbionts is commonly involved in protection against parasitoids, pathogens, RNA viruses, heat stress or provision of compounds not available from the P-symbiont [17, 24-32].

Unfortunately, research communities working on eukaryotic organelles and arthropod symbioses do not share reviews, conferences or terminology, which sometimes leads to misunderstandings mostly due to different usage of terms primary and secondary symbiosis. To avoid these misunderstandings, I will hereafter either substituted these terms by more general terms *obligate* and *facultative* or use well-recognized abbreviated form as P/S-symbionts.

According to the current rules of bacterial nomenclature, description of a new species requires an *in vitro* culture [33, 34] to accept the species as valid. Since most of insect symbionts are uncultivable, they are commonly named under provisional *Candidatus* status. Considering that complete genomes (as much richer source of information about organism' biology than cultivation can ever provide) are available for many of these bacteria, and to simplify the text, I intentionally omit the *Candidatus* status in the following text.

1.1.3 Co-symbioses, transitions, losses and replacements

Because of low effect of immunity in symbiotic tissue, facultative bacteria are commonly found within bacteriocytes of obligate symbionts, within sheath cells close to them or within so-called secondary bacteriocytes. Therefore, if a loss or degradation of an essential metabolic pathway from the obligate symbiont occurs, these facultative symbionts can complement the pathway, provide intermediate products or even cooperate with the obligate symbiont in a step-by-step interdependent biosynthetic patchwork [35, 36]. This cooperation can eventually lead to a situation when the originally facultative bacterium losses genes needed for facultative lifestyle, becomes dependent on the host and either completely replaces the original obligate symbiont or turns into an obligate co-symbiont.

Based on modeling of genome size decrease in the course of evolution, the obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*, was suggested to have evolved from such a facultative bacterium [37], possibly through symbiotic replacement. Complete replacement of the obligate symbiont in aphids, *Buchnera aphidicola*, was proved experimentally by facultative *Serratia* bacteria [38] and obligate co-symbiosis of exactly the same partners was recently confirmed in *Cinara cedri* aphid [35, 39]. Other well-known complete replacements are known from weevils, where ancient *Nardonella* sp. symbiont was replaced by a *Sodalis* lineage in grain-feeding *Sitophilus* lineage [40, 41] or from mealybugs where ancient *Tremblaya* linege was replaced by Bacteroidetes bacteria in Rhizoecini and *Cryptococcus*/*Rastrococcus* lineages [42]. Similar scenario can involve one or more replacements applied on several other cases of ancient obligate co-symbioses such as those in Auchenorrhyncha [43-49] or mealybugs [15, 42], although it is mostly unknown what the original phenotypes of additional symbiotic partners were.

1.1.4 Arthropods as hosts for intracellular symbionts

Bacterial symbionts have colonized various niches within arthropod hosts and symbiotic organs originated convergently multiple times in various arthropod groups [4, 5, 47, 50]. Three typical localizations of symbiotic organs can be distinguished: 1, bacteriocytes or bacteriome(s) localized freely in haemocoel (e.g. in sap-sucking insects); 2, a specialized segment of gut (Fig. 1A, B), gut caeca and capsules or malphigic tubules (e.g. in some blood-sucking insects, true

bugs, beetles); 3, bacteriocytes or bacteriome(s) in fat body (e.g. in cockroaches and ants). In most cases, symbiotic tissue is surrounded by rich tracheal system to transport gases from and to this metabolically highly active tissue (Fig. 1B).

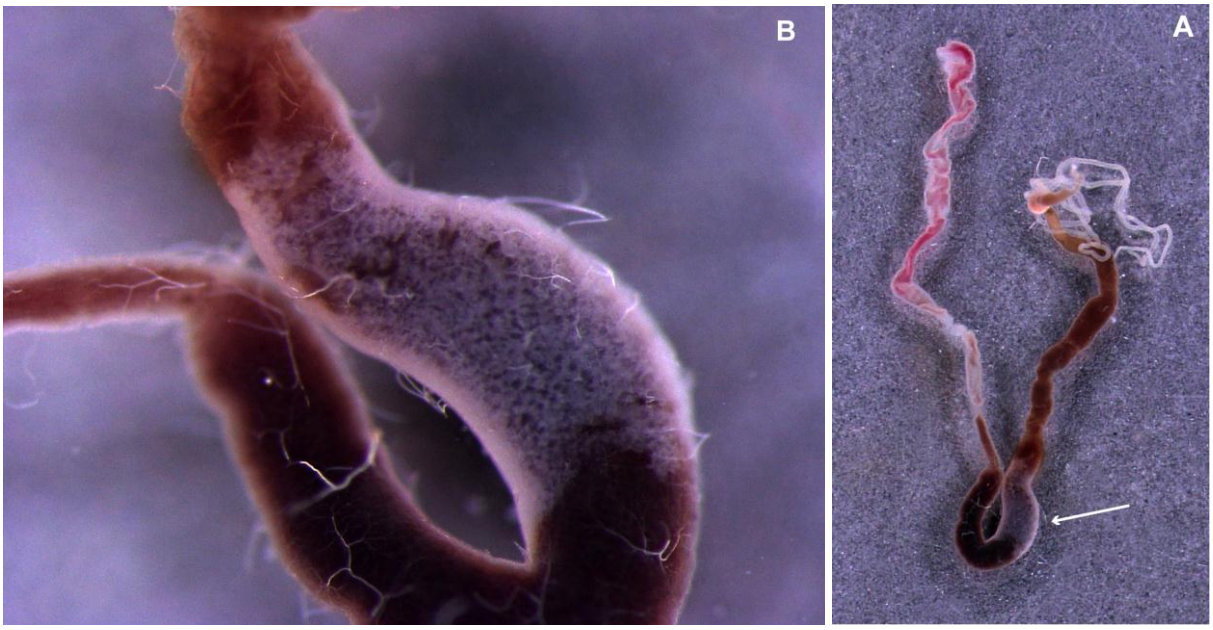


Figure 1. Dissected gut (A) of blood-sucking fly *Melophagus ovinus* (Diptera: Hippoboscidae) showing midgut section (bacteriome) with enlarged cells (bacteriocytes) harboring obligate symbiotic bacteria (B).

Not only localization within the host body, but also localization within the host cells is remarkably variable. Bacteria can be localized freely within the cell cytoplasm or surrounded by host-derived symbiosomal membrane. Several symbiotic bacteria were shown to share single host cell [51] and there is even a case of intrabacterial symbionts localized within another intracellular bacterium [52]. Moreover, symbiotic bacteria can also be localized within various cell structures and organelles such as nucleus [53], mitochondrion [54], Golgi apparatus and endoplasmic reticulum [55, 56].

Maternal transmission of obligate symbionts to offspring is certainly one of the least known phases in development of symbiotic bacteria in arthropods. Three different general routes of transmission are recognized [4]. The first is based on external smearing of eggs with symbionts and ingestion of symbionts during hatching of larvae. This mode of transmission is typical for beetles and some true bugs. The second and the most common route of transmission is transfer of bacteriocytes or bacteria (or active migration of bacteria) to the ovary and incorporation into the oocytes. The last route of transmission is present in viviparous

Hippoboscoidea (tsetse flies, louse flies and bat flies), which exploits for the symbiont transfer milk glands nourishing the evolving larva (Fig 2A, B, C). In case of active migration, bacteria can use flagellum [57], but mode of transfer for bacteria without a flagellum (such as *Riesia pediculicola* in lice) remains a mystery.

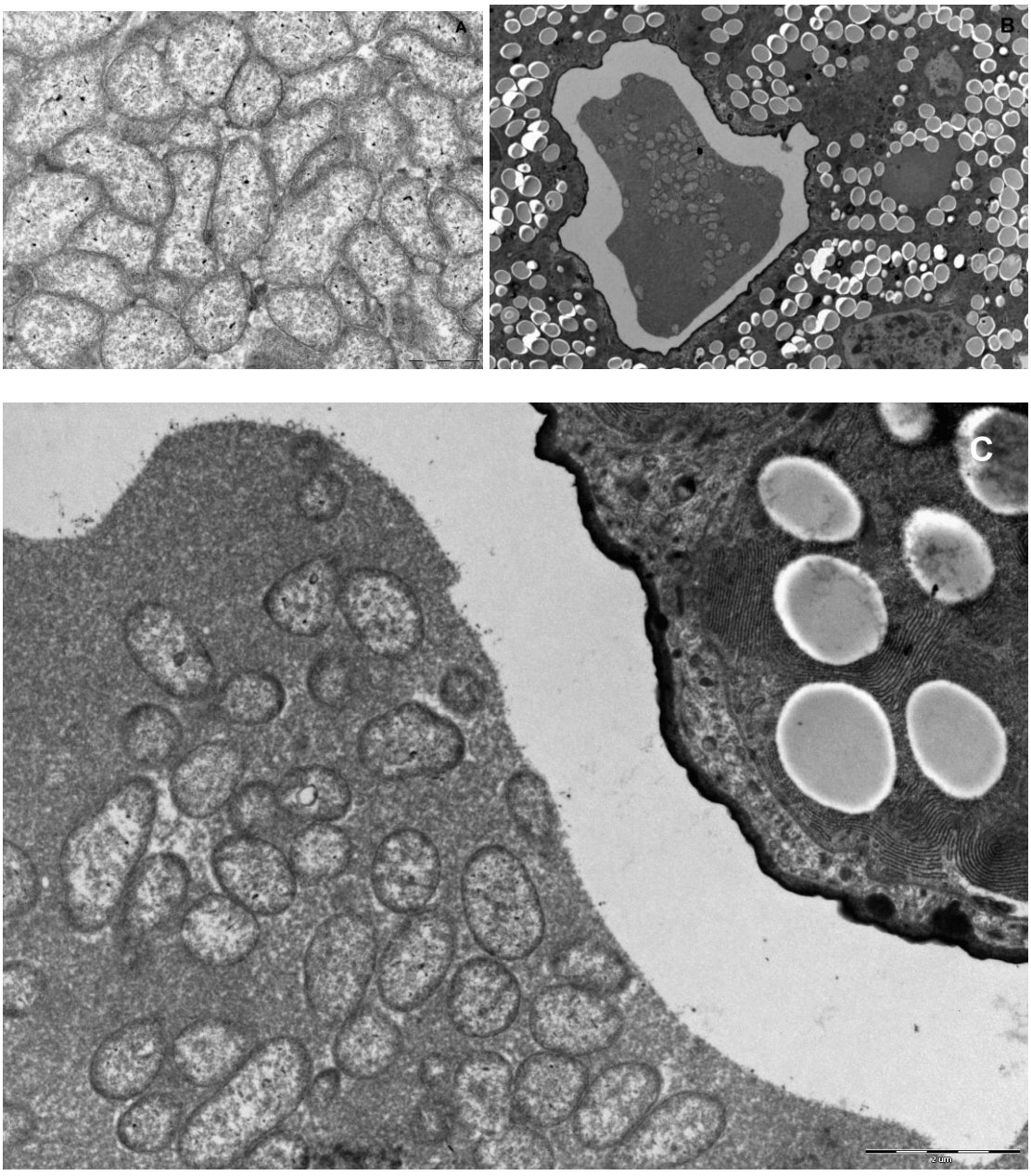


Figure 2. Transmission electron micrographs of *Melophagus ovinus* bacteriocytes from the bacteriome showing typical spherical shape of P-symbiont cells (A). Milk gland section showing vertical transmission of obligate symbiotic bacteria *Arsenophonus melophagi* through milk glands (B) and detail of transffered bacteria within milk gland secretions (C).

1.1.5 Phylogenetic overview of intracellular symbionts in arthropods

Intracellular insect symbionts are without question polyphyletic and originated many times from various free-living bacterial ancestors [7-9]. However, one phylogenetically interesting pattern observed within arthropod-bacterial symbiosis (Table 1) is that the currently known diversity of bacterial symbionts is scattered within only a few bacterial lineages and tend to form assemblages. This pattern may be due to several reasons. The two most important methodological reasons are bias in description of new lineages and phylogenetic artifacts. Biological explanations include horizontal transfer of a few established facultative symbionts across phylogenetically independent hosts (e.g. *Arsenophonus* and *Sodalis* clades), and functional and ecological background of a few bacterial lineages that makes them suitable to become intracellular symbionts. Traits of free-living bacteria helpful in symbiosis establishment are for example cell invasion apparatus (secretion systems) or pre-symbiotic intimate association with the host species (e.g. gut bacteria or pathogenic bacteria).

In this chapter, diversity of bacterial lineages which have evolved symbiotic associations with arthropods is overviewed. The term *symbiotic association* is used in its broad sense, including a broad spectrum of forms ranging from facultative commensals to obligate mutualists. Clades of typical reproductive manipulators are included either because they contain lineages confirmed to be mutualistic (e.g. *Wolbachia*) or because of their supposedly close relationship to obligate symbionts (e.g. *Flavobacterium*).

Table 1. Obligate symbiotic associations in insects. Symbiotic lineages are members of Enterobacteriales (γ-proteobacteria) if not stated otherwise.

Insect group		**Diet**	**Symbiotic lineages**
Blattaria	Cockroaches (+ *Mastotermes darwiniensis* termite)	omnivores, wood	*Blattabacterium cuenoti* (Bacteroidetes)
Psocoptera	Book lice	various	*Rickettsia* spp. (α-proteobacteria)
Thysanoptera	Thrips	plant material	*Stammerula tephritidis* – bacteria are localized extracellularly, but externally to the peritrophic membrane in midgut.
Phthiraptera	Anoplura (sucking lice)	blood	*Riesia* (*Arsenophonus*) *pediculicola* (only in Pediculidae) *Legionella* sp. (γ-proteobacteria: Legionellales; only in *Polyplax* spp.) several unnamed Enterobacterial lineages (in *Haematopinus*, *Solenopotes*, *Linognathus*, *Pedicinus* spp.)
	Rhynchophthirina (=*Haematomyzus* spp.)	blood	unnamed Enterobacteriales bacterium
	Ischnocera (chewing lice)	feather, skin debris, skin, blood	*Sodalis* sp. (in *Columbicola* spp.)
	Amblycera (chewing lice)		bacteria - no molecular data
Hemiptera: Sternorrhyncha	Coccoidea (scale insects)	plant sap	*Tremblaya princeps (*β-proteobacteria) + co-symbiont (*Moranella endobia* or other Enterobacteriaceae) in Pseudococcidae: Pseudococcinae *Tremblaya phenacola (*β-proteobacteria) in Pseudococcidae: Phenacoccinae *Uzinura diaspidicola* (Bacteroidetes) in Diaspididae *Brownia rhizoecola* (Bacteroidetes) in Pseudococcidae: Rhizoecini unnamed Bacteroidetes in *Rastrococcus*/*Cryptococcus* lineage unnamed Bacteroidetes in Monobhlebidae (*Icerya* spp. + *Drosicha* spp.) unnamed Enterobacteriales in *Drosicha* spp. unnamed Enterobacteriales in *Puto* spp. Fungi
	Aphidoidea (aphids)		*Buchnera aphidicola* *Buchnera aphidicola* + *Serratia symbiotica* *Pyrenomycetes* fungi
	Psylloidea (psyllids)		*Carsonella ruddii* (γ-proteobacteria: Oceanospirillales) *Carsonella ruddii* + Gammproteobacterial co-symbionts?
	Aleyrodoidea (whiteflies)		*Portiera aleyrodidarum* (γ-proteobacteria: Oceanospirillales) *Portiera aleyrodidarum* + Gammaproteobacterial co-symbionts?
Hemiptera: Heteroptera *****Pentatomoidea symbionts are extracellular, but obligate and vertically transferred, therefore included.	Triatomidae (Triatomid bugs)	blood	*Arsenophonus triatominarum* ?
	Cimicidae (Cimicids)		*Wolbachia* sp. (α-proteobacteria)
	Lygaeoidea (Seed bugs)	seeds	*Rohrkolberia cinguli* *Kleidoceria schneideri* *Schneideria nysicola*
	Pentatomoidea: Pentatomidae (stink bugs)	plant sap	several unnamed Gammaproteobacterial lineages
	Pentatomoidea: Acanthosomatidae (shield bugs)		*Rosenkranzia clausaccus*
	Pentatomoidea: Plataspidae (plataspid bugs)		*Ishikawaella capsulata*
	Pentatomoidea: Parastrachidae		*Benitsuchiphilus tojoi*
	Pentatomoidea: Scutelleridae (jewel bugs)		Unnamed Gammaproteobacteria + *Sodalis* sp.
Hemiptera : Auchenorrhyncha	Cicadoidea (Cicadas)	plant sap	*Sulcia muelleri* (Bacteroidetes) + *Hodgkinia cicadicola* (α-proteobacteria)
	Cercopoidea (spittlebugs)		*Sulcia muelleri* (Bacteroidetes) + *Zinderia insecticola (*β-proteobacteria)
	Membracoidea: Cicadellidae (leafhoppers)		*Sulcia muelleri* (Bacteroidetes) + *Baumannia cicadellinicola*
	Fulgoroidea (planthoppers)		*Sulcia muelleri* (Bacteroidetes) + *Vidania fulgoroideae* (α-proteobacteria) + *Purcelliella pentastirinorum* *Sulcia muelleri* (Bacteroidetes) + *Purcelliella pentastirinorum*

			Pyrenomycetes fungi
	Membracidae (treehoppers)		bacteria - no molecular data *Sulcia muelleri* (Bacteroidetes) + ?
Hymenoptera	Camponotini (carpenter ants)	omnivores	*Blochmannia* spp.
	Formicinae		*Sodalis* related bacterium? (in *Plagiolepis* spp.) unnamed Gammaproteobacterial lineage in some *Formica* species
	Pseudomyrmecinae		*Bartonella* (α-proteobacteria) related bacterium? (in *Tetraponeura* spp.)
Coleoptera	Throscidae	plant material, wood	unnamed Bacteroidetes + *Sodalis* related bacterium
	Nosodendridae	tree sap	bacteria - no molecular data
	Bostrychidae	wood	bacteria - no molecular data
	Lyctidae		variable bacteria - no molecular data
	Anobiidae		fungi - no molecular data
	Cerambycidae		Ascomycetes fungi *Sodalis* sp. (only in *Tetropium castaneum*)
	Chrysomelidae	plant material	bacteria in *Cassida* and *Bromius* spp. – no molecular data monophyletic symbiotic lineage in Donaciinae (one subclade was named *Macropleicola* spp.)
	Silvanidae: *Oryzaephilus* spp. only	stored products	bacteria - no molecular data
	Curculionidae	plant material, grains	*Nardonella* sp. *Sodalis* sp. *Curculioniphilus buchneri* unnamed gammaproteobacterial lineage
Diptera: Hippoboscoidea	Glossinidae (tsetse flies)	blood	*Wigglesworthia glossinidia* *Wigglesworthia glossinidia + Sodalis glossinidius*
	Hippoboscidae (louse flies		*Arsenophonus* sp. *Arsenophonus* sp. + *Sodalis* sp.
	Nycteribiidae + Streblidae (bat flies)		*Aschnera* (*Arsenophonus*) *chinzeii*
Diptera: Ceratopogonidae	*Dasyhelea* sp.	tree sap	bacteria - no molecular data

Alphaproteobacteria

Wolbachia (Rickettsiales)

The most species-rich hotspot of intracellular bacteria is the genus of reproductive manipulators, *Wolbachia*, with its high prevalence in arthropods (estimated to be over 66% in insects) and nematodes [58]. Many arthropod species can, moreover, harbor multiple infections; up to 5 different *Wolbachia* strains have been reported in a single host [59]. In at least three cases, *Wolbachia* evolved into an obligatory associate, namely in bedbugs, parasitic wasps [60-62] and filarial nematode lineage [63], supporting the hypothesis of transmission from parasitic or facultative to obligate symbiotic lifestyle [64, 65].

However, no reliable phylogeny of *Wolbachia* clade is currently available [66], obligate relationship with its host is rarely tested, and new lineages are only assigned to the phylogenetically related supergroups. It is therefore still uncertain how many origins of obligate *Wolbachia* exist, what is its "free-living" ancestor, and whether the huge *Wolbachia* cluster is

monophyletic or not. Many of currently known *Wolbachia* species can theoretically be involved in facultative mutualism, e.g. inducing resistance to RNA-viruses as described in *Drosophila* [24]. Such a transition from parasitic to mutualistic effect of *Wolbachia* was in natural populations of *Drosophila simulans* shown to take only 20 years [67].

Rickettsia+Midichloria+Rickettsia-like (Rickettsiales)

Within Rickettsiales, there are two more symbiotic lineages with insects and acari. The first is genus *Rickettsia*, known as a facultative bacterium or reproductive manipulator of various insects and mites [68, 69]. One *Rickettsia* lineage described from whiteflies is very likely an obligatory co-symbiont contributing to the provision of essential nutrients to the host [51, 70]. Moreover, there are numerous lineages with unknown effect, some of them likely nonpathogenic. The last symbiotic lineage of *Rickettsia* is known from different, phylogenetically distant hosts, booklice (Psocoptera). This lineage is characterized by peculiar intranuclear localization and is essential for the host [53].

A bacterium closely related to *Rickettsia*, which is very common in natural populations of various ticks and is localized within mitochondria, was described as *Midichloria mitochondrii*. Its function is currently unknown, although the genome data and its 100% prevalence in *Ixodes ricinus* females suggest that it might supply B-vitamins, cofactors or heme during starvation of its hosts [71]. This theory would explain *Midichloria* losses occurring in tick laboratory colonies with high frequency of blood-meals and correlates with the fact that ticks cannot produce harem [54, 71-74]. In addition to *Midichloria*, some species of ticks harbor *Rickettsia* with unknown effect on their vertebrate or tick host and uncertain phylogenetic position in respect to *Midichloria* [75]. This makes it currently impossible to determine how many independent symbiotic lineages have arisen within the ticks.

Lariskella arthropodarum (Rickettsiales)

Recently, facultative alphaproteobacterial associates from stinkbugs, ticks and fleas were included into a novel lineage *Lariskella arthropodarum* [76]. However, much more Rickettsiales species are needed in future to assess putative monophyly of this questionable lineage.

Bartonella+Bartonella-like (Rhizobiales)

Several studies have reported presence of bacteria closely related to the genus *Bartonella*

from different ant species [77, 78]. Genomes of these bacteria were also sequenced as a contamination in ant genome projects (L. Guy personal communication). However, more detailed studies are needed to determine if these bacteria are symbiotic or not.

Sheep infecting bacterium *Bartonella melophagi* was previously suggested to be in symbiosis with its vector *Melophagus ovinus* because of its 100% prevalence in both adults and larvae [79]. Nevertheless, microscopical and genomic analyses (Husník et al., unpublished results) suggest that these bacteria are located extracellularly along the microvilli of the midgut section containing the bacteriome. The high prevalence of this bacterium is probably caused by exploitation of the milk glands for vertical transfer. Genomic data did not reveal any strong evidence for obligate mutualism or cooperation with obligate *Arsenophonus* and facultative *Sodalis* symbionts, but complementary provision of B-vitamins cannot be excluded because the genome retains several pathways for B-vitamins biosynthesis (unpublished results).

Hodgkinia (Rhizobiales)

A co-symbiont of *Sulcia muelleri* in xylem-feeding cicadas, *Hodgkinia cicadicola*, is a bacterium with one of the most extremely reduced genomes (144 kb), but with an unprecedented genome GC content of 58.4% and alternative (UGA=stop) genetic code [44, 80]. Since mutations are universally biased toward AT in bacteria [81, 82], random genetic drift should lead towards the higher AT content as exemplified by genomes of most of insect endosymbionts [39, 45, 83]. The unusually high GC content of *Hodgkinia* (62.5 % at third positions of fourfold degenerate codons) is thus very likely driven by selection [84]. This calls into question genetic theories and models that assume that selection is less effective in populations of very small size (such as the populations of insect symbionts affected every generation by a strong bottleneck).

Betaproteobacteria

Tremblaya (Burkholderiales)

Interestingly, the smallest bacterial genome sequenced to date (139 kb) has been determined for a symbiont in Pseudococcidae mealybugs, *Tremblaya princeps*, which has also unusually high genome GC content of 58.8%. This Betaproteobacterium hosts inside its cells symbiotic Gammaproteobacteria, *Moranella endobia*, with a genome almost four times larger (538 kb) and both symbionts provide in concert its host with essential amino acids [15, 36, 85-

87]. Phylogenetic data on basal mealybug lineage Phenacoccinae confirmed morphological observations [47] that this lineage harbors *Tremblaya* (*Tremblaya phenacola*) without its intrabacterial symbionts and that this ancestral state was complemented in Pseudococcidae by intrabacterial Gammaproteobacteria and replaced by Bacteroidetes in Rhizoecini and *Cryptococcus*/*Rastrococcus* lineages [42].

Zinderia (Burkholderiales: Oxalobacteraceae)

The lowest GC content (13.5 %) yet observed within any cellular genome is that of a co-symbiont in spittlebugs - *Zinderia* insecticola [45]. This striking AT bias had profound effects on its proteome in which 36.1 % of amino acids are either isoleucine or lysine. Moreover, one change from universal code has occurred and *Zinderia* uses alternative genetic code in which UGA codes for tryptophan instead of stop identically as in *Hodgkinia*.

Vidania fulgoroideae (Burkholderiales: ?)

A probable co-symbiont of *Sulcia* and *Purcelliella* in Cixiidae planthoppers was found to cluster with bacteria associated with ticks and close to *Zinderia* [48]. Two possible scenarios can explain this topology. First, *Vidania* and *Zinderia* are members of one ancient lineage infecting Auchenorrhyncha, which was repeatedly lost or originated from closely related free-living Betaproteobacterium. Second, presented topology is artifactual because of similarly low GC content in both *Zinderia* and *Vidania*.

Gammaproteobacteria

Most of gammaproteobacterial symbionts originated within Enterobacteriaceae, but at least three symbiotic lineages are known to originate outside of this group.

Rickettsiella (Legionellales: Coxiellaceae)

Bacteria of the genus *Rickettsiella* were recently relocated from Alphaproteobacteria to Gammaproteobacteria, close to *Coxiella* and *Legionella* [88, 89]. These pathogenic bacteria and facultative symbionts are known from numerous arthropod groups including crustaceans, crickets, cockroaches, flies, beetles, spiders, mites and aphids, where they were shown to modify the aphid color and thus change its susceptibility to predators and parasites [90].

Clustering of *Rickettsiella* with *Coxiella*-like symbionts and *Diplorickettsia massiliensis* from ticks [91] implies that these bacteria might be members of a single clade.

Legionella (Legionellales: Legionellaceae)

Bacteriocyte-associated *Legionella* spp. were sequenced from two lice species (*Polyplax serrata* and *Polyplax spinulosa*) and appear to be obligate based on its location in host and vertical transmission [92, 93].

Carsonella+Portiera (Oceanospirillales?)

Although obligate symbionts of psyllids (*Carsonella ruddii*) and whiteflies (*Portiera aleyrodidarum*) were inferred to be related to *Pseudomonas* clade [94, 95], there is currently no reliable (non 16S rDNA) multi-locus phylogeny confirming such relationships or confirming that these symbionts are sister species. Unfortunately, *Portiera* genome is not yet available and availability of the 160 kb genome of *Carsonella ruddii* did not change this situation [83]. This species with its extreme AT bias (16.6% GC) and rapid evolutionary rate is usually excluded from phylogenetic analyses and if included, it is either attracted to AT-rich species within Enterobacteriales [96] or must be constrained to Oceanospirillales [7]. Strikingly, no realistic evolutionary model has ever been used to figure out its topology, although exclusion of AT-rich species has also suggested its placement within Oceanospirillales [96].

Enterobacteriales

Since the attached study [3] is devoted to detail genome-based phylogeny of Enterobacteriales, only the lineages for which complete genome is not available will be briefly discussed here. For most of these lineages, phylogenetic position is highly unstable and cannot be evaluated in respect to free-living enterobacteria. (The following bacteria are included and more thoroughly discussed in the attached phylogenetic study: *Arsenophonus* (incl. *Riesia*, *Phlomobacter* and *Aschnera*), *Baumannia*, *Blochmannia*, *Buchnera*, *Ishikawaella*, *Sodalis*, *Serratia*, *Hamiltonella*, *Regiella* and *Wigglesworthia*).

Recently, several intracellular enterobacterial lineages symbiotic in Lygaeoidea seed bugs (*Kleidoceria schneideri*, *Rohrkolberia cinguli*, *Schneideria nysicola*) and extracellular symbiotic lineages in Pentatomoidea bugs (*Benitsuchiphilus tojoi*, *Rosenkranzia clausaccus*)

have been reported [50, 97-102]. Although all these lineages are convincingly obligate mutualists, their phylogenetic position was almost exclusively based on 16S rRNA gene and is thus highly uncertain. Genome data are needed to distinguish how many times intracellular symbiosis originated within plant-feeding Heteroptera and if some lineages originated from gut associated symbionts.

Three distinct lineages of obligate symbionts are currently known from beetles: *Curculioniphilus buchneri* in *Curculio* weevils [103], *Nardonella* sp. in Dryopthoridae weevils [40, 41, 104, 105] and an unnamed monophyletic lineage in Donaciinae reed beetles (Chrysomelidae), subclade of which was named *Macropleicola* [106, 107].

Several endosymbiotic lineages of lice originated within Enterobacteriales, namely *Puchtella pedicinophila* from *Pedicinus* lice [108] and unnamed lineages from *Haematomyzus*, *Haematopinus*, *Solenopotes* and *Linognathus* genera [92].

A co-symbiont of Cixiidae planthoppers, typically for Auchenorrhyncha housed in separate bacteriomes from *Sulcia muelleri*, was named *Purcelliella pentastirinorum* [49].

In addition to the symbiotic lineages mentioned above, unnamed enterobacterial lineages with unknown function were sequences from whiteflies, psyllids or scale insects [14-16, 42].

Bacteroidetes

Several lineages of insects harbor symbionts which originated within Bacteroidetes. Relationship among these lineages is uncertain and it is not known if this group is monophyletic or not.

Cardinium hertigii (Bacteroidales)

Bacteria responsible for sex manipulation in arthropods are not only members of the *Wolbachia* clade, but several other lineages such as *Cardinium hertigii* has also evolved molecular mechanisms causing cytoplasmic incompatibility, parthenogenesis, feminization or male killing [109-111]. Although no obligate mutualists have been reported from this genus, its widespread prevalence in arthropods makes it tempting to speculate that some obligate mutualists will be found in future similarly to *Wolbachia*.

Flavobacterium (Flavobacteriales)

Phylogenetic position of an unnamed *Flavobacterium* causing male-killing in two lady bugs [112, 113] suggests that this bacterium is closely related to the putative Bacteroidetes clade of obligate mutualists containing *Sulcia*, *Blattabacterium*, *Brownia*, *Uzinura* and several other symbiotic lineages. On the other hand, support for this huge symbiotic clade is either weak or there are no free-living taxa used in phylogenetic datasets excluding any discussion about independent origins [42].

Uzinura (Flavobacteriales?)

Armored scale insects (Coccoidea: Diaspididae) have established an evolutionary stable symbiotic relationship with obligate symbiont *Uzinura diaspidicola,* undergoing at least 60 million years of strict coevolution [114, 115]. Random distribution in fat bodies might suggest that this symbiont shares some homologous traits with *Blattabacterium* (e.g. uric acid recycling).

Brownia (Flavobacteriales?)

Two lineages of mealybugs were found to replace their original obligate symbiont *Tremblaya phenacola* with Bacteroidetes bacteria. These replacements (in Rhizoecini and *Rastrococcus*/*Cryptococcus* clades) are thought to originate from two different bacterial lineages, the Rhizoecini-infecting lineage was named *Brownia rhizoecola* [42].

Blattabacterium (Flavobacteriales)

Four genome analyses of *Blattabacterium cuenoti,* fat bodies associated symbionts of cockroaches and basal *Mastotermes darwiniensis* termite, have confirmed its role in recycling nitrogen from urea or ammonia into glutamate [116-119]. In addition, loss of several pathways in wood-diet shifted lineages of *Mastotermes darwiniensis* and *Cryptocercus punctulatus* suggests that products of these pathways might be complemented by hindgut microbiota.

Sulcia (Flavobacteriales)

One of the most ancient (at least 260 million years old) and stable symbiotic relationship in arthropods is between *Sulcia muelleri* and Auchenorrhyncha group [120]. This symbiotic lineage forms dual or tripartite co-symbioses with other bacteria depending on host group (Table

1) and provides the hosts with essential amino acids [43-46, 48, 49, 121, 122]. It is interesting to note that *Sulcia* was named to honor Karel Šulc, a Moravian scientist who as one of the first authors (in 1909) recognized the bacteriome as an organ harboring microorganism [47, 120].

Tenericutes: Mollicutes

Spiroplasma (Entomoplasmatales: Spiroplasmataceae)

Numerous cases of *Spiroplasma* bacteria have been reported from arthropods mainly because of its male-killing effect [123], but several mutualistic lineages providing protection against parasitic nematodes, parasitoid wasps and cold were also described [124-126].

Chlamydiae

Fritschea (Chlamydiales: Simkaniaceae)

Two bacteriocyte-associated members of Chlamydiae closely related to *Simkania negevensis* were identified from phloem sap sucking insects: *Fritschea bemisiae* from a whitefly *Bemisia tabaci* and *Fritschea eriococci* from a scale insect *Eriococcus spurius* [127, 128]. No biological data for these species are available.

1.2 Evolutionary implications of the intracellular lifestyle

Because of uncultivable nature of the obligate symbionts, evolutionary implications of their intracellular lifestyle are mainly inferred from *in silico* analyses of genome sequence. It is important to mention that this approach leads to a certain level of uncertainty because of lack of the knowledge on the host genome and functional data from transcriptomes or proteomes, currently available only for the model species *Buchnera aphidicola* [129-131]. Although gene annotations in symbiotic genomes are generally of very high quality with only a few hypothetical genes of unknown function, assessment of pseudogenes is an extremely difficult step and commonly leads to misannotation of genes which produce a functional protein as pseudogenes or vice versa.

1.2.1 Endosymbiotic horizontal gene transfer (EGT)

It is generally assumed that evolutionary transmission from a symbiont to an organelle is

accompanied by transfers of the symbiont genes to the host chromosome and consequent targeting of proteins into the organelle. This hypothesis was corroborated by numerous studies concerning unicellular eukaryotes [132] and became one of definitions of organelles. However, in the case of arthropod symbionts, the transfer of functional genes was not validated by the two sequenced insect genomes with obligate bacteria: *Pediculus humanus* [133] and *Acyrthosiphon pisum* [131]. In the human lice genome, there were no sequences of bacterial origin found, but the exact methodological procedure was not described, which is unusual and calls for reanalysis. In the pea aphid genome, rigorous analyses revealed 12 genes of bacterial origin [131, 134, 135]. Out of these 12 genes, two pseudogenes (*dnaE* and *atpH*) appear to be transferred from *Buchnera* and the rest was transferred from Rickettsiales, probably *Wolbachia* (three LD-carboxypeptidases-one pseudogenized, five rare lipoprotein As, N-acetylmuramoyl-L-alanine amidase and 1,4-beta-N-acetylmuramidase).

The only available sequencing evidence for a functional gene transfer is from filarial nematodes [136] which are universally associated with an ancient obligate *Wolbachia* providing them with riboflavin and heme. In a few filarial lineages, this cooperation was found to be lost and the sequencing data of these species clearly suggest that parts of the original *Wolbachia* genome were transferred to the host chromosomes.

Remarkably, most cases of symbiont-to-arthropod gene transfers are known to originate from reproductive manipulators of the genus *Wolbachia* (Alphaproteobacteria: Rickettsiales). *Wolbachia* are commonly present at high density in germ cells and are transferred mostly through the egg cytoplasm, which provides opportunity for gene transfers to the host genome. Transfers of whole *Wolbachia* chromosomes or single genes are known to occur in two beetle species [137-139], several *Drosophila* species [140], mosquitos [141, 142] and parasitoid *Nasonia* wasps [143].

1.2.2 Genome streamlining

Numerous complete genomes of bacterial symbionts now available (Table 2) allowed for generalizing the changes accompanying the shift towards intracellular obligate symbiosis. The most striking is the genome reduction reaching from about 0.8 Mb to the most extremely reduced genomes smaller than 200 Kb. All bacterial features nonessential in the host cell environment are discarded; lipopolysaccharide and peptidoglycan biosynthetic pathways are

eroded in a way that some symbionts (mostly those residing within host symbiosomal membrane) retain only a fragile cell envelope. Rod cell shape genes are also lost and cell shape is usually spherical (Fig 2A), elongated tubular (up to 200 μm) or forms irregular blobs [7].

In some lineages, the genome reduction inevitably leads also to loss of many essential pathways that provide nucleotides, ATP, amino acids, B-vitamins and cofactors. All these compounds must then be supplemented by the host.

DNA repair and recombination genes are depleted in symbiont genomes, which together with small population size and severe bottlenecks increase mutational bias. Therefore, deleterious and slightly deleterious mutations accumulate; proteins of symbiotic bacteria have lower thermal stability and must be buffered by heat shock proteins [144]. Constitutive overexpression of heat shock proteins (*GroL*, *GroS*, *DnaJ*, *DnaK*, and *GrpE*) in the absence of stress is thus one of features typical for all bacterial symbionts with reduced genomes.

As mentioned above, bacteria have universal AT mutational bias [81, 82]. In symbiont populations with rapid evolution and strong effect of random genetic drift, genomes become AT rich (e.g. 13.5 % GC in *Zinderia*), presumably due to Muller's ratchet and absence of a repairing mechanism [145-147]. Nevertheless, the two most reduced genomes, *Tremblaya* (139 kb) and *Hodgkinia* (144 kb) have both high GC content (>58 %) implying that selection may play a role even in populations of insect symbionts with tiny effective population sizes [84, 148].

Table 2. Taxonomic designation, genome size and GC content of genome sequences of symbiotic bacteria of arthropods. If more than one strain has been sequenced from the same host (*Buchnera* strains from *Acyrthosiphon pisum* and *Tremblaya* strains from *Planococcus citri*), only the first published is presented here. Only complete genomes are included from the genus *Wolbachia* because of high number of its draft sequencing projects.

Bacterial group	**Symbiotic bacterium** (source hosts in brackets)	**Taxonomy**	**Genome size**	**GC content**	**Reference**
α-proteobacteria	*Hodgkinia cicadicola* str. Ds (*Diceroprocta semicincta*)	Rhizobiales: ?	143,795 bp	58.4 %	[44]
	Midichloria mitochondrii (*Ixodes ricinus*)	Rickettsiales	1,183,732 bp	36.6 %	[71]
	Wolbachia pipientis str. wMel (*Drosophila melanogaster*)	Rickettsiales: Anaplasmataceae	1,267,782 bp	35.2 %	[149]
	Wolbachia pipientis str. wPip (*Culex quinquefasciatus*)	Rickettsiales: Anaplasmataceae	1,482,455 bp	34.2 %	[150]
	Wolbachia pipientis str. wRi (*Drosophila simulans*)	Rickettsiales: Anaplasmataceae	1,445,873 bp	35.2 %	[151]
β-proteobacteria	*Tremblaya princeps* str. TPPCIT (*Planococcus citri*)	Burkholderiales: Burkholderiaceae?	138,927 bp	58.8 %	[36, 87]
	Zinderia insecticola str. Ca (*Clastoptera arizonana*)	Burkholderiales: Oxalobacteraceae?	208,564 bp	13.5 %	[45]
γ-proteobacteria	*Arsenophonus nasoniae* (*Nasonia vitripennis*)	Enterobacteriales	~3.3 Mbp (draft) ~100 Kbp plasmid(s) ~200 Kbp phage(s)	37.7 %	[152, 153]
	Baumannia cicadellinicola (*Homalodisca coagulata*)	Enterobacteriales	686,192 bp	33.2 %	[121]
	Blochmannia floridanus (*Camponotus floridanus*)	Enterobacteriales	705,557 bp	27.4 %	[154]
	Blochmannia pennsylvanicus (*Camponotus pennsylvanicus*)	Enterobacteriales	791,654 bp	29.6 %	[155]
	Blochmannia vafer (*Camponotus vafer*)	Enterobacteriales	722,593 bp	27.5 %	[156]
	Buchnera aphidicola str. APS (*Acyrthosiphon pisum*)	Enterobacteriales	640,681 bp + 7,786 plasmid pLeu + 7258 plasmid pTrp	26.2 %	[157]
	Buchnera aphidicola str. Ak (*Acyrthosiphon kondoi*)	Enterobacteriales	641,794 bp + 7,784 plasmid pLeu + 3,645 plasmid pTrp	25.7 %	[158]
	Buchnera aphidicola str. Bp (*Baizongia pistaciae*)	Enterobacteriales	615,980 bp plasmid not determined	25.3 %	[159]
	Buchnera aphidicola str. Cc (*Cinara cedri*)	Enterobacteriales	416,380 bp + 6,054 plasmid pLeu	20.1 %	[39]
	Buchnera aphidicola str. Ct (*Cinara tujafilina*)	Enterobacteriales	444,925 bp + 8,069 bp plasmid pLeu/Trp	23.0 %	[160]
	Buchnera aphidicola str. Sg (*Schizaphis graminum*)	Enterobacteriales	641,454 bp + 7,967 plasmid pLeu + 3580 plasmid pTrp	26.3 %	[161]
	Buchnera aphidicola str. Ua (*Uroleucon ambrosiae*)	Enterobacteriales	615,380 bp + 7,689 plasmid pLeu + 4,884 plasmid pTrp	24.1 %	[158]
	Carsonella ruddii (*Pachypsylla venusta*)	Oceanospirillales ?	159,662 bp	16.5 %	[83]
	Hamiltonella defensa str. 5AT (*Acyrthosiphon pisum*)	Enterobacteriales	2,110,331 bp (draft) + 59,032 plasmid	40.1 %	[162]
	Moranella endobia str. MEPC (*Planococcus citri*)	Enterobacteriales	538,294 bp	43.5 %	[36]
	Regiella insecticola str. LSR1 (*Acyrthosiphon pisum*)	Enterobacteriales	2,067,400 bp (draft)	42.5 %	[163]
	Regiella insecticola str. R5.15 (*Myzus persicae*)	Enterobacteriales	2,013,072 bp (draft)	42.6%	[164]
	Riesia pediculicola (*Pediculus humanus humanus*)	Enterobacteriales	574,526 bp + 7,628 bp plasmid	28.5 %	[133]

	Serratia symbiotica str. APS (*Acyrthosiphon pisum*)	Enterobacteriales	2,789,218 bp (draft)	52.0 %	[165]
	Serratia symbiotica str. Cc (*Cinara cedri*)	Enterobacteriales	1,762.765 bp	29.2 %	[35]
	Sodalis glossinidius str. Gm (*Glossina morsitans*)	Enterobacteriales	4,171,146 bp + 83,306 bp plasmid 1 + 27,240 bp plasmid 2 + 10,810 bp plasmid 3	54.7 %	[166]
	Wigglesworthia glossinidia str. Gb (*Glossina brevipalpis*)	Enterobacteriales	697,724 bp + 5,200 bp plasmid	22.5 %	[167]
	Wigglesworthia glossinidia str. Gm (*Glossina morsitans*)	Enterobacteriales	719,535 bp + 5,198 bp plasmid	25.2%	[57]
Bacteroidetes	*Sulcia muelleri* str. Ca (*Clastoptera arizonana*)	Flavobacteriales	276,511 bp	21.1 %	[45]
	Sulcia muelleri str. Ds (*Diceroprocta semicincta*)	Flavobacteriales	276,984 bp	22.6%	[44]
	Sulcia muelleri str. Dm (*Draeculacephala minerva*)	Flavobacteriales	243,933 bp	22.5 %	[122]
	Sulcia muelleri str. Hc (*Homalodisca coagulata*)	Flavobacteriales	245,530 bp	22.4 %	[43, 121]
	Blattabacterium cuenoti str. Bg (*Blattella germanica*)	Flavobacteriales	636,850 bp + 4,085 bp plasmid	28.2 %	[116]
	Blattabacterium cuenoti str. Cp (*Cryptocercus punctulatus*)	Flavobacteriales	605,745 bp + 3,816 bp plasmid	23.8 %	[118]
	Blattabacterium cuenoti str. Md (*Mastotermes darwiniensis*)	Flavobacteriales	587,248 bp + 3,088 bp plasmid	27.5 %	[119]
	Blattabacterium cuenoti str. Pa (*Periplaneta americana*)	Flavobacteriales	636,994 + 3,448 bp plasmid	27.1 %	[117]

1.2.3 Genetic information processing

Apart from the genes related to the provision of nutrients to the host and heat shock proteins, another category of genes is relatively highly retained. It is the category of genetic information processing, which is one of the reasons why the symbionts with reduced genomes are still considered bacteria rather than organelles. *Tremblaya princeps* will not be included and discussed here because of its composite structure with *Moranella endobia*.

The most gene variable genetic process in symbiont genomes is replication with rich set of genes in the genomes larger than 500 kb, but only the 5′-to-3′ DNA polymerase subunit (DNA pol. III α-subunit; *dnaE*), and its associated 3′-to-5′ proofreading exonuclease subunit (DNA pol. III ε-subunit; *dnaQ*) retained in the most reduced symbiont genomes. For transcription, all symbiont genomes contain three core subunits of RNA polymerase (*rpoABC*) along with its sigma factor (*rpoD*). The substantial part of genome is usually devoted to essential ribosomal RNAs and rRNA modification (*rlu* genes) and transfer RNAs and tRNA modification (*mnmAEG*). For translation, core structure of both ribosomal subunits is consistently retained along with translation initiation factors (*infABC*), elongation factors (*fusA*, *tsf*), protein release factors (*prfAB*), ribosome recycling factor (*frr*) and peptide deformylase (*def*). Aminoacyl-tRNA synthetases may not be retained for all tRNAs, but at least eight of them are always retained [7].

1.2.4 Host-symbiont cooperation

The most insightful transcriptomic study on host role in arthropod-symbiosis [129] untangled the intimate symbiotic interface in the pea aphid-*Buchnera* system and confirmed the previously suggested host-symbiont cooperation in the production of essential amino acids [168] and incorporation of ammonium nitrogen into glutamate (GOGAT cycle). Results of this study were further corroborated by proteomic approach; no evidence for the selective transfer of proteins among the symbiotic partners was detected [130], although previously proposed to be undertaken by flagellar bodies [169]. Nevertheless, no transcriptomic study has so far been published for a symbiotic system with several bacteria or from a blood-sucking host.

A study on response of aphid transcriptome on infection of secondary symbionts (*Serratia symbiotica*) revealed only a few differentially expressed genes, so the metabolic impact of facultative symbiont is predicted to be the result of the symbiont itself [170].

Host dependence on the long-lasting symbiosis between *Buchnera* and aphids and the intimacy of the association is exemplified by losses of several essential pathways such as the urea cycle and arginine biosynthesis or purine salvage pathway [171-173].

Hos-symbiont cooperation is exceptionally dependent on well working transport of compounds to and from the bacteriocytes and to and from the symbiont cells. Transporters were so far studied only for the aphid-*Buchnera* model system and two published papers conclude that symbiotic bacteria retain only a few general transporters, some of which very likely lost its substrate specificity [174] and that the host transporters involved in amino acid transport were extensively duplicated and specialized for bacteriocyte transfer [175].

Probably as a result of long-term association with bacteria, aphids, as the main model for the insect symbiosis studies, have low number of antimicrobial immunity genes [131]. For this reason, control and maintenance of vertically transmitted obligate symbionts is mainly studied in weevils of the genus *Sitophilus* and their recently acquired obligate *Sodalis* symbionts [176, 177]. This symbiotic system is in early phase of symbiont domestication and antimicrobial peptides were shown to keep symbionts under control and if these genes were silenced by RNA interference, symbionts were escaping from bacteriocytes and spreading into host tissues [178].

1.3 Conclusion and future prospects

The presented overview arises several general evolutionary questions concerning arthropods-bacteria symbiosis. In the attached study, focused on phylogenetic relationships within Enterobacteriaceae, I tried to use the most advanced phylogenetic methods to analyze phylogeny, and address various evolutionary issues within Enterobacteriaceae. Below, I briefly cover additional questions and likely answers emerging from the current knowledge. Finally, I summarize the topics to be addressed in a complex manner using yet additional molecular data.

1) How frequently have intracellular symbioses originated among different groups of bacteria and arthropods?

   Answer: There are at least twenty independent origins within bacteria and numerous lineages with uncertain position.
2) Are some taxonomical or ecological groups predisposed to form intracellular symbiotic associations?

   Answer: Yes, very likely. Mainly those commonly associated with arthropods (e.g. gut bacteria, animal and plant pathogens).
3) How common is transition between pathogenic and symbiotic lifestyle or vice versa?

   Answer: Sometimes happens, e.g. in *Wolbachia* and *Arsenophonus* lineages.
4) Which changes affect symbiotic associations and how common are losses, replacements or complementations of established symbionts.

   Answer: All these changes affect symbiotic associations and are very common.
5) How intracellular lifestyle adjusts genomes, transcriptomes or proteomes of both symbiotic partners?

   Answer: Mainly genome reduction in bacteria; gene loss, expansion and specialization followed by intimate cooperation in arthropods.
6) Are symbionts with extremely reduced genomes bacteria, organelles or something in between?

   Answer: Weird bacteria with exception of *Tremblaya princeps*, which is a composite organism with *Moranella endobia*.

Future phylogenetic goals

1) Are obligate symbionts within the Bacteroidetes (*Blattabacterium*, *Brownia*, *Sulcia* and *Uzinura*) monophyletic or not, what are their closest free-living ancestors and what is their relation to reproductive manipulators within this group?
2) Is the suggested position of *Hodgkinia* within the Rhizobiales correct and how many times has obligate symbiosis originated within the Rickettsiales?
3) What are the closest free-living betaproteobacterial ancestors of *Tremblaya* and *Zinderia*? Is *Zinderia* and *Vidania* an identical lineage of an ancient co-symbiont of *Sulcia* within Auchenorrhyncha?
4) How did the symbionts replacements in Auchenorrhyncha, weevils and scale insects take place?
5) Is *Carsonella* a sister species to *Portiera*, did they originate within Oceanospirillales, and what are their free-living relatives?
6) Will future genome projects and additional phylogenomic analyses expand number of independent origins of endosymbiosis within the Enterobacteriales?
7) Will future analyses and additional free-living lineages break the pattern of species rich symbiotic clades within Gammaproteobacteria and Bacteroidetes?

1.4 References

1. Douglas AE: **The symbiotic habit**: Princeton University Press; 2010.
2. Zchori-Fein E, Bourtis K: **Manipulative Tenants: Bacteria Associated with Arthropods**: CRC Press; 2011.
3. Husník F, Chrudimský T, Hypša V: **Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches**. *BMC Biol* 2011, **9**:87.
4. Douglas AE: **Mycetocyte symbiosis in insects**. *Biol Rev Camb Philos Soc* 1989, **64**(4):409-434.
5. Baumann P: **Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects**. *Annu Rev Microbiol* 2005, **59**:155-189.
6. Wernegreen JJ: **Genome evolution in bacterial endosymbionts of insects**. *Nat Rev Genet* 2002, **3**(11):850-861.
7. McCutcheon JP, Moran NA: **Extreme genome reduction in symbiotic bacteria**. *Nat Rev Microbiol* 2012, **10**(1):13-26.
8. Moran NA, McCutcheon JP, Nakabachi A: **Genomics and evolution of heritable bacterial symbionts**. *Annu Rev Genet* 2008, **42**:165-190.
9. Moya A, Pereto J, Gil R, Latorre A: **Learning how to live together: genomic insights into prokaryote-animal symbioses**. *Nat Rev Genet* 2008, **9**(3):218-229.
10. Douglas AE, Francois CLMJ, Minto LB: **Facultative 'secondary' bacterial symbionts and the nutrition of the pea aphid, *Acyrthosiphon pisum***. *Physiol Entomol* 2006, **31**(3):262-269.
11. Fukatsu T, Nikoh N, Kawai R, Koga R: **The secondary endosymbiotic bacterium of the pea aphid *Acyrthosiphon pisum* (Insecta: Homoptera)**. *Appl Environ Microbiol* 2000, **66**(7):2748-2758.
12. Russell JA, Latorre A, Sabater-Munoz B, Moya A, Moran NA: **Side-stepping secondary symbionts: widespread horizontal transfer across and beyond the Aphidoidea**. *Mol Ecol* 2003, **12**(4):1061-1075.
13. Sandstrom JP, Russell JA, White JP, Moran NA: **Independent origins and horizontal transfer of bacterial symbionts of aphids**. *Mol Ecol* 2001, **10**(1):217-228.
14. Thao ML, Clark MA, Baumann L, Brennan EB, Moran NA, Baumann P: **Secondary endosymbionts of psyllids have been acquired multiple times**. *Curr Microbiol* 2000, **41**(4):300-304.
15. Thao ML, Gullan PJ, Baumann P: **Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts**. *Appl Environ Microbiol* 2002, **68**(7):3190-3197.
16. Thao ML, Baumann P: **Evidence for multiple acquisition of *Arsenophonus* by whitefly species (Sternorrhyncha: Aleyrodidae)**. *Curr Microbiol* 2004, **48**(2):140-144.
17. Oliver KM, Degnan PH, Burke GR, Moran NA: **Facultative symbionts in aphids and the horizontal transfer of ecologically important traits**. *Annu Rev Entomol* 2010, **55**:247-266.
18. Pontes MH, Dale C: **Culture and manipulation of insect facultative symbionts**. *Trends Microbiol* 2006, **14**(9):406-412.
19. 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* 1999, **49 Pt 1**:267-275.

20. Matthew CZ, Darby AC, Young SA, Hume LH, Welburn SC: **The rapid isolation and growth dynamics of the tsetse symbiont *Sodalis glossinidius***. *FEMS Microbiol Lett* 2005, **248**(1):69-74.
21. Sabri A, Leroy P, Haubruge E, Hance T, Frere I, Destain J, Thonart P: **Isolation, pure culture and characterization of *Serratia symbiotica* sp nov., the R-type of secondary endosymbiont of the black bean aphid *Aphis fabae***. *Int J Syst Evol Microbiol* 2011, **61**:2081-2088.
22. 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* 2006, **72**(4):2997-3004.
23. Degnan PH, Bittleston LS, Hansen AK, Sabree ZL, Moran NA, Almeida RP: **Origin and examination of a leafhopper facultative endosymbiont**. *Curr Microbiol* 2011, **62**(5):1565-1572.
24. Teixeira L, Ferreira A, Ashburner M: **The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster***. *PLoS Biol* 2008, **6**(12):e2.
25. Oliver KM, Degnan PH, Hunter MS, Moran NA: **Bacteriophages encode factors required for protection in a symbiotic mutualism**. *Science* 2009, **325**(5943):992-994.
26. Oliver KM, Moran NA, Hunter MS: **Variation in resistance to parasitism in aphids is due to symbionts not host genotype**. *Proc Natl Acad Sci U S A* 2005, **102**(36):12795-12800.
27. Oliver KM, Moran NA, Hunter MS: **Costs and benefits of a superinfection of facultative symbionts in aphids**. *Proc Biol Sci* 2006, **273**(1591):1273-1280.
28. Oliver KM, Russell JA, Moran NA, Hunter MS: **Facultative bacterial symbionts in aphids confer resistance to parasitic wasps**. *Proc Natl Acad Sci U S A* 2003, **100**(4):1803-1807.
29. Hansen AK, Jeong G, Paine TD, Stouthamer R: **Frequency of secondary symbiont infection in an invasive psyllid relates to parasitism pressure on a geographic scale in California**. *Appl Environ Microbiol* 2007, **73**(23):7531-7535.
30. Chen DQ, B. MC, H. PA: **Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid, *A. kondoi***. *Entomol Exp Appl* 2000, **95**(3):315-323.
31. Montllor CB, Maxmen A, Purcell AH: **Facultative bacterial endosymbionts benefit pea aphids *Acyrthosiphon pisum* under heat stress**. *Ecol Entomol* 2002, **27**(2):189-195.
32. Scarborough CL, Ferrari J, Godfray HC: **Aphid protected from pathogen by endosymbiont**. *Science* 2005, **310**(5755):1781.
33. Murray RG, Schleifer KH: **Taxonomic notes: a proposal for recording the properties of putative taxa of procaryotes**. *Int J Syst Bacteriol* 1994, **44**(1):174-176.
34. Murray RG, Stackebrandt E: **Taxonomic note: implementation of the provisional status *Candidatus* for incompletely described procaryotes**. *Int J Syst Bacteriol* 1995, **45**(1):186-187.
35. Lamelas A, Gosalbes MJ, Manzano-Marin A, Pereto J, Moya A, Latorre A: ***Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont**. *Plos Genet* 2011, **7**(11):e1002357.
36. McCutcheon JP, von Dohlen CD: **An interdependent metabolic patchwork in the nested symbiosis of mealybugs**. *Curr Biol* 2011, **21**(16):1366-1372.

37. Khachane AN, Timmis KN, Martins dos Santos VA: **Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes**. *Mol Biol Evol* 2007, **24**(2):449-456.
38. 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 R Soc Lond B* 2003, **270**(1533):2543-2550.
39. Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A: **A small microbial genome: the end of a long symbiotic relationship?** *Science* 2006, **314**(5797):312-313.
40. Lefevre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A: **Endosymbiont phylogenesis in the Dryophthoridae weevils: evidence for bacterial replacement**. *Mol Biol Evol* 2004, **21**(6):965-973.
41. Conord C, Despres L, Vallier A, Balmand S, Miquel C, Zundel S, Lemperiere G, Heddi A: **Long-term evolutionary stability of bacterial endosymbiosis in Curculionoidea: additional evidence of symbiont replacement in the Dryophthoridae family**. *Mol Biol Evol* 2008, **25**(5):859-868.
42. Gruwell ME, Hardy NB, Gullan PJ, Dittmar K: **Evolutionary relationships among primary endosymbionts in the mealybug subfamily Phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae)**. *Appl Environ Microbiol* 2010, **76**(22):7521-7525.
43. McCutcheon JP, Moran NA: **Parallel genomic evolution and metabolic interdependence in an ancient symbiosis**. *Proc Natl Acad Sci U S A* 2007, **104**(49):19392-19397.
44. McCutcheon JP, McDonald BR, Moran NA: **Convergent evolution of metabolic roles in bacterial co-symbionts of insects**. *Proc Natl Acad Sci U S A* 2009, **106**(36):15394–15399.
45. McCutcheon JP, Moran NA: **Functional convergence in reduced genomes of bacterial symbionts spanning 200 million years of evolution**. *Genome Biol Evol* 2010, **2**:708-718.
46. Takiya DM, Tran PL, Dietrich CH, Moran NA: **Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts**. *Mol Ecol* 2006, **15**(13):4175-4191.
47. Buchner P: **Endosymbiosis of animals with plant microorganisms**. New York: Interscience Publishers; 1965.
48. Gonella E, Negri I, Marzorati M, Mandrioli M, Sacchi L, Pajoro M, Crotti E, Rizzi A, Clementi E, Tedeschi R, Bandi C, Alma A, Daffonchio D: **Bacterial endosymbiont localization in *Hyalesthes obsoletus*, the insect vector of Bois Noir in *Vitis vinifera***. *Appl Environ Microbiol* 2010, **77**(4):1423-1435.
49. Bressan A, Arneodo J, Simonato M, Haines WP, Boudon-Padieu E: **Characterization and evolution of two bacteriome-inhabiting symbionts in cixiid planthoppers (Hemiptera: Fulgoromorpha: Pentastirini)**. *Environ Microbiol* 2009, **11**(12):3265-3279.
50. Matsuura Y, Kikuchi Y, Hosokawa T, Koga R, Meng XY, Kamagata Y, Nikoh N, Fukatsu T: **Evolution of symbiotic organs and endosymbionts in lygaeid stinkbugs**. *ISME J* 2012, **6**(2):397-409.
51. Gottlieb Y, Ghanim M, Gueguen G, Kontsedalov S, Vavre F, Fleury F, Zchori-Fein E: **Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies**. *FASEB J* 2008, **22**(7):2591-2599.
52. von Dohlen CD, Kohler S, Alsop ST, McManus WR: **Mealybug β-proteobacterial**

endosymbionts contain γ-proteobacterial symbionts. *Nature* 2001, **412**(6845):433-436.

53. Perotti MA, Clarke HK, Turner BD, Braig HR: ***Rickettsia* as obligate and mycetomic bacteria**. *FASEB J* 2006, **20**(13):1646-1656.
54. Sassera D, Beninati T, Bandi C, Bouman EAP, Sacchi L, Fabbi M, Lo N: **'*Candidatus* Midichloria mitochondrii', an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle**. *Int J Syst Evol Microbiol* 2006, **56**:2535-2540.
55. Cho KO, Kim GW, Lee OK: ***Wolbachia* bacteria reside in host Golgi-related vesicles whose position is regulated by polarity proteins**. *Plos One* 2011, **6**(7):e22703.
56. Voronin D, Dudkina N, Kiseleva EV: **A new form of symbiotic bacteria *Wolbachia* found in the endoplasmic reticulum of early embryos of *Drosophila melanogaster*.** *Dokl Biol Sci* 2004, **396**(4):227-229.
57. Rio RV, Symula RE, Wang J, Lohs C, Wu YN, Snyder AK, Bjornson RD, Oshima K, Biehl BS, Perna NT, Hattori M, Aksoy S: **Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia***. *mBio* 2012, **3**(1):e00240-00211.
58. Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH: **How many species are infected with *Wolbachia*? - a statistical analysis of current data**. *FEMS Microbiol Lett* 2008, **281**(2):215-220.
59. Reuter M, Keller L: **High levels of multiple *Wolbachia* infection and recombination in the ant *Formica exsecta***. *Mol Biol Evol* 2003, **20**(5):748-753.
60. Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T: ***Wolbachia* as a bacteriocyte-associated nutritional mutualist**. *Proc Natl Acad Sci U S A* 2010, **107**(2):769-774.
61. Pannebakker BA, Loppin B, Elemans CP, Humblot L, Vavre F: **Parasitic inhibition of cell death facilitates symbiosis**. *Proc Natl Acad Sci U S A* 2007, **104**(1):213-215.
62. Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Bouletreau M: **Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp**. *Proc Natl Acad Sci U S A* 2001, **98**(11):6247-6252.
63. Taylor MJ, Bandi C, Hoerauf A: ***Wolbachia* bacterial endosymbionts of filarial nematodes**. *Adv Parasitol* 2005, **60**:245-284.
64. Aanen DK, Hoekstra RF: **The evolution of obligate mutualism: if you can't beat 'em, join 'em**. *Trends Ecol Evol* 2007, **22**(10):506-509.
65. Saridaki A, Bourtzis K: ***Wolbachia*: more than just a bug in insects genitals**. *Curr Opin Microbiol* 2010, **13**(1):67-72.
66. Bordenstein SR, Paraskevopoulos C, Hotopp JC, Sapountzis P, Lo N, Bandi C, Tettelin H, Werren JH, Bourtzis K: **Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say**. *Mol Biol Evol* 2009, **26**(1):231-241.
67. Weeks AR, Turelli M, Harcombe WR, Reynolds KT, Hoffmann AA: **From parasite to mutualist: rapid evolution of *Wolbachia* in natural populations of *Drosophila***. *PLoS Biol* 2007, **5**(5):e114.
68. Chen DQ, Campbell BC, Purcell AH: **A new *Rickettsia* from a herbivorous insect, the pea aphid *Acyrthosiphon pisum* (Harris)**. *Curr Microbiol* 1996, **33**(2):123-128.
69. Perlman SJ, Hunter MS, Zchori-Fein E: **The emerging diversity of *Rickettsia***. *Proc R Soc B* 2006, **273**(1598):2097-2106.
70. Himler AG, Adachi-Hagimori T, Bergen JE, Kozuch A, Kelly SE, Tabashnik BE, Chiel E, Duckworth VE, Dennehy TJ, Zchori-Fein E, Hunter MS: **Rapid spread of a bacterial symbiont in an invasive whitefly is driven by fitness benefits and female bias**. *Science* 2011, **332**(6026):254-256.

71. Sassera D, Lo N, Epis S, D'Auria G, Montagna M, Comandatore F, Horner D, Pereto J, Luciano AM, Franciosi F, Ferri E, Crotti E, Bazzocchi C, Daffonchio D, Sacchi L, Moya A, Latorre A, Bandi C: **Phylogenomic evidence for the presence of a flagellum and cbb3 oxidase in the free-living mitochondrial ancestor**. *Mol Biol Evol* 2011, **28**(12):3285-3296.
72. Epis S, Sassera D, Beninati T, Lo N, Beati L, Piesman J, Rinaldi L, McCoy KD, Torina A, Sacchi L, Clementi E, Genchi M, Magnino S, Bandi C: ***Midichloria mitochondrii* is widespread in hard ticks (Ixodidae) and resides in the mitochondria of phylogenetically diverse species**. *Parasitology* 2008, **135**(4):485-494.
73. Sassera D, Lo N, Bouman EA, Epis S, Mortarino M, Bandi C: **"*Candidatus* Midichloria" endosymbionts bloom after the blood meal of the host, the hard tick *Ixodes ricinus***. *Appl Environ Microbiol* 2008, **74**(19):6138-6140.
74. Lara FA, Lins U, Bechara GH, Oliveira PL: **Tracing heme in a living cell: hemoglobin degradation and heme traffic in digest cells of the cattle tick *Boophilus microplus***. *J Exp Biol* 2005, **208**(Pt 16):3093-3101.
75. Gillespie JJ, Joardar V, Williams KP, Driscoll T, Hostetler JB, Nordberg E, Shukla M, Walenz B, Hill CA, Nene VM, Azad AF, Sobral BW, Caler E: **A *Rickettsia* genome overrun by mobile genetic elements provides insight into the acquisition of genes characteristic of an obligate intracellular lifestyle**. *J Bacteriol* 2012, **194**(2):376-394.
76. Matsuura Y, Kikuchi Y, Meng XY, Koga R, Fukatsu T: **A novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods**. *Appl Environ Microbiol* 2012, **[Epub ahead of print]**.
77. Stoll S, Gadau J, Gross R, Feldhaar H: **Bacterial microbiota associated with ants of the genus *Tetraponera***. *Biol J Linn Soc* 2007, **90**(3):399-412.
78. Zientz E, Feldhaar H, Stoll S, Gross R: **Insights into the microbial world associated with ants**. *Arch Microbiol* 2005, **184**(4):199-206.
79. Halos L, Jamal T, Maillard R, Girard B, Guillot J, Chomel B, Vayssier-Taussat M, Boulouis HJ: **Role of Hippoboscidae flies as potential vectors of *Bartonella* spp. infecting wild and domestic ruminants**. *Appl Environ Microbiol* 2004, **70**(10):6302-6305.
80. McCutcheon JP, McDonald BR, Moran NA: **Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont**. *Plos Genet* 2009, **5**(7):e1000565.
81. Hershberg R, Petrov DA: **Evidence that mutation is universally biased towards AT in bacteria**. *Plos Genet* 2010, **6**(9):e1001115.
82. Hildebrand F, Meyer A, Eyre-Walker A: **Evidence of selection upon genomic GC-content in bacteria**. *Plos Genet* 2010, **6**(9):e1001107.
83. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M: **The 160-kilobase genome of the bacterial endosymbiont *Carsonella***. *Science* 2006, **314**(5797):267.
84. Van Leuven JT, McCutcheon JP: **An AT mutational bias in the tiny GC-rich endosymbiont genome of *Hodgkinia***. *Genome Biol Evol* 2012, **4**(1):24-27.
85. Baumann L, Thao ML, Hess JM, Johnson MW, Baumann P: **The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects**. *Appl Environ Microbiol* 2002, **68**(7):3198-3205.
86. Kono M, Koga R, Shimada M, Fukatsu T: **Infection dynamics of coexisting Beta- and Gammaproteobacteria in the nested endosymbiotic system of mealybugs**. *Appl Environ Microbiol* 2008, **74**(13):4175-4184.

87. Lopez-Madrigal S, Latorre A, Porcar M, Moya A, Gil R: **Complete genome sequence of "*Candidatus* Tremblaya princeps" strain PCVAL, an intriguing translational machine below the living-cell status**. *J Bacteriol* 2011, **193**(19):5587-5588.
88. Leclerque A: **Whole genome-based assessment of the taxonomic position of the arthropod pathogenic bacterium *Rickettsiella grylli***. *FEMS Microbiol Lett* 2008, **283**(1):117-127.
89. Cordaux R, Paces-Fessy M, Raimond M, Michel-Salzat A, Zimmer M, Bouchon D: **Molecular characterization and evolution of arthropod-pathogenic *Rickettsiella* bacteria**. *Appl Environ Microbiol* 2007, **73**(15):5045-5047.
90. Tsuchida T, Koga R, Horikawa M, Tsunoda T, Maoka T, Matsumoto S, Simon JC, Fukatsu T: **Symbiotic bacterium modifies aphid body color**. *Science* 2010, **330**(6007):1102-1104.
91. Mediannikov O, Sekeyova Z, Birg ML, Raoult D: **A novel obligate intracellular gamma-proteobacterium associated with ixodid ticks, *Diplorickettsia massiliensis*, gen. nov., sp nov.** *Plos One* 2010, **5**(7):e11478.
92. Hypša V, Křížek J: **Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura)**. *Microb Ecol* 2007, **54**(2):242-251.
93. Volf P: **Postembryonal development of mycetocytes and symbionts of the spiny rat louse *Polyplax spinulosa***. *J Invertebr Pathol* 1991, **58**(1):143-146.
94. Spaulding AW, von Dohlen CD: **Phylogenetic characterization and molecular evolution of bacterial endosymbionts in psyllids (Hemiptera: Sternorrhyncha)**. *Mol Biol Evol* 1998, **15**(11):1506-1513.
95. Thao ML, Baumann P: **Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts**. *Appl Environ Microbiol* 2004, **70**(6):3401-3406.
96. Williams KP, Gillespie JJ, Sobral BW, Nordberg EK, Snyder EE, Shallom JM, Dickerman AW: **Phylogeny of Gammaproteobacteria**. *J Bacteriol* 2010, **192**(9):2305-2314.
97. Kuchler SM, Dettner K, Kehl S: **Molecular characterization and localization of the obligate endosymbiotic bacterium in the birch catkin bug *Kleidocerys resedae* (Heteroptera: *Lygaeidae*, Ischnorhynchinae)**. *FEMS Microbiol Ecol* 2010, **73**(2):408-418.
98. Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Hironaka M, Fukatsu T: **Phylogenetic position and peculiar genetic traits of a midgut bacterial symbiont of the stinkbug *Parastrachia japonensis***. *Appl Environ Microbiol* 2010, **76**(13):4130-4135.
99. Kuechler SM, Dettner K, Kehl S: **Characterization of an obligate intracellular bacterium in the midgut epithelium of the bulrush bug *Chilacis typhae* (Heteroptera, Lygaeidae, Artheneinae)**. *Appl Environ Microbiol* 2011, **77**(9):2869-2876.
100. Kikuchi Y, Hosokawa T, Nikoh N, Meng XY, Kamagata Y, Fukatsu T: **Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs**. *BMC Biol* 2009, **7**:2.
101. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M, Fukatsu T: **Primary gut symbiont and secondary *Sodalis*-allied symbiont in the scutellerid stinkbug *Cantao ocellatus***. *Appl Environ Microbiol* 2010, **76**(11):3486-3494.
102. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M, Fukatsu T: **Bacterial symbionts of the giant jewel stinkbug *Eucorysses grandis* (Hemiptera:**

Scutelleridae). *Zool Sci* 2011, **28**(3):169-174.
103. Toju H, Hosokawa T, Koga R, Nikoh N, Meng XY, Kimura N, Fukatsu T: **"*Candidatus* Curculioniphilus buchneri", a novel clade of bacterial endocellular symbionts from weevils of the genus *Curculio***. *Appl Environ Microbiol* 2009, **76**(1):275-282.
104. Hosokawa T, Fukatsu T: ***Nardonella* endosymbiont in the West Indian sweet potato weevil *Euscepes postfasciatus* (Coleoptera: Curculionidae)**. *Appl Entomol Zool* 2010, **45**(1):115-120.
105. Kuriwada T, Hosokawa T, Kumano N, Shiromoto K, Haraguchi D, Fukatsu T: **Biological role of *Nardonella* endosymbiont in its weevil host**. *Plos One* 2010, **5**(10):e13101.
106. Kolsch G, Matz-Grund C, Pedersen BV: **Ultrastructural and molecular characterization of endosymbionts of the reed beetle genus *Macroplea* (Chrysomelidae, Donaciinae), and proposal of "*Candidatus* Macropleicola appendiculatae" and "*Candidatus* Macropleicola muticae"**. *Can J Microbiol* 2009, **55**(11):1250-1260.
107. Kolsch G, Pedersen BV: **Can the tight co-speciation between reed beetles (Col., Chrysomelidae, Donaciinae) and their bacterial endosymbionts, which provide cocoon material, clarify the deeper phylogeny of the hosts?** *Mol Phylogenet Evol* 2010, **54**(3):810-821.
108. Fukatsu T, Hosokawa T, Koga R, Nikoh N, Kato T, Hayama S, Takefushi H, Tanaka I: **Intestinal endocellular symbiotic bacterium of the macaque louse *Pedicinus obtusus*: Distinct endosymbiont origins in anthropoid primate lice and the old world monkey louse**. *Appl Environ Microbiol* 2009, **75**(11):3796-3799.
109. Duron O, Bouchon D, Boutin S, Bellamy L, Zhou LQ, Engelstadter J, Hurst GD: **The diversity of reproductive parasites among arthropods: *Wolbachia* do not walk alone**. *BMC Biol* 2008, **6**:27.
110. Zchori-Fein E, Perlman SJ: **Distribution of the bacterial symbiont *Cardinium* in arthropods**. *Mol Ecol* 2004, **13**(7):2009-2016.
111. Zchori-Fein E, Perlman SJ, Kelly SE, Katzir N, Hunter MS: **Characterization of a 'Bacteroidetes' symbiont in *Encarsia wasps* (Hymenoptera : Aphelinidae): proposal of '*Candidatus* Cardinium hertigii'**. *Int J Syst Evol Microbiol* 2004, **54**:961-968.
112. Hurst GDD, Hammarton TC, Bandi C, Majerus TMO, Bertrand D, Majerus MEN: **The diversity of inherited parasites of insects: the male-killing agent of the ladybird beetle *Coleomegilla maculata* is a member of the Flavobacteria**. *Genet Res* 1997, **70**(1):1-6.
113. Hurst GD, Bandi C, Sacchi L, Cochrane AG, Bertrand D, Karaca I, Majerus MEN: ***Adonia variegata* (Coleoptera : Coccinellidae) bears maternally inherited Flavobacteria that kill males only**. *Parasitology* 1999, **118**:125-134.
114. Gruwell ME, Morse GE, Normark BB: **Phylogenetic congruence of armored scale insects (Hemiptera: Diaspididae) and their primary endosymbionts from the phylum Bacteroidetes**. *Mol Phylogenet Evol* 2007, **44**(1):267-280.
115. Gruwell ME, Flarhety M, Dittmar K: **Distribution of the primary endosymbiont (*Candidatus* Uzinura Diaspidicola) within host insects from the scale insect family Diaspididae**. *Insects* 2012, **3**(1):262-269.
116. Lopez-Sanchez MJ, Neef A, Pereto J, Patino-Navarrete R, Pignatelli M, Latorre A, Moya A: **Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica***. *Plos Genet* 2009, **5**(11):e1000721.
117. Sabree ZL, Kambhampati S, Moran NA: **Nitrogen recycling and nutritional**

provisioning by *Blattabacterium*, the cockroach endosymbiont. *Proc Natl Acad Sci U S A* 2009, **106**(49):19521-19526.
118. Neef A, Latorre A, Pereto J, Silva FJ, Pignatelli M, Moya A: **Genome economization in the endosymbiont of the wood roach *Cryptocercus punctulatus* due to drastic loss of amino acid synthesis capabilities**. *Genome Biol Evol* 2011, **3**:1437-1448.
119. Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA: **Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis***. *Appl Environ Microbiol* 2011, **78**(9):204-210.
120. Moran NA, Tran P, Gerardo NM: **Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes**. *Appl Environ Microbiol* 2005, **71**(12):8802-8810.
121. 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: **Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters**. *PLoS Biol* 2006, **4**(6):e188.
122. Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W, Lapidus A, Wu D, McCutcheon JP, McDonald BR, Moran NA, Bristow J, Cheng JF: **One bacterial cell, one complete genome**. *Plos One* 2010, **5**(4):e10314.
123. Anbutsu H, Fukatsu T: ***Spiroplasma* as a model insect endosymbiont**. *Env Microbiol Rep* 2011, **3**(2):144-153.
124. Jaenike J, Unckless R, Cockburn SN, Boelio LM, Perlman SJ: **Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont**. *Science* 2010, **329**(5988):212-215.
125. Moya-Raygoza G, Palomera-Avalos V, Galaviz-Mejia C: **Field overwintering biology of *Spiroplasma kunkelii* (Mycoplasmatales : Spiroplasmataceae) and its vector *Dalbulus maidis* (Hemiptera : Cicadellidae)**. *Ann Appl Biol* 2007, **151**(3):373-379.
126. Xie J, Vilchez I, Mateos M: **Spiroplasma bacteria enhance survival of *Drosophila hydei* attacked by the parasitic wasp *Leptopilina heterotoma***. *Plos One* 2010, **5**(8):e12149.
127. Thao ML, Baumann L, Hess JM, Falk BW, Ng JC, Gullan PJ, Baumann P: **Phylogenetic evidence for two new insect-associated Chlamydia of the family *Simkaniaceae***. *Curr Microbiol* 2003, **47**(1):46-50.
128. Everett KD, Thao M, Horn M, Dyszynski GE, Baumann P: **Novel chlamydiae in whiteflies and scale insects: endosymbionts '*Candidatus* Fritschea bemisiae' strain Falk and '*Candidatus* Fritschea eriococci' strain Elm**. *Int J Syst Evol Microbiol* 2005, **55**(Pt 4):1581-1587.
129. Hansen AK, Moran NA: **Aphid genome expression reveals host-symbiont cooperation in the production of amino acids**. *Proc Natl Acad Sci U S A* 2011, **108**(7):2849–2854.
130. Poliakov A, Russell CW, Ponnala L, Hoops HJ, Sun Q, Douglas AE, van Wijk KJ: **Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis**. *Mol Cell Proteomics* 2011, **10**(6):M110.007039.
131. International Aphid Genomics C: **Genome sequence of the pea aphid Acyrthosiphon pisum**. *PLoS Biol* 2010, **8**(2):e1000313.
132. Keeling PJ, Palmer JD: **Horizontal gene transfer in eukaryotic evolution**. *Nat Rev Genet* 2008, **9**(8):605-618.
133. Kirkness EF, Haas BJ, Sun WL, Braig HR, Perotti MA, Clark JM, Lee SH, Robertson

HM, Kennedy RC, Elhaik E, Gerlach D, Kriventseva EV, Elsik CG, Graur D, Hill CA, Veenstra JA, Walenz B, Tubio JMC, Ribeiro JMC, Rozas J, Johnston JS, Reese JT, Popadic A, Tojo M, Raoult D, Reed DL, Tomoyasu Y, Krause E, Mittapalli O, Margam VM, Li HM, Meyer JM, Johnson RM, Romero-Severson J, VanZee JP, Alvarez-Ponce D, Vieira FG, Aguade M, Guirao-Rico S, Anzola JM, Yoon KS, Strycharz JP, Unger MF, Christley S, Lobo NF, Seufferheld MJ, Wang NK, Dasch GA, Struchiner CJ, Madey G, Hannick LI, Bidwell S, Joardar V, Caler E, Shao RF, Barker SC, Cameron S, Bruggner RV, Regier A, Johnson J, Viswanathan L, Utterback TR, Sutton GG, Lawson D, Waterhouse RM, Venter JC, Strausberg RL, Berenbaum MR, Collins FH, Zdobnov EM, Pittendrigh BR: **Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle**. *Proc Natl Acad Sci U S A* 2010, **107**(27):12168-12173.

134. Nikoh N, McCutcheon JP, Kudo T, Miyagishima SY, Moran NA, Nakabachi A: **Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host**. *Plos Genet* 2010, **6**(2):e1000827.
135. Nikoh N, Nakabachi A: **Aphids acquired symbiotic genes via lateral gene transfer**. *BMC Biol* 2009, **7**:12.
136. McNulty SN, Foster JM, Mitreva M, Dunning Hotopp JC, Martin J, Fischer K, Wu B, Davis PJ, Kumar S, Brattig NW, Slatko BE, Weil GJ, Fischer PU: **Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer**. *Plos One* 2010, **5**(6):e11029.
137. Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T: **Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect**. *Proc Natl Acad Sci U S A* 2002, **99**(22):14280-14285.
138. Nikoh N, Tanaka K, Shibata F, Kondo N, Hizume M, Shimada M, Fukatsu T: ***Wolbachia* genome integrated in an insect chromosome: Evolution and fate of laterally transferred endosymbiont genes**. *Genome Res* 2008, **18**(2):272-280.
139. Aikawa T, Anbutsu H, Nikoh N, Kikuchi T, Shibata F, Fukatsu T: **Longicorn beetle that vectors pinewood nematode carries many *Wolbachia* genes on an autosome**. *P R Soc B* 2009, **276**(1674):3791-3798.
140. Hotopp JCD, Clark ME, Oliveira DCSG, Foster JM, Fischer P, Torres MC, Giebel JD, Kumar N, Ishmael N, Wang SL, Ingram J, Nene RV, Shepard J, Tomkins J, Richards S, Spiro DJ, Ghedin E, Slatko BE, Tettelin H, Werren JH: **Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes**. *Science* 2007, **317**(5845):1753-1756.
141. Klasson L, Kambris Z, Cook PE, Walker T, Sinkins SP: **Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti***. *BMC Genomics* 2009, **10**:33.
142. Woolfit M, Iturbe-Ormaetxe I, McGraw EA, O'Neill SL: **An ancient horizontal gene transfer between mosquito and the endosymbiotic bacterium *Wolbachia pipientis***. *Mol Biol Evol* 2009, **26**(2):367-374.
143. Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Nasonia Genome Working G, Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Beukeboom LW, Desplan C, Elsik CG, Grimmelikhuijzen CJ, Kitts P, Lynch JA, Murphy T, Oliveira DC, Smith CD, van de Zande L, Worley KC, Zdobnov EM, Aerts M, Albert S, Anaya VH, Anzola JM, Barchuk AR, Behura SK, Bera AN, Berenbaum MR, Bertossa RC, Bitondi MM, Bordenstein SR, Bork P, Bornberg-Bauer E, Brunain M, Cazzamali G, Chaboub L, Chacko J, Chavez D, Childers CP, Choi JH, Clark ME, Claudianos C, Clinton RA, Cree AG, Cristino AS, Dang PM, Darby AC, de Graaf

DC, Devreese B, Dinh HH, Edwards R, Elango N, Elhaik E, Ermolaeva O, Evans JD, Foret S, Fowler GR, Gerlach D, Gibson JD, Gilbert DG, Graur D, Grunder S, Hagen DE, Han Y, Hauser F, Hultmark D, Hunter HCt, Hurst GD, Jhangian SN, Jiang H, Johnson RM, Jones AK, Junier T, Kadowaki T, Kamping A, Kapustin Y, Kechavarzi B, Kim J, Kim J, Kiryutin B, Koevoets T, Kovar CL, Kriventseva EV, Kucharski R, Lee H, Lee SL, Lees K, Lewis LR, Loehlin DW, Logsdon JM, Jr., Lopez JA, Lozado RJ, Maglott D, Maleszka R, Mayampurath A, Mazur DJ, McClure MA, Moore AD, Morgan MB, Muller J, Munoz-Torres MC, Muzny DM, Nazareth LV, Neupert S, Nguyen NB, Nunes FM, Oakeshott JG, Okwuonu GO, Pannebakker BA, Pejaver VR, Peng Z, Pratt SC, Predel R, Pu LL, Ranson H, Raychoudhury R, Rechtsteiner A, Reese JT, Reid JG, Riddle M, Robertson HM, Romero-Severson J, Rosenberg M, Sackton TB, Sattelle DB, Schluns H, Schmitt T, Schneider M, Schuler A, Schurko AM, Shuker DM, Simoes ZL, Sinha S, Smith Z, Solovyev V, Souvorov A, Springauf A, Stafflinger E, Stage DE, Stanke M, Tanaka Y, Telschow A, Trent C, Vattathil S, Verhulst EC, Viljakainen L, Wanner KW, Waterhouse RM, Whitfield JB, Wilkes TE, Williamson M, Willis JH, Wolschin F, Wyder S, Yamada T, Yi SV, Zecher CN, Zhang L, Gibbs RA: **Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species**. *Science* 2010, **327**(5963):343-348.

144. Fares MA, Ruiz-Gonzalez MX, Moya A, Elena SF, Barrio E: **GroEL buffers against deleterious mutations**. *Nature* 2002, **417**(6887):398-398.

145. Moran NA: **Accelerated evolution and Muller's rachet in endosymbiotic bacteria**. *Proc Natl Acad Sci U S A* 1996, **93**(7):2873-2878.

146. Rispe C, Moran NA: **Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection**. *Am Nat* 2000, **156**(4):425-441.

147. Pettersson ME, Berg OG: **Muller's ratchet in symbiont populations**. *Genetica* 2007, **130**(2):199-211.

148. Allen JM, Light JE, Perotti MA, Braig HR, Reed DL: **Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet**. *Plos One* 2009, **4**(3):e4969.

149. 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: **Phylogenomics of the reproductive parasite *Wolbachia pipientis w*Mel: A streamlined genome overrun by mobile genetic elements**. *PLoS Biol* 2004, **2**(3):327-341.

150. 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: **Genome evolution of *Wolbachia* strain *w*Pip from the *Culex pipiens* group**. *Mol Biol Evol* 2008, **25**(9):1877-1887.

151. Klasson L, Westberg J, Sapountzis P, Naslund K, Lutnaes Y, Darby AC, Veneti Z, Chen L, Braig HR, Garrett R, Bourtzis K, Andersson SG: **The mosaic genome structure of the *Wolbachia w*Ri strain infecting *Drosophila simulans***. *Proc Natl Acad Sci U S A* 2009, **106**(14):5725-5730.

152. Darby AC, Choi JH, Wilkes T, Hughes MA, Werren JH, Hurst GD, Colbourne JK: **Characteristics of the genome of *Arsenophonus nasoniae*, son-killer bacterium of the wasp *Nasonia***. *Insect Mol Biol* 2010, **19**(Sp. Iss. SI Suppl. 1):75-89.

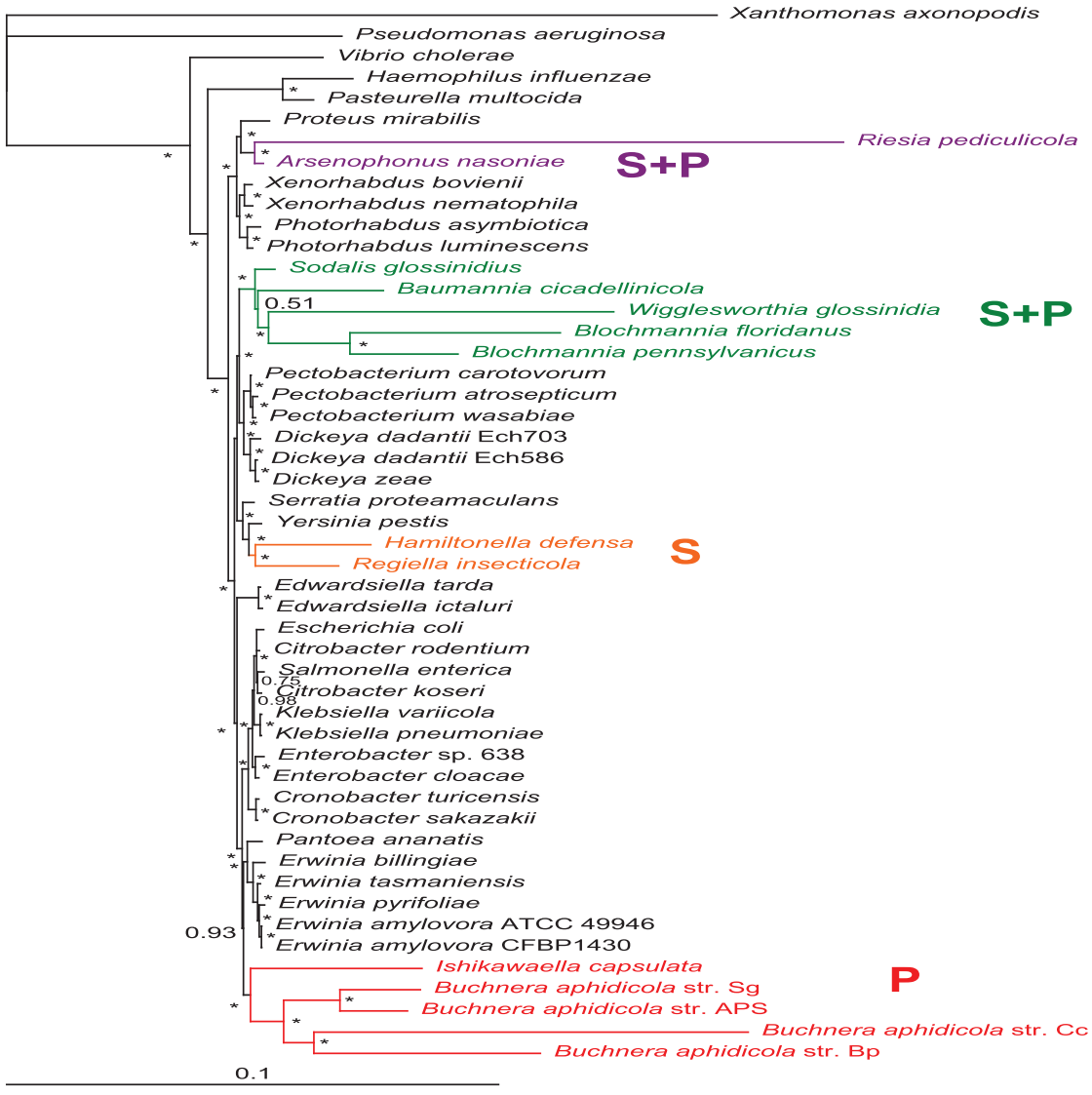
153. Wilkes TE, Darby AC, Choi JH, Colbourne JK, Werren JH, Hurst GD: **The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia***

***vitripennis*, reveals genes associated with virulence and symbiosis**. *Insect Mol Biol* 2010, **19**(Sp. Iss. SI Suppl. 1):59-73.

154. Gil R, Silva FJ, Zientz E, Delmotte F, Gonzalez-Candelas F, Latorre A, Rausell C, Kamerbeek J, Gadau J, Holldobler B, van Ham RC, Gross R, Moya A: **The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes**. *Proc Natl Acad Sci U S A* 2003, **100**(16):9388-9393.
155. Degnan PH, Lazarus AB, Wernegreen JJ: **Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects**. *Genome Res* 2005, **15**(8):1023-1033.
156. Williams LE, Wernegreen JJ: **Unprecedented loss of ammonia assimilation capability in a urease-encoding bacterial mutualist**. *BMC Genomics* 2010, **11**:687.
157. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H: **Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS**. *Nature* 2000, **407**(6800):81-86.
158. Degnan PH, Ochman H, Moran NA: **Sequence conservation and functional constraint on intergenic spacers in reduced genomes of the obligate symbiont *Buchnera***. *Plos Genet* 2011, **7**(9):e1002252.
159. 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: **Reductive genome evolution in *Buchnera aphidicola***. *Proc Natl Acad Sci U S A* 2003, **100**(2):581-586.
160. Latorre A, Lamelas A, Gosalbes MJ, Moya A: **New clues about the evolutionary history of metabolic losses in bacterial endosymbionts, provided by the genome of *Buchnera aphidicola* from the aphid *Cinara tujafilina***. *Appl Environ Microbiol* 2011, **77**(13):4446-4454.
161. Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SG: **50 million years of genomic stasis in endosymbiotic bacteria**. *Science* 2002, **296**(5577):2376-2379.
162. Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA: ***Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors**. *Proc Natl Acad Sci U S A* 2009, **106**(22):9063-9068.
163. Degnan PH, Leonardo TE, Cass BN, Hurwitz B, Stern D, Gibbs RA, Richards S, Moran NA: **Dynamics of genome evolution in facultative symbionts of aphids**. *Environ Microbiol* 2009, **12**(8):2060–2069.
164. Hansen AK, Vorburger C, Moran NA: **Genomic basis of endosymbiont-conferred protection against an insect parasitoid**. *Genome Res* 2011, **22**(1):106-114.
165. Burke GR, Moran NA: **Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids**. *Genome Biol Evol* 2011, **3**:195-208.
166. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S: **Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host**. *Genome Res* 2006, **16**(2):149-156.
167. Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M, Aksoy S: **Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia***. *Nat Genet* 2002, **32**(3):402-407.
168. Nakabachi A, Shigenobu S, Sakazume N, Shiraki T, Hayashizaki Y, Carninci P, Ishikawa H, Kudo T, Fukatsu T: **Transcriptome analysis of the aphid bacteriocyte, the symbiotic host cell that harbors an endocellular mutualistic bacterium, *Buchnera***. *Proc Natl Acad Sci U S A* 2005, **102**(15):5477-5482.

169. Maezawa K, Shigenobu S, Taniguchi H, Kubo T, Aizawa S, Morioka M: **Hundreds of flagellar basal bodies cover the cell surface of the endosymbiotic bacterium *Buchnera aphidicola* sp. strain APS**. *J Bacteriol* 2006, **188**(18):6539-6543.
170. Burke GR, Moran NA: **Responses of the pea aphid transcriptome to infection by facultative symbionts**. *Insect Mol Biol* 2011, **20**(3):357-365.
171. Ramsey JS, MacDonald SJ, Jander G, Nakabachi A, Thomas GH, Douglas AE: **Genomic evidence for complementary purine metabolism in the pea aphid, *Acyrthosiphon pisum*, and its symbiotic bacterium *Buchnera aphidicola***. *Insect Mol Biol* 2010, **19** (Suppl 2):241-248.
172. Wilson AC, Ashton PD, Calevro F, Charles H, Colella S, Febvay G, Jander G, Kushlan PF, Macdonald SJ, Schwartz JF, Thomas GH, Douglas AE: **Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola***. *Insect Mol Biol* 2010, **19**(Suppl 2):249-258.
173. Gerardo NM, Wilson AC: **The power of paired genomes**. *Mol Ecol* 2011, **20**(10):2038-2040.
174. Charles H, Balmand S, Lamelas A, Cottret L, Perez-Brocal V, Burdin B, Latorre A, Febvay G, Colella S, Calevro F, Rahbe Y: **A genomic reappraisal of symbiotic function in the aphid/*Buchnera symbiosis*: reduced transporter sets and variable membrane organisations**. *Plos One* 2011, **6**(12):e29096.
175. Price DRG, Duncan RP, Shigenobu S, Wilson ACC: **Genome expansion and differential expression of amino acid transporters at the aphid/*Buchnera* symbiotic interface**. *Mol Biol Evol* 2011, **28**(11):3113-3126.
176. Anselme C, Perez-Brocal V, Vallier A, Vincent-Monegat C, Charif D, Latorre A, Moya A, Heddi A: **Identification of the Weevil immune genes and their expression in the bacteriome tissue**. *BMC Biol* 2008, **6**:43.
177. Vallier A, Vincent-Monegat C, Laurencon A, Heddi A: **RNAi in the cereal weevil *Sitophilus* spp: systemic gene knockdown in the bacteriome tissue**. *BMC Biotechnol* 2009, **9**:44.
178. Login FH, Balmand S, Vallier A, Vincent-Monegat C, Vigneron A, Weiss-Gayet M, Rochat D, Heddi A: **Antimicrobial peptides keep insect endosymbionts under control**. *Science* 2011, **334**(6054):362-365.

BMC Biology
incorporating Journal of Biology



Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches

Husník *et al.*

BMC Biology
incorporating Journal of Biology

RESEARCH ARTICLE Open Access

Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches

Filip Husník[1*], Tomáš Chrudimský[1] and Václav Hypša[1,2*]

Abstract

Background: The bacterial family Enterobacteriaceae gave rise to a variety of symbiotic forms, from the loosely associated commensals, often designated as secondary (S) symbionts, to obligate mutualists, called primary (P) symbionts. Determination of the evolutionary processes behind this phenomenon has long been hampered by the unreliability of phylogenetic reconstructions within this group of bacteria. The main reasons have been the absence of sufficient data, the highly derived nature of the symbiont genomes and lack of appropriate phylogenetic methods. Due to the extremely aberrant nature of their DNA, the symbiotic lineages within Enterobacteriaceae form long branches and tend to cluster as a monophyletic group. This state of phylogenetic uncertainty is now improving with an increasing number of complete bacterial genomes and development of new methods. In this study, we address the monophyly versus polyphyly of enterobacterial symbionts by exploring a multigene matrix within a complex phylogenetic framework.

Results: We assembled the richest taxon sampling of Enterobacteriaceae to date (50 taxa, 69 orthologous genes with no missing data) and analyzed both nucleic and amino acid data sets using several probabilistic methods. We particularly focused on the long-branch attraction-reducing methods, such as a nucleotide and amino acid data recoding and exclusion (including our new approach and slow-fast analysis), taxa exclusion and usage of complex evolutionary models, such as nonhomogeneous model and models accounting for site-specific features of protein evolution (CAT and CAT+GTR). Our data strongly suggest independent origins of four symbiotic clusters; the first is formed by *Hamiltonella* and *Regiella* (S-symbionts) placed as a sister clade to *Yersinia*, the second comprises *Arsenophonus* and *Riesia* (S- and P-symbionts) as a sister clade to *Proteus*, the third *Sodalis*, *Baumannia*, *Blochmannia* and *Wigglesworthia* (S- and P-symbionts) as a sister or paraphyletic clade to the *Pectobacterium* and *Dickeya* clade and, finally, *Buchnera* species and *Ishikawaella* (P-symbionts) clustering with the *Erwinia* and *Pantoea* clade.

Conclusions: The results of this study confirm the efficiency of several artifact-reducing methods and strongly point towards the polyphyly of P-symbionts within Enterobacteriaceae. Interestingly, the model species of symbiotic bacteria research, *Buchnera* and *Wigglesworthia*, originated from closely related, but different, ancestors. The possible origins of intracellular symbiotic bacteria from gut-associated or pathogenic bacteria are suggested, as well as the role of facultative secondary symbionts as a source of bacteria that can gradually become obligate maternally transferred symbionts.

* Correspondence: filip.husnik@gmail.com; vacatko@paru.cas.cz
[1]Faculty of Science, University of South Bohemia, Branišovská 31, České Budějovice 37005, Czech Republic
Full list of author information is available at the end of the article

Background

One of the most fundamental evolutionary questions concerning insect-bacteria symbiosis is the origin and phylogenetic relationships of various symbiotic lineages. This knowledge is necessary for understanding the dynamics and mechanisms of symbiosis establishment and maintenance within the host. For instance, close relationships between symbionts and pathogenic bacteria suggests a transition from pathogenicity to symbiosis; polyphyly of the symbionts within a single host group is evidence of their multiple independent origins and close relationships among symbionts of different biology indicate high ecological flexibility within a given symbiotic group [1-6]. These implications are particularly important within Enterobacteriaceae, the group containing a broad spectrum of symbiotic lineages and forms described from various groups of insects. Their biology varies from loosely associated facultative symbionts (often called Secondary (S) symbionts) to obligatory mutualists of a highly derived nature, called Primary (P) symbionts [7-9]. However, the concept of the P- and S-symbionts and the associated terminology are a major oversimplification and they become inadequate for the description of the ever increasing complexity of the symbiotic system within Enterobacteriaceae. This complexity is manifested by such phenomena as the presence of multiple symbionts in a single host [10], occurrence of intermediate symbiotic forms and the replacement of symbionts within a host [11-14] or close phylogenetic relationships between typical S- and P-symbionts revealing their high ecological versatility [15]. A good example of such a complex system is provided by the occurrence of multiple obligate symbionts within Auchenorrhyncha [10], universally harboring *Sulcia muelleri* (Bacteroidetes) [16] with either *Hodgkinia cicadicola* (α-Proteobacteria) in cicadas, *Zinderia insecticola* (β-Proteobacteria) in spittlebugs or *Baumannia cicadellinicola* (γ-Proteobacteria) in sharpshooters. All of these latter symbionts are obligate and have been cospeciating with their hosts for millions of years [17-21]. A close phylogenetic relationship between typical S- and P-symbionts has been so far demonstrated in two well defined and often studied groups, the enterobacterial genera *Arsenophonus* and *Sodalis* [5,22,23]. The general capability of S-symbionts to supplement the metabolic functions of P-symbionts or even replace them was demonstrated experimentally by replacement of *Buchnera* with *Serratia* in aphids [24].

It is obvious that all these fascinating processes can only be studied on a reliable phylogenetic background [9,25-28]. Unfortunately, under current conditions, the phylogeny within Enterobacteriaceae and the placement of various symbiotic lineages are very unstable. Particularly, the P-symbionts present an extremely difficult challenge to phylogenetic computation due to their strongly modified genomes [9]. There are several root problems that are responsible for this dissatisfactory state. Traditionally, 16S rDNA was frequently used as an exclusive molecular marker for the description of a new symbiont. Many lineages are thus represented only by this gene, which has been shown within Enterobacteriaceae to be inadequate for inferring a reliable phylogeny [29]. In addition, it is notoriously known that the phylogenetic information of symbiotic bacteria is often seriously distorted due to the conditions associated with the symbiotic lifestyle. The effect of strong bottlenecks accompanied by reduced purifying selection and the overall degeneration of symbiotic genomes have been thoroughly discussed in many studies [30-33]. As a result of these degenerative processes, symbiotic lineages may experience parallel changes of their DNAs and these convergences produce the main source of phylogenetic artifacts. Among the most important features are biased nucleotide composition favoring adenine-thymine bases and rapid sequence evolution. While the compositional bias leads to the introduction of homoplasies at both nucleotide and amino acid levels, the accelerated evolution is a well known source of the long-branch attraction phenomenon [34,35]. Due to these circumstances, symbionts almost always appear as long branches in phylogenetic trees and tend to cluster together [36].

Various methodological approaches have been tested to overcome these difficulties (Additional file 1). They are based mainly on the concatenation of a large number of genes through the whole genome [37-39], the supertree and the consensus approach [37], exclusion of amino acids (FYMINK: phenylalanine, tyrosine, methionine, isoleucine, asparagine and lysine) most affected by nucleotide bias [37], modifications of sequence evolution models [11,12,36,40] and use of the genome structure as a source of phylogenetic data [41]. Phylogenomic studies based on large concatenated sets frequently imply monophyly of the typical P-symbionts (Additional file 1). However, due to the limited number of available genomes, these studies are usually based on inadequate taxon sampling. For example, secondary symbionts and plant pathogens that were shown to break the P-symbiont monophyly in the analysis using a nonhomogeneous model [40] could not be included into these phylogenomic studies. It is important to note that P-symbionts are probably only distantly related to the *Escherichia*/*Salmonella*/*Yersinia* clade. Therefore, the monophyly of P-symbionts derived from such a phylogenomic dataset is logically inevitable, but does not carry any evolutionary information.

The non-monophyletic nature of P-symbionts has been recently suggested in several studies. Perhaps the most inspiring is a study based on a nonhomogeneous model that separates P-symbionts into two independent lineages [40]. As an alternative, a paraphyletic arrangement of these symbionts in respect to several free-living taxa has been revealed from gene-order analysis based on break-point and inversion distances [41]. Most recently, Williams *et al.* [42] performed a 'telescoping' multiprotein phylogenomic analysis of 104 γ-Proteobacterial genomes. The phylogeny of Enterobacteriaceae endosymbionts was difficult to resolve, although it appeared that there were independent origins of at least the *Sodalis* and *Buchnera* lineages.

Thus, there is now a spectrum of hypotheses on the phylogeny of insect symbionts, ranging from complete polyphyly with multiple independent origins to complete monophyly with one common origin. In this study, we take advantage of current progress in computational methods to investigate phylogenetic relationships among the symbiotic lineages. One of the promising recent methodological advances is the introduction of a site-heterogeneous non-parametric mixture CAT model that allows for site-specific features of protein evolution [43]. This model was shown to solve the long-branch attraction (LBA) artifacts and outperform the previous models [44-47]. Similarly, the slow-fast method based on removal of the fastest evolving sites was shown to reduce phylogenetic artifacts [48-54], as well as purine/pyrimidine (RY) data recoding [55-58] or amino acid data recoding [59,60]. We used these methods as the core of a complex approach and tried to investigate series of methods, models and parameters to detect common trends in changes of the topologies. To do this, we applied two parallel approaches, one based on the application of recently developed algorithms and the other on the removal or recoding of the positions most affected by rapid sequence evolution and/or compositional (AT) bias. In addition, we paid particular attention to the sampling and used as much of a complete set of both symbiotic and free-living lineages as possible. This approach is particularly important to avoid interpretation uncertainties due to the absence of phylogenetically important lineages.

Results

The complete methodological design of this study and the resulting topologies are depicted in Figure 1. All matrices, alignments and phylogenetic trees are available in the TreeBASE database http://purl.org/phylo/treebase/phylows/study/TB2:S11451, as supplementary material, or on the webpage http://users.prf.jcu.cz/husnif00.

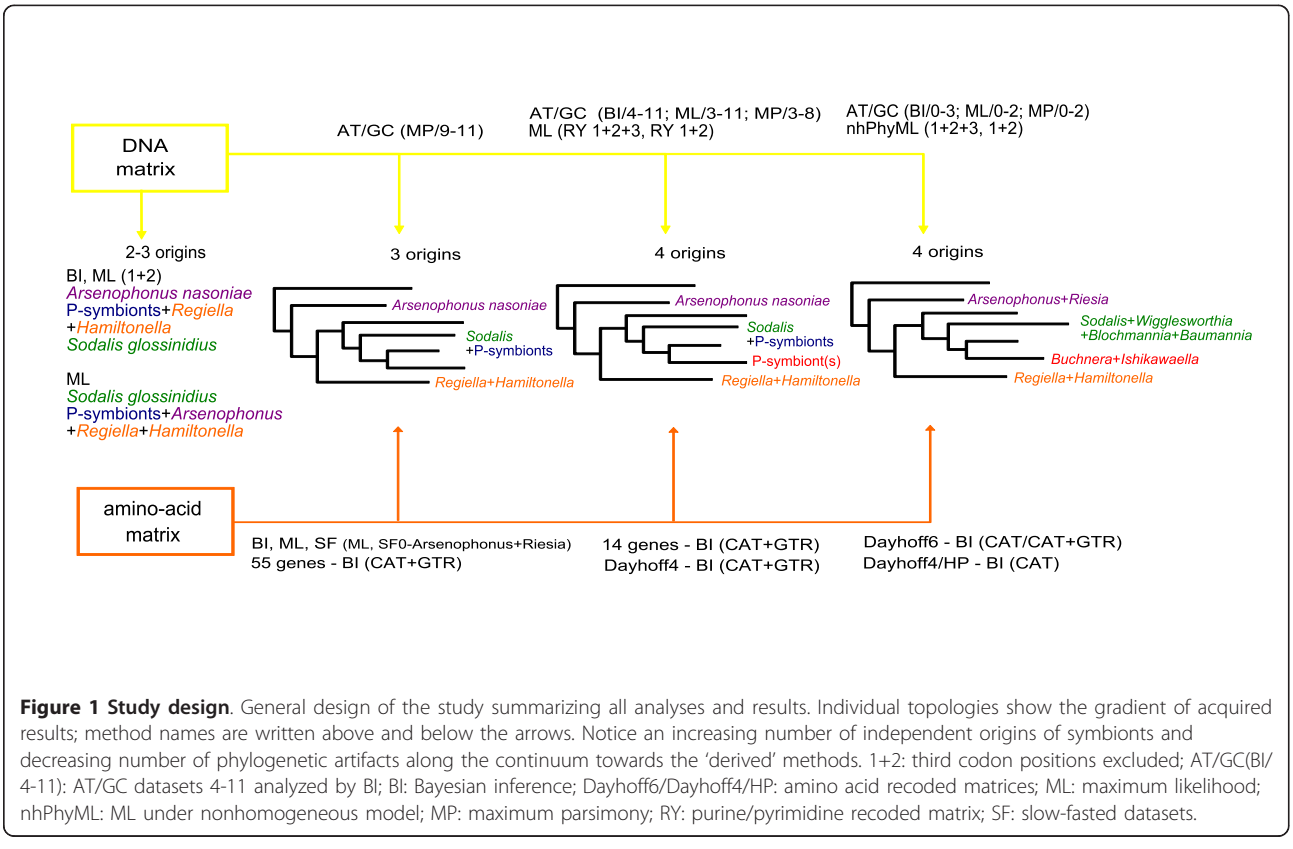


Figure 1 Study design. General design of the study summarizing all analyses and results. Individual topologies show the gradient of acquired results; method names are written above and below the arrows. Notice an increasing number of independent origins of symbionts and decreasing number of phylogenetic artifacts along the continuum towards the 'derived' methods. 1+2: third codon positions excluded; AT/GC(BI/4-11): AT/GC datasets 4-11 analyzed by BI; BI: Bayesian inference; Dayhoff6/Dayhoff4/HP: amino acid recoded matrices; ML: maximum likelihood; nhPhyML: ML under nonhomogeneous model; MP: maximum parsimony; RY: purine/pyrimidine recoded matrix; SF: slow-fasted datasets.

Standard maximum likelihood and Bayesian inference

The single gene maximum likelihood (ML) analyses of both nucleic and amino acid data provided an array of mutually exclusive topologies. The majority consensus based on amino acid data (Additional file 2a) groups almost all symbionts into polytomy with only two pairs of sister symbiotic species being resolved (*Buchnera* and *Blochmannia*). Phylogenetic trees inferred by ML and Bayesian inference (BI) from the nucleic acid concatenated data using the General Time Reversible model with an estimated proportion of invariable sites (I) and heterogeneity of evolutionary rates modeled by the four substitution rate categories of the gamma (Γ) distribution with the gamma shape parameter (alpha) estimated from the data (GTR+I+Γ) were apparently affected by phylogenetic artifacts, as demonstrated by placement of *Riesia* and *Wigglesworthia* within the *Buchnera* cluster with high posterior probabilities in the BI tree (Figure 2) and the attraction of two outgroup species (*Haemophilus* and *Pasteurella*) in the ML tree with high bootstrap support (Additional file 2b). Similar topologies were also retrieved from the amino acid concatenate by ML and BI using the LG+I+Γ, WAG+I+Γ and GTR+I+Γ models (Figure 3). Nevertheless, in contrast to the nucleotide-derived results, the monophyly of the *Buchnera* clade was not disrupted and *Hamiltonella* and *Regiella* were unambiguously separated from the other symbionts and clustered with *Yersinia*.

PhyloBayes, non-homogenous PhyML and modified matrices

The phylogenetic trees acquired under the CAT+GTR PhyloBayes model from 14 and 55 concatenated genes (Figure 4 and Additional file 2p) split symbiotic bacteria into four and three independent lineages, respectively. First, *Arsenophonus nasoniae* is a sister species to *Proteus mirabilis*; second, *Hamiltonella* and *Regiella* form a sister clade to *Yersinia pestis*; third, the *Sodalis*, *Baumannia*, *Blochmannia*, *Wigglesworthia*, *Riesia* and *Buchnera* clade form a sister clade to *Dickeya*/*Pectobacterium*. The position of *Ishikawaella* differs between the two datasets. In the 14-gene dataset, *Ishikawaella* forms a sister clade to *Pantoea* (Figure 4) and in the 55-gene dataset, it is attracted to the P-symbiont cluster (Additional file 2p).

A topology with four independent symbiotic clades resulted from the trees derived from dayhoff6 and dayhoff4 recoded amino acid data sets analyzed by CAT and CAT+GTR models (Figure 5, Additional file 2r, q) and partially with the hp (hydrophobic-polar) recoded dataset (Additional file 2c) - which was on the other hand affected by the substantial loss of phylogenetic information. The first clade is *Buchnera+Ishikawaella* as a sister clade to the *Erwinia*/*Pantoea* clade, the second clade is *Riesia+Arsenophonus* as a sister clade to *Proteus*, the third clade is *Hamiltonella+Regiella* as a sister clade to *Yersinia*, and the last clade is composed of *Sodalis*, *Baumannia*, *Blochmannia* and *Wigglesworthia*.

The analyses testing each symbiont independently, using a CAT+GTR model on the dayhoff6 recoded datasets, resulted in topologies supporting multiple origins of endosymbiosis (Additional file 2s). *Arsenophonus* clusters with *Proteus*; *Hamiltonella* clusters with *Yersinia*; *Regiella* clusters with *Yersinia*; and *Sodalis*, *Blochmannia*, *Baumannia*, *Riesia* and *Wigglesworthia* grouped into polytomy with the basal enterobacterial clades. Most importantly, the *Buchnera* clade clusters as a sister clade to the *Erwinia* clade and *Ishikawaella* is placed in polytomy with the *Pantoea* and *Erwinia* clade.

The non-homogenous (nh) PhyML nucleotide analyses with two different starting trees resulted in two different topologies (Figure 6 and Additional file 2d, e, f). When compared by the approximately unbiased (AU) test, the topology with four independent origins of symbiotic bacteria prevailed ($P = 1$) over the topology with monophyly of P-symbionts, which therefore corresponds to a local minimum due to a tree search failure (complete matrix: $P = 2 \times 10^{-67}$; matrix without the third positions: $P = 9 \times 10^{-87}$). The only incongruence in topologies based on the complete matrix (Additional file 2d) and the matrix without the third positions (Figure 6) was the placement of the *Sodalis+Baumannia+Blochmannia+Wigglesworthia* clade as a sister clade to the *Edwardsiella* or *Dickeya*/*Pectobacterium* clades.

Matrices obtained by removing positions according to the AT/GC contents produced trees covering the whole continuum illustrated in Figure 1. The most severe restrictions, that is, removal of all positions that contain both AT and GC categories or relaxing for up to three taxa (see BI trees in Additional file 2g, h, i, j), yielded topologies compatible with the results of the CAT model applied on the recoded amino acid data and of the nhPhyML analysis. Further relaxing the restriction rule led to a variety of trees along the Figure 1 continuum, with a less clear relation between the used parameter and the resulting topology (Additional file 3).

Compared to the ML analysis of all nucleotide positions, the analysis of first plus second positions reduced the obvious artifact of outgroup attraction (Additional file 2k). Nevertheless, it also sorted symbionts according to their branch length. Analysis of the RY recoded nucleotide matrix produced a tree compatible with the results of the CAT+GTR model (Additional file 2l). Analysis of the RY recoded nucleotide matrix without the third positions resulted in a topology with a *Sodalis+Baumannia+Blochmannia* cluster (as a sister to the *Pectobacterium*/*Dickeya* clade) separated from the rest

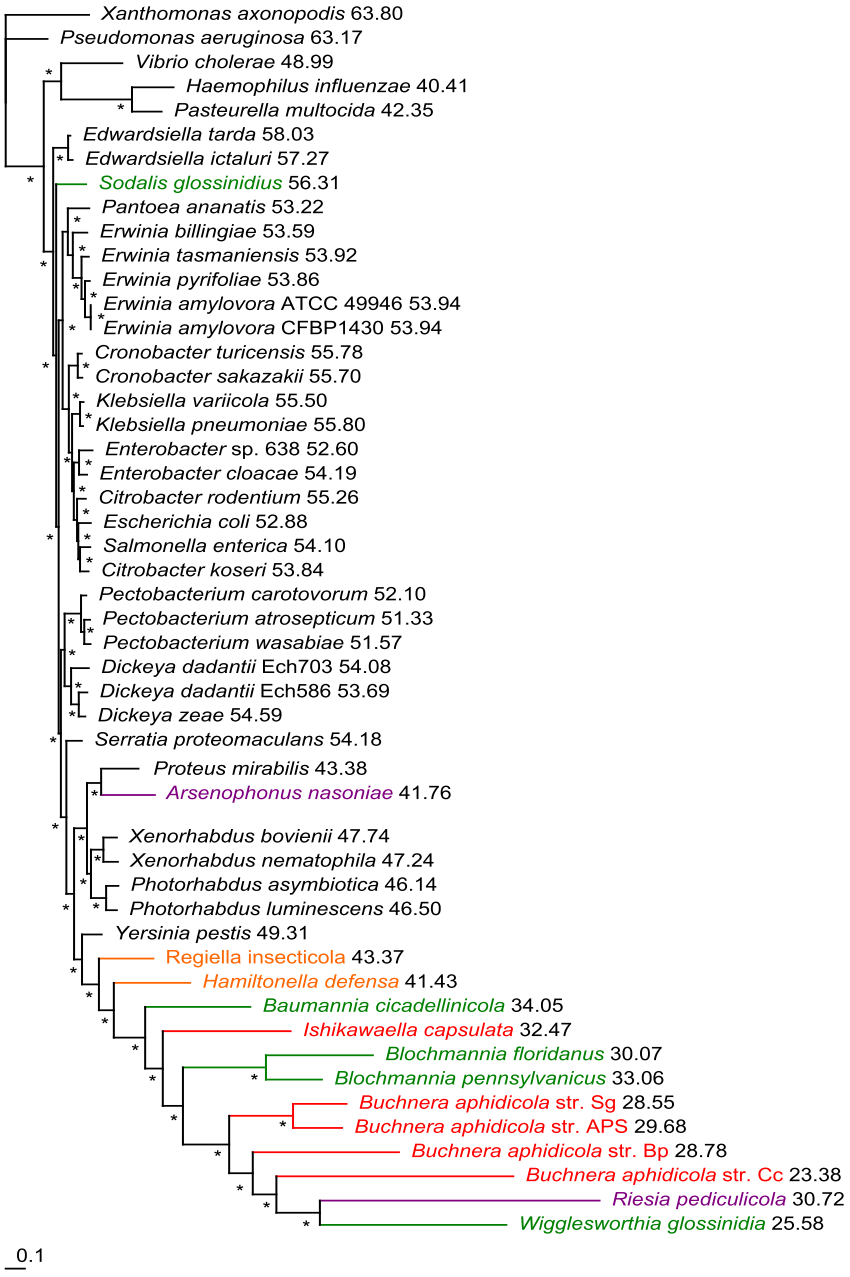


Figure 2 MrBayes phylogram - 69 genes, nucleotide matrix. Phylogenetic tree inferred from the concatenated nucleotide matrix using BI under the GTR+I+Γ model. Asterisks designate nodes with posterior probabilities equal to 1.0, values next to species names represent GC content calculated from the 69-gene dataset, genomic GC content can be found in Additional file 4. BI: Bayesian inference.

of the P-symbionts, which clustered with the *Erwinia*/*Pantoea* clade (Additional file 2m). Slow-fast analyses with gradual reduction of saturated positions did not produce the polyphyly of P-symbionts (Additional file 3; only the first five trees presented, subsequent trees are identical to the fifth tree). However, this analysis shows an increasing effect of LBA artifacts associated with the increasing number of remaining saturated positions, especially *Riesia* attraction and swapping of symbiotic branches according to their length.

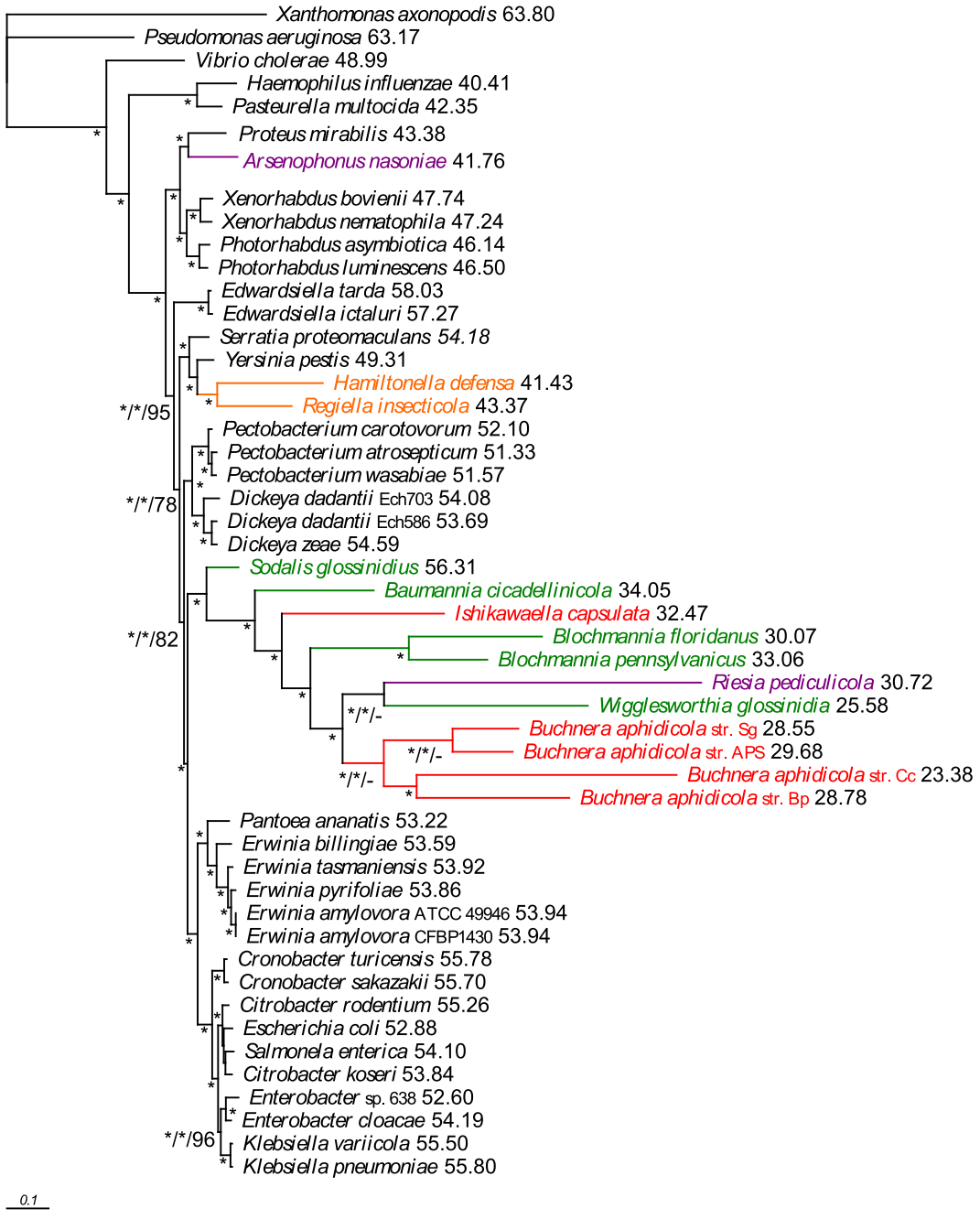


Figure 3 MrBayes phylogram - 69 genes, amino acid matrix. Phylogram inferred from the concatenated amino acid matrix using BI under the WAG+I+Γ model. Values at nodes represent posterior probabilities (WAG+I+Γ model, GTR+I+Γ protein model) and bootstrap supports from ML analysis (LG+I+Γ model). Asterisks designate nodes with posterior probabilities or bootstrap supports equal to 1.0, dashes designate values lower than 0.5 or 50, values next to species names represent GC content calculated from the 69-gene dataset, genomic GC content can be found in Additional file 4. BI: Bayesian inference. ML: maximum likelihood.

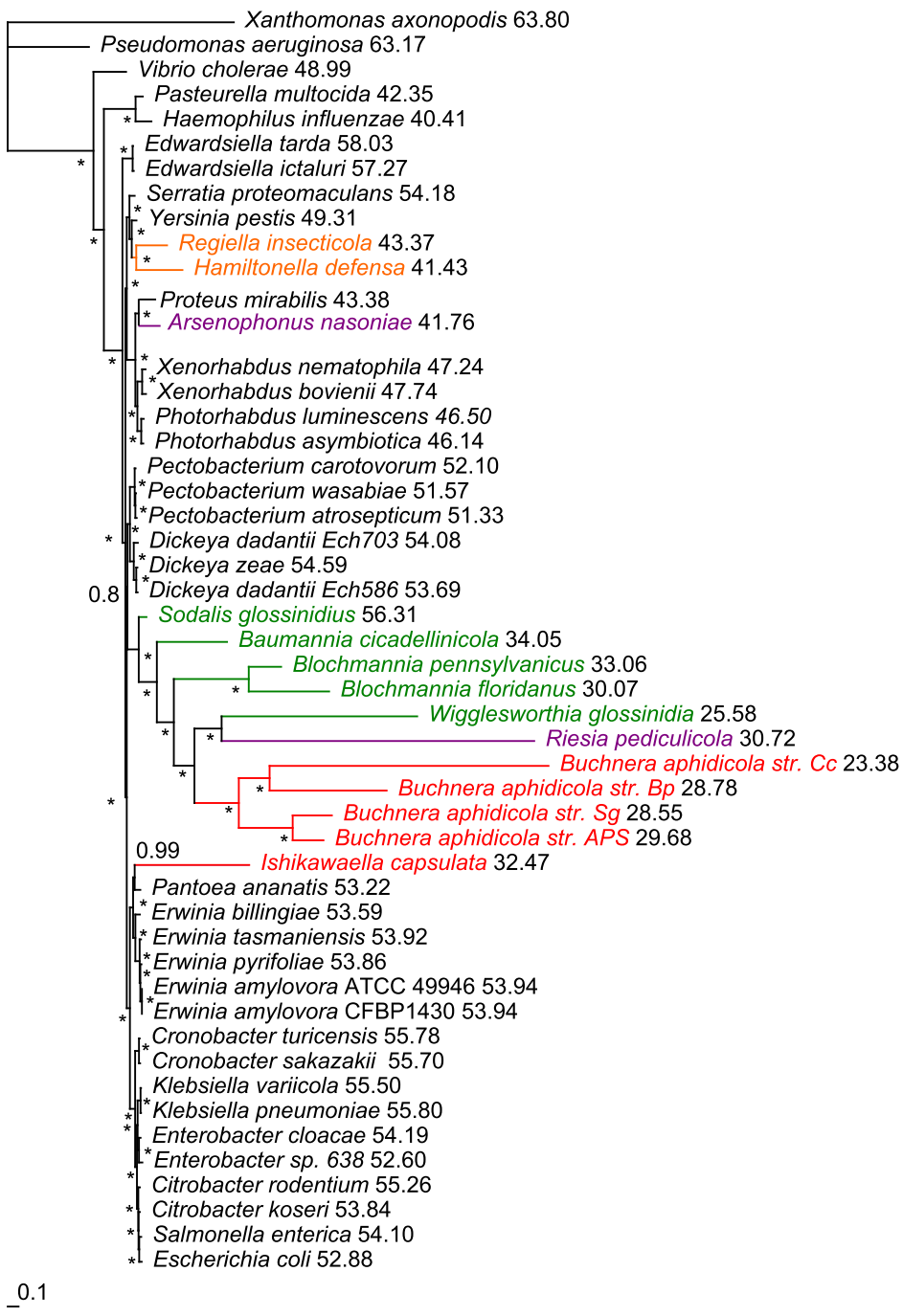


Figure 4 PhyloBayes phylogram - 14 genes, amino acid matrix. Phylogram derived from concatenation of 14 genes (*AceE*, *ArgS*, *AspS*, *EngA*, *GidA*, *GlyS*, *InfB*, *PheT*, *Pgi*, *Pnp*, *RpoB*, *RpoC*, *TrmE* and *YidC*) using PhyloBayes under the CAT+GTR model. Asterisks designate nodes with posterior probabilities equal to 1.0, values next to species names represent GC content calculated from the 69-gene dataset, genomic GC content can be found in Additional file 4.

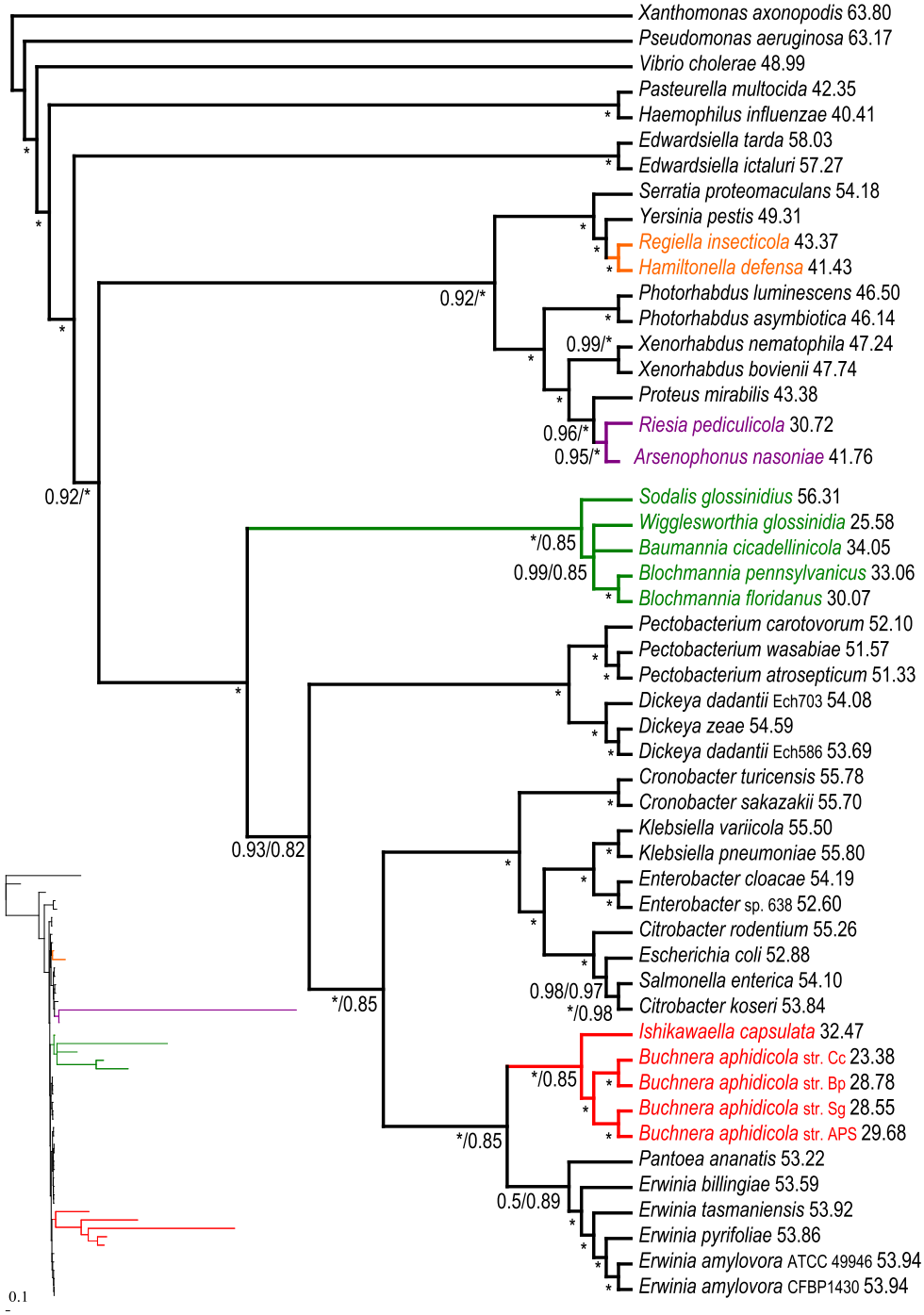


Figure 5 PhyloBayes cladogram - 69 genes, Dayhoff6 amino acid recoded matrix. Cladogram inferred from amino acid matrix recoded with Dayhoff6 scheme using PhyloBayes with the CAT and CAT+GTR model. Because of the length of symbiotic branches, phylogram is presented only as a preview (original phylogram can be found in Additional trees on our website). Values at nodes represent posterior probabilities from CAT and CAT+GTR analyses, respectively (asterisks designate nodes with posterior probabilities equal to 1.0). Values next to species names represent GC content calculated from the 69-gene dataset, genomic GC content can be found in Additional file 4.

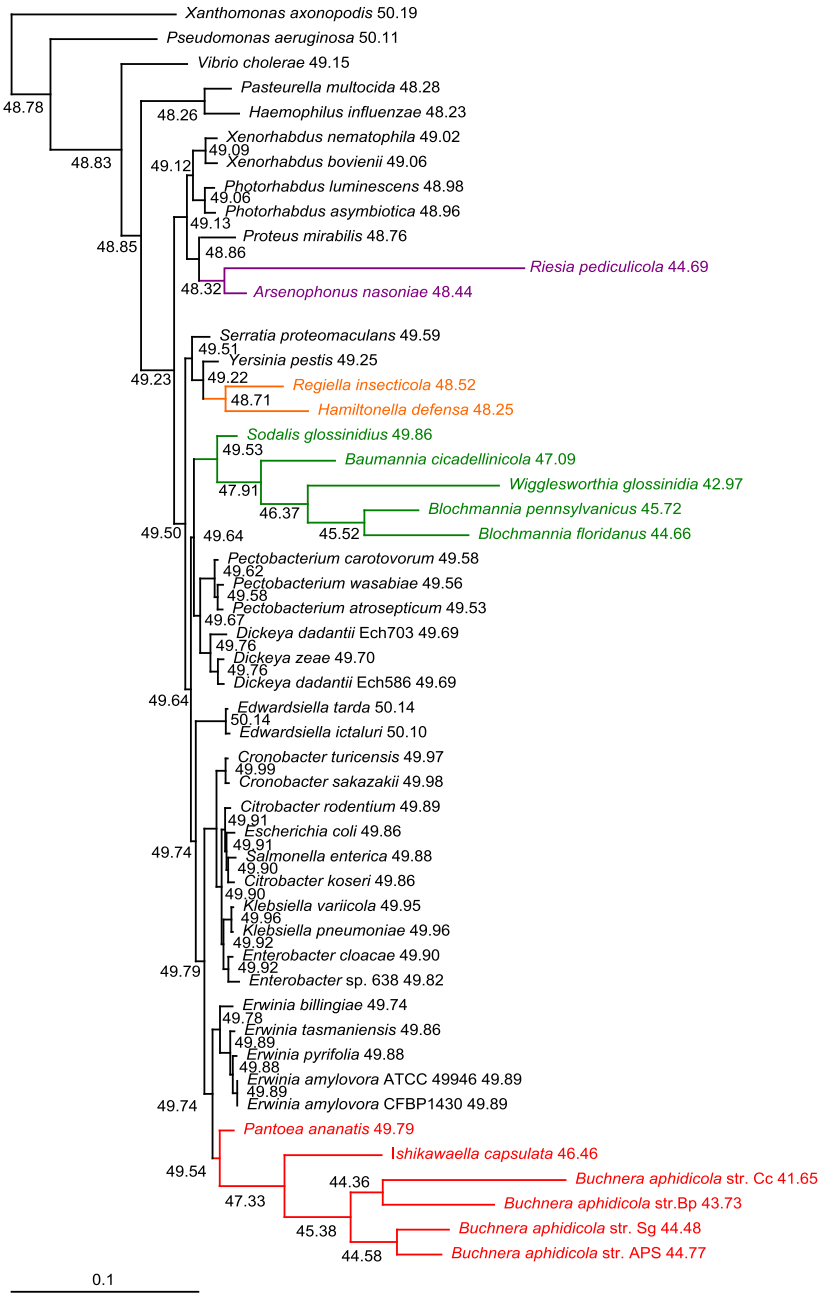


Figure 6 nhPhyML phylogram - 69 genes, nucleotide matrix, third positions excluded. Phylogram inferred from the concatenated nucleotide matrix without third codon positions using the nonhomogeneous model of evolution as implemented in nhPhyML. Values at nodes and branches represent GC content.

Discussion

Performance of the methods: convergence towards non-monophyly

The results obtained in this study strongly indicate that the frequently retrieved monophyly of P-symbionts is an artifact caused by their highly modified genomes. None of the most widely used methods, that is, ML and BI with different models used on nucleic (GTR+I+Γ) and amino acid (GTR/LG/WAG+I+Γ) data, were capable of resolving deep phylogenetic relationships and correct placement of the symbiotic taxa. This conclusion is evidenced by obvious artifacts, such as the inclusion of *Riesia* into the P-symbiotic lineage or the even more conspicuous distorted placing of *Wigglesworthia* within the *Buchnera* cluster. The arrangement of such trees suggests that these methods sort the symbionts according to their branch lengths and/or AT contents and attach the whole symbiotic cluster to the longest branch available. While the difficulty with placement of the most aberrant taxa, such as *Riesia*, *Wigglesworthia* and *Buchnera* (*Cinara cedri*) was also observed when using the mixture model accounting for site specific characteristics of protein evolution (Figure 4; Additional files 2p and 5), these artifacts disappeared after amino acid data recoding followed by CAT and CAT+GTR model analysis and the application of a nonhomogeneous model.

Additional support for the non-monophyly view stems from the second, parallel approach based on the restricted matrices. While our newly developed method shares the basic principles with the slow-fast and recoding methods, such as the removal of the positions that are likely to distort the phylogenetic relationships due to their aberrant evolution, it differs in the criteria of their removal and thus produces different input data. It is therefore significant that this method led independently to the same picture, the non-monophyly of the P-symbionts with clustering identical to the above analyses: *Ishikawaella+Buchnera* and *Sodalis+Baumannia+Blochmannia+Wigglesworthia*. The removal of the heteropecillous sites was recently shown to have similar effectiveness as our new method [61], which further supports the results. Moreover, this topology was obtained even under the maximum parsimony (MP) criterion (Additional file 3), which is known to be extremely sensitive to LBA [34]. On the other hand, although slow-fast analysis is generally considered a powerful tool for resolving relationships among taxa with different rates of evolution, we show in our data that the mere exclusion of the fast evolving sites is not sufficient when using empirical models and should be followed by analysis using some of the complex models, such as the CAT-like models. In addition, since this method usually requires an *a priori* definition of monophyletic groups, it should be used and interpreted with caution. Similar to the slow-fast method, RY recoding and exclusion of third codon positions were not sufficient for resolving deep symbiont phylogeny. However, all these methods can remove at least some of the artifacts and provide insight for further analyses.

Summarizing the topologies obtained in this study (Figure 1), a convergence can be detected towards a particular non-monophyletic arrangement of P-symbionts, as revealed under the most 'derived' methods. This result strongly supports the view of multiple origins of insect endosymbionts, as first revealed by the nonhomogeneous model of sequence evolution [40], and is partially congruent with the analyses of gene order [41] and phylogenomics of Gammaproteobacteria [42]. It is also important to note that, apart from multiple symbiont clustering, the arrangement of the non-symbiotic taxa corresponds to most of the phylogenomic analyses using *Escherichia*/*Salmonella*/*Yersinia* taxon sampling [37-39].

Biological significance of P-symbionts non-monophyly

Considering that most of the 'artifact-resistant' analyses point towards the non-monophyly of enterobacterial P-symbionts, the questions of how many symbiotic lineages are represented by the known symbiotic diversity and what are their closest free-living relatives now becomes of particular importance. It is not clear whether the split of the original P-symbiotic cluster into two lineages is definite or these two groups will be further divided after yet more sensitive methods and more complete data are available. At the moment there are still several clusters composed exclusively of derived symbiotic forms. In principle, three different processes may be responsible for the occurrence of such clusters: first, horizontal transmission of established symbiotic forms among host species; second, inadequate sampling with missing free-living relatives; or third, phylogenetic artifacts. All of these factors are likely to play a role in the current topological patterns. Being the main issues of this study, the role of methodological artifacts has been discussed above. Horizontal transmission, as the basis of non-artificial symbiotic clusters, is likely to take part at least in some cases. Perhaps the most convincing example is the *Wolbachia* cluster [62]: while within Enterobacteriaceae it may apply to *Arsenophonus*, *Sodalis* and possibly some other S-symbionts.

Recognition of the third cause, the incomplete sampling, and identification of the closest free-living relatives, now becomes a crucial step in future research. It is often assumed that symbionts originate from bacteria common to the environment typical for a given insect group. For example, cicadas spend most of their life cycle underground and feed primarily on plant roots.

Consequently, their α-Proteobacterial symbiont *Hodgkinia cicadicola* originated within Rhizobiales [19]. A similar ecological background can be noticed in yet different hosts, the ixodid and argasid ticks. Several reports have shown that some of the tick-transmitted pathogens are related to their symbiotic fauna [63-65]. Many of the insect taxa associated with symbiotic Enterobacteriaceae are phytophagous, and plant pathogens thus fit well into this hypothesis as hypothetical ancestors of various insect symbionts lineages. The presence of a type III secretion system, which is used in pathogenic bacteria for host cell invasion, in secondary symbionts [66-69] and its remnant in the primary symbiont of *Sitophilus* spp. weevils [70] could further support the theory of pathogenic ancestors of insect symbionts. It can only be speculated that these bacteria first became S-symbiont type and were horizontally transferred to various other insect species. Within some of the infected species, facultative symbionts eventually became obligatory primary symbionts. An identical situation can be observed in symbiotic clades with numerous species, such as *Wolbachia* [71,72], *Sodalis* [23,73,74] or *Arsenophonus* [5].

In our study, we gave particular attention to the sampling of free-living Enterobacteriaceae to provide as complete a background for the symbiotic lineages as possible under the current state of knowledge (that is, the availability of the genomic data). The most consistent picture derived from the presented analyses places the four main symbiotic clusters into the following positions. First, for the *Buchnera* cluster, its previously suggested relationship to *Erwinia* was confirmed. *Erwinia*, as a genus of mostly plant pathogenic bacteria, has been previously suggested to represent an ancestral organism, which upon ingestion by aphids at least 180 million years ago [75] turned into an intracellular symbiotic bacterium [76]. However, it is not known whether it was primarily pathogenic to aphids, similar to *Erwinia aphidicola* [77], or a gut associated symbiotic bacterium as in pentatomid stinkbugs [78], thrips [79,80] or Tephritidae flies [81-83]. *Ishikawaella capsulata*, an extracellular gut symbiont of plataspid stinkbugs [84], was the only symbiotic bacterium that clustered in our 'derived' analyses with the *Buchnera* clade. However, several single-gene studies indicate that this group contains some additional symbiotic lineages for which sequenced genome data is not currently available. These are, in particular, the extracellular symbionts of acanthosomatid stinkbugs [85], parastrachid stinkbugs [86], scutellerid stinkbugs [87,88] and some of the symbionts in pentatomid stinkbugs [78].

The second clade, represented in our analysis by *Sodalis+Baumannia+Blochmannia+Wigglesworthia*, is likely to encompass many other P- and S-symbionts [89-92]. The possible single origin of these symbionts has to be further tested, however the interspersion of both forms, together with basal position of *Sodalis*, seem to support a transition from a secondary to primary symbiotic lifestyle [15]. In our analysis, the whole clade was placed between pathogenic bacteria of plants and animals, the *Edwardsiella* and *Pectobacterium*/*Dickeya* clades, or as a sister to the latter group. Recently, another symbiotic bacterium (called BEV, *Euscelidius variegatus* host) was shown to be a sister species to *Pectobacterium* [93].

Two additional independent origins of insect symbionts are represented by the *Arsenophonus*/*Riesia* clade and *Hamiltonella+Regiella*. Both of these clades clustered in our analyses in the positions indicated by previous studies, that is, as related to *Proteus* and *Yersinia*, respectively [5,67,93-97].

While the position of individual symbiotic lineages is remarkably consistent across our 'artifact-resistant' analyses and are well compatible with some of the previous studies, the topology can only provide a rough picture of the relationships within Enterobacteriaceae. To get a more precise and phylogenetically meaningful background for an evolutionary interpretation, the sample of free-living bacteria as a possible source of symbiotic lineages has to be much improved. An illuminating example is provided by the bacterium *Biostraticola tofi*, described from water biofilms. When analyzed using 16S rDNA, this bacterium seemed to be closely related to *Sodalis* [98]. Its position as a sister group to the *Sodalis*/*Baumannia*/*Blochmannia*/*Wigglesworthia* clade was also retrieved in our single-gene analysis (*groEL*, data not shown). If confirmed by more precise multigene approach, *Biostraticola* would represent the closest bacterium to the large symbiotic cluster.

Conclusions

The topologies obtained by several independent approaches strongly support the non-monophyletic view of enterobacterial P-symbionts. Particularly, they show that at least three independent origins led to highly specialized symbiotic forms, the first giving rise to *Sodalis*, *Baumannia*, *Blochmannia* and *Wigglesworthia* (S- and P-symbionts), the second to *Buchnera* and *Ishikawella* and the last to *Riesia* and *Arsenophonus* (S- and P-symbionts). This separation of symbiotic clusters poses an interesting question as to whether the presented disbandment of the P-symbiotic cluster is definite or if it will continue after yet more complete data are available and more realistic evolutionary models [99-101] are applied. One obvious drawback of the current state is that many additional symbiotic lineages already known within Enterobacteriaceae cannot be at the moment included into serious phylogenetic analyses due to the lack of sufficient molecular data and will have to be revisited once complete genomic data are available.

These bacteria include symbionts of mealybugs [89,102], psyllids [90,103], lice [2,91], weevils [11,12,92], reed beetles [104,105], true bugs [78,84-88,106,107] and symbionts of leeches [108,109]. Similarly, the importance of free-living bacteria and variety of S-symbionts as possible ancestors of P-symbionts should not be underestimated when assembling datasets for phylogenetic analyses. The shift from polymerase chain reaction-based gene-centered sequencing towards high-throughput next-generation sequencing may soon provide sufficient data for more complete analyses of the Enterobacteriaceae phylogeny.

Methods

Matrices and multiple sequence alignments

The genes used in this study were extracted from 50 complete genome sequences of γ-Proteobacteria available in GenBank (Additional file 4), including 14 endosymbiotic Enterobacteriaceae. We did not include *Carsonella ruddii* [110] since this psyllid symbiotic bacterium does not appear to be a member of the Enterobacteriaceae clade [90,111] and is only attracted there by the AT rich taxa. After removal of the AT rich lineages from the analysis, *Carsonella ruddii* clusters with the genus *Pseudomonas* [42]. Also, we did not include *Serratia symbiotica* [95] because its genome only became available after completion of our datasets. However, the phylogenetic position of this symbiotic bacterium within *Serratia* genus is robust and was confirmed in several studies [6,14,112].

To minimize the introduction of a false phylogenetic signal, we compared the genomes of all symbiotic bacteria and selected only single-copy genes present in all of the included symbiotic and free-living taxa. Such strict gene exclusion was also necessary regarding the usage of computationally demanding methods; it was one of our goals to produce a taxonomically representative data set of efficient size with no missing data. Altogether, 69 orthologous genes, mostly involved in translation, ribosomal structure and biogenesis (Additional file 4) were selected according to the Clusters of Orthologous Groups of proteins (COGs) [113,114]. Single-gene nucleotide data sets were downloaded via their COG numbers from a freely available database (MicrobesOnline [115]).

All protein coding sequences were translated into amino acids in SeaView version 4 [116], aligned by the MAFFT version 6 L-INS-i algorithm [117] and toggled back to the nucleotide sequences. Ambiguously aligned positions (codons) were excluded by Gblocks v0.91b [118,119] with the following parameters: minimum number of sequences for a conserved position: 26; minimum number of sequences for a flanking position: 43; maximum number of contiguous nonconserved positions: 8; minimum length of a block: 10; allowed gap positions: with half. The resulting trimmed alignments were checked and manually corrected in BioEdit v7.0.5 [120]. Alignments were concatenated in SeaView. The 69 gene concatenate resulted in an alignment of 63, 462 nucleic acid positions with 42, 481 parsimony-informative and 48, 527 variable sites and 21, 154 amino acid positions with 12, 735 parsimony-informative and 15, 986 variable sites.

Phylogenetic analyses

We used two different approaches to deal with the distortions caused by the highly modified nature of symbiotic genomes, which are the main source of the phylogenetic artifacts in phylogenetic analyses.

First, we applied complex models of molecular evolution. Using PhyloBayes 3.2f [121], we applied non-parametric site heterogeneous CAT and CAT+GTR models [43]. For all PhyloBayes analyses, we ran two chains with an automatic stopping option set to end the chain when all discrepancies were lower than 0.3 and all effective sizes were larger than 100. Under the CAT and CAT+GTR models, the four independent PhyloBayes runs were stuck in a local maximum (maxdiff = 1) even after 25, 000 and 10, 000 cycles, respectively, and we were not able to reach Markov Chain Monte Carlo (MCMC) convergence. Therefore, we present these trees only as supplementary material (although they mostly point toward multiple origins of symbiosis; Additional file 5) and we ran the CAT+GTR analyses with the reduced dataset based on 14 genes with the number of parsimony-informative amino acid positions higher than 300 (*AceE*, *ArgS*, *AspS*, *EngA*, *GidA*, *GlyS*, *InfB*, *PheT*, *Pgi*, *Pnp*, *RpoB*, *RpoC*, *TrmE* and *YidC*). To check for compatibility of these arbitrary selected 14 genes with the rest of the data, we also analyzed, in a separate analysis, the remaining 55-gene dataset under the CAT +GTR model. Using nhPhyML [122], we applied a non-homogeneous nonstationary model of sequence evolution [123,124], which can deal with artifacts caused by compositional heterogeneity [40,125,126]. We used two different starting trees (Additional file 2n) and ran the analyses with and without the third codon positions. The resulting trees were evaluated by an AU test in CONSEL [127].

The second approach relies on the selective restriction of the data matrix. We used four previously established methods of data weighting and/or exclusion (see Background): RY data recoding, amino acid data recoding, exclusion of third codon positions and slow-fast analysis, and developed one additional method: since transition from G/C to A/T at many positions is a common homoplasy of symbiotic genomes, we removed from the matrix all positions containing both the G/C and A/T

states. All substitutions considered in the subsequent analyses thus included exclusively transversions within the A/T or G/C categories. To analyze an effect of this restriction on the reduction of the data, we prepared 11 matrices with a partially relaxed rule (removing all positions with AT+GC, allowing for one taxon exception, two taxa exception, and so on, up until a 10 taxa exception). Since this method has never been tested, we analyzed the restricted matrices by the BI, ML (parameters as for standard analyses) and MP using PAUP* 4.0b10 with the tree bisection and reconnection algorithm [128]. Four other types of data weighting and/or exclusion were used to increase the phylogenetic signal to noise ratio and determine the robustness of our results. First, the third codon positions were removed in SeaView. Second, RY recoding was performed on all and first plus second positions. Third, saturated positions were excluded from the concatenated data sets by Slow-Faster [129]. To assign substitutional rates to individual positions, unambiguously monophyletic groups were chosen on a polytomic tree (Additional file 2o), positions with the highest rates were gradually excluded and 21 restricted matrices were produced. These weighted data sets were analyzed by ML. Fourth, amino acid data recoding was performed in PhyloBayes with hp (A, C, F, G, I, L, M, V, W) (D, E, H, K, N, P, Q, R, S, T, Y), dayhoff4 (A, G, P, S, T) (D, E, N, Q) (H, K, R) (F, Y, W, I, L, M, V) (C = ?) and dayhoff6 (A, G, P, S, T) (D, E, N, Q) (H, K, R) (F, Y, W) (I, L, M, V) (C) recoding schemes. In addition, we have prepared 10 dayhoff6 recoded matrices to test individual symbiotic lineages without the presence of other symbionts. Amino acid recoded matrices were analyzed using the CAT and CAT+GTR models, which are more immune to phylogenetic artifacts than one-matrix models.

To allow for comparison of the results with previously published studies, as well as to separate the effect of newly used models and methods from changes due to the extended sampling, we also used standard procedures of phylogenetic inference, ML and BI. The following programs, algorithms and parameters were used in the ML and BI analyses. ML was applied to single-gene and concatenated alignments of both nucleotides and amino acids using PhyML v3.0 [130] with the subtree pruning and regrafting tree search algorithm. BI was performed in MrBayes 3.1.2 [131] with one to five million generations and tree sampling every 100 generations. Exploration of MCMC convergence and burn-in determination was performed in AWTY and Tracer v1.5 [132,133]. Evolutionary substitution models for proteins were selected by ProtTest 2.4 [134] and for DNA by jModelTest 0.1.1 [135] according to the Akaike Information Criterion. For DNA sequences, the GTR+I+Γ model was used [136-138]. Transition and transversion models [139] were used with I+Γ under ML for the first two AT/GC datasets. LG+I+Γ [140], WAG+I+Γ [141] and GTR+I+Γ models were used for amino acid data. A cross-validation method implemented in PhyloBayes [121,142] was used to estimate the fit of CAT-like models. For both datasets, the 14 selected genes as well as the complete 69 genes set, the cross-validation was performed according to the PhyloBayes manual in 10 replicates each with 1, 100 cycles. The CAT-Poisson model had significantly better fit to the data than the GTR model (Δl 157.37 ± 56.9379 for the 14-gene matrix and Δl 3923.9 ± 1963.5 for the 69-gene matrix); of the CAT-like models, the CAT+GTR model was found to be significantly better than the CAT-Poisson model (Δl 536.71 ± 32.8341 for the 14-gene matrix and Δl 1633.4 ± 123.482 for the 69-gene matrix) in all 10 replicates.

Additional material

Additional file 1: Summary of 20 studies on symbionts phylogeny.

Additional file 2: Additional phylogenetic trees.

Additional file 3: All phylogenetic trees derived from AT-GC and SF datasets. A rar file of all phylogenetic trees obtained under BI, ML and MP from 11 AT/GC datasets, and under ML from five slow-fasted datasets. Trees are in phylip and nexus formats and can be viewed, for example, in TreeView http://taxonomy.zoology.gla.ac.uk/rod/treeview.html or Mesquite http://mesquiteproject.org/mesquite/mesquite.html.

Additional file 4: List of the taxa and orthologous genes used in the study.

Additional file 5: Additional phylogenetic trees inferred from CAT and CAT+GTR unconverged chains.

Acknowledgements
This work was supported by the Grant Agency of Academy of Sciences of the Czech Republic (grant number P505/10/1401 to VH); the Student Grant Agency of the Faculty of Science, University of South Bohemia (grant number SGA2009002 to FH); and the Ministry of Education, Youth, and Sports of the Czech Republic (grant numbers LC06073 and MSM 60076605801 to VH). Access to the MetaCentrum computing facilities provided under the 'Projects of Large Infrastructure for Research, Development, and Innovations' program LM2010005 funded by the Ministry of Education, Youth, and Sports of the Czech Republic is highly appreciated.

Author details
[1]Faculty of Science, University of South Bohemia, Branišovská 31, České Budějovice 37005, Czech Republic. [2]Institute of Parasitology, Biology Centre of ASCR, Branišovská 31, České Budějovice 37005, Czech Republic.

Authors' contributions
FH carried out the sequence alignments and phylogenetic analyses, and participated in the study design, evolutionary interpretation of the results and preparation of the manuscript. TCH compiled and analyzed the AT/GC reduced matrices. VH conceived of the study and participated in its design, evolutionary interpretation of the results and preparation of the manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 27 July 2011 Accepted: 28 December 2011
Published: 28 December 2011

References

1. Gottlieb Y, Ghanim M, Gueguen G, Kontsedalov S, Vavre F, Fleury F, Zchori-Fein E: **Inherited intracellular ecosystem: symbiotic bacteria share bacteriocytes in whiteflies.** *FASEB J* 2008, **22(7)**:2591-2599.
2. Hypša V, Křížek J: **Molecular evidence for polyphyletic origin of the primary symbionts of sucking lice (Phthiraptera, Anoplura).** *Microb Ecol* 2007, **54(2)**:242-251.
3. Bordenstein SR, Paraskevopoulos C, Hotopp JC, Sapountzis P, Lo N, Bandi C, Tettelin H, Werren JH, Bourtzis K: **Parasitism and mutualism in *Wolbachia*: what the phylogenomic trees can and cannot say.** *Mol Biol Evol* 2009, **26(1)**:231-241.
4. Caspi-Fluger A, Zchori-Fein E: **Do plants and insects share the same symbionts?** *Isr J Plant Sci* 2010, **58(2)**:113-119.
5. Nováková E, Hypša V, Moran NA: ***Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution.** *BMC Microbiol* 2009, **9**:143.
6. Moran NA, Russell JA, Koga R, Fukatsu T: **Evolutionary relationships of three new species of Enterobacteriaceae living as symbionts of aphids and other insects.** *Appl Environ Microbiol* 2005, **71(6)**:3302-3310.
7. Baumann P: **Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects.** *Annu Rev Microbiol* 2005, **59**:155-189.
8. Douglas AE: **Mycetocyte symbiosis in insects.** *Biol Rev Camb Philos Soc* 1989, **64(4)**:409-434.
9. Moran NA, McCutcheon JP, Nakabachi A: **Genomics and evolution of heritable bacterial symbionts.** *Annu Rev Genet* 2008, **42**:165-190.
10. Takiya DM, Tran PL, Dietrich CH, Moran NA: **Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts.** *Mol Ecol* 2006, **15(13)**:4175-4191.
11. Lefevre C, Charles H, Vallier A, Delobel B, Farrell B, Heddi A: **Endosymbiont phylogenesis in the Dryophthoridae weevils: evidence for bacterial replacement.** *Mol Biol Evol* 2004, **21(6)**:965-973.
12. Conord C, Despres L, Vallier A, Balmand S, Miquel C, Zundel S, Lemperiere G, Heddi A: **Long-term evolutionary stability of bacterial endosymbiosis in Curculionoidea: additional evidence of symbiont replacement in the Dryophthoridae family.** *Mol Biol Evol* 2008, **25(5)**:859-868.
13. Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A: **A small microbial genome: the end of a long symbiotic relationship?** *Science* 2006, **314(5797)**:312-313.
14. Lamelas A, Perez-Brocal V, Gomez-Valero L, Gosalbes MJ, Moya A, Latorre A: **Evolution of the secondary symbiont "*Candidatus* Serratia symbiotica" in aphid species of the subfamily Lachninae.** *Appl Environ Microbiol* 2008, **74(13)**:4236-4240.
15. Wernegreen JJ, Kauppinen SN, Brady SG, Ward PS: **One nutritional symbiosis begat another: phylogenetic evidence that the ant tribe Camponotini acquired *Blochmannia* by tending sap-feeding insects.** *BMC Evol Biol* 2009, **9(1)**:292.
16. Moran NA, Tran P, Gerardo NM: **Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes.** *Appl Environ Microbiol* 2005, **71(12)**:8802-8810.
17. 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: **Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters.** *PLoS Biol* 2006, **4(6)**:e188.
18. McCutcheon JP, Moran NA: **Functional convergence in reduced genomes of bacterial symbionts spanning 200 million years of evolution.** *Genome Biol Evol* 2010, **2**:708-718.
19. McCutcheon JP, McDonald BR, Moran NA: **Convergent evolution of metabolic roles in bacterial co-symbionts of insects.** *Proc Natl Acad Sci USA* 2009, **106(36)**:15394-15399.
20. McCutcheon JP, McDonald BR, Moran NA: **Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont.** *Plos Genet* 2009, **5(7)**:e1000565.
21. McCutcheon JP, Moran NA: **Parallel genomic evolution and metabolic interdependence in an ancient symbiosis.** *Proc Natl Acad Sci USA* 2007, **104(49)**:19392-19397.
22. Snyder AK, McMillen CM, Wallenhorst P, Rio RV: **The phylogeny of *Sodalis*-like symbionts as reconstructed using surface-encoding loci.** *FEMS Microbiol Lett* 2011, **317(2)**:143-151.
23. Fukatsu T, Koga R, Smith WA, Tanaka K, Nikoh N, Sasaki-Fukatsu K, Yoshizawa K, Dale C, Clayton DH: **Bacterial endosymbiont of the slender pigeon louse, *Columbicola columbae*, allied to endosymbionts of grain weevils and tsetse flies.** *Appl Environ Microbiol* 2007, **73(20)**:6660-6668.
24. 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 R Soc Lond B* 2003, **270(1533)**:2543-2550.
25. Oliver KM, Degnan PH, Burke GR, Moran NA: **Facultative symbionts in aphids and the horizontal transfer of ecologically important traits.** *Annu Rev Entomol* 2010, **55**:247-266.
26. Moya A, Pereto J, Gil R, Latorre A: **Learning how to live together: genomic insights into prokaryote-animal symbioses.** *Nat Rev Genet* 2008, **9(3)**:218-229.
27. Gosalbes MJ, Latorre A, Lamelas A, Moya A: **Genomics of intracellular symbionts in insects.** *Int J Med Microbiol* 2010, **300(5)**:271-278.
28. Toft C, Andersson SGE: **Evolutionary microbial genomics: insights into bacterial host adaptation.** *Nat Rev Genet* 2010, **11(7)**:465-475.
29. Naum M, Brown EW, Mason-Gamer RJ: **Is 16S rDNA a reliable phylogenetic marker to characterize relationships below the family level in the Enterobacteriaceae?** *J Mol Evol* 2008, **66(6)**:630-642.
30. Mira A, Moran NA: **Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria.** *Microb Ecol* 2002, **44(2)**:137-143.
31. Wernegreen JJ, Moran NA: **Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes.** *Mol Biol Evol* 1999, **16(1)**:83-97.
32. Lambert JD, Moran NA: **Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria.** *Proc Natl Acad Sci USA* 1998, **95(8)**:4458-4462.
33. Allen JM, Light JE, Perotti MA, Braig HR, Reed DL: **Mutational meltdown in primary endosymbionts: selection limits Muller's ratchet.** *Plos One* 2009, **4(3)**:e4969.
34. Bergsten J: **A review of long-branch attraction.** *Cladistics* 2005, **21(2)**:163-193.
35. Ruano-Rubio V, Fares MA: **Artifactual phylogenies caused by correlated distribution of substitution rates among sites and lineages: the good, the bad, and the ugly.** *Syst Biol* 2007, **56(1)**:68-82.
36. Charles H, Heddi A, Rahbe Y: **A putative insect intracellular endosymbiont stem clade, within the Enterobacteriaceae, infered from phylogenetic analysis based on a heterogeneous model of DNA evolution.** *C R Acad Sci Ser III Sci Vie* 2001, **324(5)**:489-494.
37. Comas I, Moya A, Gonzalez-Candelas F: **From phylogenetics to phylogenomics: the evolutionary relationships of insect endosymbiotic γ-Proteobacteria as a test case.** *Syst Biol* 2007, **56(1)**:1-16.
38. Lerat E, Daubin V, Moran NA: **From gene trees to organismal phylogeny in prokaryotes: the case of the γ-Proteobacteria.** *PLoS Biol* 2003, **1(1)**:e19.
39. Canback B, Tamas I, Andersson SG: **A phylogenomic study of endosymbiotic bacteria.** *Mol Biol Evol* 2004, **21(6)**:1110-1122.
40. Herbeck JT, Degnan PH, Wernegreen JJ: **Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the Enterobacteriales (γ-Proteobacteria).** *Mol Biol Evol* 2005, **22(3)**:520-532.
41. Belda E, Moya A, Silva FJ: **Genome rearrangement distances and gene order phylogeny in γ-Proteobacteria.** *Mol Biol Evol* 2005, **22(6)**:1456-1467.
42. Williams KP, Gillespie JJ, Sobral BW, Nordberg EK, Snyder EE, Shallom JM, Dickerman AW: **Phylogeny of Gammaproteobacteria.** *J Bacteriol* 2010, **192(9)**:2305-2314.
43. Lartillot N, Philippe H: **A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process.** *Mol Biol Evol* 2004, **21(6)**:1095-1109.
44. Lartillot N, Brinkmann H, Philippe H: **Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model.** *BMC Evol Biol* 2007, **7**:S4.
45. Lartillot N, Philippe H: **Improvement of molecular phylogenetic inference and the phylogeny of Bilateria.** *Philos Trans R Soc Lond B Biol Sci* 2008, **363(1496)**:1463-1472.
46. Nesnidal MP, Helmkampf M, Bruchhaus I, Hausdorf B: **Compositional heterogeneity and phylogenomic inference of metazoan relationships.** *Mol Biol Evol* 2010, **27(9)**:2095-2104.
47. Philippe H, Brinkman FS, Martinez P, Riutort M, Baguna J: **Acoel flatworms are not platyhelminthes: evidence from phylogenomics.** *Plos One* 2007, **2(8)**:e717.

48. Brinkmann H, Philippe H: **Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies.** *Mol Biol Evol* 1999, **16(6)**:817-825.
49. Brochier C, Philippe H: **A non-hyperthermophilic ancestor for bacteria.** *Nature* 2002, **417(6886)**:244.
50. Hampl V, Čepička I, Flegr J, Tachezy J, Kulda J: **Critical analysis of the topology and rooting of the parabasalian 16S rRNA tree.** *Mol Phylogenet Evol* 2004, **32(3)**:711-723.
51. Bapteste E, Susko E, Leigh J, Ruiz-Trillo I, Bucknam J, Doolittle WF: **Alternative methods for concatenation of core genes indicate a lack of resolution in deep nodes of the prokaryotic phylogeny.** *Mol Biol Evol* 2008, **25(1)**:83-91.
52. Philippe H, Lopez P, Brinkmann H, Budin K, Germot A, Laurent J, Moreira D, Muller M, Le Guyader H: **Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions.** *Proc R Soc Lond B* 2000, **267(1449)**:1213-1221.
53. Hirt RP, Logsdon JM, Healy B, Dorey MW, Doolittle WF, Embley TM: **Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins.** *Proc Natl Acad Sci USA* 1999, **96(2)**:580-585.
54. Philippe H, Lartillot N, Brinkmann H: **Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia.** *Mol Biol Evol* 2005, **22(5)**:1246-1253.
55. Phillips MJ, Delsuc F, Penny D: **Genome-scale phylogeny and the detection of systematic biases.** *Mol Biol Evol* 2004, **21(7)**:1455-1458.
56. Delsuc F, Phillips MJ, Penny D: **Comment on "Hexapod origins: monophyletic or paraphyletic?".** *Science* 2003, **301(5639)**:1482.
57. Phillips MJ, Penny D: **The root of the mammalian tree inferred from whole mitochondrial genomes.** *Mol Phylogenet Evol* 2003, **28(2)**:171-185.
58. Gibson A, Gowri-Shankar V, Higgs PG, Rattray M: **A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods.** *Mol Biol Evol* 2005, **22(2)**:251-264.
59. Hrdý I, Hirt RP, Doležal P, Bardoňová L, Foster PG, Tachezy J, Embley TM: ***Trichomonas*** **hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I.** *Nature* 2004, **432(7017)**:618-622.
60. Embley TM, van der Giezen M, Horner DS, Dyal PL, Foster P: **Mitochondria and hydrogenosomes are two forms of the same fundamental organelle.** *Philos Trans R Soc Lond B Biol Sci* 2003, **358(1429)**:191-201, discussion 201-203.
61. Philippe H, Roure B: **Site-specific time heterogeneity of the substitution process and its impact on phylogenetic inference.** *BMC Evol Biol* 2011, **11**:17.
62. Raychoudhury R, Baldo L, Oliveira DC, Werren JH: **Modes of acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the Nasonia species complex.** *Evolution* 2009, **63(1)**:165-183.
63. Perlman SJ, Hunter MS, Zchori-Fein E: **The emerging diversity of *Rickettsia*.** *Proc R Soc B* 2006, **273(1598)**:2097-2106.
64. Noda H, Munderloh UG, Kurtti TJ: **Endosymbionts of ticks and their relationship to *Wolbachia* spp. and tick-borne pathogens of humans and animals.** *Appl Environ Microbiol* 1997, **63(10)**:3926-3932.
65. Sassera D, Beninati T, Bandi C, Bouman EAP, Sacchi L, Fabbi M, Lo N: **'*Candidatus* Midichloria mitochondrii', an endosymbiont of the tick *Ixodes ricinus* with a unique intramitochondrial lifestyle.** *Int J Syst Evol Microbiol* 2006, **56**:2535-2540.
66. Degnan PH, Leonardo TE, Cass BN, Hurwitz B, Stern D, Gibbs RA, Richards S, Moran NA: **Dynamics of genome evolution in facultative symbionts of aphids.** *Environ Microbiol* 2009, **12(8)**:2060-2069.
67. Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA: ***Hamiltonella defensa*****, genome evolution of protective bacterial endosymbiont from pathogenic ancestors.** *Proc Natl Acad Sci USA* 2009, **106(22)**:9063-9068.
68. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S: **Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host.** *Genome Res* 2006, **16(2)**:149-156.
69. Wilkes TE, Darby AC, Choi JH, Colbourne JK, Werren JH, Hurst GD: **The draft genome sequence of *Arsenophonus nasoniae*, son-killer bacterium of *Nasonia vitripennis*, reveals genes associated with virulence and symbiosis.** *Insect Mol Biol* 2010, **19(Sp. Iss. SI Suppl. 1)**:59-73.
70. Dale C, Plague GR, Wang B, Ochman H, Moran NA: **Type III secretion systems and the evolution of mutualistic endosymbiosis.** *Proc Natl Acad Sci USA* 2002, **99(19)**:12397-12402.
71. Hosokawa T, Koga R, Kikuchi Y, Meng XY, Fukatsu T: ***Wolbachia*** **as a bacteriocyte-associated nutritional mutualist.** *Proc Natl Acad Sci USA* 2010, **107(2)**:769-774.
72. Dedeine F, Vavre F, Fleury F, Loppin B, Hochberg ME, Bouletreau M: **Removing symbiotic *Wolbachia* bacteria specifically inhibits oogenesis in a parasitic wasp.** *Proc Natl Acad Sci USA* 2001, **98(11)**:6247-6252.
73. 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* 1998, **47(1)**:52-61.
74. Nováková E, Hypša V: **A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*.** *FEMS Microbiol Lett* 2007, **269(1)**:131-135.
75. Moran N, Munson M, Baumann P, Ishikawa H: **A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts.** *Proc R Soc Lond B* 1993, **253(1337)**:167-171.
76. Harada H, Oyaizu H, Ishikawa H: **A consideration about the origin of aphid intracellular symbiont in connection with gut bacterial flora.** *J Gen Appl Microbiol* 1996, **42(1)**:17-26.
77. Harada H, Oyaizu H, Kosako Y, Ishikawa H: ***Erwinia aphidicola*****, a new species isolated from pea aphid, *Acyrthosiphon pisum*.** *J Gen Appl Microbiol* 1997, **43(6)**:349-354.
78. Prado SS, Almeida RP: **Phylogenetic placement of pentatomid stink bug gut symbionts.** *Curr Microbiol* 2009, **58(1)**:64-69.
79. Chanbusarakum L, Ullman D: **Characterization of bacterial symbionts in *Frankliniella occidentalis* (Pergande), Western flower thrips.** *J Invertebr Pathol* 2008, **99(3)**:318-325.
80. De Vries EJ, Van der Wurff AWG, Jacobs G, Breeuwer JAJ: **Onion thrips, *Thrips tabaci*, have gut bacteria that are closely related to the symbionts of the western flower thrips, *Frankliniella occidentalis*.** *J Insect Sci* 2008, **8**:1-11.
81. Mazzon L, Martinez-Sanudo I, Simonato M, Squartini A, Savio C, Girolami V: **Phylogenetic relationships between flies of the Tephritinae subfamily (Diptera, Tephritidae) and their symbiotic bacteria.** *Mol Phylogenet Evol* 2010, **56(1)**:312-326.
82. Mazzon L, Piscedda A, Simonato M, Martinez-Sanudo I, Squartini A, Girolami V: **Presence of specific symbiotic bacteria in flies of the subfamily Tephritinae (Diptera Tephritidae) and their phylogenetic relationships: proposal of '*Candidatus* Stammerula tephritidis'.** *Int J Syst Evol Microbiol* 2008, **58(Pt 6)**:1277-1287.
83. Capuzzo C, Firrao G, Mazzon L, Squartini A, Girolami V: **'*Candidatus* Erwinia dacicola', a coevolved symbiotic bacterium of the olive fly *Bactrocera oleae* (Gmelin).** *Int J Syst Evol Microbiol* 2005, **55(Pt 4)**:1641-1647.
84. Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T: **Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria.** *PLoS Biol* 2006, **4(10)**:e337.
85. Kikuchi Y, Hosokawa T, Nikoh N, Meng XY, Kamagata Y, Fukatsu T: **Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs.** *BMC Biol* 2009, **7**:2.
86. Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Hironaka M, Fukatsu T: **Phylogenetic position and peculiar genetic traits of a midgut bacterial symbiont of the stinkbug *Parastrachia japonensis*.** *Appl Environ Microbiol* 2010, **76(13)**:4130-4135.
87. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M, Fukatsu T: **Primary gut symbiont and secondary *Sodalis*-allied symbiont in the scutellerid stinkbug *Cantao ocellatus*.** *Appl Environ Microbiol* 2010, **76(11)**:3486-3494.
88. Kaiwa N, Hosokawa T, Kikuchi Y, Nikoh N, Meng XY, Kimura N, Ito M, Fukatsu T: **Bacterial symbionts of the giant jewel stinkbug *Eucorysses grandis* (Hemiptera: Scutelleridae).** *Zool Sci* 2011, **28(3)**:169-174.
89. Thao ML, Gullan PJ, Baumann P: **Secondary (γ-Proteobacteria) endosymbionts infect the primary (β-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts.** *Appl Environ Microbiol* 2002, **68(7)**:3190-3197.
90. Spaulding AW, von Dohlen CD: **Phylogenetic characterization and molecular evolution of bacterial endosymbionts in psyllids (Hemiptera: Sternorrhyncha).** *Mol Biol Evol* 1998, **15(11)**:1506-1513.
91. Fukatsu T, Hosokawa T, Koga R, Nikoh N, Kato T, Hayama S, Takefushi H, Tanaka I: **Intestinal endocellular symbiotic bacterium of the macaque louse *Pedicinus obtusus*: Distinct endosymbiont origins in anthropoid**

primate lice and the old world monkey louse. *Appl Environ Microbiol* 2009, **75(11)**:3796-3799.
92. Toju H, Hosokawa T, Koga R, Nikoh N, Meng XY, Kimura N, Fukatsu T: "*Candidatus* Curculioniphilus buchneri", a novel clade of bacterial endocellular symbionts from weevils of the genus *Curculio*. *Appl Environ Microbiol* 2009, **76(1)**:275-282.
93. Degnan PH, Bittleston LS, Hansen AK, Sabree ZL, Moran NA, Almeida RP: Origin and examination of a leafhopper facultative endosymbiont. *Curr Microbiol* 2011, **62(5)**:1565-1572.
94. Allen JM, Reed DL, Perotti MA, Braig HR: Evolutionary relationships of "*Candidatus* Riesia spp., " endosymbiotic *Enterobacteriaceae* living within hematophagous primate lice. *Appl Environ Microbiol* 2007, **73(5)**:1659-1664.
95. Burke GR, Moran NA: Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol* 2011, **3**:195-208.
96. Gherna RL, Werren JH, Weisburg W, Cote R, Woese CR, Mandelco L, Brenner DJ: *Arsenophonus nasoniae* gen. nov, sp. nov, the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. *Int J Syst Bacteriol* 1991, **41(4)**:563-565.
97. 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* 2006, **72(4)**:2997-3004.
98. Verbarg S, Fruhling A, Cousin S, Brambilla E, Gronow S, Lunsdorf H, Stackebrandt E: *Biostraticola tofi* gen. nov., spec. nov., a novel member of the family *Enterobacteriaceae*. *Curr Microbiol* 2008, **56(6)**:603-608.
99. Blanquart S, Lartillot N: A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol* 2008, **25(5)**:842-858.
100. Blanquart S, Lartillot N: A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol Biol Evol* 2006, **23(11)**:2058-2071.
101. Foster PG: Modeling compositional heterogeneity. *Syst Biol* 2004, **53(3)**:485-495.
102. Kono M, Koga R, Shimada M, Fukatsu T: Infection dynamics of coexisting Beta- and Gammaproteobacteria in the nested endosymbiotic system of mealybugs. *Appl Environ Microbiol* 2008, **74(13)**:4175-4184.
103. Thao ML, Clark MA, Baumann L, Brennan EB, Moran NA, Baumann P: Secondary endosymbionts of psyllids have been acquired multiple times. *Curr Microbiol* 2000, **41(4)**:300-304.
104. Kolsch G, Matz-Grund C, Pedersen BV: Ultrastructural and molecular characterization of endosymbionts of the reed beetle genus *Macroplea* (Chrysomelidae, Donaciinae), and proposal of "*Candidatus* Macropleicola appendiculatae" and "*Candidatus* Macropleicola muticae". *Can J Microbiol* 2009, **55(11)**:1250-1260.
105. Kolsch G, Pedersen BV: Can the tight co-speciation between reed beetles (Col., Chrysomelidae, Donaciinae) and their bacterial endosymbionts, which provide cocoon material, clarify the deeper phylogeny of the hosts? *Mol Phylogenet Evol* 2010, **54(3)**:810-821.
106. Kuchler SM, Dettner K, Kehl S: Molecular characterization and localization of the obligate endosymbiotic bacterium in the birch catkin bug *Kleidocerys resedae* (Heteroptera *Lygaeidae*, Ischnorhynchinae). *FEMS Microbiol Ecol* 2010, **73(2)**:408-418.
107. Kuechler SM, Dettner K, Kehl S: Characterization of an obligate intracellular bacterium in the midgut epithelium of the bulrush bug *Chilacis typhae* (Heteroptera, Lygaeidae, Artheneinae). *Appl Environ Microbiol* 2011, **77(9)**:2869-2876.
108. Kikuchi Y, Fukatsu T: Endosymbiotic bacteria in the esophageal organ of glossiphoniid leeches. *Appl Environ Microbiol* 2002, **68(9)**:4637-4641.
109. Perkins SL, Budinoff RB, Siddall ME: New Gammaproteobacteria associated with blood-feeding leeches and a broad phylogenetic analysis of leech endosymbionts. *Appl Environ Microbiol* 2005, **71(9)**:5219-5224.
110. Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA, Hattori M: The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* 2006, **314(5797)**:267.
111. Thao ML, Baumann P: Evolutionary relationships of primary prokaryotic endosymbionts of whiteflies and their hosts. *Appl Environ Microbiol* 2004, **70(6)**:3401-3406.
112. Burke GR, Normark BB, Favret C, Moran NA: Evolution and diversity of facultative symbionts from the aphid subfamily Lachninae. *Appl Environ Microbiol* 2009, **75(16)**:5328-5335.
113. Tatusov RL, Koonin EV, Lipman DJ: A genomic perspective on protein families. *Science* 1997, **278(5338)**:631-637.
114. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 2003, **4**:41.
115. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP: MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res* 2010, , **38 Database:** D396-D400.
116. Gouy M, Guindon S, Gascuel O: SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 2010, **27(2)**:221-224.
117. Katoh K, Toh H: Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 2008, **9(4)**:286-298.
118. Talavera G, Castresana J: Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007, **56(4)**:564-577.
119. Castresana J: Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000, **17(4)**:540-552.
120. Hall TA: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 1999, **41**:95-98.
121. Lartillot N, Lepage T, Blanquart S: PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 2009, **25(17)**:2286-2288.
122. Boussau B, Gouy M: Efficient likelihood computations with nonreversible models of evolution. *Syst Biol* 2006, **55(5)**:756-768.
123. Galtier N, Gouy M: Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Mol Biol Evol* 1998, **15(7)**:871-879.
124. Galtier N, Gouy M: Inferring phylogenies from DNA sequences of unequal base compositions. *Proc Natl Acad Sci USA* 1995, **92(24)**:11317-11321.
125. Tarrio R, Rodriguez-Trelles F, Ayala FJ: Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. *Mol Biol Evol* 2001, **18(8)**:1464-1473.
126. Galtier N, Tourasse N, Gouy M: A nonhyperthermophilic common ancestor to extant life forms. *Science* 1999, **283(5399)**:220-221.
127. Shimodaira H, Hasegawa M: CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 2001, **17(12)**:1246-1247.
128. Swofford DL: PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, Massachusetts: Sinauer Associates; 2002.
129. Kostka M, Uzlíková M, Čepička I, Flegr J: SlowFaster, a user-friendly program for slow-fast analysis and its application on phylogeny of *Blastocystis*. *BMC Bioinformatics* 2008, **9**:341.
130. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003, **52(5)**:696-704.
131. Ronquist F, Huelsenbeck JP: MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003, **19(12)**:1572-1574.
132. Nylander JA, Wilgenbusch JC, Warren DL, Swofford DL: AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics* 2008, **24(4)**:581-583.
133. Rambaut A, Drummond AJ: Tracer v1.5. 2009 [http://tree.bio.ed.ac.uk/software/tracer/].
134. Abascal F, Zardoya R, Posada D: ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 2005, **21(9)**:2104-2105.
135. Posada D: jModelTest: phylogenetic model averaging. *Mol Biol Evol* 2008, **25(7)**:1253-1256.
136. Tavaré S: Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci* 1986, **17**:57-86.
137. Lanave C, Preparata G, Saccone C, Serio G: A new method for calculating evolutionary substitution rates. *J Mol Evol* 1984, **20(1)**:86-93.
138. Yang Z: Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 1994, **39(3)**:306-314.
139. Posada D: Using MODELTEST and PAUP* to select a model of nucleotide substitution. *Curr Protoc Bioinformatics* 2003, 6.5.1-6.5.14.
140. Le SQ, Gascuel O: An improved general amino acid replacement matrix. *Mol Biol Evol* 2008, **25(7)**:1307-1320.
141. Whelan S, Goldman N: A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 2001, **18(5)**:691-699.

142. Smyth P: **Model selection for probabilistic clustering using cross-validated likelihood.** *Stat Comput* 2000, **10(1)**:63-72.

doi:10.1186/1741-7007-9-87
Cite this article as: Husník *et al.*: Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. *BMC Biology* 2011 **9**:87.