Current Biology *21*, 1366–1372, August 23, 2011 DOI 10.1016/j.cub.2011.06.051

Report

An Interdependent Metabolic Patchwork in the Nested Symbiosis of Mealybugs

John P. McCutcheon[1,2,*] and Carol D. von Dohlen[3]
[1]Center for Insect Science and Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA
[2]Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA
[3]Department of Biology, Utah State University, Logan, UT 84322, USA

Summary

Highly reduced genomes of 144–416 kilobases have been described from nutrient-provisioning bacterial symbionts of several insect lineages [1–5]. Some host insects have formed stable associations with pairs of bacterial symbionts that live in specialized cells and provide them with essential nutrients; genomic data from these systems have revealed remarkable levels of metabolic complementarity between the symbiont pairs [3, 4, 6, 7]. The mealybug *Planococcus citri* (Hemiptera: Pseudococcidae) contains dual bacterial symbionts existing with an unprecedented organization: an unnamed gammaproteobacteria, for which we propose the name *Candidatus* Moranella endobia, lives inside the betaproteobacteria *Candidatus* Tremblaya princeps [8]. Here we describe the complete genomes and metabolic contributions of these unusual nested symbionts. We show that whereas there is little overlap in retained genes involved in nutrient production between symbionts, several essential amino acid pathways in the mealybug assemblage require a patchwork of interspersed gene products from *Tremblaya*, *Moranella*, and possibly *P. citri*. Furthermore, although *Tremblaya* has the smallest cellular genome yet described, it contains a genomic inversion present in both orientations in individual insects, starkly contrasting with the extreme structural stability typical of highly reduced bacterial genomes [4, 9, 10].

Results

The *Tremblaya* Genome Is Extremely Small and Degenerate

With only 138,927 base pairs (bp) and 121 protein-coding genes, *Tremblaya* has the smallest cellular genome yet described (see Figure S1 available online). There are two copies of the 16S-23S-5S ribosomal DNA operon contained within an identical 5702 bp duplication (genome coordinates 46,430–52,131 and 124,024–129,725), as expected from previous work sequencing rDNA-positive clones from a mealybug genomic library [11]. The *Tremblaya* genome encodes 44 ribosomal proteins, a number similar to other highly reduced bacterial symbiont genomes [12]. In contrast to these other tiny genomes, however, *Tremblaya* lacks many translation-related gene homologs outside of the ribosome itself. Its genome encodes no functional aminoacyl-tRNA synthetases and is missing homologs for both translational release factors (*prfA* and *prfB*), elongation factor EF-Ts (*tsf*), ribosome recycling factor (*frr*), and peptide deformylase (*def*). Similar to the alphaproteobacterial insect symbiont *Candidatus* Hodgkinia cicadicola [1], which at 144 kilobases (kb) is the next largest cellular genome, *Tremblaya* has the unusual property of having a small genome with a high guanine + cytosine (GC) content (58.8%) [11], although unlike *Hodgkinia*, it uses the standard bacterial genetic code. Compared to most highly reduced bacterial genomes, the *Tremblaya* genome is unusually gene sparse, with a coding density of only 72.9%, an average intergenic length of 389 bp, and 19 recognizable pseudogenes (Table 1; Table 2).

The *Tremblaya* Genome Exists in Two Forms within Single Insects

One of the most striking features of the *Tremblaya* genome is a 7,032 bp region flanked by two 71 bp inverted repeats, which is found in both orientations in the population (Figure 1). During genome assembly, reads overlapping either end of this 7 kb region were found contiguous with both ends of an assembly gap, suggesting genomic polymorphism in the 12 insects that were pooled for genome sequencing. PCR amplification on DNA extractions from 10 additional insects confirmed that both orientations of the inversion exist in the *Tremblaya* population in individual insects (data not shown). It is currently unknown whether both orientations are present within individual *Tremblaya* cells. *Tremblaya* encodes only six genes involved in DNA replication, recombination, or repair—the α, β, ε, and γ/τ subunits of DNA polymerase III, the replicative DNA helicase, and the DNA replication inhibitor CspD (*dnaENQXB* and *cspD*)—and thus at this evolutionary stage does not appear to be able to catalyze this inversion itself, because no genes known to be directly involved in recombination are present in the genome (e.g., *recA*, *recBCD*, *recE*, *recF*, or *recJ* [13]).

Tremblaya Retains Several Genes Involved in Essential Amino Acid Biosynthesis but Does Not Have Any Complete Pathways of Its Own

Most insects that feed exclusively on plant sap have developed symbioses with microorganisms to supplement their nutritionally deficient diet [14, 15]. Genomic work on the symbionts of sap-feeding insects has clearly shown that essential amino acid production is the primary nutritional role of these intracellular bacteria [2–7, 9, 16, 17]. Because mealybugs feed exclusively on phloem sap, it has been assumed that at least one of their two intracellular bacterial symbionts is involved in provisioning essential amino acids [18]. This assumption leads to the prediction that either *Tremblaya* or *Moranella* (or both) would retain genes necessary for the production of at least some essential amino acids. Analysis of the *Tremblaya* genome supports this hypothesis, because it encodes several gene homologs (29 of 121 genes, or 22% of the genome by nucleotide count) involved in the synthesis of the ten essential amino acids, although no one single pathway is complete in *Tremblaya* alone (Figure 2). The leucine and valine pathways are nearly complete, although gene homologs for the branched-chain amino acid aminotransferase (BCA) and the

*Correspondence: john.mccutcheon@umontana.edu

Table 1. Gene Number and Density for Several Intracellular and Free-Living Bacterial Genomes

Organism	Genome Size	Number of Genes	Coding Density (%)	Average Intergenic Length
Tremblaya	138,927	140	72.9	389
Hodgkinia	143,795	187	94.3	79
Carsonella	159,662	213	97.3	68
Zinderia	208,564	231	92.6	89
Sulcia	276,984	275	92.7	100
Buchnera	416,380	393	86.9	145
Moranella	538,294	452	79.0	268
Riesia	574,390	582	82.5	192
Mycoplasma	580,076	519	92.0	141
Rickettsia	1,111,523	870	76.0	354
Mycobacterium	3,268,203	1,655	49.8	1,188
Sodalis	4,171,146	2,523	51.7	913
Escherichia	4,639,675	4,440	87.7	155

Coding density was calculated using all protein-coding, ribosomal RNA, and tRNA genes in the genome. The average intergenic length includes all nonoverlapping intergenic regions. The GenBank accession numbers for the genomes used in these calculations were *Hodgkinia* (CP001226.1), *Carsonella* (AP009180.1), *Zinderia* (CP002161.1), *Sulcia* (CP001605.1), *Buchnera* (CP000263.1), *Riesia* (CP001085), *Mycoplasma* (L43967.2), *Rickettsia* (AJ235269.1), *Mycobacterium* (NC_002677.1), *Sodalis* (AP008232.1), and *Escherichia* (U00096.2).

small subunit (IlvN) of acetohydroxybutanoate synthase/acetolactate synthase are both missing. In *Escherichia coli*, the active sites for both acetohydroxybutanoate synthase and acetolactate synthase are on the large subunit (IlvB), and this subunit is functional in the absence of IlvN, although the V_{max} for both reactions is greatly reduced [19]. It is therefore possible that these reactions are catalyzed by IlvB alone in *Tremblaya*. The aphid symbiont *Buchnera* is also missing the BCA homolog, and it has been predicted that this reaction is carried out by the insect in the *Buchnera*-pea aphid symbiosis [20]. The recent completion of the pea aphid genome confirmed that a BCA gene homolog is encoded in the aphid genome [20]; it is also highly expressed in aphid bacteriocytes at both the mRNA [21] and protein [22] levels. Because the BCA activity is encoded in all other sequenced insect genomes [20], it seems likely that this reaction is ancestrally present in insects and retained by *Planococcus citri*. Evidence for this hypothesis comes from a BCA homolog found in an expressed sequence tag (EST) library of *Maconellicoccus hirsutus* (GenBank accession number EH218655.1), a mealybug belonging to the same subfamily as *P. citri* [23]. Additionally, a few low-coverage contigs in the present work assumed to be from the mealybug genome have top hits to animal BCA homologs in the GenBank nonredundant (nr) database, indicating that a gene coding for the BCA activity is present in the *P. citri* genome (Table S1). As in aphids [21], the mechanism used by the mealybug system in the production of methionine is not clear (Figure 2). Certain sequences in the *M. hirsutus* EST library (EH216258.1, EH218567.1, and EH211908.1) and in the *P. citri* assembly (Table S1) correspond to gene homologs of cystathionine gamma-synthase (CGL), cystathionine beta-lyase (CBL), and homocysteine S-methyltransferase (MetE), although in this case the precise annotation of these genes is uncertain. Because *Tremblaya* also encodes a homolog of *metE*, the production of methionine may involve a combination of gene products from *Tremblaya* and *P. citri*, although further work is needed to identify the exact genes involved in this pathway.

Moranella Lives Inside *Tremblaya* and Complements Several Essential Amino Acid Biosynthesis Genes Missing in *Tremblaya*

At 538,294 bp in length, the *Moranella* genome is almost four times larger than the *Tremblaya* genome (Figure S1), although

Table 2. Recognizable Pseudogenes in the *Tremblaya* Genome

Pseudogene	Description	Pathway	Present in *Moranella*
mnmG	tRNA modification protein	translation	yes
mnmE	tRNA modification protein	translation	yes
argS	arginyl-tRNA synthetase	translation	yes
mraW	16S rRNA methyltransferase	translation	yes
cysS	cysteinyl-tRNA synthetase	translation	yes
mnmA	tRNA methyltransferase	translation	yes
rplV	50S ribosomal protein L22	translation	yes
carB	carbamoyl phosphate synthetase, large subunit	amino acid synthesis	yes
argF	ornithine carbamoyltransferase	amino acid synthesis	yes
argH	argininosuccinate lyase	amino acid synthesis	yes
grpE	GrpE protein	protein folding and stability	yes
ssb	ssDNA binding protein	replication and repair	yes
rpoD	RNA polymerase, sigma 70 subunit	transcription	yes
ndk	nucleoside diphosphate kinase	nucleotide metabolism	no
mviN	lipid II flippase	cell envelope synthesis	yes
lpd	lipoamide dehydrogenase	general metabolism	no
aceF	E2 component of pyruvate dehydrogenase	general metabolism	no
TPPCIT_150	hypothetical protein	unknown	no
TPPCIT_152	hypothetical protein	unknown	no

Note: two copies of the *mraW* pseudogene exist as part of the rDNA duplication.

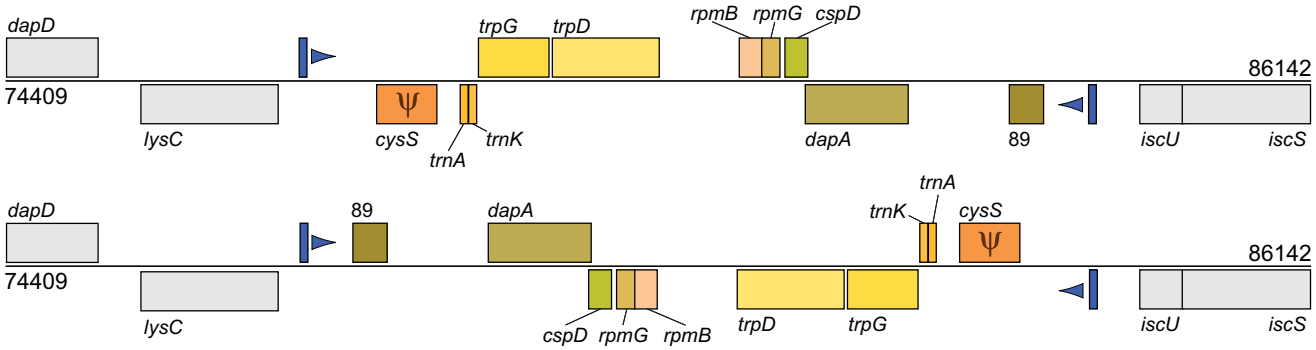

Figure 1. Schematic of the Inverted Repeat Region of the *Tremblaya* Genome

Genes flanking the inverted repeat region are shown in gray, the 71 bp inverted repeat is shown in dark blue with an arrowhead, and the genes inside the inverted repeat are represented in shades of red, yellow, and green and are shown in the two orientations observed in the population. The gene labeled 89 is TPPCIT_089, and the ψ symbol associated with *cysS* indicates a pseudogene. The orientation on top was submitted to GenBank. Complete circular diagrams of the *Tremblaya* and *Moranella* genomes can be found in Figure S1.

it codes for only 15 gene homologs involved in essential amino acid production, compared with 29 in *Tremblaya* (Figure 2). As in *Tremblaya*, no single essential amino acid pathway is complete in *Moranella* alone, and only three gene homologs involved in essential amino acid biosynthesis are present in both the *Tremblaya* and *Moranella* genomes. Strikingly, complete synthesis of tryptophan and threonine requires a patchwork of gene products from *Tremblaya* and *Moranella*, and phenylalanine, arginine, and isoleucine biosynthesis require genes from *Tremblaya*, *Moranella*, and possibly the mealybug host (Figure 2). As for the BCA homolog discussed above, genes that encode aspartate aminotransferase (AAT), ornithine aminotransferase (OAT), and threonine dehydratase (TDH) activities can be found in several insect genomes [20, 21], and copies of AAT and OAT orthologs are present in the EST library from *M. hirsutus* (GenBank accession numbers EH215711.1 [AAT] and EH214276.1 [OAT]). Additionally, a few low-coverage contigs in the present work have top hits to animal AAT, OAT, and TDH homologs in the nr database, indicating the possible presence of these genes in the *P. citri* genome (Table S1). Therefore, the last step in phenylalanine synthesis, the production of ornithine for arginine biosynthesis, and the threonine dehydratase activity of the isoleucine pathway may be performed by the insect, as has been predicted in the *Buchnera*-pea aphid symbiosis [20–22]. Our results suggest that lysine and histidine are not produced by this symbiotic partnership, because none of the missing genes in the bacterial pathways are thought be present in animal genomes [24], and no obvious homologs can be identified in the low-coverage sequence contigs. Phloem sap of various plants is known to vary widely in the levels of essential amino acids present [25], and it is possible that lysine and histidine are acquired at sufficient levels in the mealybug diet.

Gene Expression

Although the pattern of complementary gene loss and retention in the mealybug assemblage implies a functional symbiosis, especially in the context of the extreme genome reduction observed in *Tremblaya*, we sought to verify gene expression from a pathway that involved contributions from all three

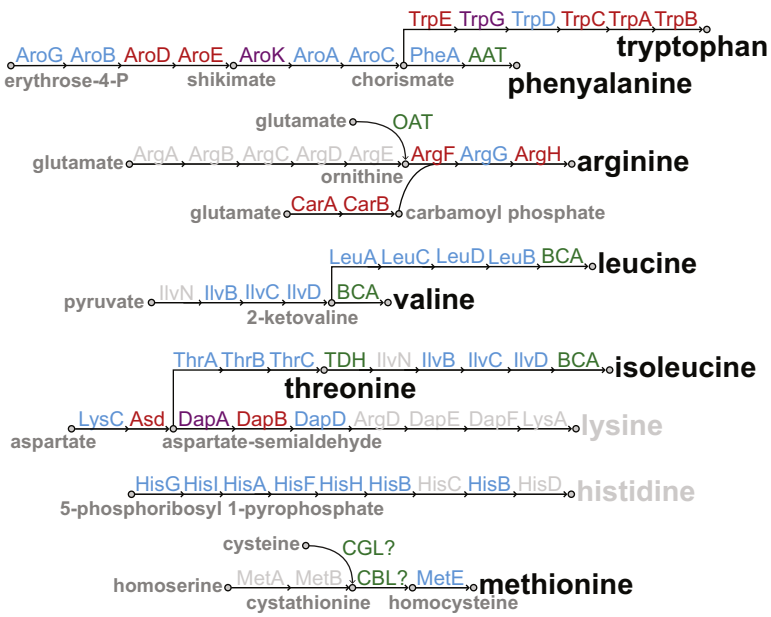

Figure 2. Predicted Essential Amino Acid Metabolic Contributions of the Mealybug-*Tremblaya*-*Moranella* Symbiosis

Gene homologs found in the *Tremblaya* genome are blue; the *Moranella* genome, red; both the *Tremblaya* and *Moranella* genomes, purple; neither the *Tremblaya* nor the *Moranella* genome, gray; activities not found in either bacterial genome but predicted to be encoded in the mealybug genome, green. Copies of *argF*, *argH*, and *carB* are found in the *Tremblaya* genome as pseudogenes. A transmission electron micrograph of the mealybug symbiosis is shown in Figure S3, and descriptions and sequences of contigs containing mealybug genes shown in green can be found in Table S1. All ten genes in the phenylalanine pathway were shown to be expressed by RT-PCR; see Figure S2.

organisms in the symbiosis. We tested for gene expression in all genes in the phenylalanine pathway—which requires gene products from *Tremblaya*, *Moranella*, and *P. citri* (Figure 2)—by RT-PCR on RNA extracted from purified mealybug bacteriomes. We targeted a region of the *P. citri* AAT mRNA predicted to be spliced (Figure S2), reasoning that a functional transcript should be processed to mature mRNA in tissues where it is actively expressed. All ten genes, including the spliced form of AAT mRNA from the low-coverage *P. citri* genome assembly, were expressed (Figure S2).

Discussion

Multispecies Nutritional Symbioses

Several multispecies symbiotic communities that provision nutrients to their hosts have been studied using genomic methods [3, 4, 6, 7, 26, 27]. In most of these previous examples, complete or near complete pathways for the synthesis of individual nutrients exist exclusively in one member of the community [3, 4, 7, 27]. For example, in both glassy-winged sharpshooter and cicada, *Sulcia muelleri* (Bacteroidetes) has near complete pathways for the production of 8 of the 10 essential amino acids, and the partnering symbiont produces the remaining two [3, 7] (in glassy-winged sharpshooter, the partner is the gammaproteobacteria *Baumannia cicadellinicola* [6]; in cicada, the partner is the alphaproteobacteria *Hodgkinia cicadicola* [1]). In the related spittlebug system, *Sulcia* has lost the ability to make tryptophan—thereby retaining the ability to make 7 instead of 8 of the essential amino acids—and the partnering bacterial symbiont, the betaproteobacteria *Zinderia insecticola*, has retained a perfectly complementary set of gene homologs for the synthesis of the remaining three essential amino acids [4]. In all of these cases, complete amino acid biosynthetic pathways are partitioned between bacterial partners; for example, gene homologs comprising the tryptophan biosynthetic pathway exist in either one symbiont or the other, never in both [4].

One system that seems to be an exception is the *Buchnera*/*Serratia* dual symbiosis in the aphid *Cinara cedri* [5, 26]. In this example, the first two steps in the tryptophan pathway (*trpEG*) are present on a plasmid in *Buchnera*, and the remaining tryptophan biosynthetic genes are found on a fragment of the *Serratia* genome, suggesting that an intermediate of the pathway must pass between *Buchnera* and *Serratia* for tryptophan to be produced [26]. However, because the *Serratia* genome is not yet complete [26], it is not known definitively that *trpEG* homologs are not encoded in the *Serratia* genome. Nevertheless, assuming that the steps in tryptophan biosynthesis are divided between *Buchnera* and *Serratia* in *C. cedri*, this example does not compare to the complexity of the step-by-step metabolic interdependency shown in the mealybug system (Figure 2).

Coordination of Essential Amino Acid Biosynthesis between *Tremblaya* and *Moranella*

As in other dual-symbiont insect systems [3, 4, 6, 7], it is unclear how transport of metabolites occurs between cosymbionts. Even when only considering phenylalanine and tryptophan synthesis, *Tremblaya* and *Moranella* would have to pass a minimum of five metabolites between the two partners to produce tryptophan, and two metabolite exchanges would be required for the synthesis of phenylalanine (not including the last step, putatively catalyzed by the mealybug AAT activity). *Moranella* encodes only a handful of genes involved in membrane transport, and none are specific for amino acids or their precursors (data not shown). If proteins were imported or exported between the cells to accomplish phenylalanine and tryptophan biosynthesis, then six proteins would have to be transported in either direction. Some components of the Sec translocation machinery are present in the *Moranella* genome (*secAEGY* and *yidC*; *secB* exists as a pseudogene), and it is possible that these are used to transport some proteins across *Moranella*'s inner membrane. A search for signal peptides in the *Moranella* proteome revealed 27 proteins with N-terminal secretory signal peptides; however, none was involved in essential amino acid biosynthesis (data not shown). Because the *Tremblaya* genome encodes no predicted transporters, it is unlikely that either *Moranella* or *Tremblaya* is capable of specifically controlling the import and export of the metabolites or enzymes needed for amino acid biosynthesis.

Translation in *Tremblaya*

Similar to other highly reduced bacterial genomes [12], *Tremblaya* has retained 19 of 21 small subunit and 25 of 33 large subunit ribosomal protein genes, suggesting that translation still occurs in *Tremblaya* cells. Although other highly reduced bacterial symbiont genomes are missing certain aminoacyl-tRNA synthetases [1–4], *Tremblaya* is the first bacterial genome published without any functional aminoacyl-tRNA synthetase copies in its genome (copies of the arginyl- and cysteinyl-tRNA synthetases exist as pseudogenes). Furthermore, *Tremblaya* is missing several gene homologs for translation-related functions—for example, both translational release factors—that are normally found in the smallest bacterial genomes [12]. Because the bacterial translational machinery is significantly different from the eukaryotic machinery (in particular the release factors [28]), it seems unlikely that all of these missing functions could be complemented by host-encoded activities. It is possible that, as in organelles [29], some *Tremblaya* genes have been transferred to the host genome and their products subsequently reimported; this would have to be accomplished by host-encoded transporters. However, in two insects that have stably associated symbiotic bacteria for which complete insect genomes exist (pea aphid and body louse), no transfer of functional genes between symbiont and host has been found [30, 31], suggesting that this may not be a common solution to missing genes in insect symbiont genomes.

The *Tremblaya* Inverted Repeat

Like other highly reduced symbiont genomes [12], *Tremblaya* encodes only a few gene homologs involved in DNA replication, and none are directly implicated in DNA repair or recombination. The loss of recombinogenic activities, combined with a dearth of repetitive sequence, is thought to explain the unusual and extreme structural stability found in reduced symbiont genomes [9, 32], where bacterial genomes separated by 20–200 million years show no rearrangements or duplications [7, 9, 10, 17]. This extreme structural stability is restricted to bacteria with severely reduced genomes that are subject to strict vertical transmission within hosts. In contrast, other obligate intracellular bacteria with larger genomes and more promiscuous transmission routes can harbor large numbers of mobile genetic elements and are more structurally dynamic [33, 34]. Because *Tremblaya* has the smallest reported bacterial genome, it is surprising to observe an inversion in its genome (Figure 1). At present, the age and origin of

this inversion is unknown. If no evolutionary advantage exists for the inversion occurring in one orientation versus the other, then it is likely a recent event, because evolutionarily old selectively neutral polymorphisms are expected to be lost in small populations, such as those that exist for intracellular symbionts [35, 36], as a result of genetic drift. If the inversion is somehow advantageous, it could be either ancient or recent, and maintained through selection; it is currently unclear what the selective advantage of this genomic polymorphism would be in *Tremblaya*.

The Nested Structure of the Mealybug Symbionts Is Likely Controlled by the Host

In the 1960s, summarizing decades of microscopical work carried out by numerous investigators, Paul Buchner described the symbiotic structures in *P. citri* bacteriocytes as "roundish or longish mucilaginous globules [now known to be *Tremblaya*] in which the symbionts [*Moranella*] are thickly embedded" [37] (Figure S3). He noted at least two morphological forms of *Moranella*: a reproductive form in which cells were small in size and in the process of dividing, and a degenerative phase in which cells became unevenly shaped and elongated [37]. The particular *Moranella* form was dependent on the life stage of the insect and seemed to be synchronized within a bacteriocyte [37].

Adding evidence to the idea that *Moranella* has life stages distinct from *Tremblaya*, recent work has shown that the infection levels of *Tremblaya* and *Moranella* are uncoupled in mealybugs, at least in males [38]. Specifically, during male development, the number of *Moranella* cells relative to *Tremblaya* cells drops significantly as the insects age, whereas in female insects, the levels of the two symbionts remain roughly equivalent over the entire life cycle [38]. These results led Kono et al. to suggest that *Tremblaya* plays an essential role in controlling the levels of the gammaproteobacteria (*Moranella*) and that a novel bacterial endo- and exocytosis-like mechanism might be involved in the *Tremblaya-Moranella* symbiosis [38]. Given that *Tremblaya* has an extremely limited coding capacity that is largely devoted to essential amino acid biosynthesis and translation, and given that only seven genes are of completely unknown function, it seems impossible that *Tremblaya* itself controls any structural aspect of the symbiosis. Likewise, the *Moranella* genome does not encode any genes involved in traditional infective strategies (such as type III or IV secretion systems) and does not indicate any obvious pathway by which it could be an active participant involved in seeking out the *Tremblaya* cytoplasm. Thus, it seems likely that the host is largely in control of the structure and organization of this bacteria-within-a-bacterium symbiosis.

The Problem of Missing Genes in Highly Reduced Symbiont Genomes, and the Nature of the *Tremblaya-Moranella* Symbiosis

The recent discoveries of bacterial symbiont genomes that are reduced in size and gene content far beyond what was once considered possible raise important but difficult-to-answer questions about how—and in some cases, whether—these bacteria carry out the necessary functions needed for their own survival, as well as for the survival of their symbiotic partners [2, 12, 39]. The missing activities resulting from such extensive gene loss could be compensated by several possible mechanisms, such as (1) gene products or metabolites of either host or bacterial origin imported from the host [30, 40, 41], (2) gene products or metabolites imported directly from the cosymbiont (if present), or (3) genetic coadaptations to the loss of genes within the reduced genome itself. *Tremblaya* presents these same issues, but with the added complexity of having a bacterial symbiont with a relatively rich gene set residing in its cytoplasm. This adds another possible mechanism—extremely speculative, to be sure—for the solution to gene loss in *Tremblaya*: (4) the direct use of *Moranella* gene products as a result of a simple, passive mechanism such as *Moranella* cell lysis within the cell membrane system of *Tremblaya*. Were the host able to control the system so that selective lysing of *Moranella* cells were possible (or if those cells occasionally spontaneously lysed on their own), this would give *Tremblaya* ready access to *Moranella* gene products that could catalyze the diverse and numerous functions missing from the *Tremblaya* genome. This scenario would obviate the need for multiple host-encoded transport processes involving molecules as potentially diverse as aminoacylated tRNAs (or their synthetases), translational control factors, several metabolites or enzymes involved in essential amino acid production, and possibly even enzymes that catalyze the *Tremblaya* genomic inversion. Evidence that the levels of *Tremblaya* and *Moranella* cells are decoupled in mealybugs [38] would seem to at least indirectly support this hypothesis. In any case, further work is needed to understand how *Tremblaya* can survive with such a limited gene set.

Genome Reduction Is an Ongoing Process in *Tremblaya*, Despite Its Already Diminutive Size

Bacterial genomes are in general compact and gene dense, with typical coding densities of 80%–90% [42]. Reduced bacterial genomes tend to be, but are not always, more gene dense than average (Table 1). Excluding *Tremblaya*, the four smallest bacterial genomes have coding densities in the range of 93%–97%, and so it is surprising that *Tremblaya* has a coding density of only 73%, making it the smallest but also one of the least gene-dense bacterial genomes published (Table 1). The other examples of bacteria with very low coding densities—*Mycobacterium leprae* (50%), *Sodalis glossinidius* (52%), and *Rickettsia prowazekii* (76%)—are also intracellular bacteria undergoing reductive genome evolution but have genomes 8–30 times larger than *Tremblaya*. These larger gene-sparse genomes are likely in the phase of genome reduction associated with the shift from a free-living to an obligate intracellular lifestyle, where the constant exposure to the stable and rich environment of the host cell combined with a severe reduction in population size (and subsequent reduction in the efficacy of purifying selection) allows large numbers of pseudogenes to accumulate [42, 43]. These pseudogenes are eventually purged from the genome through mutational patterns favoring deletions [44], leading to small gene-dense genomes such as those from insect nutritional symbionts.

As the smallest described cellular genome, *Tremblaya* is a surprising exception to this rule. One possible explanation is that the *Tremblaya* genome was gene dense prior to the acquisition of *Moranella*, and that establishment of the symbiosis relaxed the selective constraints on *Tremblaya* genes that were redundant with the more gene-rich *Moranella* genome. Basal lineages of mealybugs in the same subfamily as *P. citri* (Pseudococcinae) seem to contain *Tremblaya* without the intracellular gammaproteobacterial endosymbiont [23, 45], indicating that *Moranella* was acquired after the establishment of *Tremblaya*. The patterns of gene pseudogenization also fit this hypothesis, as most pseudogenized *Tremblaya* genes

have functional *Moranella* homologs (Table 2). Importantly, all three *Tremblaya* pseudogenes involved in essential amino acid biosynthesis (*carB*, *argF*, and *argH*) are present in the *Moranella* genome as functional copies, exactly as would be expected in such a complementary and interdependent nutritional symbiosis.

Candidatus Moranella endobia

We propose the name *Candidatus* Moranella endobia for the gammaproteobacterial symbiont living in the cytoplasm of the betaproteobacteria *Candidatus* Tremblaya princeps, which itself lives in the cytoplasm of the mealybug *Planococcus citri* bacteriocytes. Previous phylogenetic work has established *Moranella* as a member of the Enterobacteriaceae, with the tsetse fly symbiont *Sodalis glossinidius* as its closest relative [38]. Because of the uncertain relationships of the gammaproteobacterial symbionts of mealybugs [38, 45], we are only naming the gammaproteobacterial symbiont present in the genus *Planococcus*, although this name could be used in other taxa if future work warrants the designation. *Moranella* refers to the American evolutionary biologist Nancy A. Moran, and *endobia* reflects the unusual property of living exclusively in the *Tremblaya* cytoplasm (endo = inside, bia = living; feminine form). Unique properties of *Moranella* include its exclusive existence in *Tremblaya* cells and the 16S rDNA sequence GTCTTGAACTGTGGCTTTCGTAGTT (positions 839–863, *E. coli* numbering).

Accession Numbers

The *Tremblaya* and *Moranella* genome sequence data have been submitted to the DNA Data Bank of Japan, European Molecular Biology Laboratory, and GenBank databases under the accession numbers CP002244 and CP002243.

Supplemental Information

Supplemental Information includes three figures, one table, and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.cub.2011.06.051.

Acknowledgments

We thank N.A. Moran for helpful discussions and suggestions throughout the course of this work, M. Riley and O.A. Duhl for assistance in naming *Moranella*, and N. Fukushima for performing the RT-PCR experiments. This work was funded by University of Arizona Center for Insect Science National Institutes of Health Training Grant 1K12GM00708 and National Science Foundation Montana EPSCoR grant EPS-0701906 (J.P.M.).

Received: February 17, 2011
Revised: May 23, 2011
Accepted: June 21, 2011
Published online: August 11, 2011

References

1. McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2009). Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. PLoS Genet. *5*, e1000565.
2. Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H.E., Moran, N.A., and Hattori, M. (2006). The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. Science *314*, 267.
3. McCutcheon, J.P., and Moran, N.A. (2007). Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc. Natl. Acad. Sci. USA *104*, 19392–19397.
4. McCutcheon, J.P., and Moran, N.A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. Genome Biol. Evol. *2*, 708–718.
5. Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., Silva, F.J., Moya, A., and Latorre, A. (2006). A small microbial genome: The end of a long symbiotic relationship? Science *314*, 312–313.
6. Wu, D., Daugherty, S.C., Van Aken, S.E., Pai, G.H., Watkins, K.L., Khouri, H., Tallon, L.J., Zaborsky, J.M., Dunbar, H.E., Tran, P.L., et al. (2006). Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. *4*, e188.
7. McCutcheon, J.P., McDonald, B.R., and Moran, N.A. (2009). Convergent evolution of metabolic roles in bacterial co-symbionts of insects. Proc. Natl. Acad. Sci. USA *106*, 15394–15399.
8. von Dohlen, C.D., Kohler, S., Alsop, S.T., and McManus, W.R. (2001). Mealybug beta-proteobacterial endosymbionts contain gamma-proteobacterial symbionts. Nature *412*, 433–436.
9. Tamas, I., Klasson, L., Canbäck, B., Näslund, A.K., Eriksson, A.S., Wernegreen, J.J., Sandström, J.P., Moran, N.A., and Andersson, S.G. (2002). 50 million years of genomic stasis in endosymbiotic bacteria. Science *296*, 2376–2379.
10. Degnan, P.H., Lazarus, A.B., and Wernegreen, J.J. (2005). Genome sequence of *Blochmannia pennsylvanicus* indicates parallel evolutionary trends among bacterial mutualists of insects. Genome Res. *15*, 1023–1033.
11. Baumann, L., Thao, M.L., Hess, J.M., Johnson, M.W., and Baumann, P. (2002). The genetic properties of the primary endosymbionts of mealybugs differ from those of other endosymbionts of plant sap-sucking insects. Appl. Environ. Microbiol. *68*, 3198–3205.
12. McCutcheon, J.P. (2010). The bacterial essence of tiny symbiont genomes. Curr. Opin. Microbiol. *13*, 73–78.
13. Shen, P., and Huang, H.V. (1986). Homologous recombination in *Escherichia coli*: Dependence on substrate length and homology. Genetics *112*, 441–457.
14. Douglas, A.E. (1989). Mycetocyte symbiosis in insects. Biol. Rev. Camb. Philos. Soc. *64*, 409–434.
15. Moran, N.A., Plague, G.R., Sandström, J.P., and Wilcox, J.L. (2003). A genomic perspective on nutrient provisioning by bacterial symbionts of insects. Proc. Natl. Acad. Sci. USA *100* (*Suppl 2*), 14543–14548.
16. Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., and Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature *407*, 81–86.
17. van Ham, R.C., Kamerbeek, J., Palacios, C., Rausell, C., Abascal, F., Bastolla, U., Fernández, J.M., Jiménez, L., Postigo, M., Silva, F.J., et al. (2003). Reductive genome evolution in *Buchnera aphidicola*. Proc. Natl. Acad. Sci. USA *100*, 581–586.
18. Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. Annu. Rev. Microbiol. *59*, 155–189.
19. Weinstock, O., Sella, C., Chipman, D.M., and Barak, Z. (1992). Properties of subcloned subunits of bacterial acetohydroxy acid synthases. J. Bacteriol. *174*, 5560–5566.
20. Wilson, A.C., Ashton, P.D., Calevro, F., Charles, H., Colella, S., Febvay, G., Jander, G., Kushlan, P.F., Macdonald, S.J., Schwartz, J.F., et al. (2010). Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. Insect Mol. Biol. *19* (*Suppl 2*), 249–258.
21. Hansen, A.K., and Moran, N.A. (2011). Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc. Natl. Acad. Sci. USA *108*, 2849–2854.
22. Poliakov, A., Russell, C.W., Ponnala, L., Hoops, H.J., Sun, Q., Douglas, A.E., and van Wijk, K.J. (2011). Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol. Cell. Proteomics *10*, M110, 007039.
23. Hardy, N.B., Gullan, P.J., and Hodgson, C.J. (2008). A subfamily-level classification of mealybugs (Hemiptera: Pseudococcidae) based on integrated molecular and morphological data. Syst. Entomol. *33*, 51–71.
24. Payne, S.H., and Loomis, W.F. (2006). Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. Eukaryot. Cell *5*, 272–276.
25. Sandstrom, J., and Pettersson, J. (1994). Amino acid composition of phloem sap and the relation to intraspecific variation in pea aphid (*Acyrthosiphon pisum*) performance. J. Insect Physiol. *40*, 947–955.
26. Gosalbes, M.J., Lamelas, A., Moya, A., and Latorre, A. (2008). The striking case of tryptophan provision in the cedar aphid *Cinara cedri*. J. Bacteriol. *190*, 6026–6029.
27. Woyke, T., Teeling, H., Ivanova, N.N., Huntemann, M., Richter, M., Gloeckner, F.O., Boffelli, D., Anderson, I.J., Barry, K.W., Shapiro, H.J.,

et al. (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. Nature *443*, 950–955.
28. Inagaki, Y., and Ford Doolittle, W. (2000). Evolution of the eukaryotic translation termination system: Origins of release factors. Mol. Biol. Evol. *17*, 882–889.
29. Timmis, J.N., Ayliffe, M.A., Huang, C.Y., and Martin, W. (2004). Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. Nat. Rev. Genet. *5*, 123–135.
30. Nikoh, N., McCutcheon, J.P., Kudo, T., Miyagishima, S.Y., Moran, N.A., and Nakabachi, A. (2010). Bacterial genes in the aphid genome: Absence of functional gene transfer from *Buchnera* to its host. PLoS Genet. *6*, e1000827.
31. Kirkness, E.F., Haas, B.J., Sun, W., Braig, H.R., Perotti, M.A., Clark, J.M., Lee, S.H., Robertson, H.M., Kennedy, R.C., Elhaik, E., et al. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. Proc. Natl. Acad. Sci. USA *107*, 12168–12173.
32. Silva, F.J., Latorre, A., and Moya, A. (2003). Why are the genomes of endosymbiotic bacteria so stable? Trends Genet. *19*, 176–180.
33. Bordenstein, S.R., and Reznikoff, W.S. (2005). Mobile DNA in obligate intracellular bacteria. Nat. Rev. Microbiol. *3*, 688–699.
34. Newton, I.L.G., and Bordenstein, S.R. (2011). Correlations between bacterial ecology and mobile DNA. Curr. Microbiol. *62*, 198–208.
35. Andersson, S.G., and Kurland, C.G. (1998). Reductive evolution of resident genomes. Trends Microbiol. *6*, 263–268.
36. Moran, N.A., and Wernegreen, J.J. (2000). Lifestyle evolution in symbiotic bacteria: Insights from genomics. Trends Ecol. Evol. (Amst.) *15*, 321–326.
37. Buchner, P. (1965). Endosymbiosis of Animals with Plant Microorganisms (New York: Interscience).
38. Kono, M., Koga, R., Shimada, M., and Fukatsu, T. (2008). Infection dynamics of coexisting beta- and gammaproteobacteria in the nested endosymbiotic system of mealybugs. Appl. Environ. Microbiol. *74*, 4175–4184.
39. Tamames, J., Gil, R., Latorre, A., Peretó, J., Silva, F.J., and Moya, A. (2007). The frontier between cell and organelle: Genome analysis of *Candidatus* Carsonella ruddii. BMC Evol. Biol. *7*, 181.
40. Kondo, N., Nikoh, N., Ijichi, N., Shimada, M., and Fukatsu, T. (2002). Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. Proc. Natl. Acad. Sci. USA *99*, 14280–14285.
41. Dunning Hotopp, J.C., Clark, M.E., Oliveira, D.C., Foster, J.M., Fischer, P., Muñoz Torres, M.C., Giebel, J.D., Kumar, N., Ishmael, N., Wang, S., et al. (2007). Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science *317*, 1753–1756.
42. Ochman, H., and Davalos, L.M. (2006). The nature and dynamics of bacterial genomes. Science *311*, 1730–1733.
43. Andersson, J.O., and Andersson, S.G. (2001). Pseudogenes, junk DNA, and the dynamics of *Rickettsia* genomes. Mol. Biol. Evol. *18*, 829–839.
44. Mira, A., Ochman, H., and Moran, N.A. (2001). Deletional bias and the evolution of bacterial genomes. Trends Genet. *17*, 589–596.
45. Thao, M.L., Gullan, P.J., and 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*, 3190–3197.