Each host species is associated with the anciently acquired "degeneration proceeds, we compared the genomes of symbionts in two leafhopper species, Homalodisca vitripennis reduced genomes and have lost many genes considered to be essential in other bacteria. To understand how symbiont genome

Candidatus Baumannia cicadellinicola" (Ca).

ABSTRACT Obligate symbioses with bacteria allow insects to feed on otherwise unsuitable diets. Some symbionts have extremely reduced genomes and have lost many genes considered to be essential in other bacteria. To understand how symbiont genome degeneration proceeds, we compared the genomes of symbionts in two leafhopper species, Homalodisca vitripennis (glassy-winged sharpshooter [GWSS]) and Graphocephala atropunctata (blue-green sharpshooter [BGSS]) (Hemiptera: Cicadellidae). Each host species is associated with the recently acquired "Candidatus Sulcia muelleri" (Bacteroidetes) and the more recently acquired "Candidatus Baumannia cicadellinicola" (Gammaproteobacteria). BGSS “Ca. Baumannia” retains 89 genes that are absent from GWSS “Ca. Baumannia”; these underlie central cellular functions, including cell envelope biogenesis, cellular replication, and stress response. In contrast, “Ca. Sulcia” strains differ by only a few genes. Although GWSS “Ca. Baumannia” cells are spherical or pleomorphic (a convergent trait of obligate symbionts), electron microscopy reveals that BGSS “Ca. Baumannia” maintains a rod shape, possibly due to its retention of genes involved in cell envelope biogenesis and integrity. Phylogenomic results suggest that "Ca. Baumannia" is derived from the clade consisting of Sodalis and relatives, a group that has evolved symbiotic associations with numerous insect hosts. Finally, the rates of synonymous and nonsynonymous substitutions are higher in "Ca. Baumannia" than in "Ca. Sulcia," which may be due to a lower mutation rate in the latter. Taken together, our results suggest that the two "Ca. Baumannia" genomes represent different stages of genome reduction in which many essential functions are being lost and likely compensated by hosts. “Ca. Sulcia” exhibits much greater genome stability and slower sequence evolution, although the mechanisms underlying these differences are poorly understood.

IMPORTANCE In obligate animal-bacterial symbioses, bacteria experience extreme patterns of genome evolution, including massive gene loss and rapid evolution. However, little is known about this process, particularly in systems with complementary bacterial partners. To understand whether genome evolution impacts symbiont types equally and whether lineages follow the same evolutionary path, we sequenced the genomes of two coresident symbiotic bacteria from a plant sap-feeding insect and compared them to the symbionts from a related host species. We found that the older symbiont has a highly reduced genome with low rates of mutation and gene loss. In contrast, the younger symbiont has a larger genome that exhibits higher mutation rates and varies dramatically in the retention of genes related to cell wall biogenesis, cellular replication, and stress response. We conclude that while symbiotic bacteria evolve toward tiny genomes, this process is shaped by different selection intensities that may reflect the different ages and metabolic roles of symbiont types.

Obligate mutualisms with bacteria provide insects with the capability to feed on nutritionally deficient diets. Such symbionts are often housed within specialized host cells (bacteriocytes) and transmitted to offspring through transovarial mechanisms (1). This maternally transmitted, symbiotic lifestyle imposes a reduced effective population size that results in increased fixation of deleterious mutations, including inactivation of nonessential genes (2, 3). As a result, the genomes of some bacterial symbionts are among the smallest known, with several being <200 kb (4–8). Despite this extreme reduction, genes are retained to support nutritional requirements of the host, such as the synthesis of B vitamins or essential amino acids (EAA), depending on the insect’s diet (9–13). Among the lost genes, some correspond to bacterial functions considered to be essential, and the mechanisms compensating for these losses remain poorly understood (reviewed in reference 14).

The genome sizes vary among related symbiont strains from different host species (reviewed in reference 15). In the symbiont genus “Candidatus Blattabacterium,” found in cockroaches and primitive termites, the genome sizes vary by >50 kb due mainly to differential losses of vitamin and EAA pathways. The host switch from omnivory to wood feeding was accompanied by the acquisition of a stable gut microbiota hypothesized to provide functions previously assumed by “Ca.
Blattabacterium," potentially enabling further gene loss (16, 17). In some cases, the acquisition of a cosymbiont enables an ancestral symbiont to lose genes that encode redundant functions. For example, in the aphid Cinara cedri, Buchnera aphidicola has a genome ∼225 kb smaller than that of other Buchnera species, apparently the consequence of acquiring an additional symbiont, Serratia symbiotica (5, 18). Even more extreme genome reduction has occurred in the mealybug symbiont, “Candidatus Tremblaya princeps,” which has adopted its own intracellular bacterium (8, 19, 20). Finally, even when symbiont lineages inhabit related hosts with similar feeding habits, their genome sizes vary due to gene losses underlying essential bacterial functions (e.g., information processing, energy metabolism, and cell envelope synthesis) (15). Examples include Buchnera aphidicola in aphids, “Candidatus Carsonella ruddi” in psyllids, “Ca. Blattabacterium” species in cockroaches, and “Candidatus Portiera aleyrodidarum” in whiteflies (4, 16, 21–28). Ongoing gene losses likely change host-symbiont interactions in a lineage-specific fashion, but these patterns remain largely unexplored.

Members of the plant sap-feeding insect suborder Auchenorrhyncha (Hemiptera) have relied on at least two obligate bacterial symbionts for EAA synthesis for over 260 million years. This insect clade is broadly associated with “Candidatus Sulcia muelleri” (Bacteroidetes), which synthesizes a set of seven or eight EAAs, and one of several coprimary symbionts, which synthesize the two or three remaining EAAs (reviewed in reference 13). “Ca. Sulcia muelleri” is a highly conserved symbiont, hypothesized to be present in a shared ancestor of the Auchenorrhyncha and retained in most lineages (29, 30). It retains a central role in EAA metabolism and a conserved genomic architecture (reviewed in reference 31) but shows variations in genome size, mostly affecting capabilities for tryptophan biosynthesis, transcription/translation, and energy metabolism (7, 13). The ancestral Auchenorrhyncha also possessed a second symbiont, in the Betaproteobacteria, named “Candidatus Nasuia deltocephalinicola” in leaffoppers and “Candidatus Zinderia insecticola” in spittlebugs; this symbiont codiversified with “Ca. Sulcia muelleri” and hosts since its origin (7, 30, 32) but was lost and replaced in numerous host lineages. It was replaced by “Candidatus Hodgkinia cicadica” (Alphaproteobacteria) in cicadas, a Sodalis-like symbiont (Gammaproteobacteria) in the spittlebug subfamily Philaenini, and by “Candidatus Baumannia cicadellinicola” (Gammaproteobacteria) in leaffoppers (6, 31, 33). The acquisition of “Ca. Baumannia cicadellinicola” may have permitted leaffoppers to transition from phloem to xylem sap (11, 34). The genome of “Ca. Baumannia cicadellinicola” from Homalodisca vitripennis (glassy-winged sharpshooter [GWSS], formerly named Homalodisca coagulata) is one of the largest genomes sequenced from obligate symbionts of insects (11).

To address the question of the genomic stability of obligate symbionts living together within the same host, we sequenced complete “Ca. Baumannia cicadellinicola” and “Ca. Sulcia muelleri” genomes from Graphocephala atropunctata (Cicadellinae), commonly known as the blue-green sharpshooter (BGSS). BGSS was selected because it is distantly related (tribe: Cicadellini) to GWSS (tribe: Proconini) (34). (These symbiont strains are referred to hereafter as S-BGSS and S-GWSS for the two strains of “Ca. Sulcia muelleri” and as B-BGSS and B-GWSS for the two strains of “Ca. Baumannia cicadellinicola.”) Both insect species vector Pierce’s disease (Xylella fastidiosa) in the United States, causing millions of dollars in agricultural losses (35). A comparative genomic approach targeting the differences between the symbiotic systems of these two leaffopper species offers the opportunity to (i) discern the patterns of genome evolution in a symbiont lineage that may be at an intermediate stage of genome streamlining and (ii) elucidate the evolutionary path to highly reduced genomes.
RESULTS AND DISCUSSION

Symbiont genome characteristics for BGSS. The genome sizes for S-BGSS and B-BGSS are 244,618 and 759,425 bp, respectively (Fig. 1). Both symbionts have low GC content, which is typical in obligate symbioses: the GC content is 17% for S-BGSS and 39% for B-BGSS. B-BGSS contains 694 predicted protein-coding sequences (CDS), of which 20 encode hypothetical proteins. It contains two ribosomal cassettes, 39 tRNAs, and 10 CDS that are truncated or pseudogenized with uncertain function (see Table S2 in the supplemental material). S-BGSS has 225 CDS, with 10 encoding hypothetical proteins, a single ribosomal cassette, 31 tRNAs, and no identifiable pseudogenes (Table S3). Figure 2 shows clusters of orthologous groups (COGs) assignments for genes in both symbiont genomes from BGSS and GWSS. For B-BGSS, 69 genes could not be assigned.

S-BGSS genome and function. To date, three sequenced genomes are available for "Ca. Sulcia" in sharpshooter leafhoppers, including S-BGSS (this study), S-GWSS (12), and "Ca. Sulcia muelleri" from Draaculacephala minerva (36). The three genomes are perfectly syntenic and vary in content by only a few genes involved in the transcriptional and translational machinery (Fig. 2). S-BGSS has lost the phenylalanyl-tRNA synthetase gene (pheST), and "Ca. Sulcia muelleri" from D. minerva has lost a ribosomal subunit gene (rplJ) and the 16S methyltransferase gene (rsmD). Otherwise, the identical metabolic functions of "Ca. Sulcia muelleri" within sharpshooters have recently been reviewed (12, 36). Genes that encode tRNA synthetases are routinely purged from symbionts with tiny genomes, including the other "Ca. Sulcia muelleri" (7, 8, 27). Similarly, some ribosomal proteins are often lost from symbiont genomes, as are other rRNA processing and methylation genes (reviewed in reference 15). The extreme genomic conservation of sharpshooter "Ca. Sulcia muelleri" is striking given that its hosts diverged over 70 million years ago. This stands in stark contrast to its companion symbiont, "Ca. Baumannia cicadellinicola" (reviewed below).
Functional overview of “Ca. Baumannia cicadellinicola” strains. The B-BGSS genome is 73 kb larger than the previously sequenced genome of B-GWSS (686 kb) and retains 89 additional genes (Fig. 1 and 3). Aside from deleted regions, the “Ca. Baumannia cicadellinicola” genomes are perfectly syntenic. Unique B-BGSS genes are distributed in nearly every COG category but are enriched for functions related to cell envelope synthesis (category M; n = 18), coenzyme and inorganic metabolite synthesis and transport (categories H and P; n = 18), and cell cycle and division (category D; n = 10) (Fig. 2; see Table S2 in the supplemental material). Thus, B-BGSS appears to retain more autonomy involving capabilities in these categories (Fig. 3A to D). For example, B-BGSS retains the complete pathway for an additional ubiquinone coenzyme involved in the electron transport chain, transcriptional regulation genes (e.g., metJ, hha, cpdA, etc.), and transporters (e.g., iron [feoABC], glutamate/aspartate [gltP], and sialic acid [nanT] transporter genes). In the sections below, we compare and contrast the major metabolic differences implied by the “Ca. Baumannia cicadellinicola” genomes.

B-GWSS retains only 15 unique genes but no major pathways that are absent from B-BGSS. These differences include genes for proteins involved in translational machinery (epmAB), outer membrane (OM) synthesis and integrity (lipoprotein genes bamBE and lpp, respectively), a tricarboxylic acid (TCA) cycle enzyme (succinyl-CoA synthetase β-subunit gene [sucC]), and methionine synthesis (metAB) and three other putative genes with uncertain functions (e.g., genes for two membrane-bound pro-
teins and endonuclease). Thus, the capabilities of B-GWSS are largely a subset of those of B-BGSS.

B-BGSS additionally contains a plasmid (pB-BGSS) that is 6,564 bp in length, with GC content of 32.0% (see Fig. S1 in the supplemental material). No plasmid was reported for the B-GWSS genome sequence. pB-BGSS carries five genes that are shared with several plasmids sequenced from *B. aphidicola* in aphids (37, 38), including genes encoding two replication initiation protein variants (RepA1, RepA2), two Hsp20 small heat shock proteins (IbpA), and a highly conserved integral membrane protein (YqhA). The functional presence of two replication initiation proteins is unclear, as is the overall role of *yqhA*. IbpA is known to stabilize denatured proteins for refolding under heat stress (39). pB-BGSS encodes an additional protein, QmcA, which can suppress the activities of FtsH and HtpX (40), which are encoded on the main chromosome. FtsH and HtpX are involved in the proper folding and degradation of aberrant proteins in the membrane or misfolded proteins that result from heat stress, respectively (reviewed in reference 41). Thus, it appears that pB-BGSS may play a regulatory role in mitigating protein malfunction in *Ca. Baumannia cicadellinicola,* possibly during stress.

**Essential amino acid synthesis.** One of the most notable gene losses in B-BGSS relative to the genome of B-GWSS involves the first two steps of methionine biosynthesis—one of the few inferred requirements of *Ca. Baumannia cicadellinicola* by the leafhopper host. B-BGSS has lost *metA* and *metB* appears to have been recently pseudogenized, as it exists as several fragments with interspersed deleted regions (verified with Sanger sequencing). Together, *metAB* are responsible for the initiating steps in methionine synthesis that catalyze the reaction leading from homoserine to cystathionine. In the absence of these genes, it is unclear how methionine synthesis is initiated. Some other symbionts, including *B. aphidicola* and *Ca. Tremblaya princeps,* retain only *metE* (8, 42), underlying the final step in methionine synthesis. Recent genomic and transcriptomic data have revealed that the host might be able to initiate the first few steps of methionine synthesis from sulfate (43, 44). In both aphid and mealybug bacteria, cystathionine, cystathionine-g-lyase (see references 19 and 44). These findings add to evidence that insect hosts are likely to be intimately involved in these pathways. Whether the host can supplement methionine pathway losses remains untested for *Ca. Baumannia cicadellinicola,* but this difference provides evidence for a greater level of host-symbiont complementation in BGS than in GWSS with respect to amino acid synthesis.

**Bacterial morphology and cell envelope synthesis.** The cell shape differs markedly between the two *Ca. Baumannia cicadellinicola* symbionts: B-BGSS retains a rod-shaped cell resembling those of other *Gammaproteobacteria,* whereas B-GWSS occurs as spheroid cells (Fig. 4B and C) (29). The basis for the difference is not clear. Although both genomes retain genes encoding rod shape-determining proteins (Fig. 3B, MrdAB and MreBCD), B-BGSS retains more genetic pathways involving cell envelope integrity and cellular division, which are discussed further below. In other ancient obligate symbionts with highly reduced genomes, altered and bizarre cellular morphologies are often observed, including spheroid (*B. aphidicola*), pleomorphic (*Ca. Zinderia,* *Ca. Nasuia,* *Ca. Tremblaya princeps,* and *Candidatus Moranella endobia*), and large straplike (*Ca. Carsonella* species, *Ca. Hodgkinia* species, and *Ca. Sulcia* species) shapes (1, 8, 13, 14, 29, 45, 46). This morphological degeneration is correlated with the loss of genes that underlie the production of components of the outer and inner membranes and cell wall. For example, the obligate enlarged and amorphous cells of symbionts from mealybugs, psyllids, and most members of the Auchenorrhyncha no longer encode genes for the synthesis of peptidoglycan and phospholipids, nor do they have genes for synthesizing any membrane-associated lipids (4, 8, 12, 13). The larger genome of *B. aphidicola* in pea aphids, on the other hand, encodes genes for synthesizing cell wall components but not for the essential phospholipid membrane (42). In these cases, the host must provide a membrane, but how this functional transition evolved remains an open question.

Both B-GWSS and B-BGSS retain considerable machinery for cell envelope biogenesis, as their genomes encode the capabilities to synthesize fatty acids (*acc* and *fab* clusters), phospholipid membrane (*p55BC*), and cell wall components (*mur* cluster) (Fig. 3). However, B-BGSS retains additional capabilities to synthesize part of an outer membrane (OM). The OM is characteristic of free-living Gram-negative bacteria; it provides protection from the environment, and it is the etiological agent of septic shock in animals (reviewed in references 47 and 48). The OM is formed by an inner phospholipid leaflet and an outer lipopolysaccharide (LPS) leaflet that comprises a basal lipid A layer, lipopolysaccharide core, and surface antigens. B-BGSS but not B-GWSS encodes pathways for the partial synthesis and assembly of lipid A (*lpx* cluster and *msbA*), but both have lost nearly all genes for the synthesis of the core LPS chains (*waa* and *ipt* clusters, respectively). B-BGSS but not B-GWSS also contains a partial O antigen synthesis pathway (*rfa*AB) and a nearly complete set of genes for the assembly of an enterobacterial common antigen (ECA; *rfl* cluster, *wzy* elongation factor, and *wzx*E flippase). Bacterial antigens are typically anchored to the LPS core. Finally, both B-GWSS and B-BGSS contain genes coding for the transport and assembly of integral OM proteins (e.g., *hlpA,* encoding a periplasmic transporter, protein-folding *surA,* and OMP85 family genes [*tanaA*]). B-GWSS may still embed exogenous proteins in its OM, but the molecular origin of the OM is unknown.

The B-BGSS genome encodes additional pathways that contribute to the structure and integrity of the cell envelope. It has a nearly complete Tol-Pal system that anchors the outer and inner membranes to the peptidoglycan cell wall (Fig. 3B) (49–51). Mutants with knockouts of the Tol genes in *Erwinia chrysanthemi* (Enterobacteriaceae) can have altered bacterial cell shape (52). Thus, it is plausible that the rod shape observed in B-BGSS but not B-GWSS is linked to the retention of the Tol-Pal system in B-BGSS.

Our results suggest that the capability of *Ca. Baumannia* strains to synthesize a cell envelope is in flux. As symbionts lose the ability to produce their own cell envelope, the host presumably compensates, which may yield more direct control over the symbiont cell population size, metabolite exchange, and symbiont localization. This suggests that this ability is being ceded to the host, as appears to have already occurred in other ancient symbioses in insects (7, 12, 13, 26, 27, 53). The B-GWSS and B-BGSS genomes may provide clues to the earlier steps during this process. For example, apparent loss of the ability to synthesize the core region of LPS and the retention of the ECA are shared features between B-BGSS and the other recently established Sodalis-like endosymbionts of grain weevils and Philaenine spittlebugs (31, 54). Possibly there is early selection to denude LPS molecules because they
are endotoxic and elicit host immune responses, as is observed in the grain weevil-\textit{Sodalis} symbiosis (54, 55). The host may use the remaining surface molecules to recognize and control symbiont localization to the bacteriocyte.

\textbf{Symbiont replication and division}. The genome of B-BGSS retains several genetic pathways and genes that are involved in accomplishing cellular replication and division. Broadly, cellular replication requires initiation, chromosome replication, cell envelope expansion, and cell division (reviewed in reference 51). B-BGSS retains the capability to initiate and complete chromosome replication; in contrast, this ability has been stripped from B-GWSS (Fig. 3A) (11). These genes include \textit{dnaA}, which is responsible for initiating DNA replication, and \textit{fis} and \textit{ihfAB}, which increase the precision and stability of chromosomal replication (56, 57). All three genes function at the origin of replication (\textit{oriC}), which appears to be absent from the genomes of both B-GWSS and B-BGSS, as there is no obvious GC skew (Fig. 1). Genome scans with Ori-Finder identify five \textit{oriC} regions under liberal settings, but none are identified under more strict defaults. However, genomic Blast searches against the DoriC database did not reveal any \textit{oriC} orthologs. To date, the only other symbiont associated with the Auchenorrhyncha that contains recognizable genes for

**FIG 4** Phylogenetic relationships and transmission electron micrographs (TEM) of “\textit{Ca}. Baumannia cicadellinicola” strains. (A to C) Phylogenetic relationships of “\textit{Ca}. Baumannia cicadellinicola” from across sharpshooter leafhoppers inferred with 16S rRNA sequences from previous studies (33, 34). Colored arrows show the focal “\textit{Ca}. Baumannia cicadellinicola” strains and corresponding TEM images for B-BGSS (B) and B-GWSS (C). (D) Phylogenomic placement of the investigated “\textit{Ca}. Baumannia cicadellinicola” genomes within the \textit{Enterobacteriaceae} (\textit{Gammaproteobacteria}) and as sister to the \textit{Sodalis}-like clade (\textit{n} = 4, see Fig. S3 for specific \textit{Sodalis} species). Asterisks next to symbiont names show possible artifacts of long-branch attraction (66). \textit{Riesia pediculicola} is “\textit{Candidatus} Riesia pediculicola.”
Differential Genome Evolution of Companion Symbionts

initiating its own DNA replication is the relatively young Sodalis-like symbiont in spittlebugs (Fig. 3A) (31). Thus, the loss of these abilities is an eventual outcome in obligate symbioses.

The propagation and establishment of obligate symbionts in successive host insect generations requires symbiont cell replication and division. However, this remains one of the major essential functions typically lost from bacterial symbionts with small genomes. In this respect, B-BGSS retains vastly more genetic capabilities and autonomy than does B-GWSS (Fig. 3A). Specifically, during cell elongation and septation (splitting), a series of divisome proteins in the Fts cluster and other auxiliary proteins consecutively bind to form the septal ring at the division site (58). While both the B-GWSS and the B-BGSS genome retain the genes for the FtsZ and MinCDE proteins that initialize and localize septal ring formation (59), only B-BGSS encodes nearly all of the divisome proteins. B-GWSS has lost the genes for two early-stage proteins (FtsA and ZapA) and the entire downstream half of the pathway (Fig. 3A, FtsQBLWIN). Many of these genes are universally conserved in bacteria, and experimental evidence demonstrates that the loss of any is highly detrimental to cellular division in Escherichia coli (60). It is unlikely that B-GWSS is able to initiate and carry out cellular division on its own and, in contrast to the capabilities of B-BGSS, some other mediating source is likely required.

Bacterial cell growth requires the peptidoglycan sacculus extension, structural maintenance, and cleavage. While both the B-GWSS and the B-BGSS genome are genetically depauperate in these categories, only B-BGSS clearly retains some genes that can accomplish these tasks. B-BGSS retains the bifunctional transglycosylase/transpeptidase, mrcB, that is essential in cell growth and is one of the main peptidoglycan sacculus synthesis genes (reviewed in reference 61). Thus, it is unclear how B-GWSS accomplishes cell wall growth. Structural maintenance and integrity of the cell envelope during growth are at least partially maintained by the tol-pal pathway, which is also involved in outer membrane constriction during cell division (62). As discussed above, it is missing from the B-GWSS genome. Experimental evidence suggests that the tol-pal pathway is not essential in cell growth and viability because it is partially redundant with lpoB-mrcB functions (63). However, the latter genes are missing from B-GWSS as well, and their deletion usually results in severe cell division defects and lethality. The final cell division step is to cleave budding cells. Both the B-GWSS and the B-BGSS genome retain the gene for only one of three enzymes that cleave peptidoglycan sacculi, AmiB. Combinatorial knockouts of the missing amiA and amiC result in a high percentage of undivided E. coli cells (64). This morphological defect has not been observed in either B-GWSS or B-BGSS, indicating that either AmiB or an exogenous source efficiently cleaves cells. As an example of the latter case, the mealybug has horizontally acquired amiD from other bacteria, which is capable of cleaving peptidoglycan and is one of the most highly expressed genes in the bacteriocyte (19).

**Stress response.** Among the increased capabilities of B-BGSS is the regulation of metabolic functions under stress conditions and nutrient limitation. In particular, B-BGSS retains the spoT gene that is required for operating the stringent response (SR). The SR suppresses the transcription of proteins involved in metabolic processes, chromosome replication, cellular growth, and division when bacteria are starved for amino acids and other required elements (e.g., iron and carbon) (reviewed in reference 65). This is accomplished by modulating the levels of the alarmone, ppGpp, which in abundance acts as a transcriptional down-regulator. During normal environmental and growth conditions, spoT operates to degrade ppGpp, which is constitutively synthesized by relA and spoT depending on the nutritive conditions. Notably, B-GWSS retains relA, which indicates that it can produce ppGpp, but the regulation of its function is unknown. Possibly the host insect plays a role in controlling the levels of ppGpp. This in turn would provide the opportunity to control symbiont metabolism and growth. The retention of spoT-relA in B-BGSS, in contrast to other reduced-genome symbionts, provides further evidence that this symbiont retains an unusual degree of autonomy over its own metabolism. Conceivably, the maintenance of the SR can be beneficial to the host and symbiont, as it would render B-BGSS less metabolically active and taxing to the host during times of nutritional stress.

**Origins of “Ca. Baumannia cicadellinicola.”** To ascertain the potential origins of “Ca. Baumannia cicadellinicola” strains, we reconstructed their phylogenetic relationships within sharpshooters and across Gammaproteobacteria. The phylogenetic relationships for “Ca. Baumannia” strains inferred with 16S RNA sequences support the placement of B-BGSS with other “Ca. Baumannia cicadellinicola” strains previously sequenced from other Graphocephala lineages in the tribe Cicadellini (Fig. 4A, bootstrap support [BS] = 58 to 100) (34). The previously sequenced B-GWSS lineage is placed within the tribe Proconini (BS = 100), which is paraphyletic according to symbiont molecular data. These results largely corroborate the previous findings of Takiya et al. (34), who also demonstrated the one-to-one codiversification of “Ca. Baumannia” and sharpshooters.

Our phylogenomic results support the placement of “Ca. Baumannia cicadellinicola” in the Enterobacteriaceae (Gammaproteobacteria) (Fig. 4D, BS = 100). There is strong support for a monophyletic clade that includes “Ca. Baumannia cicadellinicola” and some other insect symbionts together with the Sodalis cluster of symbionts (Fig. 4D, BS = 100). However, this result should be treated with caution, as variations in rates of evolution and biases in base substitution patterns may confound the results. For example, the result that symbionts from disparate insect groups, such as ants and sharpshooters, derive from a single ancestor could be specious due to the high levels of divergence of these symbiont lineages and inadequate sampling of free-living diversity in the same group. Furthermore, previous work by Husnik et al. (66), using an amino acid recoding approach aimed at reducing the effects of biased molecular evolution (e.g., Dayhoff6 models) in a Bayesian framework, suggested that some taxa in the monophyletic symbiont clade resolved here (and reported in other papers, e.g., references 67 and 68) derive from independent origins; i.e., B. aphidicola in aphids from a Pantoea/Erwinia-like ancestor and “Candidatus Riesia pediculicola” from an Arsenophonus-like ancestor (Fig. 4D, asterisks). Representatives from both bacterial lineages were included in this study, but recoding efforts did not change the tree topology.

A previous phylogenomic study also suggested that “Ca. Baumannia cicadellinicola” is derived from Sodalis-like ancestors (66). Sodalis species are common insect symbionts found in a diversity of insect orders that sometimes have replaced more ancient symbionts (67, 69). The acquisition and replacement of obligate symbionts by Sodalis-like bacteria has been reported in giant scales (Llaveia spp.), weevils (Sitophilus spp.), and spittlebugs...
The long-term coprimary symbionts of some psyllids and mealybugs also appear to be derived from a *Sodalis*-like ancestor (8, 27).

**Molecular evolution in sharpshooter symbionts.** We analyzed the genomewide rates of synonymous (*dS*) and nonsynonymous (*dN*) substitutions for the two symbionts (Fig. 5; see Table S4 in the supplemental material). Because "*Ca. Baumannia cicadellinicola*" and "*Ca. Sulcia muelleri*" in these hosts diverged in synchrony with their host lineage and with each other (34), the time of divergence for B-GWSS and B-BGSS is the same as that for S-GWSS and S-BGSS, implying that differences in pairwise divergences are due to differences in rates of nucleotide base substitution. In fact, the rates are higher in "*Ca. Baumannia*:" the average pairwise identities in each symbiont’s genomes are 77.0% and 98.6% for "*Ca. Baumannia cicadellinicola*" and "*Ca. Sulcia muelleri*," respectively. The average *dS* rates are higher in "*Ca. Baumannia cicadellinicola*" (1.41) than in "*Ca. Sulcia muelleri*" (0.02). Of these, 128 "*Ca. Baumannia cicadellinicola*" genes are in saturation for *dS* (>2.0). The average *dN* is also higher in "*Ca. Baumannia cicadellinicola*" (0.13) than in "*Ca. Sulcia muelleri*" (0.005). Compared with "*Ca. Baumannia cicadellinicola*," all genes have extremely depressed rates of molecular evolution (Fig. 5; Table S4). Higher rates in "*Ca. Baumannia cicadellinicola*" than in "*Ca. Sulcia muelleri*" were noted previously for 16S rRNA sequences (34); our analysis shows that this is a genomewide difference.

The disparity in the rates of molecular evolution between symbionts is particularly enigmatic considering that "*Ca. Baumannia cicadellinicola*" retains more genes involved in DNA replication and repair than does "*Ca. Sulcia muelleri*" (DNA polymerase complex, complete mutSHL, recA, and uvrD). In *B. aphidicola*, accelerated rates of evolution are thought to reflect the loss of DNA repair enzymes (71). However, given that "*Ca. Sulcia muelleri*" has lower rates of evolution and fewer repair enzymes than

**FIG 5** Distribution of genomewide synonymous (*dS*) and nonsynonymous (*dN*) rates of molecular evolution in "*Ca. Sulcia muelleri*" (blue) and "*Ca. Baumannia cicadellinicola*" (orange). Rates are inferred from pairwise comparisons of all shared orthologs between the symbiont strains obtained from BGSS and GWSS. Adjacent box plots show median, quartiles, and maximum and minimum distributions for *dS* and *dN* values on the x- and y-axes, respectively. Box plots are color-coded according to genome.
does “Ca. Baumannia cicadellinicola,” this hypothesis does not readily explain these results. An alternative explanation that could account for the lower $dS$ in “Ca. Sulcia muelleri” is its lower number of tRNA genes, which potentially enforces more constraints on codon usage (“Ca. Baumannia cicadellinicola” has 39 tRNA genes, and “Ca. Sulcia muelleri” has 31 tRNA genes). Indeed, “Ca. Sulcia muelleri” does have stronger codon usage bias, reflecting its stronger bias in genomic base composition (as has been suggested for “Ca. Buchnera aphidicola” in aphids [72]), and potentially, this relates to tRNA retention in the genome and abundance in cells. However, a similar disparity in substitution rates is observed in the rRNA genes: the “Ca. Baumannia” 23S and 16S percent identities are 92% and 94%, respectively, while “Ca. Sulcia” 23S and 16S are both >98% identical. Thus, the elevated rates for both protein and rRNA genes in “Ca. Baumannia cicadellinicola” may reflect mechanisms associated with the fidelity of DNA replication.

**Conclusion.** The majority of symbionts of the Auchenorrhyncha have been stripped of most independent capabilities, regardless of the ecology and diet type of the host. Thus, it appears that in heritable, obligate nutritional mutualisms, all bacterial players are slated for extreme genome reduction. The genomic differences between “Ca. Baumannia cicadellinicola” and “Ca. Sulcia muelleri” strains indicate that these symbionts are at different phases in this process. In “Ca. Sulcia muelleri,” one of the oldest known bacterial symbionts, the rates of gene loss and the rates of molecular evolution are low. In “Ca. Baumannia cicadellinicola,” on the other hand, gene loss is ongoing and affects cell envelope synthesis, cell replication, and gene expression and transport. Furthermore, sequenced “Ca. Baumannia cicadellinicola” strains retain only partial metabolic pathways and many pseudogenes, suggesting a relatively recent inactivation of some functions. The piecemeal loss of these capabilities must render symbiotic bacteria increasingly dependent on the host. In contrast, “Ca. Baumannia cicadellinicola” strains (see Table S1 for full taxon sampling) (11, 33, 34). Single-locus 16S rRNA phylogenies were reconstructed to infer its placement among previously sequenced “Ca. Baumannia cicadellinicola” cell morphology was observed with transmission electron microscopy (TEM). Preliminary dissections were conducted to confirm the structure of bacteriomes and location of bacterial symbionts, following the method of Moran et al. (33). Collected tissue was preserved in 4% glutaraldehyde and 2% formaldehyde (EM grade). Dissected bacteriome tissues were embedded in 1% agarose and fixed in an aldehyde mixture containing 4% glutaraldehyde and 2% formaldehyde in 0.1 M cacodylate with 2 mM Ca and 4 mM Mg. The following steps were done with a microwave-vacuum procedure (Pelco BioWave Pro), following the manufacturer’s protocol. Samples were postfixed in reduced osmium tetroxide (2% osmium tetroxide and 2% potassium ferrocyanide in 0.1 M cacodylate buffer), washed with distilled water 5 times for 10 min each, and then dehydrated in 50%, 70%, and 95% ethyl alcohol (EtOH) rinses and 2 times in a 100% EtOH rinse, with 2% uranyl acetate in the 70% step. After two rinses in 100% acetone, the specimens were infiltrated with epoxy resin (Hard-Plus 812; Electron Microscopy Sciences) by soaking in resin-acetone mixtures of 30% and 60%, followed by two steps in 100% resin. Samples were placed in silica molds and incubated for 48 h at 60°C. Thin sections 60 to 70 nm thick were picked up on copper mesh grids and imaged with an FEI Tecnai transmission electron microscope.

**Phylogenetics and molecular evolution of sharpshooter symbionts.** Two approaches were taken to determine the phylogenetic relationships of B-BGSS. First, single-locus 16S rRNA phylogenies were reconstructed to infer its placement among previously sequenced “Ca. Baumannia cicadellinicola” strains (see Table S1 for full taxon sampling) (11, 33, 34). The sequences were aligned using Muscle (81) and reconstructed with maximum likelihood (ML) using RAxML, version 7.4.4, as described below (82).

Second, a phylogenomics approach with amino acid sequences was used to determine the placement of “Ca. Baumannia cicadellinicola” within the Gammaproteobacteria, using 176 sequenced genomes of Gammaproteobacteria. Conserved orthologous genes were identified using the Phylo_anthrope2 suite, which implements HMMER3 markercan (E value cutoff of 0.01) (83). Four outgroups were selected from the Betaproteobacteria and Alphaproteobacteria. Matrices were aligned with MAFFT, version 7.017, with the L-INS-I model (84). Redundant taxa and alignments with <50% pairwise identity or with >10% missing taxa were removed. This resulted in an alignment of 64 proteins (17,041 columns) for 180 taxa. Protест3 was used to assign each gene alignment a unique amino acid likelihood model (85). Phylogenetic trees were reconstructed using ML implemented in RAxML. Genes were concatenated into a single
matrix, partitioned by loci, and assigned an independent general time reversible (GTR)-based substitution model. ML searches were done for 100 bootstrap partitions and a final ML tree search. In an attempt to reduce the effect of compositional heterogeneity and long-branch attraction (LBA), ML searches were also run with a recoded amino acid matrix using the Dayhoff model (86). While this has been effective at breaking LBA and symbiont clustering in other analyses (66), it did not alter the tree topology in our ML analyses and it reduced the bootstrap support at internal nodes (data not shown).

To determine the dynamics of molecular evolution in these two symbiont types, we estimated the genomewide rates of synonymous (dS) and nonsynonymous (dN) substitutions. Orthologous genes were extracted and aligned using a codon-based model in TranslatorX (87). Pairwise dN/dS values were estimated with the CODEML module in PHYLML (settings: Mode = -2, CodonFreq = 2) (88). Genes for which dS was near saturation (dS > 2) or for which dN/dS was indistinguishable (e.g., dS = 0) were removed from subsequent analyses. All data matrices and batch analyses were run and parsed with custom Python scripts.

**Nucleotide sequence accession numbers.** Fully annotated genomes were submitted to GenBank under accession numbers CP008985 and CP008986 for B-BGSS and S-BGSS, respectively.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01697-14/-/DCSupplemental.

**REFERENCES**


Bennett et al.

September/October 2014 Volume 5 Issue 5 e01697-14

10 mBio mbio.asm.org


