Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis

Filip Husnik, Filip Husnik*, John P. McCutcheon, and John P. McCutcheon*

*Division of Biological Sciences, University of Montana, Missoula, MT 59812; †Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, Ceske Budejovice 37005, Czech Republic; ‡Faculty of Science, University of South Bohemia, Ceske Budejovice 37005, Czech Republic; and §Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity, Toronto, ON, Canada M5G 1Z8

Edited by Jeffrey D. Palmer, Indiana University, Bloomington, IN, and approved July 19, 2016 (received for review March 8, 2016)

Stable endosymbiosis of a bacterium into a host cell promotes cellular and genomic complexity. The mealybug Planococcus citri has two bacterial endosymbionts with an unusual nested arrangement: the γ-proteobacterium Moranella endobia lives in the cytoplasm of the β-proteobacterium Tremblaya princeps. These two bacteria, along with genes horizontally transferred from other bacteria to the P. citri genome, encode gene sets that form an interdependent metabolic patchwork. Here, we test the stability of this three-way symbiosis by sequencng host and symbiont genomes for five diverse mealybug species and find marked fluidity over evolutionary time. Although Tremblaya is the result of a single infection in the ancestor of mealybugs, the γ-proteobacterial symbionts result from multiple replacements of inferred different ages from related but distinct bacterial lineages. Our data show that symbiont replacement can happen even in the most intricate symbiotic arrangements and that preexisting horizontally transferred genes can remain stable on genomes in the face of extensive symbiont turnover.

Significance

Mealybugs are plant sap-sucking insects with a nested symbiotic arrangement, where one bacterium lives inside another bacterium, which together live inside insect cells. These two bacteria, along with genes transferred from other bacteria to the insect genome, allow the insect to survive on its nutrient-poor diet. Here, we show that the innermost bacterium in this nested symbiosis was replaced several times over evolutionary history. These results show that highly integrated and interdependent symbiotic systems can experience symbiont replacement and suggest that similar dynamics could have occurred in building the mosaic metabolic pathways seen in mitochondria and plastids.

Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis

Filip Husnik, Filip Husnik*, John P. McCutcheon, and John P. McCutcheon*

*Division of Biological Sciences, University of Montana, Missoula, MT 59812; †Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, Ceske Budejovice 37005, Czech Republic; ‡Faculty of Science, University of South Bohemia, Ceske Budejovice 37005, Czech Republic; and §Canadian Institute for Advanced Research, Program in Integrated Microbial Biodiversity, Toronto, ON, Canada M5G 1Z8

Edited by Jeffrey D. Palmer, Indiana University, Bloomington, IN, and approved July 19, 2016 (received for review March 8, 2016)

Stable endosymbiosis of a bacterium into a host cell promotes cellular and genomic complexity. The mealybug Planococcus citri has two bacterial endosymbionts with an unusual nested arrangement: the γ-proteobacterium Moranella endobia lives in the cytoplasm of the β-proteobacterium Tremblaya princeps. These two bacteria, along with genes horizontally transferred from other bacteria to the P. citri genome, encode gene sets that form an interdependent metabolic patchwork. Here, we test the stability of this three-way symbiosis by sequencng host and symbiont genomes for five diverse mealybug species and find marked fluidity over evolutionary time. Although Tremblaya is the result of a single infection in the ancestor of mealybugs, the γ-proteobacterial symbionts result from multiple replacements of inferred different ages from related but distinct bacterial lineages. Our data show that symbiont replacement can happen even in the most intricate symbiotic arrangements and that preexisting horizontally transferred genes can remain stable on genomes in the face of extensive symbiont turnover.

Significance

Mealybugs are plant sap-sucking insects with a nested symbiotic arrangement, where one bacterium lives inside another bacterium, which together live inside insect cells. These two bacteria, along with genes transferred from other bacteria to the insect genome, allow the insect to survive on its nutrient-poor diet. Here, we show that the innermost bacterium in this nested symbiosis was replaced several times over evolutionary history. These results show that highly integrated and interdependent symbiotic systems can experience symbiont replacement and suggest that similar dynamics could have occurred in building the mosaic metabolic pathways seen in mitochondria and plastids.

Author contributions: F.H. and J.P.M. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The nine complete endosymbiont genomes, five draft assemblies of insect genomes, and raw data have been deposited into the European Nucleotide Archive (ENA; accession nos.: Macronellicoccus hirsutus; PRJEB12066; Ferrisia virgata; PRJEB12067; Pseudococcus longispinus; PRJEB12068; Paracoccus marginatus; PRJEB12069; and Trinymus perrisi; PRJEB12071). Unannotated draft genomes of two Enterobacteriaceae symbionts from P. longispinus mealybugs and a B-supergroup Wolbachia strain sequenced from M. hirsutus mealybugs were deposited in Figshare (accession nos. 10.6084/m9.figshare.2010393 and 10.6084/m9.figshare.2010390).

1To whom correspondence may be addressed. Email: filip.husnik@gmail.com or john.mccutcheon@umontana.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1603910113/-/DCSupplemental.
symbionts themselves (9–11, 14). Another solution to avoid host extinction is to replace the degenerating symbiont with a fresh one or supplement it with a new partner. Examples of symbiont replacement and supplementation are replete in insects, occurring in at least the sap-feeding Auchenorrhyncha (23, 32–34), psyllids (22, 35), aphids (25, 36, 37), lice (38), and weevils (39, 40). When viewed over evolutionary time, it becomes clear that endosymbioses can be dynamic—both genes and organisms come and go. It follows that any view of a symbiotic system established from just one or a few host lineages might provide only a snapshot of the complexity that built the observed relationship.

Mealybugs (Hemiptera: Cocidae: Pseudococcidae) are a group of phloem-sap sucking insects that contain most of the symbiotic complexity described above. All of these insects depend on bacterial endosymbionts to provide them with essential amino acids missing from their diets, but nutrient provisioning is accomplished in dramatically different ways in different mealybug lineages. One subfamily, the Phenacocinae, has a single β-proteobacterial endosymbiont called *Tremblaya phenacola*, which provides essential amino acids and vitamins to the host insect (9, 41). In the other subfamily of mealybugs, the Pseudococcinae, *Tremblaya* has been supplemented with a second bacterial endosymbiont, a γ-proteobacterium named *Moranella endobia* in the mealybug Planococcus citri (PCIT). Although symbiont supplementation is not uncommon, what makes this symbiosis unique is its structure: *Moranella* stably resides in the cytoplasm of its partner bacterial symbiont, *Tremblaya princeps* (42–45).

The organisms in the nested three-way *P. citri* symbiosis are intimately tied together at the metabolic level. *T. princeps* PCIT has one of the smallest bacterial genomes ever reported, totaling 139 kb in length, encoding only 120 protein-coding genes, and lacking many translation-related genes commonly found in the most extremely reduced endosymbiont genomes (42). Many metabolic genes missing in *Tremblaya* are present on the *M. endobia* PCIT genome. Together with their host insect, these two symbionts are thought to work as a “metabolic patchwork” to produce nutrients needed by all members of the consortium (42). The symbiosis in *P. citri* is further supported by numerous HGTs from several different bacterial donors to the insect genome, but not from *Tremblaya* or *Moranella*. These genes are up-regulated in the insect’s symbiotic tissue (the bacteriome) and fill in many of the remaining metabolic gaps inferred from the bacterial endosymbiont genomes (9).

Other data suggest additional complexity in the mealybug symbiosis. Phylogenetic analyses of the intra-*Tremblaya* endosymbionts show that, although different lineages of mealybugs in the Pseudococcinae all possess γ-proteobacterial endosymbionts related to *Sodalis*, these bacteria do not show the coevolutionary patterns typical of many long-term endosymbionts (43, 44, 46). Developmental studies suggest that *Tremblaya* and its resident γ-proteobacteria can be differentially regulated by the host (44, 47). These data raise the possibility that the innermost bacterium of this symbiosis is labile and may have resulted from separate acquisitions, or that the original intra-*Tremblaya* symbiont has been replaced in different mealybug lineages. What is not clear is when these acquisitions may have occurred and what effect they have had on the symbiosis. Here, we use host and symbiont genome sequencing from seven mealybug species (five generated for this study) to better understand how complex interdependent symbioses may develop over time in the context of gene and organism acquisition and loss.

**Results**

**Overview of Our Sequencing Efforts.** We generated genome data for five diverse Pseudococcinae mealybug species, in total closing nine symbiont genomes into single circular-mapping molecules (five genomes from *Tremblaya* and four from the *Sodalis*-allied γ-proteobacterial symbionts) (Table 1). Unexpectedly, we detected γ-proteobacterial symbionts in *Maconellicoccus hirsutus* (MHIR), which was not previously reported to harbor intrabacterial symbionts inside *Tremblaya* cells (Figs. 1–3 and Fig. S1). We also found that *Pseudococcus longispinus* (PLON) harbored two γ-proteobacterial symbionts, each with a complex genome larger than 4 Mbp; these genomes were left as a combined draft assembly of 231 contigs with a total size of 8,191,698 bp and an N50 of 82.6 kbp (Table 1).

We also assembled five mealybug draft genomes (Table 1). Because our assemblies were generated only from short-insert paired end data, the insect draft genomes consisted primarily of numerous short scaffolds (Fig. S2 and Table S1).

**Verifying the Intra-*Tremblaya* Location for the γ-Proteobacterial Endosymbionts.** The intra-*Tremblaya* location of the γ-proteobacterial symbionts has been established for mealybugs in the genera *Planococcus* (44, 45), *Pseudococcus* (44, 48), *Crisicoccus* (49), *Antonina*, *Antonella*, *Rhodania*, *Trionymus*, and *Ferrisia* (50). However, to our knowledge, the organization of *Tremblaya* and its partner γ-proteobacteria has never been investigated in *Maconellicoccus* or *Paracoccus*. We therefore verified that both *M. hirsutus* and *Paracoccus marginatus* (PMAR) had the expected γ-proteobacteria inside *Tremblaya* structure using FISH microscopy (Fig. S3).

**Tremblaya Genomes Are Stable in Size and Structure; the γ-Proteobacterial Genomes Are Not.** Genomes from all five *T. princeps* species (those that have a γ-proteobacterial symbiont) are completely syntenic and similar in size, ranging from 138 to 143 kb (Fig. 1). The gene contents are also similar, with 107 protein-coding genes shared in all five *Tremblaya* genomes. All differences in gene content come from gene loss or nonfunctionalization in different lineages (Fig. 1). Four pseudogenes (argS, mnnmG, lpd, and rmnH) are shared in all five *T. princeps* genomes, indicating that some pseudogenes can be retained in *Tremblaya* for long periods of time. Pseudogene numbers were notably higher and coding densities were lower in *T. princeps* genomes from *P. marginatus* and *Trionymus perrissii* (TPER) (Fig. 1 and Table 1).

In contrast to the genomic stability observed in *Tremblaya*, the genomes of the γ-proteobacterial symbionts vary dramatically in size, coding density, and gene order (Figs. 1 and Table 1). These genomes range in size from 353 to ~4,000 kb (*P. longispinus* contains two ~4,000-kb genomes from different γ-proteobacteria) and are all notably different from the 539-kb *Moranella* genome of *P. citri* (42).

**Phylogenetic Analyses Confirm the Intra-*Tremblaya* γ-Proteobacterial Symbionts Result from Multiple Infections.** The lack of conservation in γ-proteobacterial genome size and structure, combined with data showing that their phylogeny does not mirror that of their mealybug or *Tremblaya* hosts (43, 44) (Fig. S1), supports early hypotheses that the γ-proteobacterial symbionts of diverse mealybug lineages result from multiple unrelated infections (43, 44). Although the *Sodalis*-allied clade is extremely hard to resolve because of low taxon sampling of facultative and free-living relatives, nucleotide bias, and rapid evolution in obligate symbionts, none of our analyses indicate a monophyletic group of mealybug symbionts congruent with the host and *Tremblaya* trees (Fig. 2 and Fig. S1).

**Draft Insect Genomes Reveal the Timing of Mealybug HGTs.** Gene annotation of low-quality draft genome assemblies is known to be problematic (51). We therefore verified that our mealybug assemblies were sufficient for our purpose of establishing gene presence or absence by comparing our gene sets with databases containing core eukaryotic [Core Eukaryotic Genes Mapping Approach (CEGMA)] and Arthropod [Benchmarking Universal Single-Copy Orthologs (BUSCO)] gene sets. CEGMA scores surpass 98% in all of our assemblies, and BUSCO Arthropoda scores range from 66 to 76% (Table S1). We note that the low scores against the BUSCO database likely reflect the hemipteran origin of mealybugs rather than our fragmented assembly; the high-quality
Table 1. Genome statistics for mealybug endosymbionts and draft mealybug genomes

<table>
<thead>
<tr>
<th>Mealybug species</th>
<th>P. avenae</th>
<th>M. hirsutus</th>
<th>F. virgata</th>
<th>P. citri</th>
<th>P. longispinus</th>
<th>P. tennisi</th>
<th>P. marginatus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mealybug abbreviation</td>
<td>PAVE</td>
<td>MHIR</td>
<td>FVIR</td>
<td>PCIT</td>
<td>PLON</td>
<td>TPER</td>
<td>PMAR</td>
</tr>
<tr>
<td>Total assembly size (bp)</td>
<td>170,756</td>
<td>304,570</td>
<td>377,829</td>
<td>284,990</td>
<td>237,582</td>
<td>191,208</td>
<td></td>
</tr>
<tr>
<td>Total o. of scaffolds</td>
<td>NA</td>
<td>12,889</td>
<td>32,723</td>
<td>167,514</td>
<td>66,857</td>
<td>80,386</td>
<td>60,102</td>
</tr>
<tr>
<td>N50</td>
<td>NA</td>
<td>47,025</td>
<td>22,300</td>
<td>12,551</td>
<td>7,078</td>
<td>10,126</td>
<td>4,908</td>
</tr>
<tr>
<td>BUSCOs Arthropoda (n=1,675)</td>
<td>76%</td>
<td>76%</td>
<td>71%</td>
<td>70%</td>
<td>66%</td>
<td>72%</td>
<td></td>
</tr>
<tr>
<td>BUSCOs Eukaryota (n=429)</td>
<td>85%</td>
<td>84%</td>
<td>80%</td>
<td>78%</td>
<td>77%</td>
<td>82%</td>
<td></td>
</tr>
<tr>
<td>CEGMA (n=248; including partial)</td>
<td>NA</td>
<td>99.19%</td>
<td>97.98%</td>
<td>98.79%</td>
<td>98.39%</td>
<td>99.6%</td>
<td>98.79%</td>
</tr>
</tbody>
</table>

**Treblaya symbiont**

<table>
<thead>
<tr>
<th>Genome size (plasmid size if present)</th>
<th>T. phenacola</th>
<th>T. princeps</th>
<th>T. princeps</th>
<th>T. princeps</th>
<th>T. princeps</th>
<th>T. princeps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average fragment coverage</td>
<td>NA (454 data)</td>
<td>795</td>
<td>663</td>
<td>374</td>
<td>1,326</td>
<td>2,364</td>
</tr>
<tr>
<td>G &lt; C (%)</td>
<td>42.2</td>
<td>61.8</td>
<td>58.3</td>
<td>58.8</td>
<td>58.9</td>
<td>57.9</td>
</tr>
<tr>
<td>CDS (pseudogenes)</td>
<td>178 (3)</td>
<td>136 (7)</td>
<td>132 (13)</td>
<td>125 (16)</td>
<td>134 (15)</td>
<td>116 (31)</td>
</tr>
<tr>
<td>CDS coding density (%)</td>
<td>86.3</td>
<td>77.2</td>
<td>69.3</td>
<td>66.0</td>
<td>70.7</td>
<td>59.2</td>
</tr>
<tr>
<td>mRNAs</td>
<td>3</td>
<td>[14]</td>
<td>[3]</td>
<td>[10]</td>
<td>[6]</td>
<td>[16]</td>
</tr>
<tr>
<td>Protein-coding DNA sequence</td>
<td>Not present</td>
<td>D. endobia</td>
<td>G. endobia</td>
<td>Mo. endobia</td>
<td>PLON1 and PLON2</td>
<td></td>
</tr>
<tr>
<td>Genome size (plasmid size)</td>
<td>834,723 bp (11,828 bp)</td>
<td>938,041 bp</td>
<td>538,294 bp</td>
<td>8,190,816+</td>
<td>628,221 bp (8,492 bp)</td>
<td>352,837 bp</td>
</tr>
<tr>
<td>Average fragment coverage</td>
<td>NA</td>
<td>121 (38)</td>
<td>372</td>
<td>827</td>
<td>30</td>
<td>559 (312; 1,750)</td>
</tr>
<tr>
<td>G &lt; C (%)</td>
<td>NA</td>
<td>44.2</td>
<td>28.9</td>
<td>43.5</td>
<td>53.9</td>
<td>42.8</td>
</tr>
<tr>
<td>CDS (pseudogenes)</td>
<td>NA</td>
<td>554 (99)</td>
<td>461 (30)</td>
<td>419 (24)</td>
<td>NA (NA)</td>
<td>510 (16)</td>
</tr>
<tr>
<td>CDS coding density (%)</td>
<td>NA</td>
<td>59.8</td>
<td>48.1</td>
<td>77.4</td>
<td>NA</td>
<td>80.4</td>
</tr>
<tr>
<td>rRNAs</td>
<td>3</td>
<td>[40]</td>
<td>14</td>
<td>3</td>
<td>39</td>
<td>8</td>
</tr>
<tr>
<td>rrnDNA</td>
<td>NA</td>
<td>3</td>
<td>[3]</td>
<td>[10]</td>
<td>3</td>
<td>41</td>
</tr>
</tbody>
</table>

**References**

9 This study

**H. endobia** codes two plasmids of 3,244 and 5,248 bp. Extended assembly metrics for draft mealybug genomes are available as Table S2.

Combined assembly size for both γ-proteobacterial symbionts in PLON. CDS, protein-coding DNA sequence; NA, not applicable; ncRNA, noncoding RNA; PAVE, Phenacoccus avenae.

We conclude that our mealybug draft assemblies are sufficient for determining the presence or absence of bacterial HGTs.

We first sought to confirm that the HGTs found previously in the *P. citri* genome (9) were present in other mealybug species (Tables S2 and S3) and establish the timing of these transfers. Consistent with our previous findings (9), there were no well-supported HGTs of *Treblaya* origin detected in any of our mealybug assemblies. Our data show that the acquisition of some HGTs with our previous findings (9), there were no well-supported HGTs of *Treblaya* origin detected in any of our mealybug assemblies.

We note that several HGTs were likely acquired after the divergence of the *Macconellaclis* clade [cytolytic synthase A (cysK), betalactamase (b-lact), type III effector (T3ef), and d-alanine-d-alanine ligase B (ddlB)]. One of these genes, cysK, clusters with sequences from other *Sodalis*-allied bacteria, consistent with a possible origin from an early γ-proteobacterial intrabacterial symbiont (Dataset S1). We note that cysK has undergone tandem duplication in *P. longispinus, Ferrisia virgata* (FVIR), and *P. citri* (Fig. S24 and Tables S2 and S3), which was also observed for several other HGTs (tms, b-lact, T3ef, chiA, ankyrin repeat proteins, and AAA-ATPases). Most of the HGTs found in only one or two mealybug species are related to peptidoglycan metabolism and were assembled on shorter scaffolds with few insect genes on them. Possible HGT losses of tms in FVIR and ddlB in *P. marginatus* were detected based on our assemblies. Except in three cases (amiD, murC, and DURI), HGT candidates detected from several mealybug species shared a significant amount of sequence similarity and clustered as a single clade in our phylogenies (Dataset S1), suggesting that these transfers resulted from single events.

Evolution of the Metabolic Patchwork. We previously found complementary patterns of gene loss and retention between *Treblaya, Moranella*, and the mealybug host in the *P. citri* symbiosis (9, 42). Our comparative genomic data allow us to see how genes are retained or lost in different genomes in multiple lineages that have γ-proteobacterial symbionts of different inferred ages (Fig. 3). These data also allow us to observe how new symbionts evolve in response to the presence of both preexisting symbionts and horizontally transferred genes.

Overall, our data point to an extremely complex pattern of gene loss and retention in the mealybug symbiosis (Fig. 3). Some pathways, such as those for the production of lysine, phenylalanine, and methionine, show a relatively similar patchwork pattern in all mealybugs, with gene retention interspersed between *Treblaya*, its γ-proteobacterial endosymbiont, and/or the host. Gene retention patterns from many other pathways, however, show much less
predictable patterns. The isoleucine, valine, leucine, threonine, and histidine pathways show a tendency toward \( \gamma \)-proteobacterial dominance in \( M. \text{hirsutus} \), \( F. \text{virgata} \), and \( P. \text{citri} \) (that is, gene retention in \( \gamma \)-proteobacterial genome) but with a clear shift toward \( \gamma \)-proteobacterial-dominated biosynthesis in \( P. \text{marginatus} \) and \( T. \text{perrisii} \). Other pathways, such as tryptophan, show \( \gamma \)-proteobacterial dominance in all mealybug symbioses but with reliance on at least one \( \beta \)-proteobacterial gene in \( P. \text{citri} \), \( P. \text{marginatus} \), and \( T. \text{perrisii} \). In the arginine pathway, gene retention is dominated by \( \beta \)-proteobacterial \( T. \text{perrisii} \) gene but with reliance on at least one \( \gamma \)-proteobacterial endosymbiont in all other lineages, with sporadic loss of \( \text{Tremblaya} \) genes in different lineages. Overall, \( M. \text{hirsutus} \) encodes the most \( \gamma \)-proteobacterial genes, whereas TPER shows the opposite pattern.

Gene Retention Patterns for Translation-Related Genes in \( \text{Tremblaya} \).

In contrast to metabolic genes involved in nutrient production, the retention patterns for genes involved in translation vary little between mealybug species (Fig. 3). As first shown in \( \text{Tremblaya} \) PCIT (42), none of the additional \( \text{Tremblaya} \) genomes that we report here encode any functional aminoacyl tRNA synthetase, with an exception of one likely functional gene (\( \text{cysS} \)) in \( \text{P. princeps} \) PLON, which is present as a pseudogene in several other lineages of \( \text{Tremblaya} \). Furthermore, all \( \text{Tremblaya} \) genomes have lost key translational control proteins that are typically retained even in the smallest endosymbiont genomes, such as ribosome recycling factor, 1-methionyl-tRNA\( ^{\text{Met}} \), \( N \)-formyltransferase, and peptide deformylase. The translational release factors RF-1 and RF-2 (\( \text{prfAB} \)) and elongation factor (EF) EF-Ts (\( \text{uf} \)) are present only in the gene-rich \( T. \text{princeps} \) MHIR genome and absent or pseudogenized in all other \( T. \text{princeps} \) genomes. Initiation factors (IFs) IF-1, IF-2, and IF-3 (\( \text{infABC} \)) and EFs EF-Tu and EF-G (\( \text{ufA} \) and \( \text{ufX} \)) are retained in all \( \text{Tremblaya} \) genomes, as are most ribosomal proteins (Dataset S24).

Taxonomy of Mealybug Endosymbionts. The naming convention in the field of insect endosymbiosis has been to keep the species names constant for lineages of endosymbiotic bacteria resulting from single infections, even if they exist in different species of host insects. The host is denoted by appending a specific abbreviation to the end of the endosymbiont name (e.g., \( T. \text{princeps} \) PCIT for \( T. \text{princeps} \) from \( P. \text{citri} \)). However, our data show that the intra-\( \text{Tremblaya} \) \( \gamma \)-proteobacterial symbionts are not from the same infection; they result from independent endosymbiotic events from clearly discrete lineages within the \( \text{Sodalis} \) clade (Fig. 2). Following convention, we have chosen to give these \( \gamma \)-proteobacteria different genus names but unite them by retaining the “endobia” species denomination for each one (such as in \( \text{Moranella endobia} \)).

We propose the following Candidatus status names for four lineages of intra-\( \text{Tremblaya} \) \( \gamma \)-proteobacterial symbionts of mealybugs for which we have completed a genome. First, \( \text{Candidatus} \) \( \text{Doolittlea} \) endobia MHIR is for the endosymbiont from \( M. \text{hirsutus} \). This name honors the American evolutionary biologist W. Ford Doolittle (1941–) for his contributions to our understanding of HGT and endosymbiosis. Second, \( \text{Candidatus} \) \( \text{Gullanella} \) endobia FVIR is for the endosymbiont from \( F. \text{virgata} \). This name honors the Australian entomologist Penny J. Gullan.
(1952) for her contributions to numerous aspects of mealybug biology and taxonomy. Third, Candidatus Mikella endobia PMAR is for the endosymbiont from P. marginatus. This name honors the Canadian biochemist Michael W. Gray (1943–) for his contributions to our understanding of organelle evolution. Fourth, Candidatus Hoaglandella endobia TPER is for the endosymbiont from T. perrisii. This name honors the American biochemist Mahlon B. Hoagland (1921–2009) for his contributions to our understanding of the genetic code, including the codiscovery of tRNA. All of the names that we propose could be extendible to related mealybugs species (e.g., G. endobia for other members of the Ferussia clade) if future phylogenetic analyses show that these symbionts result from the same infection. For simplicity, we use all endosymbiont names without the Candidatus denomination.

**Discussion**

**Diversity of Intra-Tremblaya Symbiont Genomes Suggests Multiple Replacements.** Phylogenetic analyses based on rRNA and protein-coding genes from the γ-proteobacterial endosymbionts of mealybugs first indicated their origins from multiple unrelated bacteria (43, 44). What was unclear from these data was the order and timing of the γ-proteobacterial infections and how these infections affected the other members of the symbiosis. We imagine three possible scenarios that could explain these phylogenetic and genomic data (Fig. 5). The first is that there was a single γ-proteobacterial acquisition in the ancestor of the Pseudococcinae that has evolved idiosyncratically as mealybugs diversified over time, leading to seemingly unrelated genome structures and coding capacities (the “idosyncratic” scenario) (Fig. 5A). The second is that the γ-proteobacterial infections occurred independently, each establishing symbioses inside Tremblaya in completely unrelated and separate events (the “independent” scenario) (Fig. 5B). The third is that there was a single γ-proteobacterial acquisition in the Pseudococcinae ancestor that has been replaced in some mealybug lineages over time (the “replacement” scenario) (Fig. 5C). The idiosyncratic scenario is easy to disregard, because although acquisition of a symbiont followed by rapid diversification of the host might result in different patterns of genome evolution in different lineages, it should result in monophyletic clustering in phylogenetic trees. Previous phylogenetic work as well as our phylogenomic data (Fig. 2) show that the γ-proteobacteria that have infected different mealybugs have originated from clearly distinct (and well-supported) bacterial lineages.

The independent and replacement scenarios are more difficult to tell apart with our data, and the true history of the symbiosis may have involved both. However, we favor symbiont replacement as the main mechanism that generated the complexity that we see in mealybugs, primarily because of the large differences in size observed in the γ-proteobacterial genomes (Fig. 1 and Table 1). Genome size is strongly correlated to endosymbiotic age in bacteria, especially at the onset of symbiosis, when genome reduction can be rapid (53–57). Most relevant to our argument here is the speed with which genome reduction has been shown to take place in Sodalis-allied bacteria closely related to the γ-proteobacterial symbionts of mealybugs (34, 58, 59). It has been estimated that as much as 55% of an ancestral Sodalis genome was lost on the transition to endosymbiosis in a mere ~28,000 y, barely enough time for 1% sequence divergence to accumulate between the new symbiont and a free-living relative (58). Our general assumption is, therefore, that recently established endosymbionts should have larger genomes than older symbionts. However, we note that genome reduction is not a deterministic process related to time, especially as the symbiosis ages. It is clear that, in some insects housing pairs of ancient symbionts with highly reduced genomes, the older endosymbiont can have a larger genome than the newer symbiont (60).

The evidence for recent replacement is most obvious in P. longispinus (Fig. 3 and Table 1). This symbiosis harbors two related γ-proteobacterial symbionts (61), each with a rod-like cell shape, although it is currently unclear if both bacteria reside within Tremblaya (48). Both of these genomes are about 4 Mb in size (Table 1), approximately the same size as the recently acquired Sodalis symbionts from tsetse fly (4.3 Mb) (62) and rice weevil (4.5 Mb) (59). These morphological and genomic features as well as their relatively short branches in Fig. 2 all suggest that...
The γ-proteobacterial symbionts are recent acquisitions in the P. longispinus symbiosis. The P. longispinus replacement seems so recent that the stereotypical complementary patterns of gene loss and retention have not had time to accumulate between the γ-proteobacteria and Tremblaya (Fig. 3). However, Tremblaya PLON is missing the same translation-related genes (aside from cysS) as all other Tremblaya, indicating that it has long ago adapted to the presence of a (now eliminated) bacterium living in its cytoplasm. Comprehensive analyses of the two γ-proteobacterial genomes from P. longispinus are ongoing and will be published elsewhere.

We hypothesize that the larger, gene-rich γ-proteobacterial genomes that we describe here are the result of symbiont replacements of an ancestral γ-proteobacterial endosymbiont rather than completely independent infections in different mealybug lineages. We suspect that the massive loss in key translation-related genes...
How Did the Bacteria Within a Bacterium Structure Start, and Why Does It Persist? In extreme cases of endosymbiotic genome reduction, genes required for the generation of a cell envelope, along with other fundamental processes, are lost (12, 13). This phenomenon is seen in Tremblaya, where even the largest genome (from Phenacococcinae/Pseudococcinae divergence—that is, right at the acquisition of the first γ-proteobacterial symbiont)—much richer taxon sampling would need to test the hypothesis that this was, in fact, the original symbiont lineage (Fig. 2). We also note that, in at least one other case, bacteria from the Sodalis group have established multiple repeated infections in a replacement-like pattern (38).

![Fig. 5. Three possible scenarios that built the mealybug symbiosis. Independent γ-proteobacterial acquisitions are shown as arrows, and replacements are noted with Rs above the arrow. Colors represent the different γ-proteobacterial genomes shown in Fig. 1. (A) The idiosyncratic scenario, where a single γ-proteobacterial acquisition evolved differently as mealybugs diverged, leading to different genome sizes and structures in extant mealybugs. (B) The independent scenario, where the different sizes and structures of the γ-proteobacterial genomes shown in Fig. 1 result from completely independent acquisitions. (C) The replacement scenario, where the different sizes and structures of the γ-proteobacterial genomes shown in Fig. 1 result from several replacements of an ancestral γ-proteobacterial symbiont.](image)

Evolution of Organelles and the Timing of HGT. It is widely accepted that the mitochondria found across eukaryotes are related back to a single common α-proteobacterial ancestor (68) and that the plastids resulted from a single cyanobacterial infection (69). What is less clear is what happened before these endosymbiont lineages were fixed into organelles. The textbook concept is that a bacterium was taken up by a host cell, transferred most of its genes, and became the mitochondrion or plastid (70). This idea becomes more complicated when the taxonomic affiliation of bacterial genes on eukaryotic genomes is examined (71–74). For example, only about 20% of mitochondria-related horizontally transferred genes have strong α-proteobacterial phylogenetic affinities (72). The signals for the remaining 80% are either too weak to confidently place the gene or show clear affiliation with other bacterial groups (71, 72). Hypotheses that explain these data fall roughly into two camps. Some imagine a gradual process where multiple taxonomically diverse endosymbioses may have occurred—and transferred genes—before the final α-proteobacterial symbiont was fixed. That is, the mitochondria arrived rather late in the evolution of a eukaryotic-like cell that already contained many bacterial genes resulting from HGT of previous symbionts (75, 76). Others favor a more abrupt “mitochondria early” scenario, where an endosymbiont with a taxonomically diverse mosaic genome made the transition to becoming the mitochondrion in a single endosymbiotic event, transferring its genes during the process. In this scenario, the mosaic nature of the extant eukaryotic genomes resulted from the “inherited chimerism” of the lone mitochondria bacterial ancestor because of the propensity of bacteria to participate in HGT with distantly related groups (73, 77, 78).

We suggest that the data reported here indirectly support the gradualist or mitochondria late view of organelle evolution. We find that the majority of nutrient-related HGTs occurred before the divergence of the Phenacococcinae and Pseudococcinae (Figs. 3 and 4) and therefore before the establishment of any γ-proteobacterial symbiont. In particular, HGTs in the riboflavin and lysine pathways were retained on the insect genomes as the first γ-proteobacterial symbiont was established and new γ-proteobacterial symbionts replaced old ones (Figs. 2 and 3). Our results make it clear that HGTs can remain stable on host genomes for millions of years, even after the addition or replacement of symbionts that share pathways with these genes, and directly show how mosaic metabolic pathways can be built gene by gene as symbionts come and go over time. We note that the “shopping bag” hypothesis (79), which argues that establishment of an endosymbiosis should be regarded as a continuous process involving a number of partners rather than a single event involving two partners, fits our data remarkably well. Of course, our data do not rule out inherited chimerism as a contributor to the taxonomic diversity of genes that support organelle function, because many bacterial genomes are taxonomically mosaic because of HGT (73). As with most solutions to endosymbiotic problems, the true answer is likely a complicated mixture of both processes.

Using Symbiont Supplementation and Replacement to Claw Out of the Rabbit Hole. At the onset of a nutritional symbiosis, a new organism comes on board and allows access to a previously inaccessible food source. Rapid adaptation and diversification can occur—the new symbiont adapts to the host, the host adapts to the symbiont, and the entire symbiosis expands in the newly available ecological niche. However, cases where a bacterial symbiont takes up stable residence in a host cell also seem to lead to irreversible

cently suggested to sometimes reside within Sulcia cells in the leaf-hoppers Cicadella viridis and Macrossteles laevis (66, 67). Although these studies were based only on EM imaging and not confirmed by specific probes (e.g., with FISH), it is possible that symbioses formed by bacteria taking up residence inside of degenerate symbionts with host-derived cell envelopes are not uncommon.
degeneration and codenpendence between host and symbiont (26, 28, 80, 81). What HGT, symbiont supplementation, and symbiont replacement may offer is a way out—at least temporarily, but perhaps permanently—of this degenerative ratchet.

However, new symbionts may also provide ecological opportunity in addition to evolutionary reinvigoriation. We note that the mealybug with one of the broadest host ranges is also the richest with the most recent γ-proteobacterial replacement. *P. longispinus*. *P. longispinus* is an important agricultural pest and known to feed on plants from 82 families (scalenet.info/catalogue/pseudococcus%20longispinus/). It seems possible that fresh symbionts with large genomes could provide novel functions unavailable in more degenerate symbionts, again propelling the symbioses into new niches.

### Materials and Methods

Samples of the mealybug species *M. hirsuta* (pink hibiscus mealybug; MHIR; collection locality: Helwan, Egypt), *F. virginig* (striped mealybug; FIV; collection locality: Helwan, Egypt), and *P. marginatus* (papaya mealybug; PMAR; collection locality: Mayotte, Comoro Islands) were identified and provided by Thibaut Malausa, Institut National de la Recherche Agronomique, Sophia, France. *T. perrissii* (TPER; collection locality: Poland) samples were provided by Malgorzata Kalandy-Kolodzieczyk, University of Silesia, Katowice, Poland. *P. longispinus* samples (long-tailed mealybug; PLOM) were collected by F.H. in a winter garden of the Faculty of Science, University of South Bohemia. DNA vouchers and insect vouchers of adult females for slide mounting are available from F.H. DNA was isolated from three to eight whole insects of all species by the Qiagen QiAamp DNA Micro Kit, and each library was prepared on two-thirds of an Illumina HiSeq 2000 lane and sequenced as 100-bp paired end reads. The *M. hirsuta* sample was sequenced on an entire MiSeq lane with v3 chemistry and 300-bp paired end mode. Both approaches generated sufficient coverage for both symbiont genomes and draft insect genomes. Adapter clipping and quality filtering were carried out in the Trimmomatic package (82) using default settings. Read error correction (BayesHammer), de novo assembly (k-mers 21, 33, 55, and 77 for 100-bp data and K99 and K127 for 300-bp data), and mismatch/short-indel correction were performed by the SPAdes assembler, v3.5.0 (83). Additional endosymbiont-targeted long k-mer (91 and 241 bp) assemblies generated by the Ray v2.3.1 (84) and PRICE v1.2.85 assemblies were used to improve assemblies of complex endosymbiont regions.

Additional information on the computational and microscopy methods can be found in SI Materials and Methods. General Tremblaya primers are shown in Table S4.

### ACKNOWLEDGMENTS

We thank the Genomics Core Facility at the University of Montana, the DNA Sequencing Facility at the University of Utah, and the European Molecular Biology Laboratory Genomics Core Facility in Heidelberg for sequencing services. F.H. was funded by the Fulbright Commission and Grant Agency of the University of South Bohemia Grant 04-001/2014PF. J.P.M. was funded by National (NSF-DBI-1556580 and IOS-1553529), National Aeronautics and Space Administration Astrobiology Institute Award NNA15BB04A, and NSF-Experimental Program to Stimulate Competitive Research Award NSF-IIA-1443108 (to the Montana Institute on Ecosystems).
