Evolutionary Relationships among Primary Endosymbionts of the Mealybug Subfamily Phenacoccinae (Hemiptera: Coccoidea: Pseudococcidae)†

Matthew E. Gruwell,1* Nate B. Hardy,2 Penny J. Gullan,3 and Katharina Dittmar4

Penn State Erie School of Science, P-1 Prischak Building, 4205 College Drive, Erie, Pennsylvania 16563-02033; Queensland Primary Industries and Fisheries, Entomology, 80 Meiers Road, Indooroopilly, QLD 4068, Australia5; UC Davis, Department of Entomology, One Shields Avenue, Davis, California 95616-85846; and Department of Biological Sciences, 109 Cooke Hall, SUNY at Buffalo, Buffalo, New York 142607

Received 7 June 2010/Accepted 8 September 2010

Mealybugs (Coccoidea: Pseudococcidae) are sap-sucking plant parasites that harbor bacterial endosymbionts within specialized organs. Previous studies have identified two subfamilies, Pseudococcinae and Phenacoccinae, within mealybugs and determined the primary endosymbionts (P-endosymbionts) of the Pseudococcinae to be Betaproteobacteria (“Candidatus Tremblaya princeps”) containing Gammaproteobacteria secondary symbionts. Here, the P-endosymbionts of phenacocine mealybugs are characterized based on 16S rRNA from the bacteria of 20 species of phenacocine mealybugs and four outgroup Puto species (Coccoidea: Putoidea) and aligned to more than 100 published 16S rRNA sequences from symbiotic and free-living bacteria. Phylogenetic analyses recovered three separate lineages of bacteria from the Phenacoccinae, and these are considered to be the P-endosymbionts of their respective mealybug hosts, with those from (i) the mealybug genus Rastroccocus belonging to the Bacteroidetes, (ii) the subterranean mealybugs, tribe Rhizocoeini, also within Bacteroidetes, in a clade sister to cockroach endosymbionts (Blattabacterium), and (iii) the remaining Phenacoccinae within the Betaproteobacteria, forming a well-supported sister group to “Candidatus Tremblaya princeps.” Names are proposed for two strongly supported lineages: “Candidatus Brownia rhizoecola” for P-endosymbionts of Rhizocoeini and “Candidatus Tremblaya phacoca” for P-endosymbionts of Phenacoccinae excluding Rastroccocus and Rhizocoeini. Rates of nucleotide substitution among lineages of Tremblaya were inferred to be significantly faster than those of free-living Betaproteobacteria. Analyses also recovered a clade of Gammaproteobacteria, sister to the P-endosymbiont lineage of aphids (“Candidatus Buchnera aphidicola”), containing the endosymbionts of Putoidea, the secondary endosymbionts of pseudococine mealybugs, and the endosymbionts of several other insect groups.

Many insect species possess specialized cells that contain obligate mutualist bacteria, primary endosymbionts (P-endosymbionts). P-endosymbionts usually show an evolutionary history of codiversification with an insect host and are thought to be essential to the survival of host species (4, 27). Insects with P-endosymbionts are phylogenetically and ecologically heterogeneous but commonly have a nutritionally unbalanced diet. Endosymbiotic bacteria can provide nutritional supplements to their host, permitting exploitation of otherwise inadequate food sources. For example, tsetse fly P-endosymbionts (Wigglesworthia glossinidia) synthesize B vitamins that are lacking in vertebrate blood (1, 3, 32, 33). Several additional examples of blood-feeding insect groups that depend upon P-endosymbionts are known, but the greatest diversity of P-endosymbionts is found in insects that feed on plant sap, a food source poor in amino acids. More than 80% of the approximately 63,000 described insect species with P-endosymbionts feed on plant sap and depend on P-endosymbionts to enrich their diet with amino acids (4, 34, 45, 46).

Insect P-endosymbionts are vertically transmitted (4, 9, 43), and in general, the relationships among the bacteria are highly congruent with those among host taxa due to codiversification. Diverse clades of sap-sucking insects are associated with specific P-endosymbiont lineages. Major sap-sucking groups occur in the hemipteran suborders Auchenorrhyncha (11) and Sternorrhyncha (18). Auchenorrhynchan bugs have codiversified with a Bacteroidetes P-endosymbiont lineage, Sulcia muelleri, with just a few exceptions that are interpreted as losses (28). In addition, a few auchenorrhynchan groups, mainly the sharpshooters, house a second P-endosymbiogen, Baumannia cicadellicina (Proteobacteria), in the same bacteriome as S. muelleri (24, 27, 28, 38). Sternorrhynchan insect groups follow the pattern of old and persistent associations with P-endosymbiont lineages but are more heterogeneous, with each of several family level or other higher taxa characterized by unique P-endosymbiont lineages: Aphididae hosts with Buchnera aphidicola P-endosymbionts (23), Diapridae with Uzunia diaspidica (16), Pseudococcidae with Tremblaya princeps (5, 13), Psyllidea with Carsonella ruddii (40), and Aleyrodoidea with Portiera aleyrodicarum (39).

Mealybugs (Pseudococcidae) are the second largest family of scale insects (Hemiptera: Coccoidea), with ca. 2,000 identified species in more than 270 genera (http://www.sel.barc.usda
TABLE 1. Primers designed for 16S rRNA gene sequences from bacteria

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Description</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>8FBAC</td>
<td>Universal bacteria</td>
<td>AGAGTTTATGATCCTGAGCTAG</td>
</tr>
<tr>
<td>1492R Bac</td>
<td>Universal bacteria</td>
<td>GGTTACCTGTTAGCACTT</td>
</tr>
<tr>
<td>S30 Buch16S</td>
<td>Universal bacteria</td>
<td>CGGCGCGAAGCTTCACATGCAATGC</td>
</tr>
<tr>
<td>A1446 Buch16S</td>
<td>Universal bacteria</td>
<td>TCCCTAGTGTTGACCGCGTTG</td>
</tr>
<tr>
<td>868DIA S</td>
<td>Universal bacteria</td>
<td>GAAATGTAGTGTGCAGGTCAGAG</td>
</tr>
<tr>
<td>TremblINFT</td>
<td>Tremblaya internal</td>
<td>CTACTGCGCAACGACGCGCCGG</td>
</tr>
<tr>
<td>TremblIR</td>
<td>Tremblaya internal</td>
<td>CCGGGCGGGTAGGTG</td>
</tr>
<tr>
<td>RHIZINTF</td>
<td>Brownia internal</td>
<td>GCQUACAGAATGATCACTAG</td>
</tr>
<tr>
<td>RHIZINTR</td>
<td>Brownia internal</td>
<td>CGCGTTACATTATTCGTTG</td>
</tr>
<tr>
<td>RastINTF</td>
<td>Rastroccocus internal</td>
<td>GTAGATCATCCTGGTTC</td>
</tr>
<tr>
<td>RastINR</td>
<td>Rastroccocus internal</td>
<td>GTAGATCATCCTGTTACGGC</td>
</tr>
</tbody>
</table>

*Universal primers were used for amplification, and internal primers were used for sequencing.*

.gov/SCALENET/classif.htm). They have a worldwide distribution and feed on a broad range of hosts but are concentrated in the tropics and favor grasses, legumes, and composites (6, 12). Several mealybug species are serious pests, especially of grapes, figs, apples, citrus, dates, bananas, avocados, and mangos (25). Phylogenetic analyses have identified two primary subgroups within mealybugs, the subfamilies Pseudococcinae and Phenacoccinae (19). Munson et al. (30) and Thao et al. (41) reported monophyly for the mealybug P-endosymbionts, and Downie and Gullan (13) inferred mealybug–P-endosymbiont codiversification, but these studies included samples from only the Pseudococcinae (19). Buchner (8, 9) observed microscopic anatomical differences between the P-endosymbionts of mealybug genera that today are placed in either Pseudococcinae or Phenacoccinae and also considerable variation among phenacoccine species. Hardy et al. (19) hypothesized that a bacteriome containing Tremblaya P-endosymbionts is a synapomorphy (shared feature due to common ancestry) of the Pseudococcinae and that a bacteriome with Tremblaya P-endosymbionts containing Gamma-proteobacteria S-endosymbionts (44) is a synapomorphy of the Pseudococcinae, excluding Maconellicoccus and the Ferrisia group. Here, we estimate relationships among 16S rRNA sequences from the bacterial endosymbionts of Phenacoccinae and other insects, as well as some free-living bacteria.

**MATERIALS AND METHODS**

**Nomenclature.** Most symbiotic bacteria of insects cannot be named formally because the difficulty of culturing them outside their host means that characteristics required for description according to the Bacteriological Code 1990 Revision (22) are not known. Instead, they are given names in the category “Candidatea” (22, 27). For example, the P-endosymbiont of Pseudococcinae should be referred to as “Candidateatus Tremblaya princeps.” However, most authors drop the “Candidateatus” part of the name and italicize the genus and species name (e.g., Tremblaya princeps), which is a practice that we have followed in this paper.

**Taxon sampling.** The data set included 140 sequences, of which 27 were bacterial sequences from 23 species of Phenacoccinae and Putoidea. *Puto,* the only genus in Putoidea, once was included in Pseudococcidae but now is recognized as the only genus in Putoidae, which is a practice that we have followed in this paper.

**DNA extraction and gene amplification.** All genomic DNA samples of Phenacoccinae and Putoidea were the same as those used by Hardy et al. (19). Sequences of 16S rRNA were amplified with two sets of general PCR primers, 16S 8F8BAC and 16S 8FBAC (19), which cover almost the entire gene (1,200 bp). Two PCR protocols were used: a standard three-step PCR protocol with a 52°C annealing temperature, and the EFTD touchdown protocol of Moulton and Wiegmann (29). If initial amplifications did not produce bands that could be sequenced, reamplifications were done using PCR products as the DNA template and the internal primers S30/Buch and a1271Diasp (15, 16). Internal primers were used for double coverage in sequencing: TremblINTF, TremblIR (Tremblaya-like bacteria), RhizINTF, RhizINR (bacteria from Rhizococcus), RastINTF, RastINR (bacteria from Rastroccoccus), and S688Diasp (designed from *Uzinua duspidicola,* the primary symbiont of armored scale insects). All primer sequences are given in Table 1.

Each PCR was run with a total volume of 25 μl made up of 13 μl double-distilled water (ddH₂O), 2 μl 50 mM MgCl₂, 3 μl 10× buffer, 2.5 μl of 10 mM dGTP, dATP, dCTP, and dTTP, 2 μl loading dye, 1 μl each primer (10 pmol/μl), 0.1 μl Taq polymerase, and 1 μl of genomic DNA. Each PCR was run with negative controls. Reactions were confirmed as successful using agarose gel electrophoresis and sent to the High Throughput Genomics Unit (Seattle, WA) for cleaning and DNA sequencing.

The samples PG48, NH12, NH59, PG96, and NH57 were cloned using a TOPO One Shot cloning kit (Invitrogen Corp.) and sequenced with M13 forward and reverse primers. The samples were chosen for cloning to obtain coverage across the Phenacoccinae and Putoidea. Only PG48 returned multiple distinct sequences during cloning.

**Phylogenetic analysis.** Phylogenies were estimated using maximum likelihood (ML), maximum parsimony, and Bayesian methods via the Cipres Portal (http://cipres.ornl.gov). ML trees were reconstructed using RAxML (37) with a GTR + G + I nucleotide substitution model (56) and 100 bootstrap pseudoreplicates under GTR with CAT approximation of gamma-distributed branch length variation. ML trees were estimated using PhyML (38). Parsimony trees were reconstructed using PAUP* (version 4.0b10) with the neighbor joining method of Saitou and Nei (43). Trees were filtered for length and duplication.
Estimated branch lengths among samples of Tremblaya were considerably greater than those among closely related free-living lineages of Betaproteobacteria. To test if the rate of 16S rRNA evolution was significantly different between Tremblaya and its free-living sister group, we compared two molecular clock models: a null 1-rate model and an alternative 2-rate model in which Tremblaya and its sister group were each assigned a unique rate (as in reference 36). A rooted ML topology was inferred from an alignment of Betaproteobacteria 16S rRNA sequences with RAxML (detailed phylogenetic tree available; see Fig. S1 posted at http://bio.bd.psu.edu/gruwell/supplemental) (37). The RAxML analysis consisted of 500 bootstrap replicates under GTR + CAT, with every fifth bootstrap tree used as a starting tree for more thorough optimization under GTR + γ. We used the baseline module of PAML 4.2b (47) to estimate branch lengths under HKY + γ and to calculate the likelihoods of the null and alternative clock models. We evaluated the difference between the likelihoods of our clock models with a likelihood ratio test (14).

Assigning P-endosymbiont status. It was unclear that the amplified 16S rRNA sequences were from bacteria occurring within the host bacteriome. Direct assessment of endosymbiotic properties of bacteria was not undertaken. Endosymbiont status was allocated to bacterial sequences on the basis of three criteria: (i) relationships among P-endosymbiont sequences should reflect those of their hosts, (ii) P-endosymbiont lineages should be related to other known P-endosymbiont lineages, and (iii) it should be possible to amplify P-endosymbiont sequences consistently from closely related specimens. These premises assume that (i) P-endosymbionts are vertically transmitted, (ii) the likelihood of a host’s successful acquisition of an endosymbiont increases if the candidate bacterium is already an endosymbiont, and (iii) P-endosymbionts are always present in the host bacteriome.

RESULTS

P-symbiont identities. Three groups of bacteria were recovered from phenacocine mealybugs (Fig. 1), as detailed in the next three paragraphs.

Bacterial sequences from the Phenacocinae, excluding Rastrococcus and the Rhizoecini, belonged within the Betaproteobacteria, as sister to Tremblaya princeps, the pseudococcine endosymbiont lineage (bootstrap support [BS] = 100; posterior probabilities [PP] = 100). Relationships inferred among these bacterial sequences reflected those inferred among their hosts (19). We propose the designation “Candidatus Tremblaya phenacola” for the P-endosymbiont of mealybug species in the Phenacocinae, excluding Rastrococcus and Rhizoecini. This lineage has unique sequences at the following 16S rRNA sites that distinguish it from Tremblaya princeps (homologous to the Escherichia coli positions): positions 466 to 469 are AATT (ACCA in T. princeps), 817 is G (C in T. princeps), 831 is A (G in T. princeps), 848 and 849 are TT (AA in T. princeps), and 1257 to 1261 are AGTT (GGTC in T. princeps).

Bacterial sequences from the mealybug tribe Rhizoecini were recovered as a clade within Bacteroidetes. A sister relationship with cockroach endosymbionts (Blattabacterium) was weakly supported. We propose the designation “Candidatus Brownia rhizoecola” for the P-endosymbiont of mealybug species in the tribe Rhizoecini. The generic name is in honor of Spencer W. Brown, who was a pioneer of scale insect cytogenetics, and the species name is in reference to the host taxon. This bacterial lineage has sequences that differ from its sister taxon Blattabacterium at the following loci, which correlate to the E. coli positions of 16S rRNA: positions 729 to 736 are GGCAGCT (AGCAGGT in Blattabacterium), 1037 to 1048 are AACGCAAG (AGTACAGG in Blattabacterium), 1117 to 1121 are GTC (ATTT in Blattabacterium), and 1138 to 1140 are GCA (GCTG in Blattabacterium).

Bacterial sequences from the genus Rastrococcus were recovered within Bacteroidetes, in a clade containing a bacterial sequence from Cryptococcus ulmi (Eriococcidae) (15). Strong support (BS = 76; PP = 100) was discovered for a sister relationship between this clade and a group of symbiont sequences from two other scale insect families (Monopileididae and Diaspididae). In this case, relationships among the bacteria do not mirror those among hosts. Therefore, we do not propose a name for this putative P-endosymbiont of Rastrococcus.

Most sequences from species of Putoideae were recovered from within the Gammaproteobacteria, in a weakly supported group comprised of symbiont sequences from Tetropium castaneum (Cerambycidae) (S. Gruenwald, unpublished data), Sitophilus oryzae (Dryophthoridae) (20), Glossina austeni (the symbiont Wigglesworthia glossinidia) (2), Paromenia isabellina (the symbiont Baumannia cicadellinicolica) (38), and Pseudococcinae mealybugs (secondary symbionts) (41). Sequences from a psylloid secondary symbiont (Arsenophonus) (40, 42) and an uncharacterized soft scale (Coccidae) symbiont were sisters to this group with weak support. Monophyly was not identified for the psylloid-associated bacterial sequences within this group. One sequence from Puto albicans was quite distantly related; it was recovered as sister to a Spiroplasma (Mollicutes) lineage from Drosophila with an unknown function.

Rate differences. A significant difference (P < 0.0001) in the rate of 16S rRNA evolution between Tremblaya lineages and free-living Betaproteobacteria was inferred from comparison of the null 1-rate clock model (log likelihood score [lnL] = −10,007.11) to an alternative 2-rate model (lnL = −9,942.20) in which the rates of Tremblaya and its sister group were independent. The estimated local clock rate for Tremblaya was more than 7-fold faster than that of the included free-living Betaproteobacteria.

DISCUSSION

Sap-sucking insect lineages have had relationships with P-endosymbionts characterized by persistence and stability. Nearly all of Auchenorrhyncha share a P-endosymbiont in Sulcia muelleri, and three of four Sternorrhyncha superfamilies are each thought to have a single lineage of P-endosymbiont. In contrast, among scale insects (Coccoidea), mealybugs alone have three distantly related P-endosymbiont lineages. Contrary to the hypothesis of Hardy et al. (19) that Tremblaya P-endosymbionts were a synapomorphy of the Pseudococcinae, many species in the Phenacocinae also possess Tremblaya P-endosymbionts. It is the specific lineage of Tremblaya princeps that is synapomorphic of the Pseudococcinae, with Tremblaya phenacola occurring in all Phenacocinae mealybugs other than Rastrococcus and the Rhizoecini. This suggests an endosymbiotic relationship between Tremblaya and mealybugs that evolved in conjunction with or predates the mealybug radiation.

The drastically elevated rates of 16S rRNA evolution observed among Tremblaya samples in comparison to rates among closely related Betaproteobacteria lineages are consistent with the results of other empirical studies (e.g., references 26 and 42) and with theoretical predictions centered on small populations sizes and a lack of lateral gene transfer in endosymbiont lineages (reviewed in reference 45).

This work further establishes the mealybugs and in particular the Phenacocinae as a system in which to study compara-
Obligate relationships with P-endosymbionts have been more fundamental in the evolutionary history of mealybugs than historical biogeography or host associations (19). Therefore, mealybugs should be an ideal model for researchers wishing to isolate the history of endosymbiosis from other historical and ecological processes.

ACKNOWLEDGMENTS

We thank the generous support from the scale insect community in obtaining specimens. We gratefully acknowledge the laboratory support provided by undergraduates Doug Brummell and Emily DiBlasi (SUNY at Buffalo). P.J.G. thanks Paul Baumann for initiating her interest in mealybug endosymbionts. M.E.G. thanks Benjamin B. Normark for inspiring interest in scale insects and endosymbionts and for continued support and direction during the development of this project. This paper benefited from the critical comments of two anonymous reviewers.

This research was supported by funds from the State University of New York at Buffalo, Pennsylvania State University at Erie, Behrend College, and Hatch funding from the California Agricultural Experiment Station.

REFERENCES


