# Small, Smaller, Smallest: The Origins and Evolution of Ancient Dual Symbioses in a Phloem-Feeding Insect

Gordon M. Bennett* and Nancy A. Moran

Department of Ecology and Evolutionary Biology & Microbial Diversity Institute, Yale University

*Corresponding author: E-mail: gordon.bennett@yale.edu.

## Abstract

Many insects rely on bacterial symbionts with tiny genomes specialized for provisioning nutrients lacking in host diets. Xylem sap and phloem sap are both deficient as insect diets, but differ dramatically in nutrient content, potentially affecting symbiont genome evolution. For sap-feeding insects, sequenced symbiont genomes are available only for phloem-feeding examples from the suborder Sternorrhyncha and xylem-feeding examples from the suborder Auchenorrhyncha, confounding comparisons. We sequenced genomes of the obligate symbionts, *Sulcia muelleri* and *Nasuia deltocephalinicola*, of the phloem-feeding pest insect, *Macrosteles quadrilineatus* (Auchenorrhyncha: Cicadellidae). Our results reveal that *Nasuia*-ALF has the smallest bacterial genome yet sequenced (112 kb), and that the *Sulcia*-ALF genome (190 kb) is smaller than that of *Sulcia* in other insect lineages. Together, these symbionts retain the capability to synthesize the 10 essential amino acids, as observed for several symbiont pairs from xylem-feeding Auchenorrhyncha. *Nasuia* retains genes enabling synthesis of two amino acids, DNA replication, transcription, and translation. Both symbionts have lost genes underlying ATP synthesis through oxidative phosphorylation, possibly as a consequence of the enriched sugar content of phloem. Shared genomic features, including reassignment of the UGA codon from Stop to tryptophan, and phylogenetic results suggest that *Nasuia*-ALF is most closely related to *Zinderia*, the betaproteobacterial symbiont of spittlebugs. Thus, *Nasuia/Zinderia* and *Sulcia* likely represent ancient associates that have co-resided in hosts since the divergence of leafhoppers and spittlebugs >200 Ma, and possibly since the origin of the Auchenorrhyncha, >260 Ma.

**Key words:** gene loss, genome evolution, leafhopper, nutritional symbioses, *Nasuia deltocephalinicola*, *Sulcia muelleri*.

## Introduction

Obligate symbioses with bacteria have enabled insects to exploit novel niches, previously out of adaptive reach (Moran 2007). Prominent examples include plant sap-feeding insects in the order Hemiptera that have evolved partnerships with a diverse consortium of bacterial symbionts (Buchner 1965). By providing essential nutrients lacking in xylem sap or phloem sap, these relationships have permitted insect hosts to tap into these specialized diets (Buchner 1965; Baumann 2005; Moran et al. 2008). Symbiont acquisition has thereby enabled the Hemiptera to diversify into a species-rich and geographically ubiquitous insect order.

Specialization on xylem or phloem in sap-feeding Hemiptera is a conserved trait, corresponding to ancient clades (Grimaldi and Engel 2005; Cryan and Urban 2012). Both diet types are deficient in most essential amino acids and

many cofactors, but otherwise they have vastly different nutritive qualities (Mattson 1980; Andersen et al. 1989; Sandström and Petterson 1994; Sandström and Moran 1999). In general, phloem is more nutritious, with 10-fold greater levels of available organic nitrogen and a 1000-fold greater level of carbohydrates, mostly sugars (Mattson 1980; Andersen et al. 1989). However, the influence of phloem versus xylem sap diets on the evolution of symbioses remains unknown. Because phloem-feeding systems have been studied only in the hemipteran suborder, Sternorrhyncha (e.g., *Buchnera*-aphid symbiosis; Shigenobu and Wilson 2011), and xylem-feeding examples only in the Cicadomorpha (McCutcheon and Moran 2010), which is one of the two clades comprising the hemipteran suborder Auchenorrhyncha, diet type and phylogenetic relationships are confounded in comparisons.

Species in the Auchenorrhyncha rely on two obligate symbionts: *Sulcia muelleri* (Bacteroidetes), which has been present

within Auchenorrhyncha since it emerged 260–280 Ma (Moran et al. 2005), and an array of co-primary symbionts from different bacterial divisions (fig. 1). In several groups of xylem feeders, genomes of *Sulcia* and its partner symbiont have been sequenced; these include cicadas (*Sulcia* plus *Hodgkinia cicadicola* [Alphaproteobacteria]; McCutcheon et al. 2009b), spittlebugs (*Sulcia* plus *Zinderia insecticola* [Betaproteobacteria]; McCutcheon and Moran 2010; Koga et al. 2013), and sharpshooter leafhoppers (*Sulcia* plus *Baumannia cicadellinicola* [Gammaproteobacteria]; Wu et al. 2006). In each case, *Sulcia* can produce most of the 10 essential amino acids, but is unable to produce histidine or methionine. And, in each case, these missing pathways are perfectly complemented by those encoded in the co-primary symbionts (reviewed by McCutcheon and Moran 2010). This is in contrast to the aphid-*Buchnera* model system in which the provisioning of all 10 essential amino acids is accomplished by a single obligate symbiont (Shigenobu et al. 2000, reviewed by Baumann 2005). Like aphids, most groups of Auchenorrhyncha are phloem feeders, including the treehoppers and most leafhopper groups (Membracoidea: Membracidae, most Cicadellidae), and most planthopper groups (e.g., Fulgoroidea: Cixiidae, Delphacidae, Flatidae, Fulgoridae) (Dietrich 2003). Most of these are also host to dual bacterial symbionts (fig. 1), based on microscopy and 16S rRNA screening (Buchner 1965; Noda et al. 2012; Urban and Cryan 2012; Koga et al. 2013), but to date none has been studied using genomic approaches.

A central aspect of symbiont evolution is extreme genomic reduction resulting from maternal transmission and the internment of symbionts within specialized host cells (bacteriocytes) for millions of years. Repeatedly, symbionts lose pathways and genes involved in every functional category, and particularly in functions related to cellular regulation, DNA repair, cell envelope biosynthesis, and secondary metabolism (Shigenobu et al. 2000; Tamames et al. 2007; Moran et al. 2008; McCutcheon and Moran 2010, 2012). Indeed, auchenorrhynchan symbionts possess the smallest known cellular genomes. In some cases, these genomes appear insufficient to support cellular life (e.g., McCutcheon and von Dohlen 2011; McCutcheon and Moran 2012), yet they retain a subset of genes for synthesizing essential amino acids required by hosts (McCutcheon and Moran 2012).

The similarity in amino acid biosynthetic abilities of different *Sulcia* genomes suggests that *Sulcia* lost pathways for biosynthesis of methionine and histidine early during the diversification of the Auchenorrhyncha. *Sulcia* has continuously maintained associations with symbionts providing these functions, even as these partners were sometimes replaced over evolutionary time in particular lineages (fig. 1; Buchner 1965; Moran et al. 2005; McCutcheon and Moran 2010). It remains unclear whether the original partner of *Sulcia* persists in any modern Auchenorrhyncha. However, *Zinderia* of spittlebugs, *Nasuia deltocephalinicola* of deltocephaline leafhoppers (Noda

et al. 2012), and *Vidania fulgoroideae* of fulgorid planthoppers (Gonella et al. 2011) form a clade within Betaproteobacteria in 16S rDNA-based phylogenies. This *BetaSymb* lineage has been proposed to represent the ancestral partner of *Sulcia* that has been replaced in several descendant lineages of the Auchenorrhyncha, perhaps in response to dietary or ecological shifts by hosts (see fig. 1; Koga et al. 2013). Conclusions from sequence-based phylogenies are hampered by extreme acceleration in sequence evolution and extreme nucleotide base compositional biases of symbionts. Shared genomic features across *Zinderia*, *Nasuia*, and *Vidania* could provide more definitive support for *BetaSymb* as a clade and as an ancestral symbiont of Auchenorrhyncha. But, to date, only a single *Zinderia* genome has been available.

To address the questions of 1) the pattern of genome evolution in dual symbionts of a phloem-feeding host, and 2) the origin for the *BetaSymb* lineage, we have sequenced the complete genomes of symbionts of the Aster Leafhopper (ALF), *Macrosteles quadrilineatus* (Cicadellidae: Deltocephalinae). This species is a widespread pest of North American agriculture (Frost et al. 2011). As found in other members of Deltocephalinae, it harbors *Sulcia* and the betaproteobacterial co-primary symbiont, *Nasuia* (Noda et al. 2012; Koga et al. 2013).

## Materials and Methods

### Material Acquisition and Sequencing

Samples of *M. quadrilineatus* were field-collected from Yale West Campus, West Haven, CT, USA, on *Symphyotrichum lanceolatum* (Asteraceae). Specimens were identified according to Kwon (1988) and with the barcoding mitochondrial locus Cytochrome Oxidase I (following Le Roux and Rubinoff 2009). In order to confirm the bacteriome association of the targeted bacterial lineages, whole-specimen Fluorescent in situ Hybridization was performed following Koga et al. (2009). Fluorescing probes were designed as follows: Bet940 (5′-TTAATCCACATCATCCACCG-3′) for the Betaproteobacteria labeled with 5′ AlexaFlour-555 probe modification, and Sulc664R (5′-CCMCACATTCCAGMTACTCC-3′) for *Sulcia*, labeled with a 5′ AlexaFlour-647 modification (Invitrogen). Briefly, specimens were dissected in 70% EtOH, fixed in Carnoy's solution, and incubated in alcoholic hydrogen peroxide for 2 weeks. Material was then rinsed and stored in 100% ethanol, rehydrated in PBSTx, and rinsed with a hybridization buffer. Specimens were incubated in a hybridization-probe solution, containing the oligo-probes and DAPI counterstain. Specimens were imaged with a Nikon Eclipse TE2000-U epifluorescence microscope (fig. 2).

Ten individual insects were pooled for genomic DNA (gDNA) sequencing. Total gDNA was extracted using Qiagen DNeasy extraction kit (Qiagen Corp.) in a 50 μl volume. Genomic sequencing was done using Illumina's MiSeq platform. gDNA was sheared into library insert sizes of 500 bp,
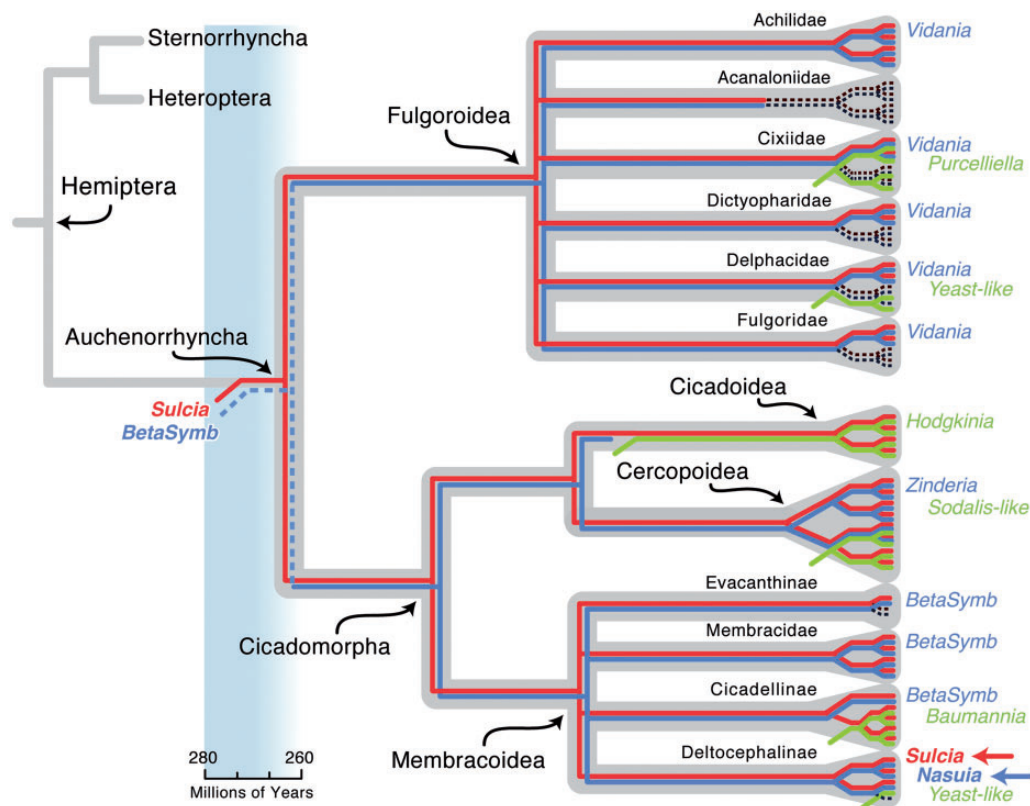
**Fig. 1.**—Schematic summary of the evolution of symbiotic associations in major lineages of Auchenorrhyncha, with emphasis on the Cicadomorpha (cicadas, spittlebugs, leafhoppers, and treehoppers). The focal symbionts, *Nasuia* and *Sulcia* from *Macrosteles quadrilineatus* (Cicadellidae: Deltocephalinae), are demarcated with a blue and red arrow, respectively. Phylogenies show evolutionary events in symbioses based on molecular phylogenetic studies (Moran et al. 2003, 2005; Takiya et al. 2006; Urban and Cryan 2012; Noda et al. 2012; Bressan and Mulligan 2013; Hou et al. 2013; Koga et al. 2013). Events are color-coded as follows: Host insect phylogeny (from Cryan and Urban 2012) gray background; *Sulcia*, red; betaproteobacterial symbiont lineage (*BetaSymb* = *Zinderia* + *Nasuia* + *Vidania*), blue; symbiont loss, black dashed line; and symbiont replacement, green. Identified symbiont names are given at the tips. This schematic is not a full account of host relationships and symbiont associations, as many species and genera remain to be explored, and some inferred losses may reflect incomplete knowledge. *Sulcia* is hypothesized to have been associated with the Auchenorrhyncha since its emergence 260–280 Ma (see time scale and corresponding blue box; Moran et al. 2005). Although further evidence is required to test the hypothesis that the Auchenorrhyncha share a common betaproteobacterial symbiont (*BetaSymb*) that was acquired early in its diversification, this potential relationship is shown here with a blue dashed line. The hypothesis that the Cicadomorpha do share the closely related *BetaSymb* (*Zinderia* + *Nasuia*), as evidenced by our genomic and phylogenetic results, is shown in a solid blue line (as is *Vidania* found throughout the Fulgoroidea based on Urban and Cryan 2012).

and library amplification (eight polymerase chain reaction [PCR] cycles) was performed with high-fidelity polymerase (KAPA Bio HiFi) to reduce known problems of biased amplification in AT-rich symbiont genomes (Aird et al. 2011; Sloan and Moran 2012a). The prepared library was sequenced using paired-end orientation on a 2 × 250 bp lane. Library preparation and sequencing were done at the Yale Center for Genome Analysis.

## Assembly

Because sequenced samples include a complex metagenomic mixture of host and bacterial symbionts, we implemented a multi-tiered approach to produce preliminary genome contigs, collect associated bacterial reads, and reassemble and improve

scaffolds (Sloan et al. 2013). A first pass assembly was done on 1/16th of the read data (1,937,157 reads) using Velvet v1.2.08 with a kmer length of 63 (Zerbino and Birney 2008). Whole genome assembly returned nearly 12,000 contigs greater than 500 bp, the majority of which had relatively low kmer coverage. Velvet yielded 10 large contigs of over 10 kb in length, which were identified using BlastN searches against the NCBI nr database. The three largest contigs, one at 190 kb and two at ~56 kb, were unambiguously assigned to *Sulcia* and to Betaproteobacteria, respectively (supplementary fig S1, Supplementary Material online). We hereafter refer to these symbiont lineages as *Sulcia*-ALF (of the Aster Leafhopper) and *Nasuia*-ALF. The remaining large contigs represent insect nuclear DNA and the complete mitochondrial genome.
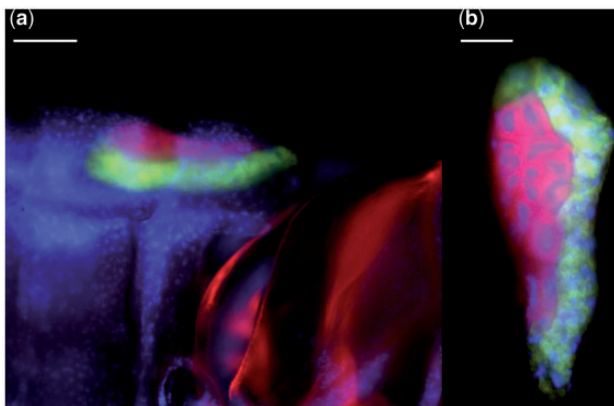
FIG. 2.—Fluorescence in situ hybridization of the bacteriome of *Macrosteles quadrilineatus* (Cicadellidae: Deltocephalinae), housing *Sulcia muelleri* (green; Bacteroidetes), and *Nasuia deltocephalinicola* (red; Betaproteobacteria). Blue fluorescence shows host nuclear DNA. (*a*) Localization of the bacteriome along the anterolateral abdominal segments. (*b*) The dissected bacteriome comprising two distinct bacteriocyte types that house *Sulcia*-ALF and *Nasuia*-ALF. White scale bars equal 100 μm (*a*) and 50 μm. (*b*) NOTE.—Red autofluorescence of the host thoracic plates.

To remove symbiont-derived reads from the total read pool, data were mapped to the Velvet contigs using SOAP2 v2.21 (settings: -m 10 -×1000 -g 3 -r 2; Li et al. 2009). Isolated symbiont reads were then reassembled using MIRA (Chevreux et al. 1999) with paired-end insert size set to 100–600 bp and normal quality settings. MIRA confirmed the 190 kb scaffold of *Sulcia*-ALF, but broke *Nasuia*-ALF sequences into three large contigs: 12, 42, and ~54 kb. MIRA also produced a 5 kb contig including a 1.1 kb tandem repeat that assembled with the ends of the two largest contigs, overlapping each end by over 450 bp. The Velvet and Mira contigs were assembled into a consensus scaffold that filled most ambiguous sections introduced by Velvet (e.g., strings of Ns). This merged the two largest *Nasuia*-ALF contigs, leaving two large scaffolds of ~102 kb and ~12 kb in length. These were assembled using PCR and Sanger sequencing (primers: 98836F-ACGCCGCCGAGAATAAGTG and 100182R-ATACAC AACGCGCCCTTCG). The remaining ends of the *Nasuia*-ALF scaffold contained a 400 bp overlap that was further supported by paired-end read data that bridged the two ends.

To assess the quality of the merged scaffolds, to improve genome assemblies, and to determine overall coverage for the symbiont genomes, total read data were mapped to the symbiont scaffolds using SOAP2. Read mapping was run iteratively until it converged on a consensus genome map, using custom Perl scripts (Sloan and Moran 2012a). This corrected indels and base pair substitutions within initial assemblies. To verify complete coverage of the circular bacterial genomes, scaffolds were broken, reoriented, and reads were remapped to confirm closure. *Sulcia*-ALF read coverage was relatively

uniform, whereas *Nasuia*-ALF mapping revealed a coverage spike of roughly 3-fold, starting at the nucleotide position 55,964, and supporting the potential repeat region assembled by Mira. The 1.1 kb repeat is too large to verify with the 250 bp reads, so PCR amplification and Sanger sequencing of the region was used to confirm the boundaries where the assembly initially broke (primers: 55201F-ATTCGCC TCATCCCCTCTCG + 56233R-AGCCATAACCTTTTGCGTCGT) and to confirm the existence of the internal repeat with inverse primers (primers: 56954F-AAATTAAGAGAGCTAGG + 56087R-TTTAACATCTATCGCTTG). The copy number of genomic repeats is difficult to verify and likely varies within insect hosts and within the symbionts that are polyploid (see Woyke et al. 2010), as has been shown in *Portiera*, the endosymbiont of whiteflies (Sloan and Moran 2013). Thus, the reported genome includes a single copy of the repeating motif flagged with a repeat annotation.

To determine whether other symbionts are present in the samples, and to search for potentially unintegrated *Sulcia*-ALF and *Nasuia*-ALF sequences, Velvet contigs longer than 500 bp were identified using translated Blastx searches against the NCBI nr database and taxonomically binned with MEGAN4 (Huson et al. 2011). A vast majority of the reads were placed in insect groups or were unassignable; these likely represent high copy insect genomic regions. Twenty-four contigs (500–700 bp in length) were close matches to regions of the genome of *Arsenophonus nasoniae* (Gammaproteobacteria). *Arsenophonus* is a cluster of bacterial symbionts present in many insect groups and in some plants (Nováková et al. 2009). Given the extremely fragmented genomic contigs for *Arsenophonus* and its low relative coverage, it is unlikely to represent an obligate primary symbiont of *Macrosteles*.

## Genome Annotation

First pass annotations were done with RAST and the Joint Genome Institute IMG ER pipelines (Aziz et al. 2008; Markowitz et al. 2009), using the general bacterial translation table. These pipelines produced a relatively complete annotated *Sulcia*-ALF genome sequence, but yielded many unidentifiable short open reading frames (ORFs) for *Nasuia*-ALF. Previous research indicated that the small genomes of *Hodgkinia* and *Zinderia* use an alternative genetic code that has reassigned the UGA stop codon to tryptophan (McCutcheon et al. 2009b; McCutcheon and Moran 2010). To see if this is the case for *Nasuia*, Glimmer3 was used to predict coding regions with the modified translation table 4 (Delcher et al. 2007). Use of the alternative table extended many previously truncated protein-coding genes to their expected lengths, providing strong evidence that *Nasuia* uses a UGA Stop → Trp recoding. Many genes utilize UGA-Trp codons at UGG-Trp conserved in other species with the standard bacterial code (e.g., *hisl*, *rpoB*, *dnaN*, etc.; see also McCutcheon and Moran 2010). All inferred genes for both symbionts were manually checked and

curated using phmmer and hmmerscan searches with available databases (NCBI nr, TIGRfam, SwissProt, and Pfam) in HMMER3 (Finn et al. 2011). The EcoCyc and KEGG databases were used further to determine gene pathways and function (Moriya et al. 2007; Keseler et al. 2011). RNA coding regions were identified using RNAammer, tRNAscan-SE, and searches against the Rfam database (Schattner et al. 2005; Lagesen et al. 2007; Burge et al. 2012). Genomes were manually annotated with Geneious Pro v5.5.8 (Drummond et al. 2010).

## Phylogenomics

To infer phylogenetic relationships of *Nasuia*-ALF, we used Phyla_Amphora2, which identifies and aligns conserved genomic markers from a set of phylum-specific genes (Wang and Wu 2013). Conserved orthologs were determined using Amphora's marker scan with HMMER3 (E-value cutoff of 0.01), which generated a 42 gene amino acid alignment (approx. 10k columns) for 102 Betaproteobacterial taxa (supplementary table S1, Supplementary Material online). The Betaproteobacterial symbiont of the mealybug *Planococcus citri*, *Tremblaya princeps* (placed in Betaproteobacteria: Burkholderiaceae; see Alves et al. 2013), was not included in the analysis as many of its genes were missing and our match threshold excluded nearly half of the available genes. Four outgroups were selected from Gammaproteobacteria and Alphaproteobacteria, and homologous genes were extracted using Blastp searches. Individual gene sets were aligned using MAFFT v7.017 with the L-INS-i model (Katoh and Standley 2013), and manually curated to remove large gaps and ambiguously aligned regions. Amino acid likelihood substitution models were determined for each gene matrix with ProtTest3 (Darriba et al. 2011). Phylogenetic trees were inferred using maximum likelihood criteria with RAxML v7.4.4 (Stamatakis 2006). A concatenated matrix was partitioned by loci and assigned gene-specific GTR-based amino acid model. Tree searches were conducted for 1000 bootstraps with a final maximum likelihood search.

# Results

## Genome Assembly

The Illumina MiSeq run yielded 30,994,516 quality-filtered reads. Sequence assembly and PCR verification produced complete genomes for *Sulcia*-ALF and *Nasuia*-ALF, with each genome comprising roughly 1.3% of total sequencing data. Coverage mapping of the merged assemblies for *Sulcia*-ALF and the *Nasuia*-ALF produced unbroken scaffolds with high mean sequencing depths of 500× and 850×, respectively. Genome sizes for *Sulcia*-ALF and *Nasuia*-ALF are 190,733 and 112,091 bp (counting only a single copy of the 1.1 kb tandem repeat; see Materials and Methods), respectively. Both are the smallest in their phyla, and *Nasuia*-ALF represents the smallest bacterial genome reported to date (fig. 3;

McCutcheon and Moran 2012). The GC content of both symbionts is low: *Sulcia*-ALF at 24.0% and *Nasuia*-ALF at 17.1%.

## *Sulcia*: Gene Loss in a Stable Symbiont

*Sulcia*-ALF contains a genome of only 190 kb, whereas previous *Sulcia* genomes, from several xylem-feeding auchenorrhynchan species, are 245–280 kb in size (fig. 4; McCutcheon and Moran 2007, 2010; McCutcheon et al. 2009a). The genome is perfectly syntenic with those of other *Sulcia* lineages; however, protein-coding content is dramatically reduced (figs. 4 and 5). *Sulcia*-ALF encodes 190 predicted protein-coding genes (five hypothetical and two pseudogenes), a single ribosomal RNA operon, noncoding RNA (*rnpB*), ssra RNA, and 30 tRNAs capable of recognizing all codons (fig. 5 and supplementary table S2, Supplementary Material online). In contrast, *Sulcia* of the sharpshooter leafhoppers and cicadas contain 226–228 and 246 genes, respectively (figs. 4 and 5; McCutcheon and Moran 2007; McCutcheon et al. 2009a; Woyke et al. 2010). Compared with previously sequenced *Sulcia* genomes, the *Sulcia*-ALF genome retains the same set of orthologs for the synthesis of eight essential amino acids: leucine, isoleucine, threonine, lysine, arginine, tryptophan, phenylalanine, and valine (fig. 5) (McCutcheon and Moran 2007, 2010; McCutcheon et al. 2009a; Woyke et al. 2010).

*Sulcia*-ALF has lost many genes underlying oxidative phosphorylation (OXPHOS) and DNA replication and repair (figs. 4 and 5). It lacks nearly all OXPHOS-related genes, including those encoding the complete cytochrome C oxidase complex, NADH dehydrogenase, menaquinone and ubiquinone biosynthesis proteins, and the molybdopterin oxioreductase complex that is involved in the terminal electron acceptor, DMSO reductase (Kisker et al. 1999). *Sulcia*-ALF does, however, retain genes encoding F-Type ATP synthase machinery that could function to synthesize ATP from the built-up electrochemical gradient generated by missing enzymes. In the absence of most protein complexes in the OXPHOS pathway, its function remains unclear.

Further gene losses involve DNA replication and repair (figs. 4 and 5), including genes encoding the DNA polymerase III beta (DnaN) subunit and DNA gyrase (GyrAB), which are responsible for the efficient replication of DNA by clamping the polymerase and relaxing coiling ahead of the DNA helicase (DnaB), respectively (Johanson and McHenry 1980; Sugino et al. 1980). Based on studies in model systems, these losses are expected to reduce replication efficiency (Burgers and Kornberg 1982; Smelkova and Marians 2001). Additionally, *Sulcia*-ALF retains only a single gene (*mutS*) encoding a subunit of the methyl-directed DNA mismatch repair complex, which is responsible for correcting mismatches and frameshift associated insertions and deletions in *Escherchia coli* (Schofield and Hsieh 2003; Polosina et al. 2009). The gene *mutL* is pseudogenized and is identifiable in *Sulcia*-ALF as a
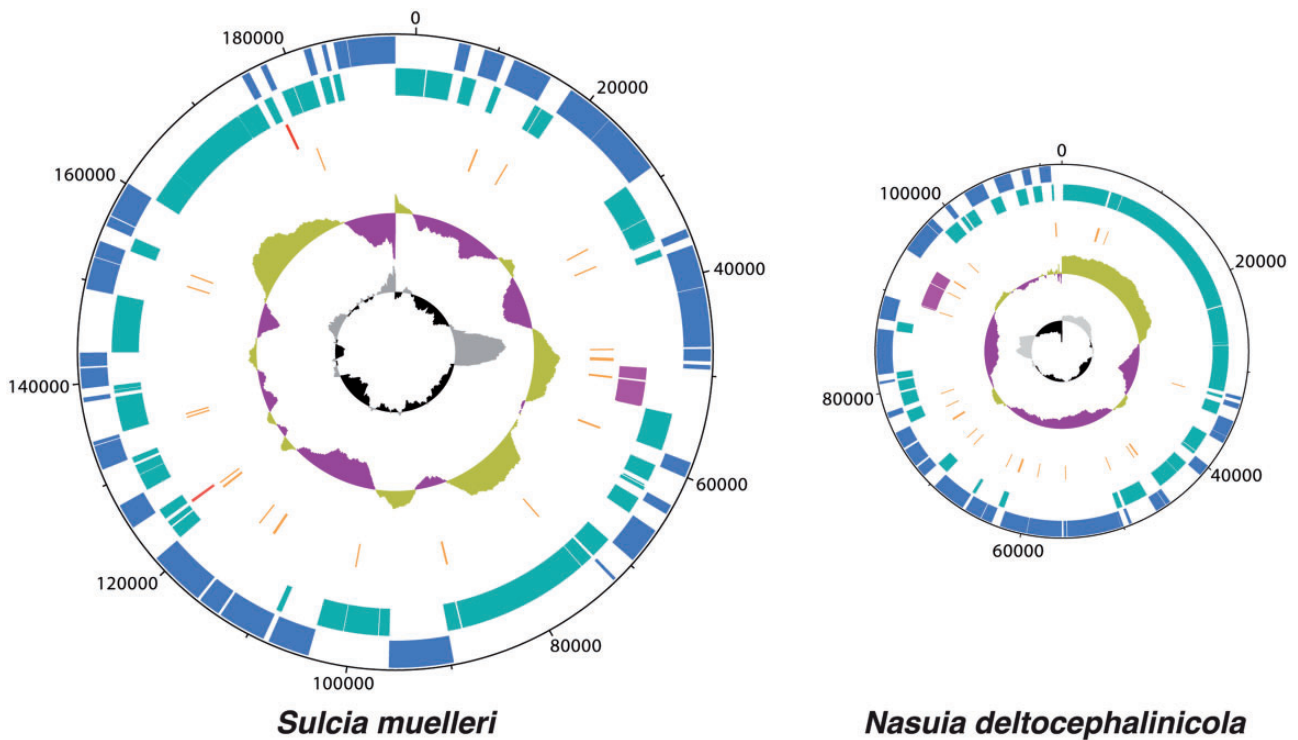
**Fig. 3.**—Genome map of the co-primary symbionts of *Macrosteles quadrilineatus*: *Sulcia*-ALF (left) and *Nasuia*-ALF (right). Genomes are scaled relative to nucleotide count. The outermost track (blue) shows genes coded on the plus strand, whereas the adjacent tracks (teal) show genes coded on the minus strand. The other tracks show RNA genes, including the rRNA operons (purple), ncRNA (red), and tRNA (orange). The two graphs show GC skew (outermost: green = values above genome average, purple = below genome average) and GC content (innermost: gray = values above genome average, black = below genome average). Plots were created with DNAPlotter (Carver et al. 2009).
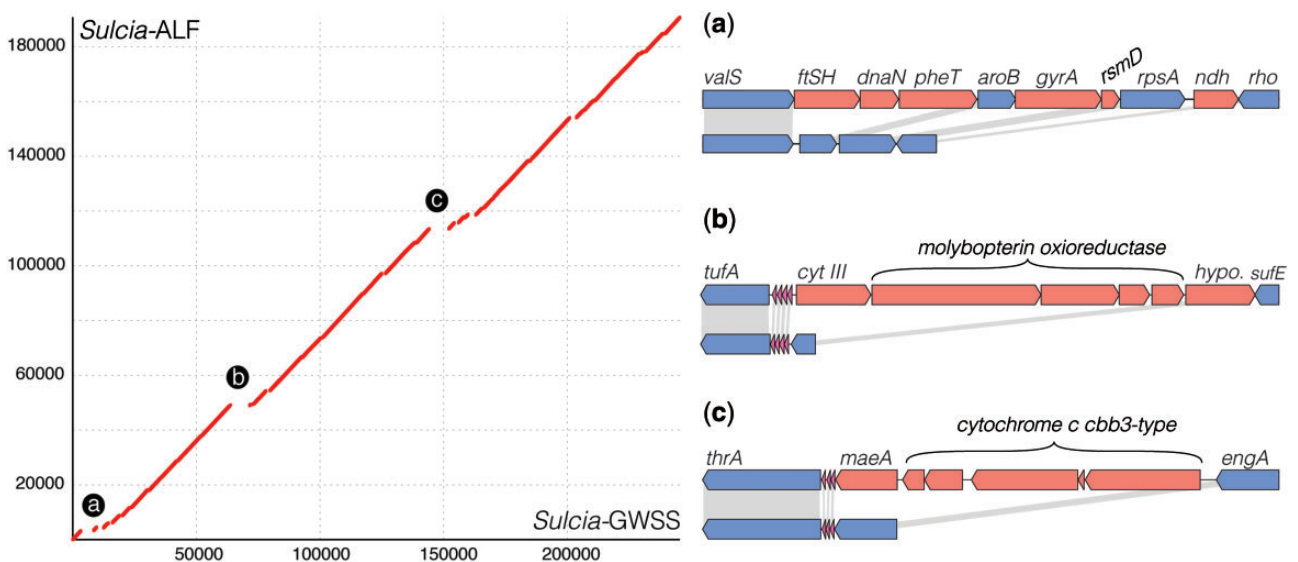


**Fig. 4.**—Synteny plot showing gene order conservation and gene loss between *Sulcia*-ALF (*Macrosteles quadrilineatus*) and *Sulcia*-GWSS (*Homalodisca coagulata*). Axes are in the number of nucleotides. Lettered breaks correspond to gene maps on the right hand side. Plots *a*–*c* show specific gene loss between *Sulcia*-GWSS (top) and *Sulcia*-ALF (bottom). Conserved orthologs are colored blue connected with gray shading, whereas genes that are lost in *Sulcia*-ALF are shown in red.
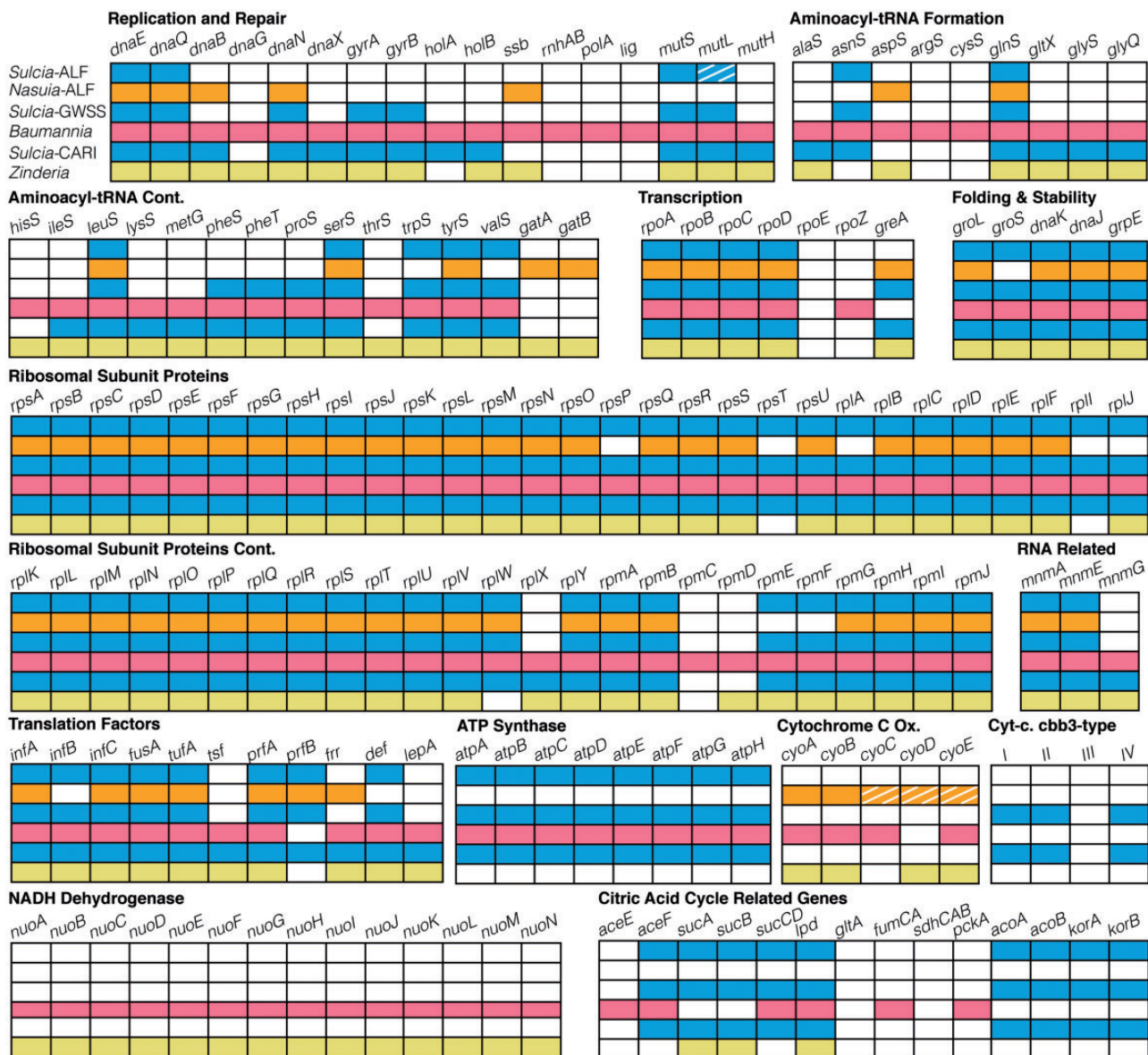
Fig. 5.—Genomic content of the co-primary symbionts from the leafhoppers, *Macrosteles quadrilineatus* (*Sulcia*-ALF and *Nasuia*-ALF) and *Homalodisca coagulata* (*Sulcia*-GWSS and *Baumannia*); and, the spittlebug, *Clastoptera arizonana* (*Sulcia*-CARI & *Zinderia*). The major pathways include DNA replication, DNA repair, translation and processing, protein stability, and energy metabolism. Colored boxes indicate genes that are present, and white boxes indicate gene absence. Striped boxes indicate genes that are present, but possibly pseudogenized, or with uncertain function.

135 bp segment, whereas homologs in other *Sulcia* lineages are 1.1 kb in length (McCutcheon et al. 2009a).

### *Nasuia*: The Functions of the Smallest Symbiont Genome

The *Nasuia*-ALF genome is 112 kb in size, 27 kb smaller than the next smallest known bacterial genome, that of *Tremblaya princeps*, a symbiont of the mealybug *P. citri* (139 kb; fig. 3; McCutcheon and von Dohlen 2011), and 30 kb smaller than the genome of *Hodgkinia cicadicola*, the symbiont that co-resides with *Sulcia* in cicadas (143 kb; McCutcheon et al.

2009b). The *Nasuia*-ALF genome encodes 137 predicted protein-coding genes, a single ribosomal operon, and 29 tRNAs predicted to be capable of recognizing all necessary codons (fig. 5 and supplementary table S3, Supplementary Material online). Fifteen predicted ORFs are unassignable to class or function. *Nasuia*-ALF contains a 1.1 kb tandem repeat that encodes part of MetF, a hypothetical protein, and a cold shock protein (CspA). Similar large genomic repeats (ranging from 150 bp to 6.1 kb) and major structural inversions have been found in other reduced symbiont genomes including those of *Carsonella* (psyllids), *Portiera* (whiteflies), and

*Tremblaya* (McCutcheon and von Dohlen 2011; Sloan and Moran 2012b, 2013).

*Nasuia*-ALF uses an alternative genetic code in which UGA is reassigned from Stop to Trp. Roughly 51% of all protein-coding genes contain at least one UGA codon. This reassignment also has been found in two other bacterial symbionts, *Zinderia* and *Hodgkinia* (McCutcheon et al. 2009b; McCutcheon and Moran 2010). *Nasuia*-ALF is further missing *prfB* encoding the peptide release RF2 that is responsible for recognizing the UGA stop codon (Capecchi 1967), as also observed in *Zinderia* and *Hodgkinia*.

*Nasuia*-ALF retains the pathways for synthesizing both histidine and methionine, as found in other symbionts that co-reside with *Sulcia* (McCutcheon and Moran 2010). *Nasuia*-ALF's histidine pathway is complete. For methionine biosynthesis, *Nasuia*-ALF uses a cobalamin-independent direct sulfhydrylation pathway as previously found for *Zinderia* (McCutcheon and Moran 2010). Both retain CysHI enzymes for the sulfhydrylation pathway and methionine synthase (MetE; Hacham et al. 2003; McCutcheon and Moran 2010). Although both symbionts retain genes encoding homoserine acetyltransferase (MetX), they appear to use two different sulfhydrylases: *Nasuia*-ALF encodes succinylhomoserine sulfhydrylase ortholog (MetB), whereas *Zinderia* uses acetyl-homoserine ortholog (MetY; McCutcheon and Moran 2010). *Nasuia*-ALF appears able to initiate acetyl-driven methionine biosynthesis initiated by MetX, but not the succinyl alternative. MetB is a versatile enzyme shown to utilize multiple homoserine-esterified intermediates, including both succinylhomoserine and acetylhomoserine (Hacham et al. 2003), and it may function to catabolize the latter. The plant-sap diet seems unlikely to furnish the missing metabolites, as plants use a phosphorylhomoserine pathway (Macnicol et al. 1981), and it has been suggested that the requisite homoserine is obtained from *Sulcia*'s threonine synthesis pathways (McCutcheon and Moran 2007).

*Nasuia*-ALF retains few other large pathways, but does retain genes underlying basic processes of replication, transcription, and translation, including genes encoding DNA and RNA polymerase subunits, RNA modification enzymes, ribosomal proteins, and some tRNA synthetases (fig. 5). It appears able to perform a limited part of the OXPHOS pathway. It encodes the machinery to synthesize and assemble a cytochrome C oxidase, including both precursors (e.g., CcsB) and assembly proteins (e.g., SCO1-like proteins), although many of these proteins have incurred AT-biased substitutions and are difficult to identify specifically, or may be pseudogenes. Finally, *Nasuia*-ALF retains genes encoding several enzymes that function in the metabolism of carbohydrate and sugars, including flavodoxin reductase (Fpr) and phosphoglycerate mutase (GpmA); however, these pathways are greatly reduced and appear incomplete. Other gene losses have left *Nasuia* without a complete set of aminoacyl tRNA synthetases (*Nasuia*-ALF has lost 14 additional genes for tRNA synthetases

relative to *Zinderia*), most TCA cycle-related genes (e.g., *sucAB* and *lpd*), and genes for several components of the DNA replication machinery (*dnaG*, *dnaX*, and *holB*).

## Phylogenomics

Phylogenetic analyses based on amino acid sequences of 42 proteins place *Zinderia* and *Nasuia* as sister lineages with high bootstrap support (BS = 100; fig. 6). Both lineages exhibit high levels of sequence divergence relative to other species of Betaproteobacteria. Both are placed within the family Oxalobacteriaceae (BS = 75). This placement was supported in previous phylogenetic results for Betaproteobacteria, including analyses based on 16S rRNA including *Zinderia* and *Nasuia* representatives (Noda et al. 2012; Koga et al. 2013) and analyses based on proteins from genome sequences including only *Zinderia* (McCutcheon and Moran 2010; Alves et al. 2013).

## Discussion

### The Origins of the Auchenorrhyncha Symbionts

*Macrosteles quadrilineatus* is associated with two obligate symbionts, *Sulcia*-ALF and *Nasuia*-ALF, as previously observed for some other members of the Deltocephalinae (Noda et al. 2012; Ishii et al. 2013). Most groups of Auchenorrhyncha are associated with *Sulcia*; however, symbionts that co-reside with *Sulcia* vary, implying multiple evolutionary transitions in major auchenorrhynchan clades (see fig. 1; Moran et al. 2003; McCutcheon et al. 2009b; Koga et al. 2013). Phylogenomic results and other biological characteristics support a close relationship of *Nasuia* with *Zinderia*, which co-resides with *Sulcia* in most cercopoid spittlebugs (Koga et al. 2013). Sequence-based phylogenetic analyses should be interpreted cautiously in the inference of bacterial symbiont relationships, as apparent relationships might reflect from a combination of convergence and long branch attraction due to the extreme AT nucleotide bias, rapid sequence evolution, and ancient divergence of symbionts relative to other bacteria (see McCutcheon and Moran 2012). Despite this, our genomic results provide additional biological support for a close relationship between *Nasuia* and *Zinderia*. First, *Nasuia* and *Zinderia* share cobalamin-independent, direct methionine synthesis pathways not found in the other co-primary symbionts of the Cicadomorpha (McCutcheon and Moran 2010). Second, and more distinctively, the reassignment of UGA observed in *Nasuia* and *Zinderia* has occurred only a few times in the history of life, including the small AT-rich genomes of mycoplasmas, several mitochondrial lineages, and *Hodgkinia* (Yamao et al. 1985; Knight et al. 2001; McCutcheon et al. 2009b; McCutcheon and Moran 2010). The reassignment is absent from most other reduced, often AT-rich genomes, such as those of *Blattabacterium*, *Blochmannia*, *Buchnera*, *Carsonella*, *Portiera*, *Uzinura*, and *Tremblaya* (Shigenobu et al. 2000; Clark et al. 2001; Gil et al. 2003; McCutcheon et al. 2009b; Sabree
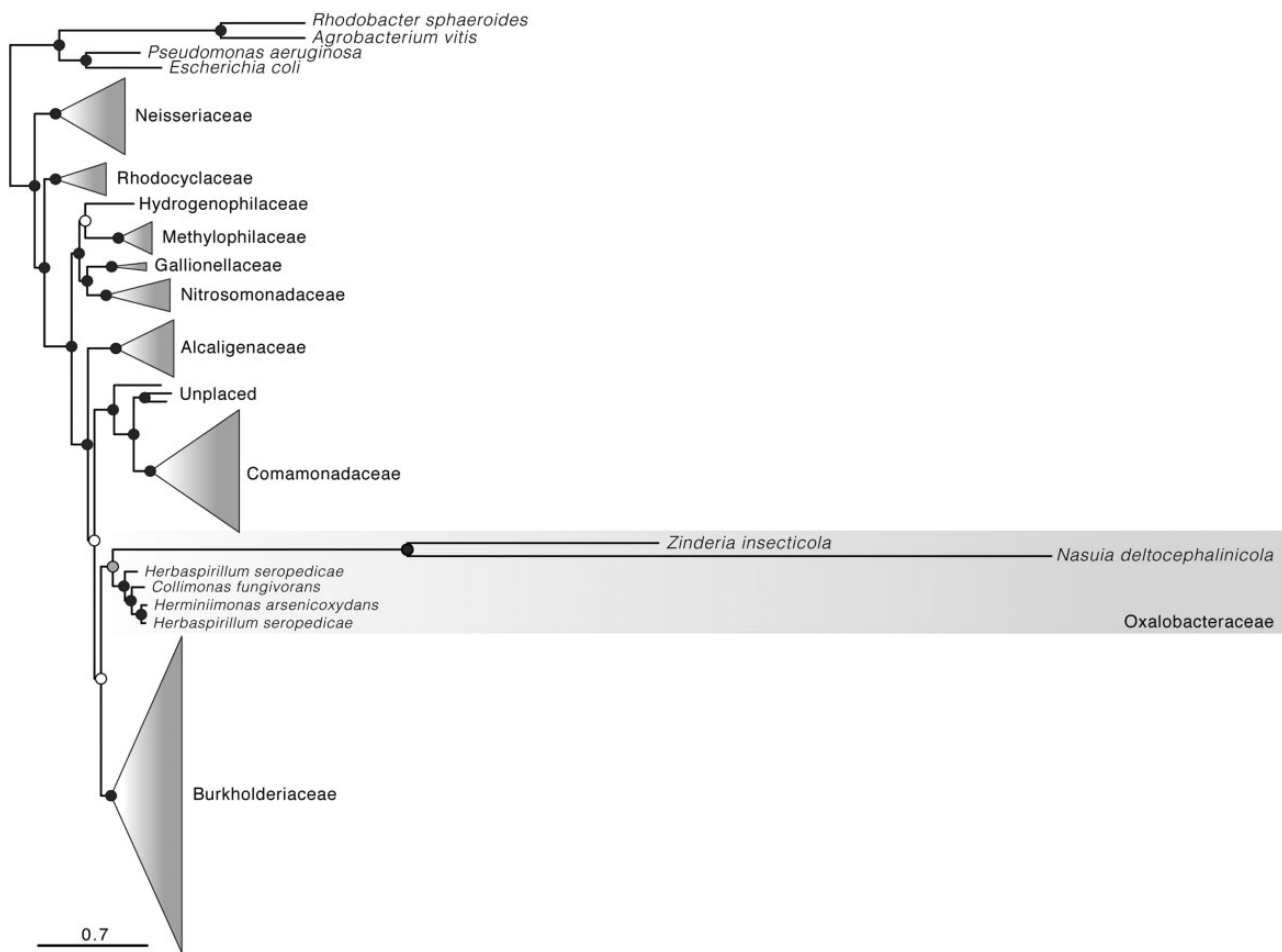
Fig. 6.—Relationships between *Nasuia deltocephalinicola* and other Betaproteobacteria. Phylogenetic tree was inferred with maximum likelihood criteria for 107 taxa for 42 proteins (approx. 10,000 amino acid columns). Circles at nodes indicate bootstrap support values: black = >75, gray = 55–75, and white = <55. See supplementary table S1, Supplementary Material online, for complete taxon sampling, and supplementary figure S2, Supplementary Material online, for full phylogeny.

et al. 2009, 2013; McCutcheon and von Dohlen 2011; Sloan and Moran 2012b).

Several hypotheses exist to explain the evolution of codon reassignment, including the complete extinction of the UGA stop codon and its reemergence as a Trp codon, along with UGG (Osawa and Jukes 1988). However, even small and AT-biased genomes contain many UGA Stop codons. An alternative hypothesis, proposed for *Hodgkinia*, is that the reassignment is triggered by an inactivating mutation in RF2 (*prfB*), which causes some peptides to be extended, but is not lethal (McCutcheon et al. 2009b). Such a mutation would resemble other deleterious mutations and gene inactivations that are fixed through genetic drift in reduced symbiont genomes. Selection would then favor replacing UGA stop codons with fully functional stop codons (UAA or UAG), freeing up UGA to bind with the tRNA-Trp anticodon. The tRNA-Trp may evolve to enhance the weak binding already present between its anticodon and UGA. Because all cases of this recoding occur

in genomes with extensive gene loss (*Nasuia*, *Zinderia*, *Hodgkinia*, *Mycoplasma*, some mitochondrial genomes), the initial loss of *prfB* as a triggering event is a potential commonality.

Our results bolster support for a symbiont clade containing *Nasuia* and *Zinderia*, and suggest that this *BetaSymb* clade represents an ancient associate of *Sulcia*, with these two symbionts co-residing in hosts at least since the origin of the Cicadomorpha (figs. 1 and 6). Preliminary phylogenetic evidence further suggests that the fulgorid symbiont, *Vidania*, also may be derived from a common ancestor shared with *Nasuia* and *Zinderia* (Urban and Cryan 2012; Koga et al. 2013). Both *Vidania* and *Zinderia* have co-diversified with their respective host groups for at least 200 million years (McCutcheon and Moran 2010; Urban and Cryan 2012). We hypothesize that *Vidania* and *Zinderia* represent subclades of *BetaSymb*, and share an ancestor that was present in the shared ancestor of the fulgoroid and cicadomorph lineages

prior to their split 260 Ma (see fig. 1; Scherbakov 2002). A genomic sequence of *Vidania*, not yet available, could provide further evidence for or against this hypothesis.

In earlier work on Auchenorrhyncha symbionts, based largely on light microscopy, Buchner (1965) and his student Müller (1940, 1962) hypothesized the existence of the symbiont later characterized with molecular data and named *Sulcia* (Moran et al. 2005). They further hypothesized companion symbionts for different auchenorrhynchan groups, but, lacking molecular evidence, did not recognize common ancestry of betaproteobacterial symbionts present in several auchenorrhynchan groups.

The hypothesis that *Sulcia* and *BetaSymb* were co-resident in the shared ancestor of Cicadomorpha implies that replacements of *BetaSymb* with alternative symbionts have occurred multiple times and at different evolutionary depths (summarized in fig. 1). These transitions include switches to *Hodgkinia* (Alphaproteobacteria) in cicadas and to the *Sodalis*-like symbiont (Gammaproteobacteria) in philaenine spittlebugs (Koga et al. 2013). Within the leafhoppers, the replacement of *Nasuia* with *Baumannia* appears to be correlated with the origin and diversification of the monophyletic xylem-feeding sharpshooters, 25–40 Ma (Moran et al. 2003; Takiya et al. 2006). *Baumannia* retains one of the largest obligate symbiont genomes to date, encoding nearly complete pathways for energy metabolism and vitamin biosynthesis (Wu et al. 2006), and contrasting with the reduced genome of *Nasuia*-ALF in other leafhoppers. Xylem feeding creates serious challenges, requiring physical and metabolic compensations to overcome negative tension and low energy content (Brodbeck et al. 1993; Zimmermann et al. 1994; Novotny and Wilson 1997). By providing new gene repertoires, the acquisition of a novel bacterial symbiont may have enabled hosts to overcome these challenges, allowing the diversification of xylem-specialized insects (Takiya et al. 2006). However, symbiont replacements and losses have sometimes also occurred in phloem-feeding lineages. For example, *Scaphoideus titanus*, another phloem feeder classified in the Deltocephalinae, lacks both *Sulcia* and *Nasuia* and instead possesses transovarially transmitted *Cardinium* (Bacteroidetes) symbionts plus a yeast-like symbiont related to those found in certain lineages of planthoppers and aphids (Sacchi et al. 2008). Some other leafhopper groups, such as the subfamily Typhlocybinae that feeds on more nutritious parenchyma plant tissues, appear to have lost their symbionts entirely (Buchner 1965; Moran et al. 2005). Nutritional aspects of these reported cases of loss or replacements have not been investigated.

## Extreme Conservation of Essential Amino Acid Synthesis

A main role of these bacterial symbionts is to provision host insects with the essential amino acids that their hosts can neither synthesize, nor obtain in sufficient quantities from their plant sap diets. In general, plant sap offers low concentrations of the essential amino acids relative to the nonessential amino acids (Brodbeck et al. 1993; Sandström and Moran 1999). Essential amino acid synthesis by the symbionts of *M. quadrilineatus* is achieved by the conserved complementation pattern previously observed in dual symbiont systems: *Sulcia*-ALF synthesizes 8 of the 10 required by the host, whereas *Nasuia*-ALF synthesizes the remaining 2 (reviewed by McCutcheon and Moran 2010). The only exception to this pattern reported so far is in spittlebugs, where *Sulcia*-CARI has additionally lost the tryptophan pathway, which is retained by *Zinderia* along with histidine and methionine synthesis pathways (McCutcheon and Moran 2010). All previous examples of this complementation in the Auchenorrhyncha were from xylem feeders (McCutcheon and Moran 2010). Our study shows that a similar pattern occurs in phloem feeders. Although phloem is far more nutritious than is xylem, having 10- to 100-fold higher concentration of free amino acids, the amino acid profile is usually unbalanced relative to animal nutritional needs, and it varies with plant species and growth stage (Sandström and Moran 1999). A high relative concentration of single amino acids can render sap noxious or unusable to insects (Brodbeck and Strong 1987), and insects have been shown to perform better on sap in which amino acids' profiles resemble experimentally determined nutritional requirements for insects (Dadd 1985).

The amino acid pathways in the symbionts of *M. quadrilineatus*, and also in the other Auchenorrhyncha, are relatively complete and do not appear to require the host or companion symbionts to contribute intermediate steps. The only proposed exception is the transfer of homoserine produced by *Sulcia* to its co-primary symbionts for biosynthesis of methionine (McCutcheon and Moran 2012). In contrast, in symbiont systems in the sister suborder, Sternorrhyncha (e.g., psyllids, whiteflies, aphids, and scale insects), some pathways are missing intermediate and terminal catabolic steps. For example, many primary symbionts in the suborder are missing the aspartate amino transferase (AspC) required for the final step in phenylalanine synthesis, and, with the exception of *Carsonella* in psyllids, they have also lost the branched-chain aminotransferase (IlvE) required to complete the synthesis of isoleucine, leucine, and valine (Shigenobu et al. 2000; Nakabachi et al. 2006; McCutcheon and von Dohlen 2011; Sloan and Moran 2012b). For the few systems for which genome sequences are available for insect hosts, host enzymes capable of completing these pathways are encoded (Wilson et al. 2010; McCutcheon and von Dohlen 2011), and these genes are upregulated in aphid bacteriocytes (Hansen and Moran 2011; Poliakov et al. 2011). In addition, the synthesis of tryptophan and/or arginine is accomplished by pathway integration between co-occurring symbionts in psyllids (Sloan and Moran 2013), the aphid *Cinara cedri* (Gosalbes et al. 2008), and mealybugs (McCutcheon and von Dohlen 2011). The fact that all sequenced symbionts of the Auchenorrhyncha, despite millions of years of divergence

and several replacements, encode relatively autonomous pathways suggests that the host and companion symbionts cannot contribute interspersed steps in the synthesis of essential amino acids. The complementary, but separate, synthesis of amino acids in the Auchenorrhyncha may reflect the ancient segregation of symbionts into different bacteriocytes (Buchner 1965; Moran et al. 2005; Koga et al. 2013).

Finally, the difference in the methionine synthesis pathways between *Zinderia* and *Nasuia* remains difficult to interpret, as only these two genomes are available for betaproteobacterial symbionts of Auchenorrhyncha and amino acid sequence identity for both loci to available sequences is low. (*Zinderia*'s MetY is 55% identical to that of *Herbaspirillum* and *Nasuia*'s MetB is 35% identical to that of *Burkholderia*.) Given the hypothesized antiquity of *Zinderia* and *Nasuia*, it is possible that their shared ancestor had both copies and that one or the other has been lost in the different lineages. Several available genomes for Betaproteobacteria indicate that multiple copies of sulfhydrylase can be present in members of Oxalobacteraceae (family in which *Zinderia* and *Nasuia* are placed) (Pedrosa et al. 2011) and the related Burkholderiaceae (Holden et al. 2004). An alternative hypothesis, that one of the sulfhydrase genes was acquired through horizontal transfer, seems less likely, as reduced symbiont genomes generally do not acquire foreign genes and the gene sequences display no signs of foreign origin.

## The Shared Loss of Energy Metabolism Genes

The most striking loss of genes from *Sulcia*-ALF and *Nasuia*-ALF involve the pathways underlying energy metabolism and particularly the OXPHOS pathway. Both symbionts retain only a single OXPHOS gene complex, and neither retain genes required to synthesize NADH dehydrogenase, which is a unique loss in *M. quadrilineatus*' symbionts compared with symbionts of previously studied auchenorrhynchan species. These losses indicate that both symbionts have a reduced capacity to synthesize ATP. Although ATP demands may be reduced, as basic cellular functions have been stripped from both symbionts, ATP is required for numerous pathways in both symbionts, raising the question of how they generate energy.

Speculatively, the loss of these functions, relative to other auchenorrhynchan symbiont systems, is related to the enriched sugars and energy-rich nature of phloem. Carbohydrate concentrations in phloem are several orders of magnitude higher (~1000×) than those in xylem (Mattson 1980). Evidence from the excreta of the xylem- and phloem-feeding insects indicates that less than 1% of xylem feeders' waste is organic carbon, but that it may be as high as 70% in the waste of phloem feeders (van Hook et al. 1980; Andersen et al. 1989). This excess of sugar in phloem sap potentially reduces the need for OXPHOS-generated ATP in symbionts, leading to the gross reduction of these pathways in the *Sulcia*-ALF and *Nasuia*-ALF genomes. As expected under this

hypothesis, the genomes of the other xylem-feeding Auchenorrhyncha do retain more complete OXPHOS pathways, including the missing NADH dehydrogenase (Wu et al. 2006; McCutcheon and Moran 2007, 2010).

In the Sternorrhyncha, OXPHOS genes are variably retained in symbiont genomes. *Buchnera* of aphids and *Portiera* of whiteflies have near-complete sets of OXPHOS-related genes, but several have lost the ATP synthase complex (e.g., Shigenobu et al. 2000; Perez-Brocal et al. 2006). Similar losses are observed in the *Tremblaya*–*Moranella* dual symbionts of the mealybug *P. citri* (McCutcheon and von Dohlen 2011). Sequenced genomes of *Carsonella* from several psyllid species reveal that they have completely lost NADH and other genes from OXPHOS protein complexes (Sloan and Moran 2012b). Sets of genes lost show some overlap between the symbiont systems of phloem-feeding Auchenorrhyncha and Sternorrhyncha, although differences are evident. Perhaps much of energy metabolism is a nonessential symbiont trait that the host is able to supplant when pathways are lost.

In order to metabolize phloem's carbohydrate-rich diet, symbionts would need to retain the genetic mechanisms to operate the glycolytic, pyruvate metabolic, and citric acid pathways. Both *Sulcia*-ALF and *Nasuia*-ALF do possess some genes for metabolizing carbohydrates; and, for *Sulcia*-CARI of cicadas, several of these (e.g., *gapA*, *lpd*, *sucAB*) are among the most highly expressed genes based on proteome analyses (McCutcheon et al. 2009a). However, these genetic repertoires in *Macrosteles*' symbionts are incomplete, calling into question whether or how well each symbiont can metabolize sugars. Possibly, the host exchanges intermediate metabolites required for carbohydrate metabolism directly with the bacteriocytes. Insects can metabolize sugars independently of their symbionts, and they maintain carbohydrate and fatty acid stores in nearby fat tissues (Kunieda et al. 2006; Arrese and Soulages 2010). In the obligate intracellular symbionts of the grain weevil, *Sitophilus zeamis* (Curculionidae), host-encoded sugar transport proteins are indeed expressed in bacteriocyte tissues (Heddi et al. 2005). It is hypothesized that these transporters allow sugars to enter bacteriocytes along a chemiosmotic gradient where they are metabolized in response to the carbohydrate-rich nature of the host insect's diet (e.g., wheat; Heddi et al. 2005).

An alternative explanation is that the host may supply its symbionts with ATP directly, or that the required missing OXPHOS genes are provided by one of the symbiotic partners. *Sulcia*-ALF retains F-type ATP synthase, which requires an electrochemical gradient to drive the synthesis of ATP (Kasimoglu et al. 1996). However, in the absence of all other members of the OXPHOS chain, it is not clear how the electrochemical gradient is achieved by *Sulcia*-ALF. Possibly, the existing OXPHOS genes are shared between co-primary symbionts because *Sulcia*-ALF and *Nasuia*-ALF encode complementary components, although how these would be integrated is not clear. Proteins or protein complexes may be shared

between the host and symbiont partners through the endoplasmic reticulum (ER), which is closely associated with host-derived membranes that encase intracellular symbionts (reviewed by Houk and Griffiths 1980). The loss and retention of single, yet complementary, genes in the OXPHOS pathway by *Sulcia*-ALF and *Nasuia*-ALF is a mystery, requiring further work to understand symbiont and host contributions to the metabolism of energy in this symbiosis.

## Supplementary Material

Supplementary figure S1 and tables S1–S3 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

## Acknowledgments

## Literature Cited

Aird D, et al. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 2:R18.

Alves JMP, et al. 2013. Genome evolution and phylogenomic analysis of *Candidatus* Kinetoplastibacterium, the betaproteobacterial endosymbiont of *Strigomonas* and *Angomonas*. Genome Biol Evol. 5:138–350.

Andersen PC, Brodbeck BV, Mizell RF. 1989. Metabolism of amino acids, organic acids and sugars extracted from the xylem fluid of four host plants by adult *Homalodisca coagulata*. Entomol Exp Appl. 50:149–159.

Arrese EL, Soulages JL. 2010. Insect fat body: energy, metabolism, and regulation. Annu Rev Entomol. 55:207–225.

Aziz RK, et al. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75.

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

Bressan A, Mulligan KL. 2013. Localization and morphological variation of three bacteriome-inhabiting symbionts within a planthopper of the genus *Oliarus* (Hemiptera: Cixiidae). Environ Microbiol Rep. 5:499–505.

Brodbeck B, Strong S. 1987. Amino acid nutrition of herbivorous insects and stress to host plants. In: Barbosa P, Schultz JC, editors. Insect outbreaks. New York: Academic Press. p. 347–364.

Brodbeck VB, Mizell RF, Andersen PC. 1993. Physiological and behavioral adaptations of three species of leafhoppers in response to the dilute nutrient content of xylem fluid. J Insect Physiol. 39:73–81.

Buchner P. 1965. Endosymbiosis of animals with plant microorganisms. New York: Wiley Interscience.

Burge SW, et al. 2012. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. 41:D226–D232.

Burgers PM, Kornberg A. 1982. ATP activation of DNA polymerase III holoenzyme of *Escherichia coli*. I. ATP-dependent formation of an initiation complex with a primed template. J Biol Chem. 10:11468–11473.

Capecchi MR. 1967. Polypeptide chain termination in vitro: isolation of a release factor. Biochemistry 58:1144–1151.

Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25:119–120.

Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)* 99:45–56.

Clark MA, Baumann L, Thao ML, Moran NA, Baumann P. 2001. Degenerative minimalism in the genome of a psyllid endosymbiont. J Bacteriol. 183:1853–1861.

Cryan JR, Urban JM. 2012. Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? Syst Entomol 37:7–21.

Dadd RH. 1985. Nutrition: organisms. In: Kerkut GA, Gilbert LT, editors. Comprehensive insect physiology, biochemistry and pharmacology. New York: Pergamon Press. p. 313–390.

Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165.

Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679.

Dietrich CH. 2003. Auchenorrhyncha (Cicadas, spittlebugs, leafhoppers, treehoppers and planthoppers). In: Resh VH, Cardé RT, editors. Encyclopedia of insects. New York: Academic Press. p. 66–74.

Douglas A. 1998. Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria *Buchnera*. Annu Rev Entomol. 43:17–37.

Drummond AJ, et al. 2010. Geneious v5.1. [cited 2013 Aug 21]. Available from: http://www.geneious.com.

Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39:W29–W37.

Frost KE, Willis DK, Groves RL. 2011. Detection and variability of aster yellows phytoplasma titer in its insect vector, *Macrosteles quadrilineatus* (Hemiptera: Cicadellidae). J Econ Entomol. 104:1800–1815.

Gil R, et al. 2003. The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. Proc Natl Acad Sci U S A. 100:9388–9393.

Gonella E, et al. 2011. Bacterial endosymbiont localization in *Hyalesthes obsoletus*, the insect vector of Bois noir in *Vitis vinifera*. Appl Environ Microbiol. 77:1423–1435.

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

Grimaldi D, Engel MS. 2005. Evolution of the insects. New York: Cambridge University Press.

Hacham Y, Gophna U, Amir R. 2003. In vivo analysis of various substrates utilized by cystathionine c-synthase and O-acetylhomoserine sulfhydrylase in methionine biosynthesis. Mol Biol Evol. 20:1513–1520.

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc Natl Acad Sci U S A. 108:2849–2854.

Heddi A, et al. 2005. Molecular and cellular profiles of insect bacteriocytes: mutualism and harm at the initial evolutionary step of symbiogenesis. Cell Microbiol. 7:293–305.

Holden MT, et al. 2004. Genomic plasticity of the causative agent of melioidosis, *Burkholderia pseudomallei*. Proc Natl Acad Sci U S A. 101:14240–14245.

Hou Y, Ma Z, Dong S, Chen Y, Yu X. 2013. Analysis of yeast-like symbiote diversity in the brown planthopper (BPH), *Nilaparvata lugens* (Stal), using a novel nested PCR-DGGE protocol. Curr Microbiol. 67:263–270.

Houk EJ, Griffiths GW. 1980. Intracellular symbiotes of the Homoptera. Ann Rev Entomol. 25:161–187.

Huson DH, Mitra S, Ruscheweyh H, Weber N, Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res. 21:1552–1560.

Ishii Y, Matsuura Y, Kakizawa S, Nikoh N, Fukatsu T. 2013. Diversity of bacterial endosymbionts associated with *Macrosteles* leafhoppers vectoring phytopathogenic phytoplasmas. Appl Environ Microbiol. 79: 5013–5022.

Johanson KO, McHenry CS. 1980. Purification and characterization of the beta-subunit of the DNA polymerase III holoenzyme of *Escherichia coli*. J Biol Chem. 22:10984–10990.

Kasimoglu E, Park SJ, Malek J, Tseng CP, Gunsalus RP. 1996. Transcriptional regulation of the proton-translocating atpase (*atpIBEFHAGDC*) operon of *Escherichia coli*: control by cell growth rate. J Bacteriol. 178: 5563–5567.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Keseler IM, et al. 2011. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res. 39:D583–D590.

Kisker C, et al. 1999. A structural comparison of molybdenum cofactor-containing enzymes. FEMS Microbiol Rev. 22:503–521.

Knight RD, Freeland SJ, Landweber LF. 2001. Rewiring the keyboard: evolvability of the genetic code. Nat Rev Genet. 2:49–58.

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, Tsuchida T, Fukatsu T. 2009. Quenching autofluorescence of insect tissues for in situ detection of endosymbionts. Appl Entomol Zool. 44:281–291.

Kunieda T, et al. 2006. Carbohydrate metabolism genes and pathways in insects: insights from the honeybee genome. Insect Mol Biol. 15: 563–576.

Kwon YJ. 1988. Taxonomic revision of the leafhopper genus *Macrosteles* Fieber of the world [PhD thesis]. Cardiff (United Kingdom): University of Wales.

Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35:3100–3108.

Le Roux JJ, Rubinoff D. 2009. Molecular data reveals California as the potential source of an invasive leafhopper species, *Macrosteles sp. nr. severini* transmitting the aster yellows phytoplasma in Hawaii. Ann Appl Biol. 154:429–439.

Li R, et al. 2009. SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966–1967.

Macnicol PK, Datko AH, Giovanelli J, Mudd SH. 1981. Homocysteine biosynthesis in green plants: physiological importance of the trans-sulfuration pathway in *Lemna paucicostata*. Plant Physiol. 68:619–625.

Markowitz VM, et al. 2009. IMG ER: a system for microbial genome annotation, expert review and curation. Bioinformatics 25:2271–2278.

Mattson WJ. 1980. Herbivory in relation to plant nitrogen content. Annu Rev Ecol Syst. 11:119–161.

McCutcheon JP, McDonald BR, Moran NA. 2009a. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc Natl Acad Sci U S A. 106:15394–15399.

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

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

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

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

McCutcheon JP, von Dohlen CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Curr Biol. 21: 1366–1372.

Moran NA. 2007. Symbiosis as an adaptive process and source of phenotypic complexity. Proc Natl Acad Sci U S A. 104:8627–8633.

Moran NA, Dale C, Dunbar H, Smith WA, Ochman H. 2003. Intracellular symbionts of sharpshooters (Insecta: Hemiptera: Cicadellinae) form a distinct clade with a small genome. Environ Microbiol. 5: 116–126.

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

Moran NA, Tran P, Gerardo NM. 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. Appl Environ Microbiol. 71: 8802–8810.

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.

Müller HJ. 1940. Die symbiose der Fulgoroiden (Homoptera: Cicadina). Zoologica 98:111–220.

Müller HJ. 1962. Neuere vorstellungen über verbreitung und phylogenie der endosymbiosen der zikaden. Zeitschrift für Morphologie und Ökologie der Tiere. 51:190–210.

Nakabachi A, et al. 2006. The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science 314:267.

Noda H, et al. 2012. Bacteriome-associated endosymbionts of the green rice leafhopper *Nephotettix cincticeps* (Hemiptera: Cicadellidae). Appl Entomol Zool. 47:217–225.

Nováková E, Hypsa V, Moran NA. 2009. *Arsenophonus*, an emerging clade of intracellular symbionts with a broad host distribution. BMC Microbiol. 9:143.

Novotny V, Wilson MR. 1997. Why are there no small species among xylem-sucking insects? Evol Ecol 11:419–437.

Osawa S, Jukes TH. 1988. Evolution of the genetic code as affected by anticodon content. Trends Genet. 4:191–198.

Pedrosa F, et al. 2011. Genome of *Herbaspirillum seropedicae* strain SmR1, a specialized diazotrophic endophyte of tropical grasses. PLoS Genet. 7:e1002064.

Perez-Brocal V, et al. 2006. A small microbial genome: the end of a long symbiotic relationship? Science 314:312–313.

Poliakov A, et al. 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol Cell Proteomics. 10: 110007039.

Polosina YY, Mui J, Pitsikas P, Cupples CG. 2009. The *Escherichia coli* mismatch repair protein MutL recruits the Vsr and MutH endonucleases in response to DNA damage. J Bacteriol. 191:4041–4043.

Sabree ZL, Huan CY, Okusu A, Moran NA, Normark BB. 2013. The nutrient supplying capabilities of *Uzinura*, an endosymbiont of armored scale insects. Environ Microbiol. 15:1988–1999.

Sabree ZL, Kambhampati S, Moran NA. 2009. Nitrogen recycling and nutritional provisioning in *Blattabacterium*, the cockroach endosymbiont. Proc Natl Acad Sci U S A. 106:19521–19526.

Sacchi L, et al. 2008. Multiple symbiosis in the leafhopper Scaphoideus titanus (Hemiptera: Cicadellidae): details of transovariol transmission of *Cardinium* and yeast-like symbionts. Tissue Cell. 40:231–242.

Sandström J, Moran NA. 1999. How nutritionally imbalanced is phloem sap for aphids? Entomol Exp Appl. 91:203–210.

Sandström J, Pettersson J. 1994. Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrthosiphon pisum*) performance. J Insect Physiol. 40:947–955.

Schattner P, Brooks A, Lowe TM. 2005. The tRNAsca-se, snoscan, snoGPS web servers for detection of tRNAs ad snoRNAs. Bioinformatics 33: w686–w689.

Scherbakov D. 2002. The 270 million year history of Auchenorrhyncha. Denisia 4:29–36.

Schofield MJ, Hsieh P. 2003. DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol. 57:579–608.

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

Shigenobu S, Wilson ACC. 2011. Genomic rebelations of mutualism: the pea aphid and its obligate bacterial symbiont. Cell Mol Life Sci. 68: 1297–1309.

Sloan DB, Bennett GB, Engel P, Williams D, Ochman H. 2013. Disentangling associated genomes. Method Enzymol. (In press).

Sloan DB, Moran NA. 2012a. Endosymbiotic bacteria as a source of carotenoids in whiteflies. Biol Lett. 8:986–989.

Sloan DB, Moran NA. 2012b. Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. Mol Biol Evol. 29:3781–3792.

Sloan DB, Moran NA. 2013. The evolution of genomic instability in the obligate endosymbionts of whiteflies. Gen Biol Evol. 5:783–793.

Smelkova N, Marians JJ. 2001. Timely release of both replication forks from *oriC* requires modulation of origin topology. J Biol Chem. 276: 39186–3919.

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

Sugino A, Higgins NP, Cozzarelli NR. 1980. DNA gyrase subunit stoichiometry and the covalent attachment of subunit A to DNA during DNA cleavage. Nucleic Acid Res. 8:3865–3874.

Takiya DM, Tran PL, Dietrich CH, Moran NA. 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. Mol Ecol. 15:4175–4191.

Tamames J, et al. 2007. The frontier between cell and organelle: genome analysis of *Candidatus Carsonella ruddii*. BMC Evol Biol. 7:181.

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

van Hook RI, Nielsen MG, Shugart HH. 1980. Energy and nitrogen relations for a *Macrosiphum liriodendri* (Homoptera: Aphididae) population in an east Tennessee *Liriodendron tulipifera* stand. Ecology 61: 960–975.

Wang Z, Wu M. 2013. A phylum-level bacterial phylogenetic marker database. Mol Biol Evol. 30:1258–1262.

Wilson AC, et al. 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. Insect Mol Biol. 19:249–258.

Woyke T, et al. 2010. One bacterial cell, one complete genome. PLoS One 5:e10314.

Wu D, et al. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. 4:e188.

Yamao F, et al. 1985. UGA is read as tryptophan in *Mycoplasma capricolum*. Proc Natl Acad Sci U S A. 82:2306–2309.

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Zimmermann U, et al. 1994. Xylem water transport: is the available evidence consistent with the cohesion theory? Plant Cell Environ. 17: 1169–1181.

**Associate editor:** John McCutcheon