

Phylogenomics of flavobacterial insect nutritional endosymbionts with implications for Auchenorrhyncha phylogeny

Yanghui Cao  and Christopher H. Dietrich* 

Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, IL, 61820, USA

Accepted 28 May 2021

Abstract

“*Candidatus* Sulcia muelleri” (Sulcia) is a diverse lineage of intracellular nutritional endosymbiotic bacteria strictly associated with auchenorrhynchos hemipteran insects. Sulcia has undergone long-term codiversification with its insect hosts but the phylogeny of these endosymbionts, their relationships to other bacteria, and the extent of their occurrence within various groups of Auchenorrhyncha remain inadequately explored. Comprehensive phylogenetic analyses of Sulcia and related bacteria were performed to elucidate its position relative to other members of Phylum Bacteroidetes and the degree of congruence to the phylogeny of its auchenorrhynchos hosts. Maximum likelihood (ML) and maximum parsimony (MP) analyses of *Flavobacteriales* based on genomic data from 182 bacterial strains recover a monophyletic Sulcia within a larger clade of flavobacterial insect endosymbionts, closely related to *Weeksellaceae*. Molecular divergence time analysis of Sulcia dates the origin of Sulcia at approximately 339.95 million years ago (Myr) and the initial divergence within Sulcia at approximately 256.91 Myr but these are considered underestimates due to the tendency for endosymbionts to evolve at higher rates compared to their free-living relatives. Screening of 96 recently sequenced hemipteran transcriptomes revealed that 73 of these species, all Auchenorrhyncha, harbored Sulcia. Phylogenetic analysis of 131 orthologous genes plus *16S rRNA* for 101 Sulcia strains, representing six fulgoroidean families and all the families of Cicadomorpha except Tettigarctidae, recover largely congruent phylogenies between Sulcia and Auchenorrhyncha. The phylogeny of Sulcia strongly supports the superfamily relationships Fulgoroidea + (Cicadoidea + (Cercopoidea + Membracoidea)). Relationships within individual superfamilies are also largely concordant, with the few areas of apparent incongruence between Sulcia and insect genes attributable to low branch support in one or both datasets. These results suggest that analysis of Sulcia phylogeny may contribute to resolution of contentious aspects of Auchenorrhyncha phylogeny.

© 2021 Willi Hennig Society.

Introduction

Bacterial symbionts occur in a wide range of insects (Baumann et al., 2006; Kikuchi, 2009) and often provide fitness benefits to the hosts, including provision of essential nutrients (Moran et al., 2003; McCutcheon and Moran, 2007), enhancing digestion of food (Tokuda and Watanabe, 2007; Salem et al., 2017), increasing resistance to parasitism (Oliver et al., 2008; Eleftherianos et al., 2018) and insecticides (Kikuchi et al., 2012), and improving tolerance to abiotic stresses (Montllor et al., 2002; Neelakanta et al., 2010;

Engl et al., 2018). In turn, host insects afford a stable living environment for their bacterial partners (Nováková et al., 2013). In many cases of such mutualism, endosymbionts and hosts have formed long-term associations and even mirror the phylogenies of each other (Clark et al., 2000; Lo et al., 2003; Moran et al., 2005; Gruwell et al., 2007; Liu et al., 2013). Codiversification has most often been observed in primary endosymbionts, also known as obligate endosymbionts, due to their strictly vertical transmission (Clark et al., 2010). This suggests that the phylogeny of endosymbionts could be used to elucidate the evolutionary history of their hosts (Takiya et al., 2006; Kölsch and Pedersen, 2010; Nováková et al., 2013).

*Corresponding author:

E-mail address: chdietri@illinois.edu

“*Candidatus* Sulcia muelleri” (hereafter called ‘Sulcia’) is one such group of vertically transmitted, intracellular primary endosymbiotic bacteria obligately associated with a wide range of sap-feeding hemipteran insects, including planthoppers, cicadas, spittlebugs, leafhoppers and treehoppers (Müller, 1962; Moran et al., 2005). Within the insect body, Sulcia and other primary endosymbionts inhabit specialized clusters of cells in the abdomen called bacteriomes where they exchange nutrients with the host. They are transmitted vertically, from parent to offspring, and there is no evidence that they can be horizontally transmitted. Thus, functionally, these bacteria act as specialized organelles of their host insects and their phylogeny is expected to mirror that of the hosts. Sulcia belongs to Phylum Bacteroidetes and is closely related to *Flavobacteriales* (Moran et al., 2005; McCutcheon and Moran, 2007; Matsuura et al., 2018). So far, Sulcia is only definitely recorded from the hemipteran suborder Auchenorrhyncha, with doubtful reports from a moss bug (Coleorrhyncha) and a true bug (Heteroptera) (Müller, 1962; Koohpayma et al., 2018). During the long-term association with auchenorrhynchan insects, the Sulcia genome has undergone remarkable reduction (< 0.3 Mb), but retained conserved functions for nutritional synthesis (Wu et al., 2006; McCutcheon et al., 2009). Although ability to synthesize essential amino acids (EAAs) varies among lineages, Sulcia and at least one other primary endosymbiont collaboratively provide the ten EAAs to their host insects (Bennett and Moran, 2013; Koga and Moran, 2014; Mao et al., 2017; Bennett and Mao, 2018; Matsuura et al., 2018). Moran et al. (2005) reconstructed a phylogeny of Sulcia based on 16S *rRNA* which is largely congruent with the phylogeny of Auchenorrhyncha. Based on this result, they suggested that the origin of Sulcia dated back to the origin of the common ancestor of Auchenorrhyncha (260–280 million years ago (Myr) based on fossils). Later, strict codiversification was shown to have occurred between Sulcia and several major lineages within Auchenorrhyncha, including leafhoppers (Takiya et al., 2006), planthoppers (Urban and Cryan, 2012) and cicadas (Matsuura et al., 2018; Simon et al., 2019). Therefore, this endosymbiont may provide additional evidence relevant to estimating the phylogeny of Auchenorrhyncha.

Based on microscopic observations of Sulcia (denoted “a-symbiont”) and other bacteriome-inhabiting endosymbionts, Müller (1962) suggested the phylogeny Coleorrhyncha + (Fulgoromorpha + Cicadomorpha). However, although the endosymbionts of Coleorrhyncha and Auchenorrhyncha appear similar morphologically, the presence of Sulcia in Coleorrhyncha has not been confirmed by molecular methods. A previous 16S-based phylogeny of Sulcia supported the sister

relationship between Fulgoromorpha and Cicadomorpha, and Cicadoidea as sister to Cercopoidea plus Membracoidea (Moran et al., 2005). However, the dataset included only 20 Sulcia strains representing a limited number of Auchenorrhyncha hosts and the phylogeny was based on data from a single gene region. Recent genomic studies of many organisms harboring Sulcia endosymbionts (Wu et al., 2006; McCutcheon and Moran, 2007; McCutcheon et al., 2009; McCutcheon and Moran, 2010; Woyke et al., 2010; Bennett and Moran, 2013; Bennett et al., 2014; Koga and Moran, 2014; Van Leuven et al., 2014; Campbell et al., 2015; Chang et al., 2015; Bennett et al., 2016; Jia et al., 2017; Mao et al., 2017; Ankrah et al., 2018; Bennett and Mao, 2018; Matsuura et al., 2018; Shih et al., 2019; Waneka et al., 2020) provide an ideal opportunity to conduct a comprehensive phylogenomic study of Sulcia, examine its relationships to related bacterial lineages, and rigorously test the degree to which these endosymbionts have co-diversified with their hosts. Our main questions were: (i) how is Sulcia related to free-living flavobacteria and various other lineages of flavobacterial insect endosymbionts? (ii) to what degree have Sulcia cospeciated with their auchenorrhynchan insect hosts; and (iii) do phylogenetic relationships among Sulcia help resolve contentious relationships among their host Auchenorrhyncha? To address these questions, we analyzed three large datasets: (i) a dataset comprising data from 51 single-copy genes of 182 bacterial strains was used to estimate relationships between Sulcia and *Flavobacteriales* including free-living and endosymbiont lineages; (ii) a dataset comprising 131 single-copy genes plus 16S was used to estimate relationships among Sulcia from 101 Auchenorrhyncha species representing all major lineages of this hemipteran suborder and compare the phylogeny of Sulcia with that of their Auchenorrhyncha hosts; (iii) a dataset comprising 16S sequences for 233 Sulcia and other insect endosymbionts was used to assess relationships more broadly among flavobacterial insect endosymbionts.

Methods

Retrieval of genome data and 16S *rRNA* gene

Genomes of 182 bacterial strains were downloaded from the National Center for Biotechnology Information (NCBI) assembly database to represent all the genera of *Flavobacteriales* that had genome data available as of December 8, 2020 (Table S1). For the free-living species, we chose one genome for each genus, in most cases selecting the type of strain of the type species. For the endosymbionts related to *Flavobacteriales*, all available assemblies were downloaded. Overall, the 182 genomes represent 126 free-living species of *Flavobacteriales*, 53 endosymbiont strains (Sulcia: 40; *Blattabacterium*: 11; “*Candidatus* Uzinura diaspidicola” (hereafter

Uzinura); 1; “*Candidatus* Walczuchella monophlebidarum” (hereafter Walczuchella); 1) and three species of *Bacteroidales* (outgroup). In total, 233 sequences of the bacterial *16S rRNA* gene were used in the phylogenetic analysis (accession numbers are shown in Fig. S1a, b), including 75 sequences (≥ 1400 bp) obtained from the NCBI nucleotide database, 89 retrieved from the genomes mentioned above and 69 predicted from the transcriptome assemblies of Auchenorrhyncha (see below).

Transcriptome assembly

Transcriptomic data for 96 hemipteran insects were obtained from previous studies (Tassone et al., 2017; Galetto et al., 2018; Johnson et al., 2018; Wang et al., 2019; Skinner et al., 2020) (Table 1). We reconduted *de novo* transcriptome assembly for all these data because different assembly methods may have been used for the original assemblies which may have resulted in microbial genes being discarded. Raw reads were processed using TrimmomaticPE (Bolger et al., 2014) to remove poor quality data with a minimum length of 50, leading and trailing settings of 5, slidingwindow setting of 4 : 15 and an Illuminaclip setting of 2 : 30 : 10. Trimmed reads were assembled using Trinity v2.6.6 (Haas et al., 2013) with a minimum assembled contig length of 50.

Identification of orthologues

OrthoFinder v2.5.1 (Emms and Kelly, 2019) was used to identify orthologous groups among the bacterial genomes under default parameters. To cluster the orthologous groups shared among the species of *Flavobacteriales* and related endosymbionts, we used the protein files of all downloaded genomes. Orthologues shared among Sulcia were predicted based on the genomes of 40 Sulcia strains and three outgroup species (*Blattabacterium cuenoti*, *B. clevelandi* and *B. punctulatus*). Universal or near-universal single-copy genes were selected for use in phylogenetic analysis. To increase the number of loci included we followed a relaxed strategy used in previous phylogenomic analyses (e.g., Zhan et al., 2011; Li et al., 2020) whereby absence or duplication in no more than two species was tolerated but loci duplicated in individual species were eliminated from those species in the final dataset. This screening of loci resulted in 51 orthologs of *Flavobacteriales* species and 131 orthologs of Sulcia strains, respectively.

To predict Sulcia genes within the transcriptomes of hemipteran insects, we used the full repertoires of protein sequences of four Sulcia strains (BGSS, SMAURBIH, PSPU and OLIH) and five genes from other strains as query sequences, which together represent all orthologous groups (290) shared among Sulcia genomes. TBLASTN searches were performed against the transcriptome assemblies with a cut-off E value of 10^{-5} . Candidate sequences were then used for BLASTX search against the database comprising the original query sequences and the full protein files of *Homalodisca vitripennis* and *Nilaparvata lugens*. Sequences eligible for reciprocal best BLAST hits were further confirmed through TBLASTN against the standard nucleotide databases in NCBI under default settings. Only the sequences with the best hit as a Sulcia gene and matching length ≥ 150 bp were treated as “Sulcia gene”. For phylogenomic analyses, we only chose the samples containing at least 40 of the 131 universal single-copy genes of Sulcia identified using genomic data. The final datasets of Sulcia presented 101 strains, 39 of which have genome data and the remaining samples were based on the genes retrieved from transcriptome data. We also predicted the *16S rRNA* gene of Sulcia using the *16S rRNA* of strain PSPU (NCBI accession: AP013293.1) as the query with a TBLASTN cut-off of E value 10^{-5} . Sequences with matching length ≥ 500 bp were then confirmed by a best hit of Sulcia in NCBI.

Alignment and trimming

Amino acid sequences of each orthologous group were aligned independently using PASTA v1.8.5 (Mirarab et al., 2015) under default settings. Alignments of nucleotide sequences were generated according to the amino acid alignments by a custom perl script, except the *16S rRNA* gene which was directly aligned using PASTA. To generate the concatenated data matrixes, individual alignments were concatenated first by a custom perl script then trimmed by trimAl v1.4 (Capella-Gutiérrez et al., 2009) with a gapthreshold setting of 0.3. Individual nucleotide alignments were also processed using the same trimming method before constructing gene trees. The alignment of 233 *16S* sequences was edited and trimmed manually. Aligned data used in the phylogenetic analyses are deposited in the Illinois Data Bank (https://doi.org/10.13012/B2IDB-7486289_V1).

Maximum likelihood (ML) analyses

To estimate the relationships among Sulcia and other species of *Flavobacteriales*, a concatenation analysis using the ML method was performed. Specifically, a ML tree based on the amino acid alignments of 51 single-copy genes was constructed by RAXML v8.2.8 (Stamatakis, 2014) with the “-f a” algorithm and model setting of PROTGAMMAAUTO. Branch support was evaluated using 100 bootstrap replicates. ML analysis for the *16S rRNA* gene of 233 bacterial strains was performed under the GTR + G + I model of RAXML with 1000 bootstrap replicates. Model selection was based on the best model suggested by PartitionFinder v2.1.1 (Lanfear et al., 2016). We prepared two datasets of Sulcia for ML analyses: one contained the concatenated amino acid sequences of 131 single-copy genes, and another incorporated the corresponding nucleotide alignment as well as *16S rRNA*. The nucleotide dataset was partitioned by gene using PartitionFinder v2.1.1 with *recluster_max* = 100. GTR + G + I model was determined as the best model for the majority of the partition schemes. Settings of the ML analyses were the same for *Flavobacteriales*, except that the GTR + G + I model was applied for the nucleotide dataset.

The previously sequenced Sulcia strain Neo (NCBI assembly accession: GCA_002277765.2) was excluded after the initial phylogenetic analysis due to the following observations: (i) we found a large number of duplicate genes (present in 36 orthologous groups) in Neo, which is very rare in the other strains; (ii) almost all the duplicate genes comprised two types of paralogs: with at least one copy nearly identical to the Sulcia genes from cicadas and another copy highly similar to the homologs from leafhopper-associated strains; and (iii) the two types of paralogs were never clustered, conflicting with the finding that gene duplication usually occurs in adjacent regions. Instead, the genes highly similar to the homologs from leafhoppers (also include some single-copy genes) were clustered on several scaffolds, which lack the genes highly similar to the cicada-derived Sulcia genes. Because the original study that generated the Neo strain sequenced a Sulcia strain from a leafhopper in addition to the cicada strain (Ankrah et al., 2018) we suspect that cross-contamination among samples occurred and, therefore, excluded data from this strain from further analyses.

Maximum parsimony (MP) analyses

MP analyses were performed using TNT v1.5 (Goloboff and Catalano, 2016) based on the concatenated datasets of *Flavobacteriales* (amino acid dataset) and Sulcia (nucleotide dataset) independently. All characters were weighted equally and gaps were treated as missing data. New technology searches were performed using sectorial search, ratchet, tree-drifting and tree-fusing algorithms under default settings. Ten initial random addition sequences were used and

Table 1
Prediction of Sulcia genes using transcriptomic data of hemipteran species

SRA run No.	Taxon	Number of Sulcia genes*	Number of the single-copy orthologs of Sulcia†	Presence of Sulcia 16S rRNA ‡	Reference
Coleorrhyncha					
SRR15134708	Peloridiidae: <i>Hackeriella veitchi</i>	1	0	N	Skinner et al. (2020)
SRR1821949	Peloridiidae: <i>Pelordium pomponorum</i>	0	0	N	Skinner et al. (2020)
SRR2051515	Peloridiidae: <i>Xenophyes metoponcus</i>	0	0	N	Johnson et al. (2018)
SRR921658	Peloridiidae: <i>Xenophysella greensladeae</i>	0	0	N	Misof et al. (2014)
Auchenorrhyncha					
Cercopoidea					
SRR2051468	Aphrophoridae: <i>Aphrophora alni</i>	213	114	Y	Johnson et al. (2018)
SRR1821955	Aphrophoridae: <i>Philaenus spumarius</i>	150	95	Y	Johnson et al. (2018)
SRR921578	Cercopidae: <i>Cercopis vulnerata</i>	189	107	Y	Misof et al. (2014)
SRR1821958	Cercopidae: <i>Prosapia bicincta</i>	128	78	Y	Johnson et al. (2018)
SRR2959980	Clastopteridae: <i>Clastoptera arizonana</i>	115	75	Y	Tassone et al. (2017)
SRR2496621	Clastopteridae: <i>Clastoptera obtusa</i>	183	106	Y	Skinner et al. (2020)
SRR2496653	Epipygidae: <i>Epipyga</i> sp.	236	125	Y	Skinner et al. (2020)
SRR2496648	Machaerotidae: <i>Pectinariophyes stalii</i>	195	112	Y	Skinner et al. (2020)
Cicadoidea					
SRR2496613	Cicadidae: <i>Burbunga queenslandica</i>	0	0	N	Skinner et al. (2020)
SRR2496636	Cicadidae: <i>Chilecicada</i> sp.	222	123	Y	Skinner et al. (2020)
SRR2496614	Cicadidae: <i>Guineapsaltria flava</i>	6	4	Y	Skinner et al. (2020)
SRR2496629	Cicadidae: <i>Kikihia scutellaris</i>	15	11	Y	Skinner et al. (2020)
SRR2496631	Cicadidae: <i>Maoricicada tenuis</i>	0	0	N	Skinner et al. (2020)
SRR1821973	Cicadidae: <i>Megathibicen dorsatus</i>	143	83	Y	Skinner et al. (2020)
SRR921625	Cicadidae: <i>Okanagana villosa</i>	213	117	Y	Misof et al. (2014)
SRR2496615	Cicadidae: <i>Tamasa doddi</i>	26	16	Y	Skinner et al. (2020)
SRR2496640	Cicadidae: <i>Tettigades auropilosa</i>	212	113	Y	Skinner et al. (2020)
SRR2496625	Tettigarctidae: <i>Tettigarcta crinita</i>	0	0	N	Skinner et al. (2020)
Fulgoroidea					
SRR1821891	Acanaloniidae: <i>Acanalonia conica</i>	0	0	N	Johnson et al. (2018)
SRR2496620	Achilidae: <i>Catonia nava</i>	58	44	Y	Skinner et al. (2020)
SRR1821902	Caliscelidae: <i>Bruchomorpha oculata</i>	74	53	Y	Johnson et al. (2018)
SRR2051474	Caliscelidae: <i>Caliscelis bonellii</i>	77	54	Y	Johnson et al. (2018)
SRR1821943	Cixiidae: <i>Melanoliarus placitus</i>	113	81	Y	Johnson et al. (2018)
SRR2051512	Cixiidae: <i>Tachycixius pilosus</i>	88	64	Y	Johnson et al. (2018)
SRR1821950	Delphacidae: <i>Idiosystatus acutusculus</i>	97	69	Y	Johnson et al. (2018)
SRR921622	Delphacidae: <i>Nilaparvata lugens</i>	0	0	N	Johnson et al. (2018)
SRR2496623	Derbidae: <i>Omolona uhleri</i>	0	0	N	Skinner et al. (2020)
SRR1821915	Dictyopharidae: <i>Chondrophana gavi</i>	55	41	Y	Johnson et al. (2018)
SRR2051482	Dictyopharidae: <i>Dictyophara europaea</i>	68	51	Y	Johnson et al. (2018)
SRR1821952	Dictyopharidae: <i>Phylloscelis atra</i>	59	46	Y	Johnson et al. (2018)
SRR5134743	Dictyopharidae: <i>Yucanda albida</i>	21	16	Y	Skinner et al. (2020)
SRR2496650	Eurybrachidae: <i>Platybrachys</i> sp.	44	35	Y	Skinner et al. (2020)
SRR4949629–30	Flatidae: <i>Geisha distinctissima</i>	0	0	N	Wang et al. (2019)
SRR2496657	Flatidae: <i>Jamella australiae</i>	0	0	N	Skinner et al. (2020)
SRR2051504	Flatidae: <i>Metcalfia pruinosa</i>	0	0	N	Johnson et al. (2018)
SRR1821947	Flatidae: <i>Ormenoides venusta</i>	0	0	N	Johnson et al. (2018)

Table 1
(Continued)

SRA run No.	Taxon	Number of Sulcia genes*	Number of the single-copy orthologs of Sulcia†	Presence of Sulcia 16S rRNA ‡	Reference
SRR1821913	Fulgoridae: <i>Cyrtopoptus belfragei</i>	19	13	Y	Johnson et al. (2018)
SRR4949627–28	Fulgoridae: <i>Laternaria candelaria</i>	0	0	N	Wang et al. (2019)
SRR5134712	Fulgoridae: <i>Lycorma delicatula</i>	83	58	Y	Skinner et al. (2020)
SRR5010201–2	Issidae: <i>Gergithus</i> sp.	0	0	N	Wang et al. (2019)
SRR2496635	Issidae: <i>Thionia simplex</i>	0	0	N	Skinner et al. (2020)
SRR5134723	Nogodiniidae: <i>Bladina</i> sp.	8	5	N	Skinner et al. (2020)
SRR2496646	Nogodiniidae: <i>Lipocallia australensis</i>	27	19	Y	Skinner et al. (2020)
SRR5134739	Ricaniidae: <i>Ricania speculum</i>	0	0	N	Skinner et al. (2020)
SRR2496651	Ricaniidae: <i>Scolypopa</i> sp.	0	0	N	Skinner et al. (2020)
SRR5134718	Tettigometridae: <i>Tettigometra bipunctata</i>	2	1	N	Skinner et al. (2020)
SRR1821928	Tropiduchidae: <i>Ladella</i> sp.	0	0	N	Johnson et al. (2018)
	Membracidae				
SRR5134719	Aetalionidae: <i>Aetalion reticulatum</i>	162	112	Y	Skinner et al. (2020)
SRR5134711	Aetalionidae: <i>Lophyraspis</i> sp.	185	124	Y	Skinner et al. (2020)
SRR1821894	Cicadellidae: <i>Agallia constricta</i>	46	26	Y	Johnson et al. (2018)
SRR5134720	Cicadellidae: <i>Agudus</i> sp.	139	98	Y	Skinner et al. (2020)
SRR5134722	Cicadellidae: <i>Aphrodes bicincta</i>	189	128	Y	Skinner et al. (2020)
SRR2959665	Cicadellidae: <i>Cuerna arida</i>	210	121	Y	Tassone et al. (2017)
SRR1821917	Cicadellidae: <i>Empoasca fabae</i>	0	0	N	Johnson et al. (2018)
SRR2496644	Cicadellidae: <i>Euacanthella palustris</i>	128	74	Y	Skinner et al. (2020)
SRR5816888	Cicadellidae: <i>Euscelidius variegatus</i>	176	125	Y	Galetto et al. (2018)
SRR2895344	Cicadellidae: <i>Graphocephala atropunctata</i>	131	83	Y	Tassone et al. (2017)
SRR2051489	Cicadellidae: <i>Graphocephala fennahi</i>	136	87	Y	Johnson et al. (2018)
SRR5134740	Cicadellidae: <i>Haldorus</i> sp.	63	42	Y	Skinner et al. (2020)
SRR1821922	Cicadellidae: <i>Hespenedra chilensis</i>	76	52	Y	Johnson et al. (2018)
SRR2959971	Cicadellidae: <i>Homalodisca liturata</i>	72	50	Y	Tassone et al. (2017)
SRR2496637	Cicadellidae: <i>Idiocerus rotundens</i>	139	95	Y	Skinner et al. (2020)
SRR2496638	Cicadellidae: <i>Macropsis decisa</i>	124	94	Y	Skinner et al. (2020)
SRR2496639	Cicadellidae: <i>Neocaelidia tumidifrons</i>	69	51	Y	Skinner et al. (2020)
SRR2496634	Cicadellidae: <i>Nonia palmeri</i>	186	117	Y	Skinner et al. (2020)
SRR5134736	Cicadellidae: <i>Penestragamia robusta</i>	165	117	Y	Skinner et al. (2020)
SRR5134735	Cicadellidae: <i>Penthimia</i> sp.	120	93	Y	Skinner et al. (2020)
SRR1821957	Cicadellidae: <i>Ponana quadrilaba</i>	204	121	Y	Johnson et al. (2018)
SRR2496652	Cicadellidae: <i>Stenocotis depressa</i>	156	105	Y	Skinner et al. (2020)
SRR2496641	Cicadellidae: <i>Tinobregmus viridescens</i>	90	68	Y	Skinner et al. (2020)
SRR5134733	Cicadellidae: <i>Tituria crinita</i>	15	12	Y	Skinner et al. (2020)
SRR2051513	Cicadellidae: <i>Ulopa reticulata</i>	171	114	Y	Johnson et al. (2018)
SRR1821980	Cicadellidae: <i>Viduaflava flavomaculata</i>	194	117	Y	Johnson et al. (2018)
SRR2496643	Cicadellidae: <i>Xestocephalus desertorum</i>	55	37	Y	Skinner et al. (2020)
SRR1821930	Melizoderidae: <i>Llanquihuea pilosa</i>	157	105	Y	Johnson et al. (2018)
SRR5134721	Membracidae: <i>Anastris</i> sp.	127	89	Y	Skinner et al. (2020)
SRR2051475	Membracidae: <i>Centrotus cornutus</i>	164	112	Y	Johnson et al. (2018)
SRR5134724	Membracidae: <i>Chelyoidea</i> sp.	178	118	Y	Skinner et al. (2020)
SRR5134725	Membracidae: <i>Cyphonia clavata</i>	186	124	Y	Skinner et al. (2020)
SRR1821904	Membracidae: <i>Enchenopa latipes</i>	147	100	Y	Johnson et al. (2018)

Table 1
(Continued)

SRA run No.	Taxon	Number of Sulcia genes*	Number of the single-copy orthologs of Sulcia†	Presence of Sulcia 16S rRNA ‡	Reference
SRR2496655	Membracidae: <i>Heteronotus</i> sp.	66	46	Y	Skinner et al. (2020)
SRR1821924	Membracidae: <i>Holgatiella chepuensis</i>	131	91	Y	Johnson et al. (2018)
SRR5134729	Membracidae: <i>Lycoderes burneisteri</i>	149	104	Y	Skinner et al. (2020)
SRR5134730	Membracidae: <i>Membracis tectigera</i>	157	109	Y	Skinner et al. (2020)
SRR2496633	Membracidae: <i>Microcentrus caryae</i>	95	63	Y	Skinner et al. (2020)
SRR1821942	Membracidae: <i>Nessorhinus gibberulus</i>	132	95	Y	Johnson et al. (2018)
SRR5134734	Membracidae: <i>Notocera</i> sp.	140	99	Y	Skinner et al. (2020)
SRR5134738	Membracidae: <i>Procyrtia</i> sp.	149	98	Y	Skinner et al. (2020)
SRR1821970	Membracidae: <i>Stictocephala bisonia</i>	71	52	Y	Johnson et al. (2018)
SRR5134742	Membracidae: <i>Tolania</i> sp.	106	68	Y	Skinner et al. (2020)
SRR2496642	Membracidae: <i>Umbonia crassicornis</i>	25	16	Y	Skinner et al. (2020)
SRR1821937	Myerslopiidae: <i>Mapuchea</i> sp.	156	93	Y	Johnson et al. (2018)

*Sequence with best hit of NCBI BLASTN as a Sulcia gene and at least matches 150 bp.

†Number of sequences matching the 131 universal single-copy orthologues of Sulcia.

‡Sequence with best hit of NCBI BLASTN as a 16S rRNA gene of Sulcia and at least matches 500 bp.

minimum length was found five times. To generate a strict consensus tree and evaluate the branch supports, bootstrap resampling was performed with 100 replicates using new technology search (parameter settings as mentioned above).

Multispecies coalescent analysis

We also performed multispecies coalescent analysis to construct the phylogeny of Sulcia. Gene trees of 16S rRNA and the 131 single-copy protein-coding genes were constructed using nucleotide alignments by the “-f a” algorithm of RAxML under the GTR + G + I model with 100 bootstrap replicates. Then the individual gene trees were employed to infer the coalescent-based species tree using ASTRAL v 4.10.5 (Mirarab and Warnow, 2015) with the full annotation option.

Divergence time estimation

Divergence times among *Flavobacteriaceae* lineages were estimated using r8s v1.81 (Sanderson, 2003) based on the ML phylogeny of 51 protein-coding genes. Substitution rates might be different among different lineages of *Flavobacteriales*, especially between the free-living microbes and endosymbionts. Therefore, we chose the penalized likelihood (PL) method, which permits rate variation, with the truncated Newton (TN) algorithm for optimization. The optimal value of the smoothing parameter was determined as 10 by a cross-validation analysis. Due to the absence of fossils for providing independent estimates of node ages in this group of bacteria, four nodes within groups of bacteria related, but external, to the clade comprising flavobacterial insect endosymbionts were calibrated using information obtained from the comprehensive TimeTree database (<http://www.timetree.org/>, Kumar et al., 2017): (i) the root age (split of *Bacteroidales* and *Flavobacteriales*) was fixed at 676 Myr; (ii) the crown age of *Flavobacteriaceae* was fixed at 429 Myr; (iii) the divergence time between *Luteibaculum* and *Cryomorpha* was constrained at 404 Myr; and (iv) the crown age of *Crocinitomicaceae* was constrained as 345 Myr (Marin et al., 2017).

Cophylogenetic pattern

To test the congruence between the phylogenies of Auchenorrhyncha and Sulcia, we used the best ML tree of Auchenorrhyncha constructed by Skinner et al. (2020), which is based on the degeneracy-coded nucleotide dataset of 2139 genes, and the best ML tree of Sulcia based on the nucleotide sequences of 132 genes. Auchenorrhyncha-Sulcia association patterns were represented using the R package cophyloplot. The congruence between the phylogenies of Auchenorrhyncha and Sulcia was tested using the global-fit method in PACo (Balbuena et al., 2013). Significance was assessed by the goodness-of-fit test with 100 000 permutations.

Results

Datasets

A total of 51 universal single-copy genes were identified among the 182 genomes of Bacteroidetes, the concatenated amino acid alignment of which contained 10 526 positions (8751 variable sites and 8087 parsimony-informative sites). The alignment of 16S rRNA from 233 bacterial lineages comprised 1455 positions. We identified 131 universal single-copy genes

among the genomes of 40 *Sulcia* strains and three *Blattabacterium* species, which were used to infer the phylogeny of *Sulcia* in addition to *16S rRNA*. An additional 62 *Sulcia* strains were included in the phylogenetic analysis based on data newly retrieved from insect transcriptomes. Among the 96 transcriptome datasets screened for the presence of *Sulcia* (Table 1), 76 had one or more *Sulcia* genes present. Auchenorrhynchan groups lacking *Sulcia* genes mostly occurred among the Cicadoidea (cicadas) and Fulgoroidea (planthoppers). The two concatenated datasets of 101 *Sulcia* strains contained 41 970 amino acid positions (28 491 variable sites and 24 434 parsimony-informative sites) and 127 339 nucleotide positions (74 622 variable sites and 61 154 parsimony-informative sites) respectively, both representing six planthopper families (Achilidae, Caliscelidae, Cixiidae, Delphacidae, Dictyopharidae and Fulgoridae), a cicada family (Cicadidae), five spittlebug families (Aphrophoridae, Cercopidae, Clastopteridae, Epipygidae and Machaerotidae), two leafhopper families (Cicadellidae and Myerslopiidae) and three treehopper families (Aetalionidae, Melizoderidae, Membracidae).

Phylogeny of Flavobacteriales and related endosymbionts

Maximum likelihood and maximum parsimony analyses based on the concatenated alignment of 51 protein-coding genes recovered largely concordant higher-level relationships of *Flavobacteriales* with high bootstrap values (Fig. 1 and Fig. S2). MP analysis using new technology search algorithms yielded 16 most parsimonious trees of length 174 397 (consistency index = 0.256; retention index = 0.690). The strict consensus tree generated by bootstrap resampling is presented in Fig. S2. Both ML and MP analyses grouped flavobacterial insect endosymbionts as a single clade with maximum bootstrap support. This clade is sister to the free-living bacterial group *Weeksellaceae* according to the ML topology (Fig. 1). However, MP analysis recovered a sister relationship between the clade of insect endosymbionts and a single species of *Ichthyobacteriaceae*, which together formed a sister clade to *Weeksellaceae* (Fig. S2). Within the clade of endosymbionts, three well-supported sublineages were recovered, representing the symbiotic bacteria associated with scale insects (Walczuchella and Uzinura), cockroaches (*Blattabacterium*) and Auchenorrhyncha (*Sulcia*), respectively. Surprisingly, *Sulcia* is recovered as sister to *Blattabacterium*, which conflicts with the closer relationship between the two hemipteran lineages (Auchenorrhyncha and scale insects). Within *Sulcia*, strains associated with Cicadoidea, Membracoidea and Cercopoidea were recovered as three monophyletic groups with maximum bootstrap

support. The Cicadoidea group is sister to the Membracoidea plus Cercopoidea groups.

The maximum likelihood tree based on *16S rRNA* showed relationships generally similar to the topology based on genome data but with lower bootstrap values for many branches (Fig. S1a, b). The major differences were observed among the free-living groups, which are not a focus of our study. Even with the much larger taxon sample in the *16S* dataset, flavobacterial endosymbionts of insects were still consistently recovered as a well-supported monophyletic group sister to *Weeksellaceae*, which is congruent with the ML phylogeny based on genomic data. As in the genomic dataset, we recovered *Sulcia* and *Blattabacterium* as monophyletic although the *Sulcia* clade received only moderate support. These two clades were not recovered as sister groups although the branches separating them are very short with low bootstrap values. Endosymbionts of scale insects were placed in two clades, one of which contained only “*Candidatus Brownia*”, and another one comprised Walczuchella, Uzinura and the remaining strains. This result agrees with the previous finding that flavobacterial endosymbionts of scale insects have independent origins (Rosenblueth et al., 2012). Our results also suggested multiple origins of the endosymbionts within beetles. Male-killing bacteria of ladybugs (Coccinellidae) were placed on a clade near the base and the endosymbionts of grain pest beetles formed two lineages, including a large one sister to *Sulcia* and a small one sister to *Blattabacterium*. Four large subclades were recovered within *Sulcia*, corresponding to the strains associated with Fulgoroidea, Cicadoidea, Cercopoidea and Membracoidea, respectively. As in the trees constructed from protein-coding genes (Fig. 1 and Fig. S2), the Fulgoroidea group is sister to the remaining *Sulcia* strains and the Cicadoidea clade is sister to a clade comprising symbionts of Membracoidea and Cercopoidea.

Divergence time estimates

Divergence time analysis was performed based on the ML tree constructed from 51 orthologous genes. The split between the common ancestor of flavobacterial endosymbionts and the free-living group *Weeksellaceae* was estimated at approximately 504.92 Myr (Fig. 2). The crown clade of these endosymbionts was dated approximately 395.56 Myr. Divergence between Walczuchella and Uzinura was estimated at approximately 294.33 Myr, while the split between *Blattabacterium* and *Sulcia* was dated approximately 339.95 Myr. The initial divergence of *Sulcia*, corresponding to the split between Fulgoroidea- and Cicadomorpha-associated strains was estimated at approximately 256.91 Myr. Within the Cicadomorpha-clade the origin of the Cicadoidea group was dated in the Cretaceous

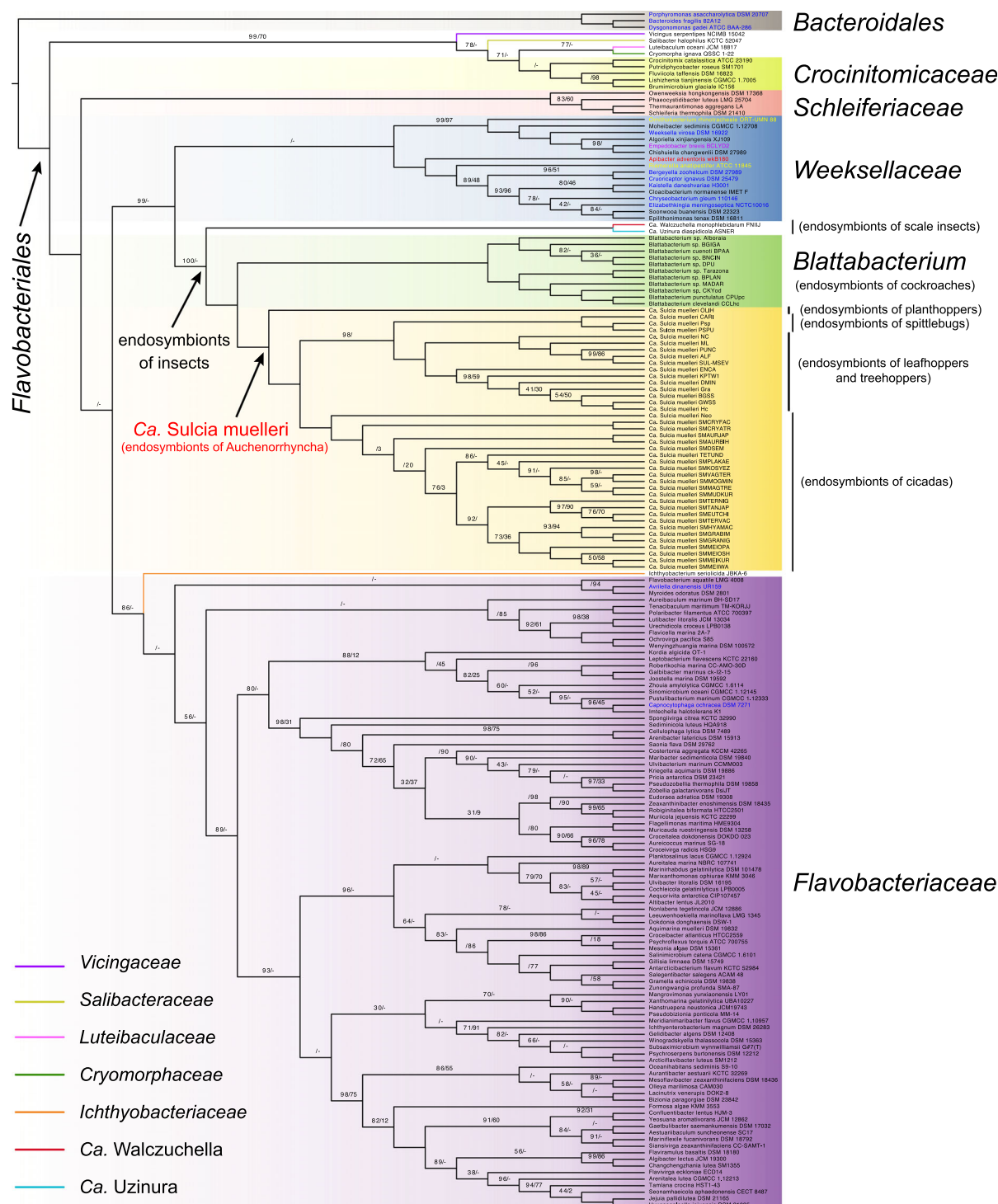


Fig. 1. Phylogeny of *Flavobacteriales* and related endosymbionts based on genome data. The topology is based on the best ML tree constructed using the amino acid sequences of 51 single-copy orthologous genes. Bootstrap values from ML and MP analyses are indicated above the branches as ML/MP (maximum support values (100) are not shown), relationships not supported by the MP analysis are indicated as “-”. Most of the free-living bacteria studied here were isolated from marine environments or fresh-water ecosystems. Free-living microbes with colored tips were not collected from aquatic environment, with blue, yellow, red and purple tips representing the species isolated from human, bird, insect and plant, respectively. Detailed information about isolation sources or hosts is listed in Table S1.

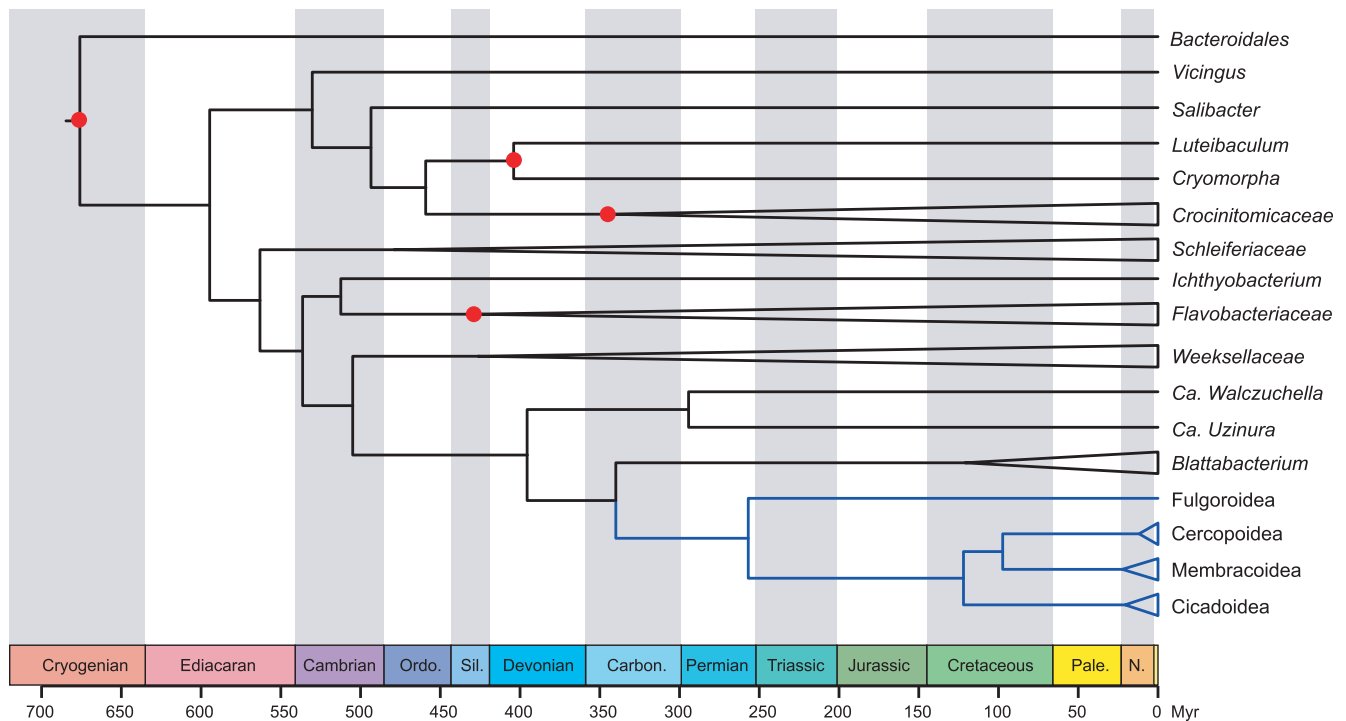


Fig. 2. Timetree chronogram of “*Ca. Sulcia muelleri*” and other *Flavobacteriales*. Red dots indicate the calibrated nodes. Hosts are used as the tip labels for the four sublineages of Sulcia (marked in navy).

(~121.95 Myr), followed by the split between the Membracoidea and Cercopoidea groups at approximately 97.37 Myr. Divergences within each major Sulcia clade (the Cicadoidea, Membracoidea and Cercopoidea groups) were dated within the Neogene.

Prediction of *Sulcia* genes based on insect transcriptome data

To identify the potential hosts of Sulcia, we screened transcriptome datasets of 96 hemipteran insects, including 92 species of Auchenorrhyncha and four species of Coleorrhyncha. As expected, Sulcia genes were present in most species of Auchenorrhyncha but nearly absent in Coleorrhyncha (Table 1). Although a short sequence (207 bp) from the moss bug *Hackeriella veitchi* matched a Sulcia gene in NCBI, many other sequences were identified as “*Candidatus Evansia muelleri*” (data not shown), the obligate endosymbiont of Coleorrhyncha (Kuechler et al., 2013). Therefore, our results only supported the association between Sulcia and Auchenorrhyncha. In Auchenorrhyncha, Sulcia protein-coding genes were retrieved from 75 species, although the number of genes predicted greatly varies among species (Table 1). *16S rRNA* of Sulcia (sequences shown in Table S2) were identified in 73 species, all of which harbor Sulcia protein-coding genes. Among these *16S rRNA* sequences, 59 were full-length *16S* and the rest were longer than 1435 bp.

We failed to predict Sulcia genes in nearly half of the planthoppers (13 of 29), most of which belong to four closely related families of “higher planthoppers”, Acanaloniidae, Issidae, Flatidae and Ricaniidae (Skinner et al., 2020). So far, only one Sulcia genome has been sequenced from a planthopper (Bennett and Mao, 2018), and it has the smallest known Sulcia genome so far (157 kb, 152 protein genes). Consistent with this prior observation, we retrieved relatively small numbers of Sulcia genes (≤ 113) in this group. In Cicadomorpha, Sulcia genes were absent in only four species, including three cicadas (*Maoricicada tenuis*, *Burbunga queenslandica* and *Tettigarcta crinita*) and one microleafhopper (*Empoasca fabae*). At least 115 protein-coding genes of Sulcia were predicted in the species of Cercopoidea. In Membracoidea, 32 of the 45 species screened contained over 100 Sulcia genes. Four species of Cicadoidea yielded large numbers of Sulcia genes (≥ 143), while the remaining species yielded no more than 30 Sulcia genes.

Phylogeny of *Sulcia* and congruence with Auchenorrhyncha phylogeny

We performed ML analyses using both amino acid (Fig. S3) and nucleotide sequences (Fig. 3 and Fig. S4), as well as a MP analysis (Fig. S5) and a multispecies coalescent analysis (Fig. S6) based on the nucleotide dataset. MP analysis using new technology

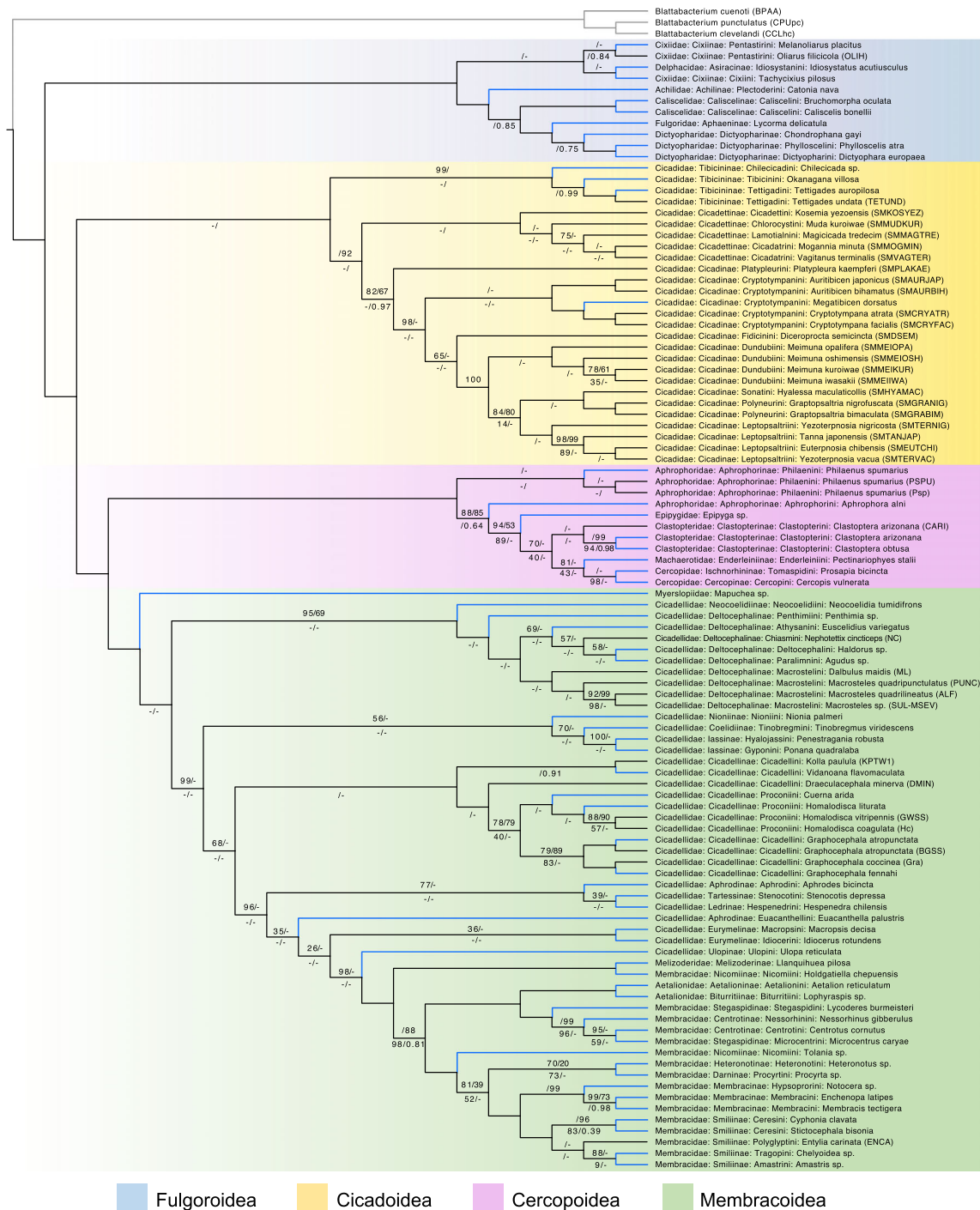


Fig. 3. Phylogeny of “*Ca. Sulcia muelleri*” based on phylogenomic analyses. The topology shows the best tree from ML analysis based on 16S *rRNA* and the nucleotide sequences of 131 protein-coding genes. Bootstrap values for ML analyses based on nucleotide (nucl) and amino acid (aa) datasets are shown above branches as nucl/aa, while support values of MP analysis and local posterior probability values of coalescent analysis are indicated below branches as MP/coalescent. Maximum support values (100 for ML and MP, 1 for coalescent) are not shown, relationships not supported by a certain analysis are indicated as “-”. The primary endosymbionts of cockroaches (*Blattabacterium*) were used as outgroups (clades in grey, labeled with bacterial species names). Terminals of *Sulcia* were labeled with host taxon and available strain names (in parentheses, using genomic data). Bacterial genes of the blue branches were retrieved from insect transcriptomes.

search algorithms yielded three most parsimonious trees of length 173 318 (consistency index = 0.648; retention index = 0.916). All analyses consistently recovered the monophyly of the four groups corresponding to the four superfamilies of Auchenorrhyncha with the relationships Fulgoroidea + (Cicadoidea + (Cercopoidea + Membracoidea) having maximum bootstrap support. In addition, relationships with high support within each major group were largely concordant among all analyses. Based on the best ML trees of Auchenorrhyncha and Sulcia (Fig. 4), the global-fit test indicated highly significant congruence between the phylogenies of Auchenorrhyncha and Sulcia ($m2 = 409.7564$, $P < 0.0001$). The few areas of possible incongruence occur in branches that received less than maximum branch support in one or both phylogenies and/or were unstable among trees resulting from different analyses (Fig. 4). For simplicity, we use host names to represent different groups of Sulcia below.

The phylogeny of Sulcia based on the large datasets recovers consistent, well-supported relationships among six families of Fulgoroidea. Cixiidae and Delphacidae form a sister clade to the remaining fulgoroids. Despite the different placements of Delphacidae in the two ML trees (Figs S3, S4), Cixiidae were consistently recovered as paraphyletic. Achilidae is sister to the clade containing Caliscelidae, Fulgoridae and Dictyopharidae. Dictyopharidae is recovered as a monophyletic group sister to Fulgoridae. Except for the paraphyly of Cixiidae, other relationships presented here are congruent with the phylogeny of the fulgoroid hosts (Fig. 4). The phylogeny based on *16S rRNA* (Fig. S1a) also supports the sister relationship between the Cixiidae-Delphacidae clade and the other fulgoroids and the paraphyly of Cixiidae. Relationships among Achilidae, Caliscelidae, Fulgoridae and Dictyopharidae are different but branches separating these families received low bootstrap support.

Within Cicadoidea, the MP tree differs from results of the other analyses in the earliest diverging branches. The topologies of ML and coalescent analyses strongly support Tibicininae as a monophyletic clade sister to Cicadettinae plus Cicadinae (Figs S3, S4, S6), which is consistent with the previous study of cicada-associated Sulcia (Matsuura et al., 2018). The phylogeny based on *16S* also places Tibicininae on earlier diverging lineages but recovered this subfamily as paraphyletic (Fig. S1b). Based on the MP analysis, however, a species of Cicadinae (*Platyleura kaempferi*) is placed as sister to the remaining Cicadidae followed by Cicadettinae as next earliest diverging clade (Fig. S5). The remaining species of Cicadinae form a sister clade to Tibicininae (Fig. S5). Despite the inconsistent subfamily-level relationships among different phylogenomic analyses, several tribes of Cicadidae are strongly supported as monophyletic, including Cicadatrini,

Cryptotympanini, Dundubiini, Leptopsaltriini, Polyneurini and Tettigadini. Within Tibicininae, Chilecadinini is sister to Tibicinini plus Tettigadini (Figs S3–S6). Cicadettini is sister to the other tribes of Cicadettinae. However, the relationships among Chlorocystini, Lamotialnini and Cicadatrini are inconsistent among analyses. Platyleurini is placed as sister to the remaining Cicadinae based on the ML and coalescent analyses (Figs S3, S4, S6). The placement of Fidicinini is unstable, either sister to Cryptotympanini (Figs S3, S5, S6) or grouped with the other tribes (Fig. S4). Dundubiini, Leptopsaltriini, Polyneurini and Sonatini are consistently placed in the same clade, with Sonatini sister to Polyneurini.

Phylogenetic relationships among cercopoids vary among analyses except that the status of Aphrophoridae is stable in the phylogenomic analyses. This family is paraphyletic, with *Philaenus spumarius* placed as sister to the remaining Cercopoidea followed by *Aphrophora alni* as next earliest diverging branch. The concatenated analyses uniformly recovered Epipygidae as sister to the clade containing Clastopteridae, Cercopidae and Machaerotidae, but placed Machaerotidae sister to either Cercopidae (Fig. S4, moderate support; Fig. S5, low support) or Clastopteridae (Fig. S3, low support). The coalescent analysis placed Cercopidae near the root of Cercopoidea and Machaerotidae sister to Epipygidae plus Clastopteridae (Fig. S6). The ML and MP analyses based on the nucleotide dataset yielded generally higher bootstrap values than the other analyses.

Within Membracoidea, Myerslopiidae is sister to the remaining taxa and Cicadellidae is paraphyletic with respect to a clade containing three treehopper families, Aetalionidae, Membracidae and Melizoderidae. Several leafhopper and treehopper subfamilies are strongly supported as monophyletic by all analyses, including Deltocephalinae, Cicadellinae, Iassinae, Membracinae and Smiliinae. The largest leafhopper subfamily Deltocephalinae is sister to Neocoelidiinae in the ML trees (Figs S3, S4), which is consistent with Dietrich et al. (2017) and some analyses of Skinner et al. (2020). However, this relationship was not recovered by the MP and coalescent analyses (Figs S5, S6). Within Cicadellinae, the tribe Cicadellini is paraphyletic with respect to the monophyletic tribe Proconiini. As in previous studies (Dietrich et al., 2017; Skinner et al., 2020), relationships among major lineages of leafhoppers were unresolved due to contradictory results among different analyses as well as very short, deep internal branches with low branch support values. The treehopper clade is sister to the leafhopper subfamily Ulopinae with high support. Treehoppers split into two sublineages: one containing Melizoderidae plus *Holdgatiella chepuensis* (Membracidae: Nicomiinae) and another one comprising Aetalionidae and the remaining Membracidae. Aetalionidae is well

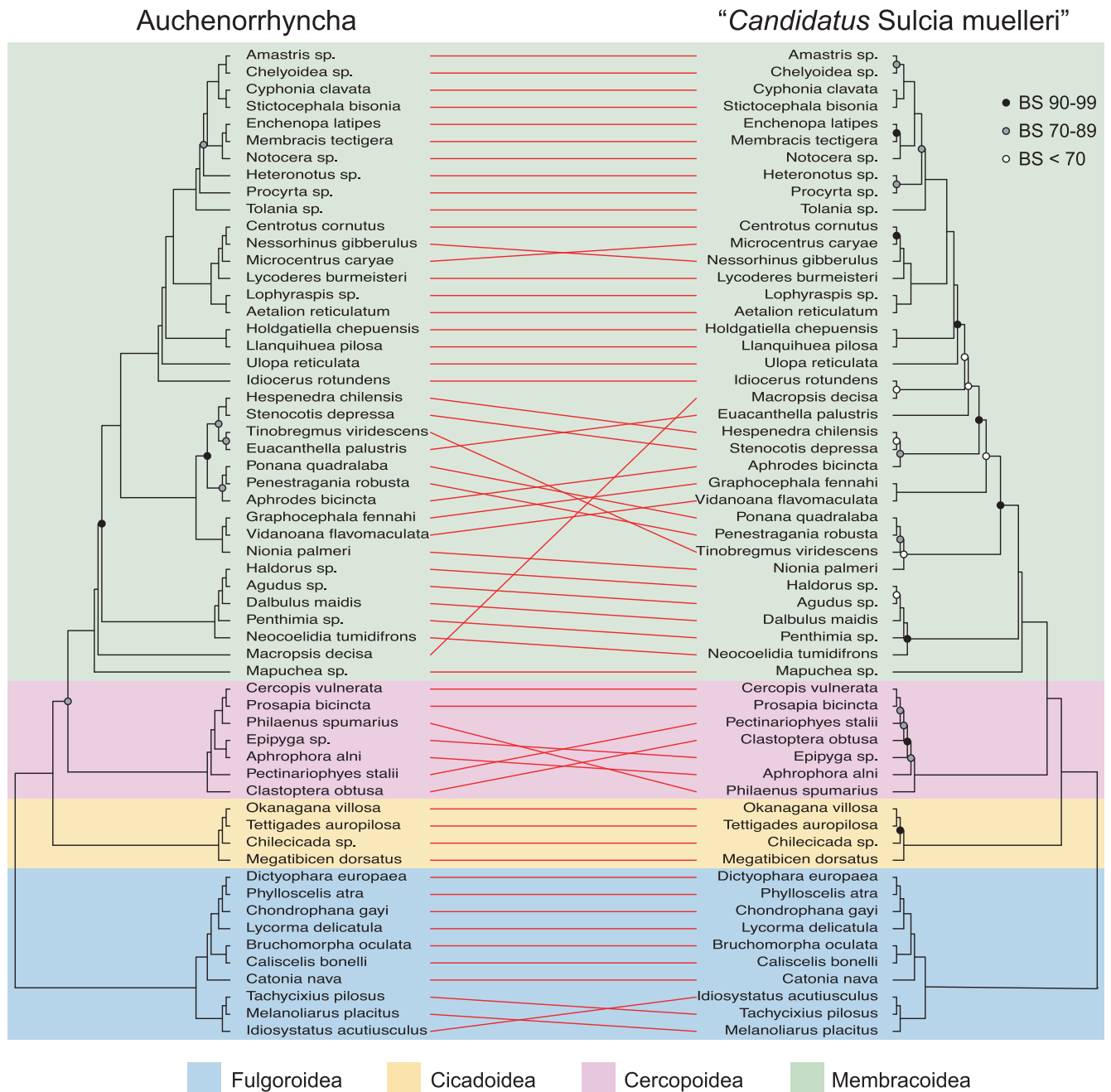


Fig. 4. Cophylogeny of "*Ca. Sulcia muelleri*" (right) and their Auchenorrhyncha hosts (left). Terminals of *Sulcia* were labeled with the name of corresponding host taxon. The phylogeny of Auchenorrhyncha is the best tree of ML analysis constructed by Skinner et al. (2020) based on the degeneracy-coded nucleotide dataset of 2139 arthropod orthologs. Bootstrap values (BS) are 100 unless indicated by dots.

supported as a monophyletic group sister to a clade comprising Stegaspidinae and Centrotinae, with the former subfamily consistently paraphyletic. Another species of Nicomiinae, *Tolania* sp., is sister to the remaining Membracidae with relationships resolved either as ((Darninae + Heteronotinae) + (Membracinae + Smiliinae)) (Figs S3–S5) or (Darninae + (Heteronotinae + (Membracinae + Smiliinae))) (Fig. S6).

Discussion

Evolution of flavobacterial endosymbionts

Sulcia was placed in *Flavobacteriaceae* by Matsuura et al. (2018) based on analysis of the *16S rRNA* sequences of *Sulcia*, some other related endosymbionts and a few species of *Flavobacteriaceae*. With the

addition of more free-living bacteria to the analyses of both 16S sequences and protein-coding genes, our results indicate that *Sulcia* does not belong to *Flavobacteriaceae*, sensu stricto, but to a large sister clade of this family within *Flavobacteriales*. Interestingly, this clade strictly comprises *Sulcia* and endosymbionts of various other insects. In the analysis of phylogenomic data, *Sulcia* is strongly supported as sister to *Blattabacterium*, a genus of cockroach endosymbionts and, together, these two clades are sister to a small clade comprising two other insect endosymbionts. Phylogenomic data are not yet available for a broader sample of flavobacterial insect endosymbionts but, according to our analysis of *16S rRNA* data for a much larger taxon sample, the lineage comprising *Blattabacterium* and *Sulcia* includes additional groups of flavobacterial endosymbionts of insects including two lineages of scale insect endosymbionts, two lineages of grain beetle endosymbionts, and a lineage of male-killing endosymbionts of Coccinellidae (Coleoptera). This entire lineage of insect endosymbionts forms a clade closely related to *Weeksellaceae*. The habitats of *Weeksellaceae* are more diverse than those of the other free-living bacteria included in this study. Although the ancestral habitat of *Flavobacteriales* is most likely to be aquatic environments, given that many free-living species were isolated from seawater, river, marine/river sediment, aquatic organisms and even sewage water (Table S1), species of *Weeksellaceae* have been isolated from a wide range of terrestrial organisms, including plants, insects, birds and humans. Therefore, it seems possible that the common ancestor of *Weeksellaceae* and the flavobacterial insect endosymbionts had already colonized the terrestrial environment, facilitating association with various terrestrial plants and animals. The monophyly of the clade comprising flavobacterial insect endosymbionts suggests that this lineage could have arisen through a single infection of the common ancestor of hemipteran insects, cockroaches and beetles. However, this would require numerous losses of endosymbionts in various descendant insect clades, and the deep phylogenetic relationships among flavobacterial endosymbionts, as reconstructed here, are inconsistent with the phylogeny of their hosts (Misof et al., 2014). Also, some bostrichid beetles harbor two different but distantly related flavobacterial endosymbionts, indicating that multiple, independent symbiont acquisitions have occurred (Engl et al., 2018). Thus, despite the grouping of most known flavobacterial endosymbionts into a single clade, available evidence suggests that such endosymbionts have infected insects multiple times independently. Further sampling and characterization of insect microbiomes are needed to explore more fully the diversity of flavobacterial endosymbionts across the Insecta.

Divergence times of Sulcia

The association between *Sulcia* and Auchenorrhyncha is most likely due to a single ancient infection, since the monophyly of Auchenorrhyncha was supported by several recent studies (Cryan and Urban, 2012; Johnson et al., 2018; Skinner et al., 2020), *Sulcia* was also recovered as monophyletic with strong support and the phylogeny of *Sulcia* is highly congruent with that of its hosts, consistent with the hypothesis that these endosymbionts and their hosts co-diversified over millions of years. Previous studies estimated the time of origin of *Sulcia* according to the age of Auchenorrhyncha. Using fossil data for auchenorrhynchan insects, Moran et al. (2005) suggested *Sulcia* diversified 260 Myr. Recent publications (Mao et al., 2017; Waneka et al., 2020) assumed an even earlier time of origin of *Sulcia* (~300 Myr). Here, we present the first explicit attempt to estimate the age of *Sulcia* and its included lineages using molecular divergence time estimation. Our timetree of *Flavobacteriales* estimates the origin of flavobacterial endosymbionts at approximately 504.92 Myr somewhat earlier than the estimated age of Insecta (~479 Myr, Misof et al., 2014). The initial divergence time of flavobacterial endosymbionts (~395.56 Myr) is comparable to the estimated divergence time between Blattodea and Hemiptera (~389 Myr, Misof et al., 2014). The first split of *Sulcia* is estimated to have occurred in the late Permian (~256.91 Myr), in agreement with the estimate of Moran et al. (2005) but this is much later than the initial divergence time of Auchenorrhyncha estimated by the divergence time analysis of insect phylogenomic data (Johnson et al., 2018). Our timetree also suggests that the three major *Sulcia* lineages (the Cicadoidea, Membracoidea and Cercopoidea groups) are much younger than their hosts. This is consistent with the results of previous studies, which have shown that the evolutionary rates of endosymbionts are generally higher than those of related free-living microbes (Itoh et al., 2002; Moran et al., 2008). In the case of *Sulcia*, it would be difficult to account for this possible rate shift in the divergence time analysis without introducing node calibrations internal to the *Sulcia* lineage. Unfortunately, the only data for such calibrations are the age estimates of the host insects, which are themselves based on very limited fossil information. Given the strong correspondence between the phylogenies of *Sulcia* and its hosts and the strictly vertical mode of transmission of these endosymbionts, it seems highly unlikely that the disparity in divergence time estimates between *Sulcia* and Auchenorrhyncha could be due to multiple independent acquisitions of *Sulcia* by different lineages of Auchenorrhyncha. More likely, the discrepancy in ages reflects a shift to a higher evolutionary rate in *Sulcia* following the transition from a free-living to endosymbiotic lifestyle.

Association between *Sulcia* and hemipteran insects

The role of *Sulcia* as a primary endosymbiont of Auchenorrhyncha is well established. Other than Auchenorrhyncha, only one moss bug and one true bug were previously reported as hosts of *Sulcia* (Müller, 1962; Koohpayma et al., 2018), but both of these records are questionable. Prior to the availability of molecular methods for characterizing bacteria, Müller (1962) grouped the endosymbiont found in a species of Peloridiidae with the most common symbiont of Auchenorrhyncha (“a-symbiont”, later identified as *Sulcia*) based on morphological similarity. Moran et al. (2005) questioned this result based on amplification of *Betaproteobacteria*/*Microbacteriaceae*-related *16S rRNA* from two species of Peloridiidae. Later, Kuechler et al. (2013) examined more species of Peloridiidae and identified the same primary endosymbiont among different samples, which they proposed as a new species of *Gammaproteobacteria* (“*Candidatus* Evansia muelleri”). Our screening of the transcriptomes of four species of Peloridiidae failed to detect the presence of *Sulcia*. Koohpayma et al. (2018) reported *Sulcia* from the true bug *Nysius cymoides* based on a *16S* sequence highly similar (99% identity, but without reporting matching length or depositing the sequence in a public database) to the *16S rRNA* of a *Sulcia* strain from the leafhopper *Orosius albicinctus*. Phylogenetic analysis placed this sequence in the clade of four Deltocephalinae leafhopper-associated strains with high support, although the sister relationship between the strains from *Nysius cymoides* and *Orosius albicinctus* was not strongly supported. Based on the phylogenetic placement and the lack of other published records of *Sulcia* from Hemiptera, we strongly suspect that the detection of *Sulcia* in *Nysius cymoides* reported by Koohpayma et al. (2018) resulted from sample contamination. Further studies of non-auchenorrhynchan Hemiptera are needed to confirm the absence of *Sulcia* or *Sulcia*-like endosymbionts in these groups.

In this study, we obtained full-length or nearly full-length *16S rRNA* sequences of *Sulcia* in 73 auchenorrhynchan species. The phylogeny based on these *Sulcia* sequences was largely congruent with the phylogeny of Auchenorrhyncha (Fig. S1a, b), indicating that our predictions of *Sulcia* from transcriptomes of particular Auchenorrhyncha species are reliable. Thus, we consider the presence of *Sulcia 16S rRNA* as strong evidence for the association between *Sulcia* and the examined species, even though the number of *Sulcia* protein-coding genes was low for some species. Our work strongly supports the presence of *Sulcia* in eight planthopper families, one cicada family and all recognized families of spittlebugs, leafhoppers and treehoppers. Including the previously reported hosts of *Sulcia*

(Müller, 1962; Urban and Cryan, 2012; Simon et al., 2019), *Sulcia* is present in all the families of Cicadomorpha and at least the following 13 families of Fulgoromorpha: Achilidae, Caliscelidae, Cixiidae, Delphacidae, Derbidae, Dictyopharidae, Eurybrachidae, Fulgoridae, Issidae, Kinnaridae, Lophopidae, Nogodinidae and Tettigometridae.

In Cicadomorpha, the loss of *Sulcia* was only observed in a few sublineages of Membracoidea, including leafhopper subfamilies Evacanthinae, Ledrinae, Typhlocybinae and the treehopper tribe Terentiini (Müller, 1962; Moran et al., 2005). We failed to detect *Sulcia* genes in three cicadas, *M. tenuis*, *B. queenslandica* and *T. crinita*, as well as in the microleafhopper *E. fabae*. However, both *T. crinita* and *E. fabae* were reported as *Sulcia* hosts in previous studies (DeLay, 2013; Simon et al., 2019). Our failure to detect *Sulcia* using insect transcriptomes could be due either to the absence of *Sulcia* or low-quality RNA samples. Since *Sulcia* provides essential nutrients to its hosts, it is less likely that our failure to detect *Sulcia* in these species resulted from lack of expression of any *Sulcia* genes. Based on the study of Skinner et al. (2020), the transcriptomes of *T. crinita* and *E. fabae* showed only moderate rates of missing data for the arthropod orthologue set used in that study (39.2–43.2%), suggesting relatively good RNA quality. We were able to retrieve *Sulcia* genes even from transcriptomes with otherwise high levels of missing data (89.6–91.6%). Given the discrepancy between our study and previous reports of *Sulcia* in *T. crinita* and *E. fabae*, it is possible that loss of *Sulcia* has occurred only within certain populations of these species. This hypothesis needs further testing. Previous authors suggested that microleafhoppers (Typhlocybinae) have frequently lost *Sulcia* and other nutritional symbionts common among other leafhoppers because, unlike most leafhoppers, typhlocybines feed preferentially on more nutritious parenchyma cell contents rather than nutrient-poor phloem or xylem sap (Bennett and Moran, 2013).

Compared to Cicadomorpha, losses of *Sulcia* appear to have been more common in Fulgoromorpha. Previous studies (Müller, 1962; Moran et al., 2005; Urban and Cryan, 2012) and our work all failed to detect *Sulcia* in Acanaloniidae, Flatidae, Ricaniidae and Tropiduchidae. Urban and Cryan (2012) also reported the absence of *Sulcia* in Caliscelidae, Derbidae, Eurybrachidae and Issidae. However, Müller (1962) and/or our study indicate at least some of the species in these families harbor *Sulcia*. Similarly, *Sulcia* has been identified in some species of Delphacidae by Müller (1962) and Urban and Cryan (2012) but was not detected by Moran et al. (2005) and Tang et al. (2010) in some other species in this family. Here, we found cases of both the presence and absence of *Sulcia* in Delphacidae and Fulgoridae. These results suggest that multiple

independent losses of Sulcia occurred in planthoppers. Interestingly, most of these families, including Acanaloniidae, Caliscelidae, Eurybrachidae, Flatidae, Issidae, Ricaniidae and Tropicodidae, were placed in a single, relatively derived sublineage of Fulgoroidea according to a recent phylogeny of Auchenorrhyncha (Skinner et al., 2020), implying recurrent loss of Sulcia in this sublineage. Additional studies of endosymbiont loss and replacement in this lineage are needed to reveal the evolutionary patterns of such changes in the microbiome and the identities and phylogenetic affinities of replacement endosymbionts.

Implications for Auchenorrhyncha phylogeny

The evolutionary history of the diverse sap-sucking hemipteran suborder Auchenorrhyncha (> 43 000 described species) has been contentious for decades (Cryan and Urban, 2012; Bartlett et al., 2018; Skinner et al., 2020). The suborder comprises two infraorders: Fulgoromorpha, including a single extant superfamily Fulgoroidea (planthoppers, 21 families) and Cicadomorpha consisting of three superfamilies, Cicadoidea (cicadas, two families), Cercopoidea (spittlebugs, five families) and Membracoidea (leafhoppers and treehoppers, five families). Based on some morphological synapomorphies, Hennig (1981) and Zrzavý (1992) suggested that Auchenorrhyncha is monophyletic, although they suggested different placements of this group within Hemiptera (sister to either Sternorrhyncha or Heteropteroidea). However, other early studies suggested that Auchenorrhyncha is not monophyletic based on other morphological characters (Hamilton, 1981), fossil evidence (Hamilton, 1996; Szewo, 2002), molecular data (Campbell et al., 1995) or combined evidence (Bourgoin and Campbell, 2002). A recent phylogeny based on mitochondrial genomes also failed to recover a monophyletic Auchenorrhyncha, placing Cicadomorpha as the sister to Fulgoroidea plus Coleorrhyncha (Li et al., 2017). Nevertheless, Auchenorrhyncha was repeatedly recovered as monophyletic in several recent studies which used either Sanger sequence data from several genes (Cryan and Urban, 2012) or transcriptomes (Misof et al., 2014; Johnson et al., 2018; Skinner et al., 2020). The transcriptome-based studies consistently supported a sister relationship between Auchenorrhyncha and Coleorrhyncha (Johnson et al., 2018; Skinner et al., 2020). Within Auchenorrhyncha, the two infraorders and the four superfamilies are widely accepted as monophyletic, although Emelyanov (1987) and Song et al. (2017) suggested Fulgoroidea plus Membracoidea is sister to Cercopoidea plus Cicadoidea based on morphological data and mitochondrial genomes, respectively. However, the relationships within Cicadomorpha remain controversial, with various studies having proposed all

three possible alternative relationships among the three included superfamilies, i.e., (i) Cercopoidea + (Cicadoidea + Membracoidea); (ii) Cicadoidea + (Cercopoidea + Membracoidea) and (iii) Membracoidea + (Cercopoidea + Cicadoidea). Each of these alternatives appears to have at least some morphological support (Evans, 1963, 1977; Hamilton, 1981, 1999; Emelyanov, 1987; Blocker, 1996; Liang and Fletcher, 2002; Rakhov, 2002). The first hypothesis has also been defended based on fossil evidence (Hamilton, 1996; Szewo, 2002). Early molecular phylogenies based on 18S rRNA suggested either Cicadoidea (von Dohlen and Moran, 1995) or Membracoidea (Campbell et al., 1995; Ouvrard et al., 2000) as sister to the remaining Cicadomorpha. Combining morphological, fossil and genetic evidence, Bourgoin and Campbell (2002) suggested the third topology as the most likely. Li et al. (2017) also proposed Membracoidea as sister to Cercopoidea plus Cicadoidea using mitochondrial genomes. Surprisingly, the last two topologies were both recovered by phylogenomic analyses based on the same genes (> 2000 genes) but using alternative methods of coding characters and conducting the analyses (Johnson et al., 2018; Skinner et al., 2020). Maximum likelihood analyses using amino acid alignments and a degeneracy-coded nucleotide dataset recovered Cicadoidea as the sister to Cercopoidea plus Membracoidea with high to maximum support, while concatenated and coalescent analyses using complete nucleotide sequences (third codon positions unaltered) fully supported the third hypothesis. Skinner et al. (2020) also conducted a separate analysis by grouping genes with different guanine-cytosine (GC) content in the third codon position, which favored a sister relationship between Cercopoidea and Cicadoidea.

The phylogeny of Sulcia is largely concordant with the most recent phylogeny of Auchenorrhyncha based on transcriptomic data (Skinner et al., 2020) with most areas of incongruence attributable to low branch support or topological instability in one or both analyses. Our results strongly supported several phylogenetic hypotheses of Auchenorrhyncha, including: (i) monophyly of Auchenorrhyncha with Fulgoromorpha sister to Cicadomorpha; (ii) monophyly of the four auchenorrhynchan superfamilies, Fulgoroidea, Cicadoidea, Cercopoidea and Membracoidea; and (iii) paraphyly of Cicadellidae with respect to three treehopper families. The higher-level phylogeny of Cicadomorpha remains controversial, although several recent studies recovered Membracoidea as the sister to Cicadoidea plus Cercopoidea (Cryan and Urban, 2012; Li et al., 2017; Johnson et al., 2018; Skinner et al., 2020). Our phylogeny of Sulcia consistently recovered a sister relationship between Membracoidea and Cercopoidea with these two groups together sister to Cicadoidea. This relationship was also recovered in some of the analyses

presented by Johnson et al. (2018; analysis of concatenated amino acid dataset) and Skinner et al. (2020; analyses of concatenated amino acid dataset and nucleotide data with degeneracy-coded third codon positions). It is also supported by some morphological synapomorphies shared among extant Membracoidea and Cercopoidea, such as the presence of subgenital plates and absence of the median ocellus (Hamilton, 1981; Blocker, 1996). Nevertheless, Skinner et al. (2020) reported extensive gene tree conflict in relationships among Cicadoidea, Cercopoidea and Membracoidea, so this aspect of Auchenorrhyncha phylogeny must be considered inadequately resolved at present.

The planthopper lineage (Fulgoroidea) comprises 21 extant families with more than 12 000 species (Bartlett et al., 2018). Although a few early morphology-based studies considered Tettigometridae as the most plesiomorphic fulgoroid family (Muir, 1923; Emelyanov, 1991), many subsequent analyses recovered Cixiidae plus Delphacidae as sister to the other fulgoroids with Tettigometridae in a relatively derived position (Yeh et al., 2005; Urban and Cryan, 2007; Song and Liang, 2013; Skinner et al., 2020). The latter hypothesis was supported by a previous *16S*-based phylogeny of Sulcia (Urban and Cryan, 2012), although the support values were generally low. Our phylogenomic results strongly supported the sister relationship between the Cixiidae-Delphacidae lineage and the remaining Fulgoroidea, but we did not recover Sulcia from our tettigometrid transcriptome dataset and the strain previously reported from Tettigometridae (Urban and Cryan, 2012) was not included in our *16S*-based phylogenetic analysis. Yeh et al. (2005) and Urban and Cryan (2007) questioned the monophyly of Cixiidae and Delphacidae. Here, we consistently recovered the Cixiidae-associated Sulcia as paraphyletic. In agreement with most prior molecular phylogenies of Fulgoroidea (Yeh et al., 2005; Urban and Cryan, 2007; Skinner et al., 2020), our analyses based on Sulcia genes supported the placement of Achilidae near the root of the tree and also recovered a close relationship between Dictyopharidae and Fulgoridae. Although relationships among the remaining taxa were poorly resolved in previous analyses, many of the families were placed in a large, derived clade, referred to as “advanced” or “higher” fulgoroids by Yeh et al. (2005) and Bartlett et al. (2018), which usually contains Acanaloniidae, Caliscelidae, Flatidae, Issidae, Nogodinidae, Ricaniidae, Tettigometridae and Tropiduchidae (Urban and Cryan, 2007; Skinner et al., 2020). As mentioned above, many members of this lineage were found to lack Sulcia.

The phylogeny of Cicadoidea was not studied comprehensively until very recently. Based on five molecular markers, Marshall et al. (2018) constructed a phylogeny for 139 cicada species. They confirmed the classification

scheme proposed by Moulds (2005) (two families in Cicadoidea and three subfamilies in Cicadidae) and established a new subfamily. Based on their results, Tettigarctidae and Cicadidae are sister groups and Tibicininae is a monophyletic sister clade to the remaining Cicadidae. However, relationships among the monophyletic groups Cicadinae and Cicadettinae as well as the paraphyletic group Tettigomyiinae were not resolved. Based on mitochondrial genomes, Łukasik et al. (2019) placed a poorly known genus *Derotettix* as sister to the remaining Cicadidae and recovered Tibicininae as the sister to Cicadettinae plus Cicadinae. More recently, Simon et al. (2019) reconstructed the phylogeny of Cicadoidea using anchored hybrid enrichment data. They recognized a fifth cicadid subfamily to include *Derotettix* and suggested the following relationships: Tettigarctidae + (Derotettiginae + (Tibicininae + (Tettigomyiinae + (Cicadettinae + Cicadinae))). They also recovered a similar phylogeny using ten Sulcia genes, which however were unable to resolve the relationships among Tettigomyiinae, Cicadinae and Cicadettinae. Only three subfamilies of Cicadidae are represented in our Sulcia dataset, including Tibicininae, Cicadettinae and Cicadinae. Our phylogenomic analyses of Sulcia, except the MP results, support Tibicininae as the sister to Cicadettinae plus Cicadinae, which is highly consistent with the phylogeny of cicadas. As in Łukasik et al. (2019), Chilecicadini is sister to Tibicinini plus Tettigadini. Relationships within Cicadettinae differ slightly among studies. In Marshall et al. (2018), Cicadettini grouped with Chlorocystini, and Lamotialnini was placed in the same clade with Cicadatrini. In contrast, Łukasik et al. (2019) and our study recovered Cicadettini as sister to the other three tribes, but both failed to resolve the relationships among Chlorocystini, Lamotialnini and Cicadatrini. Deeper relationships within Cicadinae were inconsistent among analyses using different data but the sister relationship between Leptopsaltriini and Sonatini plus Polyneurini, and their close relationship to Dundubiini, were strongly supported by previous studies (Marshall et al., 2018; Matsuura et al., 2018; Łukasik et al., 2019) as well as by the ML analyses in our study.

Cercopoidea is a relatively small superfamily of Auchenorrhyncha but remains the least well studied with respect to phylogeny. Previous studies have failed to achieve consensus in either classification schemes or phylogeny of this group. Morphology-based studies classified cercopoids into as many as five families, including Aphrophoridae, Cercopidae, Clastopteridae, Epipygidae and Machaerotidae (Baker, 1927; Metcalf, 1960). Using morphological criteria, Hamilton (2001, 2012, 2015) recognized only three of the five families by treating Machaerotinae as a subfamily of Clastopteridae and Aphrophorinae as a subfamily of the Cercopidae, while other recent authors retained the

five-family system (Bartlett et al., 2018; Paladini et al., 2018; Skinner et al., 2020). The most extensive phylogenetic study of cercopoids (Cryan and Svenson, 2010) suggested Machaerotidae is the earliest diverging lineage, sister to the remaining cercopoids. This relationship was also supported by Cryan and Urban (2012). However, some other analyses recovered Clastopteridae plus Machaerotidae as sister to the remaining cercopoids (Cryan, 2005; Bell et al., 2014; Paladini et al., 2018). In the recent transcriptome-based phylogenetic study of Auchenorrhyncha (Skinner et al., 2020), different analyses recovered either Clastopteridae or Machaerotidae as sister to the remaining Cercopoidea but never placed these two families in a single clade. Our phylogenomic analysis of Sulcia is consistent with several previous studies (Cryan, 2005; Cryan and Svenson, 2010; Cryan and Urban, 2012; Bell et al., 2014; Skinner et al., 2020) in recovering Aphrophoridae as paraphyletic with respect to Epipygidae and Cercopidae, but most of the branches received less than maximum bootstrap support.

Membracoidea is the most speciose group in Auchenorrhyncha, with > 25 000 extant species (Bartlett et al., 2018) placed in five families, Aetalionidae, Cicadellidae, Melizoderidae, Membracidae and Myrslopiidae. Hamilton (1999) suggested Myrslopiidae, a lineage distinct from leafhoppers and treehoppers, is the earliest diverging extant group based on some morphological characters shared with cicadooids and cercopoids but absent in other membracoids. This hypothesis was supported by many subsequent studies (Dietrich, 2002; Cryan and Urban, 2012; Dietrich et al., 2017; Johnson et al., 2018; Skinner et al., 2020) as well as our results. The leafhopper family Cicadellidae was traditionally recognized as sister to treehoppers (Evans, 1946) but phylogenetic studies of morphological and molecular data have consistently shown that Cicadellidae is paraphyletic with respect to the treehopper clade (including Aetalionidae, Melizoderidae and Membracidae) (Hamilton, 1999; Dietrich et al., 2001, 2017; Cryan and Urban, 2012; Johnson et al., 2018; Skinner et al., 2020). Our results agree with these prior studies. Treehoppers were assumed to be most closely related to the leafhopper group Ulopinae, as they both lack the rows of macrosetae on the hind tibia characteristic of other leafhoppers (Hamilton, 1999). Recent molecular phylogenies indicated that the treehopper clade is sister to one or both of the leafhopper subfamilies Megophthalminae and Ulopinae (Cryan and Urban, 2012; Dietrich et al., 2017; Skinner et al., 2020). Although Megophthalminae was not included in our phylogenomic analyses of Sulcia, the analyses recovered the sister relationship between the treehopper lineage and Ulopinae. In the 16S tree for Sulcia, the treehopper lineage was sister to Megophthalminae plus Ulopinae,

in agreement with prior phylogenomic analyses (Dietrich et al., 2017; Skinner et al., 2020). Prior analyses have consistently failed to resolve relationships among most major lineages (subfamilies) of Cicadellidae, possibly due to an ancient rapid radiation in this group (Dietrich et al., 2017; Skinner et al., 2020). Our phylogeny based on Sulcia genes also failed to clarify the relationships among major lineages of leafhoppers, suggesting coincident rapid radiation between leafhoppers and associated Sulcia lineages. Within the clade of treehoppers, although most relationships appear to be well resolved by previous analyses, the relationships among the three families remain contentious. Dietrich and Deitz (1993) treated the absence of parapsidal clefts on the mesonotum as a synapomorphy for Aetalionidae and Membracidae, and suggested that Melizoderidae, which resemble leafhoppers in retaining parapsidal clefts, are sister to Aetalionidae plus Membracidae. However, using anchored hybrid enrichment data, Dietrich et al. (2017) recovered Aetalionidae as sister, or paraphyletic with respect, to Melizoderidae plus Membracidae. Transcriptome-based analysis consistently recovered a sister relationship between Melizoderidae and a membracid *Holdgatiella chepuensis* (Skinner et al., 2020). However, the placements of this clade and Aetalionidae varied among analyses with some placing Aetalionidae as sister to the remaining treehoppers, and others placing Melizoderidae + *Holdgatiella* as sister to all other treehoppers. Our study consistently recovered Melizoderidae + *Holdgatiella* as sister to the other treehoppers and Aetalionidae as sister to the Stegaspidinae-Centrotinae clade, which is concordant with previous insect transcriptome-based analyses using either the concatenated amino acid alignment or the degeneracy-coded nucleotide dataset (Skinner et al., 2020).

Conclusions

The most taxon- and character-rich phylogenetic study of Sulcia and related flavobacterial insect endosymbionts so far confirms that the phylogeny of Sulcia is largely congruent with that of its hosts. Although the phylogeny based on Sulcia genes did not completely resolve the relationships within Auchenorrhyncha, it yielded estimates that are well resolved and largely congruent with recently published phylogenies based on large numbers of insect orthologue sequences. Thus, phylogenetic analyses of Sulcia endosymbiont genomic data are generally informative of phylogenetic relationships of their hosts and contribute to the testing of some contentious hypotheses concerning the phylogeny of Auchenorrhyncha. In agreement with some prior studies, our results strongly suggest that Sulcia occurs exclusively within the

hemipteran suborder Auchenorrhyncha but that this primary endosymbiont has been lost independently in several auchenorrhynchan lineages. Our results also show for the first time that *Sulcia* belongs to a much larger group of flavobacterial insect endosymbionts that includes symbionts of scale insects, cockroaches, and beetles but incongruence in the phylogenetic relationships among these major symbiont lineages and those of their hosts suggests that some combination of multiple independent acquisition and loss has occurred. Further screening and phylogenetic analyses of the microbiota of insects belonging to these and other lineages are needed to further elucidate patterns of acquisition and loss of flavobacterial endosymbionts during the evolution of the group.

Acknowledgment

This study was funded by U.S. National Science Foundation grant DEB 16-39601.

Conflicts of interest

The authors declare no conflicts of interest.

References

- Ankrah, N.Y.D., Chouaia, B. and Douglas, A.E., 2018. The cost of metabolic interactions in symbioses between insects and bacteria with reduced genomes. *MBio* 9, e01433-18.
- Baker, C.F. 1927. Some Philippine and Malaysian Machaerotidae (Cercopoidea). *Philipp. J. Sci.* 32, 529–548.
- Baumann, P., Moran, N.A., Baumann, L. and Dworkin, M. 2006. Bacteriocyte-associated endosymbionts of insects. *Prokaryotes* 1, 403–438.
- Balbuena, J.A., Míguez-Lozano, R. and Blasco-Costa, I., 2013. PAcO: a novel procrustes application to cophylogenetic analysis. *PLoS One* 8, e61048.
- Bartlett, C.R., Deitz, L.L., Dmitriev, D.A., Sanborn, A.F., Soulier-Perkins, A. and Wallace, M.S., 2018. The diversity of the true hoppers (Hemiptera: Auchenorrhyncha). In: Footitt, R. G. & Adler, P. H. (Eds.), *Insect Biodiversity*. John Wiley & Sons Ltd, Chichester, pp. 501–590.
- Bell, A.J., Svenson, G.J. and Cryan, J.R., 2014. The phylogeny and revised classification of Machaerotidae, the tube-making spittlebugs (Hemiptera: Auchenorrhyncha: Cercopoidea). *Syst. Entomol.* 39, 474–485.
- Bennett, G.M., Abbà, S., Kube, M. and Marzachi, C., 2016. Complete genome sequences of the obligate symbionts “*Candidatus Sulcia muelleri*” and “*Ca. Nasuia deltocephalinicola*” from the pestiferous leafhopper *Macrostelus quadripunctulatus* (Hemiptera: Cicadellidae). *Genome Announc.* 4, e1604–15. <https://doi.org/10.1128/genomeA.01604-15>
- Bennett, G.M. and Mao, M., 2018. Comparative genomics of a quadripartite symbiosis in a planthopper host reveals the origins and rearranged nutritional responsibilities of anciently diverged bacterial lineages. *Environ. Microbiol.* 20, 4461–4472.
- Bennett, G.M., McCutcheon, J.P., MacDonald, B.R., Romanovicz, D. and Moran, N.A., 2014. Differential genome evolution between companion symbionts in an insect-bacterial symbiosis. *MBio* 5, e01697–14. <https://doi.org/10.1128/mBio.01697-14>
- Bennett, G.M. and Moran, N.A., 2013. Small, smaller, smallest: the origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biol. Evol.* 5, 1675–1688.
- Blocker, H.D., 1996. Origin and radiation of the Auchenorrhyncha. In: Schaefer, C. W. (Ed.), *Studies on Hemipteran Phylogeny*. Thomas Say Publications in Entomology, Entomological Society of America, Lanham, MD, pp. 46–64.
- Bolger, A.M., Lohse, M. and Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.
- Bourgoin, T. and Campbell, B.C., 2002. Inferring a phylogeny for Hemiptera: falling into the “autapomorphic trap”. *Denisia* 4, 67–82.
- Campbell, B.C., Steffen-Campbell, J.D., Sorensen, J.T. and Gill, R.J., 1995. Paraphyly of Homoptera and Auchenorrhyncha inferred from 18S rDNA nucleotide sequences. *Syst. Entomol.* 20, 175–194.
- Capella-Gutiérrez, S., Silla-Martínez, J.M. and Gabaldón, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25, 1972–1973.
- Campbell, M.A., Leuven, J.T.V., Meister, R.C., Carey, K.M., Simon, C. and McCutcheon, J.P., 2015. Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont *Hodgkinia*. *Proc. Natl. Acad. Sci.* 112, 10192–10199.
- Chang, H.H., Cho, S.T., Canale, M.C., Mugford, S.T., Lopes, J.R.S., Hogenhout, S.A. and Kuo, C.-H., 2015. Complete genome sequence of “*Candidatus Sulcia muelleri*” ML, an obligate nutritional symbiont of maize leafhopper (*Dalbulus maidis*). *Genome Announc.* 3, e01483–14. <https://doi.org/10.1128/genomeA.01483-14>
- Clark, E.L., Karley, A.J. and Hubbard, S.F., 2010. Insect endosymbionts: manipulators of insect herbivore trophic interactions? *Protoplasma* 244, 25–51.
- Clark, M.A., Moran, N.A., Baumann, P. and Wernegreen, J.J., 2000. Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution* 54, 517–525.
- Cryan, J.R., 2005. Molecular phylogeny of Cicadomorpha (Insecta: Hemiptera: Cicadoidea, Cercopoidea and Membracoidea): adding evidence to the controversy. *Syst. Entomol.* 30, 563–574.
- Cryan, J.R. and Svenson, G.J., 2010. Family-level relationships of the spittlebugs and froghoppers (Hemiptera: Cicadomorpha: Cercopoidea). *Syst. Entomol.* 35, 393–415.
- Cryan, J.R. and Urban, J.M., 2012. Higher-level phylogeny of the insect order Hemiptera: is Auchenorrhyncha really paraphyletic? *Syst. Entomol.* 37, 7–21.
- DeLay, B.D., 2013. Symbionts associated with the salivary glands of the potato leafhopper, *Empoasca fabae*, and their function when feeding on leguminous hosts. (Doctoral dissertation).
- Dietrich, C.H., 2002. Evolution of Cicadomorpha (Insecta, Hemiptera). *Denisia* 4, 155–170.
- Dietrich, C.H., Allen, J.M., Lemmon, A.R., Lemmon, E.M., Takiya, D.M., Evangelista, O., Walden, K.K.O., Grady, P.G.S. and Johnson, K.P., 2017. Anchored hybrid enrichment-based phylogenomics of leafhoppers and treehoppers (Hemiptera: Cicadomorpha: Membracoidea). *Insect Syst. Divers.* 1, 57–72.
- Dietrich, C.H. and Deitz, L.L., 1993. Superfamily Membracoidea (Homoptera: Auchenorrhyncha). II. Cladistic analysis and conclusions. *Syst. Entomol.* 18, 297–311.
- Dietrich, C.H., Rakitov, R.A., Holmes, J.L. and Black, W.C., 2001. Phylogeny of the major lineages of Membracoidea (Insecta: Hemiptera: Cicadomorpha) based on 28S rDNA sequences. *Mol. Phylogenet. Evol.* 18, 293–305.
- von Dohlen, C.D. and Moran, N.A., 1995. Molecular phylogeny of the Homoptera: a paraphyletic taxon. *J. Mol. Evol.* 41, 211–223.
- Eleftherianos, I., Yadav, S., Kenney, E., Cooper, D., Ozakman, Y. and Patnogi, J., 2018. Role of endosymbionts in insect–parasitic nematode interactions. *Trends Parasitol.* 34, 430–444.
- Emelyanov, A.F., 1987. The phylogeny of the Cicadina (Homoptera, Cicadina) based on comparative morphological data. *Trans. All-Union Entomol. Soc.* 69, 19–109.

- Emelyanov, A.F., 1991. An attempt of construction of the phylogenetic tree of the planthoppers (Homoptera, Cicadina). *Entomol. Rev.* 70, 24–28.
- Emms, D.M. and Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238.
- Engl, T., Eberl, N., Gorse, C., Krüger, T., Schmidt, T.H.P., Plarre, R., Adler, C. and Kaltenpoth, M., 2018. Ancient symbiosis confers desiccation resistance to stored grain pest beetles. *Mol. Ecol.* 27, 2095–2108.
- Evans, J.W., 1946. A natural classification of leaf-hoppers (Homoptera, Jassoidea). Part 2: Aetalionidae, Hylcidae, Eurymelidae. *Trans. R. Entomol. Soc. Lond.* 97, 39–54.
- Evans, J.W., 1963. The phylogeny of the Homoptera. *Annu. Rev. Entomol.* 8, 77–94.
- Evans, J.W., 1977. The leafhoppers and froghoppers of Australia and New Zealand (Homoptera: Cicadelloidea and Cercopoidea) part 2. *Rec. Aust. Mus.* 31, 83–129.
- Galetto, L., Abbà, S., Rossi, M., Vallino, M., Pesando, M. and Marzachi, C., 2018. Two phytoplasmas elicit different responses in the insect vector *Euscelidius variegatus* Kirschbaum. *Infect. Immun.* 86, e00042–18.
- Goloboff, P.A. and Catalano, S.A., 2016. TNT version 1.5, including a full implementation of phylogenetic morphometrics. *Cladistics* 32, 221–238.
- Gruwell, M.E., Morse, G.E. and Normark, B.B., 2007. Phylogenetic congruence of armored scale insects (Hemiptera: Diaspididae) and their primary endosymbionts from the phylum Bacteroidetes. *Mol. Phylogenet. Evol.* 44, 267–280.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512.
- Hamilton, K.G.A., 1981. Morphology and evolution of the rhyngotan head (Insecta: Hemiptera, Homoptera). *Can. Entomol.* 113, 953–974.
- Hamilton, K.G.A., 1996. Cretaceous Homoptera from Brazil: implications for classification. In: Schaefer, C. W. (Ed.), *Studies on Hemipteran Phylogeny*. Thomas Say Publications in Entomology, Entomological Society of America, Lanham, MD, pp. 89–110.
- Hamilton, K.G.A., 1999. The ground-dwelling leafhoppers Myerslopiidae, new family, and Sagmatiini, new tribe (Homoptera: Membracoidea). *Invertebr. Taxon.* 13, 207–235.
- Hamilton, K.G.A., 2001. A new family of froghoppers from the American tropics (Hemiptera: Cercopoidea: Epipygidae). *Biodiversity* 2, 15–21.
- Hamilton, K.G.A., 2012. Revision of Neotropical aphrophorine spittlebugs, part 1: Ptyelini (Hemiptera, Cercopoidea). *Zootaxa* 3497, 41–59.
- Hamilton, K.G.A., 2015. A new tribe and species of Clastopterinae (Hemiptera: Cercopoidea: Clastopteridae) from Africa, Asia and North America. *Zootaxa* 3946, 151–189.
- Hennig, W., 1981. *Insect Phylogeny*. John Wiley & Sons Ltd, New York, NY.
- Itoh, T., Martin, W. and Nei, M., 2002. Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc. Natl. Acad. Sci.* 99, 12944–12948.
- Jia, D., Mao, Q., Chen, Y., Liu, Y., Chen, Q., Wu, W., Zhang, X., Chen, H., Li, Y. and Wei, T., 2017. Insect symbiotic bacteria harbour viral pathogens for transovarial transmission. *Nat. Microbiol.* 2, 1–7.
- Johnson, K.P., Dietrich, C.H., Friedrich, F., Beutel, R.G., Wipfler, B. and Yoshizawa, K., 2018. Phylogenomics and the evolution of hemipteroid insects. *Proc. Natl. Acad. Sci.* 115, 12775–12780.
- Kikuchi, Y., 2009. Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes Environ.* 24, 195–204.
- Kikuchi, Y., Hayatsu, M., Hosokawa, T., Nagayama, A., Tago, K. and Fukatsu, T., 2012. Symbiont-mediated insecticide resistance. *Proc. Natl. Acad. Sci.* 109, 8618–8622.
- Koga, R. and Moran, N.A., 2014. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. *ISME J.* 8, 1237–1246.
- Kölsch, G. and Pedersen, B.V., 2010. Can the tight co-speciation between reed beetles (Col., Chrysomelidae, Donaciinae) and their bacterial endosymbionts, which provide cocoon material, clarify the deeper phylogeny of the hosts? *Mol. Phylogenet. Evol.* 54, 810–821.
- Koohpayma, F., Bagheri, A., Fallahzadeh, M., Dousti, A.F. and Askari-Seyahooei, M., 2018. *Nysius cymoides* (Hemiptera: Lygaeidae), a new economically important pest on *Acacia tortilis* and its intracellular bacterial endosymbionts. *Entomol. News.* 128, 11–23.
- Kuechler, S.M., Gibbs, G., Burckhardt, D., Dettner, K. and Hartung, V., 2013. Diversity of bacterial endosymbionts and bacteria–host co-evolution in Gondwanan relict moss bugs (Hemiptera: Coleorrhyncha: Peloridiidae). *Environ. Microbiol.* 15, 2031–2042.
- Kumar, S., Stecher, G., Suleski, M. and Hedges, S.B., 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. and Evol.* 34, 1812–1819.
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. and Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773.
- Liang, A.P. and Fletcher, M.J., 2002. Morphology of the antennal sensilla in four Australian spittlebug species (Hemiptera: Cercopoidea) with implications for phylogeny. *Aust. J. Entomol.* 41, 39–44.
- Li, F.W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z. and Szövényi, P., 2020. *Anthoceros* genomes illuminate the origin of land plants and the unique biology of hornworts. *Nat. Plants* 6, 259–272.
- Li, H., Leavengood, J.M., Chapman, E.G., Burkhardt, D., Song, F., Jiang, P., Liu, J., Zhou, X. and Cai, W., 2017. Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *Proc. Royal Soc. B.* 284, 20171223.
- Liu, L., Huang, X., Zhang, R., Jiang, L. and Qiao, G., 2013. Phylogenetic congruence between *Mollitrachosiphum* (Aphididae: Greenideinae) and Buchnera indicates insect–bacteria parallel evolution. *Syst. Entomol.* 38, 81–92.
- Lo, N., Bandi, C., Watanabe, H., Nalepa, C. and Beninati, T., 2003. Evidence for cocoladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Mol. Biol. Evol.* 20, 907–913.
- Lukasik, P., Chong, R.A., Nazario, K., Matsuura, Y., Bublitz, D.A.C. and McCutcheon, J.P., 2019. One hundred mitochondrial genomes of cicadas. *J. Hered.* 110, 247–256.
- Mao, M., Yang, X., Poff, K. and Bennett, G., 2017. Comparative genomics of the dual-obligate symbionts from the treehopper, *Entylia carinata* (Hemiptera: Membracidae), provide insight into the origins and evolution of an ancient symbiosis. *Genome Biol. Evol.* 9, 1803–1815.
- Marin, J., Battistuzzi, F.U., Brown, A.C. and Hedges, S.B., 2017. The timetree of prokaryotes: new insights into their evolution and speciation. *Mol. Biol. Evol.* 34, 437–446.
- Marshall, D.C., Moulds, M., Hill, K.B.R., Price, B.W., Wade, E.J. and Simon, C., 2018. A molecular phylogeny of the cicadas (Hemiptera: Cicadidae) with a review of tribe and subfamily classification. *Zootaxa* 4424, 1–64.
- Matsuura, Y., Moriyama, M., Lukasik, P., Vanderpool, D., Tanahashi, M., Meng, X.Y., McCutcheon, J.P. and Fukatsu, T., 2018. Recurrent symbiont recruitment from fungal parasites in cicadas. *Proc. Natl. Acad. Sci.* 115, E5970–E5979.
- McCutcheon, J.P., McDonald, B.R. and Moran, N.A., 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci.* 106, 15394–15399.
- McCutcheon, J.P. and Moran, N.A., 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci. USA* 104, 19392–19397.

- McCutcheon, J.P. and Moran, N.A., 2010. Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol. Evol.* 2, 708–718.
- Metcalf, Z.P., 1960. General catalogue of the Homoptera, Fascicle VII: Cercopoidea, Part 1: Machaerotidae. Waverly Press Inc., Baltimore, MD.
- Mirarab, S., Nguyen, N., Guo, S., Wang, L.-S., Kim, J. and Warnow, T., 2015. PASTA: Ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J. Comput. Biol.* 22, 377–386.
- Mirarab, S. and Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31, i44–i52.
- Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A. and Zhou, X., 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346, 763–767.
- Montllor, C.B., Maxmen, A. and Purcell, A.H., 2002. Facultative bacterial endosymbionts benefit pea aphids *Acyrtosiphon pisum* under heat stress. *Ecol. Entomol.* 27, 189–195.
- Moran, N.A., McCutcheon, J.P. and Nakabachi, A., 2008. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42, 165–190.
- Moran, N.A., Plague, G.R., Sandström, J.P. and Wilcox, J.L., 2003. A genomic perspective on nutrient provisioning by bacterial symbionts of insects. *Proc. Natl. Acad. Sci. USA* 100, 14543–14548.
- Moran, N.A., Tran, P. and Gerardo, N.M., 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum Bacteroidetes. *Appl. Environ. Microbiol.* 71, 8802–8810.
- Moulds, M., 2005. An appraisal of the higher classification of cicadas (Hemiptera: Cicadoidea) with special reference to the Australian fauna. *Rec. Aust. Mus.* 57, 375–446.
- Muir, F., 1923. On the classification of the Fulgoroidea (Homoptera). *Proc. Hawaii. Entomol. Soc.* 5, 205–247.
- Müller, H.J., 1962. Neuere Vorstellungen über Verbreitung und Phylogenie der Endosymbiosen der Zikaden. *Z. Morphol. Oekol. Tiere* 51, 190–210.
- Neelakanta, G., Sultana, H., Fish, D., Anderson, J.F. and Fikrig, E., 2010. *Anaplasma phagocytophilum* induces *Ixodes scapularis* ticks to express an antifreeze glycoprotein gene that enhances their survival in the cold. *J. Clin. Invest.* 120, 3179–3190.
- Nováková, E., Hypša, V., Klein, J., Footitt, R.G., von Dohlen, C.D. and Moran, N.A., 2013. Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Mol. Phylogenet. Evol.* 68, 42–54.
- Oliver, K.M., Campos, J., Moran, N.A. and Hunter, M.S., 2008. Population dynamics of defensive symbionts in aphids. *Proc. Royal Soc. B* 275, 293–299.
- Ouvrard, D., Campbell, B.C., Bourgoin, T. and Chan, K.L., 2000. 18S rRNA Secondary structure and phylogenetic position of (Insecta, Hemiptera). *Mol. Phylogenet. Evol.* 16, 403–417.
- Paladini, A., Takiya, D.M., Urban, J.M. and Cryan, J.R., 2018. New World spittlebugs (Hemiptera: Cercopidae: Ischnorhininae): dated molecular phylogeny, classification, and evolution of aposematic coloration. *Mol. Phylogenet. Evol.* 120, 321–334.
- Rakitov, R.A., 2002. Structure and function of the malpighian tubules, and related behaviors in juvenile cicadas: evidence of homology with spittlebugs (Hemiptera: Cicadoidea & Cercopoidea). *Zool. Anz.* 241, 117–130.
- Rosenbluth, M., Sayavedra, L., Sámano-Sánchez, H., Roth, A. and Martínez-Romero, E., 2012. Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). *J. Evol. Biol.* 25, 2357–2368.
- Salem, H., Bauer, E., Kirsch, R., Berasategui, A., Cripps, M., Weiss, B., Koga, R., Fukumori, K., Vogel, H., Fukatsu, T. and Kaltenpoth, M., 2017. Drastic genome reduction in an herbivore's pectinolytic symbiont. *Cell* 171, 1520–1531.e13.
- Sanderson, M.J., 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19, 301–302.
- Shih, H.-T., Su, C.-C., Chang, C.-J., Vargas, S., Dai, Z. and Chen, J., 2019. Draft genome sequence of “*Candidatus Sulcia muelleri*” strain KPTW1 from *Kolla paulula*, a vector of *Xylella fastidiosa* causing Pierce's disease of grapevine in Taiwan. *Microbiol. Resour. Announc.* 8, e01347-18.
- Simon, C., Gordon, E.R.L., Moulds, M.S., Cole, J.A., Haji, D. and Łukasik, P., 2019. Off-target capture data, endosymbiont genes and morphology reveal a relict lineage that is sister to all other singing cicadas. *Biol. J. Linn. Soc.* 128, 865–886.
- Skinner, R.K., Dietrich, C.H., Walden, K.K.O., Gordon, E., Sweet, A.D., Podsiadlowski, L., Petersen, M., Simon, C., Takiya, D.M. and Johnson, K.P., 2020. Phylogenomics of Auchenorrhyncha (Insecta: Hemiptera) using transcriptomes: examining controversial relationships via degeneracy coding and interrogation of gene conflict. *Syst. Entomol.* 45, 85–113.
- Song, N., Cai, W. and Li, H., 2017. Deep-level phylogeny of Cicadomorpha inferred from mitochondrial genomes sequenced by NGS. *Sci. Rep.* 7, 10429.
- Song, N. and Liang, A.-P., 2013. A preliminary molecular phylogeny of planthoppers (Hemiptera: Fulgoroidea) based on nuclear and mitochondrial DNA sequences. *PLoS One* 8, e58400.
- Stamatakis, A., 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Szwedo, J., 2002. Amber and amber inclusions of planthoppers, leafhoppers and their relatives (Hemiptera, Archaeorrhyncha et Clypeorrhyncha). *Denisia* 4, 37–56.
- Takiya, D.M., Tran, P.L., Dietrich, C.H. and Moran, N.A., 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta: Hemiptera: Cicadellidae) and their dual bacterial symbionts. *Mol. Ecol.* 15, 4175–4191.
- Tang, M., Lv, L., Jing, S., Zhu, L. and He, G., 2010. Bacterial symbionts of the brown planthopper, *Nilaparvata lugens* (Homoptera: Delphacidae). *Appl. Environ. Microbiol.* 76, 1740–1745.
- Tassone, E.E., Cowden, C.C. and Castle, S.J., 2017. *De novo* transcriptome assemblies of four xylem sap-feeding insects. *GigaScience* 6, giw007.
- Tokuda, G. and Watanabe, H., 2007. Hidden cellulases in termites: revision of an old hypothesis. *Biol. Lett.* 3, 336–339.
- Urban, J.M. and Cryan, J.R., 2007. Evolution of the planthoppers (Insecta: Hemiptera: Fulgoroidea). *Mol. Phylogenet. Evol.* 42, 556–572.
- Urban, J.M. and Cryan, J.R., 2012. Two ancient bacterial endosymbionts have coevolved with the planthoppers (Insecta: Hemiptera: Fulgoroidea). *BMC Evol. Biol.* 12, 87.
- Van Leuven, J.T., Meister, R.C., Simon, C. and McCutcheon, J.P., 2014. Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell*, 158, 1270–1280.
- Waneka, G., Vasquez, Y.M., Bennett, G.M. and Sloan, D.B., 2020. Mutational pressure drives differential genome conservation in two bacterial endosymbionts of sap feeding insects. *bioRxiv*. (preprint). <https://doi.org/10.1101/2020.07.29.225037>
- Wang, Y.H., Wu, H.Y., Rédei, D., Xie, Q., Chen, Y. and Bu, W.J., 2019. When did the ancestor of true bugs become stinky? Disentangling the phylogenomics of Hemiptera-Heteroptera. *Cladistics* 35, 42–66.
- Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A. and Cheng, J.F., 2010. One bacterial cell, one complete genome. *PLoS One* 5, e10314.
- Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L. and Eisen, J.A., 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* 4, e188.
- Yeh, W., Yang, C.-T. and Hui, C.F., 2005. A molecular phylogeny of planthoppers (Hemiptera: Fulgoroidea) inferred from mitochondrial 16S rDNA sequences. *Zool. Stud.* 44, 519–535.
- Zhan, S., Merlin, C., Boore, J.L. and Reppert, S.M., 2011. The monarch butterfly genome yields insights into long-distance migration. *Cell* 147, 1171–1185.

Zrzavý, J., 1992. Evolution of antennae and historical ecology of the hemipteran insects (Paraneoptera). *Acta Entomol. Bohemoslov.* 89, 77–86.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig S1A. Best tree constructed by maximum likelihood analysis of “*Ca. Sulcia muelleri*” and related endosymbionts based on *16S rRNA* (top portion).

Fig S1B. Best tree constructed by maximum likelihood analysis of “*Ca. Sulcia muelleri*” and related endosymbionts based on *16S rRNA* (bottom portion).

Fig S2. Strict consensus tree of *Flavobacteriales* constructed by bootstrap resampling algorithm of maximum parsimony analysis based on the amino acid sequences of 51 orthologues.

Fig S3. Best tree constructed by maximum likelihood analysis of “*Ca. Sulcia muelleri*” based on the amino acid sequences of 131 orthologues.

Fig S4. Best tree constructed by maximum likelihood analysis of “*Ca. Sulcia muelleri*” based on *16S rRNA* and the nucleotide sequences of 131 orthologues.

Fig S5. Strict consensus tree of “*Ca. Sulcia muelleri*” constructed by bootstrap resampling algorithm of maximum parsimony analysis based on *16S rRNA* and the nucleotide sequences of 131 orthologues.

Fig S6. Species tree of “*Ca. Sulcia muelleri*” resulting from the coalescent analysis of 132 gene trees (*16S rRNA* plus 131 orthologues, nucleotide datasets).

Table S1. Information of the 182 bacterial genomes used in this study.

Table S2. Predicted *16S rRNA* of *Sulcia* based on the transcriptomes of Auchenorrhyncha.