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

Filip Husnik[a,b,c,1] and John P. McCutcheon[a,d,1]

[a]Division of Biological Sciences, University of Montana, Missoula, MT 59812; [b]Institute of Parasitology, Biology Centre of the Czech Academy of Sciences, Ceske Budejovice 37005, Czech Republic; [c]Faculty of Science, University of South Bohemia, Ceske Budejovice 37005, Czech Republic; and [d]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 sequencing 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.

Sodalis | organelle | horizontal gene transfer | scale insect

Many organisms require intracellular bacteria for survival. The oldest and most famous example is the eukaryotic cell, which depends on mitochondria (and in photosynthetic eukaryotes, the chloroplasts or plastids) for the generation of biochemical energy (1–4). However, several more evolutionarily recent examples exist, where intracellular bacteria are involved in nutrient production from unbalanced host diets. For example, deep sea tube worms, some protists, and many sap-feeding insects are completely dependent on intracellular bacteria for essential nutrient provisioning (5–7). Some of these symbioses can form highly integrated organismal and genetic mosaics that, in many ways, resemble organelles (8–11). Like organelles, these endosymbionts have genomes encoding few genes (12, 13), rely on gene products of bacterial origin that are encoded on the host genome (9–11, 14, 15), and in some cases, import protein products encoded by these horizontally transferred genes back into the symbiont (16, 17). The names given to these bacteria—endosymbiont, protoorganelle, or bona fide organelle—are a matter of debate (18–21). What is not in doubt is that long-term interactions between hosts and essential bacteria generate highly integrated and complex symbioses.

Establishment of a nutritional endosymbiosis is beneficial for a host by allowing access to previously inaccessible food sources. However, strict dependence on intracellular bacteria can come with a cost: endosymbionts that stably associate with and provide essential functions to hosts often experience degenerative evolution (22–25). This degenerative process is thought to be driven by long-term reductions in effective population size (N_e) caused by the combined effects of asexuality [loss of most recombination and lack of new DNA through horizontal gene transfer (HGT)] and host restriction (e.g., frequent population bottlenecks at transmission in vertically transmitted bacteria) (26). The outcomes of these processes are clearly reflected in the genomes of long-term endosymbionts. These genomes are the smallest of any bacterium that is not an organelle, have among the fastest rates of evolution measured for any bacterium (12, 13), and are predicted to encode proteins and RNAs with decreased structural stability (26, 27). In symbioses where the endosymbiont is required for normal host function, such as in the bacterial endosymbionts of sap-feeding insects, this degenerative process can trap the host in a symbiotic "rabbit hole," where it depends completely on a symbiont which is slowly degenerating (28).

Unimpeded, the natural outcome of this degenerative process would seem to be extinction of the entire symbiosis. However, extinction, if it does happen, is difficult to observe, and surely is not the only solution to dependency on a degenerating symbiont. For example, organelles are bacterial endosymbionts that have managed to survive for billions of years (2). Despite the reduced N_e of organelle genomes relative to nuclear genomes, eukaryotes are able to purge deleterious mutations that arise on organelle genomes, perhaps through a combination of host-level selection and the strong negative selective effects of substitutions on gene-dense organelle genomes (29, 30). Extant organelle genomes also encode few genes relative to most bacteria, and it is also likely that a long history of moving genes to the nuclear genome has helped slow or stop organelle degeneration (21, 31). Some of the most degenerate insect endosymbionts also seem to have adopted a gene transfer strategy, although the number of transferred genes is far smaller compared with organelles. In aphids, mealybugs, psyllids, and whiteflies, some genes related to endosymbiont function are encoded on the nuclear genome, although in most cases, these genes have been transferred from other bacteria and not the

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.: *Maconellicoccus hirsutus*: PRJEB12066; *Ferrisia virgata*: PRJEB12067; *Pseudococcus longispinus*: PRJEB12068; *Paracoccus marginatus*: PRJEB12069; and *Trionymus perrisii*: 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).

[1]To 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: Cocoidea: 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 Phenacoccinae, 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 upregulated 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 *N*50 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*, *Antoniella*, *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*, *mnmG*, *lpd*, and *rsmH*) 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 perrisii* (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 3 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

Mealybug species	*P. avenae*	*M. hirsutus*	*F. virgata*	*P. citri*	*P. longispinus*	*T. perrisii*	*P. marginatus*
Mealybug abbreviation	PAVE	MHIR	FVIR	PCIT	PLON	TPER	PMAR
Total assembly size (bp)	NA	163,044,544	304,570,832	377,829,872	284,990,201	237,582,518	191,208,351
Total o. of scaffolds	NA	12,889	32,723	167,514	66,857	80,386	60,102
*N*50 | *N*75	NA	47,025 | 22,300	25,562 | 12,551	7,078 | 3,639	10,126 | 4,908	4,681 | 2,689	6,799 | 3,788
BUSCOs Arthropoda (*n*=2,675)	NA	76%	76%	71%	70%	66%	72%
BUSCOs Eukaryota (*n*=429)	NA	85%	84%	80%	78%	77%	82%
CEGMA (*n*=248; including partial)	NA	99.19%	97.98%	98.79%	98.39%	99.6%	98.79%
Tremblaya symbiont	*T. phenacola*	*T. princeps*	*T. princeps*	*T. princeps*	*T. princeps*	*T. princeps*	*T. princeps*
Genome size (plasmid size if present)	170,756 bp (744 bp)	138,415 bp	141,620 bp	138,927 bp	144,042 bp	143,340 bp	140,306 bp
Average fragment coverage	NA (454 data)	795	663	374	1,326	2,364	787
G + C (%)	42.2	61.8	58.3	58.8	58.9	57.8	58.3
CDS (pseudogenes)	178 (3)	136 (7)	132 (13)	125 (16)	134 (15)	116 (31)	124 (17)
CDS coding density (%)	86.3	77.2	69.3	66.0	70.7	59.2	67.0
rRNAs | tRNAs | ncRNAs	4 | 31 | 3	6 | 14 | 3	6 | 14 | 3	6 | 10 | 3	6 | 16 | 3	6 | 12 | 3	6 | 17 | 3
γ-Proteobacterial symbiont	Not present	*D. endobia*	*G. endobia*	*Mo. endobia*	PLON1 and PLON2	*H. endobia*	*Mi. endobia*
Genome size (plasmid size)	NA	834,723 bp (11,828 bp)	938,041 bp	538,294 bp	8,190,816*	628,221 bp (8,492 bp)	352,837 bp
Average fragment coverage	NA	121 (38)	372	827	30	559 (312; 1,750)	620
G + C (%)	NA	44.2	28.9	43.5	53.9	42.8	30.6
CDS (pseudogenes)	NA	564 (99)	461 (30)	419 (24)	NA (NA)	510 (16)	273 (8)
CDS coding density (%)	NA	59.8	48.1	77.4	NA	80.4	75.5
rRNAs | tRNAs | ncRNAs	NA	3 | 40 | 14	3 | 39 | 8	5 | 41 | 9	NA	3 | 41 | 10	3 | 41 | 5
Reference	9	This study	This study	42	This study	This study	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*.

pea aphid genome (52) scores 72% using identical settings. 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 *Tremblaya* origin detected in any of our mealybug assemblies.] Our data show that the acquisition of some HGTs [*bioABD*, *ribAD*, *dapF*, *lysA*, tryptophan 2-monooxygenase oxidoreductase (*tms*), and ATPases associated with diverse cellular activities (AAA-ATPases)] predated the Phenacoccinae/Pseudococcinae divergence and thus the acquisition of any γ-proteobacterial endosymbiont (Fig. 3). These old HGTs mostly involve amino acid and B vitamin metabolism, are usually found on longer insect scaffolds that contain several essential insect genes, and are syntenic across mealybug species (Fig. 4). In each of these cases, no other bacterial genes or pseudogenes were found within the scaffolds (Tables S2 and S3), suggesting that these HGTs resulted from the transfer of small DNA fragments or that flanking bacterial DNA from larger fragments was lost after the transfer was established. The origin of some of these transfers [7,8-diaminopelargonic acid synthase and biotin synthase (*bioAB*)] likely predates the entire mealybug lineage, because they are found in the genome of the whitefly *Bemisia tabaci* (11).

We find that several HGTs were likely acquired after the divergence of the *Maconellicoccus* clade [cysteine synthase A (*cysK*), beta-lactamase (*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*F*). We note that *cysK* has undergone tandem duplication in *P. longispinus*, *Ferrisia virgata* (FVIR), and *P. citri* (Fig. S2*A* 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 DUR1), 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 *Tremblaya*, *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 *Tremblaya*, its γ-proteobacterial endosymbiont, and/or the host. Gene retention patterns from many other pathways, however, show much less

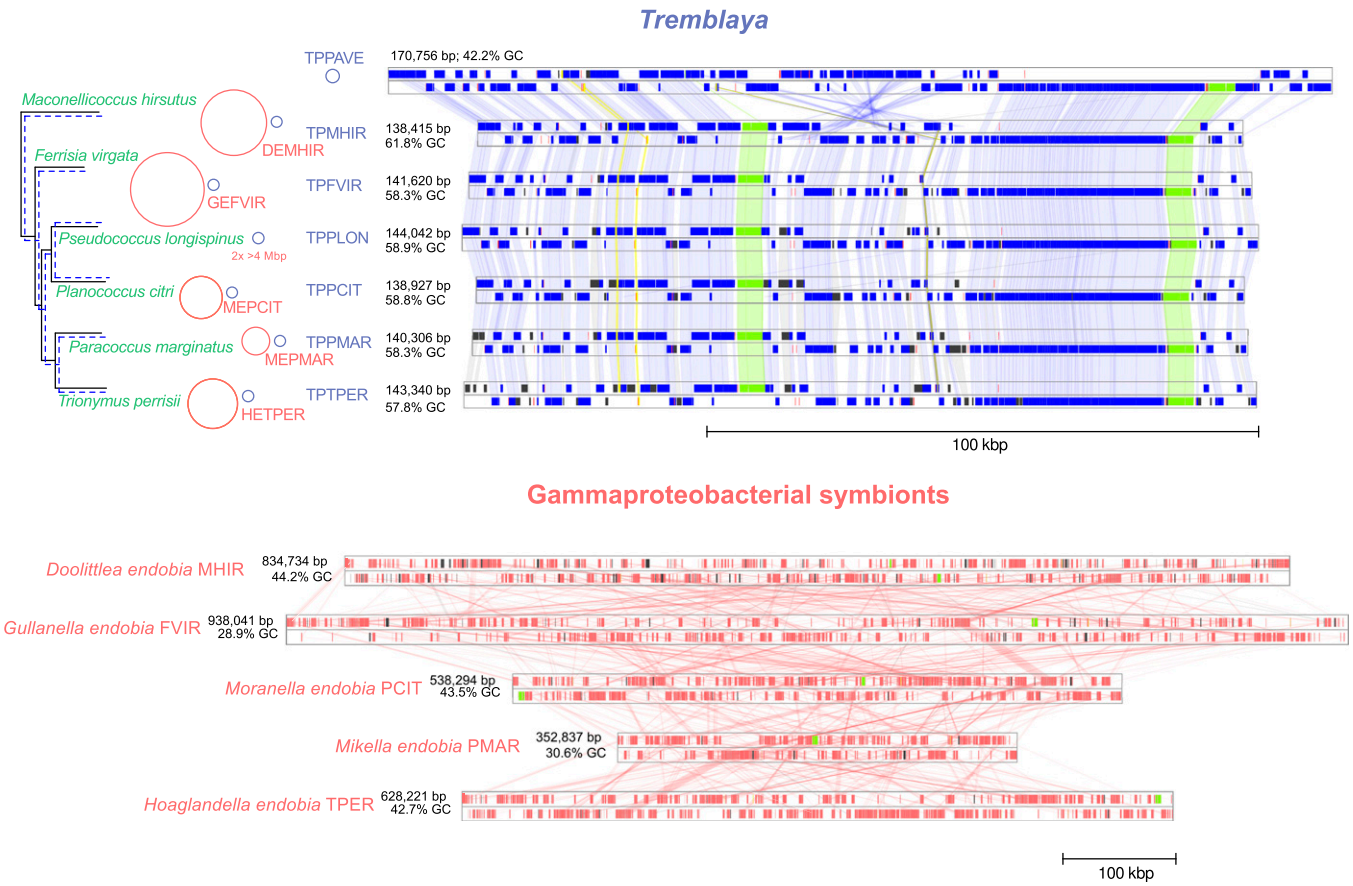

Fig. 1. Genome size and structure of the mealybug endosymbionts. Linear genome alignments of (*Upper*) seven *Tremblaya* genomes (blue) are contrasted with linear genome alignments of (*Lower*) five genomes of their respective γ-proteobacterial symbionts (red). The *T. princeps* genomes are perfectly collinear and similar in size, whereas the γ-proteobacterial genomes are highly rearranged and different in size. Alignments are ordered based on a schematic mealybug/*Tremblaya* phylogeny (original phylogenies are in Fig. S1) and accompanied by basic genome statistics (detailed genome statistics are in Table 1). Gene boxes are colored according to their category: proteins in blue, pseudogenes in gray, rRNAs in green, noncoding RNAs in yellow, and tRNAs in red.

predictable patterns. The isoleucine, valine, leucine, threonine, and histidine pathways show a tendency toward *Tremblaya*-dominated biosynthesis in *M. hirsutus*, *F. virgata*, and *P. citri* (that is, gene retention in *Tremblaya* and gene loss in the γ-proteobacterial symbiont) but with a clear shift toward γ-proteobacterial–dominated biosynthesis in *P. marginatus* and *T. perrisii*. Other pathways, such as tryptophan, show γ-proteobacterial dominance in all mealybug symbioses but with reliance on at least one *Tremblaya* gene in *P. citri*, *P. marginatus*, and *T. perrisii*. In the arginine pathway, gene retention is dominated by *Tremblaya* in *M. hirsutus* but by the γ-proteobacterial endosymbiont in all other lineages, with sporadic loss of *Tremblaya* genes in different lineages. Overall, *M. hirsutus* encodes the most *Tremblaya* genes and the fewest γ-proteobacterial genes, whereas TPER shows the opposite pattern.

Gene Retention Patterns for Translation-Related Genes in *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 *Tremblaya* PCIT (42), none of the additional *Tremblaya* genomes that we report here encode any functional aminoacyl tRNA synthetase, with an exception of one likely functional gene (*cysS*) in *T. princeps* PLON, which is present as a pseudogene in several other lineages of *Tremblaya*. Furthermore, all *Tremblaya* genomes have lost key translational control proteins that are typically retained even in the smallest endosymbiont genomes, such as ribosome recycling factor, L-methionyl-tRNA^{fMet} *N*-formyltransferase, and peptide deformylase. The translational release factors RF-1 and RF-2 (*prfAB*) and elongation factor (EF) EF-Ts (*tsf*) are present only in the gene-rich *T. princeps* MHIR genome and absent or pseudogenized in all other *T. princeps* genomes. Initiation factors (IFs) IF-1, IF-2, and IF-3 (*infABC*) and EFs EF-Tu and EF-G (*tufA* and *fusA*) are retained in all *Tremblaya* genomes, as are most ribosomal proteins (Dataset S2*A*).

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. princeps* PCIT for *T. princeps* from *P. citri*). However, our data show that the intra-*Tremblaya* γ-proteobacterial symbionts are not from the same infection; they result from independent endosymbiotic events from clearly discrete lineages within the *Sodalis* clade (Fig. 2). Following convention, we have chosen to give these γ-proteobacteria different genus names but unite them by retaining the "*endobia*" species denomination for each one (such as in *Moranella endobia*).

We propose the following Candidatus status names for four lineages of intra-*Tremblaya* γ-proteobacterial symbionts of mealybugs for which we have completed a genome. First, *Candidatus* Doolittlea endobia MHIR is for the endosymbiont from *M. hirsutus*. This name honors the American evolutionary biologist W. Ford Doolittle (1941–) for his contributions to our understanding of HGT and endosymbiosis. Second, *Candidatus* Gullanella endobia FVIR is for the endosymbiont from *F. virgata*. This name honors the Australian entomologist Penny J. Gullan

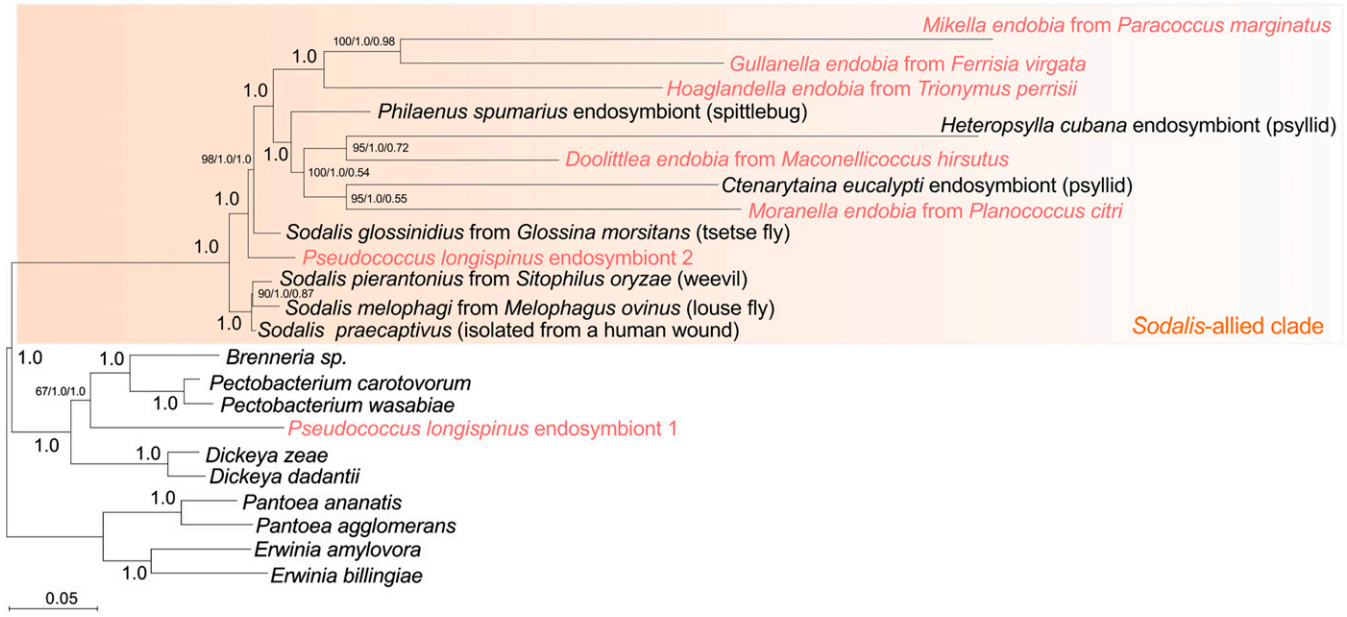

Fig. 2. The intra-*Tremblaya* mealybug symbionts are members of the *Sodalis* clade of γ-proteobacteria. A multigene phylogeny of *Sodalis*-allied insect endosymbionts and closely related Enterobacteriaceae (γ-proteobacteria) was inferred from 80 concatenated proteins under the LG + G evolutionary model in RaxML v8.2.4. Mealybug endosymbionts are highlighted in red. Values at nodes represent bootstrap pseudoreplicates from the maximum likelihood (ML) analysis, posterior probabilities from Bayesian inference (BI) topology inferred under the LG + I + G model, and posterior probabilities from BI topology inferred from the Dayhoff6 recoded dataset under the CAT + GTR + G model in PhyloBayes, respectively.

(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 *Ferrisia* 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. 5*A*). 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. 5*B*). 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. 5*C*). The idosyncratic 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

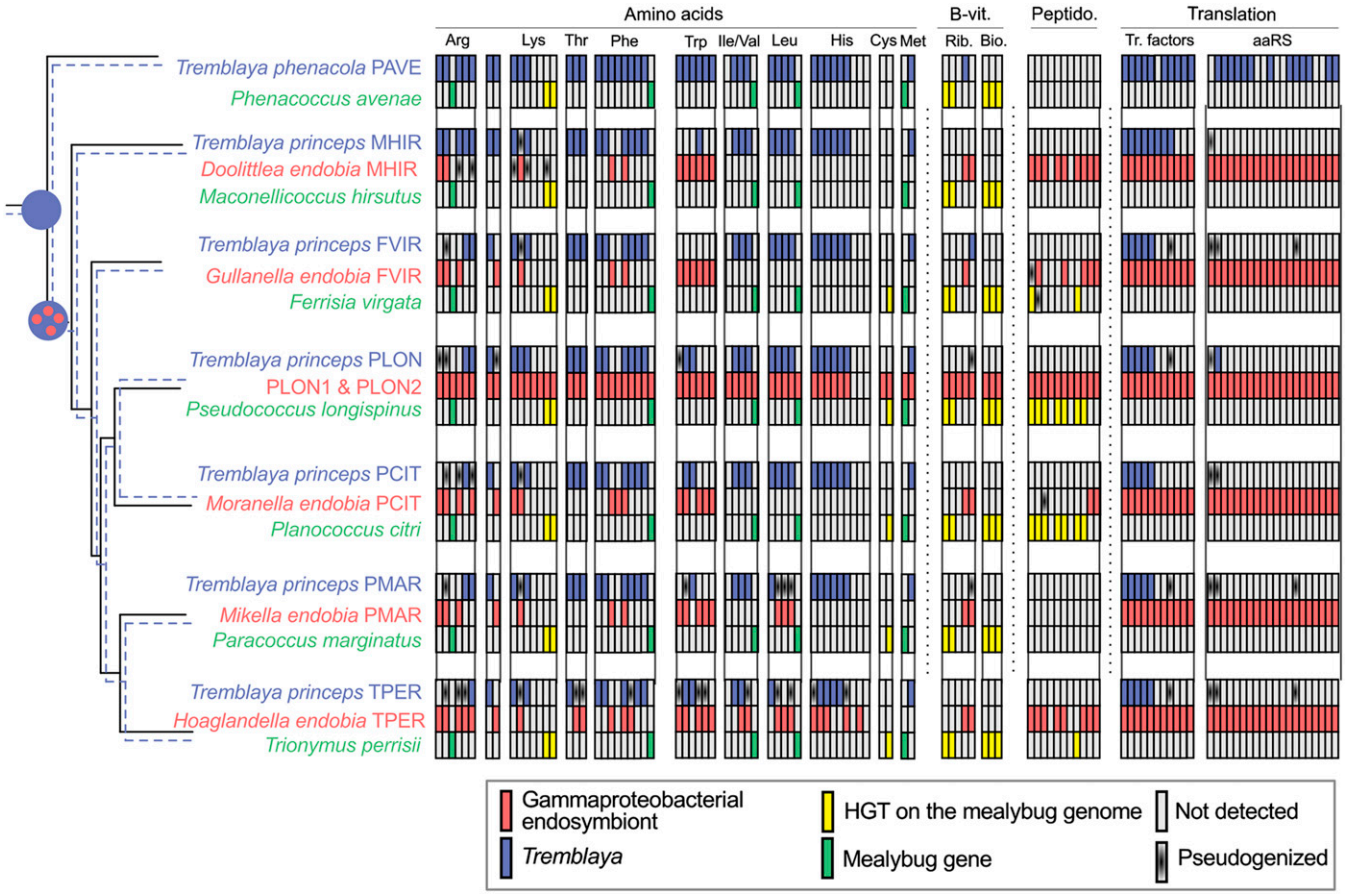

Fig. 3. A complex history of gene retention, loss, and acquisition in the mealybug symbiosis. Retention of selected biosynthetic pathways, such as amino acids, B vitamins (B-vit.), peptidoglycan (Peptido.), translation-related genes [various initiation, elongation, and termination factors (Tr. factors)], and HGTs. For each of seven mealybug species, boxes in row 1 represent *Tremblaya* genes (blue), row 2 represents its γ-proteobacterial symbionts (red), and row 3 represents the host genome (insect genes in green and HGTs in yellow). Missing genes are shown in gray, and recognizable pseudogenes are shown with black radial gradient. Raw data used here (including gene names) are available in Dataset S2*B*. Bio., biotin; Rib., riboflavin.

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-proteobactia 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

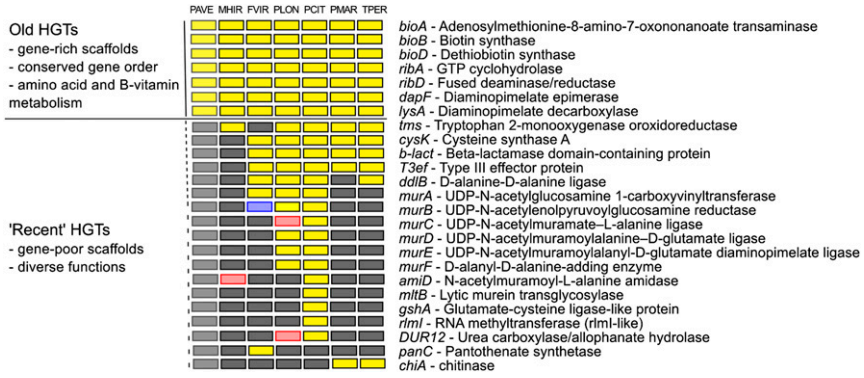

Fig. 4. HGTs detected in individual mealybug species. Retention of HGT candidates detected across all mealybug species (blue, possible pseudogene; gray, gene not detected; red, different phylogenetic origin; yellow, gene present).

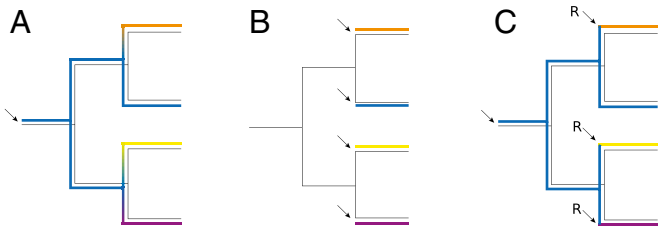

Fig. 5. Three possible scenarios that built the mealybug symbiosis. Independent γ-proteobacterial acquisitions are shown as arrows, and replacements are noted with Rs above the arrow. Colors represent the different γ-proteobacterial genomes shown in Fig. 1. (*A*) The idiosyncratic scenario, where a single 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.

(Fig. 3) in *Tremblaya* occurred in response to the first γ-proteobacterial infection, which then required all subsequent replacement events to also reside within the *Trembalay* cytoplasm. It is tempting to speculate that the 353-kb *Mikella* PMAR genome is the ancestral intra-*Tremblaya* symbiont lineage that has not been replaced or at least has not been recently replaced. However, because the relevant clades split right after the Phenacoccinae/Pseudococcinae divergence—that is, right at the acquisition of the first γ-proteobacterial symbiont—much richer taxon sampling would be needed 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).

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 *Phenacoccus avenae*, which lacks a γ-proteobacterial symbiont) encodes no genes for the production of fatty acids or peptidoglycan (9). We assume that the envelope that defines the *Tremblaya* cytoplasm is made by the host, because it cannot be made by *Tremblaya*. These data suggest that when the first γ-proteobacterial endosymbiont established residence in *Tremblaya*, it invaded a membrane system that was perhaps more eukaryotic than bacterial in nature (even if it ultimately ended up in a "bacterial" cytoplasm). Bacteria in the *Sodalis* group are very good at establishing intracellular infections in insect cells (38, 63, 64), and we suggest that their propensity to infect *Tremblaya* might simply reflect this ability. The cytoplasm vs. envelope distinction is important, because the mealybug symbiosis has been held up by many—including us—as a rare example of a stable bacteria within a bacterium symbiosis. Although this description might be apt if one considers the *Tremblaya* cytoplasm bacterial in nature, it may not be if one considers the types of membranes that the innermost bacteria had to cross to get there.

But why did the first γ-proteobacterial endosymbiont end up inside *Tremblaya*? We can think of two related possibilities. The first is that it was easier to use the established transport system between the insect cell and *Tremblaya* (65) than to evolve a new one. The second is that the insect immune system likely does not target *Tremblaya* cells, and so the *Tremblaya* cytoplasm is an ideal hiding place for a newly arrived symbiont. After the loss of critical translation-related genes in *Tremblaya*, the symbiosis would persist with a bacteria within a bacterium structure because no other structure is possible. We note that *Sodalis*- and *Arsenophonus*-allied symbionts were recently suggested to sometimes reside within *Sulcia* cells in the leafhoppers *Cicadella viridis* and *Macrosteles 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.

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 Phenacoccinae 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-proteobactieral 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

degeneration and codependence 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 reinvigoration. We note that the mealybug with one of the broadest host ranges is also the species 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. hirsutus* (pink hibiscus mealybug; MHIR; collection locality: Helwan, Egypt), *F. virgata* (striped mealybug; FVIR; 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. perrisii* (TPER; collection locality: Poland) samples were provided by Małgorzata Kalandyk-Kołodziejczyk, University of Silesia, Katowice, Poland. *P. longispinus* samples (long-tailed mealybug; PLON) 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 multiplexed on two-thirds of an Illumina HiSeq 2000 Lane and sequenced as 100-bp paired end reads. The *M. hirsutus* 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 K21, K33, K55, and K77 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) assemblers 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 04-001/2014/P. J.P.M. was funded by National Science Foundation (NSF) Grants IOS-1256680 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).

1. Gray MW, Doolittle WF (1982) Has the endosymbiont hypothesis been proven? *Microbiol Rev* 46(1):1–42.
2. Palmer JD (1997) Organelle genomes: Going, going, gone! *Science* 275(5301):790–791.
3. Martin W, Müller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392(6671):37–41.
4. Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440(7084):623–630.
5. Douglas AE (1989) Mycetocyte symbiosis in insects. *Biol Rev Camb Philos Soc* 64(4): 409–434.
6. Nowack ECM, Melkonian M (2010) Endosymbiotic associations within protists. *Philos Trans R Soc Lond B Biol Sci* 365(1541):699–712.
7. Stewart FJ, Newton ILG, Cavanaugh CM (2005) Chemosynthetic endosymbioses: Adaptations to oxic-anoxic interfaces. *Trends Microbiol* 13(9):439–448.
8. Nakayama T, Ishida K (2009) Another acquisition of a primary photosynthetic organelle is underway in *Paulinella chromatophora*. *Curr Biol* 19(7):R284–R285.
9. Husnik F, et al. (2013) Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153(7):1567–1578.
10. Sloan DB, et al. (2014) Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol Biol Evol* 31(4):857–871.
11. Luan J-B, et al. (2015) Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol* 7(9):2635–2647.
12. McCutcheon JP, Moran NA (2011) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10(1):13–26.
13. Moran NA, Bennett GM (2014) The tiniest tiny genomes. *Annu Rev Microbiol* 68: 195–215.
14. Nikoh N, et al. (2010) Bacterial genes in the aphid genome: Absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet* 6(2):e1000827.
15. Nowack ECM, et al. (2011) Endosymbiotic gene transfer and transcriptional regulation of transferred genes in *Paulinella chromatophora*. *Mol Biol Evol* 28(1):407–422.
16. Nowack ECM, Grossman AR (2012) Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc Natl Acad Sci USA* 109(14):5340–5345.
17. Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima SY (2014) Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Curr Biol* 24(14):R640–R641.
18. Theissen U, Martin W (2006) The difference between organelles and endosymbionts. *Curr Biol* 16(24):R1016–R1017.
19. Keeling PJ, Archibald JM (2008) Organelle evolution: What's in a name? *Curr Biol* 18(8):R345–R347.
20. McCutcheon JP, Keeling PJ (2014) Endosymbiosis: Protein targeting further erodes the organelle/symbiont distinction. *Curr Biol* 24(14):R654–R655.
21. Keeling PJ, McCutcheon JP, Doolittle WF (2015) Symbiosis becoming permanent: Survival of the luckiest. *Proc Natl Acad Sci USA* 112(33):10101–10103.
22. Sloan DB, Moran NA (2012) Genome reduction and co-evolution between the primary and secondary bacterial symbionts of psyllids. *Mol Biol Evol* 29(12):3781–3792.
23. Bennett GM, Moran NA (2013) Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a Phloem-feeding insect. *Genome Biol Evol* 5(9):1675–1688.
24. Nakabachi A, et al. (2013) Defensive bacteriome symbiont with a drastically reduced genome. *Curr Biol* 23(15):1478–1484.
25. Manzano-Marín A, Latorre A (2014) Settling down: The genome of *Serratia symbiotica* from the aphid *Cinara tujafilina* zooms in on the process of accommodation to a cooperative intracellular life. *Genome Biol Evol* 6(7):1683–1698.
26. Moran NA (1996) Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc Natl Acad Sci USA* 93(7):2873–2878.
27. Fares MA, Barrio E, Sabater-Muñoz B, Moya A (2002) The evolution of the heat-shock protein GroEL from *Buchnera*, the primary endosymbiont of aphids, is governed by positive selection. *Mol Biol Evol* 19(7):1162–1170.
28. Bennett GM, Moran NA (2015) Heritable symbiosis: The advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci USA* 112(33):10169–10176.
29. Popadin KY, Nikolaev SI, Junier T, Baranova M, Antonarakis SE (2013) Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. *Mol Biol Evol* 30(2):347–355.
30. Cooper BS, Burrus CR, Ji C, Hahn MW, Montooth KL (2015) Similar efficacies of selection shape mitochondrial and nuclear genes in both *Drosophila melanogaster* and *Homo sapiens*. *G3 (Bethesda)* 5(10):2165–2176.
31. Smith DR, Keeling PJ (2015) Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. *Proc Natl Acad Sci USA* 112(33):10177–10184.
32. McCutcheon JP, Moran NA (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc Natl Acad Sci USA* 104(49):19392–19397.
33. Koga R, Bennett GM, Cryan JR, Moran NA (2013) Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. *Environ Microbiol* 15(7): 2073–2081.
34. Koga R, Moran NA (2014) Swapping symbionts in spittlebugs: Evolutionary replacement of a reduced genome symbiont. *ISME J* 8(6):1237–1246.
35. Thao ML, et al. (2000) Secondary endosymbionts of psyllids have been acquired multiple times. *Curr Microbiol* 41(4):300–304.
36. Lamelas A, et al. (2011) *Serratia symbiotica* from the aphid *Cinara cedri*: A missing link from facultative to obligate insect endosymbiont. *PLoS Genet* 7(11):e1002357.
37. Vogel KJ, Moran NA (2013) Functional and evolutionary analysis of the genome of an obligate fungal symbiont. *Genome Biol Evol* 5(5):891–904.
38. Smith WA, et al. (2013) Phylogenetic analysis of symbionts in feather-feeding lice of the genus *Columbicola*: Evidence for repeated symbiont replacements. *BMC Evol Biol* 13(1):109.
39. Lefèvre C, et al. (2004) Endosymbiont phylogenesis in the dryophthoridae weevils: Evidence for bacterial replacement. *Mol Biol Evol* 21(6):965–973.
40. Toju H, Tanabe AS, Notsu Y, Sota T, Fukatsu T (2013) Diversification of endosymbiosis: Replacements, co-speciation and promiscuity of bacteriocyte symbionts in weevils. *ISME J* 7(7):1378–1390.
41. Gruwell ME, Hardy NB, Gullan PJ, Dittmar K (2010) Evolutionary relationships among primary endosymbionts of the mealybug subfamily phenacoccinae (hemiptera: Coccoidea: Pseudococcidae). *Appl Environ Microbiol* 76(22):7521–7525.
42. McCutcheon JP, von Dohlen CD (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* 21(16):1366–1372.
43. Thao ML, Gullan PJ, Baumann P (2002) Secondary (gamma-Proteobacteria) endosymbionts infect the primary (beta-Proteobacteria) endosymbionts of mealybugs multiple times and coevolve with their hosts. *Appl Environ Microbiol* 68(7):3190–3197.
44. Kono M, Koga R, Shimada M, Fukatsu T (2008) Infection dynamics of coexisting beta- and gammaproteobacteria in the nested endosymbiotic system of mealybugs. *Appl Environ Microbiol* 74(13):4175–4184.

45. von Dohlen CD, Kohler S, Alsop ST, McManus WR (2001) Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature* 412(6845):433–436.
46. López-Madrigal S, et al. (2014) Molecular evidence for ongoing complementarity and horizontal gene transfer in endosymbiotic systems of mealybugs. *Front Microbiol* 5:449.
47. Parkinson JF, Gobin B, Hughes WOH (2016) Heritability of symbiont density reveals distinct regulatory mechanisms in a tripartite symbiosis. *Ecol Evol* 6(7):2053–2060.
48. Gatehouse LN, Sutherland P, Forgie SA, Kaji R, Christeller JT (2012) Molecular and histological characterization of primary (betaproteobacteria) and secondary (gammaproteobacteria) endosymbionts of three mealybug species. *Appl Environ Microbiol* 78(4):1187–1197.
49. Koga R, Nikoh N, Matsuura Y, Meng XY, Fukatsu T (2013) Mealybugs with distinct endosymbiotic systems living on the same host plant. *FEMS Microbiol Ecol* 83(1): 93–100.
50. Buchner P (1965) *Endosymbiosis of Animals with Plant Microorganisms* (Interscience Publishers, New York), p 909.
51. Denton JF, et al. (2014) Extensive error in the number of genes inferred from draft genome assemblies. *PLOS Comput Biol* 10(12):e1003998.
52. International Aphid Genomics Consortium (2010) Genome sequence of the pea aphid *Acyrthosiphon pisum*. *PLoS Biol* 8(2):e1000313.
53. Moran NA, Mira A (2001) The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol* 2(12):H0054.
54. Frank AC, Amiri H, Andersson SG (2002) Genome deterioration: Loss of repeated sequences and accumulation of junk DNA. *Genetica* 115(1):1–12.
55. Moran NA (2002) Microbial minimalism: Genome reduction in bacterial pathogens. *Cell* 108(5):583–586.
56. Moran NA, McCutcheon JP, Nakabachi A (2008) Genomics and evolution of heritable bacterial symbionts. *Annu Rev Genet* 42:165–190.
57. Moya A, Peretó J, Gil R, Latorre A (2008) Learning how to live together: Genomic insights into prokaryote-animal symbioses. *Nat Rev Genet* 9(3):218–229.
58. Clayton AL, et al. (2012) A novel human-infection-derived bacterium provides insights into the evolutionary origins of mutualistic insect-bacterial symbioses. *PLoS Genet* 8(11):e1002990.
59. Oakeson KF, et al. (2014) Genome degeneration and adaptation in a nascent stage of symbiosis. *Genome Biol Evol* 6(1):76–93.
60. McCutcheon JP, Moran NA (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol Evol* 2:708–718.
61. Rosenblueth M, Sayavedra L, Sámano-Sánchez H, Roth A, Martínez-Romero E (2012) Evolutionary relationships of flavobacterial and enterobacterial endosymbionts with their scale insect hosts (Hemiptera: Coccoidea). *J Evol Biol* 25(11):2357–2368.
62. Toh H, et al. (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Res* 16(2):149–156.
63. Hosokawa T, Kaiwa N, Matsuura Y, Kikuchi Y, Fukatsu T (2015) Infection prevalence of *Sodalis symbionts* among stinkbugs. *Zoological Lett* 1(1):5.
64. Dale C, Young SA, Haydon DT, Welburn SC (2001) The insect endosymbiont S*odalis glossinidius* utilizes a type III secretion system for cell invasion. *Proc Natl Acad Sci USA* 98(4):1883–1888.
65. Duncan RP, et al. (2014) Dynamic recruitment of amino acid transporters to the insect/symbiont interface. *Mol Ecol* 23(6):1608–1623.
66. Michalik A, Jankowska W, Kot M, Gołas A, Szklarzewicz T (2014) Symbiosis in the green leafhopper, *Cicadella viridis* (Hemiptera, Cicadellidae). Association *in statu nascendi? Arthropod Struct Dev* 43(6):579–587.
67. Kobiałka M, Michalik A, Walczak M, Junkiert Ł, Szklarzewicz T (2016) *Sulcia* symbiont of the leafhopper *Macrosteles laevis* (Ribaut, 1927) (Insecta, Hemiptera, Cicadellidae: Deltocephalinae) harbors *Arsenophonus* bacteria. *Protoplasma* 253(3):903–912.
68. Wang Z, Wu M (2014) Phylogenomic reconstruction indicates mitochondrial ancestor was an energy parasite. *PLoS One* 9(10):e110685.
69. Ochoa de Alda JAG, Esteban R, Diago ML, Houmard J (2014) The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat Commun* 5:4937.
70. Booth A, Doolittle WF (2015) Eukaryogenesis, how special really? *Proc Natl Acad Sci USA* 112(33):10278–10285.
71. Kurland CG, Andersson SG (2000) Origin and evolution of the mitochondrial proteome. *Microbiol Mol Biol Rev* 64(4):786–820.
72. Gray MW (2015) Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. *Proc Natl Acad Sci USA* 112(33):10133–10138.
73. Ku C, et al. (2015) Endosymbiotic gene transfer from prokaryotic pangenomes: Inherited chimerism in eukaryotes. *Proc Natl Acad Sci USA* 112(33):10139–10146.
74. Zimorski V, Ku C, Martin WF, Gould SB (2014) Endosymbiotic theory for organelle origins. *Curr Opin Microbiol* 22:38–48.
75. Ettema TJG (2016) Evolution: Mitochondria in the second act. *Nature* 531(7592): 39–40.
76. Pittis AA, Gabaldón T (2016) Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* 531(7592):101–104.
77. Ku C, et al. (2015) Endosymbiotic origin and differential loss of eukaryotic genes. *Nature* 524(7566):427–432.
78. Koonin EV (2015) Archaeal ancestors of eukaryotes: Not so elusive any more. *BMC Biol* 13(1):84.
79. Larkum AWD, Lockhart PJ, Howe CJ (2007) Shopping for plastids. *Trends Plant Sci* 12(5):189–195.
80. Fares MA, Ruiz-González MX, Moya A, Elena SF, Barrio E (2002) Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature* 417(6887):398.
81. Andersson JO, Andersson SG (1999) Insights into the evolutionary process of genome degradation. *Curr Opin Genet Dev* 9(6):664–671.
82. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.
83. Bankevich A, et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19(5):455–477.
84. Boisvert S, Laviolette F, Corbeil J (2010) Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol* 17(11):1519–1533.
85. Ruby JG, Bellare P, Derisi JL (2013) PRICE: Software for the targeted assembly of components of (Meta) genomic sequence data. *G3 (Bethesda)* 3(5):865–880.
86. Walker BJ, et al. (2014) Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9(11):e112963.
87. Hunt M, et al. (2013) REAPR: A universal tool for genome assembly evaluation. *Genome Biol* 14(5):R47.
88. Seemann T (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30(14):2068–2069.
89. Konwar KM, Hanson NW, Pagé AP, Hallam SJ (2013) MetaPathways: A modular pipeline for constructing pathway/genome databases from environmental sequence information. *BMC Bioinformatics* 14(1):202.
90. Karp PD, et al. (2010) Pathway Tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* 11(1):40–79.
91. Jones P, et al. (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30(9):1236–1240.
92. Rutherford K, et al. (2000) Artemis: Sequence visualization and annotation. *Bioinformatics* 16(10):944–945.
93. Segata N, Börnigen D, Morgan XC, Huttenhower C (2013) PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* 4:2304.
94. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9(4):286–298.
95. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.
96. Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
97. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19(12):1572–1574.
98. Lartillot N, Rodrigue N, Stubbs D, Richer J (2013) PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol* 62(4):611–615.
99. Darling AE, Mau B, Perna NT (2010) progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5(6):e11147.
100. Li L, Stoeckert CJ, Jr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178–2189.
101. Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M (2013) Blobology: Exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet* 4:237.
102. Koutsovoulos G, et al. (2016) No evidence for extensive horizontal gene transfer in the genome of the tardigrade *Hypsibius dujardini*. *Proc Natl Acad Sci USA* 113(18): 5053–5058.
103. Delmont TO, Eren AM (2016) Identifying contamination with advanced visualization and analysis practices: Metagenomic approaches for eukaryotic genome assemblies. *PeerJ* 4:e1839.
104. Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29(8):1072–1075.
105. Parra G, Bradnam K, Korf I (2007) CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23(9):1061–1067.
106. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31(19):3210–3212.
107. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res* 33(20):6494–6506.
108. Wheeler TJ, Eddy SR (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29(19):2487–2489.
109. Huerta-Cepas J, Dopazo J, Gabaldón T (2010) ETE: A python environment for tree exploration. *BMC Bioinformatics* 11(1):24.
110. Van Leuven JT, Meister RC, Simon C, McCutcheon JP (2014) Sympatric speciation in a bacterial endosymbiont results in two genomes with the functionality of one. *Cell* 158(6):1270–1280.
111. Schindelin J, et al. (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods* 9(7):676–682.