Cladogenesis and Genomic Streamlining in Extracellular Endosymbionts of Tropical Stink Bugs

Alejandro Otero-Bravo[1], Shana Goffredi[2], and Zakee L. Sabree[1,*]

[1]Department of Evolution, Ecology and Organismal Biology, Ohio State University

[2]Department of Biology, Occidental College, Los Angeles, California

*Corresponding author: E-mail: sabree.8@osu.edu.

Accepted: February 5, 2018

Data deposition: These Whole Genome Shotgun projects have been deposited at GenBank under the accessions PDKR00000000, PDKS00000000, PDKT00000000, and PDKU00000000. The versions described in this paper are versions PDKR01000000, PDKS01000000, PDKT01000000, and PDKU01000000. Sequence alignments for phylogenetic reconstruction as well as scripts used in this paper are available under DOI 10.6084/m9.figshare.5700718 and 10.6084/m9.figshare.5700721.

Abstract

Phytophagous stink bugs are globally distributed and many harbor vertically inherited bacterial symbionts that are extracellular, yet little is known about how the symbiont's genomes have evolved under this transmission strategy. Genome reduction is common in insect intracellular symbionts but limited genome sampling of the extracellular symbionts of distantly related stink bugs has precluded inferring patterns of extracellular symbiont genome evolution. To address this knowledge gap, we completely sequenced the genomes of the uncultivable bacterial symbionts of four neotropical stink bugs of the *Edessa* genus. Phylogenetic and comparative analyses indicated that the symbionts form a clade within the *Pantoea* genus and their genomes are highly reduced (~0.8 Mb). Furthermore, genome synteny analysis and a jackknife approach for phylogenetic reconstruction, which corrected for long branch attraction artifacts, indicated that the *Edessa* symbionts were the result of a single symbiotic event that was distinct from the symbiosis event giving rise to *Candidatus* "Pantoea carbekii," the extracellular symbiont of the invasive pentatomid stink bug, *Halyomorpha halys*. Metabolic functions inferred from the *Edessa* symbiont genomes suggests a shift in genomic composition characteristic of its lifestyle in that they retained many host-supportive functions while undergoing dramatic gene loss and establishing a stable relationship with their host insects. Given the undersampled nature of extracellular insect symbionts, this study is the first comparative analysis of these symbiont genomes from four distinct *Edessa* stink bug species. Finally, we propose the candidate name "*Candidatus* Pantoea edessiphila" for the species of these symbionts with strain designations according to their host species.

Key words: genome evolution, *Pantoea*, primary symbiont, neotropical stink bug, genome degradation.

Introduction

True bugs (Hemiptera), including aphids, cicadas, leafhoppers, and stink bugs, as a group are known to maintain ancient mutualist associations (>100 Myr old) with 1–2 obligately intracellular, vertically inherited bacterial species, and their study has yielded surprising discoveries about genome evolution and the minimal requirements for cell subsistence (Wernegreen 2015). These bacterial symbionts have often undergone dramatic losses of genes typically found in closely related free-living bacteria, and the limited set of genes they retain includes those that support their intracellular survival and their mutualism (e.g., primary metabolism and supplementation of host metabolism such as amino acid or cofactor biosynthesis) with their hosts. Many of the genes that are lost encode functions that have been inferred to be tangential to the mutualism (Moran et al. 2008; Kenyon and Sabree 2014; Wernegreen 2015). Genome reduction in insect endosymbionts has been largely characterized in hemipterans (i.e., Auchenorrhyncha and Sternorrhyncha) with strictly intracellular bacterial mutualists that are vertically transmitted within host tissues (intraegg to bacteriome), yet studies focusing on symbionts that are *extracellularly* and vertically transmitted, as observed in stink bugs (Pentatomomorpha; Heteroptera), are conspicuously lacking.

Phytophagous stink bugs are globally distributed, with some species being important agricultural pests (i.e.,

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Halyomorpha halys and *Nezara viridula*), and bacterial symbionts have been frequently detected in conspicuous distal midguts, typically referred to as the V4 or M4 region that contain invaginations or "crypts" that are specialized to house bacterial symbionts (Buchner 1965; Gordon et al. 2016; Hosokawa et al. 2016; Karamipour et al. 2016). These insects typically transmit their primary bacterial symbionts via nymphal consumption of symbiont-laden secretions that have been deposited on the surfaces of eggs by gravid females (Prado et al. 2006; Kikuchi et al. 2009; Kaiwa et al. 2010; Otero-Bravo and Sabree 2015; Salem et al. 2015). Primary symbionts traverse the gastric tract following their consumption to reside within the crypts of the V4 region (Buchner 1965; Gordon et al. 2016). Prolonged developmental time, aberrant behavior, and reduced fecundity have been observed in several stink bugs deprived of their primary symbiont and essential nutrient supplementation is among the functions that they have been hypothesized to perform (Abe et al. 1995; Prado et al. 2006; Kikuchi et al. 2009; Tada et al. 2011; Bistolas et al. 2014; Taylor et al. 2014; Duron and Noël 2016; Karamipour et al. 2016).

Molecular characterization of pentatomid symbionts, largely using the 16 S rRNA gene, has demonstrated that the primary symbionts of these insects are members of the *Pantoea* genus (Enterobacteriaceae; Gammaproteobacteria) (Bansal et al. 2014; Bistolas et al. 2014; Duron and Noël 2016). Successful efforts thus far to cultivate *Pantoea* symbionts from pentatomids have been limited to phyllospheric *Pantoea* species that could infect *Nezara viridula* digestive tissues (Esquivel and Medrano 2014) and four *Pantoea* species that cooccurred with one of two additional uncultivable *Pantoea* (Hosokawa et al. 2016). Interestingly, the draft genomes of the uncultivable *Plautia stali Pantoea* symbionts were up to 50% reduced in comparison to the cultivable *P. stali Pantoea* symbionts, which supports previous work demonstrating that primary insect mutualists with reduced genomes have lost the genes necessary for growth independent of their hosts (Wernegreen 2015). Unlike the relatively well-sampled primary insect mutualists of members of the Auchenorrhyncha and Sternorrhyncha, genomic sequencing of stink bug primary insect mutualists has been limited to a few exemplars, including those associated with *P. stali*, spanning several stink bug families: "*Can*. Ishikawaella capsulate" (*Megacopta cribaria*; Plataspididae) (Hosokawa et al. 2006), "*Can*. Tachikawaea gelatinosa" (*Urostylis westwoodii*; Urosylididae) (Kaiwa et al. 2014), and "*Can*. Pantoea carbekii" (*Halyomorpha halys*; Pentatomidae) (Kenyon et al. 2015). The paucity of primary insect mutualists genomes available for hosts, especially within a single insect family, precludes identifying patterns of genome evolution under the divergent symbiont transmission modality observed in stink bugs.

The *Edessa* is an exclusively neotropical stink bug genus that is comprised of over 300 species that exhibit significant morphological and ecological diversity, and the genus includes all but a few members of the subfamily Edessinae (Panizzi and Grazia 2015). Although the ecological roles of the *Edessa* are poorly described (Silva and Oliveira 2010; Panizzi and Grazia 2015), *E. rufomarginata* and *E. meditabunda* are generalists (Rizzo and Saini 1987) and potentially depend on several plants and distinct plant tissues for complete development (Panizzi and Machado-Neto 1992). Nonetheless, *Edessa* feed on plants that are likely limited in essential amino acids and vitamins, and thus may rely upon their primary bacterial symbionts to provision these and other nutrients limited in their host's diet.

Although symbiont transmission has not been previously documented in *Edessa*, maternal secretions on recently laid eggs, as is typical of pentatomids, was observed. Symbiont transmission by egg smearing is widespread throughout the Pentatomidae, and symbiont localization in the crypts has been documented for *E. bella* (Bistolas et al. 2014), both of which suggests symbiont transmission and localization to be likely the same for other *Edessa*. This study reports the de novo genome sequencing of the primary bacterial symbionts of four *Edessa* spp. (Pentatomidae). Phylogenetic and metabolic pathway inferences are used to describe *Edessa* primary symbiont evolution within the context of free-living *Pantoea* and previously described symbionts of stink bugs. This comparative genomic analysis of stink bug primary symbionts spans stink bug hosts within the *Edessa*, Pentatomidae and Pentatomorpha, and primary symbiont genome evolution is detailed in this relatively undersampled bacterial-insect symbiosis in the context of obligately intracellular insect endosymbionts.

Materials and Methods

Sample Collection and Sequencing

Four *Edessa* species local to the La Selva Biological Station, Costa Rica were studied: *E. eburatula*, *E. bella* (previously called "Edessa sp. nov 1" in Bistolas et al. 2014 and formally described in Fernandes Marin et al. 2015), *E. loxdalii*, and *Edessa n.* sp. Although the first two were previously characterized as each having a primary symbiont (Bistolas et al. 2014), the latter two edessines were found to harbor related symbionts in this study. Specimens were collected in June 2015, and preserved in 70% ethanol until dissection. Individuals were rinsed three times with filtered 70% ethanol before dissection, and the V4 region of the midgut was removed and subjected to DNA extraction using the DNEasy Blood and Tissue kit (Qiagen) with RNAse treatment. Illumina libraries were created using the Nextera XT DNA Library Prep kit and sequenced using Illumina MiSeq sequencer to generate 2×300-bp paired-end reads at the Ohio State University Molecular and Cellular Imaging Center.

Assembly and Annotation

Reads were evaluated with FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc; last accessed December 9, 2017) and trimmed using Trimmomatic 0.36 (Bolger et al. 2014) to trim Illumina indices and low quality bases. Individual libraries were assembled using SPAdes 3.5.0 (Bankevich et al. 2012). Bandage version 0.8.0 (Wick et al. 2015) was used to indicate possible plasmids as circular contigs, and infer connections between contigs and scaffolds based on the assembly deBruijn graphs (see Genome Assembly, Supplementary Material online). Additionally, all contigs were assigned to kingdom-level taxonomic bins by querying the NCBI GenBank "nt" database using BLASTN (default parameters) with the contigs and separated based on the kingdom of the subject sequence according to the NCBI GenBank "taxdb" database. Within each assembly, eight scaffolds could be distinguished from the rest due to read coverage binning and BLAST hit results, and, given that these scaffolds were linked in the deBruijn graphs, these results suggested that near-complete SoE genomes were captured (see Genome Assemblies, Supplementary Material online). Each assembly resulted in a draft genome comprised of six scaffolds ranging between 47 and 378 kb, each with an average per-base coverage ranging between 86× and 206×, and two shorter scaffolds (∼1,700 and ∼3,600 bp) of higher coverage (270× to 1,837× and 289× to 1,932×, respectively) that corresponded to the 16 S rRNA and the 5 S + 23 S rRNA gene regions (see Assembly Statistics, Supplementary Material online). No ambiguous bases were detected in scaffolds comprising the chromosomes or plasmids. As the ends of several of the genome-comprising scaffolds contained regions of the rRNA gene, it is likely that the scaffolds for each genome could not be resolved into circular chromosomes due to the rRNA gene operons being too long to allow in silico resolution of assembly conflicts. Putative plasmids were detected in all of the genome projects and were comprised of single scaffolds that had paired reads that mapped at the ends of the scaffold, had best BLAST hits either to Bacteria or previously identified plasmids and generally had average per base read coverage that was higher than observed in chromosomal scaffolds.

Coding regions were identified using Prokka (Seemann 2014), which uses Prodigal (Hyatt et al. 2010) for CDS prediction, RNAmmer (Lagesen et al. 2007) for ribosomal genes, Aragorn (Laslett and Canback 2004) for tRNAs, SignalP (Petersen et al. 2011) for signal peptides, and Infernal (Kolbe and Eddy 2011) for noncoding RNAs. Functional annotation was performed in Prokka with the UniProt (Consortium 2017), Pfam (Finn et al. 2014), and TIGRFAMs (Selengut et al. 2007) databases. Additional functional annotation was performed using the KEGG (Kanehisa et al. 2016) and COG (Galperin et al. 2015) databases with KAAS (Moriya et al. 2007) and cdd2cog (Leimbach 2016), respectively. Genomes and annotations were inspected and manually edited in Geneious v8 (Kearse et al. 2012). Metabolic reconstruction was performed using the KEGG and MetaCyc (Caspi et al. 2016) databases and web portals. Pseudogenes were identified by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) and by comparing annotated protein lengths to the reference length of orthologues in closely related bacteria (i.e., *Pantoea* spp.).

Metabolic reconstructions of the *Edessa* symbionts were compared with publicly available primary stink bug symbiont genomes: *P. carbekii* (GenBank accession number NZ_CP010907.1), "*Candidatus* Ishikawaella capsulate" (henceforth *I. capsulata*, NZ_AP010872.1), and "*Candidatus* Tachikawaea gelatinosa" (henceforth *T. gelatinosa*, NZ_AP014521.1). To compare the functional profiles of the *Edessa* symbiont genomes, we also generated the COG profiles with the same method for the five available complete reference genomes of noninsect associated *Pantoea*: *Pantoea agglomerans* (NZ_CP016889), *P. ananatis* (NC_013956), *P. stewartii* (NZ_CP017581), *P. rwandensis* (NZ_CP009454), and *P. vagans* (NC_014562). We compared the profiles using R v3.3.3 (R Core Team 2013) and used a two-tailed nested ANOVA for comparisons between symbionts and nonsymbionts.

Phylogenetic Reconstruction

Multilocus bacterial phylogenies using protein coding sequences from up to 35 members of the Enterobacteriaceae sensu lato (see supplementary table S2, Supplementary Material online, for genome accession numbers), including *Edessa* primary symbionts, were generated using RAxML (Stamatakis 2014) for maximum likelihood estimation, FastTree for approximate maximum likelihood (Price et al. 2010), and PhyloBayes (Lartillot et al. 2007) for Bayesian inference. Alignments on gene clusters were done using TranslatorX (Abascal et al. 2010) and inspected manually and concatenated in Geneious v8.

Bayesian and maximum likelihood phylogenies were reconstructed using eleven protein coding genes from all four *Edessa* primary symbionts and 35 nonprimary insect symbiont members of the Enterobacteriaceae to contextualize them within the Enterobacteriaceae (see supplementary material S1 and tables S3 and S4, Supplementary Material online, for assessment of the phylogenetic runs). Next, we inferred the placement of the *Edessa* primary symbionts within the *Pantoea* genus using maximum likelihood and approximate maximum likelihood methods on 322 shared protein coding regions between the *Edessa* primary symbionts and 122 *Pantoea* species and strains (see supplementary material S6–S10, Supplementary Material online). Additionally, individual *Edessa* primary symbiont phylogenies were reconstructed using a subset of representative Enterobacteriaceae species exclusive of all other reduced genome insect symbionts to

address potentially confounding effects of using one or more symbionts in the inference of the position of the other symbionts (Husník et al. 2011) (see supplementary material S6–S10, Supplementary Material online).

Genome Synteny and Average Nucleotide Identity

Inter-*Edessa* primary symbiont pairwise genome alignments were made using LastZ (parameters: step size = 20, –nogapped –notransition) and plotted using R to identify syntenic regions (Harris 2007). Similar methods were used to identify syntenic regions between an exemplar *Edessa* primary symbiont and four noninsect primary symbiont *Pantoea* species (i.e., *P. stewartii, P. agglomerans, P. rwandensis*, and *P. ananatis*) and *P. carbekii.* Average Nucleotide Identity (ANI) was calculated for whole genomes using JSpecies (Richter and Rosselló-Móra 2009) with the "nucmer" option as a way to assess evolutionary divergence from the previously named *Pantoea* species.

Mutation Rates

Synonymous (dS) and nonsynonymous (dN) substitution rate values were calculated for all shared genes between the four symbionts of *Edessa*, as well as *P. agglomerans, P. ananatis, P. stewartii*, and *P. rwandensis.* Genes were identified by using reciprocal best hit BLAST searches between all pairs of genome, custom Python and R scripts, and PAML (Yang 2007), using codeml (parameters: runmode -2, codonfreq 2, nssites 0, model 2). For each of these we generated the maximum likelihood nonrooted RAxML tree from five tree searches. The omega ratio (dN/dS) and dS values were compared between the symbionts of the *Edessa* and the noninsect associated *Pantoea*. Genes with saturated values (>3.0) were excluded from the analysis. Additionally, the COG category as annotated by the COG database was used to evaluate differences according to gene function. Two tailed nested ANOVA tests were done in R to evaluate significant differences between symbiotic and nonsymbiotic taxa.

Results and Discussion

General Features of the *Edessa* Primary Symbiont Genomes

Full genome data were obtained for bacterial symbionts associated with four *Edessa* species, which are hereafter referred to collectively as the "Symbionts of *Edessa*" (SoE) and individually as SoEL (symbiont of *E. loxdalii*), SoEE (symbiont of *E. eburatula*), SoEO (symbiont of *E. bella*), and SoET (symbiont of *Edessa n* sp.). The SoE chromosome sizes ranged from 0.81 to 0.83 Mb, and three plasmids, pSOE1 (12.8–13.7 kb), pSOE2 (3.4–4.3 kb), and pSOE3 (2.9–3.5 kb) were identified in all four *Edessa* genome projects, whereas a fourth plasmid, pSOE4 (2.6–3.0 kb), was identified only in SoEL, SoEE, and SoEO. Gene content and order was highly conserved across all plasmids, which suggests that these plasmids were present in the ancestral symbiont prior to host diversification. Eight scaffolds were consistently recovered in each assembly that comprised the chromosomal genome and regions of the highly conserved rRNA operons flanked each scaffold, which consistently confounded assembly of single circular chromosomes. The SoE genomes were A + T%-biased (26.9–27.8% G + C%) relative to free-living *Pantoea* and other stink bug associated symbionts (fig. 1), and they were slightly larger than more distantly related pentatomorpha stink bug primary symbionts (e.g., *I. capsulata*: 0.75 Mb; *T. gelatinosa*: 0.71 Mb) but smaller than the more closely related than other pentatomini primary symbiont genomes (e.g., *P. carbekii*: 1.2 Mb; *Plautia stali*: 2.4–5.5 Mb) (table 1). Protein coding density in SoE genomes was greater than in *P. carbekii* and generally more similar to the other pentatomorpha stink bug primary symbionts.

The SoE Represent a Distinct, Rapidly Evolving *Pantoea* Species

Phylogenies inferred by two methods (RAxML and Phylobayes) consistently placed the common ancestor of the *Edessa* symbionts within the Enterobacteriaceae sensu lato, what is now considered the Erwiniaceae (Adeolu et al. 2016). They are firmly placed among the *Pantoea* (fig. 2), similarly to *P. carbekii*, the various symbionts of *P. stali*, and previous phylogenetic reconstructions of other stink bug symbionts (Bistolas et al. 2014; Kenyon et al. 2015; Duron and Noël 2016). In order to disentangle the effects of long branch attraction artifacts, several *Pantoea* phylogenies were reconstructed using a jackknife or "leave-one-out" approach, where a single SoE symbiont was used to evaluate if the SoE shared a recent common ancestor with *P. carbekii* (see supplementary material S2–S5 and S7–S10, Supplementary Material online). By this approach, all four SoE symbionts were consistently placed at the base of a *P. ananatis–P. stewartii–P. agglomerans* clade in trees constructed either with individual or all SoE symbionts, but excluding *P. carbekii* (fig. 3). Meanwhile, similar phylogenies generated with *P. carbekii*, but excluding all SoE symbionts, placed *P. carbekii* in a distinct clade closer to *P. dispersa*, *P. rodasii*, and *P. rwandensis* (supplementary fig. S6, Supplementary Material online), indicating that this symbiont would have originated from a different *Pantoea* clade. Although *Pantoea* phylogenies that included both SoE and *P. carbekii* yielded a single clade comprised exclusively of these taxa (supplementary figs. S1 and S6, Supplementary Material online), the authors warn against inferring ancestral relationships from these results as this is likely a long branch attraction artifact, which is common in insect symbiont phylogenies and distorts conclusions made about their relatedness (Ruano-Rubio and Fares 2007; Husnik et al. 2013).

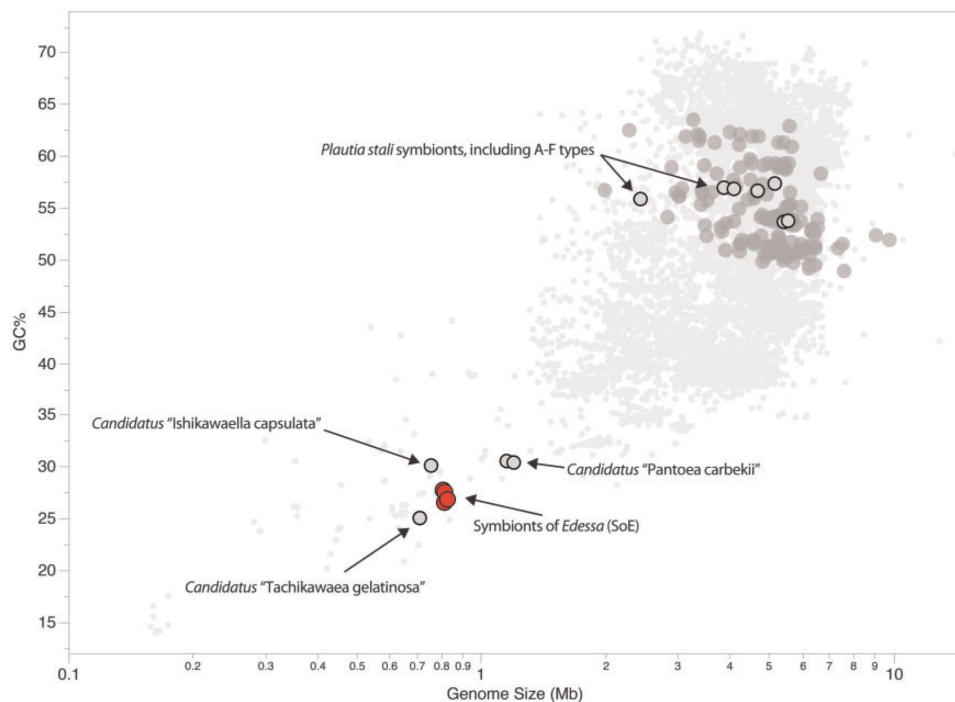

Fig. 1.—A + T% bias in *Edessa* primary symbiont genomes. G + C% for symbionts of *Edessa* genomes (red dots) were plotted with similar information from the genomes of other gammaproteobacterial stink bug symbionts (black outlined light gray circles), nonobligately host-associated and free living *Pantoea* species (dark gray circles) and other gammaproteobacteria that span several known ecological niches (small light gray circles). Data were generated from 40,356 gammaproteobacterial genomes deposited NCBI GenBank as of October 2017.

The branch of the SoE is significantly longer than the branch for any other nonsymbiotic *Pantoea* indicating an accelerated rate of evolution, which is characteristic of genome reduced symbionts (Wernegreen 2015). Synonymous substitutions along the branch leading up to all four SoE were found to be saturated (dS > 3.0), disallowing direct comparisons to nonsymbiotic members. Comparisons between rates found between symbionts and between distantly related nonsymbiotic species of *Pantoea* yielded no significant differences between groups when separated by COG category (*P* value ≫0.05) except for category R—General Function Prediction Only. However, taken together, the omega values from the SoE are slightly higher than for the noninsect associated *Pantoea* (supplementary fig. S12, Supplementary Material online). No single gene showed a dN/dS value >1. This does not necessarily mean that no genes are under strong selection, but is consistent with hypotheses implicating genetic drift as having a significant impact on insect symbiont genome evolution.

SoE Genome Exhibit Considerable Intraspecific Structural Conservation

All four SoE genomes were nearly completely syntenic to each other and exhibited a highly conserved gene order, with a single 25 kb inversion in scaffold 1 of SoEL being observed (fig. 4*A*). Conversely, gene order was not conserved when was observed when an exemplar of the SoE genomes (SoEE) was compared with *P. carbekii* (fig. 4*B*). The many observed rearrangements between the SoE and *P. carbekii* genomes are likely due to recombination events that in distinct ancestors preceding their host association, as observed in leaf nodule symbionts (Pinto-Carbó et al. 2016), and this supports the hypothesis that SoE represents an origin of symbiosis that is independent of *P. carbekii* and its likely nonhost-associated ancestor. Alignments of the *Edessa* symbionts were compared with alignments made for nonhost restricted *Pantoea* (fig. 4*C*), and although a general trend of gene order is evident, considerable gene order shuffling was observed.

Conservation of gene order has been observed for several obligate intracellular insect symbionts with slight or no deviations from complete collinearity (Tamas et al. 2002; Degnan et al. 2005; Sabree et al. 2010; Patiño-Navarrete et al. 2013) and has been attributed to the loss of phages, mobile elements, and genes encoding recombination proteins, and/or with virtually no interaction with other bacterial species due to their host-restricted lifestyle. Extracellular symbionts, on an average, have fewer mobile elements than facultative intracellular symbionts, but more than obligate intracellular symbionts (Newton and Bordenstein 2011). However, the absence of recombination events among the SoE suggests general stability despite being extracellular symbionts.

Table 1
Genome Statistics for Stinkbug Symbionts, Members of the Pantoea, and Buchnera

Bacteria	Association	Insect Host	Symbiont Genome Size[b] (Mb)	GC%	Coding Density (%)	Coding CDS	Reported Plasmids	Pseudogenes	tRNAs	Ribosomal RNAs
Pantoea carbekii	Vertically inherited insect symbiont	Halyomorpha halys (Pentatomidae; Pentatominae)	1.15	30.6	67.40	829	4	12	40	8
SoET	Vertically inherited insect symbiont	Edessa sp. 2 (Pentatomidae; Edessinae)	0.81	27.8	81.79	677	4	11	35	6[a]
SoEO	Vertically inherited insect symbiont	Edessa bella (Pentatomidae; Edessinae)	0.81	26.6	80.76	668	4	9	35	6[a]
SoEL	Vertically inherited insect symbiont	Edessa loxdalii (Pentatomidae; Edessinae)	0.82	27.6	83.67	689	4	8	35	6[a]
SoEE	Vertically inherited insect symbiont	Edessa eburatula (Pentatomidae; Edessinae)	0.83	26.9	83.71	698	3	11	35	6[a]
Ishikawaella capsulata	Vertically inherited insect symbiont	Megacopta spp. (Plataspididae)	0.75	30.2	82.10	620	1	35	37	9
Tachikawaea gelatinosa	Vertically inherited insect symbiont	Urostylis spp. (Urostylididae)	0.71	25.1	85.16	613	0	8	35	9
P. ananatis	Nonhost-restricted	–	4.70	53.7	86.69	4,282	–	153	71	22
P. rwandensis	Nonhost-restricted	–	4.33	53.9	87.51	3,941	–	57	79	22
P. vagans	Nonhost-restricted	–	4.02	55.5	87.29	3,670	–	60	80	22
P. agglomerans	Nonhost-restricted	–	4.18	55.5	86.92	3,844	–	213	76	22
Type-A symbiont of Plautia stali	Vertically inherited insect symbiont	Plautia stali (Pentatomidae; Pentatominae)	3.87	57.0	NA	3890	2	–	–	–
Type-B symbiont of Plautia stali	Vertically inherited insect symbiont	Plautia stali (Pentatomidae; Pentatominae)	2.43	55.9	NA	NA	–	–	–	–
Type-C symbiont of Plautia stali	Insect symbiont	Plautia stali (Pentatomidae; Pentatominae)	5.14	57.4	NA	4882	–	–	–	–
Type-D symbiont of Plautia stali	Insect symbiont	Plautia stali (Pentatomidae; Pentatominae)	5.54	53.8	NA	5311	–	–	–	–
Type-E symbiont of Plautia stali	Insect symbiont	Plautia stali (Pentatomidae; Pentatominae)	5.41	53.7	NA	5064	–	–	–	–
Type-F symbiont of Plautia stali	Insect symbiont	Plautia stali (Pentatomidae; Pentatominae)	4.67	56.7	NA	4457	–	–	–	–
Buchnera aphidicola str. APS (Acyrthosiphon pisum)	Vertically inherited insect symbiont	Aphids	0.66	26.4	88.00	617	2	1	32	3

NOTE.—Gray indicates previously sequenced stinkbug symbionts.
[a]Based on the number of RNAs at the edges of contigs.
[b]Symbiont genome size is estimated from the best assembly.

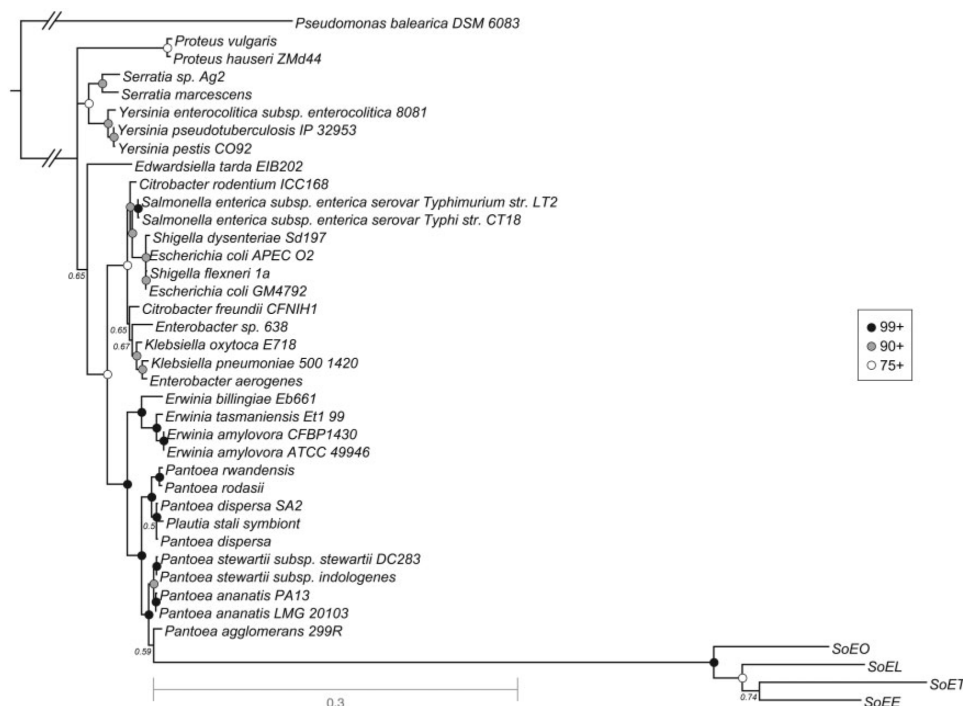

Fig. 2.—Phylogenetic reconstruction of Enterobacteriaceae. Tree was based on 11 coding sequences run using PhyloBayes model GTR + CAT with six-categories Dayhoff recoding. Support values are posterior probabilities labeled as dots as shown in the legend.

Although gene order conservation tends to be the trend in intracellular symbionts, several factors can alter this stability even in these host-restricted symbionts, such as the presence of large repetitive intergenic regions as in *Portiera* (Sloan and Moran 2013), long and variable host life as in *Hodgkinia* (Campbell et al. 2015), and genome fusion and posterior degeneration as in *Tremblaya* (Gil et al. 2018).

Bacterial symbionts of the pentatomidae are typically vertically inherited via consumption of maternal secretions by nymphs (reviewed in Otero-Bravo and Sabree 2015), and these symbionts must persist on the egg surfaces until consumption and must traverse the digestive tract to reach their residence within the distal midgut. During both of these stages of symbiont transmission there are possibilities for encountering other bacteria and recombination and/or gene flow could arise, yet no evidence this was observed in the SoE genomes. Furthermore, while extracellular transmission mechanisms could increase the chances of SoE coming in contact with other related bacteria as well as open the possibility of symbiont replacement, the observed genomic reduction in SoE suggests strong pressure from the host to maintain strict inheritance or a mechanism for symbiont sorting, as observed in bean bugs (Ohbayashi et al. 2015). Further details into the transmission of stink bug symbionts and their establishment in the midgut crypts would be incredibly informative toward understanding how symbionts with reduced genomes exhibit high fidelity to their host.

SoE Are Enriched in Functions Supportive of Host–Symbiont Association

When COG profiles of nonhost-associated *Pantoea* were compared with those of the SoE, significantly greater relative representation in the following categories were observed: general function prediction only, amino acid transport and metabolism, translation, ribosomal structure and biogenesis, energy production and conversion, inorganic ion transport and metabolism, coenzyme transport and metabolism, replication, recombination and repair, posttranslational modification and chaperones, and nucleotide transport and metabolism (fig. 5). Most of these can be explained by the host-restricted lifestyle, where only the most vital and necessary functions are maintained while those encoding secondary functions are lost (Kenyon and Sabree 2014; Moran and Bennett 2014). This was exhibited in the SoE genomes where they had a lower proportion of genes dedicated to the categories of carbohydrate metabolism and transport, function unknown, transcription, cell wall, cell membrane and envelope biogenesis, signal transduction, cell motility, intracellular trafficking, secretion, and vesicular transport. The loss of genes that provide these functional redundancies concurrent with many other loci tangential to the mutualism, and the retention of the relatively few genes essential for amino acid metabolism, has resulted in an overall dramatic reduction in the genome sizes of SoE and the preservation of the skewed proportion of genes in SoE genomes that support the mutualism.

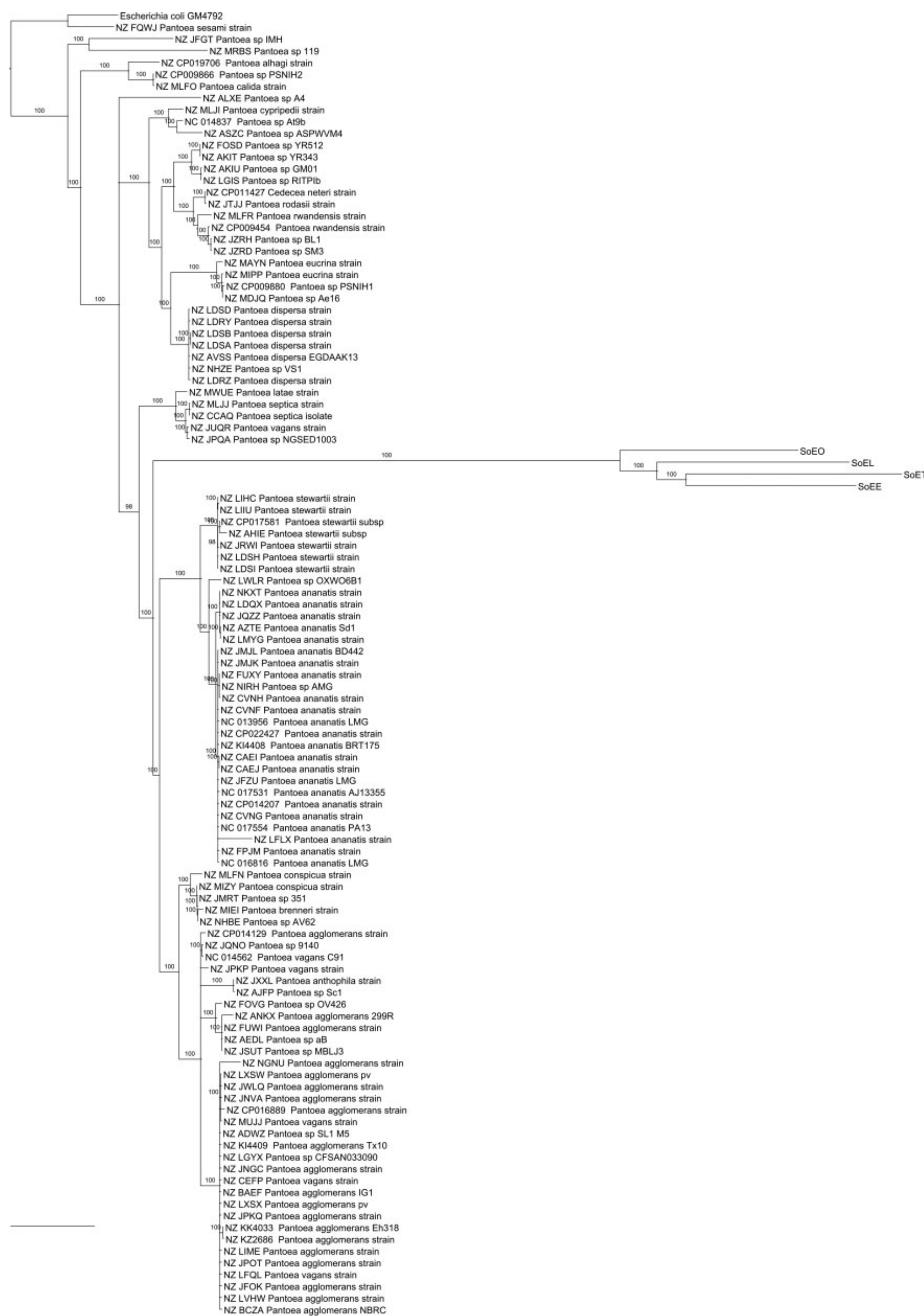

Fig. 3.—Core genome phylogenetic reconstruction of the *Pantoea* genus including the Symbionts of *Edessa*. Maximum likelihood reconstruction using RAxML for 322 loci obtained from reciprocal best hit BLASTs. Support values are based on 50 bootstrap iterations. Colors indicate highly supported clades among all reconstructions. Scale bar is proportional to 0.05 substitutions per site.

SoE Can Provide Several Amino Acids, Vitamins, and Cofactors

Stink bug symbionts are thought to generally provide the host with amino acids, vitamins, and cofactors that are underrepresented in their herbivorous diets (Hosokawa et al. 2006; Kikuchi et al. 2009; Nikoh et al. 2011; Kaiwa et al. 2014; Kenyon et al. 2015), and all of the SoE genomes encode complete or near complete canonical pathways for most essential amino acids, including branched chain amino acids (valine, isoleucine, and leucine) whose pathways are nearly complete except for the lack of the *ilvE* gene encoding the terminal aminotransferase (fig. 6). This gene is conspicuously

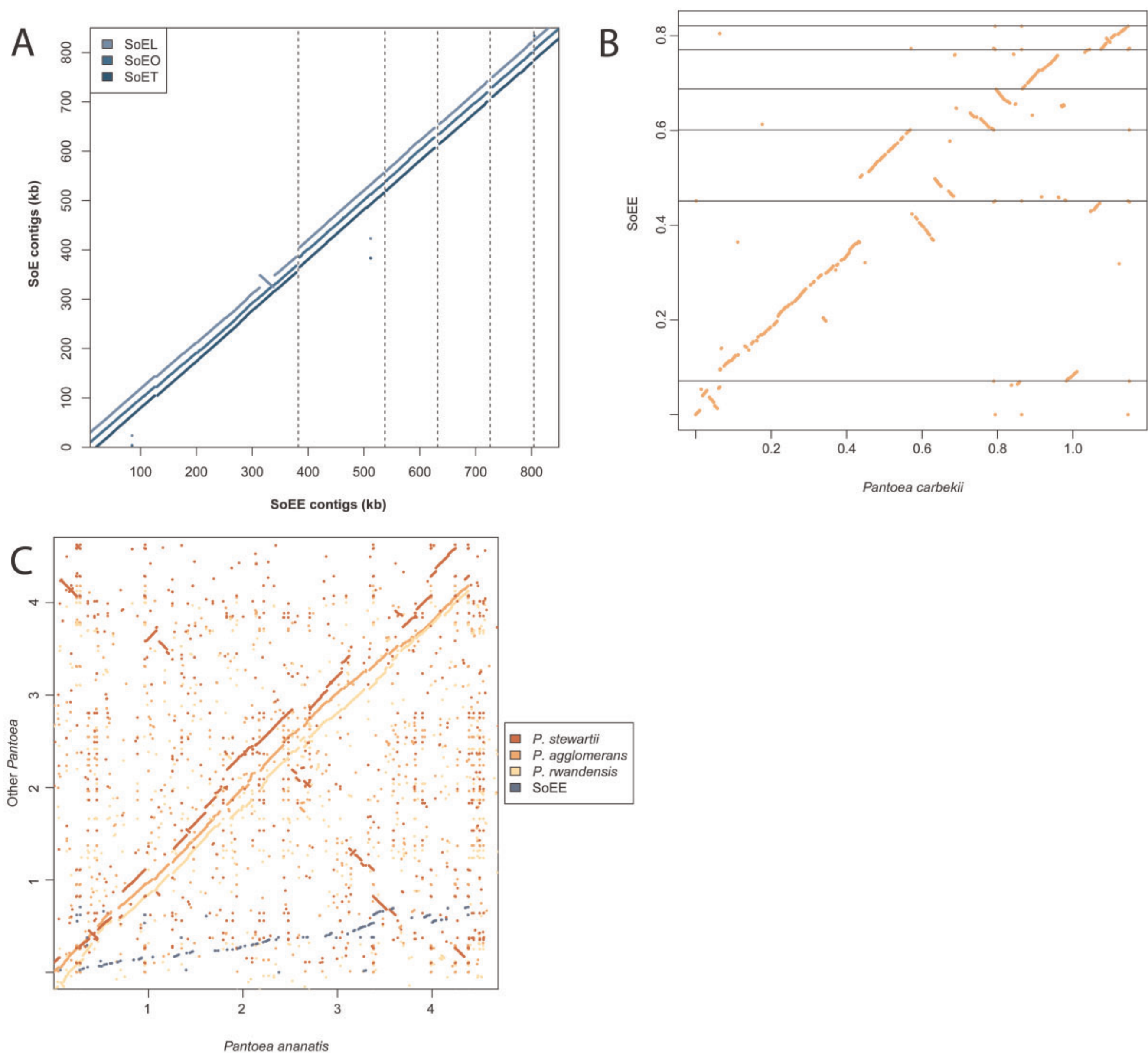

Fig. 4.—Genome synteny for the Symbionts of Edessa. Dotplots showing synteny between the four strains (*A*), synteny between SoEE and *Pantoea carbekii* (*B*), and synteny between nonhost restricted strains of *Pantoea* and SoEE (*C*).

absent in the genomes of multiple other insect mutualists, including *P. carbekii, I. capsulata, T. gelatinosa, Buchnera*, and *Candidatus* "Uzinura aspidicola," and it has been shown that a host encoded equivalent is up-regulated in the bacteriocytes of the hosts for intracellular symbionts (Hansen and Moran 2011; Husnik et al. 2013; Luan et al. 2015). *Edessa* primary symbionts can also synthetize some nonessential amino acids (i.e., glutamate, aspartate, alanine, cysteine, tyrosine) and notable gene deletions in several pathways have been observed. Canonical proline biosynthesis requires the products of *proA, proB, proC*, and *putA*, however, the latter two genes are absent in all four genomes, a pattern that is mirrored in *P. carbekii* (Kenyon et al. 2015). Although SoET and SoEL encode the complete canonical serine biosynthesis pathway that requires the products of *serA, serB*, and *serC*, the first two genes are either absent or pseudogenized in SoEO and SoEE. *P. carbekii, I. capsulata, T. gelatinosa, Buchnera, Baumannia*, and *Blochmania* also lack intact versions of either or both *serA* and *serB*. In SoEE, *serB* is missing, *serC* is retained, but *serA* appears to be pseudogenized, where four shortened ORFs can be found in the region where *serA* is found in SoET and SoEL. It is unclear if the products of these four smaller ORFs fulfill the role of SerA for this symbiont, but shows a potentially recent pseudogenization. In the

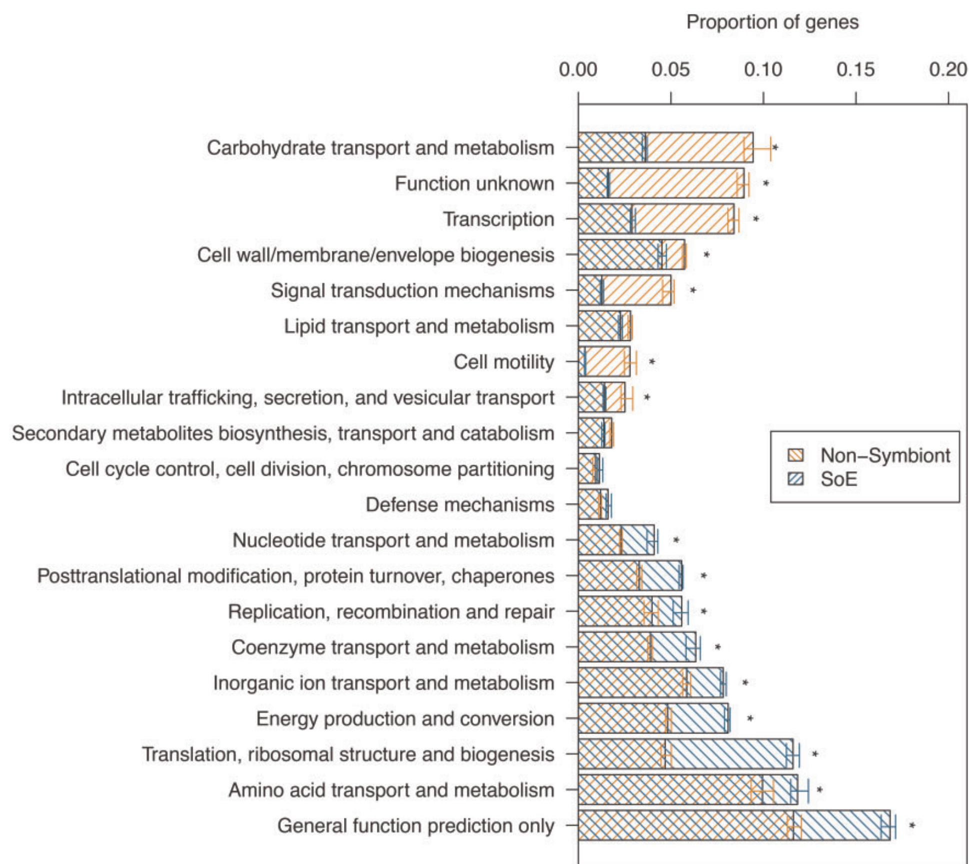

Fig. 5.—COG Profiles of SoE compared with other *Pantoea*. The proportion of genes belonging to each category is shown. Asterisks indicate significant differences in a two-tailed nested ANOVA.

case of asparagine, only SoEE encodes asparagine synthetase B (*asnB*), which is sufficient for the asparagine biosynthesis from aspartate. This enzyme is also absent in several other insect symbionts, including *I. capsulata* and *T. gelatinosa*, while being present in *Wigglesworthia*, and is pseudogenized in *P. carbekii* (Kenyon et al. 2015).

Arginine biosynthesis is accomplished in SoE by a nearly canonical pathway that includes the replacement of the acetylornithine aminotransferase ArgD with the succinylornithine transaminase AstC, which also participates in ornithine degradation (Kim and Copley 2007). This replacement appears to be consistent across the other stink bug symbionts and it is notable that enzymes that are inferred to participate in multiple metabolic pathways are retained during SoE genome reduction. This phenomenon has been observed in other bacterial endosymbionts of insects (e.g., *Sulcia muelleri* and *Sodalis;* Koga and Moran 2014) and reflects a convergence upon this evolutionary trajectory.

Thiamine and biotin can be generated by all of the SoE, with the former being an essential cofactor in all three domains of life (Costliow and Degnan 2017) and the latter being a key metabolite underlying other insect-microbe symbioses (Nikoh et al. 2014). Although all the genes for the biosynthesis of biotin are present on the SoE chromosomes, with several being within a small inverted region, several genes for thiamine biosynthesis (*thiCDEFGHS*) are on the plasmid pSOE1 that is present in all of the SoE genomes. *P. carbekii* is able to produce thiamine but not biotin, whereas the reverse is true for *I. capsulata* and *T. gelatinosa*. Interestingly, among stink bug symbionts, *T. gelatinosa* is the only one encoding a full thiamine transporter despite sharing the absence of the biosynthesis pathway with *I. capsulata*. Additionally, all SoE encode nearly all of the enzymes for producing folate, another critical vitamin for life, yet they lack *nudB* (dihydroneopterin triphosphate pyrophosphohydrolase), *pabC* (aminodeoxychorismate lyase), and *pabB* (4-amino-4-deoxychorismate synthase), which catalyze the committed step of folate synthesis in bacteria (Gabelli et al. 2007) and chorismate incorporation into folate biosynthesis, respectively. *nudB* is also missing in *P. carbekii*, *I. capsulata*, and *T. gelatinosa*, whereas *pabC* and *pabB* are missing in *T. gelatinosa.* Finally, all of the SoE encode pathways for riboflavin biosynthesis except for lacking the 5-amino-6-(5-phospho-D-ribitylamino)uracil

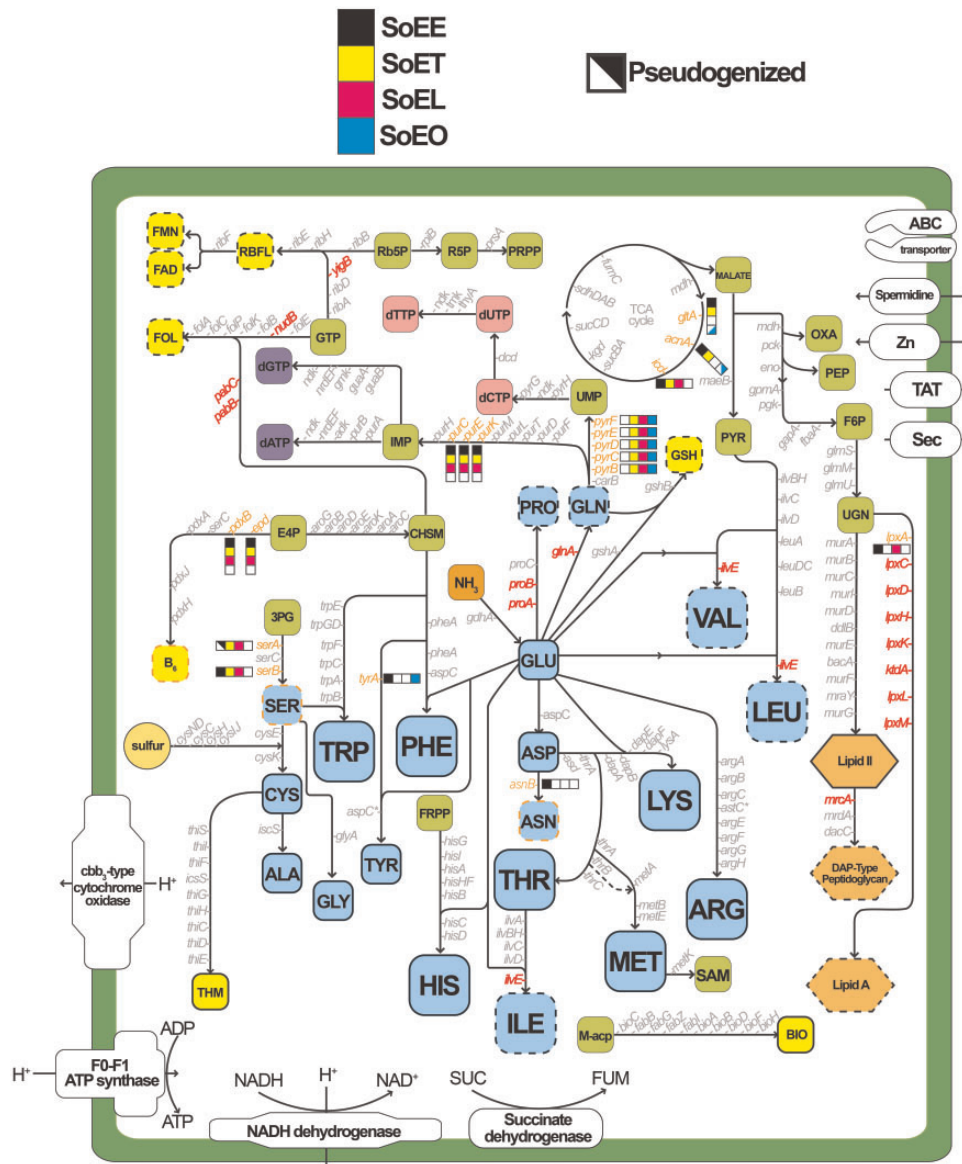

Fig. 6.—Metabolic reconstruction of the Symbionts of *Edessa*. Boxes indicate metabolites or products of bacterial metabolism. A solid black outline indicates all SoE contain the genes in the canonical pathway for the synthesis that product, a dashed black outline indicates no SoE contains all canonical enzymes for the synthesis of that product, whereas a dashed orange outline indicates some, but not all SoE contain all canonical enzymes for the synthesis of that product. Genes along an outline are colored gray if present, orange if present in some, or red if absent in all. Genes in orange also include the indication of in which SoE genomes they are present, absent, or pseudogenized. Large blue boxes indicate essential amino acids while small blue boxes indicate nonessential amino acids. Pink boxes indicate pyrimidines. Purple boxes indicate purines. Yellow boxes indicate vitamins and cofactors. Green boxes indicate other metabolites. 3PG, 3-phosphoglycerate; B6, vitamin B6; BIO, biotin; CHSM, chorismate; dATP, deoxyadenosine triphosphate; dCTP, deoxycitidine triphosphate; dGTP, deoxyguanosine triphosphate; dTTP, deoxythymidine triphosphate; dUTP, deoxyuridine triphosphate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; FAD, flavin adenine dinucleotide; FMN, flavin mononucleotide; FOL, folate; GSH, glutathione; GTP, guanosine 5′ triphosphate; IMP, inosinic acid; M-acp, malonyl-ACP; OXA, oxoaloacetate; PEP, phosphoenolpyruvate; PRPP, phosphoribosylpyrophosphate; PYR, pyruvate; R5P, ribose-5-phosphate; Rb5P, ribulose 5-phosphate; RBFL, riboflavin; SAM, s-adenosyl methionine; THM, thiamine; UGN, UDP-N-acetyl glucosamine; UMP, uridine monophosphate.

phosphatase YigB, which is present in other stink bug symbionts, yet several other phosphatases are encoded on the genomes that may complement this missing function. That many of these genes are consistently retained in light of dramatic gene losses observed across the SoE genomes suggests strong host selective pressures to retain these function, which similar to what is observed in endocellular insect symbionts.

SoE Exhibit Intraspecific Variation and Intersymbiont Convergence in Carbon Metabolism

The TCA cycle shows an interesting variation among the SoE and other symbionts. Although SoEE and SoET contain the full TCA cycle, SoEL lacks *gltA*, *acnA*, and *icd*, catalyzing half of the cycle from oxoaloacetate to 2-oxoglutarate. Additionally, SoEO also lacks *icd*, but both *gltA* and *acnA* are pseudogenized, the first being split into four coding regions, whereas the second one has an early stop codon. Whether the proteins encoded by the coding regions remaining in *gltA* can perform the role of the complete enzyme is unknown. This case shows how among genomes from four closely related symbionts, three stages of potential genomic degeneration of an important metabolic pathway are observed. Although *P. carbekii* and *I. capsulata* retain all the enzymes of the TCA cycle, *T. gelatinosa* has lost almost the entire pathway while retaining genes encoding SucAB that catalyze the single step between 2-oxoglutarate and succinyl-CoA. Additionally, the intergenic region between the genes surrounding *icd* is only 500∼ bp in SoEL, and a BLAST search against the "nt" database results in no hits with high query coverage, whereas the same intergenic region in SoEO is 1,500∼ bp and a BLAST search results in the *icd* protein in *P. carbekii* and *I. capsulata*, which is likely due to a recent loss.

Given that glutamate dehydrogenase (GdhA) can generate 2-oxoglutarate/alpha-ketoglutaric acid from glutamate, it is possible for SoE, all of whom encode *gdhA* on their plasmids, to derive sufficient energy from their incomplete TCA cycles (Koga and Moran 2014).

SoE genomes also exhibit convergence with other insect symbionts in that all SoE retain only the nonoxidative branch of the pentose phosphate cycle, which is similar to *Wigglesworthia* but unlike *P. carbekii* and *I. capsulata* that both retain the full cycle. Additionally, all SoE encode full pathways for glycolysis and the production of G3P from glucose-6-phosphate, yet lack phosphoglucomutase (*pgm*) for the transfer of the phosphate group in glucose, which is reflected in *I. capsulata*, *T. gelatinosa*, and *Buchnera*, and distinct from *P. carbekii* who retains genes encoding the entire pathway.

Proposal of Candidate Names

The SoE have been shown to be exclusively found in the midgut crypts of the *Edessa*, and comparisons to other primary symbionts of stink bugs, such as *P. carbekii, I. capsulata*, and *T. gelatinosa*, reveal several common motifs of primary unculturable symbionts. The symbiont loci recovered in this study show near perfect identity to those previously sequenced (Bistolas et al. 2014), with individuals collected 3 years apart, indicating their stability in their host's population. A protocol for distinguishing the SoE from other stink bug symbionts through FISH microscopy has established (Bistolas et al. 2014), and the reported multilocus phylogenies and extensive genome sequencing and analyses all suggest that SoE represent a distinct *Pantoea* species. These data satisfy the recommendations for the naming of *Candidatus* microbial species (Stackebrandt et al. 2002), and the description of several strains is also contained in this paper. Additionally, they satisfy the conditions recently proposed for uncultivated microbes (Konstantinidis et al. 2017) that require a near-complete genome sequence with no contamination, ecological data on the microbe's habitat, reliable rDNA sequence, as well as a picture of the microorganism. Based on these conditions, we propose the candidate name "*Candidatus* Pantoea edessiphila" for the symbionts of the *Edessa*, due to their close association with their hosts, and we also propose strain designations to distinguish the pentatomid host with which it is associated: "*Candidatus* Pantoea edessiphila" strain SoEL (symbiont of *E. loxdalii*); "*Candidatus* Pantoea edessiphila" strain SoEE (symbiont of *E. eburatula*), "*Candidatus* Pantoea edessiphila" strain SoEO (symbiont of *E. bella*); and "*Candidatus* Pantoea edessiphila" strain SoET (symbiont of *Edessa* sp.).

Conclusion

Genomic analysis of the symbionts of *Edessa* pentatomid stink bugs have revealed several typical signatures of strong host association and vertical inheritance (i.e., significant genome reduction, retention of mutualism-supportive genes, A + T% bias), yet these features are observed in symbionts that neither reside within host cells or are perpetually within host tissues. As genome reduction has been observed in extracellular gut symbionts of different stinkbugs that maintain stable relationships with their hosts, it has been suggested that endocellularity is not fundamental to this phenomenon (Hosokawa et al. 2016). Both phylogenetic and genomic structural analyses support the placement of these symbionts within a novel clade among the *Pantoea* that is distinct from previously sequenced stink bug symbionts. Among the *Edessa* symbionts, metabolic capabilities exhibit minimal variation and overall gene and metabolic pathway losses parallel genome streamlining observed in distant, unrelated insect symbionts. Characterization of additional pentatomid stink bug symbiont genomes will facilitate determining the impact of endocellularity and transmission modality on genome evolution. Furthermore, bacterial-pentatomid stink bug symbioses have occurred multiple times and exhibit varying levels host fidelity, which provides a rich opportunity to examine the genome dynamics of symbionts associated with closely related host insects.

Exploiting the availability of both insects and low-cost sequencing technology facilitates the analysis of symbionts of closely related insects to infer a more robust depiction of both the relationship between the bacterial symbionts and their hosts and the evolutionary histories of the bacterial symbionts as written in their genomes.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The authors want to express deep gratitude to Kalia Bistolas for collecting many of the specimens used in this research. We also want to thank La Selva Biological Research Station and staff, as well as the Organization of Tropical Studies (OTS) and CONAGEBio, Dr Fernandez Marin for his help with identification of specimens. Lastly, we thank the anonymous reviewers for their thoughtful comments and reviews. This work was supported by the Ohio Agricultural Research Development Center SEEDS Research Enhancement-Interdisciplinary Grant, The Ohio State University, the computation resources of the Ohio Biodiversity Conservation Partnership (OBCP) cluster, and the Colombian Administrative Department of Science, Technology and Innovation COLCIENCIAS.

Literature Cited

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38(Web Server issue):W7–13.

Abe Y, Mishiro K, Takanashi M. 1995. Symbiont of brown-winged green bug, *Plautia stali* SCOTT. Jpn J Appl Entomol Zool. 39(2):109–115.

Adeolu M, Alnajar S, Naushad S, Gupta RS. 2016. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morgane. Int J Syst Evol Microbiol. 66(12):5575–5599.

Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19(5):455–477.

Bansal R, Michel A, Sabree Z. 2014. The crypt-dwelling primary bacterial symbiont of the polyphagous pentatomid pest *Halyomorpha halys* (Hemiptera: pentatomidae). Environ Entomol. 43(3):617–625.

Bistolas KSI, Sakamoto RI, Fernandes JaM, Goffredi SK. 2014. Symbiont polyphyly, co-evolution, and necessity in pentatomid stink bugs from Costa Rica. Front Microbiol. 5:1–15.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114–2120.

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

Campbell MA, et al. 2015. Genome expansion via lineage splitting and genome reduction in the cicada endosymbiont Hodgkinia. Proc Natl Acad Sci U S A. 112(33):10192–10199.

Caspi R, et al. 2016. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 44(D1):D471–D480.

Consortium U. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45(D1):D158–D169.

Costliow ZA, Degnan PH. 2017. Thiamine acquisition strategies impact metabolism and competition in the gut microbe bacteroides thetaiotaomicron. mSystems 2(5):e00116-17.

Degnan PH, Lazarus AB, Wernegreen JJ. 2005. Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. Genome Res. 15(8):1023–1033.

Duron O, Noël V. 2016. A wide diversity of Pantoea lineages are engaged in mutualistic symbiosis and cospeciation processes with stinkbugs. Environ Microbiol. Rep. 8:715–727.

Esquivel JF, Medrano EG. 2014. Ingestion of a marked bacterial pathogen of cotton conclusively demonstrates feeding by first instar southern green stink bug (Hemiptera: pentatomidae). Environ Entomol. 43(1):110–115.

Fernandes Marin JA, Juliete Da Silva V, Oliveira Correia A, Mendes Nunes B. 2015. New species of Edessa Fabricius, 1803 (Hemiptera: pentatomidae) from Costa Rica. Zootaxa 3999:511–536.

Finn RD, et al. 2014. Pfam: the protein families database. Nucleic Acids Res. 42(Database issue):D222–D230.

Gabelli SB, et al. 2007. Structure and function of the *E. coli* dihydroneopterin triphosphate pyrophosphatase: a Nudix enzyme involved in folate biosynthesis. Structure 15(8):1014–1022.

Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43(D1):D261–D269.

Gil R, et al. 2018. Tremblaya phenacola PPER: an evolutionary beta-gammaproteobacterium collage. ISME J. 12:124–135.

Gordon ERL, McFrederick Q, Weirauch C. 2016. Phylogenetic evidence for ancient and persistent environmental symbiont reacquisition in Largidae (Hemiptera: heteroptera). Appl Environ Microbiol. 82(24):7123–7133.

Hansen AK, Moran NA. 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc Natl Acad Sci U S A. 108(7):2849–2854.

Harris RS. 2007. Improved pairwise alignment of genomic DNA [PhD thesis]. The Pennsylvania State University.

Hosokawa T, et al. 2016. Obligate bacterial mutualists evolving from environmental bacteria in natural insect populations. Nat Microbiol. 1:15011.

Hosokawa T, Kikuchi Y, Nikoh N, Shimada M, Fukatsu T. 2006. Strict host-symbiont cospeciation and reductive genome evolution in insect gut bacteria. PLoS Biol. 4(10):e337–1851.

Hosokawa T, Matsuura Y, Kikuchi Y, Fukatsu T. 2016. Recurrent evolution of gut symbiotic bacteria in pentatomid stinkbugs. Zool Lett. 2:24.

Husník F, Chrudimský T, Hypša V. 2011. Multiple origins of endosymbiosis within the Enterobacteriaceae (γ-Proteobacteria): convergence of complex phylogenetic approaches. BMC Biol. 9:87.

Husnik F, et al. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a Tripartite nested mealybug symbiosis. Cell 153(7):1567–1578.

Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

Kaiwa N, et al. 2010. Primary gut symbiont and secondary, sodalis-allied symbiont of the scutellerid stink bug cantao ocellatus. Appl Environ Microbiol. 76(11):3486–3494.

Kaiwa N, et al. 2014. Symbiont-Supplemented Maternal Investment Underpinning Host's Ecological Adaptation. Curr. Biol. 24:1–6.

Kanehisa M, et al. 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44(D1):D457–D462.

Karamipour N, Mehrabadi M, Fathipour Y. 2016. Gammaproteobacteria as essential primary symbionts in the striped shield bug, *Graphosoma Lineatum* (Hemiptera: pentatomidae). Sci Rep. 6:33168.

Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28(12):1647–1649.

Kenyon LJ, Meulia T, Sabree ZL. 2015. Habitat visualization and genomic analysis of 'Candidatus *Pantoea carbekii*,' the primary symbiont of the brown marmorated stink bug. Genome Biol Evol. 7(2):620–635.

Kenyon LJ, Sabree ZL. 2014. Obligate insect endosymbionts exhibit increased ortholog length variation and loss of large accessory proteins concurrent with genome shrinkage. Genome Biol. Evol. 6:763–75.

Kikuchi Y, et al. 2009. Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stink bugs. BMC Biol. 7:2.

Kim J, Copley SD. 2007. Why metabolic enzymes are essential or nonessential for growth of *Escherichia coli* K12 on glucose. Biochemistry 46(44):12501–12511.

Koga R, Moran NA. 2014. Swapping symbionts in spittlebugs: evolutionary replacement of a reduced genome symbiont. ISME J. 8(6):1237–1246.

Kolbe DL, Eddy SR. 2011. Fast filtering for RNA homology search. Bioinformatics 27(22):3102–3109.

Konstantinidis KT, Rosselló-Móra R, Amann R. 2017. Uncultivated microbes in need of their own taxonomy. ISME J. 11:2399–2406.

Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35(9):3100–3108.

Lartillot N, et al. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol Biol. 7(Suppl 1):S4.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32(1):11–16.

Leimbach A. 2016. bac-genomics-scripts: Bovine *E. coli* mastitis comparative genomics edition. doi: 10.5281/ZENODO.215824.

Luan J-B, et al. 2015. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. Genome Biol Evol. 7(9):2635–2647.

Moran NA, Bennett GM. 2014. The tiniest tiny genomes. Annu Rev Microbiol. 68(1):195–215.

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

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

Newton ILG, Bordenstein SR. 2011. Correlations between bacterial ecology and mobile DNA. Curr Microbiol. 62(1):198–208.

Nikoh N, et al. 2014. Evolutionary origin of insect-Wolbachia nutritional mutualism. Proc Natl Acad Sci U S A. 111(28):10257–10262.

Nikoh N, Hosokawa T, Oshima K, Hattori M, Fukatsu T. 2011. Reductive evolution of bacterial genome in insect gut environment. Genome Biol Evol. 3:702–714.

Ohbayashi T, et al. 2015. Insect's intestinal organ for symbiont sorting. Proc. Natl. Acad. Sci. 112:E5179–E5188.

Otero-Bravo A, Sabree ZL. 2015. Inside or out? Possible genomic consequences of extracellular transmission of crypt-dwelling stink bug mutualists. Front Ecol Evol. 3:1–7.

Panizzi AR, Grazia J. 2015. True bugs (Heteroptera) of the neotropics. Dordrecht: Springer Netherlands.

Panizzi AR, Machado-Neto E. 1992. Development of nymphs and feeding habits of nymphal and adult *Edessa meditabunda* (Heteroptera: pentatomidae) on soybean and sunflower. Ann Entomol Soc Am. 85(4):477–481.

Patiño-Navarrete R, Moya A, Latorre A, Peretó J. 2013. Comparative genomics of *Blattabacterium cuenoti*: the frozen legacy of an ancient endosymbiont genome. Genome Biol Evol. 5(2):351–361.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786.

Pinto-Carbó M, et al. 2016. Evidence of horizontal gene transfer between obligate leaf nodule symbionts. ISME J. 10(9):2092–2105.

Prado SS, Rubinoff D, Almeida RPP. 2006. Vertical transmission of a pentatomid Caeca-associated symbiont. Ann Entomol Soc Am. 99:577–585.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5(3):e9490.

R Core Team. 2013. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.r-project.org; last accessed December 9, 2017.

Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 106(45):19126–19131.

Rizzo HF, Saini ED. 1987. Aspectos morfologicos y biologicos de *Edessa rufomarginata* (De Geer)(Hemiptera, Pentatomidae). Rev Fac Agron. 8:51–63.

Ruano-Rubio V, Fares MA. 2007. Artifactual phylogenies caused by correlated distribution of substitution rates among sites and lineages: the good, the bad, and the ugly. Syst Biol. 56(1):68–82.

Sabree ZL, Degnan PH, Moran NA. 2010. Chromosome stability and gene loss in cockroach endosymbionts. Appl Environ Microbiol. 76(12):4076–4079.

Salem H, Florez L, Gerardo N, Kaltenpoth M. 2015. An out-of-body experience: the extracellular dimension for the transmission of mutualistic bacteria in insects. Proc Biol Sci. 282(1804):20142957.

Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14):2068–2069.

Selengut JD, et al. 2007. TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. Nucleic Acids Res. 35(Database):D260–D264.

Silva DP, Oliveira PS. 2010. Field biology of *Edessa rufomarginata* (Hemiptera: pentatomidae): phenology, behavior, and patterns of host plant use. Environ Entomol. 39(6):1903–1910.

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

Stackebrandt E, et al. 2002. Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol. 52(Pt 3):1043–1047.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9):1312–1313.

Tada A, et al. 2011. Obligate association with gut bacterial symbiont in Japanese populations of the southern green stink bug *Nezara viridula* (Heteroptera: pentatomidae). Appl Entomol Zool. 46(4):483–488.

Tamas I, et al. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. Science 80:296.

Taylor CM, Coffey PL, DeLay BD, Dively GP. 2014. The importance of gut symbionts in the development of the brown marmorated stink bug, *Halyomorpha halys* (Stal). PLoS One 9(3):e90312.

Wernegreen JJ. 2015. Endosymbiont evolution: predictions from theory and surprises from genomes. Ann N Y Acad Sci. 1360:16–35.

Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31(20):3350–3352.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

Associate editor: Richard Cordaux