Mondo, Stephen J.; Salvioli, Alessandra; Bonfante, Paola; Morton, Joseph B.; Pawlowska, Teresa E.. Nondegenerative evolution in ancient heritable bacterial endosymbionts of fungi. MOLECULAR BIOLOGY AND EVOLUTION. 33 (9) pp: 2216-2231.
DOI: 10.1093/molbev/msw086

Nondegenerative evolution in ancient heritable bacterial endosymbionts of fungi

Stephen J. Mondo, Alessandra Salvioli, Paola Bonfante, Joseph B. Morton and Teresa E. Pawlowska

Short title: Evolution in heritable endosymbionts

Key words: diversifying selection; effective population size; evolution rate; genetic drift; *Glomeribacter gigasporarum*; mutation rate; purifying selection; vertical transmission

ABSTRACT

Bacterial endosymbionts are critical to the existence of many eukaryotes. Among them, vertically transmitted endobacteria are uniquely typified by reduced genomes and molecular evolution rate acceleration relative to free-living taxa. These patterns are attributable to genetic drift-dominated degenerative processes associated with reproductive dependence on the host. The degenerative evolution scenario is well supported in endobacteria with strict vertical transmission, such as essential mutualists of insects. In contrast, heritable endosymbionts that are nonessential to their hosts and engage occasionally in horizontal transmission are expected to display deviations from the degenerative evolution model. To explore evolution patterns in such nonessential endobacteria, we focused on Candidatus Glomeribacter gigasporarum ancient heritable mutualists of fungi. Using a collection of genomes, we estimated in Glomeribacter mutation rate at 2.03×10-9 substitutions per site per year and effective population size at 1.44×108. Both fall within the range of values observed in free-living bacteria. To assess the ability of Glomeribacter to purge slightly deleterious mutations, we examined genome-wide dN/dS values and distribution patterns. We found that these dN/dS profiles cluster Glomeribacter with free-living bacteria as well as with other nonessential endosymbionts, while distinguishing it from essential heritable mutualists of insects. Finally, our evolutionary simulations revealed that the molecular evolution rate acceleration in Glomeribacter is caused by limited recombination in a largely clonal population rather than by increased fixation of slightly deleterious mutations. Based on these patterns, we propose that genome evolution in Glomeribacter is nondegenerative and exemplifies a departure from the model of degenerative evolution in heritable endosymbionts.

INTRODUCTION

Many eukaryotes rely on bacterial endosymbionts for diverse goods and services (Sachs, Skophammer, and Regus 2011). Among these prokaryotic associates, vertically transmitted endobacteria, regardless of phylogenetic origin, stand out as a group characterized by minimal genomes and rapid molecular evolution (Moran and Mira 2001; Moran, McCutcheon, and Nakabachi 2008; McCutcheon and Moran 2012). These two patterns are attributed to degenerative processes linked to reproductive dependence on the host. Endosymbionts are propagated through host reproductive structures, leading to transmission bottlenecks in every host generation (Mira and Moran 2002). In addition, endosymbiont populations associated with individual host lineages are reproductively isolated from each other, which results in endosymbiont clonality and population subdivision. All these phenomena are expected to reduce endosymbiont effective population size and magnify the impact of genetic drift relative to natural selection (Charlesworth 2009). With genetic drift acting as a major force that shapes genome evolution, slightly deleterious mutations fix rapidly (Ohta 1973), leading to the loss of gene functions, including DNA repair (McCutcheon and Moran 2012), which further accelerates accumulation of mutations. Due to the strong deletional bias prevailing in the bacterial genomes (Mira, Ochman, and Moran 2001; Kuo and Ochman 2009), mutation-compromised genes are eliminated and the genome contracts (Moran, McLaughlin, and Sorek 2009). Elevated fixation of slightly deleterious mutations is also expected to contribute to the enhanced molecular evolution rate observed in heritable endosymbionts compared to their free-living relatives (Ohta 1972). The scenario of degenerative genome size reduction and molecular evolution rate acceleration is well supported by empirical data from endobacteria with strict vertical transmission, such as essential (primary) mutualists of insects, exemplified by Buchnera aphidicola, which provides aphids with indispensable metabolites (Moran 1996; Moran, McLaughlin, and Sorek 2009; McCutcheon and Moran 2012). However, genome evolution in heritable endosymbionts of other, non-insect hosts, has not been explored deeply. Furthermore, the hypothesis of genetic drift-dominated evolution has not been tested extensively in heritable endosymbionts that are not essential to their host survival. In insects, such nonessential endobacteria include Hamiltonella defensa and Regiella insecticola defensive mutualists, which provide protection against natural enemies (Oliver et al. 2003; Scarborough, Ferrari, and Godfray 2005), as well as Wolbachia pipientis. Wolbachia endobacteria often manipulate host reproduction in favor of their own transmission and thus behave as parasites (Werren, Baldo, and Clark 2008; Engelstädter and Hurst 2009). However, they can also act as nutritional mutualists (Brownlie et al. 2009) and offer protection from viral infections (Hedges et al. 2008; Teixeira, Ferreira, and Ashburner 2008). While some nonessential heritable mutualisms may be evolutionarily stable (Mondo et al. 2012), others are believed to represent transition stages along an evolutionary trajectory leading to reciprocal partner dependence (Lamelas et al. 2011a). Thus, over evolutionary time, many nonessential mutualists are expected to become essential to their hosts' survival. Interestingly, population dynamics in nonessential endosymbionts appear to be different from those observed in essential heritable mutualists. In particular, nonessential endobacteria engage in processes known to avert effective population size reduction, including occasional horizontal transmission between hosts as well as gene exchanges and recombination (Rousset et al. 1992; Werren, Zhang, and Guo 1995; Baldo et al. 2006; Degnan and Moran 2008; Mondo et al. 2012). Yet, some of them, such as Wolbachia, show evidence of genome degeneration (Lambert and Moran 1998). Candidatus Glomeribacter gigasporarum (a beta-proteobacterium referred to hereafter as Glomeribacter) is a nonessential endosymbiont displaying evolution patterns that deviate from the degenerative evolution model (Castillo and Pawlowska 2010; Mondo et al. 2012), and, therefore, we selected it as a model to explore and document these patterns. Glomeribacter is one of many recently discovered endosymbionts of fungi, which also include Burkholderia rhizoxinica associated with Rhizopus microsporus (Partida-Martinez and

Hertweck 2005), and the unnamed endosymbiont of Mortierella elongata (Sato et al. 2010). Glomeribacter colonizes hyphae and spores of arbuscular mycorrhizal fungi (AMF) in the family Gigasporaceae (Bianciotto et al. 2003). AMF have maintained an obligate mututalism with the majority of plants since they evolved on land (Smith and Read 2008). These fungi facilitate plant mineral uptake in exchange for photosynthates. In turn, Glomeribacter is a nonessential mutualist that improves hyphal expansion of its fungal hosts prior to mycorrhizal colonization of plant roots (Lumini et al. 2007), which appears to be related to priming of fungal energy metabolism (Salvioli et al. 2016). Even though the Glomeribacter endobacteria are transmitted vertically in AMF host clonal lineages (Lumini et al. 2007), molecular evolution patterns suggest that they engage in infrequent horizontal transmission and recombination (Mondo et al. 2012). The Glomeribacter-Glomeromycota symbiosis is at least 400 million years old and, therefore, is remarkably evolutionarily stable despite its nonessential state (Mondo et al. 2012). Compared to freeliving Burkholderia relatives with genomes of 8 – 9 Mb (Winsor et al. 2008), the 1.7 Mb genome of Glomeribacter is highly reduced and exhibits signs of metabolic dependence on the fungal host (Ghignone et al. 2012). In contrast to most heritable endosymbionts with AT rich genomes (McCutcheon and Moran 2012), the Glomeribacter genome displays a moderate GC content of 54.8% (Ghignone et al. 2012). Suggestive that selection may act efficiently in Glomeribacter, the ribosomal RNA (rRNA) structures of these endosymbionts show no evidence of impaired stability due to accumulation of slightly deleterious mutations (Castillo and Pawlowska 2010). This observation is further supported by the presence of a nearly full complement of DNA repair genes in the Glomeribacter genome (Ghignone et al. 2012). Yet, paradoxically, the Glomeribacter lineage exhibits a significant acceleration of the molecular evolution rate relative to free-living Burkholderia, based on nucleotide substitution patterns in rRNA gene sequences (Castillo and Pawlowska 2010). To explore the role of genetic drift versus selection in the evolution of Glomeribacter, we estimated in these endobacteria the neutral mutation rate, μ, and effective population size, Ne. We also examined purifying selection against slightly deleterious mutations as well as diversifying selection in the Glomeribacter population. Finally, we modeled evolutionary processes under different mutation and recombination conditions in a population. We found that in Glomeribacter, μ and Ne fall within the range of values observed in free-living bacteria. We also established that these endobacteria appear to be as effective at purging slightly deleterious mutations as free-living bacteria. Our evolutionary simulations revealed that the evolutionary rate acceleration apparent in Glomeribacter may be independent of the nucleotide substitution rate and attributed to infrequent recombination in a predominantly clonal population. Overall, our findings indicate that genome evolution in Glomeribacter is driven by selection rather than by genetic drift.

RESULTS

Mutation rate in Glomeribacter is low and effective population size large.

To understand the impact of the nonessential heritable mutualist lifestyle on the magnitude of genetic drift relative to selection, we estimated in Glomeribacter the neutral mutation rate, μ, approximated from the synonymous nucleotide substitution rate, and effective population size, Ne. As these estimates can be confounded by deviations from the neutral evolution model and by changes in population demography, we conducted tests of neutrality, such as Tajima's D (Tajima 1989) as well as the Fu and Li's D* and F* (Fu and Li 1993). For these tests, we used sequences of ftsZ and pstA genes sampled from 19 Glomeribacter strains associated with seven AMF species (Mondo et al. 2012 and Fig. 1). We did not detect any significant departures from neutrality and from a constant population size (Table 1). Both μ and Ne, in addition to coalescent demographic modeling, can be inferred by computing point estimates from $\mu = Ds/2t$ and $Ne = \pi s/(2\mu)$, where Ds is synonymous site divergence between a focal population and a divergent reference population, t is the divergence time between these two populations, and πs is synonymous site diversity in the focal population. To avoid confounding effects of nucleotide substitution saturation at synonymous sites, we chose to focus our estimates on a subset of Glomeribacter strains associated with two genera of AMF, Gigaspora and Racocetera, and consider strains associated with the Cetraspora genus of AMF as a reference population. We calculated Ds = 0.8940 through comparing genomic sequences of Glomeribacter associated with AMF hosts Gigaspora margarita BEG34 (Ghignone et al. 2012), Gi. margarita JA201A-16, and R. castanea BEG1 with the genomic sequence of Glomeribacter from C. pellucida IN211. Genomic sequences of Glomeribacter from Gi. margarita JA201A-16, R. castanea BEG1, and C. pellucida IN211 were generated during this study (Fig. 2 and Fig. S1, Table S1). To estimate t, we modeled in BEAST (Bouckaert et al. 2014) the evolutionary history of 19 Glomeribacter strains using sequences of 16S rRNA, 23S rRNA, ftsZ and pstA genes (Mondo et al. 2012 and Fig. 1). This phylogenetic reconstruction revealed that the Gigaspora/Racocetra population of Glomeribacter diverged from the Cetraspora population of Glomeribacter at t = 220 MYA with the lower bound of the 95% highest posterior density (HPD) interval at 173 MYA and the upper bound of the 95% HPD interval at 276 MYA. Using the values of Ds and t, we estimated μ in the Gigaspora/Racocetra population of Glomeribacter to be $2.03\times10\text{-}9$ substitutions per site per year, with the 95% confidence interval, CI, from $1.62\times10\text{-}9$ to $2.59\times10\text{-}9$, based on the 95% HPD interval generated by BEAST for the divergence time between the Gigaspora/Racocetra versus Cetraspora populations of Glomeribacter (Fig. 1). In turn, Ne of the Gigaspora/Racocetra population of Glomeribacter was estimated at 1.44×108 (95% CI 1.13×108 – 1.80×108) based on πs = 0.5849 measured across genomic sequences of Glomeribacter from Gi. margarita BEG34 (Ghignone et al. 2012), Gi. margarita JA201A- 16, and R. castanea BEG1. In addition to point estimates of μ and Ne, we conducted in BEAST (Bouckaert et al. 2014) coalescent modeling of the Gigaspora/Racocetra population of Glomeribacter using sequences of ftsZ and pstA genes (Mondo et al. 2012). These analyses yielded a mean value of μ posterior distribution equal to $7.42\times10\text{-}10$ (95% HPD interval $4.46\times10\text{-}12$ – $2.89\times10\text{-}9$) and a mean value of the Ne posterior distribution equal to 7.67×107 (95% HPD 1.95×107 – 1.80×108). The coalescent estimates, while somewhat lower, are comparable to point estimates of μ and Ne. Importantly, as both these methods rely on the fossil record-based minimum date of the origin of the Glomeribacter-AMF symbiosis, the μ and Ne values are likely maximum estimates. For comparison, in the Buchnera aphidicola essential heritable mutualists of aphids $\mu = 2.2\times10\text{-}7$ mutations per site per year (Moran, McLaughlin, and Sorek 2009) and $Ne = 1.0\times107$ (Funk, Wernegreen, and Moran 2001). In the Escherichia coli free-living enterics, $\mu = 4.5\times10\text{-}9$ (Ochman, Elwyn, and Moran 1999) and $Ne = 2.5\times107$ (Charlesworth and Eyre-Walker 2006). In marine picocyanobacteria Prochlorococcus and Synechococcus, Ne ranges from 1.01×1011 to 1.42×1011, respectively (Baumdicker, Hess, and Pfaffelhuber 2012).

Glomeribacter is effective in purging of slightly deleterious mutations

Our findings indicating that μ and Ne in Glomeribacter are similar to those in free-living bacteria suggest that selection may be active in the Glomeribacter population. Traditionally, selection is quantified by assessing the relationship between the number of nonsynonymous substitutions per nonsynonymous site, dN, and the number of synonymous substitutions per synonymous site, dS (Nielsen 2005). Values of dN equal to dS signify neutral evolution. A deficit of dN relative to dS is indicative of purifying selection. The reverse, an excess of dN relative to dS is more challenging to interpret as it may represent evidence of diversifying selection in a population where selection is unimpaired, or accumulation of slightly deleterious mutations in a population where genetic drift is magnified relative to selection (Moran 1996). However, when these ratios are considered on a genome-wide scale, the uncertainty concerning their interpretation is easily resolved. Specifically, the excess of dN relative to dS on the single gene level may be indicative of diversifying selection. However, when observed genome-wide, it is more likely a sign of widespread amino acid change and over-accumulation of slightly deleterious mutations due to genetic drift, as it is unlikely that diversifying selection acts on every gene in the genome. We examined values of dN/dS across three Glomeribacter genomes (BEG1, BEG34 and JA201A-16) and compared them to those in bacterial lineages with different lifestyles, including nonessential heritable endosymbionts of insects (Hamiltonella/Regiella and Wolbachia), essential heritable mutualists of insects (Blochmannia, Buchnera and Sulcia), bacteria free living in the soil (Burkholderia) and in the ocean (Prochlorococcus) as well as microbes with free-living capacities forming associations with plant roots (Bradyrhizobium) and human gut, such as enteric bacteria (Enterobacter/Escherichia/Salmonella) and lactic acid bacteria (Bifidobacterium and Lactobacillus) (Table 2). These specific lineages were chosen based on relatedness, such that pairwise genome-wide dS values calculated using the NG86 method (Nei and Gojobori 1986) were greater than 0.25 and less than 1.25 when possible (Table S2). As with Glomeribacter, only high quality single copy orthologs were retained for downstream analyses. In each lineage, represented by three genomes, all single copy orthologous coding sequences were collected, and aligned to compute an average genome-wide dN/dS value. Calculations of dN/dS are expected to be sensitive to the nucleotide transition/transversion rate bias, κ, and to the codon usage bias (Yang 2007b). Moreover, the effects of these biases on dN/dS can compensate each other. Consequently, we estimated dN/dS values using three approaches that differ in handling the two biases. NG86 disregards both of them, LWL85m is designed to accommodate κ (Tzeng, Pan, and Li 2004), whereas YN00 is expected to account for both κ and codon usage (Yang and Nielsen 2000). To understand the impact of both κ and codon usage on dN/dS across different microbial lineages, we computed κ and quantified codon bias. We found that in Glomeribacter $\kappa = 5.2$ was somewhat higher than in other microbial lineages except Wolbachia with $\kappa = 10.1$ (Table 3). A several fold bias favoring transitions over transversions is common in molecular evolution and attributed to a transition/transversion bias in mutation (Stoltzfus and Norris 2015). It remains unclear why in this respect Glomeribacter and Wolbachia stand out among other microbes included in the study. To quantify codon bias in microbes with diverse lifestyles, we calculated the effective number of codons, ENC, which can extend from 20, under the extreme bias of one codon used per amino acid, to 61, under no bias with equally likely usage of alternate codons (Wright 1990), as well as the codon bias index, CBI, which can range from 0, under the uniform use of synonymous codons, to 1, under the maximum bias (Morton 1993). We found that, with ENC = 49 and CBI = 0.33, Glomeribacter appeared to experience a low codon usage bias, a pattern shared with other nonessential heritable endosymbionts and enteric bacteria (Table 3). In contrast, essential heritable mutualists and some free-living microbes exhibited a substantial bias (Table 3). While in essential mutualists the bias is likely the result of the general relaxation of selection on codon use (Wernegreen and Moran 1999; Herbeck, Wall, and Wernegreen 2003), in free-living bacteria it may be related to either GC content or rapid growth rates,

as both of these features are correlated with increased codon bias (Sharp, Emery, and Zeng 2010; Ran, Kristensen, and Koonin 2014). Consistent with previous studies (Herbeck et al. 2003), our results showed high genome-wide dN/dS ratios in essential heritable mutualists, an outcome apparent across all three estimation methods, albeit no dN/dS value could be computed for Buchnera under the YN00 model due to dS saturation (Table 3 and Table S2). In contrast, Glomeribacter displayed dN/dS values comparable to those in nonessential heritable endosymbionts. However, these values varied widely and inconsistently between the methods, suggesting that comparisons of genome-wide dN/dS estimates across lineages may not be a reliable approach for discerning the role of drift versus selection. Instead, we examined the underlying distributions of dN/dS values estimated for individual genes across the genomes. We found that in Glomeribacter and nonessential heritable endosymbionts of insects, the distribution of dN/dS across the genome was clustered around the mean regardless of the dN/dS estimation method (Fig. 3, Fig. S2 and S3). This pattern was similar to that found in free-living bacterial lineages and distinct from the more platykurtic distributions in essential heritable mutualists of insects (Fig. 3, Fig. S2 and S3). To further explore differences between distributions of dN/dS values across the microbial genomes, we conducted a quantile-quantile (QQ) plot analysis, which allowed for comparison of two distributions by plotting their quantiles against each other (Sokal and Rohlf 1998). Using the results of the NG86 and LWLm methods, we contrasted the dN/dS distribution across the genome of Glomeribacter and other nonessential heritable endosymbionts (Hamiltonella/Regiella and Wolbachia) with those in free-living bacteria (Burkholderia) and essential heritable mutualists (Buchnera). The Burkholderia species were chosen to represent freeliving bacteria for two reasons: (i) they have the largest genome sizes of all lineages surveyed, which allowed us for generating a robust reference dN/dS profile, and (ii) the genome-wide estimates of dN/dS in free-living Burkholderia (Table 3 and Table S2) were not at one of the extremes observed in free-living bacteria, which provides a more accurate representation of what may typically be expected from this group. Buchnera was selected to represent essential heritable mutualists because genome degeneration has been empirically studied and validated in this lineage (Moran 1996; Moran, McLaughlin, and Sorek 2009), making it an ideal reference for comparing patterns of mutation accumulation with nonessential heritable endosymbionts. We found a strong similarity of dN/dS distributions between nonessential heritable endosymbionts, including Glomeribacter, and free-living Burkholderia (Fig. 4A, 4C, 4E and Fig. S4A, S4C, S4E). In contrast, dN/dS distributions differed markedly between nonessential heritable endosymbionts and Buchnera essential heritable mutualists (Fig. 4B, 4D, 4F and S4B, S4D, S4F). These observations suggest that, while the footprints of purifying selection are similar between the genomes of Glomeribacter, other heritable nonessential endosymbionts and free-living bacteria, the genomes of heritable essential mutualists deviate from this pattern. In Glomeribacter, nonessential heritable endosymbionts of insects and free-living bacteria, large portions of the genome exhibit a deficit of dN relative to dS, which is strongly indicative of purifying selection. Only small fractions of these genomes show an excess of dN relative to dS, which may represent diversifying selection. In contrast, the excess of dN relative to dS in essential heritable mutualists extends across considerably larger portions of the genome compared to that in Glomeribacter and other nonessential heritable endosymbionts, suggesting that these two groups of endobacteria are accumulating nucleotide substitutions in a distinctly different manner.

Genes under diversifying selection in Glomeribacter

To explore whether, in addition to purifying selection, the Glomeribacter genomes experience diversifying selection, we conducted a genome-wide screen for positively selected genes using PAML (Yang 2007a). We discovered 68 putative genes under diversifying selection (Table S3). Many of these genes encoded outer membrane proteins or enzymes involved in energy metabolism. This pattern is similar to what was reported in Wolbachia nonessential endosymbionts (Brownlie et al. 2007). We further chose a set of 17

genes with signatures of positive selection to examine whether they are functional (Table S4). We found that 15 of them were expressed by Glomeribacter during the presymbiotic growth of its Gi. margarita BEG34 host (Fig. S5). To assess a potentially confounding impact of recombination on identifying genes under diversifying selection, we estimated ρ/θ and r/m using ClonalFrame (Didelot and Falush 2007) for all genes experiencing diversifying selection. We found that values of ρ/θ = 0.0099 (95% CI 0.00961 – 0.010047) and r/m = 0.85 (0.799 – 0.900) estimated for genes under diversifying selection were comparable to both ρ/θ = 0.0161 (0.01237 – 0.02014) and r/m = 0.92 (0.748 – 1.097) inferred for 40 randomly selected single copy gene ortholog clusters excluding genes under selection as well as to genome-wide estimates based on 402 single copy gene orthologs (Table 1). These results suggest that, while recombination may have a small effect on detection of genes under selection, this impact is likely minimal.

Molecular evolution rate acceleration in Glomeribacter

Molecular evolution rate acceleration relative to free-living taxa is one of the hallmarks of evolution in heritable endobacteria (Moran, McCutcheon, and Nakabachi 2008). This pattern is a consequence of enhanced fixation of slightly deleterious mutations in populations of small Ne (Moran 1996). The nonessential Glomeribacter mutualists also exhibit apparent evolution rate acceleration relative to freeliving Burkholderia, based on the nucleotide substitution patterns in rRNA genes (Castillo and Pawlowska 2010). By conducting Tajima's 1D relative rate tests on sequences at 16S rRNA, 23 rRNA and 25 protein-coding loci sampled from the genomes of Glomeribacter and its close relatives (Fig. 2), we confirmed that this rate acceleration is a genome-wide phenomenon (Table S5). The low μ and the large Ne of the Glomeribacter population as well as its ability to effectively purge slightly deleterious mutations suggest that the underlying cause of this apparent evolution rate acceleration is different than in essential heritable mutualists. We hypothesized that the long-term maintenance of a largely clonal population coupled with infrequent recombination (Mondo et al. 2012) is responsible for rate acceleration in Glomeribacter. To test this hypothesis, we conducted forward simulations using SFS_CODE (Hernandez 2008), which can evolve a neutral fragment of DNA over time under a variety of models. We fixed the population mutation rate, θ, at 0.1791, the value observed across the three Glomeribacter genomes (BEG1, BEG34, and JA201A-16). We then varied recombination rate, ρ, from no recombination with ρ = 0 to high recombination with ρ = 10. To simulate recombination in Glomeribacter, we included ρ = 0.001791, which we calculated based on the ρ/θ = 0.01 (Table 1) estimated by ClonalFrame (Didelot and Falush 2007). For comparison, we also included parameters for the Ca. Tremblaya princeps essential heritable mutualists of insects and close relatives of Glomeribacter (Fig. 2). Even though the symbioses formed by Tremblaya and Glomeribacter are distinctly different, development and generation times are comparable between their mealybug (Chong, Roda, and Mannion 2008) and AMF (Smith and Read 2008) hosts. The rate of evolution in Tremblaya is significantly more rapid than in Glomeribacter (Table S5), with rRNA gene sequences evolving 2.3 times faster (Castillo and Pawlowska 2010). Consequently, we set θ = 0.4119 and ρ = 0 to simulate evolutionary history of the Tramblaya lineage. Under each evolutionary scenario, we determined the patristic distance, PD, between the ancestral sequence used to initiate the simulation and each individual descendant sequence at the end of the simulation (root-to-tip distance) for ten independent simulations. We then used these data to compute a mean PD with 95% confidence interval. We found that the highest recombination rate ρ = 10 resulted in the lowest average root-to-tip distance PD = 0.189 (95% CI 0.187 – 0.190; Fig. 5A). In clonal populations with ρ = 0, branch lengths increased to PD = 0.209 (95% CI 0.201 – 0.217; Fig. 5B). Remarkably, in simulations under the parameters measured in Glomeribacter, we observed a dramatic increase in average root-to-tip branch length reaching PD = 0.284 (95% CI 0.276 – 0.291; Fig. 5C). This result suggests that low levels of recombination can play a large role in enhancing total branch length. Finally, consistent with empirical observations (Castillo and Pawlowska 2010 and Fig. 2),

branches were longest under parameters measured in Tremblaya essential heritable mutualists with PD = 0.482 (95% CI 0.465 – 0.499; Fig. 5D). Notably, simulations under low recombination rates yielded genealogies displaying a sizable variance in branch length (Fig. 5C and 5D), a pattern similar to that in genealogies of Glomeribacter (Fig. 1) and Buchnera (Lamelas et al. 2011b). Collectively, our simulations suggest that the molecular evolution rate acceleration apparent in Glomeribacter could be attributed to recombination events occurring infrequently in a largely clonal population.

DISCUSSION

Genome evolution is nondegenerative in Glomeribacter

Our results revealed that in the Glomeribacter nonessential heritable mutualists of fungi μ was low and Ne large. Furthermore, these endobacteria appeared to be effective at purging slightly deleterious mutations from their genomes. Consequently, we conclude that genome evolution in Glomeribacter is a consequence of an adaptive process rather than of genetic drift that dominates evolution in essential mutualists with strict vertical transmission. The low μ in Glomeribacter can be attributed to the preservation of DNA repair capabilities in these endobacteria (Ghignone et al. 2012). Conservation of these functions, in turn, appears to be a consequence of the large Ne of the Glomeribacter population and efficient selection against slightly deleterious mutations. For this reason, it is important to consider what features of these endobacteria permit a large Ne. Certainly, horizontal transmission and recombination are critical to increasing Ne in nonessential heritable endosymbionts. Essential heritable mutualists lack both of these characteristics. However, as is the case in Glomeribacter, horizontal transmission and recombination may occur at fairly low rates (Mondo et al. 2012). Consequently, the nonessential endosymbionts likely experience long periods of vertical transmission before the opportunities for cross-lineage recombination or host switching arise. If Ne is to remain large over extended evolutionary time, Glomeribacter must have additional mechanisms that contribute to purging of slightly deleterious mutations from their populations. One such mechanism of reducing the population load of deleterious mutations may be related to the number of bacterial cells that are transmitted from one host generation to the next. Evolutionary theory suggests that, as long as selection operates on bacterial cells within the host, increasing the number of endosymbiont cells packaged with host propagules will reduce the load of mutations deleterious to both the endosymbiont and the host (Otto and Orive 1995; Rispe and Moran 2000; Roze and Michod 2001). The number of Glomeribacter cells per AMF spore ranges from ~3,500 to 26,000 (Jargeat et al. 2004). These numbers are considerably larger than ~800 cells that are considered to represent the bottleneck size in the Buchnera essential mutualists of insects (Mira and Moran 2002). The difference in the number of bacteria transmitted in these two symbiotic systems from one host generation to the next suggests that the size of a transmission bottleneck may play an important role in increasing the effective population size in Glomeribacter.

Diversifying selection in Glomeribacter

In addition to purifying selection, the Glomeribacter genomes appear to experience diversifying selection. A surprisingly large number of positively selected genes are involved in energy metabolism, which may be related to the absence in Glomeribacter of a key enzyme of glycolysis, phosphofructokinase (Ghignone et al. 2012). With a limited capacity to utilize sugars as a source of energy, Glomeribacter may be using alternative substrates for energy production, with selection driving this pathway towards increased affinity for such substitutes. Moreover, benefits provided by Glomeribacter to their fungal hosts seem to be connected to energy metabolism (Salvioli et al. 2016), which could make selection on Glomeribacter energy-related genes important for the host. In particular, the presence of endosymbionts is known to improve the extension of fungal germ tubes (Lumini et al. 2007). Germinating fungal spores do not experience energy inputs prior to plant root colonization, and appear to benefit from Glomeribacter-mediated mobilization of resources fueling rapid presymbiotic hyphal growth.

Similarities with free-living bacteria

Molecular evolution patterns cluster Glomeribacter with free-living bacteria, with a particular similarity to pelagic bacteria, such as Prochlorococcus marinus (Baumdicker, Hess, and Pfaffelhuber 2012) and Synechococcus sp. (Dufresne et al. 2003). These bacteria are characterized by reduced genomes, mutation rates that are not different from the rates observed in other free-living bacteria, such as E. coli (Osburne et al. 2011), and large effective population sizes (Baumdicker, Hess, and Pfaffelhuber 2012). In free-living bacteria, the loss of abilities to synthesize costly metabolites that are readily available in the environment is generally associated with fitness advantages over genotypes that retain such biosynthetic functions (Zamenhof and Eichhorn 1967). Experimental evolution studies conducted in specialized environments under conditions of large effective population sizes indicate that selective pressures generated by such fitness advantages may drive elimination of genes that are dispensable in these particular environments, leading to genome size reduction (Koskiniemi et al. 2012; Lee and Marx 2012). Moreover, the resulting genotypes show improved fitness relative to their ancestors. Such outcome is distinctly different from what is observed in populations repeatedly subjected to severe transmission bottlenecks, and experiencing genome reduction in the process (Nilsson et al. 2005). While both evolutionary trajectories lead to genome contraction, accompanied by increased dependence on external resources and by the loss of lifestyle flexibility, evolution in populations of small effective sizes is associated with continued decrease in fitness (Nilsson et al. 2005; Koskiniemi et al. 2012; Lee and Marx 2012). In pelagic bacteria, genome contraction is attributed to an adaptive process in which genes involved in synthesis of metabolites already available in the environment become dispensable, so that their loss confers a fitness advantage (Morris, Lenski, and Zinser 2012). In the case of pelagic bacteria, other members of the microbial community are believed to provision products of such dispensable genes. Similarly, molecular evolution patterns apparent in Glomeribacter suggest that, even though the genomes of these microbes are highly reduced and their lifestyle is severely restricted, Glomeribacter populations are of a considerable effective size and experience efficient selection. Consequently, despite striking lifestyle differences, these heritable nonessential mutualists seem to resemble free-living pelagic bacteria rather than heritable essential mutualists of insects. Similarities with other nonessential endosymbionts Like Glomeribacter, nonessential insect endosymbionts (Hamiltonella/Regiella, and Wolbachia) appeared to share mutation accumulation patterns with free-living bacteria. However, without data on mutation rate and effective population size in these nonessential endobacteria, it remains unclear whether they share a similar mode of evolution with Glomeribacter. For example, our inferences regarding deleterious mutation accumulation in Wolbachia are not entirely consistent with destabilization of rRNA structures observed previously in these bacteria (Lambert and Moran 1998). This discrepancy is intriguing as recombination is considered to be intense in Wolbachia (Atyame et al. 2011), with the overall per-site effect of recombination relative to mutation, r/m = 3.57, and r/m values reaching over 8 in individual strains (Ellegaard et al. 2013). Rapid accumulation of mutations combined with a high recombination rate may indicate that the mechanisms of Wolbachia genome evolution are subtly different from those in nonessential mutualists and may reflect differences between the lifestyles of antagonists and mutualists. Rapid evolution and recombination are expected to be favored in antagonistic microbes, such Wolbachia, as they facilitate overcoming host defenses (Jaenike 1978; Morran et al. 2011). In contrast, maintenance of recombination in nonessential heritable mutualists is considered advantageous because it protects endosymbionts from the loss of capacity for horizontal transmission, which, in turn, permits recolonization of host lineages that may have lost their partners due to environmental variability (Mondo et al. 2012).

Evolutionary rate acceleration in Glomeribacter

In phylogenetic reconstructions, most heritable endobacteria, including nonessential endosymbionts, show an apparent molecular evolution rate acceleration relative to free-living sister taxa (Moran, McCutcheon, and Nakabachi 2008). In essential heritable mutualists, such as Buchnera, rate acceleration relative to free-living taxa, like E. coli, is caused by exceptionally high supply and rapid fixation of slightly deleterious mutations. Remarkably, generation time in Buchnera is considerably longer than in E. coli. In contrast to 100 generations produced yearly in nature by E. coli, Buchnera undergoes only 30 to 50 doublings per year (Clark, Moran, and Baumann 1999). This generation time difference underscores the dramatic disparity in mutation rate per generation that is needed to generate evolution rate acceleration in absolute time. Accordingly, mutation rate in Buchnera is estimated at 4×10-9 substitutions per site per replication (Burke et al. 2010) versus 8.9×10-11 substitutions per site per replication in E. coli (Wielgoss et al. 2011). While the generation time in Glomeribacter is unknown, its mutation rate per year appears to be comparable to that of E. coli (Ochman, Elwyn, and Moran 1999) and considerably lower than in Buchnera (Moran, McLaughlin, and Sorek 2009). These observations, together with Glomeribacter's ability to effectively purge slightly deleterious mutations, suggest that a mechanism different from rapid supply and fixation of mutations is responsible for acceleration of its molecular evolution rate. By simulating evolutionary processes under a range of population recombination conditions, we found that rate acceleration in Glomeribacter could be attributed to low levels of recombination in a largely clonal population. Importantly, these simulations were focused on the impact of mutation and recombination on phylogeny and ignored the effects of natural selection. Nevertheless, their outcomes are rather intuitive. In a predominantly clonal population, two lineages evolving separately for a prolonged period of time will accumulate many distinct mutations independently. As a consequence, a recombination event between these two lineages will introduce a large amount of novel genetic material to the recipient lineage, and lead to increased phylogenetic distance between the recombinant strain and its closest relative. This process appears evident in Glomeribacter. While the rate of recombination in these endobacteria is low, the per site effect of recombination is substantial (Mondo et al. 2012). Our genome-wide estimate of r/m = 0.48 (Table 1) indicates that nearly half of all nucleotide substitutions are likely the result of recombination rather than mutation. Consequently, we propose that in Glomeribacter enhanced nucleotide substitution rates are not required to explain the increase of molecular evolution rate relative to the free-living taxa. While we are not aware of other simulation studies linking the average root-to-tip distance of a phylogenetic tree with the recombination rate across the taxonomic units, the effects of recombination on the tree topology have been assessed before. For example, Schierup and Hein (2000) reported that trees constructed from sequences simulated under the coalescent with recombination displayed terminal branches that were longer and the time to the most common ancestor that was shorter than in trees reconstructed from sequences evolved without recombination. These patterns are consistent with tree topologies generated from sequences evolved under a range of recombination regimes in our study.

Conclusion

We found that the Glomeribacter heritable endobacteria of fungi exhibit a low mutation rate and maintain a large effective population size. Consequently, they are capable of effectively purging slightly deleterious mutations from their genomes. Elevated substitution rates are not required to explain the molecular evolution rate acceleration in Glomeribacter compared to free-living relatives. Instead, this pattern appears to result from a predominantly clonal lifestyle punctuated by rare recombination events. Overall, we found that evolution in Glomeribacter is adaptive rather than degenerative and exemplifies a departure from the model of genetic drift-dominated evolution in heritable endosymbionts.

MATERIALS & METHODS

Sequencing of the Glomeribacter genomes, assembly and annotation

To represent a broad phylogenetic diversity of Glomeribacter, we selected strains associated with G. margarita JA201A-16, S. castanea BEG1, and C. pellucida IN211 for genome sequencing. For DNA preparation, 50-150 spores of each AMF, obtained from the International Arbuscular Mycorrhizal Culture Collection (INVAM), West Virginia University, were decontaminated as described by Mondo et al. (2012), crushed in 500 µL of water, and passed through a 2 µm filter to eliminate fungal nuclei. Following filtration, spore contents were concentrated using the SpeedVac Concentrator (ThermoFisher Scientific, Waltham, MA) to a final volume of 50 µL. 2 µL of concentrated filtrate per reaction were then used for 5 independent 20 µL whole genome amplification (WGA) reactions with the Illustra GenomiPhi 2 DNA Amplification Kit (GE Healthcare Life Sciences, Piscataway, NJ). Chimeric WGA amplicons were eliminated using the approach of Gilbert et al. (2010), followed by quantification and pooling of equal amounts of DNA from each independent WGA reaction to a total of 1 µg to enhance evenness of coverage across the genome (Zhang et al. 2006). 300-400 bp sequencing libraries were prepared using the Illumina TruSeq DNA Sample Prep Kit and sequenced using the Illumina HiSeq paired-end 100 bp platform (Illumina, San Diego, CA) at the Cornell University Core Laboratories Center. Due to the large volume of sequence data recovered, raw reads were split into ten 100x coverage subsets and assembled independently using Velvet 1.1 (Zerbino and Birney 2008). Contigs were assembled using Geneious 5.4 (Biomatters Ltd.) and surveyed for predicted coding sequences using AMIGene (Bocs et al. 2003) under the Glomeribacter gene model (Ghignone et al. 2012). The recovered gene models were additionally functionally annotated using PFAM (Finn et al. 2014) and SwissProt (Bairoch and Apweiler 2000) using E-value cutoff of 1e-5 (Table S1).

Glomeribacter evolutionary history

Evolutionary history of Glomeribacter was reconstructed using nucleotide sequences at 16S rRNA, 23S rRNA and 25 protein-coding loci (nusA, pyrG, rplA, rplB, rplC, rplD, rplE, rplF, rplK, rplL, rplM, rplN, rplP, rplS, rplT, rpmA, rpoB, rpsB, rpsC, rpsE, rpsI, rpsJ, rpsK, rpsM, rpsS) sampled from the genomes of Glomeribacter and its close relatives, which include free-living Burkholderia as well endosymbionts of fungi and insects. Sequences were aligned using MUSCLE (Edgar 2004) and concatenated. Phylogenetic trees were

constructed in MrBayes 3.2 (Ronquist et al. 2012) under the Generalized Time Reversible (GTR) nucleotide substitution model plus invariant sites I and Γ rate heterogeneity (Tavaré 1986), with analyses run for 10,000,000 generations with a 25% burn-in. The average standard deviation of split frequencies was used as a convergence diagnostic. Maximum Likelihood trees were constructed using RAxML 8 (Stamatakis 2014) with 1000 bootstrap replicates under the GTR+Γ nucleotide substitution model. The divergence time, t, between the Glomeribacter strains associated with the Gigaspora and Racocetra versus Cetraspora genus of AMF was inferred in BEAST 2.1.3 (Bouckaert et al. 2014) based on the 16S rRNA, 23S rRNA, ftsZ and pstA genes as described in Mondo et al. (2012). LogCombiner 2.1.3 was used to merge three independent runs of 250,000,000 steps with the burn-in of 25% each under the uncorrelated lognormal relaxed clock model (Drummond et al. 2006), the GTR+I+Γ nucleotide substitution model (Tavaré 1986), and the calibration point at the 396 ± 12 MYA, corresponding to the origin of the Glomeribacter-Glomeromycota symbiosis (Mondo et al. 2012).

Estimating μ and Ne in Glomeribacter

We took two approaches to estimating neutral mutation rate, μ, approximated from the synonymous nucleotide substitution rate (Kasuga, White, and Taylor 2002), and effective population size, Ne, in the Gigaspora/Racocetra population of Glomeribacter. First, we made point estimates of μ and Ne from $\mu = Ds/2t$ and $Ne = \pi s/(2\mu)$, where Ds is synonymous nucleotide divergence, t is the divergence time between the Gigaspora/Racocetra versus Cetraspora populations of Glomeribacter, and πs is synonymous site diversity. Ds and πs were computed based on genome-wide comparisons involving Glomeribacter genomes from Gi. margarita BEG34 (Ghignone et al. 2012), Gi. margarita JA201A-16, R. castanea BEG1 and C. pellucida IN211. Amino acid (AA) sequences were first subjected to an All vs. All BLASTp (Altschul et al. 1990) with parameters: E value cutoff of 1e-10, maximum matches = 500. BLASTp comparisons were conducted at the Cornell University Computational Biology Service Unit (Sun et al. 2010) and then used for identification of orthologs with OrthoMCL 1.4 (Li, Stoeckert, and Roos 2003). To visualize ortholog conservation across sequenced Glomeribacter lineages, Venn diagrams were produced using the VennDiagram 1.6.16 package for R 3.2 (Fig. S1). To ensure the removal of paralogs from OrthoMCL clusters and increase the number of single copy orthologs, we used stringent parameters for OrthoMCL (mode = 3, pi_cutoff = 40, pv_cutoff = 1e30, and inflation = 5). AA orthologs were aligned using MUSCLE (Edgar 2004). Protein alignments were converted to PHYLIP codon alignments in PAL2NAL (Suyama, Torrents, and Bork 2006). Poor alignments were removed from the dataset, alignments with frame shifts were trimmed according to Orsi et al. (2008), and curated codon alignments were concatenated. Additionally, these orthologs were sorted based on the physical locations in the Glomeribacter BEG1 genome assembly (Ghignone et al. 2012) and used to infer the population-level recombination rate, ρ/θ, as well as the per-site effect of recombination relative to mutation, r/m, in ClonalFrame (Didelot and Falush 2007) run for 200,000 generations after an initial burn-in of 200,000. Second, we sampled jointly posterior distribution of μ and Ne in a demographic model of the Gigaspora/Racocetra population of Glomeribacter with no outgroup using BEAST 2.1.3 (Bouckaert et al. 2014) based on ftsZ and pstA gene sequences described in Mondo et al. (2012). LogCombiner 2.1.3 (Bouckaert et al. 2014) was used to merge three independent MCMC runs of 100,000,000 steps with the burn-in of 25% each under the uncorrelated lognormal relaxed clock model (Drummond et al. 2006), the GTR+I+Γ nucleotide substitution model (Tavaré 1986), and calibration point at 180 ± 45 MYA marking the diversification of the Gigaspora/Racocetra population of Glomeribacter (Fig. 1). DnaSP 5 (Librado and Rozas 2009) was used to extract information on nucleotide

diversity and divergence as well as to conduct tests of neutrality, including Tajima's D (Tajima 1989) as well as the Fu and Li's D* and F* (Fu and Li 1993).

Deleterious mutation accumulation across bacterial lifestyles

For each lineage surveyed (Table 2), single copy orthologs were identified as described in section "Estimating μ and Ne in Glomeribacter". For each ortholog cluster, the values of κ as well as dN, dS and dN/dS under three different methods, NG86 (Nei and Gojobori 1986), LWL85m (Tzeng, Pan, and Li 2004), and YN00 (Yang and Nielsen 2000), were estimated using the yn00 module of PAML 4.8 (Yang 2007a) under default parameters. These results were filtered to remove orthologs where dS was saturated (dS > 1.5) and orthologs that displayed little to no divergence from one another (dS < 0.1). To quantify similarities between genome-wide dN/dS distribution patterns in different microbes, we conducted aquantile-quantile (QQ) plot analysis. The degree of similarity between genome-wide dN/dS distributions was assessed by calculating the correlation coefficient R2. DnaSP 5 (Librado and Rozas 2009) was used to measure codon usage bias by calculating the codon bias index, CBI (Morton 1993), and the effective number of codons, ENC (Wright 1990).

Diversifying selection in Glomeribacter

To investigate the genes under diversifying selection in Glomeribacter, we collected single copy orthologs across the JA201A-16, BEG1, and BEG34 genomes as described above, then surveyed each ortholog individually for evidence of positive selection using PAML 4.4d (Yang 2007a). For this analysis, we used the codeml algorithm with parameters: NSsites = 0 1 2 3 7 8, CodonFreq = 2, seqtype =1, kappa = 0.3, omega = 1.3, ncatG = 10. To enhance confidence in our results, we chose the most conservative model (M7 vs M8) to test for genes experiencing positive selection and used the likelihood ratio test to determine significance (df = 2, $\chi 2$ critical value = 6, equivalent to P = 0.05). All PAML analyses were conducted on the BIOSIM cluster at the Cornell University Computational Biology Service Unit (Sun et al. 2010). Specific AA residues under positive selection were identified using the Bayes Empirical Bayes algorithm integrated into PAML (Yang 2007a). As the draft genome of Glomeribacter IN211 was substantially smaller than the other three genomes, it was excluded from this analysis. Additionally, to investigate the potential influence of recombination on predicted genes under selection, ClonalFrame (Didelot and Falush 2007) was run using default parameters for 200,000 generations after an initial burn-in of 200,000. 40 ortholog clusters not experiencing positive selection and selected at random were subjected to a similar analysis to assess the amount of recombination experienced by individual genes in the genome. To determine whether the putative positively selected Glomeribacter genes (Table S3) were not pseudogenes, we analyzed expression in a subset of 17 of them. DNA from G. margarita BEG34 wild type and cured spores was extracted according to Lumini et al. (2007). The primers (Table S4) were tested on the extracted material to assess amplification specificity, using the DNA from cured spores as a negative control. RNA was extracted from three independent batches of 100 wild type and cured germinating spores to examine expression patterns of the target genes. The extraction was performed with the RNeasy Microarray Tissue kit (Qiagen, Germany). The samples were then treated with the TURBO DNA-free DNAse (Life Technologies, Carlsbad, CA) and retrotranscribed with the Superscript II Reverse Transcriptase using random primers (Life Technologies) according to the manufacturer's instructions. The obtained cDNA was used as a template for the amplification of the target genes. Individual PCR reactions were assembled in a 20 μl reaction volume

with a mixture of 200 µM dNTPs, 0.5 µM of each oligonucleotide primer, 2 µl of 10x PCR Buffer, 0.5 µl of HotStarTaq DNA polymerase (Qiagen) and 1 µl of cDNA template. The amplification conditions included an initial step of 15 min at 95°C followed by 35 cycles of 45 sec at 94°C, 45 sec at 58°C and 2 min at 72°C, with a final extension of 10 min at 72°C. The obtained amplified fragments were visualized in a 1.5% agarose gel stained with ethidium bromide.

Molecular evolution rate acceleration in Glomeribacter

To explore whether molecular evolution rate acceleration apparent in Glomeribacter's rRNA gene sequences relative to free-living Burkholderia (Castillo and Pawlowska 2010) is a genome-wide phenomenon, we used sequences at 16S rRNA, 23S rRNA and 25 protein-coding loci to conduct Tajima's 1D relative rate test implemented in MEGA7 (Kumar, Stecher, and Tamura 2016). To assess the effects of recombination on the rate of molecular evolution, we simulated population evolutionary histories under various values of parameters θ and ρ using a forward evolution simulator SFS_CODE (Hernandez 2008). Aside from θ and ρ, each simulation was run using default parameters for at least 40 iterations on a 1 kb fragment of DNA evolving neutrally. As SFS_CODE does not offer a way to simulate recombination in a haploid model, we used the cross-over model, where ploidy = 2 was required, but conducted all subsequent analyses on the single chromosome level. Phylogenies were then constructed using MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003) under the GTR nucleotide substitution model. Analyses included runs of 1,200,000 generations with a sub-sampling frequency of 200 and burn-in of 120,000. Patristic distances were extracted from each phylogeny using PATRISTIC (Fourment and Gibbs 2006). To measure how recombination impacts branch length, we calculated the average patristic distance, PD, between the ancestral sequence (used to begin the simulation) and each individual sampled at the end of the simulation (60 individuals per iteration). Ten iterations were analyzed per condition. Accession Number. NCBI BioProject PRJNA276133.

ACKNOWLEDGMENTS

We thank R. Hernandez for advice on SDF_CODE, E. Angert, O. Lastovetsky, J. Russell, G. Turgeon and two anonymous reviewers for helpful comments. This work was supported by the National Science Foundation grant DEB-0918880 to TEP and CSBR-1349308 to JBM and the 60% project from University of Torino to PB.

REFERENCES

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Atyame, C. M., F. Delsuc, N. Pasteur, M. Weill, and O. Duron. 2011. Diversification of Wolbachia endosymbiont in the Culex pipiens mosquito. Mol. Biol. Evol. 28:2761-2772.

Bairoch, A., and R. Apweiler. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28:45-48.

Baldo, L., S. Bordenstein, J. J. Wernegreen, and J. H. Werren. 2006. Widespread recombination throughout Wolbachia genomes. Mol. Biol. Evol. 23:437-449.

Baumdicker, F., W. R. Hess, and P. Pfaffelhuber. 2012. The infinitely many genes model for the distributed genome of bacteria. Genome Biol. Evol. 4:443-456.

Bianciotto, V., E. Lumini, P. Bonfante, and P. Vandamme. 2003. 'Candidatus Glomeribacter gigasporarum' gen. nov., sp nov., an endosymbiont of arbuscular mycorrhizal fungi. Int. J. Syst. Evol. Microbiol. 53:121-124.

Bocs, S., S. Cruveiller, D. Vallenet, G. Nuel, and C. Médigue. 2003. AMIGene: Annotation of MIcrobial Genes. Nucleic Acids Res. 31:3723-3726.

Bouckaert, R., J. Heled, D. Kuhnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

Brownlie, J. C., M. Adamski, B. Slatko, and E. A. McGraw. 2007. Diversifying selection and host adaptation in two endosymbiont genomes. BMC Evol. Biol. 7:68.

Brownlie, J. C., B. N. Cass, M. Riegler, J. J. Witsenburg, I. Iturbe-Ormaetxe, E. A. McGraw, and S. L. O'Neill. 2009. Evidence for metabolic provisioning by a common invertebrate endosymbiont, Wolbachia pipientis, during periods of nutritional stress. PLoS Pathog. 5:e1000368.

Burke, G. R., H. J. McLaughlin, J. C. Simon, and N. A. Moran. 2010. Dynamics of a recurrent Buchnera mutation that affects thermal tolerance of pea aphid hosts. Genetics 186:367-372.

Castillo, D. M., and T. E. Pawlowska. 2010. Molecular evolution in bacterial endosymbionts of fungi. Mol. Biol. Evol. 27:622-636.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10:195-205.

Charlesworth, J., and A. Eyre-Walker. 2006. The rate of adaptive evolution in enteric bacteria. Mol. Biol. Evol. 23:1348–1356.

Chong, J. H., A. L. Roda, and C. M. Mannion. 2008. Life history of the mealybug, Maconellicoccus hirsutus (Hemiptera : Pseudococcidae), at constant temperatures. Environ. Entomol. 37:323-332.

Clark, M. A., N. A. Moran, and P. Baumann. 1999. Sequence evolution in bacterial endosymbionts having extreme base compositions. Mol. Biol. Evol. 16:1586-1598.

Degnan, P. H., and N. A. Moran. 2008. Evolutionary genetics of a defensive facultative symbiont of insects: exchange of toxin-encoding bacteriophage. Mol. Ecol. 17:916-929.

Didelot, X., and D. Falush. 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics 175:1251-1266.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PloS Biology 4:699-710.

Dufresne, A., M. Salanoubat, F. Partensky, F. Artiguenave, I. M. Axmann, V. Barbe, S. Duprat, M. Y. Galperin, E. V. Koonin, F. Le Gall, K. S. Makarova, M. Ostrowski, S. Oztas, C. Robert, I. B. Rogozin, D. J. Scanlan, N. T. de Marsac, J. Weissenbach, P. Wincker, Y. I. Wolf, and W. R. Hess. 2003. Genome sequence of the cyanobacterium Prochlorococcus marinus SS120, a nearly minimal oxyphototrophic genome. P. Natl. Acad. Sci. USA 100:10020-10025.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Ellegaard, K. M., L. Klasson, K. Näslund, K. Bourtzis, and S. G. E. Andersson. 2013. Comparative genomics of Wolbachia and the bacterial species concept. PLoS Genet. 9:e1003381.

Engelstädter, J., and G. D. D. Hurst. 2009. The ecology and evolution of microbes that manipulate host reproduction. Annu. Rev. Ecol. Evol. S. 40:127-149.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. Sonnhammer, J. Tate, and M. Punta. 2014. Pfam: the protein families database. Nucleic Acids Res. 42:D222-230.

Fourment, M., and M. J. Gibbs. 2006. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. BMC Evol. Biol. 6:1.

Fu, Y.-X., and W.-H. Li. 1993. Statistical tests of neutrality of mutations. Genetics 133:693-709. Funk, D. J., J. J. Wernegreen, and N. A. Moran. 2001. Intraspecific variation in symbiont genomes: Bottlenecks and the aphid-Buchnera association. Genetics 157:477-489.

Ghignone, S., A. Salvioli, I. Anca, E. Lumini, G. Ortu, L. Petiti, S. Cruveiller, V. Bianciotto, P. Piffanelli, L. Lanfranco, and P. Bonfante. 2012. The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. ISME J. 6:136-145.

Gilbert, J. A., K. Zhang, and J. D. Neufeld. 2010. Multiple displacement amplification. Pp. 4255-4263 in K. N. Timmis, ed. Handbook of Hydrocarbon and Lipid Microbiology. Springer-Verlag, Berlin Heidelberg.

Hedges, L. M., J. C. Brownlie, S. L. O'Neill, and K. N. Johnson. 2008. Wolbachia and virus protection in insects. Science 322:702-702.

Herbeck, J. T., D. J. Funk, P. H. Degnan, and J. J. Wernegreen. 2003. A conservative test of genetic drift in the endosymbiotic bacterium Buchnera: Slightly deleterious mutations in the chaperonin groEL. Genetics 165:1651-1660.

Herbeck, J. T., D. P. Wall, and J. J. Wernegreen. 2003. Gene expression level influences amino acid usage, but not codon usage, in the tsetse fly endosymbiont Wigglesworthia. Microbiology 149:2585-2596.

Hernandez, R. D. 2008. A flexible forward simulator for populations subject to selection and demography. Bioinformatics 24:2786-2787.

Jaenike, J. 1978. A hypothesis to account for the maintenance of sex within populations. J. Evol. Theory 3:191-194.

Jargeat, P., C. Cosseau, B. Ola'h, A. Jauneau, P. Bonfante, J. Batut, and G. Bécard. 2004. Isolation, freeliving capacities, and genome structure of "Candidatus Glomeribacter gigasporarum," the endocellular bacterium of the mycorrhizal fungus Gigaspora margarita. J. Bacteriol. 186:6876-6884.

Kasuga, T., T. J. White, and J. Taylor. 2002. Estimation of nucleotide substitution rates in eurotiomycete fungi. Mol. Biol. Evol. 19:2318-2324.

Koskiniemi, S., S. Sun, O. G. Berg, and D. I. Andersson. 2012. Selection-driven gene loss in bacteria. PLoS Genet. 8:e1002787.

Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. doi:10.1093/molbev/msw054.

Kuo, C.-H., and H. Ochman. 2009. Deletional bias across the three domains of life. Genome Biol. Evol. 1:145-152.

Lambert, J. D., and N. A. Moran. 1998. Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. P. Natl. Acad. Sci. USA 95:4458-4462.

Lamelas, A., M. J. Gosalbes, A. Manzano-Marin, J. Pereto, A. Moya, and A. Latorre. 2011a. Serratia symbiotica from the aphid Cinara cedri: A missing link from facultative to obligate insect endosymbiont. PLoS Genet. 7:e1002357.

Lamelas, A., M. J. Gosalbes, A. Moya, and A. Latorre. 2011b. 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. Microb. 77:4446-4454.

Lee, M.-C., and C. J. Marx. 2012. Repeated, selection-driven genome reduction of accessory genes in experimental populations. PLoS Genet. 8:e1002651.

Li, L., C. J. Stoeckert, and D. S. Roos. 2003. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178-2189.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Lumini, E., V. Bianciotto, P. Jargeat, M. Novero, A. Salvioli, A. Faccio, G. Bécard, and P. Bonfante. 2007. Presymbiotic growth and sporal morphology are affected in the arbuscular mycorrhizal fungus Gigaspora margarita cured of its endobacteria. Cell. Microbiol. 9:1716-1729.

McCutcheon, J. P., and N. A. Moran. 2012. Extreme genome reduction in symbiotic bacteria. Nat. Rev. Microbiol. 10:13-26.

Mira, A., and N. A. Moran. 2002. Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. Microb. Ecol. 44:137-143.

Mira, A., H. Ochman, and N. A. Moran. 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet. 17:589-596.

Mondo, S. J., K. H. Toomer, J. B. Morton, Y. Lekberg, and T. E. Pawlowska. 2012. Evolutionary stability in a 400-million-year-old heritable facultative mutualism. Evolution 66:2564-2576.

Moran, N. A. 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. P. Natl. Acad. Sci. USA 93:2873-2878.

Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. Annu. Rev. Genet. 42:165-190.

Moran, N. A., H. J. McLaughlin, and R. Sorek. 2009. The dynamics and time scale of ongoing genomic erosion in symbiotic bacteria. Science 323:379-382.

Moran, N. A., and A. Mira. 2001. The process of genome shrinkage in the obligate symbiont Buchnera aphidicola. Genome Biol. 2:Research0054.1-0054.12.

Morran, L. T., O. G. Schmidt, I. A. Gelarden, R. C. Parrish, and C. M. Lively. 2011. Running with the Red Queen: Host-parasite coevolution selects for biparental sex. Science 333:216-218.

Morris, J. J., R. E. Lenski, and E. R. Zinser. 2012. The Black Queen Hypothesis: Evolution of dependencies through adaptive gene loss. mBio 3:e00036-00012.

Morton, B. R. 1993. Chloroplast DNA codon use: Evidence for selection at the psb A locus based on tRNA availability. J. Mol. Evol. 37:273-280.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418-426.

Nielsen, R. 2005. Molecular signatures of natural selection. Annu Rev Genet 39:197-218.

Nilsson, A. I., S. Koskiniemi, S. Eriksson, E. Kugelberg, J. C. D. Hinton, and D. I. Andersson. 2005. Bacterial genome size reduction by experimental evolution. P. Natl. Acad. Sci. USA 102:12112-12116.

Ochman, H., S. Elwyn, and N. A. Moran. 1999. Calibrating bacterial evolution. P. Natl. Acad. Sci. USA 96:12638-12643.

Ohta, T. 1972. Population size and rate of evolution. J. Mol. Evol. 1:305-314.

Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. Nature 246:96-98.

Oliver, K. M., J. A. Russell, N. A. Moran, and M. S. Hunter. 2003. Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. P. Natl. Acad. Sci. USA 100:1803-1807.

Orsi, R. H., Q. Sun, and M. Wiedmann. 2008. Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of Listeria monocytogenes. BMC Evol. Biol. 8:233.

Osburne, M. S., B. M. Holmbeck, A. Coe, and S. W. Chisholm. 2011. The spontaneous mutation frequencies of Prochlorococcus strains are commensurate with those of other bacteria. Environ. Microbiol. Rep. 3:744-749.

Otto, S. P., and M. E. Orive. 1995. Evolutionary consequences of mutation and selection within an individual. Genetics 141:1173-1187.

Partida-Martinez, L. P., and C. Hertweck. 2005. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. Nature 437:884-888.

Ran, W., D. M. Kristensen, and E. V. Koonin. 2014. Coupling between protein level selection and codon usage optimization in the evolution of bacteria and archaea. mBio 5:e00956-00914.

Rispe, C., and N. A. Moran. 2000. Accumulation of deleterious mutations in endosymbionts: Muller's ratchet with two levels of selection. Am. Nat. 156:425-441.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A.

Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61:539-542.

Rousset, F., D. Bouchon, B. Pintureau, P. Juchault, and M. Solignac. 1992. Wolbachia endosymbionts responsible for various alterations of sexuality in arthropods. P. Roy. Soc. B-Biol. Sci. 250:91-98.

Roze, D., and R. E. Michod. 2001. Mutation, multilevel selection, and the evolution of propagule size during the origin of multicellularity. Am. Nat. 158:638-654.

Sachs, J. L., R. G. Skophammer, and J. U. Regus. 2011. Evolutionary transitions in bacterial symbiosis. P. Natl. Acad. Sci. USA 108:10800-10807.

Salvioli, A., S. Ghignone, M. Novero, L. Navazio, F. Venice, P. Bagnaresi, and P. Bonfante. 2016. Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its bioenergetic potential. ISME J. 10:130-144.

Sato, Y., K. Narisawa, K. Tsuruta, M. Umezu, T. Nishizawa, K. Tanaka, K. Yamaguchi, M. Komatsuzaki, and H. Ohta. 2010. Detection of betaproteobacteria inside the mycelium of the fungus Mortierella elongata. Microbes Environ. 25:321-324.

Scarborough, C. L., J. Ferrari, and H. C. J. Godfray. 2005. Aphid protected from pathogen by endosymbiont. Science 310:1781-1781.

Schierup, M. H., and J. Hein. 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156:879-891.

Sharp, P. M., L. R. Emery, and K. Zeng. 2010. Forces that influence the evolution of codon bias. Philos. T. Roy. Soc. B 365:1203-1212.

Smith, S. E., and D. J. Read. 2008. Mycorrhizal Symbiosis. Academic Press, New York.

Sokal, R. R., and F. J. Rohlf. 1998. Biometry. W. H. Freeman and Company, New York.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312-1313.

Stoltzfus, A., and R. W. Norris. 2015. On the causes of evolutionary transition:transversion bias. Mol. Biol. Evol. 10.1093/molbev/msv274.

Sun, Q., R. Bukowski, C. Myers, L. Ponnala, S. Stefanov, M. Howard, A. Clark, and J. Pillardy. 2010. Computational Biology Service Unit: Cornell University Core Facility for Computational Biology. J. Biomol. Tech. 21:S68.

Suyama, M., D. Torrents, and P. Bork. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609-W612.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Tavaré, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lect. Math. Life Sci. 17:57-86.

Teixeira, L., A. Ferreira, and M. Ashburner. 2008. The bacterial symbiont Wolbachia induces resistance to RNA viral infections in Drosophila melanogaster. PLoS Biology 6:2753-2763.

Tzeng, Y. H., R. Pan, and W. H. Li. 2004. Comparison of three methods for estimating rates of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 21:2290-2298.

Wernegreen, J. J., and N. A. Moran. 1999. Evidence for genetic drift in endosymbionts (Buchnera): analyses of protein-coding genes. Mol Biol Evol 16:83-97.

Werren, J. H., L. Baldo, and M. E. Clark. 2008. Wolbachia: master manipulators of invertebrate biology. Nat. Rev. Microbiol. 6:741-751.

Werren, J. H., W. Zhang, and L. R. Guo. 1995. Evolution and phylogeny of Wolbachia: reproductive parasites of arthropods. P. Roy. Soc. Lond. B Bio. 261:55-63.

Wielgoss, S., J. E. Barrick, O. Tenaillon, S. Cruveiller, B. Chane-Woon-Ming, C. Médigue, R. E. Lenski, and D. Schneider. 2011. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with Escherichia coli. G3 1:183-186.

Winsor, G. L., B. Khaira, T. Van Rossum, R. Lo, M. D. Whiteside, and F. S. Brinkman. 2008. The Burkholderia Genome Database: facilitating flexible queries and comparative analyses. Bioinformatics 24:2803-2804.

Wright, F. 1990. The 'effective number of codons' used in a gene. Gene 87:23-29.

Yang, Z. 2007a. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24:1586-1591.

Yang, Z. 2007b. Computational Molecular Evolution. Oxford University Press, New York.

Yang, Z. H., and R. Nielsen. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17:32-43.

Zamenhof, S., and H. H. Eichhorn. 1967. Study of microbial evolution through loss of biosynthetic functions: Establishment of “defective” mutants. Nature 216:456-458.

Zerbino, D. R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821-829.

Zhang, K., A. C. Martiny, N. B. Reppas, K. W. Barry, J. Malek, S. W. Chisholm, and G. M. Church. 2006. Sequencing genomes from single cells by polymerase cloning. Nat. Biotechnol. 24:680.

Table 1. Diversity levels, recombination rates, and neutrality tests in the *Glomeribacter* population.

	Individual marker data[†]		**Whole-genome data**[‡]
	ftsZ	*pstA*	
π	0.0919	0.0154	0.2185
π_N	0.0106	0.1775	0.1064
π_S	0.3789	0.0833	0.7283
Θ_W	0.0685	0.1477	0.2044
Tajima's *D* (total mutations)	1.2194 ($P > 0.1$)	-0.8434 ($P > 0.1$)	NA
Tajima's *D* (segregating sites)	1.4308 ($P > 0.1$)	0.1924 ($P > 0.1$)	NA
Fu and Li's *D** (total mutations)	0.8303 ($P > 0.1$)	-1.4254 ($P > 0.1$)	NA
Fu and Li's *D** (segregating sites)	0.8000 ($P > 0.1$)	-1.8600 ($P > 0.1$)	NA
Fu and Li's *F** (total mutations)	1.1026 ($P > 0.1$)	-1.4564 ($P > 0.1$)	NA
Fu and Li's *F** (segregating sites)	1.1378 ($P > 0.1$)	-1.4725 ($P > 0.1$)	NA
ρ/θ (95% CI)	0.1 (0.01 – 0.39)[§]		0.01 (0.001 – 0.0104)
r/*m* (95% CI)	2.5 (0.44 – 6.30)[§]		0.48 (0.41 – 0.83)

π, nucleotide diversity calculated using all sites; π_N, nonsynonymous site diversity; π_S, synonymous site diversity; Θ_W, Watterson estimator calculated using all sites; ρ/θ, the rate of recombination relative to the rate of mutation; *r*/*m*, the ratio of rates at which nucleotides become substituted as a result of recombination versus mutation; CI, confidence interval.

[†]based on sequences of *ftsZ* and *pstA* sampled from *Glomeribacter* strains associated with 19 AMF populations (Fig. 1).

[‡]based on 402 gene orthologs sampled from the *Glomeribacter* genomes BEG1, BEG34, IN211, and JA201A-16.

[§]calculations from Mondo et al. (2012) based on concatenated *ftsZ* and *pstA* gene sequences.

Table 2. Microbial genomes, lineages, and lifestyles included in the study.

Organism	GenBank accession number	Lineage	Lifestyle
Ca. Blochmannia floridanus	NC_005061.1	*Blochmannia*	Essential heritable mutualist
Ca. Blochmannia pennsylvanicus	NC_007292.1	*Blochmannia*	Essential heritable mutualist
Ca. Blochmannia vafer	NC_014909.2	*Blochmannia*	Essential heritable mutualist
Buchnera aphidicola (*Cinara tujafilina*)	NC_015662.1	*Buchnera*	Essential heritable mutualist
Buchnera aphidicola LSR1 (*Acyrthosiphon pisum*)	NZ_ACFK01000001.1	*Buchnera*	Essential heritable mutualist
Buchnera aphidicola Bp (*Baizongia pistaciae*)	NC_004545.1	*Buchnera*	Essential heritable mutualist
Ca. Sulcia muelleri CARI	NC_014499.1	*Sulcia*	Essential heritable mutualist
Ca. Sulcia muelleri DMIN	NC_014004.1	*Sulcia*	Essential heritable mutualist
Ca. Sulcia muelleri SMDSEM	NC_013123.1	*Sulcia*	Essential heritable mutualist
Ca. Glomeribacter gigasporarum BEG1	PRJNA276133	*Glomeribacter*	Nonessential heritable mutualist
Ca. Glomeribacter gigasporarum BEG34	NZ_CAFB01000000	*Glomeribacter*	Nonessential heritable mutualist
Ca. Glomeribacter gigasporarum IN211	PRJNA276133	*Glomeribacter*	Nonessential heritable mutualist
Ca. Glomeribacter gigasporarum JA201A-16	PRJNA276133	*Glomeribacter*	Nonessential heritable mutualist
Ca. Hamiltonella defensa 5AT	NC_012751.1	*Hamiltonella/Regiella*	Nonessential heritable mutualist
Ca. Hamiltonella defensa MED	GCA_000258345.1	*Hamiltonella/Regiella*	Nonessential heritable mutualist
Ca. Regiella insecticola LSR1	GCA_000143625.1	*Hamiltonella/Regiella*	Nonessential heritable mutualist
Wolbachia pipentis wMel (*Drosophila melanogaster*)	NC_002978.6	*Wolbachia*	Nonessential heritable endosymbiont
Wolbachia wPip (*Culex quinquefasciatus*)	NC_010981.1	*Wolbachia*	Nonessential heritable endosymbiont
Wolbachia wSim (*Drosophila simulans*)	GCA_000167495.1	*Wolbachia*	Nonessential heritable endosymbiont
Burkholderia ambifaria MC40-6	GCA_000019925.1	*Burkholderia*	Free-living soil microbe
Burkholderia multivorans ATCC 17616	GCA_000010545.1	*Burkholderia*	Free-living soil microbe
Burkholderia vietnamiensis G4	GCA_000016205.1	*Burkholderia*	Free-living soil microbe
Prochlorococcus marinus MIT 9215	NC_009840.1	*Prochlorococcus*	Free-living oceanic microbe
Prochlorococcus marinus MIT 9301	NC_009091.1	*Prochlorococcus*	Free-living oceanic microbe
Prochlorococcus marinus subsp. *pastoris* CCMP1986	NC_005072.1	*Prochlorococcus*	Free-living oceanic microbe
Bradyrhizobium diazoefficiens USDA 110	NC_004463.1	*Bradyrhizobium*	Free-living root symbiont
Bradyrhizobium sp. BTAi1	NC_009485.1	*Bradyrhizobium*	Free-living root symbiont
Bradyrhizobium sp. ORS278	NC_009445.1	*Bradyrhizobium*	Free-living root symbiont
Enterobacter cloacae subsp. *cloacae* NCTC 9394	NC_021046.1	*Enterobacter/Escherichia/ Salmonella* (EES)	Free-living enteric bacterium
Escherichia coli ATCC 8739	NC_010468.1	*Enterobacter/Escherichia/ Salmonella* (EES)	Free-living enteric bacterium
Salmonella enterica subsp. *enteric*	NC_006905.1	*Enterobacter/Escherichia/ Salmonella* (EES)	Free-living enteric bacterium
Bifidobacterium animalis subsp. *lactis* AD011	NC_011835.1	*Bifidobacterium*	Free-living lactic acid bacterium
Bifidobacterium catenulatum DSM 16992	GCA_000173455.1	*Bifidobacterium*	Free-living lactic acid bacterium
Bifidobacterium longum subsp. *longum* JCM 1217	NC_015067.1	*Bifidobacterium*	Free-living lactic acid bacterium
Lactobacillus acidophilus NCFM	NC_006814.3	*Lactobacillus*	Free-living lactic acid bacterium
Lactobacillus farciminis KCTC 3681	GCA_000184535.1	*Lactobacillus*	Free-living lactic acid bacterium
Lactobacillus johnsonii NCC 533	NC_005362.1	*Lactobacillus*	Free-living lactic acid bacterium

Table 3. Genome-wide estimates of the nucleotide transition/transvertion rate bias, κ, the codon bias index, CBI, the effective number of codons, ENC, as well as dN, dS, and dN/dS according to the NG86 (Nei and Gojobori 1986), LWL85m (Tzeng, Pan, and Li 2004), and YN00 (Yang and Nielsen 2000) methods in microbes with different lifestyles.

Lifestyle	*Lineage*	κ	*CBI*	*ENC*	NG86				LWL85m				YN00			
					n	d_N	d_S	$d_N/d_S \pm 95\%$ CI	*n*	d_N	d_S	$d_N/d_S \pm 95\%$ CI	*n*	d_N	d_S	$d_N/d_S \pm 95\%$ CI
Essential heritable mutualists																
	Blochmannia	3.2	0.76	32	534	0.212	0.886	0.248 ± 0.0088	512	0.222	0.896	0.261 ± 0.0096	118	0.208	1.308	0.164 ± 0.0134
	Buchnera	1.9	0.65	37	294	0.327	0.986	0.346 ± 0.0166	47	0.302	1.191	0.262 ± 0.0423	NA	NA	NA	NA
	Sulcia	1.6	0.60	39	162	0.132	0.309	0.439 ± 0.0270	158	0.135	0.445	0.364 ± 0.0285	157	0.112	0.915	0.132 ± 0.0095
Nonessential heritable mutualists																
	*Glomeribacter**	5.2	0.33	49	631	0.087	0.603	0.149 ± 0.0066	615	0.095	0.514	0.190 ± 0.0082	633	0.098	0.517	0.198 ± 0.0089
	Hamiltonella/ *Regiella*	2.6	0.34	51	211	0.210	1.146	0.187 ± 0.0143	302	0.227	1.203	0.195 ± 0.0097	257	0.239	1.161	0.207 ± 0.0145
Nonessential heritable endosymbionts																
	Wolbachia	10.1	0.37	47	481	0.080	0.477	0.169 ± 0.0111	329	0.086	0.348	0.256 ± 0.0123	475	0.079	0.521	0.166 ± 0.0107
Free-living microbes																
	Burkholderia	3.6	0.68	32	3915	0.056	0.327	0.158 ± 0.0031	3927	0.055	0.350	0.144 ± 0.0027	3876	0.044	0.730	0.062 ± 0.0017
	Prochlorococcus	2.6	0.53	40	1492	0.096	0.773	0.127 ± 00043	1351	0.095	0.727	0.143 ± 0.0048	1409	0.041	0.531	0.084 ± 0.0034
Root symbionts																
	Bradyrhizobium	2.2	0.62	35	3871	0.128	0.587	0.202 ± 0.1006	3799	0.120	0.681	0.162 ± 0.0024	3128	0.062	1.092	0.055 ± 0.0017
Enteric bacteria																
	EES	2.3	0.38	45	2151	0.082	1.065	0.074 ± 0.0048	1375	0.081	1.012	0.078 ± 0.0033	2079	0.074	1.088	0.068 ± 0.0022
Lactic acid bacteria																
	Bifidobacterium	1.7	0.49	40	888	0.212	1.053	0.199 ± 0.007	790	0.187	1.061	0.173 ± 0.0056	246	0.129	1.157	0.110 ± 0.0011
	Lactobacillus	1.6	0.48	42	521	0.283	1.201	0.233 ± 0017	288	0.218	1.190	0.181 ± 0.0119	98	0.143	1.106	0.127 ± 0.0187

**Glomeribacter* BEG1, BEG34, and JA201A-16; d_N, number of nonsynonymous substitutions per nonsynonymous site; d_S, number of synonymous substitutions per synonymous site; *n*, number of orthologs; CI, confidence interval; NA, not available due to d_S saturation.

Fig. 1. Genealogy of Ca. Glomeribacter gigasporarum with the combined posterior distribution of the divergence times estimated under the uncorrelated lognormal relaxed clock model, reconstructed using the 16S rRNA, 23S rRNA, ftsZ and pstA genes. Bayesian posterior probabilities are shown above branches; the gray bars represent the mean 95% highest posterior densities; the node with normally distributed calibration prior is marked by an asterisk; Glomeribacter genomes sequenced in the present study are marked by single arrowheads; a double arrowhead marks the previously published Glomeribacter genome (Ghignone et al. 2012).

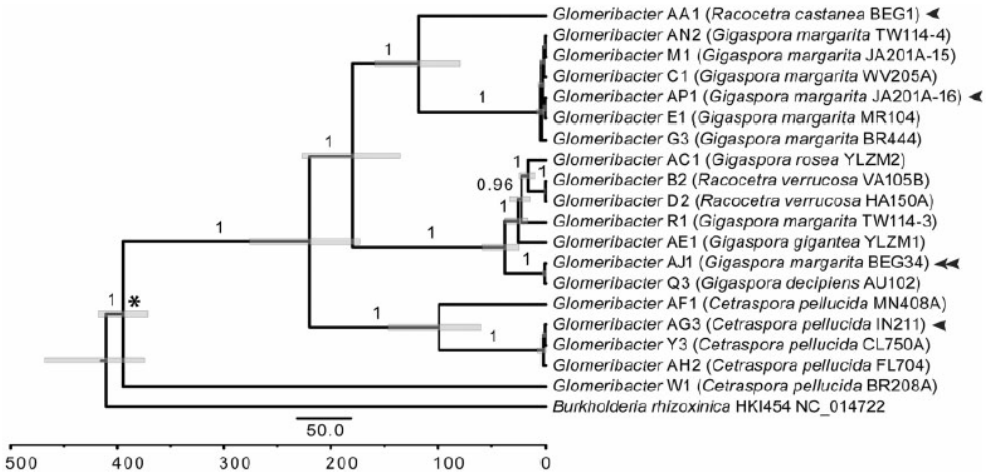

Fig. 2. Evolutionary history of Ca. Glomeribacter gigasporarum reconstructed using nucleotide sequences at 16S rRNA, 23S rRNA and 25 protein-coding loci. Bayesian posterior probabilities are shown above branches. Branches with maximum likelihood bootstrap support over 70% are thickened

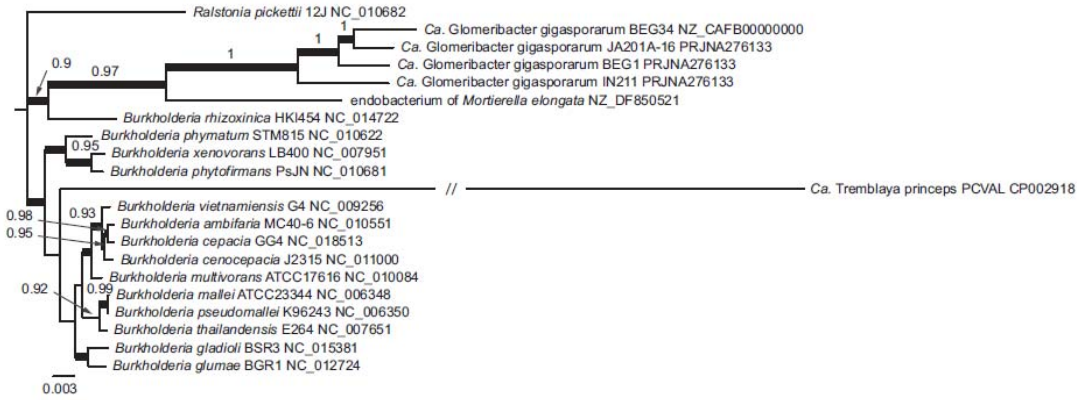

Fig. 3. Distribution of dN/dS values estimated for individual genes using the NG86 method (Nei and Gojobori 1986) across the genomes of microbes representing lineages with different lifestyles. Essential heritable mutualists Buchnera (A), Sulcia (B), Blochmannia (C); nonessential endosymbionts Glomeribacter (D), Hamiltonella/Regiella (E), and Wolbachia (F); free-living Burkholderia (G) and Prochlorococcus (H); Bradyrhizobium root symbionts (I); enteric bacteria (J); lactic acid bacteria Bifidobacterium (K) and Lactobacillus (L).

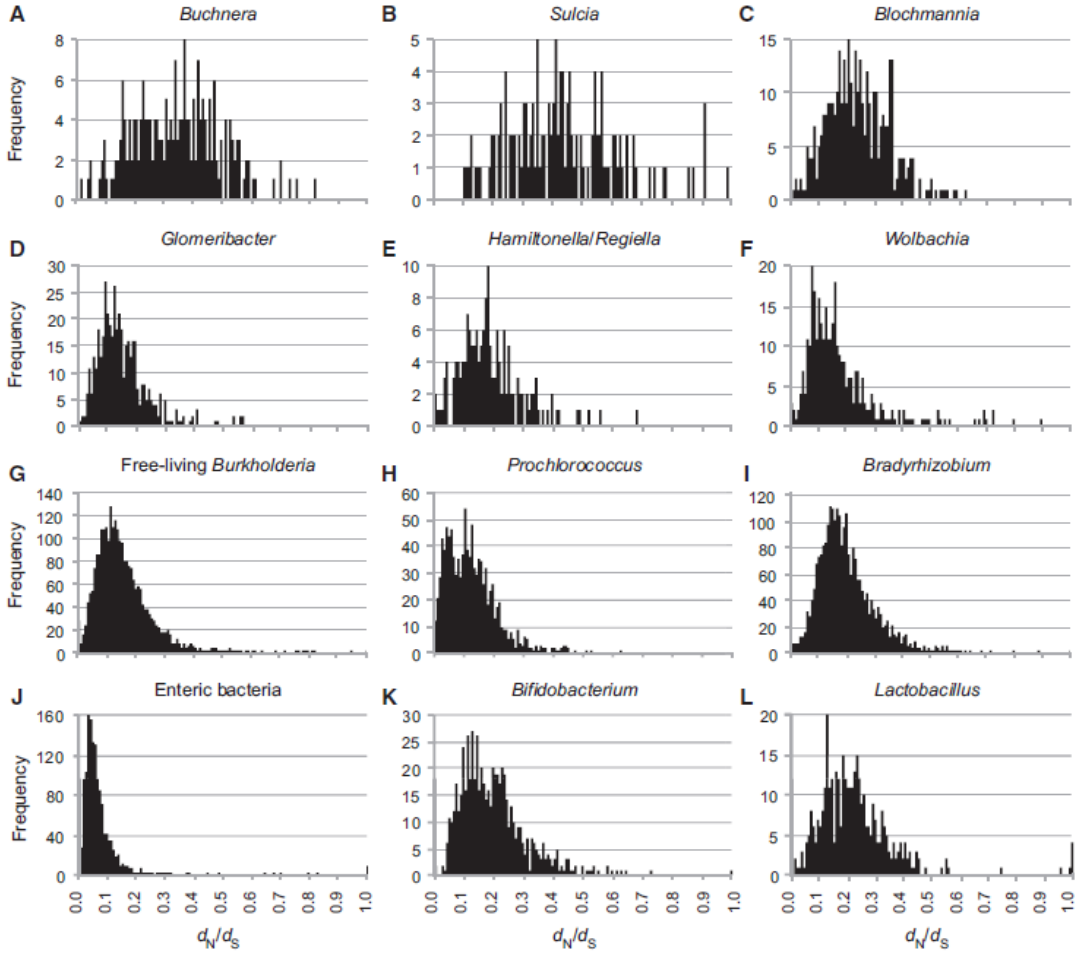

Fig. 4. QQ plots comparing NG86 dN/dS genome-wide distributions in non-essential endosymbionts to those of free-living bacteria and essential heritable mutualists. (A) Glomeribacter vs. free-living Burkholderia, (B) Glomeribacter vs. Buchnera, (C) Hamiltonella/Regiella vs. free-living Burkholderia, (D) Hamiltonella/Regiella vs. Buchnera, (E) Wolbachia vs. free-living Burkholderia, and (F) Wolbachia vs. Buchnera.

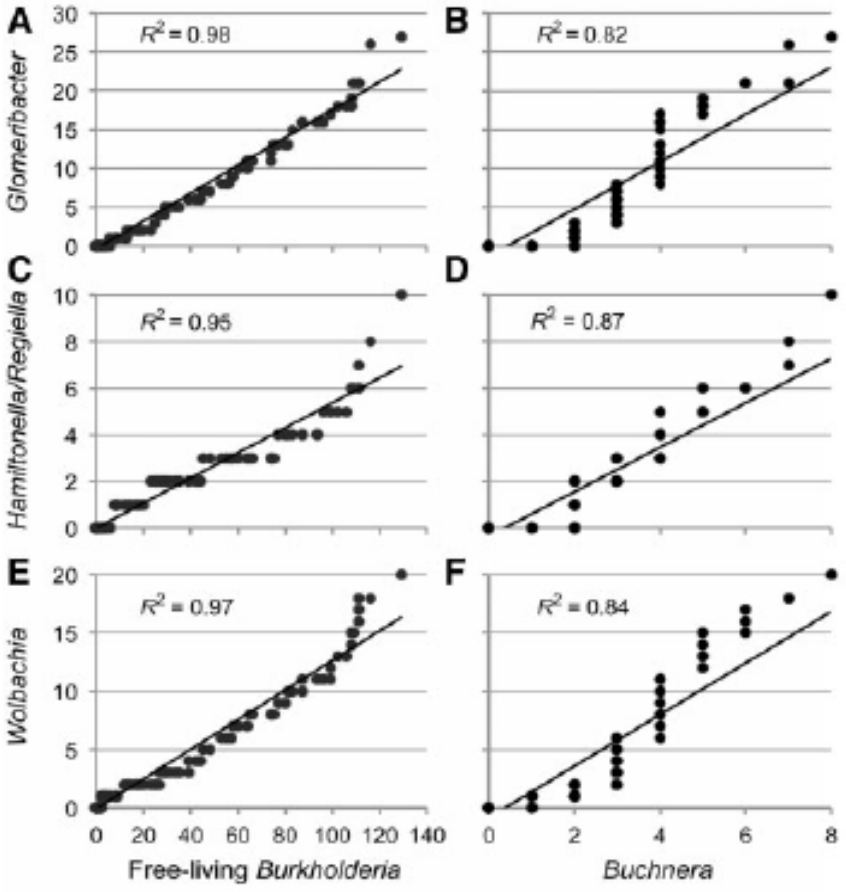

Fig. 5. Simulated evolutionary histories of populations with different mutation and recombination parameters. (A) A highly recombinant population, θ = 0.1791 as in Glomeribacter and ρ = 10. (B) A clonal population with θ = 0.1791 and ρ = 0. (C) A population with limited recombination, θ = 0.1791 and ρ = 0.001791 as in Glomeribacter. (D) A population with a high nucleotide substitution rate and parameters as in Tremblaya princeps, a relative of Glomeribacter and essential heritable mutualist of mealybugs, θ = 0.4119, ρ = 0. Branches with Bayesian posterior probabilities of 0.95 or higher are thickened.