

Habitat Visualization and Genomic Analysis of “*Candidatus Pantoea carbekii*,” the Primary Symbiont of the Brown Marmorated Stink Bug

Laura J. Kenyon¹, Tea Meulia², and Zakee L. Sabree^{1,*}

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

²Molecular and Cellular Imaging Center, Ohio Agricultural Research and Development Center and the Department of Plant Pathology, The Ohio State University

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

Accepted: January 7, 2015

Data deposition: This genome project has been deposited in the National Center for Biotechnology Information GenBank sequence database under the following accession numbers: CP010907, CP010908, CP010909, CP010910, CP010911.

Abstract

Phytophagous pentatomid insects can negatively impact agricultural productivity and the brown marmorated stink bug (*Halyomorpha halys*) is an emerging invasive pest responsible for damage to many fruit crops and ornamental plants in North America. Many phytophagous stink bugs, including *H. halys*, harbor gammaproteobacterial symbionts that likely contribute to host development, and characterization of symbiont transmission/acquisition and their contribution to host fitness may offer alternative strategies for managing pest species. “*Candidatus Pantoea carbekii*” is the primary occupant of gastric ceca lumina flanking the distal midgut of *H. halys* insects and it is acquired each generation when nymphs feed on maternal extrachorion secretions following hatching. Insects prevented from symbiont uptake exhibit developmental delays and aberrant behaviors. To infer contributions of *Ca. P. carbekii* to *H. halys*, the complete genome was sequenced and annotated from a North American *H. halys* population. Overall, the *Ca. P. carbekii* genome is nearly one-fourth (1.2 Mb) that of free-living congeners, and retains genes encoding many functions that are potentially host-supportive. Gene content reflects patterns of gene loss/retention typical of intracellular mutualists of plant-feeding insects. Electron and fluorescence in situ microscopic imaging of *H. halys* egg surfaces revealed that maternal extrachorion secretions were populated with *Ca. P. carbekii* cells. The reported findings detail a transgenerational mode of symbiont transmission distinct from that observed for intracellular insect mutualists and illustrate the potential additive functions contributed by the bacterial symbiont to this important agricultural pest.

Key words: *Halyomorpha halys*, gammaproteobacteria, pentatomid, Hemiptera, invasive pest.

Introduction

Insect-bacterial mutualisms are widespread in insects and have had significant impacts on the evolution and diversification of insects, which remains the most speciose of macrofauna (i.e., Baumann 2005; Moran et al. 2008; McFall-Ngai et al. 2013; Douglas 2014). Many insects that feed exclusively on plant vascular fluids (e.g., xylem or phloem) or parenchymal cell contents can successfully exploit these abundant and nutritionally imbalanced diets with the assistance of bacterial symbionts that can provision essential amino acids and vitamins to their host (Moran et al. 2008; Douglas 2013). Complete genome sequencing and annotation of *Buchnera aphidicola* revealed a highly reduced genome lacking many genes

typically present in free-living relatives of *B. aphidicola*, namely *Escherichia coli*, yet the presence of nutrient-yielding biosynthetic pathways. This pattern of gene loss/retention suggests that in spite of dramatic genome reduction due to gene loss, *B. aphidicola* maintains a genic repertoire that is comprised largely those encoding processes generative of host-supportive nutrients (Shigenobu et al. 2000), and the genomes of other intracellular bacterial mutualists of evolutionarily distant insects, including cockroaches, cicadas, and carpenter ants, exhibit similar patterns of genome streamlining and maintenance of host-supportive genic repertoires (reviewed in Moran et al. 2008). With a few notable exceptions, available genomes of bacterial mutualists of insects have

© The Author(s) 2015. 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

been largely derived from transovarially transmitted, intracellularly incarcerated species.

Among heteropteran insects, alternative means of intergenerational transmission and domiciling of symbionts have been observed (reviewed in Hosokawa et al. 2013). Herbivorous females of the Pentatomidae and Plataspidae have been observed to deposit gammaproteobacterial symbiont-laden gut secretions that are either smeared on eggs or encapsulated and positioned proximally to eggs (Fukatsu and Hosokawa 2002; Hosokawa et al. 2005; Tada et al. 2011; Kikuchi et al. 2012). Unlike intracellular mutualists that are present within immature tissues prior to emergence, these symbionts persist in an unknown state of activity outside of host tissues prior to nymphal acquisition by oral consumption of maternal secretions and are presumed to travel to and fill the extracellular lumina of gastric ceca located on the distal midgut region. For example, *Megacopta punctatissima* (Plataspidae) nymphs acquire the gammaproteobacterial symbiont, "*Candidatus Ishikawaella capsulata*," by consuming maternally generated capsules affixed to eggs, whereas *Plautia stali* (Pentatomidae) receive an inoculum of an unnamed gammaproteobacterial symbiont as nymphs from consuming maternal secretions smeared on eggs and both insects domicile their symbionts in the ceca of specialized crypts (Abe et al. 1995; Fukatsu and Hosokawa 2002; Hosokawa et al. 2006). Denial of either species from acquiring their symbionts resulted in delayed growth, retarded development, and reduced fecundity (Abe et al. 1995; Fukatsu and Hosokawa 2002; Hosokawa et al. 2008). Although the occurrence of gammaproteobacterial symbionts inhabiting specialized midgut ceca of stink bugs has been well-documented, relatively few complete genomes for these symbionts are currently available (e.g., Nikoh et al. 2011; Brown et al. 2014; unpublished *Pl. stali* symbiont genome GenBank: AP012551.1) to assist in inferring the specific contributions of these symbionts to their hosts (e.g., vitamins, essential amino acids, etc.) or possible genomic consequences (e.g., reduction, skewed genic profile, A+T-bias) of their host associations.

In this regard, the complete sequencing of the primary gammaproteobacterial symbiont, "*Candidatus Pantoea carbekii*" (hereafter referred to as *P. carbekii*; Bansal et al. 2014), of the highly invasive (Hoebeke and Carter 2003; Wermelinger et al. 2008; Fogain and Graff 2011; Zhu et al. 2012; Leskey et al. 2012; Vetek et al. 2014; Xu et al. 2014) and polyphagous (Seetin 2011; Bergmann et al. 2014) pentatomid, *Halyomorpha halys*, commonly known as the brown marmorated stink bug, is reported. Although the *P. carbekii* genome is reduced relative to free-living gammaproteobacteria, it encodes enzymes that can generate essential nutrients potentially limited in the host's diet and that may assist symbiont survival on the egg surfaces prior to nymphal consumption and infection of the distal midgut. As in the aforementioned stink bugs, *P. carbekii* is domiciled within

the lumina of pigmented distal midgut gastric ceca and is obtained by nymphs when they consume maternal egg secretions following hatching (Taylor et al. 2014). Prevention of symbiont acquisition by nymphs through surface-sterilization of *H. halys* eggs results in aberrant nymph behavior and developmental delays (Taylor et al. 2014). To detail the transgenerational symbiont transmission strategy in pentatomids, in situ electron and fluorescence microscopy was used to obtain high-definition spatial and taxon-specific imagery of *P. carbekii*, yielding a more detailed description of the location and structural characteristics of the egg surface inhabited by *P. carbekii*. Collectively, these data suggest that genomic characteristics typically observed in insect mutualists are evident in *P. carbekii*, and the symbiont expresses a suite of enzymes that 1) may facilitate survival on egg surfaces prior to nymphal uptake and 2) play an essential role in *H. halys* development through provisioning of nutrients limited in the host's diet.

Materials and Methods

Genome Sequencing and Annotation

Complete genome sequencing of *P. carbekii* was performed using DNA extracted from adult *H. halys* specimens collected in Wooster, OH in 2013 and briefly maintained in a lab colony. DNA was extracted from the V4 region of the midgut using the DNEasy Blood and Tissue (QIAGEN) kit. Illumina MiSeq sequencing platform with the v2 reagent kit was used to generate 7.7 million 250-bp paired-end reads with an expected 250 bp insert size. Reads were quality trimmed (parameter: Base calls <Q30 Phred score were trimmed and reads <150 bp were excluded) and assembled within the CLC Genomics Workbench (version 6, CLC Bio, parameters: mapping mode=map reads back to contigs, automatic bubble size=yes, minimum contig length=200, automatic word size=yes, scaffolding=yes, auto-detect paired distances=yes, mismatch cost=2, deletion cost=3, length fraction=0.5, similarity fraction=0.8, word size=23, and bubble size=241), generating 111,569 contigs. A tBLASTx search (e value=0.0001) versus the National Center for Biotechnology Information (NCBI) "nt" database was done for all 111,569 contigs. Only 67 contigs had significant hits and of these, only 11 had hits to bacterial species ([supplementary material S1](#), [Supplementary Material](#) online). The four largest of these (contig_3, contig_28, contig_43, and contig_139) were greater than 20 kb (range: 65–792 kb, average: 284 kb). Concurrent with our *P. carbekii* assembly efforts, an unpublished *H. halys* symbiont genome from *H. halys* specimens of unknown origin was released in GenBank in October 2013 (GenBank: NC_022547) by researchers in the Nihon University, School of Pharmacy in Japan (hereafter called *P. carbekii* JPN). Pairwise alignment of the de novo generated four large contigs with the *P. carbekii* JPN genome was performed using BLASTN (parameters: e value 0.001) to obtain a

putative genome scaffold. Evident in the contig arrangement with *P. carbekii* JPN and the sequence annotations at the ends of the four large contigs, ribosomal RNA (rRNA) operons were likely between each of these contigs and assembly of a single contiguous chromosome was failing due to the considerable sequence conservation between rRNA regions. Specifically, putative 5S and partial 23S sequences were annotated at both ends of contig_3, a partial 16S sequence was detected at one end of contig_43, a partial 16S sequence was detected at one end of contig_28 and a partial 23S was detected at one end of contig_139 whereas a partial 16S sequence was detected at the opposite end of this contig. To close gaps between these four contigs de novo, outward-oriented polymerase chain reaction (PCR) primers were designed to amplify sequence-spanning gaps between contigs using Primer3Plus (Untergasser et al. 2012) with the ends of the four large contigs as references. All primers used in this study are reported in [supplementary material S1, Supplementary Material](#) online. Both combinatorial and guided (i.e., inferred contig ordering based on alignments to *P. carbekii* JPN) primer pairing in long-range PCR reactions performed with Platinum High Fidelity Taq polymerase (Invitrogen-Thermo, MA) according to manufacturer's recommended thermocycler conditions. Successful PCR amplifications yielded amplicons that were approximately the expected size of rRNA operons based on the lengths of related *Pantoea* spp. rRNA operons (region that spans complete 23S, 5S, and 16S sequences; *Pantoea ananatis* LMG20103, NC_013956.2 average: 5.2 kb, range: 5.1–5.4 kb; *Pantoea vagans* C9-1, NC_014562.1 average: 5.2 kb, range: 5.0–5.3 kb). PCR products were purified using the Beckman Coulter Agencourt AMPure XP purification system and Sanger sequencing of all amplicons was performed at The Ohio State University Plant-Microbe Genomics Facility. Sequences were manually edited and quality-checked in Geneious (version 7.0.4, Biomatters Limited). A Sanger-based, primer-walking strategy was used to sequence gap-spanning amplicons and yielded a single, contiguous chromosome. The remaining seven contigs that were of putative bacterial origin were then mapped to the contiguous chromosome. Of these seven bacterial contigs, three mapped to the chromosome in the three rRNA operon regions. In silico translations of open reading frames (ORFs) detected using Prodigal (default parameters; Hyatt et al. 2010) on each of the remaining four contigs (contig_578, contig_74, contig_85, and contig_131) that did not assemble de novo with the *P. carbekii* chromosome were 77–91% identical (BLAST 2.2.30+, program: BLASTp, parameters: default, database: refseq_protein; Altschul et al. 1997) to gammaproteobacterial plasmid replication proteins (e.g., RepA and RepFIB) and were tested to determine whether they were plasmids. Inverse PCR (Life Technologies Platinum Taq DNA Polymerase High Fidelity; see [supplementary material S1, Supplementary Material](#) online, for primers) using pairs of outward-facing primers positioned at the ends of contig_578, contig_74,

contig_85, and contig_131 resulted in positive amplification products that, when sequenced, closed the contigs upon themselves. The average read coverage for each of the contigs was as follows: contig_578 (pBMSBPS1): 2,960 \times , contig_74 (pBMSBPS2): 7,487 \times , contig_85 (pBMSBPS3): 1,412 \times , and contig_131 (pBMSBPS4): 610 \times . The lengths for the plasmids are pBMSBPS1: 14,562 bp, pBMSBPS2: 6,287 bp, pBMSBPS3: 17,880 bp, and pBMSBPS4: 7,593 bp.

MiSeq-generated reads were quality-filtered using FASTX-Toolkit v 0.0.13 (Gordon and Hannon 2010; "fastq_quality_trimmer" was used to trim nucleotides with quality <Q30 from the ends of the reads and reads less than 100 bp after trimming were discarded, "artifacts_remove" was implemented to remove reads with only three of the four possible bases, "fastq_quality_filter" was used to remove reads with less than 80% of bases with >Q28, and "fastq_to_fasta" was used to remove reads with any "N" ambiguous base calls). These high-quality reads were mapped to the *P. carbekii* chromosome to detect any polymorphisms and to detect any inaccurate base calls using BWA (default parameters; Li and Durbin 2010) with an average per base read coverage of 745 \times as determined by the "genomeCoverageBed" program in BEDTools (Quinlan and Hall 2009). Mapping results were visually inspected using the Integrative Genomics Viewer (Thorvaldsdóttir et al. 2012) and base calls supported by less than 99% of mapped reads were manually edited in Artemis Genome Viewer (Rutherford et al. 2000), unless they were supported by the Sanger sequencing. Only ten polymorphic sites (out of 1,150,626 nt) were detected in intergenic spaces, where two different bases were supported nearly equally by the reads. In these ten cases, the base with majority (>50%) of the reads supporting it was called.

Chromosome and plasmid encoded ORFs were determined with Prodigal, transfer RNAs (tRNAs) were called using tRNA-scan and Aragorn, and transfer-messenger RNAs (tmRNAs) and rRNAs were detected using Rfam (Laslett and Canback 2004; Schattner et al. 2005; Hyatt et al. 2010; Burge et al. 2012). Annotation was completed using Pfam functional domain determination, COG assignments, KEGG Orthology Groups, TIGRFAM, and BLAST (Basic Local Alignment Search Tool) comparisons of ORFs to the "ECO," "nr," and "nt" databases (Altschul et al. 1997; Tatusov et al. 1997; Haft et al. 2003; Moriya et al. 2007; Punta et al. 2012; Zhou and Rudd 2013; BLASTP, e value=0.001). Artemis Genome Viewer and DNAPlotter were used to assist with annotation and detecting G/C skew (Rutherford et al. 2000; Carver et al. 2009). Metabolic reconstruction was done using KEGG-KAAS with information from EcoCyc and MetaCyc (Moriya et al. 2007; Caspi et al. 2008; Keseler et al. 2009). *Pantoea carbekii* proteins that were fragmented due to internal stop codon(s) and/or frameshift mutations were annotated as pseudogenes. Average-per-base coverage in pseudogenes did not vary significantly from the average-per-base coverage across the

genome (745×), confirming that these indels and frameshifts were not likely due to sequencing error.

Comparative Analyses

Biosynthesis-related gene content and general genomic characteristics of *P. carbekii* were compared with those of the following gammaproteobacteria inhabiting various ecological niches (NCBI GenBank accession numbers are parenthesized): *E. coli* str. K12 (NC_000913), *P. ananatis* (AP012032.1), *Pl. stali* symbiont (AP012551.1), *Ca. I. capsulata* symbiont of *M. punctatissima* (AP010872.1), and *B. aphidicola* str. APS (NC_002528) (hereby referred to by their generic names). General genome statistics were compared across these genomes as in Moran et al. (2008). Maximal protein length was determined using the “infoseq” program in EMBOSS (Rice et al. 2000). Pseudogenes in the genomes of the above-mentioned taxa were reported based on GenBank-reported genome annotations.

Molecular Phylogenetic Reconstruction

Orthologous protein groups from the proteomes of related gammaproteobacteria with complete and draft genomes were determined using OrthoMCL (Fischer et al. 2011). Orthologs and the sourced proteomes used for phylogenetic reconstruction are reported in [supplementary material S1, Supplementary Material](#) online. Eighty orthologs from 47 taxa were individually aligned in MAFFT using the L-INS-i algorithm (Katoh et al. 2002) and gap-containing columns were removed using a custom Perl script, leaving a total of 18,678 characters per taxon. Aligned proteins were concatenated using a custom Perl script and maximum-likelihood trees were inferred using a web-based implementation of RAxML (parameters: rapid bootstrapping under the LG protein model and DAYHOFF substitution matrix, 100 bootstrap trees; CIPRES Science Gateway; Stamatakis et al. 2008; Miller et al. 2010) and the best-supported maximum-likelihood tree was reported. FigTree v1.3.1 and MEGA5.1 were used to prepare trees for publication (Rambaut and Drummond 2010; Tamura et al. 2011).

Single Nucleotide Polymorphism and Genomic Synteny Analysis

The *P. carbekii* and *P. carbekii* JPN chromosomes were compared to identify single nucleotide polymorphisms (SNPs) using BWA (Li and Durbin 2010) and the Genome Analysis Tool Kit (“GATK”; McKenna et al. 2010). Specifically, BWA-aligned genomes were subjected to an additional local alignment within GATK to improve the accuracy of SNP calling by locally realigning areas where there might be insertions or deletions (McKenna et al. 2010). GATK and Samtools (parameters: defaults; Li et al. 2009) were simultaneously used to call SNPs and only SNPs detected by both were considered. Furthermore, SNPs were only retained for analysis if greater

than 99% of the *P. carbekii* reads matched the called-SNP. Per base read depth for the SNPs averaged 857.5 reads (range: 51–1,784×), as determined by BEDTools genomeCoverage Bed program with the “-d” flag implemented for per base read depth calls (Quinlan and Hall 2009). SNPs were determined to be transitions or transversions within Microsoft Excel. For SNPs detected in coding regions, the “diffseq” program within EMBOSS was used to detect SNPs resulting in nonsynonymous or synonymous changes (Rice et al. 2000). SNPs resulting in nonsynonymous mutations in coding regions were categorized using the COG database (Tatusov et al. 1997). Structural rearrangements between the two genomes were detected using the “nucmer” program within the MUMmer package (Delcher et al. 2002; default parameters except –maxmatch option) and visualized using the Genome Synteny Viewer (Revanna et al. 2011).

Halyomorpha halys Gut Microbiome Analysis

Three adults from an Ohio population-derived, in-house *H. halys* colony were euthanized in 70% ethanol and immediately dissected in 1X phosphate-buffered saline (PBS). DNA was extracted separately from the V1–V2, V3, and V4 midgut tissues of the digestive tract (for details, see Bansal et al. 2014) using the QIAgen DNEasy Blood and Tissue Kit and submitted to the Institute for Genomics and Systems Biology Next Generation Sequencing Core (Argonne National Laboratory, Argonne, IL) for 16S rRNA amplicon library preparation and Illumina MiSeq 2×251 bp paired-end sequencing. Illumina-generated 16S rRNA sequence reads (hereafter referred to as “iTags” after Degnan and Ochman 2012) were assembled using Pandaseq (Masella et al. 2012), and trimmed to remove base calls with less than Q30 Phred score and quality-filtered (parameters: reads with ≥1 errors in barcodes or primers and/or were less than 230 bp or greater than 260 bp in length were removed) using the CLC Genome Workbench. Quality-filtered reads were preprocessed using the mothur software package (version 1.29; Schloss et al. 2009) and operational taxonomic units (OTUs) were clustered at ≥95% identity using usearch (version 7; Edgar 2010). Sequences representing OTUs were taxonomically assigned by BLAST (program: BLASTn, parameters: default) searches of SILVA 16S small-subunit “SSU” (version 115; Pruesse et al. 2007) and NCBI GenBank nt (downloaded on March 18, 2014) databases. Best hits within these databases that aligned with greater than 99% of the OTU sequences informed OTU genus designations.

Fluorescence In Situ Hybridization and Electron Microscopy

Fluorescence In Situ Hybridization

Crude homogenates were prepared from 1) freshly laid eggs with the extrachorion matrix removed (see below), 2) freshly laid eggs soaked in 10% bleach for 10 min and rinsed in

dH₂O, 3) eggs extracted from gravid females, and 4) V4 midgut crypts for microscopic imaging. Additionally, lavages of the extrachorion matrix surrounding freshly laid eggs were also prepared for imaging by placing single eggs in 200 μ l 1X PBS solution in a 1.5-ml centrifuge tube, vortexing for 10–15 s and the extrachorion matrix was further manually disrupted by pipetting approximately ten times, while being careful not to puncture the egg. Following the removal of the lavage, eggs were further rinsed with 1X PBS to remove any residual extrachorion matrix and homogenized in 1X PBS using a combination of vigorous vortexing, pipetting, and manual crushing with a sterile pestle until little or no intact egg fragments were visible. Eggs from gravid females and bleached eggs underwent a similar process (vigorous vortexing, pipetting, and manual crushing with a sterile pestle in 1X PBS) upon collection or after treatment (i.e., bleach treatment). V4 midgut sections were dissected from colony-maintained adults euthanized in 70% ethanol in 1X PBS and immediately transferred to a 1.5- μ l microcentrifuge tube containing 200 μ l 1X PBS. V4 tissues were homogenized until little or no intact pieces remained by vigorous vortexing, pipetting, and crushing with a sterile pestle. Three experimental replicates were performed for each of the treatments, except for the V4 midgut preparations, of which there were five replicates and the egg lavage treatments, of which there were nine replicates. Preparation for fluorescence in situ hybridization (FISH) followed Osborn and Smith (2005). Specifically, approximately 10–15 μ l of each sample (egg lavage, crushed egg or V4 midgut homogenates) was placed on microscope slides, air-dried in a vacuum hood, and fixed over an open flame for 1–3 s. An amount of 40 μ l 4% paraformaldehyde was immediately added to each sample, covered with a cover-slip, and incubated for 24 h at 4 °C. The samples were then dehydrated in a series of ethanol washes for 3 min at each of the following concentrations of ethanol: 50%, 80%, and 95%. An amount of 1 μ l (50 ng/ μ l) of each probe was warmed to 48 °C and mixed with 10 μ l prewarmed hybridization buffer (180 μ l 5 M NaCl; 20 μ l Tris-HCl; 200 μ l formamide; 599 μ l ddH₂O; 1 μ l 10% sodium dodecyl sulfate, SDS) for each sample. A quantity of 10 μ l of the hybridization buffer/probe solution (final probe concentration: 5 ng/ μ l) was added to each sample, covered with a cover-slip, and incubated for 24 h at 46 °C. Cy3-labeled Enterobacteriaceae-specific ("ENT1251": 5'-Cy3-TGCTCTCGCGAGGTCGCTTCTCT-3', 50 ng/ μ l; Ootsubo et al. 2002) and TYE-563-labeled *P. carbekii*-specific ("Crbck-TEX": 5'-TYE-ATGCTGCCGTTCGACTT-3', 50 ng/ μ l; this study) FISH probes were used separately. Following incubation, samples were rinsed with washing buffer (1 ml Tris-HCl; 2.15 ml 5 M NaCl; 0.5 ml 0.5 M ethylenediaminetetraacetic acid; 46.35 ml ddH₂O) for 3 min with gentle agitation, then dipped in ice-cold ddH₂O and air-dried in a vacuum hood. Slides were counterstained with 4,6-diamidino-2-phenylindole (1 μ g/ml) prior to adding mounting buffer (10% glycerol in 1X PBS) and covered with a

cover-slip to view with a Nikon Eclipse Ti inverted epifluorescence microscope.

Transmission Electron Microscopy

Midgut gastric ceca (V4) tissues were dissected from colony-maintained adult insects in ice-cold fixative (3% glutaraldehyde, 1% paraformaldehyde in 0.1 M potassium phosphate buffer, pH 7.2) in preparation for transmission electron microscopy (TEM), as described in Bansal et al. (2014). Briefly, samples were fixed for 3 h on a rocker at 23 °C, washed with 0.1 M potassium phosphate buffer (0.1 M KH₂PO₄, 0.1 M K₂HPO₄, pH 7.2), and treated with a postfixative solution (1% osmium tetroxide, 1% uranyl acetate in distilled water) at 23 °C for 1 h. Tissues were dehydrated through a graded ethanol series, resin infiltrated using a propylene oxide-resin series, and then embedded in EM Bed-812 resin (Electron Microscopy Sciences, Hartfield, PA). Ultrathin-sections were prepared using a Leica EM UC6 ultra-microtome. Three of the prepared V4 midgut thin-sections were used for FISH microscopy, using the protocol described above. After staining with 3% aqueous uranyl acetate for 20 min, followed by Reynolds' lead citrate for 10 min (Reynolds 1963), sections were imaged using a Hitachi H-7500 transmission electron microscope and images recorded with a Optronics QuantiFire S99835 (SIA) digital camera.

Scanning Electron Micrograph

Freshly laid eggs (within 1 day of oviposition; 5–10 eggs per treatment) were subjected to one of the following treatments prior to scanning electron micrograph (SEM) imaging: 1) 10–15 s vortexing in 1X PBS; 2) soaking in 9% bleach for 2 min, followed by three washes with 1X PBS; 3) soaking in 10% SDS for 10 min, followed by three washes with 1X PBS; and 4) no treatment. Treated and control eggs were fixed in 3% glutaraldehyde, 2% paraformaldehyde, in 0.1 M potassium phosphate buffer pH 7.2 (PB) over night, and subsequently rinsed three times for 10 min with PB and postfixed in 1% OsO₄ in PB for 1 h. After two PB washes, samples were dehydrated through washes in 50%, 75%, 95%, and 100% ethanol for 15 min each. Following dehydration, samples were dried in a critical-point dryer, sputter coated with platinum, and viewed on the Hitachi S-3500N scanning electron microscope.

Results and Discussion

Pantoea carbekii Dominates the *H. halys* Crypt-Bearing Midgut and Is Abundant on Egg Surfaces

iTags were generated from adult Ohio *H. halys* gut tissues to assess the diversity of bacteria comprising the microbiome of the V4 gut region and a location-specific, relatively high concentration of *P. carbekii* was detected in *H. halys* V4 midgut tissues. More than 98% of approximately 23,000 high-quality iTags for all of the V4-midgut region tissue samples were

unambiguously assigned to *P. carbekii*. In contrast, less than 1% of the iTags generated from all of the V1–V2 tissue samples and in most of the V3 tissue samples could be assigned to this symbiont. Additionally, *P. carbekii* has been detected in DNA preparations from *H. halys* egg clusters (Bansal et al. 2014; Taylor et al. 2014) and histological methods were used to localize egg-associated *P. carbekii* cells. *Halyomorpha halys* eggs are typically laid in clusters, affixed to one another with a maternally secreted extrachorion matrix (fig. 1E). Inspection of five eggs (41 SEM images) in which the extrachorion matrix was disrupted by brief vortexing in sterile, distilled 1X PBS (fig. 1A) revealed dense patches of *P. carbekii* on the egg surface (fig. 1B). Rod-shaped bacterial cells were observed beneath the matrix (fig. 1C) that were morphologically similar to those imaged by TEM of thin-sections of the V4 midgut gastric ceca (fig. 1D). Of the eggs that were treated with 10% bleach, three were observed (eight SEM images) and none had bacterial cells and or any of the extrachorion matrix present on untreated or vortexed eggs (fig. 1F). Four of the eggs that were treated with 10% SDS were observed (13 SEM images) and they maintained some of the extrachorion matrix but no bacterial cells were observed (fig. 1H). To confirm that the bacterial cells present on egg surfaces were *P. carbekii*, FISH microscopy was performed on lavages prepared from the egg extrachorion matrix (fig. 1G) and of V4 midgut tissues (supplementary material S2A and B, Supplementary Material online) using a *P. carbekii*-specific FISH probe. Rod-shaped cells that exhibited a strong fluorescence signal and were morphologically similar (fig. 1G and supplementary fig. S2A and B, Supplementary Material online) to those in the SEM (fig. 1C) and TEM (fig. 1D) were observed. Lavages prepared from eggs that were bleached or prewashed multiple times in 1X PBS, or extracted from gravid females yielded no observable bacterial cells by FISH or light microscopic methods. Although pentatomomorphans exhibit a few types of symbiont acquisition and transmission mechanisms, varying from recruitment of environmental *Burkholderia* strains by each generation of alydids (Kikuchi et al. 2007) to vertical transmission of *Ishikawaella* through consumption of symbiont-filled capsules by plataspids (Nikoh et al. 2011), maternal egg smearing of gut fluids and subsequent nymphal feeding on these eggs has been documented in other pentatomids examined, such as *Nezara viridula* (Prado et al. 2006), *Eurydema* spp. (Kikuchi et al. 2012b) and *Sibaria englemanni* (Bistolos et al. 2014), and appears to be the mode of symbiont transfer in *H. halys* (Taylor et al. 2014).

General Characteristics of the *P. carbekii* Genome

The *H. halys* symbiont, *P. carbekii*, forms a clade with the *Pl. stali* symbiont within the *Pantoea* group (fig. 2) and its genome is significantly reduced in size compared with congeners and other gammaproteobacteria. Consisting of only 1,197,048 nucleotides, the *P. carbekii* genome is roughly

one-fourth the size of related species (table 1). Although this represents a significant genome reduction, many intracellular insect symbionts, such as *Buchnera* or *Blattabacterium*, have further reduced genomes, with many being less than 1 Mb (Moran et al. 2008). The *P. carbekii* chromosome encodes 797 protein-coding genes, 2 tmRNAs, 40 tRNAs, and two complete rRNA operons and one rRNA operon containing 16S–23S-encoding regions and lacking an identifiable 5S-encoding region. Four plasmids were detected, all encoding RepA, and one encodes a similar gene content as the plasmid detected in *Ishikawaella*. Twelve pseudogenes were predicted and annotated (table 1) and roughly 40 additional proteins appear truncated, but not pseudogenized, relative to orthologs in other *Pantoea* species. With the 40 tRNA species, *P. carbekii* is able to decode all 20 amino acids (table 1), including the translation initiating *N*-formyl-methionyl-tRNA (BMSBPS_0622), and a isoleucine-charged tRNA (BMSBPS_0773) that, following posttranscriptional lysylation of a position 34 cytidine by TilS, changes the anticodon from CAU to AUA to prevent misreading of AUG as isoleucine (Soma et al. 2003). Multiple copies of genes encoding tRNAs that recognize the mRNA codons “AUG” and “GAA” (which correspond to methionine and glutamate, respectively) and “AUC” (isoleucine) are present. As with other primary symbionts of insects, the *P. carbekii* genome exhibits reduced G+C% (30.57%) compared with free-living relatives such as *P. ananatis* (53.76%), which is a hallmark feature associated with genome reduction that accompanies elevated fixation of mutations under relaxed selection in endosymbionts with stable host-restricted lifestyles (Moran et al. 2008). Although the genome is reduced, *P. carbekii* still encodes metabolic pathways for the production of peptidoglycan, generation of ATP by aerobic respiration, and other primary metabolic processes. The genes encoding enzymes for peptidoglycan biosynthesis are often missing in intracellular-residing insect mutualists with tiny genomes (e.g., *Carsonella* and *Tremblaya*) and their presence suggests that *P. carbekii* can produce a cell wall, which might be important for persistence in the extrachorion matrix. ATP synthesis by aerobic respiration is prevalent across members of the gammaproteobacteria and that *P. carbekii* retains this efficient energy-yielding pathway reflects the oxic state of the midgut crypts.

Another trait observed that is associated with a host-restricted lifestyle is the reduction in the maximal protein length in comparison to free-living relatives, such as *P. ananatis* and *E. coli* (table 1). Some proteins involved in toxin production, secondary metabolic processes, virulence, extracellular sensing, or are of unknown function were shown to be, on average, larger than those involved in primary metabolic processes such as DNA replication, transcription, translation and essential amino acid biosynthesis, and the former, being nonessential for long-standing host-restricted lifestyles, are largely missing from insect endosymbiont genomes and their absence likely contributes to the reduced

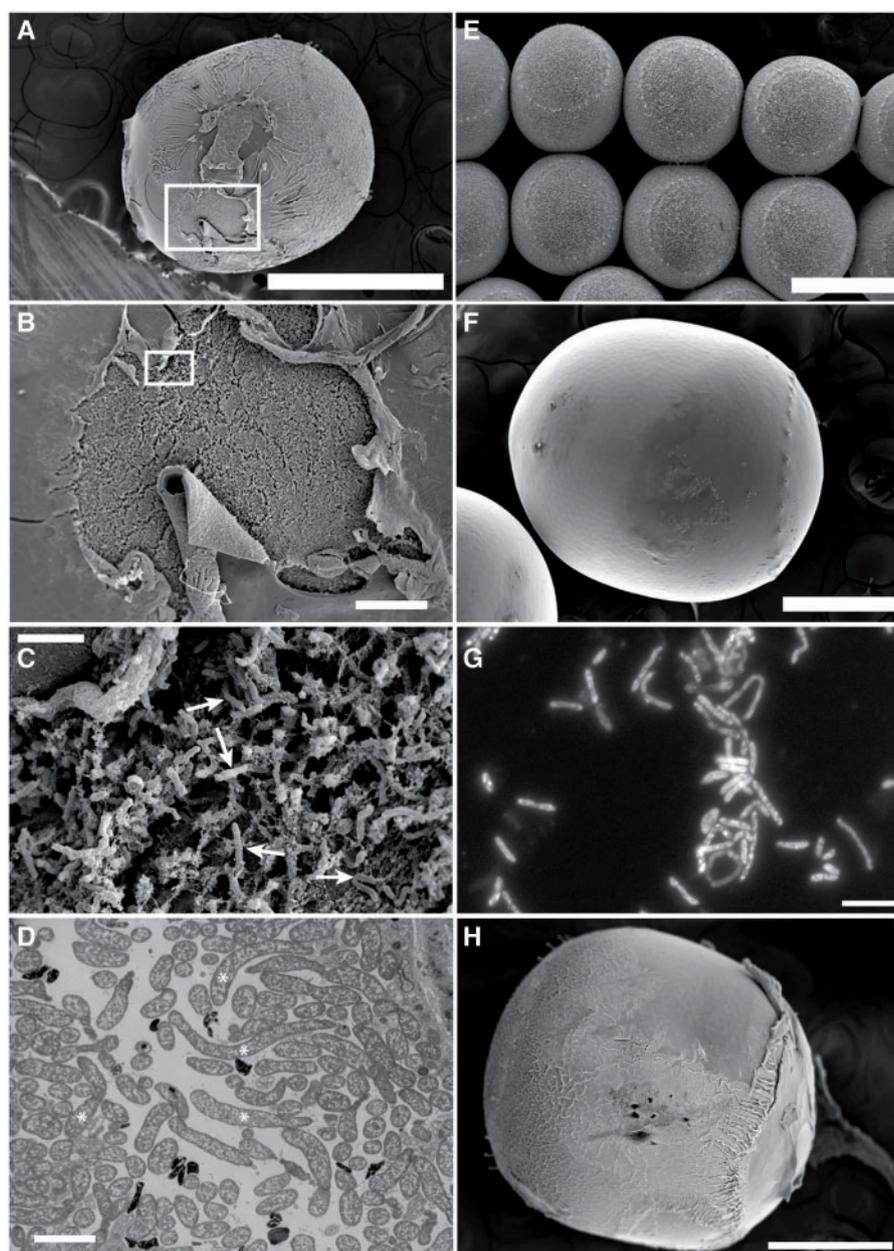


Fig. 1.—Ultrastructural imaging of *H. halys* eggs. (A) SEM of the *H. halys* egg surface with the extrachorion matrix peeling back to reveal an abundance of *P. carbekii* cells. Scale bar: 1 mm. (B) Detail of figure 1A showing a confluence of *P. carbekii* cells intercalated in extrachorion matrix. Scale bar: 0.3 mm. (C) Detail of figure 1B showing individual *P. carbekii* cells (indicated at the tips of the white arrows) in the extrachorion matrix. Scale bar: 20 μm . (D) TEM imaging of *P. carbekii* cells within the V4 midgut. Scale bar: 4 μm . (E) SEM of untreated *H. halys* eggs. Scale bar: 1 mm. (F) Bleach-treated *H. halys* egg. Scale bar: 0.5 mm. (G) FISH microscopy of the *P. carbekii* present in extrachorion matrix lavages with *P. carbekii*-specific FISH probes. Scale bar: 10 μm . (H) 10% SDS-treated *H. halys* egg. Scale bar: 0.5 mm.

genome sizes (Kenyon and Sabree 2014). *Pantoea carbekii* exhibits a somewhat similar genic profile in that the longest protein encoded by its genome is the 1,413 amino acids long beta' subunit of RNA polymerase (RpoC), which is important for transcription, and it is one-third that of the longest protein encoded in the genome of a free-living relative, *P. ananatis* AJ13355 (putative secondary metabolite

biosynthesis protein YP_005934773; 4,385 amino acids) that is of unknown function.

P. carbekii Metabolism and Putative Role in *H. halys* Physiology

Phytophagous diets are limited for essential amino acids and some vitamins and, based on the genome content, *P. carbekii*

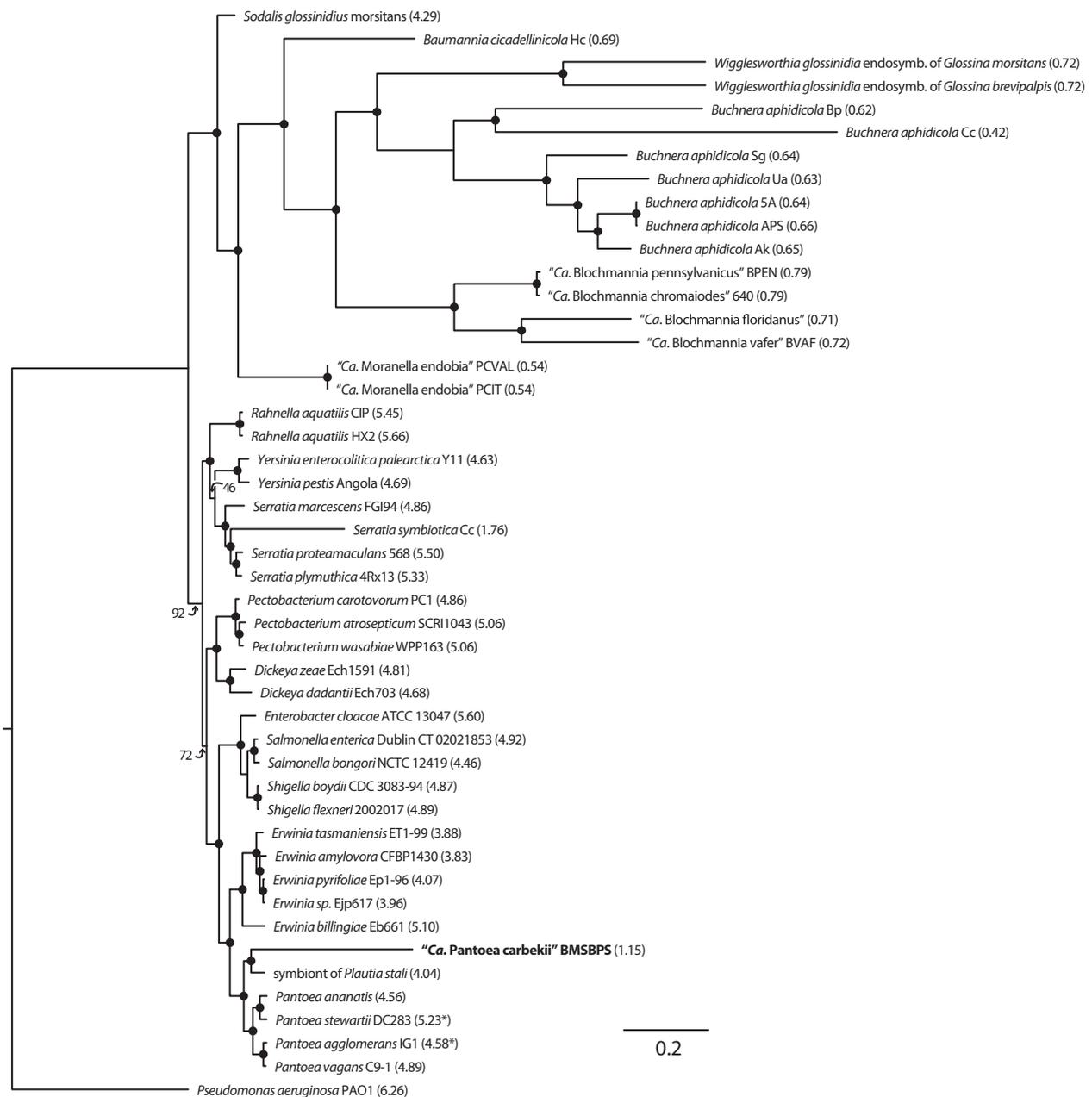


FIG. 2.—Core gene phylogeny of *Enterobacteriaceae*. Maximum-likelihood-based phylogenetic reconstruction of *P. carbekii* and other gammaproteobacteria ($n = 47$ taxa) using 80 concatenated and aligned orthologous proteins, comprising a total of 18,678 characters per taxon, was performed in RAXML. Support values generated from 100 bootstrap trees are indicated at branch points and were greater than 95% (black dot) unless otherwise noted. Parenthesized values next to species names are genome sizes in megabases; asterisks indicate estimated sizes for draft genomes. Accession numbers for the genomes from which the proteomes were derived and the orthologs used are reported in [supplementary material S1E](#), [Supplementary Material](#) online.

appears capable of supplementing the diet of *H. halys* with these nutrients. Although the genome shows evidence of reduction, *P. carbekii* encodes canonical pathways, like those typically observed in free-living gammaproteobacteria, namely *E. coli* and *P. ananatis*, for the biosynthesis of all essential and nonessential amino acids (fig. 3; [supplementary material S1](#), [Supplementary Material](#) online) except for proline, isoleucine, leucine, and valine. The proline biosynthesis

genes *proA* and *proB*, which encode glutamate-5-semialdehyde dehydrogenase and glutamate 5-kinase, respectively, are both absent from the *P. carbekii* genome and the ornithine-to-proline cyclodeaminase gene, *ocd*, is also missing. As no alternative pathways or functionally equivalent genes appear present, *P. carbekii* may not be able to synthesize proline de novo. Isoleucine, leucine, and valine biosynthesis pathways are complete except for the missing amino acid aminotransferase, *IlvE*,

Table 1Genome Characteristics Compared between *Pantoea carbekii* and Related Free-Living and Symbiotic Organisms with Varying Genome Sizes

	<i>Escherichia coli</i> K-12	<i>Pantoea ananatis</i>	Symbiont of <i>Plautia stali</i>	<i>Pantoea carbekii</i>	<i>Ishikawaella capsulata</i>	<i>Buchnera aphidicola</i> APS
Genome size (bp)	4,641,652	4,877,280 ^a	4,092,852 ^a	1,197,048 ^a	754,729 ^a	655,725 ^a
Plasmids	—	1	2	4	1	2
Chromosomal CDS	4,140	4,038	5,122	797	623	564
Plasmid CDS	—	278	26; 55	9; 5; 11; 6	8	3; 7
rRNA coding genes ^b	22 ^c	22 ^c	16 ^d	8 ^e	9 ^f	3 ^g
tRNAs	89	78	59	40	37	32
Pseudogenes	184	NA	NA	12	35	13
Maximum protein size (amino acids)	2,358	4,385	1,843	1,413	1,415	1,407
G+C content (%)	51	54 (52)	57 (48; 49)	31 (30; 27; 26; 24)	30 (28)	26 (27; 31)
Habitat	Free-living/enteric	Plant pathogen	Insect symbiont	Insect symbiont	Insect symbiont	Intracellular insect symbiont
Host insect	—	—	<i>Plautia stali</i> (Brown-winged Green Bug)	<i>Halyomorpha halys</i> (Brown Marmorated Stink Bug)	<i>Megacopta punctatissima</i> (Japanese Common Plataspid Stink Bug)	<i>Acyrtosiphon pisum</i> (Pea Aphid)
NCBI accession	NC_000913.3	AP012032.1, AP012033.1	AP012551.1, AP012552.1, AP012553.1		AP010872.1, AP010873.1	NC_002528.1, NC_002253.1, NC_002252.1

^aGenome size is the sum of the chromosome and plasmid(s).^bGenes encoding the 5S, 16S, and 23S rRNAs were counted.^cIn addition to seven 5S–23S–16S rRNA operons, an additional 5S rRNA coding region has been annotated.^dTwo 23S rRNA–5S rRNA operons are present in addition to four 5S–23S–16S rRNA operons.^eTwo 5S–23S–16S rRNA operons are annotated and additional 16S–23S rRNA operon has been annotated.^fThree complete rRNA operons are annotated.^gA 23S–5S rRNA operon and a separate 16S rRNA gene have been annotated in the genome.

which is required for the final step in branched chain amino acid biosynthesis. Absence of this gene has been observed in both *Ishikawaella* and *Buchnera* and host participation in the production of these essential amino acids has been suggested (Shigenobu et al. 2000; Nikoh et al. 2011).

Although the genome of *P. carbekii* lacks *argD* (acetylornithine aminotransferase), which catalyzes amination steps in lysine and arginine biosynthesis pathways (Ledwidge and Blanchard 1999), and *argI* (ornithine carbamoyltransferase), which catalyzes the sixth step of arginine biosynthesis (Glandsdorff et al. 1967), appears pseudogenized (frameshift around an adenosine 9-mer), the functional role of ArgD in arginine and lysine biosynthesis could be replaced by AstC (Kim and Copley 2007), which is encoded on the *P. carbekii* plasmid, pBMSBPS1, and similarly on the *Ishikawaella* plasmid (Nikoh et al. 2011). Although the gene encoding the asparagine synthetase A (*asnA*) is missing from the asparagine biosynthesis pathway, asparagine synthetase B, *asnB*, which has homologous functions in *E. coli*, is present. However, *asnB* appears pseudogenized by a frameshift around an adenosine 8-mer but previous work has shown that a subset of transcripts from frameshift-based pseudogenes can encode intact enzymes due to transcriptional slippage at homopolymeric sites (Tamas et al. 2008). ArgI production may be also rescued in a similar manner if transcriptional slippage around

the adenosine 9-mer corrects the internal frameshift. If these pseudogenized-by-frameshift genes are nonfunctional, then it is possible that host-encoded enzymes may complement the functions performed by these enzymes (Russell et al. 2013). *Pantoea carbekii* encodes complete or near-complete canonical pathways for the production of several vitamins and cofactors, including folate (vitamin B9), riboflavin (vitamin B2; although *ribC* is absent), pyridoxal-5'-phosphate (vitamin B6), glutathione, iron–sulfur clusters, and lipoate (fig. 3). Unlike the *Ishikawaella* genome, *P. carbekii* is likely unable to synthesize biotin due to the complete absence of this pathway (supplementary material S1, Supplementary Material online). Like *Ishikawaella*, it is missing genes encoding enzymes involved in molybdopterin biosynthesis (*meaA*, *meaB*, and *meaC*), but the corresponding enzymes or metabolites may be supplied by other *P. carbekii* biosynthetic pathways, the host, or the host's diet.

P. carbekii Plasmids Encode Genes Important for Nitrogen Assimilation and Thiamine Biosynthesis

Pantoea carbekii maintains four plasmids (pBMSBPS1–4) ranging from 6.3 to 17.9 kb in size and, besides all sharing the plasmid replication RepA-encoding gene, each encodes a distinct genic repertoire. pBMSBPS1 encodes for nine proteins

Among the 11 proteins encoded on pBMSBPS3, several appear to be involved in thiamine (vitamin B1) biosynthesis, which is not known to be produced *de novo* by insects (Sweetman and Palmer 1928; Craig and Hoskins 1940). Specifically, ThiSFGO are largely responsible for the production of the thiazole moiety, which, along with the pyrimidine moiety, are the two key precursors in prokaryotic thiamine biosynthesis (Du et al. 2011). In *E. coli*, the thiazole moiety is derived from condensation of tyrosine, cysteine, and 1-deoxy-D-xylulose-5-phosphate (Du et al. 2011), which involves thiazole synthase ThiH, which is not identifiable in the *P. carbekii* genome (including all plasmids). However, ThiO, or glycine oxidase, is encoded on pBMSBPS3 and is typically utilized by *Bacillus subtilis* instead of ThiH, but it uses glycine instead of tyrosine in the formation of the thiazole moiety (Du et al. 2011). Also undetectable in the *P. carbekii* genome is ThiI, a bifunctional enzyme that adenylates tRNA and mobilizes sulfur in 4-thiouridine synthesis and has been suggested to participate as a persulfide carrier in thiazole synthesis (Taylor et al. 1998). Mutational analysis of ThiI in *Salmonella* has shown that only a rhodanese domain in ThiI, which can perform the sulfur transfer to thiazole, is necessary and thus an alternative sulfur transferase could, along with cysteine, perform this function (Martinez-Gomes et al. 2011). A conserved hypothetical protein in *P. carbekii*, BMSBPS_0563, has a rhodanese-like domain (PFAM: PF00581) and thus may perform this function. The remaining genes involved in thiamine biosynthesis are present on the chromosome. Specifically, ThiCD uses 5-aminoimidazole ribotide from purine biosynthesis to generate the pyrimidine moiety, ThiE couples the thiazole and pyrimidine moieties to form thiamine monophosphate, and ThiL catalyzes the formation of the active form of thiamine.

Of the 11 protein-coding regions on the remaining two plasmids, pBMSBPS2 and pBMSBPS4, seven encode proteins with conserved domains of unknown function or are conserved hypothetical proteins.

pBMSBPS2 encodes five proteins, two of which could potentially be involved in stress-tolerance, namely the DNA mismatch repair protein MutT and inositol monophosphatase SuhB. A *mutT* deletion in *E. coli* significantly increases the spontaneous occurrence of A:T to C:G transversions a 1,000-fold over the wild-type and increases transcriptional errors (Dukan et al. 2000; Shimokawa et al. 2000). Although the physiological role of SuhB is not well understood, a *suhB* mutation is sensitive to cold, suggesting a potential role in cold-tolerance (Matsuhisa et al. 1995; Chen and Roberts 2000).

Degradation of DNA Replication and Repair Mechanisms and Abundant SNPs between *P. carbekii* Strains

A few genes encoding products involved in DNA repair, transcription, homologous recombination, and metabolite

conversions exhibit abnormal gene morphologies and/or harbor indel mutations, resulting in premature stop codons, frameshifts or truncations. Notably, DNA polymerase I, which is involved in DNA repair (Glickman 1975; Sharon et al. 1975; Smith et al. 1975), is present in the *Buchnera*, *Ishikawaella*, and *P. ananatis* genomes as single loci encoding approximately 900 amino acid enzymes, but it is split into two protein-coding regions (e.g., *polA1* and *polA2*) in different reading frames separated by stop codons in *P. carbekii*. Although the frameshifts and stop codons may interfere with production of a classical DNA polymerase I, the presence of intact functional domains within *polA1* and *polA2* suggests that, together, the products of these two genes may be capable of performing the functions of PolA. Additionally, the NAD-dependent DNA ligase, LigA, which is active during DNA replication, recombination, and repair and joins DNA fragments, closely resembles orthologs in *Ishikawaella*, *Buchnera*, and other *Pantoea* species except for lacking approximately 90 C-terminal residues that comprise a BRCT domain (Pfam: PF00533) that binds to DNA but is not essential for DNA-joining activity in *E. coli* (Wilkinson et al. 2005).

Several genes involved in DNA repair are missing from the *P. carbekii* genome. Absent are *phr*, which encodes a deoxyribodipyrimidine photolyase that acts in a light-dependent reaction to split pyrimidine dimers after UV radiation exposure (Keseler et al. 2009), and *xth* (Exonuclease III) and *rep* (Rep helicase) whose products are also involved in DNA repair. Xth is an endonuclease that repairs DNA where damaged bases have been removed or lost in *E. coli* (Gossard and Verly 1978). The loss of the regulation of *xth* results in the hypersensitivity of *E. coli* mutants to UV radiation (Sak et al. 1989). Rep helicase is required for replication in *E. coli*, preventing double-stranded breaks and acting in a replication fork restart pathway (Michel et al. 1997; Sandler 2000). The absence of *rep* in *E. coli* results in severe growth problems, namely the accumulation of DNA within a single cell (Trun 2003). We detected the 8-oxo-dGTP diphosphatase (*mutT*) gene on pBMSBPS2 and it encodes half of the base excision repair (BER) system whereas the formamidopyrimidine DNA glycosylase, encoded by *mutM* was absent; *mutM* deletions in *E. coli* have been shown to result in an increase in GC→AT transversions (Cox 1976; Cabrera et al. 1988). Finally, *recG* (RecG DNA helicase) is not detected in the *P. carbekii* genome, but it is also missing in *Ishikawaella* and *Buchnera* (supplementary material S1, Supplementary Material online).

Pairwise comparison of the reported *P. carbekii* genome and *P. carbekii* JPN highlights areas of the genome that may have accumulated recent mutations, assuming that the former is recently (i.e., in the last ~30 years) derived from a related native regional Asian *H. halys* population. An amount of 99.8% nucleotide identity between the genomes of the two *P. carbekii* strains was determined by megaBLASTn

alignment and the overall gene order between the genomes of the two *P. carbekii* strains was identical, with only two 76- and 66-bp inversions and a 100-bp deletion in intergenic regions being detected in the *P. carbekii* US genome. An SNP analysis was performed on the two strains to identify genes that may be experiencing rapid mutations and are either 1) the target of strong positive selection or 2) no longer being selected for since they are not required for maintenance of the organism (i.e., Brown et al. 2014). An average of 1 SNP per kb was observed and of 1,144 total SNPs observed in *P. carbekii*, 511 of these were in genic regions (supplementary material S1, Supplementary Material online). Most of the SNPs represented transitions (724) versus transversions (420). Over half of the SNPs in genic regions (293/511) coded for nonsynonymous protein mutations and these SNPs were distributed over 170 unique genes, with some genes having up to seven nonsynonymous mutations, but these did not result in any observable loss-of-function due to new stop codons (supplementary material S1, Supplementary Material online). Of the few SNPs that resulted in nonsynonymous changes in protein-coding regions, determining the impact of these changes on enzyme function is part of ongoing investigations. Furthermore, additional sampling in *H. halys* originating from both extant and archived populations in the United States and the Asian-Pacific regions would be useful in determining whether the observed SNPs are fixed and/or distributed by habitat.

Genes with more than one nonsynonymous SNP included *ytfN* (or *tamB*), *uvrA*, *surA*, and *polA*. *ytfN* encodes the integral inner membrane protein portion of the translocation and assembly module "TAM," which was recently found to promote efficient secretion of autotransporters in proteobacterial species (Selkrig et al. 2012). Along with TamA (also encoded within *P. carbekii* genome), the integral outer membrane protein in this complex, TAM allows for the efficient translocation of Antigen 43 and EhaA, two autotransporter proteins that are involved in biofilm formation and pathogenesis in *E. coli* (Danese et al. 2000; Wells et al. 2008; Selkrig et al. 2012). Interestingly, Antigen 43 and EhaA are not encoded by the *P. carbekii* genome. This suggests that TAM might translocate other proteins in *P. carbekii*, not yet described, that may share similar features with these biofilm-formation proteins. *uvrA*, which encodes a subunit of the UvrABC nucleotide excision repair generalized DNA repair process also appears to have acquired new mutations (Kenyon and Walker 1981). This is notable as it is involved in SOS response in *E. coli* and therefore might play a similar role in *P. carbekii*, which may experience environmental stress during its time on the *H. halys* egg surface. *surA* is yet another gene involved in DNA repair in which four substitution mutations are observed but it is unclear as to how these may affect its ability to fold outer membrane proteins or respond to high or low pH (Dartigalongue et al. 2001).

Degraded Cell Division Genes May Contribute to Nonuniform Cell Morphology

The cells of the gammaproteobacterial symbiont of *Murgantia histrionica* have been reported to exhibit nonuniform lengths (Prado et al. 2010) and in SEM, TEM, and FISH imaging of *P. carbekii* revealed cells with similar morphologies (fig. 1C, D, and G). The observed cell morphology may reflect cells undergoing cell division but the loss of the Rep helicase (*rep*) and presence of a truncated FtsK and pseudogenized FtsN gene may impair cytokinesis. Mutations in *rep* in *E. coli* resulted in severe growth defects and a near-doubling of the amount of chromosomal DNA present within each cell (Trun 2003). FtsK directs DNA translocation and chromosome segregation during cytokinesis in *E. coli* and the *P. carbekii* ortholog (932 amino acids) is approximately 80% of the length of FtsK in *P. ananatis* (1,112 amino acids) and the *P. stali* symbiont (1,143 amino acids). Although it lacks the DNA-binding C-terminal gamma domain (Pfam: PF01580), it retains the putative transmembrane FtsK-SpollIE domain that is implicated in DNA translocation in *E. coli* (Begg et al. 1995) and *Ba. subtilis* (Fleming et al. 2010). *Escherichia coli* Δ *ftsK*-encoded proteins with substitutions in the C-terminal domain show impaired DNA binding but remain capable of functioning in DNA translocation and chromosome segregation (Sivanathan et al. 2006). Additionally, *E. coli* with FtsK lacking the C-terminal domain exhibits a higher proportion of filamentous cells compared with the wild-type (Sivanathan et al. 2006). Sequence analysis alone is not conclusive on whether or not *P. carbekii* FtsK performs similarly to orthologs in *P. ananatis* or *E. coli*, or if its lack of a C-terminal domain contributes to the observed *P. carbekii* cell morphology. Although 387 nt coding for an intact N-terminal portion of the protein is detected, *ftsN* has been annotated as a pseudogene because an in silico translation of an adjacent 382 nt reveals a region coding for the peptidoglycan-binding SPOR domain found within the C-terminal portion of FtsN that is peppered with stop codons and frameshifts. FtsN is important in cell division as it interacts with another key late-stage cytokinesis protein, FtsA, to trigger septation, yet FtsA has been shown to function independently of FtsN (Bernard et al. 2007). Although *ftsN* is absent in the *Buchnera* genome, truncated and near-full length orthologs were annotated in the *Ishikawaella* (170 amino acids) and the *Pl. stali* symbiont (279 amino acids) genomes. Consequences of these mutations could be nonlethal impairment of cytokinesis within *P. carbekii*, leading to the observed cell morphologies, but further experimental work to confirm the expression and functionality of *P. carbekii* FtsK and FtsN is needed.

Extrachorion Persistence and Stress Tolerance in *P. carbekii*

During the period between oviposition and consumption by emerged nymphs (~1 week; Nielsen and Hamilton 2009),

P. carbekii lives outside of host gastric tissues within a maternally secreted matrix on the egg surface. Although beneath and intercalated within the extrachorion matrix of *H. halys* eggs, *P. carbekii* is exposed to fluctuations in environmental conditions (e.g., temperature fluctuations, UV radiation, desiccation, etc.) and biotic interactions (i.e., competition with other microbes or predation) that may result in stress in *P. carbekii* cells. A number of putative stress–response genes were detected in the *P. carbekii* genome and are discussed briefly. Protein denaturation, misfolding, and aggregation can be detrimental to cells following heat stress and sigma-32, which is encoded by *rpoH*, is at the hub of rapid responses to heat stress through the expression of enzymes involved in refolding and stabilizing denatured proteins (chaperones: GroESL, ClpB, IbpAB, and DnaKJ) or degrading irreversibly misfolded proteins (proteases: ClpAP, ClpXP, HslUV, and FtsH) (reviewed in Rosen and Ron 2002; Gunesekere et al. 2006). *Pantoea carbekii* encodes all of the aforementioned enzymes, with *ibpB* present on pBMSBPS1. IbpB binds aggregated or denatured proteins and has been shown to increase in expression in response to temperature up-shifts and during biofilm formation in *E. coli* (Lasokowska et al. 1996; Ren et al. 2004). Some of its functions are dependent on IbpA (Kuczyńska-Wiśnik et al. 2002), which is encoded on the *P. carbekii* chromosome and is present in both *Ishikawaella* and *Buchnera*. In *Buchnera*, IbpA significantly impacts heat-tolerance in pea aphids as shown by a single nucleotide deletion in *ibpA* that resulted in sharp fitness declines in pea aphids harboring *Buchnera ΔibpA* were maintained at elevated temperatures (25–30°C) (Dunbar et al. 2007). The presence of *ibpA* and the pBMSBPS1-encoded *ibpB* in *P. carbekii* may indicate that it also confers elevated temperature tolerance to *H. halys*.

Additionally, host factor one (*hfg*) is also encoded on the *P. carbekii* chromosome and it has both RNA chaperone activity and regulates the expression of, among other stress response genes, the stationary phase and environmental stress response regulator, RpoS (Muffler et al. 1996; Moll et al. 2003; Battesti et al. 2011).

Many of the aforementioned genes are not present in *Ishikawaella*, which has a similar mode of inheritance to *P. carbekii*, but it is packaged in symbiont capsules, which are deposited next to the eggs (Fukatsu and Hosokawa 2002), rather than in a surface smearings. *Ishikawaella* might not have the same exposure to abiotic and biotic pressures as *P. carbekii* does within the extrachorion matrix. It has been suggested that the *Ishikawaella* capsule conditions mimic that of the host midgut, protecting it from environmental fluctuations (Fukatsu and Hosokawa 2002; Hosokawa et al. 2005). As a result, the environmental conditions for *Ishikawaella* may not necessitate retention of the same repertoire of stress response genes as *P. carbekii*.

Conclusions

We report the first ultrastructural characterization of *P. carbekii* within the extrachorion matrix of brown marmorated stink bug eggs by SEM, and provide the complete genome sequence of this agricultural pest primary symbiont. Detection of the symbiont within this extrachorion matrix confirms that *H. halys* shares its primary symbiont transmission modality with other phytophagous stink bugs that exhibit egg-smearing behavior and harbor gammaproteobacterial symbionts within the gastric ceca of the distal midgut. Elucidating the biochemical composition of the extrachorion matrix may reveal chemical attractants that stimulate nymphal feeding behavior immediately following emergence (i.e., presence of specific attractants) as well as compounds involved in improving the survivability of *P. carbekii* outside of host tissues following oviposition and prior to nymph consumption.

Detailed genomic analysis of *P. carbekii* indicates that it has the potential to provision a wide range of dietary supplements, namely essential amino acids and vitamins, to its herbivorous host, and that the genome displays hallmarks of long-term host association, including a low G+C% and a reduced genic repertoire and genome size. If *P. carbekii* is provisioning nutrients to its host insect, then it would support the ability of *H. halys* to exploit a wide range of host plants and would explain breadth of greater than 150 host plants *H. halys* is known to feed upon (Bergmann et al. 2014). The *P. carbekii* genome is reduced in size relative to known nonhost-restricted *Pantoea* sp., but many gammaproteobacterial intracellular symbionts have genomes less than 1 Mb in size and the retention of genes involved in peptidoglycan and cell wall biosynthesis, and stress response, which are absent in many gammaproteobacterial intracellular mutualist genomes, are among those contributing to the relatively modest size reduction of the *P. carbekii* genome. The multiphasic lifestyle (e.g., within the extrachorion matrix, insect gut during migration to the gastric ceca, and intraluminal crypt-dwelling) of the *P. carbekii* may necessitate a broader genic repertoire than bacterial symbionts that exist solely within host tissues.

Supplementary Material

Supplementary materials S1 and S2 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org>).

Acknowledgments

The authors thank Andrew Michel, of the Department of Entomology and the Ohio Agricultural Research and Development Center, for *Halyomorpha halys* specimens and thoughtful comments. They also thank the anonymous reviewers for their many helpful comments.

Literature Cited

- Abe Y, Mishiro K, Takanashi M. 1995. Symbiont of brown-winged green bug, *Plautia stali* Scott. Appl Entomol Zool. 39:109–115.
- Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.
- Bansal R, Michel A, Sabree ZL. 2014. The crypt-dwelling primary bacterial symbiont of the polyphagous pentatomid pest *Halyomorpha halys* (Hemiptera: Pentatomidae). Environ Entomol. 43:617–625.
- Battesti A, Majdalani N, Gottesman S. 2011. The RpoS-mediated general stress response in *Escherichia coli*. Annu Rev Microbiol. 65:189–213.
- Baumann P. 2005. Biology bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu Rev Microbiol. 59:155–189.
- Begg KJ, Dewar SJ, Donachie WD. 1995. A new *Escherichia coli* cell division gene, ftsK. J Bacteriol. 177:6211–6222.
- Bergmann E, et al. 2014. Host plants of the brown marmorated stink bug in the U.S. [updated April 29, 2014]. Available from: <http://www.stopbmsb.org/where-is-bmsb/host-plants>.
- Bernard CS, Sadasivam M, Shiomi D, Margolin W. 2007. An altered FtsA can compensate for the loss of essential cell division protein FtsN in *Escherichia coli*. Mol Microbiol. 64:1289–1305.
- Bistolas K, Sakamoto R, Fernandes J, Goffredi S. 2014. Symbiont polyphyly, co-evolution, and necessity in pentatomid stinkbugs from Costa Rica. Front Microbiol. 5:349.
- Brown A, Huynh LY, Bolender CM, Nelson KG, McCutcheon JP. 2014. Population genomics of a symbiont in the early stages of a pest invasion. Mol Ecol. 23:1516–1530.
- Burge SW, et al. 2012. Rfam 11.0: 10 years of RNA families. Nucleic Acids Res. 41:D226–D232.
- Cabrera M, Nghiem Y, Miller JH. 1988. *mutM*, a second mutator locus in *Escherichia coli* that generates GC→TA transversions. J Bacteriol. 170:5405–5407.
- Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 25:119–120.
- Caspi R, et al. 2008. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Res. 36:D623–D631.
- Chen L, Roberts MF. 2000. Overexpression, purification, and analysis of complementation behavior of *E. coli* SuhB protein: comparison with bacterial and archaeal inositol monophosphatases. Biochemistry 39:4145–4153.
- Cox EC. 1976. Bacterial mutator genes and the control of spontaneous mutation. Annu Rev Genet. 10:135–156.
- Craig R, Hoskins WM. 1940. Insect biochemistry. Annu Rev Biochem. 9:617–640.
- Cunin R, Glandsdorff N, Piérard A, Stalon V. 1986. Biosynthesis and metabolism of arginine in Bacteria. Microbiol Rev. 50:314–352.
- Danese PN, Pratt LA, Dove SL, Kolter R. 2000. The outer membrane protein, antigen 43, mediates cell-to-cell interactions within *Escherichia coli* biofilms. Mol Microbiol. 37:424–432.
- Dartigalongue C, Missiakas D, Raina S. 2001. Characterization of the *Escherichia coli* sigma E regulon. J Biol Chem. 276:20866–20875.
- Degnan PH, Ochman H. 2012. Illumina-based analysis of microbial community diversity. ISME J. 6:183–194.
- Delcher AL, Phillippy A, Carlton J, Salzberg SL. 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 30:2478–2483.
- Douglas AE. 2013. Microbial brokers of insect-plant interactions revisited. J Chem Ecol. 39:952–961.
- Douglas AE. 2014. Lessons from studying insect symbioses. Cell Host Microbe. 10:359–367.
- Du Q, Wang H, Xie J. 2011. Thiamin (vitamin B1) biosynthesis and regulation: a rich source of antimicrobial drug targets? Int J Biol Sci. 7:41–52.
- Dukan S, et al. 2000. Protein oxidation in response to increased transcriptional or translational errors. Proc Natl Acad Sci U S A. 97:5746–5749.
- Dunbar HE, Wilson AC, Ferguson NR, Moran NA. 2007. Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. PLoS Biol. 5:1006–1015.
- Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 19:2460–2461.
- Fischer S, et al. 2011. Using OrthoMCL to assign proteins to OrthoMCL-DB groups or to cluster proteomes into new ortholog groups. Curr Protoc Bioinformatics. 35:6.12.1–6.12.19.
- Fleming TC, et al. 2010. Dynamic SpoIIIE assembly mediates septal membrane fission during *Bacillus subtilis* sporulation. Genes Dev. 24:1160–1172.
- Fogain R, Graff S. 2011. First records of the invasive pest, *Halyomorpha halys* (Hemiptera: Pentatomidae), in Ontario and Quebec. J Entomol Soc Ont. 142:45–48.
- Fukatsu T, Hosokawa T. 2002. Capsule-transmitted gut symbiotic bacterium of the Japanese common plataspid stinkbug, *Megacopta punctatissima*. Appl Environ Microbiol. 68:389–396.
- Glandsdorff N, Sand G, Verhoef C. 1967. The dual genetic control of ornithine transcarbamylase synthesis in *Escherichia coli* K12. Mutat Res. 4:743–751.
- Glickman BW. 1975. The role of DNA Polymerase I in excision-repair. Basic Life Sci. 5A:213–228.
- Gordon A, Hannon GJ. 2010. “Fastx-toolkit.” FASTQ/A short-reads pre-processing tools [updated June 1, 2014]. Available from: http://hanonlab.cshl.edu/fastx_toolkit.
- Gossard F, Verly WG. 1978. Properties of the main endonuclease specific for apurinic sites of *Escherichia coli* (endonuclease VI). Eur J Biochem. 82:321–332.
- Gunesekere IC, et al. 2006. Comparison of the RpoH-dependent regulon and general stress response in *Neisseria gonorrhoeae*. J Bacteriol. 188:4769–4776.
- Haft DH, Selengut JD, White O. 2003. The TIGRFAMs database of protein families. Nucleic Acids Res. 31:371–373.
- Hoebeker E, Carter ME. 2003. *Halyomorpha halys* (Stål) (Heteroptera: Pentatomidae): a polyphagous plant pest from Asia newly detected in North America. Proc Entomol Soc Wash. 105:225–237.
- Hosokawa T, et al. 2013. Diverse strategies for vertical symbiont transmission among subsocial stinkbugs. PLoS One 8:e65081.
- Hosokawa T, Kikuchi Y, Meng XY, Fukatsu T. 2005. The making of symbiont capsule in the plataspid stinkbug *Megacopta punctatissima*. FEMS Microbiol Ecol. 54:471–477.
- 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.
- Hosokawa T, Kikuchi Y, Shimada M, Fukatsu T. 2008. Symbiont acquisition alters behaviour of stinkbug nymphs. Biol Lett. 4:45–48.
- Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.
- Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.
- Kenyon CJ, Walker GC. 1981. Expression of the *E. coli* *uvrA* gene is inducible. Nature 289:808–810.
- 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. 6:763–775.
- Keseler IM, et al. 2009. EcoCyc: a comprehensive view of *Escherichia coli* biology. Nucleic Acids Res. 37:D464–D470.
- Kikuchi Y, et al. 2012. Primary gut symbiont and secondary, *Sodalis*-allied symbiont of the Scutellerid stinkbug *Cantao ocellatus*. Appl Environ Microbiol. 76:3486–3494.

- Kikuchi Y, Hosokawa T, Fukatsu T. 2007. Insect-microbe mutualism without vertical transmission: a stinkbug acquires a beneficial gut symbiont from the environment every generation. *Appl Environ Microbiol.* 73: 4308–4316.
- Kikuchi Y, Hosokawa T, Nikoh N, Fukatsu T. 2012. Gut symbiotic bacteria in cabbage bugs *Eurydema rugosa* and *Eurydema dominulus* (Heteroptera: Pentatomidae). *Appl Entomol Zool.* 47:1–8.
- Kim J, Copley SD. 2007. Why metabolic enzymes are essential or nonessential for growth of *Escherichia coli* K12 on glucose. *Biochemistry* 46: 12501–12511.
- Kuczyńska-Wisniak D, et al. 2002. The *Escherichia coli* small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured *in vivo* during extreme heat shock. *Microbiology* 148: 1757–1765.
- Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32: 11–16.
- Lasokowska E, Wawrzynow A, Taylor A. 1996. IbpA and IbpB, the new heat-shock proteins, bind to endogenous *Escherichia coli* proteins aggregated intracellularly by heat shock. *Biochimie.* 78: 117–122.
- Ledwidge R, Blanchard JS. 1999. The dual biosynthetic capability of N-acetylornithine aminotransferase in arginine and lysine biosynthesis. *Biochemistry* 38:3019–3024.
- Leskey T, et al. 2012. Pest status of the brown marmorated stink bug, *Halyomorpha halys* in the USA. *Outlooks Pest Manag.* 23:218–226.
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26:589–595.
- Li H, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
- López-Sánchez MJ, et al. 2009. Evolutionary convergence and nitrogen metabolism in *Blattabacterium* strain Bge, primary endosymbiont of the cockroach *Blattella germanica*. *PLoS Genet.* 5:e1000721.
- Martinez-Gomez NC, Palmer LD, Vivas E, Roach PL, Downs DM. 2011. The rhodanese domain of ThiI is both necessary and sufficient for synthesis of the thiazole moiety of thiamine in *Salmonella enterica*. *J Bacteriol.* 193(18):4582–4587.
- Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. 2012. PANDAseq: paired-end assembler for Illumina sequences. *BMC Bioinformatics* 13:31.
- Matsuhisa A, Suzuki N, Noda T, Shiba K. 1995. Inositol monophosphatase activity from the *Escherichia coli* suhB gene product. *J Bacteriol.* 177: 200–205.
- McFall-Ngai M, et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci U S A.* 110:3229–3236.
- McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:297–1303.
- Michel B, Ehrlich SD, Uezem M. 1997. DNA double-strand breaks caused by replication arrest. *EMBO J.* 16:430–438.
- Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop (GCE). Institute of Electrical and Electronics Engineers; 14 November 2010; New Orleans, Louisiana. Redhook, New York: Curran Associates, Inc., p. 1–8.
- Moll I, Afonyushkin T, Vytvytska O, Kaberdin VR, Blasi U. 2003. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. *RNA* 9:1308–1314.
- Moran NA, McCutcheon JP, Nakabachi A. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet.* 42: 165–190.
- Moriya Y, Itoh M, Okuda S, Yoshizawa A, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35:W182–W185.
- Muffler A, Fischer D, Hengge-Aronis R. 1996. The RNA-binding protein HF-1, known as host factor for phage Qbeta RNA replication, is essential for *rpoS* translation in *Escherichia coli*. *Gene Dev.* 10: 1143–1151.
- Nielsen AL, Hamilton GC. 2009. Life history of the invasive species *Halyomorpha halys* (Hemiptera: Pentatomidae) in Northeastern United States. *Ann Entomol Soc Am.* 102:608–616.
- Nikoh N, Hosokawa T, Oshima K, Hattori M, Fukatsu T. 2011. Reductive evolution of a bacterial genome in insect gut environment. *Genome Biol Evol.* 3:702–714.
- Ootsubo MT, et al. 2002. Oligonucleotide probe for detecting Enterobacteriaceae by *in situ* hybridization. *J Appl Microbiol.* 93:60–68.
- Osborn AM, Smith CJ. 2005. Molecular microbial ecology. New York: Garland Science.
- Prado SS, Rubino D, Almeida RPP. 2006. Vertical transmission of a pentatomid caeca-associated symbiont. *Arthropod Biol.* 99:577–585.
- Prado SS, Hung KY, Daugherty MP, Almeida RPP. 2010. Indirect effects of temperature on stink bug fitness, via maintenance of gut-associated symbionts. *Appl Environ Microbiol.* 76:1261–1266.
- Pruesse E, et al. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188–7196.
- Punta M, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.
- Quinlan AR, Hall IM. 2009. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- Rambaut A, Drummond A. 2010. FigTree v1.3.1. Edinburgh (United Kingdom): Institute of Evolutionary Biology, University of Edinburgh. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>.
- Ren D, Bedzyk LA, Thomas SM, Ye RW, Wood TK. 2004. Gene expression in *Escherichia coli* biofilms. *Appl Microbiol Biotechnol.* 64:515–524.
- Revanna KV, Chiu CC, Bierschank E, Dong Q. 2011. GSV: a web-based genome synteny viewer for customized data. *BMC Bioinformatics* 12: 316.
- Reynolds ES. 1963. The use of lead citrate at high pH as an experimental study. *J Cell Biol.* 17:208–212.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16:276–277.
- Rosen R, Ron EZ. 2002. Proteome analysis in the study of the bacterial heat-shock response. *Mass Spectrom Rev.* 21:244–265.
- Russell CW, Bouvaine S, Newell PD, Douglas AE. 2013. Shared metabolic pathways in a coevolved insect-bacterial symbiosis. *Appl Environ Microbiol.* 79:6117–6123.
- Rutherford K, et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16:944–945.
- Sak BD, Eisenstark A, Touati D. 1989. Exonuclease III and the catalase hyperperoxidase II in *Escherichia coli* are both regulated by the *katF* gene product. *Proc Natl Acad Sci U S A.* 86:3271–3275.
- Sandler SJ. 2000. Multiple genetic pathways for restarting DNA replication forks in *Escherichia coli* K-12. *Genetics* 155:487–497.
- Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33:W686–W689.
- Schloss PD, et al. 2009. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 75: 7537–7541.
- Seetin M. 2011. News release: losses to mid-Atlantic apple growers at \$37 million from brown marmorated stink bug. [updated April 12, 2011]. U. S. Apple Association. Available from: http://www.usapple.org/index.php?option=com_content&view=article&id=160:bmsb-loss-midatlantic&catid=8:media-category.
- Selkrig J, et al. 2012. Discovery of an archetypal protein transport system in bacterial outer membranes. *Nat Struct Mol Biol.* 19:506–510.

- Sharon R, Miller C, Ben-Ishai R. 1975. Two modes of excision repair in toluene-treated *Escherichia coli*. *J Bacteriol.* 123:1107–1114.
- Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. *APS. Nature* 407:81–86.
- Shimokawa H, Fujii Y, Furuichi M, Sekiguchi M, Nakabeppu Y. 2000. Functional significance of conserved residues in the phosphohydrolase module of *Escherichia coli* MutT protein. *Nucleic Acids Res.* 28: 3240–3249.
- Sivanathan V, et al. 2006. The FtsK γ domain directs oriented DNA translocation by interacting with KOPS. *Nat Struct Mol Biol.* 13: 965–972.
- Smith DW, Tait RC, Harris AL. 1975. DNA repair in DNA-polymerase-deficient mutants of *Escherichia coli*. *Basic Life Sci.* 5B:473–481.
- Soma A, et al. 2003. An RNA-Modifying enzyme that governs both the codon and amino acid specificities of isoleucine tRNA. *Mol Cell.* 12: 689–698.
- Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol.* 57:758–771.
- Sweetman MD, Palmer LS. 1928. Insects as test animals in vitamin research: I. Vitamin requirements of the flour beetle, *Tribolium confusum duval*. *J Biol Chem.* 77:33–52.
- Tada A, et al. 2011. Obligate association with gut bacterial symbiont in Japanese populations of the southern green stinkbug *Nezara viridula* (Heteroptera: Pentatomidae). *Appl Entomol Zool.* 46:483–488.
- Tamas I, et al. 2008. Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly (A) tracts. *Proc Natl Acad Sci U S A.* 105:14934–14939.
- Tamura K, et al. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 24:631–637.
- 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* (Stål). *PLoS One* 9:e90312.
- Taylor SV, et al. 1998. Thiamin biosynthesis in *Escherichia coli*. Identification of this thiocarboxylate as the immediate sulfur donor in the thiazole formation. *J Biol Chem.* 273:16555–16560.
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2012. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform.* 14:179–192.
- Trun N. 2003. Mutations in the *E. coli* Rep helicase increase the amount of DNA per cell. *FEMS Microbiol Lett.* 226:187–193.
- Untergasser A, et al. 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40:e115.
- Veronese FM, Boccu E, Conventi L. 1975. Glutamate dehydrogenase from *Escherichia coli*: induction, purification and properties of the enzyme. *Biochim Biophys Acta.* 377:217–28.
- Vetek G, Papp V, Haltrich A, Redei D. 2014. First record of the brown marmorated stink bug, *Halyomorpha halys* (Hemiptera: Heteroptera: Pentatomidae), in Hungary, with description of the genitalia of both sexes. *Zootaxa* 3780:194–200.
- Wells TJ, et al. 2008. EhaA is a novel autotransporter protein of enterohemorrhagic *Escherichia coli* O157: H7 that contributes to adhesion and biofilm formation. *Environ Microbiol.* 10:589–604.
- Wermelinger B, Wyniger D, Forster B. 2008. First records of an invasive bug in Europe: *Halyomorpha halys* Stål (Heteroptera: Pentatomidae), a new pest on woody ornamentals and fruit trees? *Mit Sch Ges.* 81:1.
- Wilkinson A, et al. 2005. Analysis of ligation and DNA binding by *Escherichia coli* DNA ligase (LigA). *Biochim Biophys Acta.* 1749: 113–122.
- Xu J, Fonseca DM, Hamilton GC, Hoelmer KA, Nielsen AL. 2014. Tracing the origin of US brown marmorated stink bugs, *Halyomorpha halys*. *Biol Invasions.* 16:153–166.
- Zhou J, Rudd KE. 2013. EcoGene 3.0. *Nucleic Acids Res.* 41:D613–D624.
- Zhu G, Bu W, Gao Y, Liu G. 2012. Potential geographic distribution of brown marmorated stink bug invasion (*Halyomorpha halys*). *PLoS One* 7(2):e31246.

Associate editor: Daniel Sloan