



Cicada Endosymbionts Have tRNAs That Are Correctly Processed Despite Having Genomes That Do Not Encode All of the tRNA Processing Machinery

 James T. Van Leuven,^{a*} Meng Mao,^b Denghui D. Xing,^a Gordon M. Bennett,^b  John P. McCutcheon^a

^aDivision of Biological Sciences, University of Montana, Missoula, Montana, USA

^bDepartment of Life and Environmental Sciences, University of California, Merced, Merced, California, USA

ABSTRACT Gene loss and genome reduction are defining characteristics of endosymbiotic bacteria. The most highly reduced endosymbiont genomes have lost numerous essential genes related to core cellular processes such as replication, transcription, and translation. Computational gene predictions performed for the genomes of the two bacterial symbionts of the cicada *Diceroprocta semicineta*, “*Candidatus* Hodgkinia cicadicola” (*Alphaproteobacteria*) and “*Ca. Sulcia muelleri*” (*Bacteroidetes*), have found only 26 and 16 tRNA genes and 15 and 10 aminoacyl tRNA synthetase genes, respectively. Furthermore, the original “*Ca. Hodgkinia cicadicola*” genome annotation was missing several essential genes involved in tRNA processing, such as those encoding RNase P and CCA tRNA nucleotidyltransferase as well as several RNA editing enzymes required for tRNA maturation. How these cicada endosymbionts perform basic translation-related processes remains unknown. Here, by sequencing eukaryotic mRNAs and total small RNAs, we show that the limited tRNA set predicted by computational annotation of “*Ca. Sulcia muelleri*” and “*Ca. Hodgkinia cicadicola*” is likely correct. Furthermore, we show that despite the absence of genes encoding tRNA processing activities in the symbiont genomes, symbiont tRNAs have correctly processed 5′ and 3′ ends and seem to undergo nucleotide modification. Surprisingly, we found that most “*Ca. Hodgkinia cicadicola*” and “*Ca. Sulcia muelleri*” tRNAs exist as tRNA halves. We hypothesize that “*Ca. Sulcia muelleri*” and “*Ca. Hodgkinia cicadicola*” tRNAs function in bacterial translation but require host-encoded enzymes to do so.

IMPORTANCE The smallest bacterial genomes, in the range of about 0.1 to 0.5 million base pairs, are commonly found in the nutritional endosymbionts of insects. These tiny genomes are missing genes that encode proteins and RNAs required for the translation of mRNAs, one of the most highly conserved and important cellular processes. In this study, we found that the bacterial endosymbionts of cicadas have genomes which encode incomplete tRNA sets and lack genes required for tRNA processing. Nevertheless, we found that endosymbiont tRNAs are correctly processed at their 5′ and 3′ ends and, surprisingly, that mostly exist as tRNA halves. We hypothesize that the cicada host must supply its symbionts with these missing tRNA processing activities.

KEYWORDS endosymbionts, evolutionary biology, genomics, RNAseq, symbiosis, tRNA maturation

The smallest and most gene-poor bacterial genomes are from the nutritional endosymbionts of sap-feeding insects (1, 2). These tiny genomes have lost even seemingly essential genes, such as those involved in DNA replication and translation (2). In terms of genome size and coding capacity, insect endosymbiont genomes span the

Citation Van Leuven JT, Mao M, Xing DD, Bennett GM, McCutcheon JP. 2019. Cicada endosymbionts have tRNAs that are correctly processed despite having genomes that do not encode all of the tRNA processing machinery. *mBio* 10:e01950-18. <https://doi.org/10.1128/mBio.01950-18>.

Editor Margaret J. McFall-Ngai, University of Hawaii at Manoa

Copyright © 2019 Van Leuven et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to John P. McCutcheon, john.mccutcheon@umontana.edu.

* Present address: James T. Van Leuven, Center for Modeling Complex Interactions, University of Idaho, Moscow, Idaho, USA.

Received 9 May 2019

Accepted 17 May 2019

Published 18 June 2019

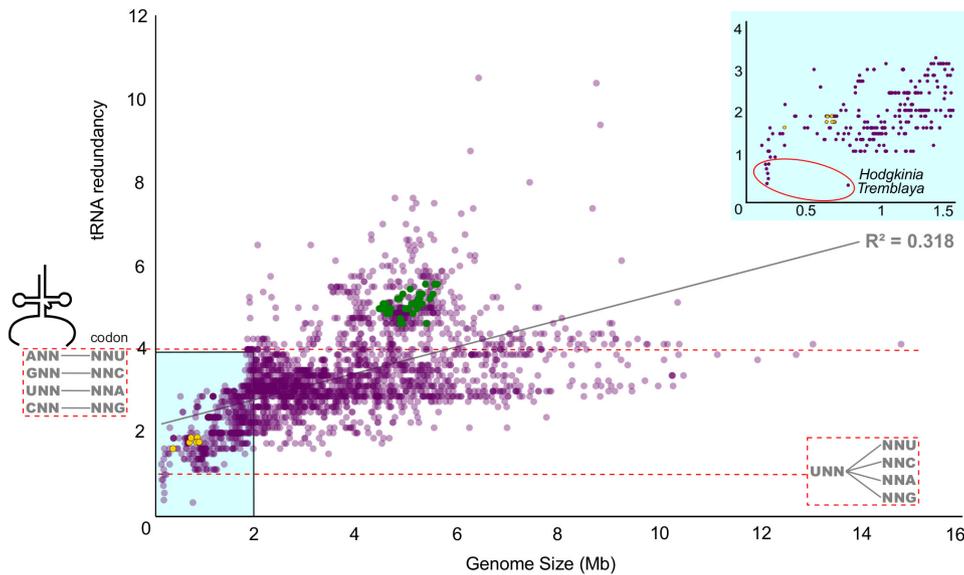


FIG 1 Genome size and tRNA redundancy are positively correlated. Each fully sequenced bacterial genome is shown as a dot ($n = 2,761$). tRNA redundancy data represent the number of total 4-box tRNA genes in a genome over the number of 4-box families. The red dashed line at $y = 1$ shows a limit where only one tRNA was found from each of the eight 4-box families. Below this limit, it is unclear if the organism has enough tRNAs for translation. The red dashed line at $y = 4$ shows one tRNA gene for each 4-box codon. *Buchnera aphidicola* and *Escherichia coli* are shown as yellow and green dots, respectively. Theoretically, all bacteria could function with a redundancy value of 1.

gap between their free-living bacterial cousins that retain the genes for performing essential cell functions and organelles of symbiotic origin that have lost nearly all of their genes (that is, the mitochondria and plastids), particularly those involved in replication, transcription, and translation (2, 3). Insect endosymbiont genomes thus provide an opportunity to learn more about key adaptations enabling codependent and highly integrated endosymbioses but in associations that are younger and more labile than the classic cellular organelles.

"*Candidatus Hodgkinia cicadicola*" (*Alphaproteobacteria*; hereafter *Hodgkinia*) and "*Ca. Sulcia muelleri*" (*Bacteroidetes*; hereafter *Sulcia*) have two of the smallest bacterial genomes published. *Sulcia* and *Hodgkinia* are obligate nutritional endosymbionts of many cicadas, where they work together to provide essential amino acids to their hosts (4). In the cicada *Diceroprocta semicincta* (Insecta: Hemiptera: Cicadoidea), the *Hodgkinia* genome is 143 kilobase pairs (kbp) in length and the *Sulcia* genome is 277 kbp. Only 16 tRNA genes and 10 of the 20 required aminoacyl-tRNA synthetase genes (aaRSs) were found in the *Hodgkinia* genome using computational methods (5). While the number of tRNA genes encoded in bacterial genomes is quite variable (Fig. 1), theoretical estimates predict that a minimum of ~20 tRNAs are required to translate all 61 codons (6, 7). *Hodgkinia* is missing tRNA genes needed to decode leucine, valine, arginine, serine, threonine, aspartic acid, asparagine, and tyrosine codons (Fig. 2). The mealybug endosymbiont "*Candidatus Tremblaya princeps*" (here *Tremblaya*) also falls below the theoretical limit of 20, encoding only 8 to 12 tRNA genes and 0 or 1 aaRSs, depending on the strain (Fig. 1; see also Table S1 in the supplemental material) (8–11). However, *Tremblaya* is unusual in hosting its own intrabacterial endosymbiont, which may provide the missing tRNAs and aaRSs (9). There is no such explanation for the apparent lack of tRNA, aaRS, and tRNA processing genes in *Hodgkinia*.

Functional tRNAs are generated by a complex, multistep process that usually requires trimming off transcribed nucleotides that precede (5' leader) and follow (3' trailer) the predicted tRNA gene, posttranscriptional nucleotide editing at numerous

positions, the addition of a terminal 3' CCA, and aminoacylation of the mature tRNA to produce a molecule that is active on the ribosome. After transcription, 5' leaders are trimmed by the nearly universal ribozyme RNase P (12, 13). The 3' trailer is cleaved off by a combination of endonucleases and/or exonucleases (14), and if a terminal 3' CCA is not encoded in the genome, one is added by a nucleotidyl transferase (15, 16). Finally, tRNA nucleotides are modified by a variety of enzymes at various conserved positions (17–19). These important modifications influence tRNA tertiary structure and interactions with cellular enzymes and proteins (17).

The original published computational annotation of the *Hodgkinia* genome from *D. semicineta* lacks most genes related to tRNA processing (5). It is missing the RNA (*rnpB*) and protein (*rnpA*) subunits of RNase P and the nucleases responsible for 3' trailer trimming. *Hodgkinia* does not encode a CCA-adding enzyme, despite having only one tRNA gene with a genome-encoded terminal CCA. This *Hodgkinia* genome contains only three genes involved in tRNA editing (*mnmE*, *mnmA*, and *mnmG*), all of which are predicted to be involved in the conversion of uridine to 5-methylaminomethyl-2-thiouridine at U34 (5, 20). Because the genes encoding aaRSs and tRNA processing enzymes are large and typically highly conserved across life, it is unlikely that these proteins were missed in the original genome annotation.

The dramatic tRNA and aaRS gene loss observed in *Hodgkinia* is extremely rare in bacteria; only *Hodgkinia* and *Tremblaya* lack the theoretical minimum number of tRNAs and aaRS genes in their genomes (Fig. 1; see also Table S1). However, detection of tRNA genes in highly degraded endosymbiont genomes—particularly in mitochondrial genomes—is notoriously difficult (21, 22). Many mitochondrial tRNAs have unusual structures, in some cases missing entire d-loops, making them easy to miss by computational gene-finding algorithms unless they are specifically trained to find them (23). Similarly, the archaeal tRNA genes in the degenerate genome of *Nanoarchaeum* were initially missed because of the unusual way in that they are split and permuted (24, 25). We therefore reasoned that our initial computational annotation of the tRNAs of *Hodgkinia* (5) might be incomplete.

The apparent loss of these key genes in *Hodgkinia* raises several questions. Were *Hodgkinia* tRNAs and other small RNAs missed during computational gene prediction? For the tRNAs present on the genome, are their 5' and 3' ends correctly processed? Are *Hodgkinia* tRNAs modified only at the U34 wobble position as expected based on the gene content of the genome? Do host enzymes complement the missing symbiont genes? Here we address these questions by sequencing eukaryotic messenger and total small RNAs from the cicada species *D. semicineta*.

RESULTS

Endosymbiont tRNA genes are correctly annotated. We sequenced small RNAs expressed in cicada bacteriome tissues to experimentally search for unannotated *Hodgkinia* and *Sulcia* tRNAs. Among 145,176,847 quality-filtered reads of length 18 to 90 nucleotides (nt), 15.6% and 28.4% map to the *Hodgkinia* and *Sulcia* genomes, respectively (see Table S2 in the supplemental material). While the mean levels of genome-wide read coverage of the *Sulcia* and *Hodgkinia* genomes were similar (224 reads/bp and 164 reads/bp, respectively), *Hodgkinia* tRNA expression levels were extremely variable (Fig. 3).

FIG 2 Legend (Continued)

lowest usage; e.g., of the four alanine codons found in *Hodgkinia* protein-coding genes, GCU is used most frequently. The nucleotide sequences for *Hodgkinia* alanine codons (which make up 13.7% of the genome) are shown as an example; all others are omitted for simplicity of display. The presence of perfectly paired tRNAs is indicated by a dark gray box. Light gray fill indicates that a tRNA could possibly be used to translate the codon by N34 wobble. The anticodon sequence of each tRNA is shown to the right of its cognate codons and is written 5' to 3'. N34 modifications that are likely needed for tRNA-codon pairing are indicated by a superscript "1." A red-colored three-letter amino acid abbreviation indicates that the genome does not encode that aaRS. tRNA abundances determined by RNAseq are shown in the "Expression rank" boxes.

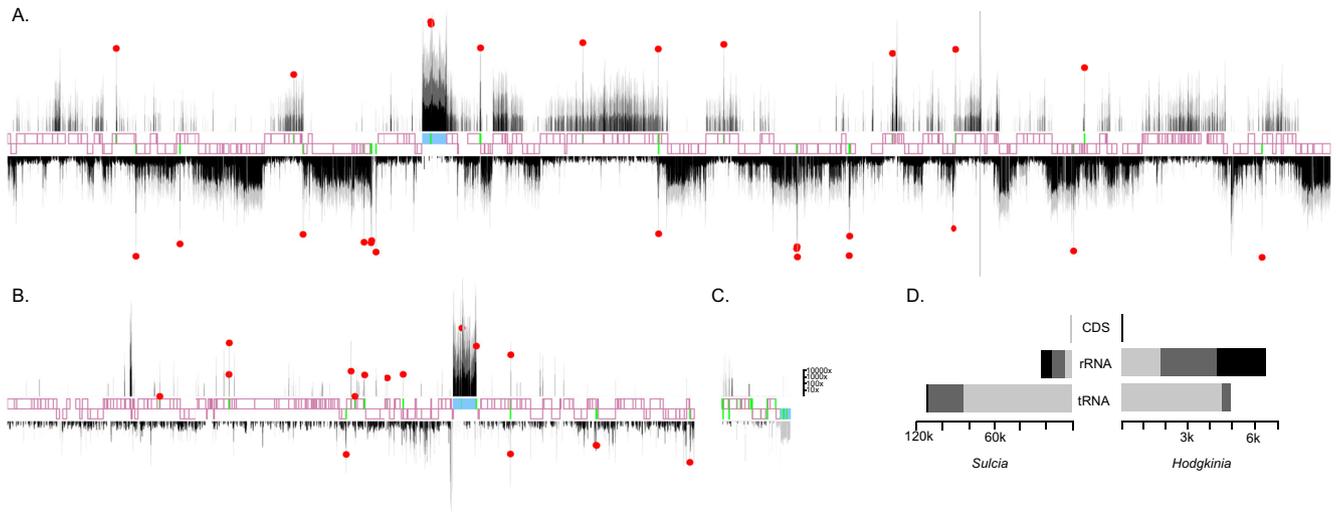


FIG 3 RNA expression patterns from the *Sulcia*, *Hodgkinia*, and mitochondrial genomes show relatively low levels of expression of *Hodgkinia* tRNAs. (A to C) Read depth plotted across the *Sulcia* (A), *Hodgkinia* (B), and mitochondrial (C) genomes. Protein-coding genes, rRNA genes, and tRNA genes on the sense and antisense strands are shown in pink, blue, and green, respectively. Red dots represent the highest read depth for each tRNA. Coverage depths for reads of lengths 18 to 47, 48 to 89, and 90 to 100 are shown in light gray, dark gray, and black, respectively, and data (drawn on a log₁₀ scale) represent results after subtraction of background (median genic) coverage. (D) Median coverage depths for *Sulcia* (left) and *Hodgkinia* (right) for each gene category and read length. The bars are colored as described for panels A to C.

In *Sulcia*, we found that >99% of the reads mapped to tRNAs, transfer-messenger RNA (tmRNA), RNase P, and ribosomal RNAs (Table S2). The median depth of the reads mapping to the rest of the genome was 380×. To identify unannotated tRNAs, we manually inspected read coverage across the genome to identify regions with pronounced spikes in coverage. Only one of these high coverage spikes was found in an intergenic region (positions 75,798 to 75,839). All other spikes occurred within the bounds of annotated protein-coding sequences (CDSs). The spike at positions 75,798 to 75,839 corresponds to a Thr^{GGT} tRNA that was unannotated in the original *Sulcia* annotation due to a small assembly error in the published genome. We updated the *Sulcia* genome to reflect this change (NCBI reference sequence [NC_012123.1](https://www.ncbi.nlm.nih.gov/nuccore/NC_012123.1); see Table S3 for primer sequences). In contrast to the tRNA reads, the reads contributing to CDS spikes did not have terminal CCAs or predicted folded structures that resembled tRNA (26). In summary, none of the spikes in coverage can be attributed to unannotated tRNAs.

In *Hodgkinia*, we found high expression of predicted tRNAs, ribosomal RNAs, and the 5' ends of CDSs. Our analysis uncovered previously unannotated RNase P and tmRNA genes (discussed below ["Discovery of unannotated RNase P and tmRNA genes in *Hodgkinia*" section]). Of the total of 16 tRNA genes of *Hodgkinia*, many are not expressed above background (Fig. 3; see also Table S4). tRNA genes Gly061 and Gly108 each have no full-length reads aligning to them, even when allowing for 5 to 8 mismatches to accommodate modified bases (Table S4).

Mapping reads to endosymbiont genomes allowed us to characterize tRNA processing patterns and to uncover expression of unannotated genes but was not well suited for identifying spliced or otherwise unconventional RNAs, such as intron-containing tRNAs. Therefore, we collapsed identical reads of length 48 to 90 nt and searched these collapsed reads for sequences that contained predicted tRNA genes, that ended in CCA, or that partially matched the *Sulcia* or *Hodgkinia* genomes. Because the number of unique (or nearly unique) reads was very high, we chose a minimum coverage cutoff value of 100×. The only transcript we found belonging to *Sulcia* or *Hodgkinia* (Blast E value, <1E−25) that was not an annotated tRNA or tmRNA was the previously unannotated *Sulcia* Thr^{GGT} transcript described above. Given these data, we conclude that the computational tRNA predictions for *Hodgkinia* and *Sulcia* are correct and that these genomes do not encode complete sets of tRNAs.

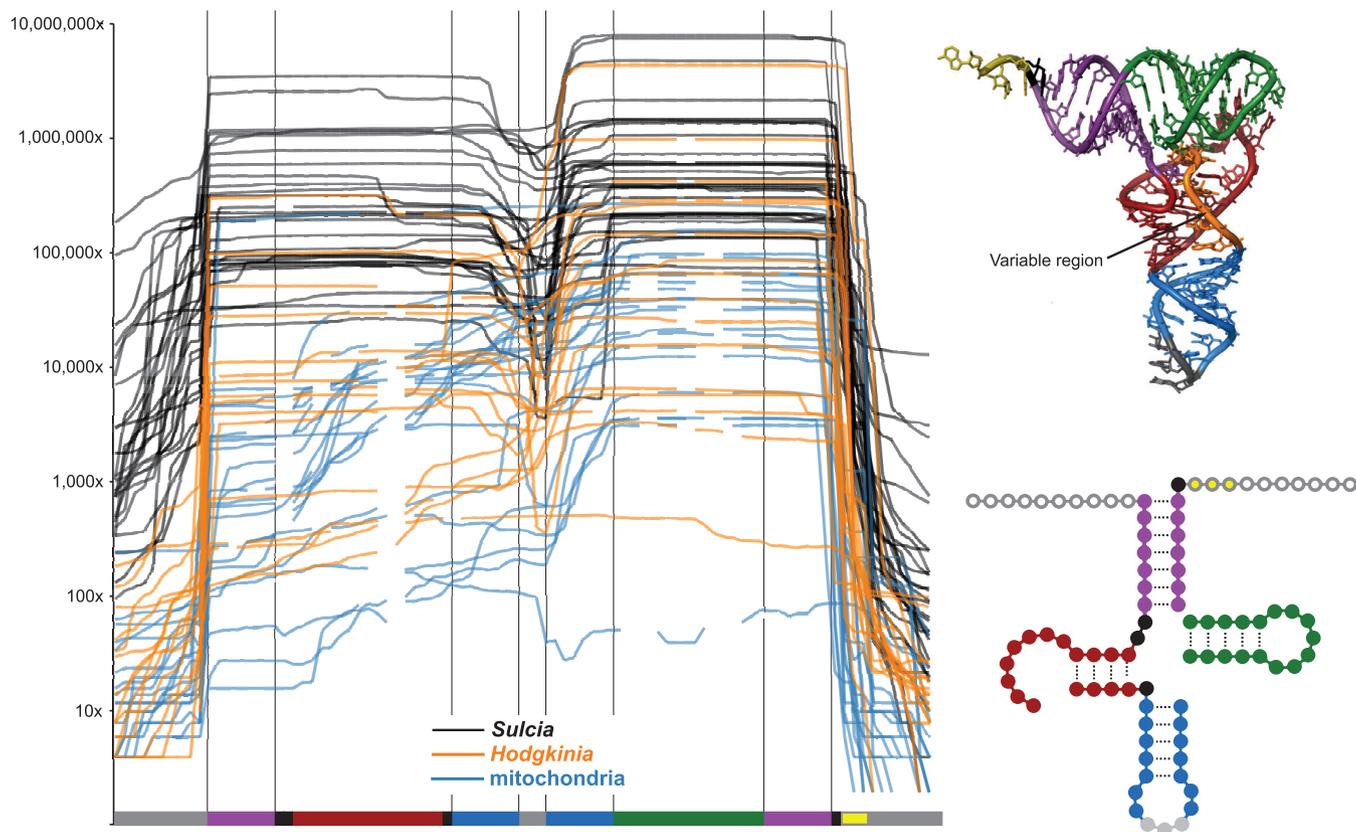


FIG 4 Dynamic range of tRNA half expression in *Sulcia* and *Hodgkinia*. Line graphs show read depth across each tRNA in *Hodgkinia* (orange), *Sulcia* (black), and the cicada mitochondria (blue). Because the tRNA lengths differed, only the bases present in all symbiont tRNA genes are shown in the structural diagrams of tRNAs. Mitochondrial tRNAs were often missing these regions, as indicated by gaps in the line graph. A total of 18 to 100 nucleotide reads were mapped for this figure.

Most tRNAs are found as tRNA halves. Among 145,176,847 quality-filtered reads that were 18 to 90 nt in length, only 0.05% (74,651) and 4.8% (2,520,749) mapped in full length to *Hodgkinia* and *Sulcia* tRNAs, respectively (Table S2). Most reads were shorter than the predicted tRNA genes (Fig. 4; see also Table S4). The abundance of short reads could have been caused by RNA degradation, PCR bias toward short amplicons during library creation, or bona fide stable tRNA halves (27–30). Because our library creation protocol included reverse transcription (RT) after 5' and 3' RNA adapter ligation, these short reads are not likely due to reverse transcriptase failing to proceed through modified nucleotides, because these truncated reverse transcription products would not include both priming sites for PCR. To verify that the large number of tRNA halves that we found was not due to a bias against full-length but modified tRNAs, we created small RNA libraries where the RNA had been treated with the demethylating enzyme AlkB prior to reverse transcription (31, 32). This enzyme removes N¹-methyladenosine (m¹A), N¹-methylguanosine (m¹G), and N³-methylcytosine (m³C) modifications from tRNAs, and treatment of tRNAs with this enzyme has been shown to help recover more tRNAs by sequencing from some organisms (31, 32). We found that AlkB treatment had no effect on the length distribution or nucleotide modification rates of reads mapped to *Sulcia*, *Hodgkinia*, and mitochondrial DNA (mtDNA) (see Fig. S1 in the supplemental material; see also Table S2), suggesting that these methylation tRNA modifications were not affecting our results in any meaningful way. We therefore conclude that there are large amounts of either tRNA degradation products or stable tRNA halves present in the adult cicada bacteriome.

Discovery of unannotated RNase P and tmRNA genes in *Hodgkinia*. By aligning small RNA reads to the *Hodgkinia* genome, we found expression of previously un-

notated RNase P (*rnpB*) and tmRNA (*ssrA*) genes at positions 25,448 to 25,794 and 92,713 to 93,140 in the *Hodgkinia* NC_012960.1 genome. Given that the 5' ends of *Hodgkinia* tRNAs are correctly processed (see the next section), and that we cannot find any other RNA nucleases in *Hodgkinia*, it seems likely that this RNase P is responsible for the observed 5' tRNA processing. The permuted *Hodgkinia* tmRNA is coded for in the reverse direction, on the antisense strand (Fig. S2). All components typically conserved in tmRNA structures can be found in the proposed tmRNA gene; however, the peptide tag does not end in the canonical YALAA sequence (33, 34). The coding RNA and acceptor RNAs are separated by a 129-nt intervening sequence containing complementary sequences needed for folding, and very few reads map to this region. Polymorphic sites indicate CCA addition at the 3' end of both the coding and acceptor RNAs, further supporting the idea of the likely functionality of tmRNAs (35), especially given the presence of the gene for its protein ligand (*smpB*) in at least two *Hodgkinia* genomes from different cicada species (36). We also observed reads of various lengths at both the 5' and 3' ends of the tmRNA gene, indicating that end trimming probably occurs.

The 5' end processing of *Hodgkinia* and *Sulcia* tRNAs. Many reads aligning to *Hodgkinia* and *Sulcia* tRNA genes extend past the predicted gene boundary, suggesting that they are transcribed with 5' leaders and that these extra nucleotides are trimmed off (Fig. 4). We previously predicted the presence of the RNA moiety of RNase P RNA in *Sulcia* (5), and we now predict the presence of this gene in *Hodgkinia* as described above. To help verify the activity of both the *Hodgkinia* and *Sulcia* RNase P RNAs, we created two pools of Illumina-compatible small RNA libraries. In the creation of both libraries, RNA adapter sequences are ligated directly to the small RNA pools at the 5' and 3' ends. Adapter ligation can be blocked by either a triphosphorylated or diphosphorylated 5' RNA end, but a functional RNase P generates 5' monophosphate ends which are and active for ligation (80). By splitting one pool of small RNAs into two groups, one treated with tobacco acid pyrophosphatase (TAP), which would generate 5' monophosphates, and one untreated, we tested the 5' processed state of bacteriome tRNAs (13, 19, 37). In both *Hodgkinia* and *Sulcia*, we found no difference between the tRNA sets from each library (Spearman's rank correlation, $P < 0.005$; Table S5), suggesting that the 5' ends of tRNAs are monophosphorylated in the cicada bacteriome. This is consistent with the presence of an active RNase P enzyme in both bacterial endosymbionts.

The 3' ends of *Hodgkinia* and *Sulcia* tRNAs are correctly processed. The processing of tRNA 3' ends is more complicated than the processing of 5' ends. If a 3' CCA is not encoded on the genome (the majority of *Sulcia* and *Hodgkinia* tRNA genes do not have encoded CCA ends), one must be added by a CCA transferase enzyme after the 3' trailer sequence is removed through processing by various RNA nucleases. Consistent with the presence of 3' trailer sequences, we found reads extending past the predicted 3' boundary of *Hodgkinia* and *Sulcia* tRNAs (Fig. 4; see also Fig. 5). *Sulcia* contains a putative RNase ([ACU52822.1](#)) that could potentially process the 3' trailer, although the gene is most similar to that corresponding to RNase Y, which is involved in mRNA decay (38). *Hodgkinia* contains no such RNA nuclease candidates. We also observed reads ending in C, CC, and CCA that mapped to *Sulcia* and *Hodgkinia* tRNA genes, indicating that the nucleotides of the terminal CCA were added one at a time to the 3' end of transcripts lacking 3' trailers (Fig. 5). *Sulcia* contains a tRNA CCA nucleotidyl transferase, but *Hodgkinia* does not. Our mRNA sequencing (mRNA-Seq) data show upregulation of a tRNA CCA nucleotidyl transferase encoded on the cicada genome in the bacteriome tissue, although we do not know if this enzyme is active on *Hodgkinia* tRNAs. However, in plants, mammals, and yeast, isoforms of this protein are localized to both the cytoplasm and to organelles (39–41), suggesting that this enzyme might be targeted to different cellular compartments in cicada cells.

tRNA modification occurs in *Hodgkinia* and *Sulcia*. The *Hodgkinia* genome encodes only three genes known to be involved in tRNA modification, all of which are

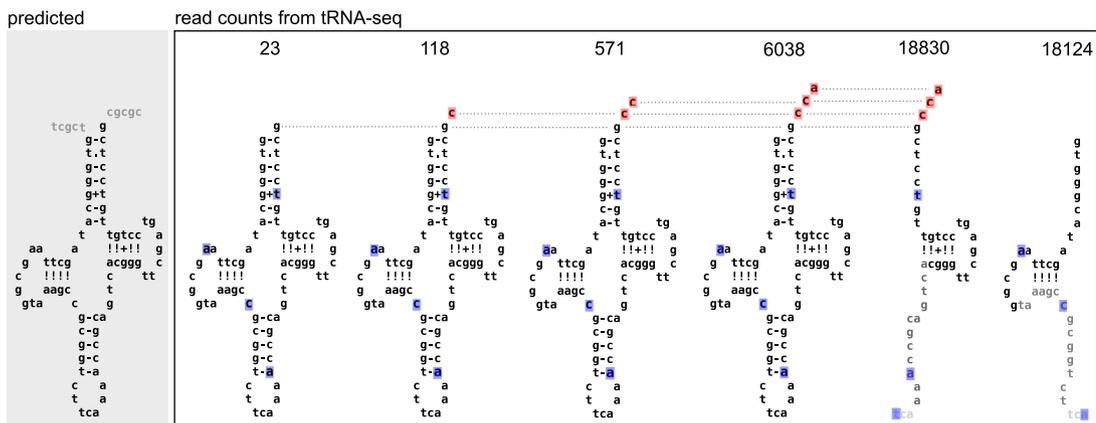


FIG 5 tRNA processing occurs in a stepwise manner, but full-length tRNAs comprise a small minority of the total reads. The majority of reads mapping to the *Hodgkinia* tRNA Trp062 gene (51,683) mapped to one of the secondary structures shown. Polymorphic sites (>2%) are shown in blue (RNA modifications) or red (CCA addition). tRNA halves are colored to indicate common sites of RNA degradation, where black letters indicate the highest read depth.

predicted to act on U34: *mnmA*, *mnmG* (*gidA*), and *mnmE* (*trmE*) (5). MnmA catalyzes the 2-thiolation of U to s2U; MnmG and MnmE form a dimer that catalyzes the conversion of s2U to nm5s2U (20). The *Sulcia* genome carries these three genes, along with *truA* and *tilS* (5). TruA modifies U38-U40 to pseudouridine, and TilS converts C34 to I34, enabling the specific recognition of Met versus Ile anticodons (20). We found sequence polymorphisms in tRNA transcripts—which we interpret as potential base modifications (31, 42–44)—at several sites other than expected position 34 in *Hodgkinia* (positions 1, 4, 9, 11, 15, 16, 20, 23, 26, 27, 28, 37, 39, 40, 42, 43, 49, 57, 58, 62, 63, 64, and 68) and expected positions 34 and 38 to 40 in *Sulcia* (positions 6, 10, 11, 16, 20, 28, 31, 37, 49, 53, 54, 60, 64, 69, and 71) (Fig. 6; see also Table S6). Importantly, we observed no differences in modification patterns between AlkB-treated and untreated tRNAs (Table S6), suggesting that *Sulcia* and *Hodgkinia* do not have meaningful levels of m¹A, m¹G, or m³C modifications in their tRNAs. For a position to be called polymorphic, we required a read depth of at least 10× and greater than 2% polymorphism at the modified site. Interestingly, *Hodgkinia* tRNAs are more highly modified than *Sulcia* tRNAs in both the diversity of modification and the total number of tRNAs modified. Among the 16 tRNAs of *Hodgkinia*, 15 have at least one putatively modified site, versus 8 of 28 in *Sulcia* and 10 of 22 in the mitochondrial genome (Fig. 6; see also Table S6).

Cicadas upregulate some aminoacyl tRNA synthetase transcripts in bacteriome tissues. We sequenced cicada mRNAs from four insects in search of tRNA processing genes that could complement those missing from the *Hodgkinia* and *Sulcia* genomes. A total of 189,137 genes were assembled by Trinity from a combined total of 255,193,489 100-bp reads. The longest and mean contig lengths were 18,931 bp and 765 bp, respectively. Seventy-eight percent of the paired-end reads properly mapped in the transcriptome assembly, giving an overall Transrate assembly score of 0.21. A total of 21,651 nonredundant annotated protein-coding genes were found on the assembled contigs using Trinotate (Table S1). BUSCO orthologs were searched, and 75% of the core arthropod genes were found in our assembly, a typical value for hemipteran genomes (8). We used edgeR to identify 1,778 genes (Table S2) that were differentially expressed between bacteriome and insect tissues (false-discovery rate [FDR], ≤0.05). Among these 1,778 differentially expressed genes (DEGs), 1,211 had higher expression values in the bacteriome samples (Table S3). Most DEGs were not annotated by Trinotate and remained hypothetical proteins. Only nine genes upregulated in the bacteriome were annotated with a function involving tRNA maturation or charging, including five copies of D-tyrosyl-tRNA deacylase; two copies of an aminoacyl tRNA synthase complex-interacting multifunctional protein, tRNA uracil(54)-C(5)-methyltransferase B; and tRNA modification GTPase MnmE (Table S7). The gene expression profiles were wildly different between

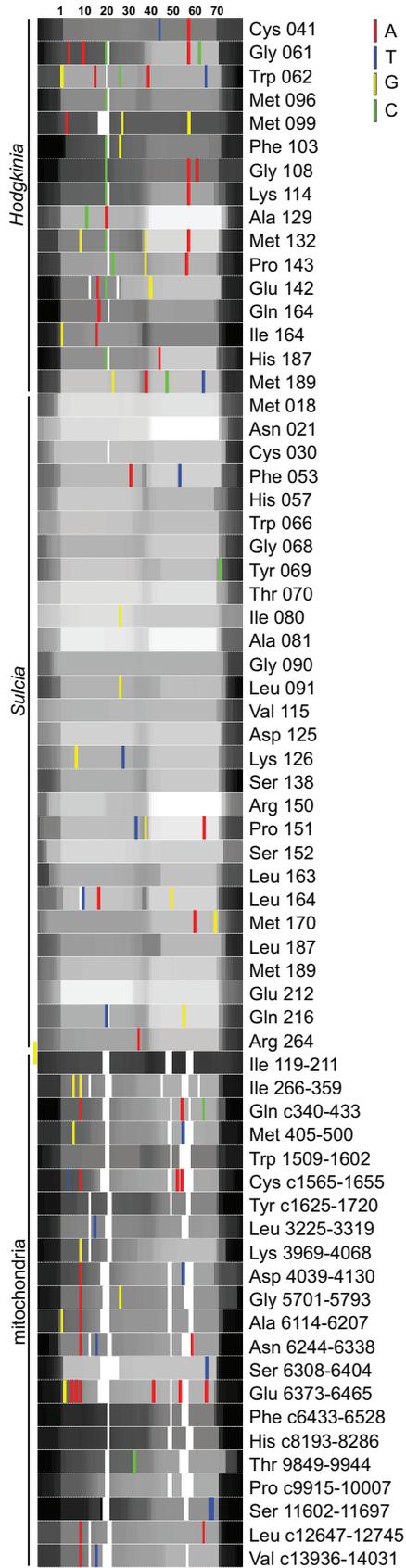


FIG 6 Expression level of individual tRNAs shown with polymorphic sites that have frequencies of greater than 2%. The per-base read depth was log transformed and is shown on a color scale (Continued on next page)

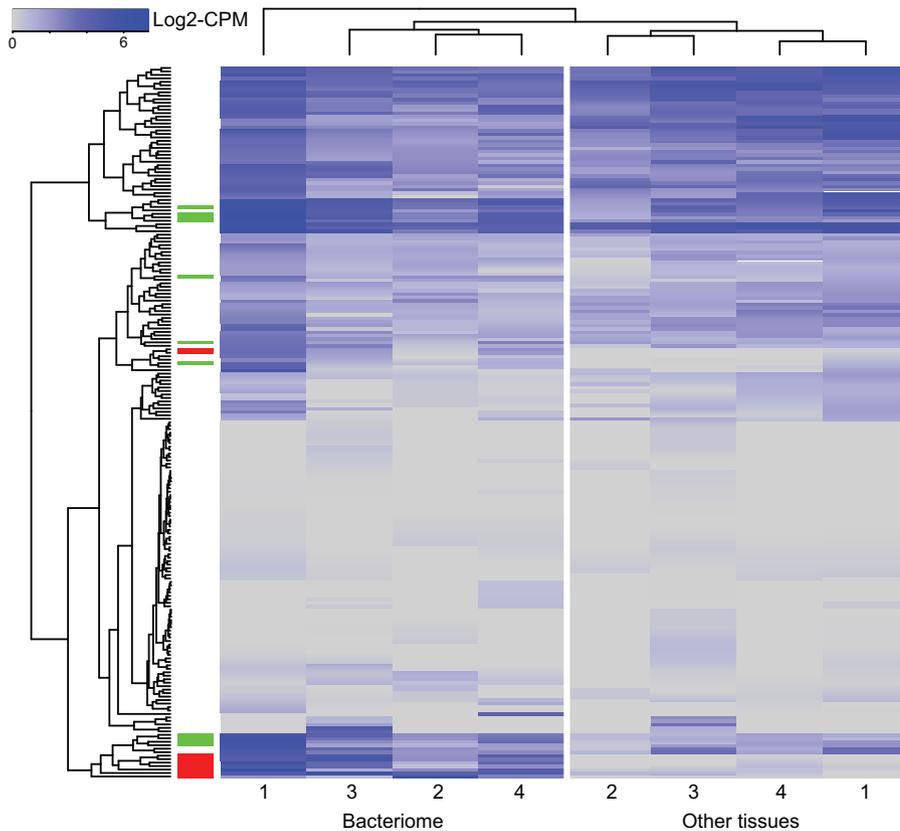


FIG 7 Host tRNA processing genes are rarely overexpressed in bacteriome tissues. Data representing expression of Trinity assembled genes whose Trinotate annotations involve tRNA processing are shown for bacteriome and nonbacteriome tissues for replicate insects 1 to 4. Genes and samples are clustered by Euclidian distance in R. Differentially expressed genes (edgeR, $P > 0.05$) are indicated by a red block in the leftmost column, and genes differentially expressed, but not significantly so, are indicated in green.

insects (Fig. 7), so we examined the most highly differentially expressed genes (top quartile) which had nonzero count-per-million values for at least three of the four replicates and were not identified as differentially expressed by edgeR. Among the 205 tRNA maturation genes identified by Trinotate, 9 not identified by edgeR were highly expressed in the bacteriome but not in other cicada tissues (Fig. 7). These genes are expressed at levels that are 50-fold to 2,586-fold higher in bacteriomes than in other tissues. Some of the genes highly expressed in bacteriomes are complementary with the functions missing from the *Hodgkinia* and *Sulcia* genomes, including the arginine, cysteine, and serine aaRSs (Table 1).

Five potential horizontal gene transfers (HGTs) from bacteria were identified in cicadas by Mao et al. (45) from the transcriptome assemblies reported here (Table S8), including a pectin lyase gene (*pel*) and two copies of a transcription regulatory gene (*yebC*), an AAA-ATPase gene, and a ribosome recycling factor gene (*frr*). None of these genes are predicted to play direct roles in tRNA production or maturation. Overall, the cicada genome seems to contain few genes from HGT that obviously complement lost endosymbiont genes (with the possible exception of *frr*), in contrast to several related sap-feeding insects (9, 45–48).

FIG 6 Legend (Continued)

corresponding to 0 to 255, making even large differences in expression levels difficult to distinguish by eye. Low expression is shown in black and high expression in white. Polymorphic sites are colored according to their genomic sequence. Ten bases of leader and trailer are shown as described for Fig. 5. Gaps are shown in white and are apparent primarily in mitochondrial RNAs (mtRNAs). See Table S2 for gene name descriptions.

TABLE 1 Expression of cicada aaRS genes in complementing *Sulcia* and *Hodgkinia*

Genome	Expression (counts per million) ^a																				
	<i>alaS</i>	<i>asnS</i>	<i>aspS</i>	<i>argS</i>	<i>cysS</i>	<i>glnS</i>	<i>gltX</i>	<i>glyS</i>	<i>hisS</i>	<i>ileS</i>	<i>leuS</i>	<i>lysS</i>	<i>metG</i>	<i>pheS</i>	<i>proS</i>	<i>serS</i>	<i>thrS</i>	<i>trpS</i>	<i>tyrS</i>	<i>valS</i>	
<i>Hodgkinia</i>	X						X	X	X	X			X	X	X				X		X
<i>Sulcia</i>	X	X				X	X	X		X	X	X	X	X	X	X			X	X	X
Cicada mitochondria	-0.1	1.2	0.5	0	16.8		-2.3		0	-4.7	-19.0		-0.4	-4.4	-0.7	-1.3		-2.2	-2.3	0.2	
Cicada cytoplasm	3.0	-9.1	3.4	44	3.8	6.2	-8.7	-10.8	9.9	-20.9	4.2	-8.3	1.4	-0.6		12.9	29.1	6.1	-3.8	0.6	

^aThe presence of aaRS genes in the *Hodgkinia* and *Sulcia* genomes is indicated by an "X." Differential expression values (counts per million) of cicada aaRS genes are shown, where positive values indicate overexpression in the bacteriome (log₂ counts per million in the bacteriome minus log₂ counts per million in other tissues).

Methodological caveats and complexities. We found several unusual results while analyzing our data. First, we isolated highly abundant transcripts containing predicted tRNAs not belonging exclusively to *Hodgkinia*, *Sulcia*, or host tRNA genes. In these cases, the two halves of the transcript aligned to separate genomic locations of the endosymbionts or even to the genomes of separate organisms (half to *Sulcia* and half to *Hodgkinia*). In all cases, these were tRNA-like sequences that were joined near the anticodon. We could not amplify these RNAs from total RNA using gene-specific RT-PCR and thus concluded that they represent by-products of the RNA ligation steps of the library preparation. This serves as a cautionary result with respect to this method. Despite this, in all cases, true *Hodgkinia* and *Sulcia* tRNAs were also amplified, cloned, and sequenced as positive controls (see Table S3 for primer sequences). Second, our protocol may have selected against heavily modified tRNAs because of reverse transcription blocking (31, 49), although we tried to rule this possibility out by treating a library with the demodifying enzyme AlkB (Fig. S1). Since we found many tRNA sequences with a polymorphism(s) at conventionally modified sites, it is possible that reverse transcriptase can proceed over some modifications, consistent with previous findings (42–44, 49).

DISCUSSION

5' tRNA processing from the genomes of *Sulcia* and *Hodgkinia* can be explained, but 3' processing cannot. We did not computationally identify RNase P in our first *Hodgkinia* genome annotation (5), but here we report the presence of this gene in the *Hodgkinia* genome. While still apparently lacking the protein component (*rnpA*), trimming of 5' tRNA leaders could still occur with the RNA component alone because it is a ribozyme (50). The 5' processing of both *Hodgkinia* and *Sulcia* tRNAs can now be explained by normal cellular processes.

The way that *Sulcia* and *Hodgkinia* trim and process the 3' ends of their tRNAs is less clear. Cleavage of the 3' trailer from pre-tRNAs can be accomplished by a variety of redundant exonucleases and/or endonucleases (14), none of which are encoded in *Sulcia* or *Hodgkinia*. In *Escherichia coli*, RNase PH, RNase T, RNase D, and RNase II can all trim back the 3' end of pre-tRNAs (14). In *Sulcia*, we identified an RNA nuclease with similarity to RNase Y. Although RNase Y is typically thought to initiate mRNA decay, it is also implicated in multiple RNA processing tasks (38). No such putative enzyme can be found in the *Hodgkinia* genome from *D. semicineta*, although a gene for the endoribonuclease YbeY is present in *Hodgkinia* genomes from other cicada species (51). After 3' tRNA trailers are trimmed, a terminal CCA is added by a CCA-adding enzyme (19). This gene is conserved across all domains of life (52), including in *Sulcia*, but is apparently missing in *Hodgkinia*. We have identified transcripts belonging to *Hodgkinia* that have terminal C, CC, and CCAs. The presence of these variants indicates that *Hodgkinia* tRNAs are exposed to an active CCA-adding enzyme. One potential candidate is the host genome-encoded mitochondrial copy that is upregulated in bacteriome tissues (see Table S7 in the supplemental material). Mitochondrial CCA-adding enzymes are known to have broad specificity and functionality (39, 53) and so may work on *Hodgkinia* tRNAs if that are somehow trafficked to *Hodgkinia* cells by the host.

Base modification, even if off-target and not biologically relevant, cannot be explained by the genes encoded in the *Sulcia* and *Hodgkinia* genomes. Base

modifications are essential for tRNA aminoacylation and codon recognition and have been well described in previous work (19), including the bacterial endosymbiont *Buchnera aphidicola* (43). tRNAs of *Buchnera* are posttranscriptionally modified at N37 and N34, but most of the observed modifications can be performed by enzymes encoded on the *Buchnera* genome (*miaA*, *miaB*, *rimN*, *trmD*, *tadA*, *queA*, *mnmA*, *mnmE*, *iscS*, *tilS*, and *gidA*). It is therefore a bit surprising that tRNA modifications were detected in *Hodgkinia* and *Sulcia* when the genes for these modification enzymes are not carried in their genomes. We note that *Hodgkinia* and mitochondrial tRNAs have more modifications in common than do those of *Hodgkinia* and *Sulcia* (Fig. 6; see also Table S6). Assuming that these “modifications” do not represent experimental artefacts of some sort, it seems possible that some of the host tRNA modification enzymes are active in *Hodgkinia* and *Sulcia*, although it is not clear if these modifications are biologically relevant.

The potential problems with old cicadas. Two results from this work suggest that using older cicadas—that is, cicadas that have emerged from the ground and are near the end of their long lives—for RNA work might be less than ideal. The first is our finding that both *Sulcia* tRNAs and *Hodgkinia* tRNAs exist primarily as tRNA halves in adult cicadas. While tRNA halves have been found in numerous other organisms and have been shown to have regulatory functions, they are often associated with organismal stress (27–30). It seems possible that harvesting cicadas as adults, near the end of their lives, might result in gene expression and RNA processing patterns associated with stress. Likewise, the mRNA expression values show wide swings in transcript abundance over our four biological replicates, and this also may be due to a breakdown of transcriptional regulation in elderly cicadas. It would be interesting to examine the state of cicada RNAs in actively growing nymphs in future work.

Translational machinery is shared between host and organelle in eukaryotes. The most gene-rich mitochondrial genomes of the Jakobid protists look very much like endosymbiont genomes and contain a full-fledged set of about 30 tRNA genes (54). In contrast, the most gene-poor mitochondrial genomes of some trypanosomatids and alveolates contain no tRNA genes (55). The range is similar in plastids, including 1 to 30 tRNA genes (56, 57). The sets of retained tRNAs in *Hodgkinia* and *Tremblaya* and in organelles overlap, but there are considerable differences (Table S1). Among 22 tRNA anticodon species in *Hodgkinia* and *Tremblaya* from the mealybug *Planococcus citri*, only trnA^{UGC}, trnI^{GAU}, trnM^{CAU}, and trnF^{UGC} are present in both genomes (Table S1). The degrees of tRNA gene conservation of *Hodgkinia*, mitochondria, and plastids are strikingly similar (58, 59).

Unlike in insect endosymbionts, organellar aaRSs have been completely transferred to the nuclear genome (60–62). The processes involved in aaRS and tRNA import into organelles are complex and are reviewed elsewhere (60, 63, 64). The mechanisms for localizing charged tRNAs to the organelle are diverse and organism specific. In humans, for example, all aaRSs except GlnRS and LysRS are bacterially derived and targeted specifically to mitochondria to charge mitochondrially encoded tRNAs (65). In contrast, only 45 aaRS genes are expressed from the *Arabidopsis thaliana* genome; 21 are found only in the cytoplasm, 21 are dually targeted, 2 are chloroplast specific, and 1 is targeted to all three cellular compartments (81). The mitochondrial genomes of apicomplexans are missing tRNAs and aaRSs. They must import aminoacylated tRNAs that were charged in the cytoplasm (66). It is worth noting that the mitochondrial and plastid genomes of the plant *A. thaliana* contain 22 and 30 tRNA genes and yet cytosolic tRNAs are still imported (67). The import of seemingly unnecessary tRNAs occurs quite frequently, and in most cases, the role of redundant tRNAs in organelles is unknown (64). Our results strongly suggest that *Hodgkinia* and possibly *Sulcia* import tRNAs from their host cicada.

Despite the massive genetic integration of organelle with host (68, 69), most mitochondria and plastids retain genomes and thus some level of genetic autonomy. Gene retention patterns in the genomes of highly reduced bacterial symbionts also suggest a hurdle to giving up independence of some processes to their hosts, espe-

cially transcription, translation, and replication (2, 3). The *Hodgkinia* genome is lacking many genes involved in these processes, and our results reported here indicate that these missing functions are likely complemented by its host. While these results support the idea that some obligate symbioses may undergo major transitions to become a highly integrated unit (70), other recent data from other cicadas show that these integrated units are not inevitably stable, as sometimes *Hodgkinia* is lost and replaced by a new fungal symbiont (71).

MATERIALS AND METHODS

Sequencing small RNAs. The bacteriomes of three wild-caught female *D. semicincta* cicadas collected around Tucson, AZ, in July 2010 and July 2012 were dissected in RNAlater. RNA was also extracted from a combination of brain, abdomen, and leg tissues as a nonbacteriome control. RNA was purified using a Roche High Pure miRNA isolation kit. Illumina libraries were made using a ScriptMiner small RNA-Seq library preparation kit (Epicenter). One sample was split into index 1 and index 2, and index 2 was treated with the supplied TAP enzyme to reduce the 5' end to a monophosphate. The remaining two samples (index 3 and index 4) were processed without TAP treatment. In a separate experiment, RNA from one additional cicada was split into two samples, and one of the two was treated with the AlkB enzyme to remove m¹A, m¹G, and m³C nucleotide modifications. AlkB demethylation was performed as previously described (31). Briefly, about 15 pmol of small RNA was treated with a mixture of purified wild-type *E. coli* AlkB (500 pmol) and mutant AlkB (a D135S substitution [32]; 625 pmol) in 100 μ l of demethylation buffer for 2 h at room temperature. The reaction was quenched by adding EDTA to reach a final concentration of 5 mM. Both the AlkB-treated RNA and the untreated RNA were purified using a Roche High Pure miRNA isolation kit. Reverse transcription was done with an adapter-specific primer using Superscript III (Invitrogen). Libraries were amplified for 15 rounds using FailSafe PCR enzyme mix (Epicenter) and the supplied primers. PCR bands that were 50 to 300 nt in size (including 113-nt adapters) were cut from an 8% polyacrylamide gel after staining with SYBR Safe (Invitrogen). Gel cutouts were shredded using 0.5-ml tubes with needle holes in the bottom and eluted in 300 μ l 0.5 M ammonium acetate for 3.5 h at 37°C, and the contents were separated using a 0.22- μ m-pore-size sterile filter. DNA was purified by isopropanol precipitation. A total of 226,712,931 100-nt single-end reads were generated on three HiSeq lanes at the Vincent J. Coates Genomics Sequencing Laboratory of the University of California (UC), Berkeley.

Transcriptome sequencing (RNAseq) read processing for small RNAs. Adapter sequences were trimmed using Cutadapt version 1.0 (-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -g AATGATAC GGCGACCACCGACAGGT TCAGAGTTCTACAGTCCGACGATC -O 7) (72) and quality filtered (fastq_quality_filter -q 20 -p 90) using FASTX-Toolkit version 0.0.12 (http://hannonlab.cshl.edu/fastx_toolkit/). Reads shorter than 18 nt were discarded because identical matches of up to 16 nt in length can be found between different symbiont tRNA genes. Reads shorter than 48 nt were assumed to be too short to make a functional tRNA (73). Reads were mapped to *Hodgkinia* and *Sulcia* genomes and tRNA genes using either bowtie-1.0.0 (-best -maqerr 150 -seedlen 18) or bwa-0.7.5 aln (-n 0.08 -i 2) (74, 75).

De novo small RNA discovery. Identical reads were compressed using FASTX-Toolkit (fastx-collapser) and aligned to *Hodgkinia* and *Sulcia* tRNA genes (blastall 2.2.25, blastn -E 1E-25). Sequences that did not align to known tRNA genes were aligned to the *Hodgkinia* and *Sulcia* full-genome sequences (blastn -e 1E-10). The remaining sequences that did not align to the bacterial genomes were considered cicada sequences, and tRNAs were predicted using tRNAscan-SE 1.21 and ARAGORN 1.2.34 (76, 77). Nearly identical sequences were grouped into contigs using CAP3 (78). Collapsed sequences with different anticodons, 5' leaders, or 3' trailers that assembled together in CAP3 were separated.

Comparing TAP-treated libraries to untreated libraries. Reads that were 20 to 100 nt and 70 to 100 nt in length were mapped to a multi-fasta file containing *Hodgkinia*, *Sulcia*, and mitochondrial tRNA genes plus 15 bp of genome sequence flanking the gene using bowtie-0.12.7 (-f -S -n 3). The orders of tRNA coverage were compared between indexes using Spearman rank correlation (see Table S5 in the supplemental material). Scripts from Trinity v20140717 packages align_and_estimate_abundance.pl, abundance_estimates_to_matrix.pl, run_DE_analysis.pl, and analyze_diff_expr.pl were used to compare differential transcription levels (with parameters -SS_lib_type F -est_method RSEM -aln_method bowtie -seedlen 18 -maqerr 150 -best). Using separate *de novo* approaches for library 1 and library 2 (with 48-nt to 100-nt reads), we normalized tRNA coverage (number of reads per tRNA/total number of reads mapping to all tRNAs). A ratio representing the difference between library 1 coverage and library 2 coverage was calculated for each tRNA. For all values less than zero, the inverse was taken and multiplied by -1. In this way, we tried to capture the relative differences in expression levels for all tRNAs from all organisms. These data were tabulated such that the source organism, paired amino acid types, anticodon sequence, and relative expression changes for every tRNA were used as inputs in a linear model in R (Table S5).

mRNAseq. Reads from isolate SRR952383 were pooled with reads from an additional three cicadas and assembled using Trinity v2.1.1 with kmer_length = 25 and min_contig_length = 48 (79). Assembled transcripts matching *Sulcia* and *Hodgkinia* were removed using bwa-mem v07.12. The remaining transcripts were annotated using Trinotate v2.0.2. Potential horizontally transferred genes (HTGs) were identified from the cicada transcripts with BLASTP searches against the bacterial nr database and verified against known HTGs from the leafhopper *Macrostelus quadrilineatus* (45), since none currently exists for *D. semicincta*. Among the 398,377 Trinotate annotated gene isoforms, 115,253 resulted in blast hits in the Swiss-Prot, TrEMBL, or EggNOG databases. A total of 1,284 contaminated transcripts were removed from

the Trinity assembly and Trinotate annotation list following identification during submission to NCBI's transcript sequence archive (TSA). Genes involved in tRNA processing were identified by searching for "tRNA" or "tRNA" in the Trinotate annotation. Expression values were analyzed and visualized in R using the edgeR, trinotateR, and ggplots2 packages.

Cloning and sequencing of specific tRNAs. tRNAs of interest were reverse transcribed using tRNA-specific primers (SuperScript III), amplified (NEB OneTaq), and cloned (Invitrogen TOPO TA cloning kit) using standard procedures. Plasmids were purified (Omega Plasmid Mini) and sequenced with the standard M13F primer.

Bioinformatics. Complete bacterial genome sequences were downloaded from <ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/all.fna.tar.gz>. Chromosomal sequences were searched for tRNA genes using tRNAscan-SE 1.21 and the bacterial model (77). tRNA redundancy was simply calculated by dividing the number of 4-box and 6-box family tRNA genes by the number of 4-box and 6-box families.

Data availability. Trinotate annotations for all transcripts assembled using Trinity are available at <https://doi.org/10.6084/m9.figshare.6687089.v1>.

Data corresponding to the expression levels of gene-level transcripts that were differentially expressed between bacteriome and other cicada tissues, shown as log₂ fold change, are available at <https://doi.org/10.6084/m9.figshare.6687089.v1>.

Trinotate annotations for transcripts that are expressed at higher levels in cicada bacteriome tissues than in other cicada tissues are available at <https://doi.org/10.6084/m9.figshare.6687089.v1>.

Raw Illumina reads are available in NCBI's SRA database for the small RNA (SRR5081152 to 5081157) and transcriptome (SRS470226 and SRR5060328 to 5060333) data sets. Assembled transcript sequences are available from the NCBI TSA database (GGPH00000000.1).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mBio.01950-18>.

FIG S1, EPS file, 0.5 MB.

FIG S2, EPS file, 0.2 MB.

TABLE S1, DOCX file, 0.01 MB.

TABLE S2, DOCX file, 0.02 MB.

TABLE S3, DOCX file, 0.01 MB.

TABLE S4, DOCX file, 0.02 MB.

TABLE S5, DOCX file, 0.01 MB.

TABLE S6, DOCX file, 0.03 MB.

TABLE S7, DOCX file, 0.02 MB.

TABLE S8, DOCX file, 0.01 MB.

ACKNOWLEDGMENTS

We thank members of the McCutcheon laboratory for technical assistance and conceptual feedback, Juan Alfanzo for many helpful discussions about tRNA, Todd Lowe and Patricia Chan for running a custom Infernal search on the *Hodgkinia* genome, Filip Husník for contributions to Table S1, Bodil Cass for cicada collection, and Jessica Warren and Dan Sloan for gifts of the AlkB enzymes.

This work was supported by National Science Foundation (NSF) grants IOS-1256680 and IOS-1553529 and by National Aeronautics and Space Administration Astrobiology Institute Award NNA15BB04A. J.T.V. was supported by National Institutes of Health (NIH) grant P20 GM104420.

REFERENCES

1. McCutcheon JP, Moran NA. 2011. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10:13–26. <https://doi.org/10.1038/nrmicro2670>.
2. Moran NA, Bennett GM. 2014. The tiniest tiny genomes. *Annu Rev Microbiol* 68:195–215. <https://doi.org/10.1146/annurev-micro-091213-112901>.
3. McCutcheon JP. 2016. From microbiology to cell biology: when an intracellular bacterium becomes part of its host cell. *Curr Opin Cell Biol* 41:132–136. <https://doi.org/10.1016/j.cceb.2016.05.008>.
4. McCutcheon JP. 2010. The bacterial essence of tiny symbiont genomes. *Curr Opin Microbiol* 13:73. <https://doi.org/10.1016/j.mib.2009.12.002>.
5. McCutcheon JP, McDonald BR, Moran NA. 2009. Origin of an alternative genetic code in the extremely small and GC-rich genome of a bacterial symbiont. *PLoS Genet* 5:e1000565. <https://doi.org/10.1371/journal.pgen.1000565>.
6. van der Gulik P, Hoff W. 2011. Unassigned codons, nonsense suppression, and anticodon modifications in the evolution of the genetic code. *J Mol Evol* 73:59–69. <https://doi.org/10.1007/s00239-011-9470-3>.
7. Osawa S, Jukes TH, Watanabe K, Muto A. 1992. Recent evidence for evolution of the genetic code. *Microbiol Rev* 56:229–264.
8. Husník F, McCutcheon JP. 2016. Repeated replacement of an intra bacterial symbiont in the tripartite nested mealybug symbiosis. *Proc Natl Acad Sci U S A* 113:E5416–E5424. <https://doi.org/10.1073/pnas.1603910113>.
9. Husník F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrög D, Wilson ACC, von Dohlen CD, Fukatsu T, McCutcheon JP. 2013. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* 153:1567–1578. <https://doi.org/10.1016/j.cell.2013.05.040>.

10. López-Madrugal S, Latorre A, Porcar M, Moya A, Gil R. 2011. Complete genome sequence of “*Candidatus Tremblaya princeps*” strain PCVAL, an intriguing translational machine below the living-cell status. *J Bacteriol* 193:5587–5588. <https://doi.org/10.1128/JB.05749-11>.
11. McCutcheon JP, von Dohlen CD. 2011. An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr Biol* 21:1366–1372. <https://doi.org/10.1016/j.cub.2011.06.051>.
12. Evans D, Marquez SM, Pace NR. 2006. RNase P: interface of the RNA and protein worlds. *Trends Biochem Sci* 31:333–341. <https://doi.org/10.1016/j.tibs.2006.04.007>.
13. Randau L, Schröder I, Söll D. 2008. Life without RNase P. *Nature* 453:120–123. <https://doi.org/10.1038/nature06833>.
14. Condon C. 2007. Maturation and degradation of RNA in bacteria. *Curr Opin Microbiol* 10:271–278. <https://doi.org/10.1016/j.mib.2007.05.008>.
15. Deutscher MP. 1990. Ribonucleases, tRNA nucleotidyltransferase, and the 3' processing of tRNA. *Prog Nucleic Acids Res Mol Biol* 39:209–240. [https://doi.org/10.1016/S0079-6603\(08\)60628-5](https://doi.org/10.1016/S0079-6603(08)60628-5).
16. Zhu L, Deutscher MP. 1987. tRNA nucleotidyltransferase is not essential for *Escherichia coli* viability. *EMBO J* 6:2473–2477. <https://doi.org/10.1002/j.1460-2075.1987.tb02528.x>.
17. Jackman JE, Alfonzo JD. 2013. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdiscip Rev RNA* 4:35–48. <https://doi.org/10.1002/wrna.1144>.
18. Limbach PA, Crain PF, McCloskey JA. 1994. Summary: the modified nucleosides of RNA. *Nucleic Acids Res* 22:2183–2196. <https://doi.org/10.1093/nar/22.12.2183>.
19. Söll D, RajBhandary UL. 1995. tRNA: structure, biosynthesis, and function. ASM Press, Washington, DC.
20. Dunin-Horkawicz S, Czerwoniec A, Gajda MJ, Feder M, Grosjean H, Bujnicki JM. 2006. MODOMICS: a database of RNA modification pathways. *Nucleic Acids Res* 34:D145–D149. <https://doi.org/10.1093/nar/gkj084>.
21. Soma A, Onodera A, Sugahara J, Kanai A, Yachie N, Tomita M, Kawamura F, Sekine Y. 2007. Permuted tRNA Genes expressed via a circular RNA intermediate in *Cyanidioschyzon merolae*. *Science* 318:450–453. <https://doi.org/10.1126/science.1145718>.
22. Wolstenholme DR, Macfarlane JL, Okimoto R, Clary DO, Wahleithner JA. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc Natl Acad Sci U S A* 84:1324–1328. <https://doi.org/10.1073/pnas.84.5.1324>.
23. de Bruijn MHL, Schreier PH, Eperon IC, Barrell BG, Chen EY, Armstrong PW, Wong JFH, Roe BA. 1980. A mammalian mitochondrial serine transfer RNA lacking the “dihydrouridine” loop and stem. *Nucleic Acids Res* 8:5213–5222. <https://doi.org/10.1093/nar/8.22.5213>.
24. Randau L, Söll D. 2008. Transfer RNA genes in pieces. *EMBO Rep* 9:623–628. <https://doi.org/10.1038/embor.2008.101>.
25. Watanabe Y, Suematsu T, Ohtsuki T. 2014. Losing the stem-loop structure from metazoan mitochondrial tRNAs and co-evolution of interacting factors. *Front Genet* 5:109. <https://doi.org/10.3389/fgene.2014.00109>.
26. Gruber AR, Lorenz R, Bernhart SH, Neubock R, Hofacker IL. 2008. The Vienna RNA Websuite. *Nucleic Acids Res* 36:W70–W74. <https://doi.org/10.1093/nar/gkn188>.
27. Choudhury S, Slotkin K, Martinez G. 2017. tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Res* 45:5142–5152. <https://doi.org/10.1093/nar/gkx103>.
28. Haiser HJ, Karginov FV, Hannon GJ, Elliot MA. 2008. Developmentally regulated cleavage of tRNAs in the bacterium *Streptomyces coelicolor*. *Nucleic Acids Res* 36:732–741. <https://doi.org/10.1093/nar/gkm1096>.
29. Jackowiak P, Nowacka M, Strozycycki PM, Figlerowicz M. 2011. RNA degradome—its biogenesis and functions. *Nucleic Acids Res* 39:7361–7370. <https://doi.org/10.1093/nar/gkr450>.
30. Thompson DM, Parker R. 2009. Stressing out over tRNA cleavage. *Cell* 138:215–219. <https://doi.org/10.1016/j.cell.2009.07.001>.
31. Zheng G, Qin Y, Clark WC, Dai Q, Yi C, He C, Lambowitz AM, Pan T. 2015. Efficient and quantitative high-throughput tRNA sequencing. *Nat Methods* 12:835–837. <https://doi.org/10.1038/nmeth.3478>.
32. Cozen AE, Quartley E, Holmes AD, Hrabeta-Robinson E, Phizicky EM, Lowe TM. 2015. ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat Methods* 12:879–884. <https://doi.org/10.1038/nmeth.3508>.
33. Giudice E, Mace K, Gillet R. 21 March 2014, posting date. Trans-translation exposed: understanding the structures and functions of tmRNA-SmpB. *Front Microbiol* <https://doi.org/10.3389/fmicb.2014.00113>.
34. Keiler KC, Shapiro L, Williams KP. 2000. tmRNAs that encode proteolysis-inducing tags are found in all known bacterial genomes: a two-piece tmRNA functions in *Caulobacter*. *Proc Natl Acad Sci U S A* 97:7778–7783. <https://doi.org/10.1073/pnas.97.14.7778>.
35. Keiler KC. 2008. Biology of trans-translation. *Annu Rev Microbiol* 62:133–151. <https://doi.org/10.1146/annurev.micro.62.081307.162948>.
36. Hudson CM, Lau BY, Williams KP. 13 August 2014, posting date. Ends of the line for tmRNA-SmpB. *Front Microbiol* <https://doi.org/10.3389/fmicb.2014.00421>.
37. Efstratiadis A, Vournakis JN, Donis-Keller H, Chaconas G, Dougall DK, Kafatos FC. 1977. End labeling of enzymatically decapped mRNA. *Nucleic Acids Res* 4:4165–4174. <https://doi.org/10.1093/nar/4.12.4165>.
38. Chen Z, Itzek A, Malke H, Ferretti JJ, Kreth J. 2013. Multiple roles of RNase Y in *Streptococcus pyogenes* mRNA processing and degradation. *J Bacteriol* 195:2585–2594. <https://doi.org/10.1128/JB.00097-13>.
39. von Braun SS, Sabetti A, Hanic-Joyce PJ, Gu J, Schlieff E, Joyce P. 2007. Dual targeting of the tRNA nucleotidyltransferase in plants: not just the signal. *J Exp Bot* 58:4083–4093. <https://doi.org/10.1093/jxb/erm267>.
40. Feng W, Hopper AK. 2002. A Los1p-independent pathway for nuclear export of intronless tRNAs in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* 99:5412–5417. <https://doi.org/10.1073/pnas.082682699>.
41. Nagaike T, Suzuki T, Tomari Y, Takemoto-Hori C, Negayama F, Watanabe K, Ueda T. 2001. Identification and characterization of mammalian mitochondrial tRNA nucleotidyltransferases. *J Biol Chem* 276:40041–40049. <https://doi.org/10.1074/jbc.M106202200>.
42. Findeiss S, Langenberger D, Stadler PF, Hoffmann S. 2011. Traces of post-transcriptional RNA modifications in deep sequencing data. *Biol Chem* 392:305–313. <https://doi.org/10.1515/BC.2011.043>.
43. Hansen AK, Moran NA. 2012. Altered tRNA characteristics and 3' maturation in bacterial symbionts with reduced genomes. *Nucleic Acids Res* 40:7870–7884. <https://doi.org/10.1093/nar/gks503>.
44. Iida K, Jin H, Zhu J-K. 2009. Bioinformatics analysis suggests base modifications of tRNAs and miRNAs in *Arabidopsis thaliana*. *BMC Genomics* 10:155. <https://doi.org/10.1186/1471-2164-10-155>.
45. Mao M, Yang X, Bennett GM. 2018. Evolution of host support for two ancient bacterial symbionts with differentially degraded genomes in a leafhopper host. *Proc Natl Acad Sci U S A* 115:E11691–E11700. <https://doi.org/10.1073/pnas.1811932115>.
46. Luan J-B, Chen W, Hasegawa DK, Simmons AM, Wintermantel WM, Ling K-S, Fei Z, Liu S-S, Douglas AE. 2015. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol* 7:2635–2647. <https://doi.org/10.1093/gbe/evv170>.
47. Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA. 2014. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol Biol Evol* 31:857–871. <https://doi.org/10.1093/molbev/msu004>.
48. Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima S. 2014. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Curr Biol* 24:R640–R641. <https://doi.org/10.1016/j.cub.2014.06.038>.
49. Torrent M, Chalancon G, de Groot NS, Wuster A, Babu MM. 2018. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. *Sci Signal* 11:eaat6409. <https://doi.org/10.1126/scisignal.aat6409>.
50. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35:849–857. [https://doi.org/10.1016/0092-8674\(83\)90117-4](https://doi.org/10.1016/0092-8674(83)90117-4).
51. Łukasik P, Nazario K, Van Leuven JT, Campbell MA, Meyer M, Michalik A, Pessacq P, Simon C, Veloso C, McCutcheon JP. 2018. Multiple origins of interdependent endosymbiotic complexes in a genus of cicadas. *Proc Natl Acad Sci U S A* 115:E226–E235. <https://doi.org/10.1073/pnas.1712321115>.
52. Yue D, Maizels N, Weiner AM. 1996. CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae*. *RNA* 2:895–908.
53. Phizicky EM, Hopper AK. 2010. tRNA biology charges to the front. *Genes Dev* 24:1832–1860. <https://doi.org/10.1101/gad.1956510>.
54. Burger G, Gray MW, Forget L, Lang BF. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout Jakobid protists. *Genome Biol Evol* 5:418–438. <https://doi.org/10.1093/gbe/evt008>.
55. Hancock K, Hajduk SL. 1990. The mitochondrial tRNAs of *Trypanosoma brucei* are nuclear encoded. *J Biol Chem* 265:19208–19215.
56. Barbrook AC, Santucci N, Plenderleith LJ, Hiller RG, Howe CJ. 2006.

- Comparative analysis of dinoflagellate chloroplast genomes reveals rRNA and tRNA genes. *BMC Genomics* 7:297. <https://doi.org/10.1186/1471-2164-7-297>.
57. Bock R. 2007. Structure, function, and inheritance of plastid genomes, p 29–63. *In* Bock R (ed), *Cell and molecular biology of plastids*. Springer, Berlin, Germany.
 58. Delannoy E, Fujii S, Des Francs-Small CC, Brundrett M, Small I. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Mol Biol Evol* 28:2077–2086. <https://doi.org/10.1093/molbev/msr028>.
 59. Lohan AJ, Wolfe KH. 1998. A subset of conserved tRNA genes in plastid DNA of nongreen plants. *Genetics* 150:425–433.
 60. Alfonzo JD, Söll D. 2009. Mitochondrial tRNA import—the challenge to understand has just begun. *Biol Chem* 390:717–722. <https://doi.org/10.1515/BC.2009.101>.
 61. Gray MW. 2012. Mitochondrial evolution. *Cold Spring Harb Perspect Biol* 4:a011403. <https://doi.org/10.1101/cshperspect.a011403>.
 62. Salinas T, Duchêne A-M, Maréchal-Drouard L. 2008. Recent advances in tRNA mitochondrial import. *Trends Biochem Sci* 33:320–329. <https://doi.org/10.1016/j.tibs.2008.04.010>.
 63. Duchêne A-M, Pujol C, Maréchal-Drouard L. 2009. Import of tRNAs and aminoacyl-tRNA synthetases into mitochondria. *Curr Genet* 55:1–18. <https://doi.org/10.1007/s00294-008-0223-9>.
 64. Salinas-Giegé T, Giegé R, Giegé P. 2015. tRNA biology in mitochondria. *Int J Mol Sci* 16:4518–4559. <https://doi.org/10.3390/ijms16034518>.
 65. Suzuki T, Nagao A, Suzuki T. 2011. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. *Annu Rev Genet* 45: 299–329. <https://doi.org/10.1146/annurev-genet-110410-132531>.
 66. Esseiva AC, Naguleswaran A, Hemphill A, Schneider A. 2004. Mitochondrial tRNA import in *Toxoplasma gondii*. *J Biol Chem* 279:42363–42368. <https://doi.org/10.1074/jbc.M404519200>.
 67. Duchêne A-M, Maréchal-Drouard L. 2001. The chloroplast-derived trnW and trnM-e genes are not expressed in *Arabidopsis* mitochondria. *Biochem Biophys Res Commun* 285:1213–1216. <https://doi.org/10.1006/bbrc.2001.5303>.
 68. Archibald JM. 2015. Endosymbiosis and eukaryotic cell evolution. *Curr Biol* 25:R911–R921. <https://doi.org/10.1016/j.cub.2015.07.055>.
 69. Embley TM, Martin W. 2006. Eukaryotic evolution, changes and challenges. *Nature* 440:623–630. <https://doi.org/10.1038/nature04546>.
 70. Kiers ET, West SA. 2015. Evolving new organisms via symbiosis. *Science* 348:392–394. <https://doi.org/10.1126/science.aaa9605>.
 71. Matsuura Y, Moriyama M, Łukasik P, Vanderpool D, Tanahashi M, Meng X-Y, McCutcheon JP, Fukatsu T. 2018. Recurrent symbiont recruitment from fungal parasites in cicadas. *Proc Natl Acad Sci U S A* 115: E5970–E5979. <https://doi.org/10.1073/pnas.1803245115>.
 72. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
 73. Klimov PB, O'Connor BM. 2009. Improved tRNA prediction in the American house dust mite reveals widespread occurrence of extremely short minimal tRNAs in acariform mites. *BMC Genomics* 10:598. <https://doi.org/10.1186/1471-2164-10-598>.
 74. Langmead B, Trapnell C, Pop M, Salzberg S. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25. <https://doi.org/10.1186/gb-2009-10-3-r25>.
 75. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
 76. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11–16. <https://doi.org/10.1093/nar/gkh152>.
 77. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.0955>.
 78. Huang X, Madan A. 1999. CAP3: a DNA sequence assembly program. *Genome Res* 9:868–877. <https://doi.org/10.1101/gr.9.9.868>.
 79. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>.
 80. Kazantsev AV, Pace NR. 2006. Bacterial RNase P: a new view of an ancient enzyme. *Nat Rev Microbiol* 4:729–740. <https://doi.org/10.1038/nrmicro1491>.
 81. Duchêne A-M, Giritch A, Hoffmann B, Cognat V, Lancelin D, Peeters NM, Zaepfel M, Maréchal-Drouard L, Small ID. 2005. Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 102:16484–16489. <https://doi.org/10.1073/pnas.0504682102>.