The ISME Journal (2014) 8, 1237–1246

www.nature.com/ismej

ORIGINAL ARTICLE

Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont

Ryuichi Koga[1,2] and Nancy A Moran[1,3]
[1]*Microbial Diversity Institute, Yale University, West Haven, CT, USA;* [2]*Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan and* [3]*Department of Integrative Biology, University of Texas at Austin, Austin, TX, USA*

Bacterial symbionts that undergo long-term maternal transmission experience elevated fixation of deleterious mutations, resulting in massive loss of genes and changes in gene sequences that appear to limit efficiency of gene products. Potentially, this dwindling of symbiont functionality impacts hosts that depend on these bacteria for nutrition. One evolutionary escape route is the acquisition of a novel symbiont with a robust genome and metabolic capabilities. Such an acquisition has occurred in an ancestor of *Philaenus spumarius*, the meadow spittlebug (Insecta: Cercopoidea), which has replaced its ancient association with the tiny genome symbiont *Zinderia insecticola* (Betaproteobacteria) with an association with a symbiont related to *Sodalis glossinidius* (Gammaproteobacteria). Spittlebugs feed exclusively on xylem sap, a diet that is low both in essential amino acids and in sugar or other substrates for energy production. The new symbiont genome has undergone proliferation of mobile elements resulting in many gene inactivations; nonetheless, it has selectively maintained genes replacing functions of its predecessor for amino-acid biosynthesis. Whereas ancient symbiont partners typically retain perfectly complementary sets of amino-acid biosynthetic pathways, the novel symbiont introduces some redundancy as it retains some pathways also present in the partner symbionts (*Sulcia muelleri*). Strikingly, the newly acquired *Sodalis*-like symbiont retains genes underlying efficient routes of energy production, including a complete TCA cycle, potentially relaxing the severe energy limitations of the xylem-feeding hosts. Although evolutionary replacements of ancient symbionts are infrequent, they potentially enable evolutionary and ecological novelty by conferring novel metabolic capabilities to host lineages.
The ISME Journal (2014) **8,** 1237–1246; doi:10.1038/ismej.2013.235; published online 9 January 2014
Subject Category: Evolutionary genetics
Keywords: Auchenorrhyncha; Cercopoidea; expansion of metabolic capability; genome reduction; symbiont replacement; xylem-feeding

Introduction

Symbioses with bacteria have been a key driver of eukaryotic evolution, enabling expansion into new ecological niches and species diversification. Among the most prominent evolutionary radiations dependent on symbiosis are sap-feeding insects, such as aphids, psyllids, cicadas and spittlebugs. All have intimate associations with maternally transmitted bacterial symbionts that provision essential nutrients and thereby enable dietary specialization on phloem or xylem sap of vascular plants (Baumann, 2005; Moran, 2007). Although such cases illustrate a role of symbiosis in ecological innovation, reliance on symbionts can eventually become an evolutionary liability. The small population sizes and asexuality of maternally transmitted symbionts result in the accumulation of mildly deleterious mutations, each of small effect but cumulatively resulting in lowered functionality and fitness (Moran, 1996). Among the most extreme examples of deteriorating genomes are those found in insect symbionts (Moran, 2007; McCutcheon and Moran, 2010, 2012; Bennett and Moran, 2013). Up to 90% of ancestral genes and associated metabolic capabilities may be lost, and retained genes show extensive sequence changes that reflect fixation of mutations deleterious to gene function. For example, symbiont gene products exhibit thermal instability (Lambert and Moran, 1998; van Ham *et al.*, 2003; Baumann, 2005; Moran, 2007), and the resulting heat sensitivity of symbionts can affect hosts, in which reproduction fails when symbionts are eliminated by heat (Moran, 1996; Dunbar *et al.*, 2007). Furthermore, obligate symbiont lineages

Correspondence: NA Moran, Department of Integrative Biology, University of Texas at Austin, 2506 Speedway, A5000, Austin, TX 78712-1211, USA.
E-mail: nancy.moran@austin.utexas.edu
Received 24 September 2013; revised 27 November 2013; accepted 30 November 2013; published online 9 January 2014

cannot regain lost functions as they do not incorporate foreign genes through the lateral gene transfer processes that are frequent in most bacteria (Tamas *et al.*, 2002; Moran, 2007; Moran *et al.*, 2008; McCutcheon and Moran, 2010, 2012; Bennett and Moran, 2013).

Hosts such as sap-feeding insects have evolved obligate dependence on symbionts despite dwindling symbiont capabilities and tolerances. Replacing an ancestral symbiont with a new one is a potential solution: symbiont services, such as essential amino-acid biosynthesis, are ubiquitous capabilities in bacteria. But, as a result of host–symbiont coadaptation, symbiont cells are intimately involved in host development (for example, Braendle *et al.*, 2003; Koga *et al.*, 2012) such that their sudden removal causes lethal developmental disruption. Nonetheless, several cases are known in which an insect host lineage has acquired a new symbiont, which has entirely replaced an ancestral one (for example, Lefèvre *et al.*, 2004; Conord *et al.*, 2008; Toenshoff *et al.*, 2012; Toju *et al.*, 2013) or has persisted along with the ancestral symbiont while assuming a subset of its functions (for example, Lamelas *et al.*, 2011).

Most sap-feeding insects of suborder Auchenorrhyncha host two symbionts, *Sulcia muelleri* (Bacteroidetes) and a partner that varies among host groups (Moran, 2007; McCutcheon and Moran, 2010; Bennett and Moran, 2013). In each case, the paired symbionts show strikingly complementary sets of genes related to production of essential amino acids (the 10 protein amino acids that are required nutrients for animals). For example, most spittlebugs (Cercopoidea) are host to *Sulcia* plus *Zinderia insecticola* (Betaproteobacteria) (Koga *et al.*, 2013), and genome sequences of symbionts of the spittlebug *Clastoptera arizonana* indicate that *Zinderia* can produce tryptophan, methionine and histidine, exactly complementing *Sulcia*'s ability to produce the other seven essential amino acids required by the hosts, which consume only xylem sap (McCutcheon and Moran, 2010). However, in the meadow spittlebug (*Philaenus spumarius*) and closely related species (tribe Philaenini), *Zinderia* has been replaced by a *Sodalis*-like symbiont (hereafter SLs, Gammaproteobacteria) (Koga *et al.*, 2013). Because both *Sulcia* and SLs are present in all populations of *P. spumarius* sampled, and because each occupies a specialized bacteriocyte type, they are inferred to be obligate symbionts for their host, similar to the dual symbionts in other groups of Auchenorrhyncha that show complementary sets of biosynthetic pathways for nutrient provisioning to hosts (McCutcheon and Moran, 2007, 2010; McCutcheon *et al.,* 2009). *P. spumarius* ingests the main transpiration xylem stream from a variety of plant species and prefers plants with high amino-acid content in the xylem fluid (Wiegert, 1964; Horsfield, 1977; Hoffman and McEvoy, 1985; Malone *et al.*, 1999). This diet lacks not only essential amino acids but also sugar or other abundant energy substrates; nonessential amino acids, primarily glutamine and asparagine, contain the primary energy content (Brodbeck *et al.*, 1990). Energy demands are heightened further, because xylem is under strong negative pressure in plant shoots (Raven *et al.*, 2005), and *P. spumarius* must use powerful muscles to extract its food.

A symbiont replacement event, such as that in *P. spumarius*, presents the possibility of expanded host functionality, as free-living bacteria possess many more metabolic capabilities than anciently associated obligate symbionts. Alternatively, the new symbiont may quickly shed genes conferring any capabilities beyond those of its predecessor, resulting in re-establishment of the same division of labor between symbiont types, though possibly with improved efficiency. The case of the SLs in the Philaenini presents a relatively recent case of symbiont replacement, providing a snapshot of how a bacterial genome changes following establishment of symbiosis, how tasks of different symbionts are sorted out during the evolution of the new partnership and whether metabolic capabilities may be expanded in the process. To address these questions, we have sequenced and analyzed the genomes of the paired *Sulcia* and SLs symbionts of *P. spumarius* and compared with those of paired *Sulcia* and *Zinderia*, the ancestral symbiont types in spittlebugs.

Materials and methods

Detailed procedures and DNA oligomer sequences are available as supporting online materials.

Insect materials

Individuals of the *P. spumarius* were collected at the Yale University West Campus (West Haven, CT, USA), Berkeley (CA, USA) and Dublin (Ireland) for this study. DNAs of seven spittlebug species, *Cosmoscarta heros*, *Aphrophora quadrinotata*, *Philaenarcys bilineata*, *Neophilaenus lineatus*, *Mesoptyelus fasciatus*, *Philaenus maghresignus* and *Philaenus tesselatus*, were the same as those used in a previous study (Koga *et al.*, 2013).

Genome sequencing and data analyses

DNA was extracted from three individuals collected in West Haven, CT, USA using Proteinase K treatment and phenol extraction. A 300-bp shotgun library was constructed with the extracted DNA and was subjected to high-throughput sequencing from both ends (2×100 bp) using Genome Analyzer IIx (Illumina, San Diego, CA, USA) at the Yale Center for Genome Analysis. The obtained paired reads were assembled and analyzed using the CLC Genomics Workbench 5.5 (CLC Bio, Aarhus, Denmark). Genes were annotated using MiGAP

(Sugawara *et al.*, 2009), KAAS (Moriya *et al.*, 2007) and MetaCyc (Caspi *et al.*, 2008). Insertion sequence (IS) finder (Siguier *et al.*, 2006) was used to identify ILs. Circular graphs of the genomic data were drawn using CIRCOS ver. 0.61 (Krzywinski *et al.*, 2009).

Cloning and sequencing of IS1 in SLs
The entire sequence of IS1 in SLs (IS1SLs) was obtained by PCR cloning with the primers IS1N5, IS1N3, IS1N-OF and IS1N-OR and Sanger sequencing essentially as described previously (Koga *et al.*, 2013).

Copy number of IS1SLs
DNAs for quantitative PCR assay were extracted by using the DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA, USA). Using DNA samples from single spittlebugs from three localities, the numbers of IS1SLs and of the single copy gene, *groEL*, were estimated using a Mastercycler ep realplex (Eppendorf, Hamburg, Germany) with SYBR Green chemistry (Moriyama *et al.*, 2012). Quantifications of IS1SLs copy numbers were conducted with the primers IS1N-QF and IS1N-QR while those of *groEL* were based on the primers PsSodGroEL-AF and PsSodGroEL-AR. Reaction mixtures were described elsewhere (Moriyama *et al.*, 2012), and thermo-cycling conditions included an initial denaturation step for 95 °C for 2 min; 32 × cycle regime of 95 °C for 30 s, 52 °C for 30 s and 72 °C for 30 s and a melting curve analysis. The copy number of IS1SLs per genome was estimated by dividing the IS1SLs copy number by the *groEL* copy number in each individual assayed.

Occurrence of IS1 copies in SLs of other spittlebug species
DNAs of eight spittlebug species were subjected to PCR with three combinations of primers, IS1N5 and IS1N3, IS1N-AF and IS1N-AR and IS1N-OF and IS1N-OR, amplifying different regions of IS1Sln. The PCR conditions were same as that for cloning of IS1SLs.

Expression of distinct 16S rRNA genes in the SLs genome
To determine which copies of SLs 16S rRNA genes were silenced and which were expressed, we used reverse transcription restriction fragment length polymorphism analyses of 16S rRNA expressed in SLs. Insect specimens were divided into two parts along the middle line. One half was subjected to RNA extraction with the combination of TRIzol (Life Technologies, Carlsbad, CA, USA) and RNeasy kit (QIAGEN), while DNA was extracted from the other part with the DNeasy Blood and Tissue Kit (QIAGEN). Complementary DNAs (cDNAs) were synthesized with Superscript II (Life Technologies) according to the manufacturer's protocol. The cDNAs and corresponding DNAs were subjected to PCR with the primers Sod628F and Sod1195R in the reaction mixture same as that for the cloning of IS1SLs with a thermo-cycling conditions, including an initial denaturation step for 95 °C for 2 min; 32 × cycle regime of 95 °C for 15 s, 50 °C for 30 s and 68 °C for 1 min. The PCR products were digested with either SacII or NcoI (New England Biolabs, Ipswich, MA, USA) according to the manufacturer's protocol.

cDNA cloning of 16S rRNA expressed in SLs
Total RNA of *P. spumarius* samples, including their bacteriomes and symbionts, was converted to cDNA using Superscript II and the primers 16SB1 and 1507R according to the manufacturer's protocol. The cDNA was converted to dsDNA form and adenylated at its 3′ ends using the mRNA-Seq Sample Preparation Kit (Illumina) as directed in the manufacturer's protocol. The synthesized double-stranded cDNA was cloned and sequenced in the same way as that for the cloning of IS1SLs.

Molecular phylogenetic analyses
Multiple alignments were assembled using MUSCLE (Edgar, 2004) implemented in MEGA 5.2 (Tamura *et al.*, 2011). Ambiguously aligned regions and sites, including missing data, were removed from the multiple alignments before phylogenetic tree constructions. For the phylogeny of SLs (Supplementary Figure S5C), multiple alignments of six protein coding genes, *alaS*, *ffh*, *pyrG*, *typA*, *uvrB* and *uvrC*, were concatenated and subjected to a phylogenetic analysis (Engel *et al.*, 2012). All phylogenetic trees were inferred with maximum likelihood method using RAxML v7.2.8 (Stamatakis, 2006).

Data deposition
The scaffolds for SLs of *P. spumarius* have been deposited at the GenBank/DDBJ/EMB under the accession nos. BASS01000001–BASS01000562 (562 entries), while the sequences of the complete genome of *S. muelleri* and the full-length IS1 are deposited as AP013293 and AB849119, respectively.

Results and discussion

Sulcia-PSPU genome features
For *Sulcia* of *P. spumarius* (*Sulcia*-PSPU), we obtained a complete circular genome, based on an average sequencing depth of ~2600 ×, resulting in a genome sequence of 285 kilobases (kb) (Table 1). The *Sulcia*-PSPU genome is highly similar to the four previously sequenced *Sulcia* genomes from different sap-feeding insects. All have identical chromosomal arrangement of shared genes (Supplementary Figure S1), highly reduced gene sets, AT-biased base

composition and single rRNA operons. Some differences in gene sets occur among *Sulcia* genomes, attributable to different gene loss events during the diversification of *Sulcia* lineages (Supplementary Figure S1). *Sulcia*-PSPU has no unique gene losses and has only three genes (*gdhA*, *nusG* and *map*) not found in any sequenced *Sulcia* genomes. As expected, it is most similar to that of the *Sulcia* previously sequenced for the spittlebug *C. arizonana* (*Sulcia*-CARI; Figure 1a).

Table 1 Features of the genomes of SLs-PSPU and *Sulcia*-PSPU

Feature	*SLs-PSPU*	*Sulcia-PSPU*
Number of scaffolds	562 (607 contigs)	1 (completed)
Genome size (megabases)	1.38[a]	0.285
Average coverage	266 ×	2985 ×
G + C%	54.2%	20.9%
Number of intact ORFs	1027	252
Number of rRNA operons	⩾2[b]	1
Number of tRNAs	37	30
Insertion sequences[c]	IS1, IS5, IS10 and IS110	Not found
Phages and prophages[c]	Enterobacteria phage Mu, *Burkholderia* phage BcepMu, *Salmonella* phage ST64B	Not found

Abbreviations: IS, insertion sequence; ORF, open reading frame; SLs, *Sodalis*-like symbiont; *Sulcia*-PSPU; *Sulcia* of *P. spumarius*.
[a]Total length of combined scaffolds and an approximate estimate of the total length of unique sequence. Because of multicopy IS elements, the actual chromosome size is likely somewhat larger.
[b]No full-length rRNA genes were assembled. However, we identified two 16S rRNA loci of *Sodalis*-like symbiont from the species, indicating that this symbiont of the species harbors at least two rRNA loci.
[c]According to the classification of IS elements and phages found in *S. glossinidius* by Belda *et al.* (2010).

SLs-PSPU genome features

SLs-PSPU genes show high sequence homology to genes of *Sodalis glossinidius*, a symbiont from tsetse fly hosts (Figure 1b), with coding regions showing nucleotide identities of 90–96%. The genomic sequence of approximately 1.38 Mb (megabases) was fragmented into 562 scaffolds (Table 1), which were recognizable as belonging to the SLs-PSPU genome based on depth of coverage and close homology to previously sequenced *Sodalis* genomes (see Supplementary Methods for detailed criteria). These scaffolds represent nearly the complete gene repertoire of SLs-PSPU, based on the presence of 203 of a defined set of 205 single copy genes commonly present in gammaproteobacterial genomes (Lerat *et al.*, 2003). The fragmentation of the assembly primarily resulted from the presence of a large number of IS elements corresponding to nine IS types; four of these IS types have close sequence similarity to IS elements previously identified in the *S. glossinidius* genome, which also contains many more IS elements than do most bacterial genomes (Belda *et al.*, 2010) (Figure 2). One abundant type, in the IS1 family, is most similar to copies in *S. glossinidius* and also related to elements in *Escherichia coli* strains (Supplementary Figure S2A). This element (IS1SLs) is present at many locations

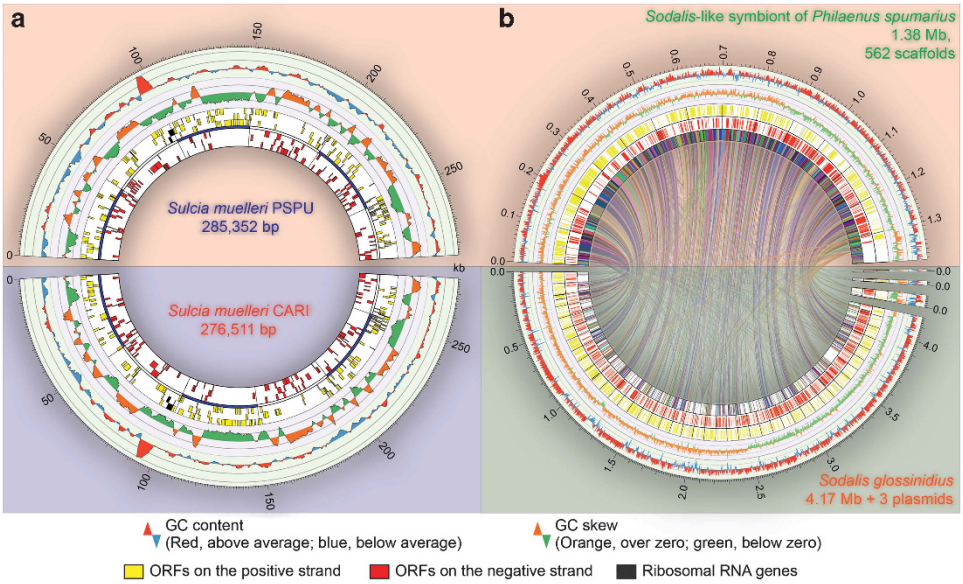

Figure 1 Comparison of genomes of the two *P. spumarius* symbionts (upper half in **a** and **b**) with previously sequenced genomes of related bacterial species (lower half in **a** and **b**). (**a**) Genome of *Sulcia muelleri*-PSPU compared with that of *S. muelleri*-CARI, showing arrangement and orientation of open reading frames, GC content, and GC skew. The two *Sulcia* genomes are completely syntenic except for a few genes present in one and not the other. (**b**) Genome of SLs-PSPU compared with that of *Sodalis glossinidius*, including its three plasmids. The innermost graph of SLs-PSPU shows concatenated scaffolds arranged according to position in the *S. glossinidius* genome; scaffolds are distinguished by colors and homology to sequences in the *S. glossinidius* is indicated by lines of the same color.

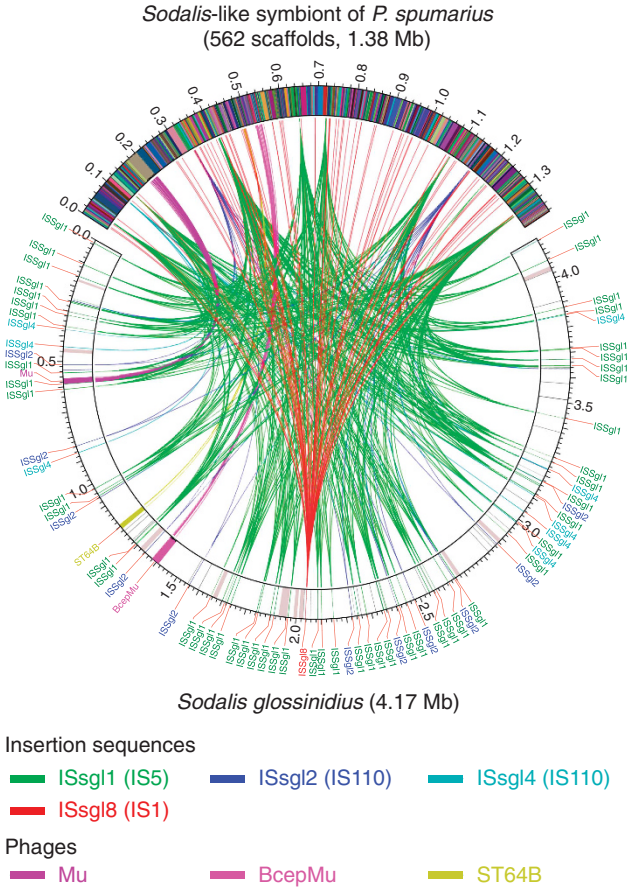

Figure 2 Sequences of SLs-PSPU that are derived from IS elements and phages, showing their relative locations in the *S. glossinidius* genome. A concatenated representation of scaffolds derived from SLs-PSPU is shown at the top with each scaffold in a different color. IS elements and phages identified within the *S. glossinidius* genome and identified by Belda *et al.* (2010) are shown with gray bands on the bar representing the *S. glossinidius* genome. Among them, those showing significant similarities to the SLs scaffolds are labeled with their name and colored as indicated at the bottom of the figure. Names of IS elements and phages proposed by Belda *et al.* (2010) were adopted. In the legends, the family name of the IS elements are in parentheses. Portions of ISsgl8 (IS1 family) are present as a single copy in the *S. glossinidius* genome but found on 235 scaffolds in the SLs-PSPU genome.

in the SLs-PSPU genome but includes only one copy in the *S. glossinidius* genome, indicating many transpositions in the SLs lineage following its divergence from the lineage leading to *S. glossinidius* (Figure 2). Using a full-length reconstruction of IS1SLs (767 bp) as a query to search the assembly for the SLs-PSPU genome, homologous sequences were found to be represented on 235 of 562 scaffolds, mostly on the terminal regions, indicating that copies of IS1SLs contributed to the fragmentation of the genomic assembly (Supplementary Figure S3). Other IS types were present on ends of many other scaffolds. Copy numbers of multicopy IS elements could not be determined directly from the assembly. Using quantitative PCR to estimate copy number for IS1SLs for individual *P. spumarius* from different localities, we found averages ranging from 277 to 1062 IS1SLs copies per genome (Supplementary Figure S2E), indicating that the elements are still actively transposing and are creating diversity among SLs genomes within *P. spumarius* populations. The sequenced SLs genome came from a population with an average of 277 copies of IS1SLs per SLs genome, close to the number of scaffolds on which IS1 sequences were present. Copies of IS1SLs were also detected from some other members of Philaenini that host a SLs (Supplementary Figure S2F).

Genomic evidence for SLs as an evolutionary recent symbiont acquisition

The genomic features of SLs-PSPU resemble those of other symbionts that have undergone relatively recent genome reduction and still retain many pseudogenes and IS copies; previous examples include *S. glossinidius* and related symbionts in grain weevils (Plague *et al.*, 2008; Clayton *et al.*, 2012) and *Serratia symbiotica*, a facultative symbiont of aphids (Burke and Moran, 2011). Although the overall length of scaffolds for SLs-PSPU is 1380 kb, the chromosome is likely somewhat larger because assemblies of multicopy IS elements collapsed into fewer copies. Nonetheless, SLs has a far smaller genome than do closely related symbionts, *S. glossinidius* (4.5 Mb) and the grain weevil symbiont SOPE (4.2 Mb), both of which are reduced compared with *Sodalis* HS (5.2 Mb), a free-living relative putatively representative of a genome similar to that of non-symbiotic ancestors of SLss (Clayton *et al.*, 2012). The SLs-PSPU genome contains at least 173 pseudogenes (containing 25 disrupted by an IS sequence). This is an underestimate, as smaller, highly degraded pseudogenes would not be recognized.

Full-length sequences for rRNA genes were not retrieved from the genome assembly for SLs-PSPU, and sequencing of cloned PCR products for 16S rRNA revealed that this reflected the occurrence of two divergent rRNA sequences for different operon copies in SLs-PSPU strains from *P. spumarius* and from other species in the tribe Philaenini (Supplementary Figure S4A), resulting in fragmented assemblies. Previously, a similar case of rRNA operon divergence was observed in the related grain weevil symbionts, in which alternative copies within a genome were more divergent than those present in symbiont strains in different host species (Dale *et al.*, 2003). Highly reduced symbiont genomes typically contain only a single copy of rRNA genes (McCutcheon and Moran, 2012), and the divergence of rRNA operons may represent an evolutionarily transitional stage during the elimination of excess rRNA gene copies (Smith *et al.*, 2013). Indeed, we found that one copy is transcriptionally

silenced (Supplementary Figures S4B and C) and represents a pseudogene.

Evolutionary succession of symbiotic partners in spittlebugs

Auchenorrhyncha appear to be derived from a Permian-age ancestor in which *Sulcia* was paired with a betaproteobacterial symbiont ancestral to *Zinderia* (Bennett and Moran, 2013; Koga *et al*., 2013). *Sulcia* was retained in a large proportion of descendant lineages, whereas the betaproteobacterial symbiont was more often lost and replaced by new symbionts, including *Hodgkinia cicadicola* (Alphaproteobacteria) in cicadas (Cicadoidea), *Baumannia cicadellinicola* (Gammaproteobacteria) in sharpshooters (Cicadellidae: Cicadellinae) and SLs in *P. spumarius* and close relatives. Where it has been retained, in some planthoppers (Fulgoridae), leafhoppers (Cicadellidae: Deltocephalinae), and spittlebugs (Urban and Cryan, 2012; Bennett and Moran, 2013; Koga *et al*., 2013), the betaproteobacterial symbiont appears to have a highly degenerate genome. For example, the *Zinderia*-CARI genome contains only 202 protein-coding genes, exhibits extreme divergence of protein and rRNA sequences, the most extreme AT nucleotide bias (13.5% G + C) yet reported, and an altered genetic code in which TGA encodes tryptophan rather than Stop (McCutcheon *et al*., 2009). Recently the betaproteobacterial symbiont of deltocephaline leafhoppers, *Nasuia deltocephalinicola*, was shown to have the smallest genome (112 kb) sequenced to date with the same altered genetic code (Bennett and Moran, 2013). Based on molecular phylogenies, *Nasuia* and *Zinderia* belong to the same symbiont clade, and they are proposed to have descended from an ancestral symbiont infecting a shared ancestor of leafhoppers (Cicadellidae) and spittlebugs (Cercopoidea) (Urban and Cryan, 2012; Bennett and Moran, 2013; Koga *et al*., 2013). Both lack many genes for energy production (Bennett and Moran, 2013), supporting elimination of these genes before the divergence of *Nasuia* and *Zinderia*, and before the acquisition of SLs by an ancestor of Philaenini (Figure 4).

The evolutionary transition from *Sulcia* + *Zinderia* to *Sulcia* + SLs probably included a stage in which *Sulcia*, *Zinderia* and SLs coexisted. Indeed, some modern spittlebugs do contain *Sulcia* + *Zinderia* + SLs (Koga *et al*., 2013). In *P. spumarius*, the SLs is housed in a bacteriocyte type that is distinct from that housing *Zinderia* (Koga *et al*., 2013). As *Sodalis* strains are widespread as maternally transmitted facultative symbionts of many insect hosts (Snyder *et al*., 2011), an ancestor of the SLs may have initially colonized an ancestor of *P. spumarius* and other Philaenini as an opportunistic symbiont, then evolved to become an obligate mutualist, enabling the complete elimination of *Zinderia*. The confinement of each symbiont to distinct bacteriocyte types is likely key for their ability to coexist within a host, without competitive exclusion.

Implications of SLs acquisition for the symbiotic metabolism of P. spumarius

The inferred symbiotic metabolism of *P. spumarius*, in which SLs replaces *Zinderia*, shows striking differences from that inferred for *C. arizonana*, which hosts *Sulcia* plus *Zinderia*. The largest change in inferred metabolic capabilities comes from the addition of novel pathways present in SLs, which though having a reduced genome still retains many more genes than are present in either *Zinderia* or the related *Nasuia*. *Sulcia*-PSPU and *Sulcia*-CARI are largely similar in encoding pathways for production of seven essential amino acids (leucine, valine, isoleucine, phenylalanine, arginine, threonine and lysine) from nonessential amino acids such as aspartate and glutamate. However, *Sulcia*-PSPU does show one important difference from *Sulcia*-CARI: *Sulcia-PSPU* possesses *gdhA*, encoding glutamate dehydrogenase, which catalyzes the production of 2-oxoglutarate from glutamate. This gene is one of only the three genes intact in *Sulcia*-PSPU but absent from other *Sulcia*, although small fragments showing sequence homology still remain at the same chromosomal position in *Sulcia*-CARI (Supplementary Figure S5A). As further evidence that the gene is retained from ancestral *Sulcia* rather than acquired through horizontal gene transfer, a phylogenetic tree shows that the most closely related sequences are from *Blattabacterium* species (Supplementary Figure S5B), cockroach symbionts that are the closest sequenced relatives of *Sulcia*.

For amino-acid biosynthesis, SLs acquisition has resulted in some redundancy of functions between paired symbionts of *P. spumarius*. *Sulcia* and *Zinderia* of *C. arizonana* show perfect complementarity of genes corresponding to enzymatic steps in the production of essential amino acids, with *Zinderia* encoding intact pathways for production of methionine, histidine and tryptophan (McCutcheon and Moran, 2010). SLs-PSPU also shows some complementarity with *Sulcia*: it retains these three pathways, thus replacing the lost *Zinderia* functions, and has lost all genes for steps in the production of four amino acids (valine, isoleucine, leucine and phenylalanine) for which *Sulcia* retains biosynthetic pathways (Figure 3a). However, in contrast to the *Zinderia*–*Sulcia* partnership, SLs-PSPU and *Sulcia*-PSPU show some redundancy in amino-acid biosynthetic abilities; both symbionts retain genes underlying the synthesis of arginine, lysine and threonine.

SLs-PSPU retains many other genes not present in *Zinderia*. These include genes underlying F-type ATPase and pathways for production of cofactors (biotin, folate, flavin, vitamin B6, ubiquinone,

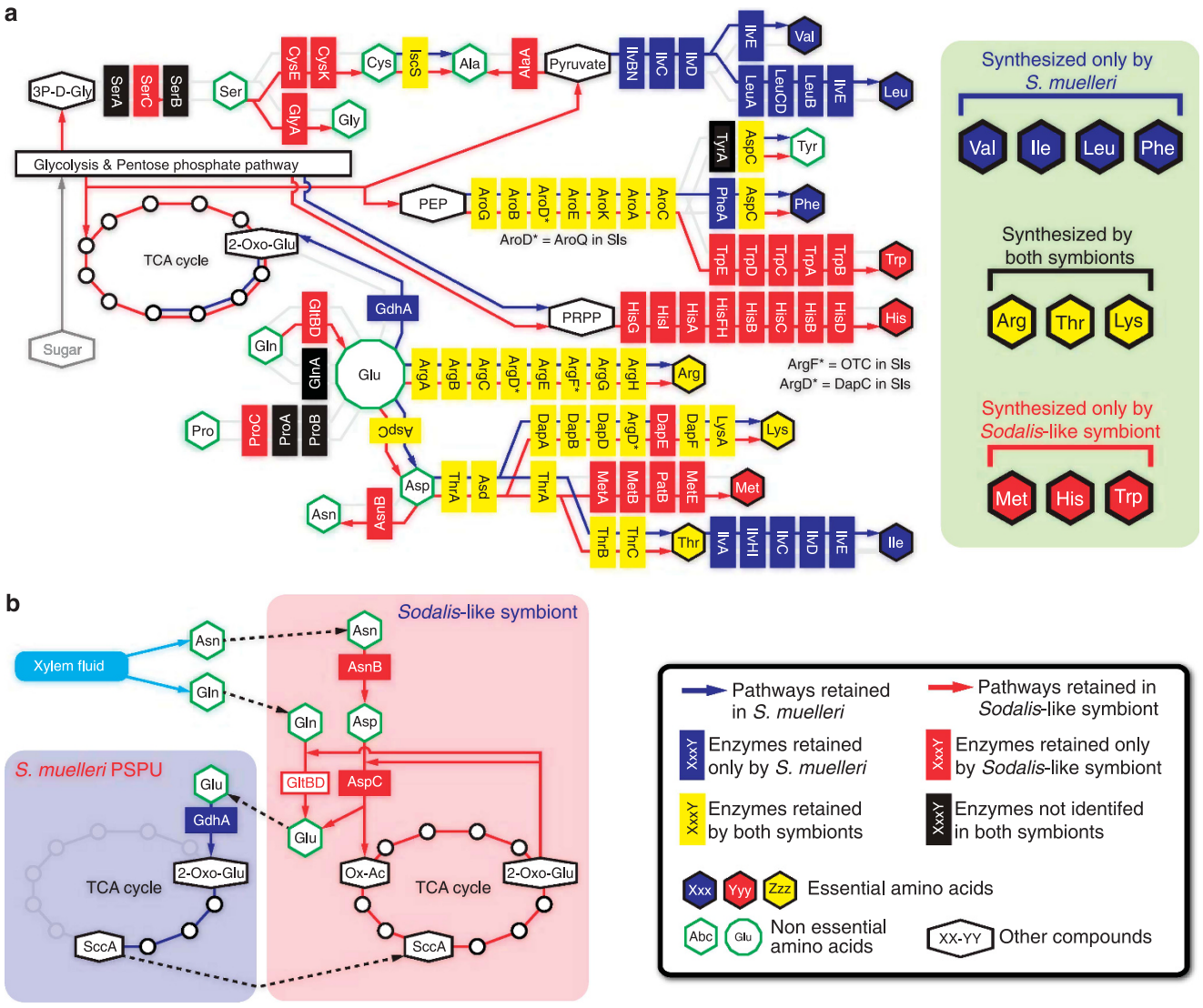

Figure 3 Reconstruction of symbiont-based metabolism in *P. spumarius* showing reactions for which SLs and *Sulcia* possess intact genes. (**a**) Pathways for amino-acid biosynthesis and energy production. (**b**) Hypothesized movement of molecules between compartments for SLs and *Sulcia*, showing route for production of glutamic acid used for producing TCA cycle intermediates.

glutathione, heme, thiamine), fatty acids, phospholipids, purine and pyrimidine nucleotides, terpenoids and components of the bacterial cell wall (Supplementary Table S1). Genes underlying steps in the TCA cycle are retained, indicating capability for producing energy from carbon skeletons derived from sugar or from amino-acid catabolism. Xylem sap lacks sugar, and amino acids are the primary energy substrates available for use by xylem-feeding insects (Brodbeck *et al.*, 1990). We hypothesize that the *Sulcia*-encoded glutamate dehydrogenase (GdhA) enables the generation of 2-oxoglutarate from glutamate and that this enters the SLs-PSPU as a substrate in the TCA cycle, providing a major source of ATP for the symbiotic system (Figure 3b). This possibility would depend on availability of excess glutamate, while glutamine and asparagine comprise most amino acids in the xylem (Brodbeck *et al.*, 1990). Glutamate could be provided by SLs PSPU, which retains intact *gltBD*, encoding glutamate synthase, the enzyme that catalyzes the production of glutamate from glutamine. In sequenced relatives of SLs PSPU (*S. glossinidius*, *Sodalis* HS and SOPE of grain weevils), *gltBD* is absent or fragmented (Supplementary Figure S5C), but it is intact in SLs PSPU. These genes are also lacking from *Zinderia*-CARI and *Nasuia* (Bennett and Moran, 2013), supporting their early loss from the genome of the betaproteobacterial partner of *Sulcia*, and suggesting that the acquisition of SLs conferred the capability to produce glutamate using glutamine, one of the most abundant organic molecules in the diet (Wiegert, 1964).

Thus, we hypothesize that the unique retention of *gltBD* in SLs PSPU and of *gdhA* in *Sulcia* PSPU may enable a novel route for exploiting relatively

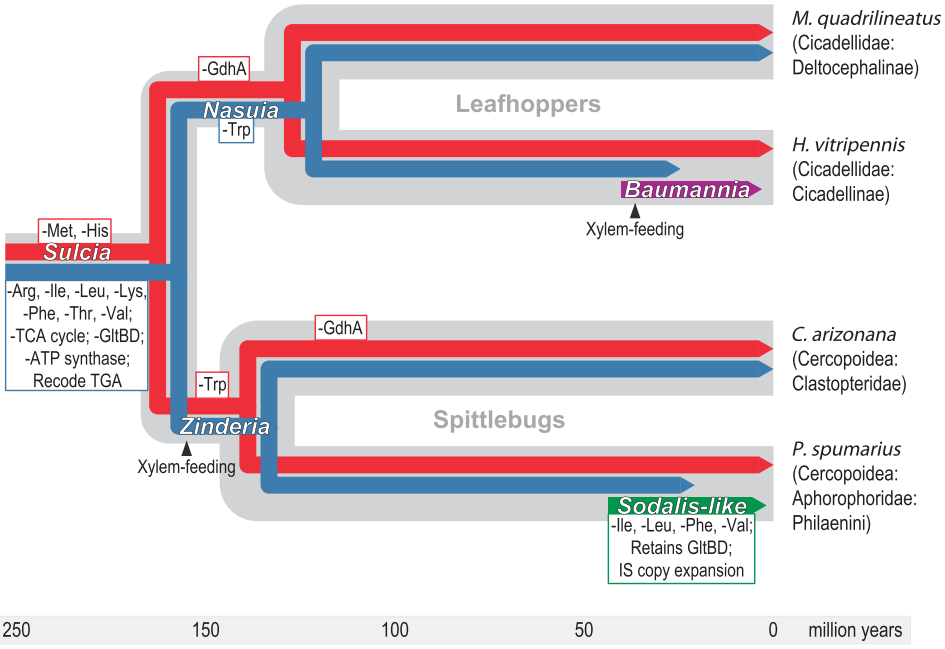

Figure 4 Hypothesized evolutionary sequence leading to current symbiotic associations of *Philaenus spumarius*, showing events discussed in text. Timescale is approximate, and dates are based on Cryan and Svenson (2010). Representative insect species for which symbiont genomes were sequenced in previous studies are *Macrosteles quadrilineatus* (Bennett and Moran, 2013), *Homalodisca vitripennis* (McCutcheon and Moran, 2007) and *Clastoptera arizonana* (McCutcheon and Moran, 2010).

abundant nonessential amino acids for energy production (Figure 4). Earlier studies on host plant utilization by *P. spumarius* show that it preferentially feeds on plants that are nitrogen fixers, particularly those that transport nitrogen in the xylem in the form of glutamine and asparagine (Thompson, 1994, 2004). In the xylem of studied hosts of *P spumarius*, amino acids comprise 98% of organic matter, and glutamine and asparagine comprise ~75% of these amino acids (Wiegert, 1964). Thus, efficient use of these amides for energy production is likely key for using the main xylem transpiration stream as a sole food source. This added route for extracting energy may have expanded the niche of *P. spumarius* and other members of the Philaenini containing the SLs (Koga *et al.*, 2013); this is among the most ecologically successful groups of spittlebugs, in terms of abundance and geographic and host plant ranges (Weaver and King, 1954; Thompson, 1994).

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank Gordon Bennett for providing the insects collected in CA, and Gordon Bennett, Dan Sloan and Philipp Engel for technical advice and discussion. Funding was from NSF 1106195 to NM and from Yale University and the University of Texas.

References

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

Belda E, Moya A, Bentley S, Silva FJ. (2010). Mobile genetic element proliferation and gene inactivation impact over the genome structure and metabolic capabilities of *Sodalis glossinidius*, the secondary endosymbiont of tsetse flies. *BMC Genomics* **11**: 449.

Bennett GM, Moran NA. (2013). Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol Evol* **5**: 1675–1688.

Braendle C, Miura T, Bickel R, Shingleton AW, Kambhampati S, Stern DL. (2003). Developmental origin and evolution of bacteriocytes in the aphid-*Buchnera* symbiosis. *PLoS Biol* **1**: E21.

Brodbeck BV, Mizell RF III, French WJ, Andersen PC, Aldrich JH. (1990). Amino acids as determinants of host preference for the xylem feeding leafhopper, *Homalodisca coagulata* (Homoptera: Cicadellidae). *Oecologia* **83**: 338–345.

Burke GR, Moran NA. (2011). Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol Evol* **3**: 195–208.

Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M *et al.* (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **36**: D623–D631.

Clayton AL, Oakeson KF, Gutin M, Pontes A, Dunn DM, Niederhausern von AC *et al.* (2012). A novel human-infection-derived bacterium provides insights into the

evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genet* **8**: e1002990.

Conord C, Despres L, Vallier A, Balmand S, Miquel C, Zundel S *et al.* (2008). Long-term evolutionary stability of bacterial endosymbiosis in Curculionoidea: additional evidence of symbiont replacement in the Dryophthoridae family. *Mol Biol Evol* **25**: 859–868.

Cryan JR, Svenson GJ. (2010). Family-level relationships of the spittlebugs and froghoppers (Hemiptera: Cicadomorpha: Cercopoidea). *Syst Entomol* **35**: 393–415.

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

Dunbar HE, Wilson ACC, Ferguson NR, Moran NA. (2007). Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol* **5**: e96.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792.

Engel P, Martinson VG, Moran NA. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proc Natl Acad Sci USA* **109**: 11002–11007.

Hoffman G, McEvoy PB. (1985). Mechanical limitations on feeding by meadow spittlebugs *Philaenus spumarius* (Homoptera: Cercopidae) on wild and cultivated host plants. *Ecol Entomol* **10**: 415–426.

Horsfield D. (1977). Relationships between feeding of *Philaenus spumarius* (L.) and the amino acid concentration in the xylem sap. *Ecol Entomol* **2**: 259–266.

Koga R, Bennett GM, Cryan JR, Moran NA. (2013). Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. *Environ Microbiol* **15**: 2073–2081.

Koga R, Meng X-Y, Tsuchida T, Fukatsu T. (2012). Cellular mechanism for selective vertical transmission of an obligate insect symbiont at the bacteriocyte-embryo interface. *Proc Natl Acad Sci USA* **109**: E1230–E1237.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D *et al.* (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* **19**: 1639–1645.

Lambert JD, Moran NA. (1998). Deleterious mutations destabilize ribosomal RNA in endosymbiotic bacteria. *Proc Natl Acad Sci USA* **95**: 4458–4462.

Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A. (2011). *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genet* **7**: e1002357.

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

Lerat E, Daubin V, Moran NA. (2003). From gene trees to organismal phylogeny in prokaryotes: the case of the γ-proteobacteria. *PLoS Biol* **1**: E19.

Malone M, Watson R, Pritchard J. (1999). The spittlebug *Philaenus spumarius* feeds from mature xylem at the full hydraulic tension of the transpiration stream. *New Phytologist* **143**: 261–271.

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

McCutcheon JP, McDonald BR, Moran NA. (2009). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc Natl Acad Sci USA* **106**: 15394–15399.

McCutcheon JP, Moran NA. (2012). Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* **10**: 13–26.

McCutcheon JP, Moran NA. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* **2**: 708–718.

Moran NA. (1996). Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* **93**: 2873–2878.

Moran NA. (2007). Symbiosis as an adaptive process and source of phenotypic complexity. *Proc Natl Acad Sci USA* **104**: 8627–8633.

Moran NA, McCutcheon JP, Nakabachi A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* **42**: 165–190.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* **35**: W182–W185.

Moriyama M, Koga R, Hosokawa T, Nikoh N, Futahashi R, Fukatsu T. (2012). Comparative transcriptomics of the bacteriome and the spermalege of the bedbug *Cimex lectularius* (Hemiptera: Cimicidae). *Appl Entomol Zool* **47**: 233–243.

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

Raven PH, Evert RF, Eichhorn SE. (2005). *Biology of Plants*, 7th edn. W. H. Freeman: New York, NY, USA.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**: D32–D36.

Smith WA, Oakeson KF, Johnson KP, Reed DL, Carter T, Smith KL *et al.* (2013). Phylogenetic analysis of symbionts in feather-feeding lice of the genus *Columbicola*: evidence for repeated symbiont replacements. *BMC Evol Biol* **13**: 109.

Snyder AK, McMillen CM, Wallenhorst P, Rio RVM. (2011). The phylogeny of *Sodalis*-like symbionts as reconstructed using surface-encoding loci. *FEMS Microbiol Lett* **317**: 143–151.

Stamatakis A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.

Sugawara H, Ohyama A, Mori H, Kurokawa K. (2009). Microbial Genome Annotation Pipeline (MiGAP) for diverse users. The 20th International Conference on Genome Informatics (GIW2009).

Tamas I, Klasson L, Canbäck B, Näslund AK, Eriksson A-S, Wernegreen JJ *et al.* (2002). 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**: 2376–2379.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Thompson V. (2004). Associative nitrogen fixation, C4 photosynthesis, and the evolution of spittlebugs (Hemiptera: Cercopidae) as major pests of neotropical sugarcane and forage grasses. *Bull Entomol Res* **94**: 189–200.

Thompson V. (1994). Spittlebug indicators of nitrogen-fixing plants. *Ecol Entomol* **19**: 391–398.

Toenshoff ER, Gruber D, Horn M. (2012). Co-evolution and symbiont replacement shaped the symbiosis between adelgids (Hemiptera: Adelgidae) and their bacterial symbionts. *Environ Microbiol* **14**: 1284–1295.

Toju H, Tanabe AS, Notsu Y, Sota T, Fukatsu T. (2013). Diversification of endosymbiosis: replacements, co-speciation and promiscuity of bacteriocyte symbionts in weevils. *ISME J* **7**: 1378–1390.

Urban JM, Cryan JR. (2012). Two ancient bacterial endosymbionts have coevolved with the planthoppers (Insecta: Hemiptera: Fulgoroidea). *BMC Evol Biol* **12**: 87.

van Ham RCHJ, Kamerbeek J, Palacios C, Rausell C, Abascal F, Bastolla U *et al.* (2003). Reductive genome evolution in *Buchnera aphidicola*. *Proc Natl Acad Sci USA* **100**: 581–586.

Weaver CF, King DR. (1954). Meadow spittlebug *Philaenus leucophthalmus* (L.). *Res Bull Ohio Agric Exp Station* **741**: 1–98.

Wiegert RG. (1964). Population energetics of meadow spittlebugs (*Philaenus spumarius* L.) as affected by migration and habitat. *Ecol Monogr* **34**: 217–241.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)