bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

A mitochondrion-like ribosome in an insect endosymbiont

Israel S. Fernández[1,2*], Jessica M. Waren[3,4], Justus Nwachukwu[3,4], Dewight Williams[5], Amaia González-Magaña[1,6], Noa Spencer[4], John P. McCutcheon[3,4*]

Understanding how the mitochondrion evolved from a bacterial endosymbiont to an organelle is obscured by time and a lack of comparative examples. It is clear that mitochondrial ribosomes are derived from bacterial ribosomes, but it is unclear whether the baroque and idiosyncratic structural changes seen in mitochondrial ribosomes are specifically driven by mitochondrial evolution or simply reflect the syndrome of being a long-term endosymbiont. Here we show that two of the most characteristic features of mitochondrial ribosomes—the acquisition of host proteins and alterations to specific structural regions—are shared by *Hodgkinia*, an ancient bacterial endosymbiont of cicadas. These results imply that many of the striking changes observed in mitochondrial ribosomes result from their common ancestry as bacterial endosymbionts rather than their uniqueness as mitochondria.

The 3 RNAs and 54 proteins of the bacterial ribosome have changed little over >3 billion years of evolution, leading to ribosome structures that are remarkably similar across a huge diversity of bacteria (*1*). While the ribosome architectures responsible for core functions such as mRNA decoding and peptidyl transferase are highly conserved in not only bacterial ribosomes but in all ribosomes reported to date, many mitochondrial ribosomes (mitoribosomes) have experienced substantial change and diversification at sites peripheral to these core functions (*2*–*5*). Mitoribosomes have lost ancestral bacterial ribosomal proteins (bRPs), contracted or expanded their ribosomal RNA (rRNA), expanded lengths of bRPs compared their homologs in bacteria, and acquired numerous new proteins of eukaryotic ancestry that are imported from the host (mRPs) (*6*). Two regions of the mitoribosomal large subunit (LSU) are repeatedly found to be altered compared to bacterial ribosomes: the central protuberance (CP) and the nascent polypeptide exit tunnel (ET). At the core of the CP in ribosomes is the highly conserved 5S rRNA, which in mitoribosomes has been reduced and modified in green algae (*4*, *7*), lost and replaced by a tRNA in animals (*8*), and lost and replaced by numerous new eukaryotic proteins in trypanosomes (*9*). Similarly, in yeast mitoribosomes, new eukaryotic proteins have been recruited to the ET which, in combination with expansions in RPs and rRNA, have moved the normally conserved ET about 35 Å away from its canonical location in bacteria (*10*). The ET in trypanosome mitoribosomes has also undergone significant structural change, with an alternative ET path formed by protein extensions of bRPs and the recruitment of new mRPs (*9*).

Many cellular and genetic features once thought to be unique to mitochondria and plastids have now been identified in bacterial endosymbionts of insects and protists (*11*, *12*). Extreme genome reduction (*13*), biochemical and genomic complementarity (*14*, *15*), and import of host proteins into endosymbiont cells (*14*, *16*, *17*) have been described for several systems containing stably associated endosymbiotic bacteria. Endosymbionts tend to lose more genes the longer they spend in a host-restricted state, but genes related to translation and the ribosome itself remain the most evolutionarily conserved and are the least susceptible to loss as genome reduction progresses (*12*, *18*–*20*). Some of the most reduced endosymbiotic genomes, however, have lost multiple tRNAs and aminoacyl-tRNA synthetase genes, along with up to 20 of the 54 bRPs (*12*, *19*–*22*). The cicada endosymbiont *Candidatus* Hodgkinia cicadicola (hereafter *Hodgkinia*) is one such example. Its tiny 144 kb genome is predicted to have lost genes for several tRNAs and aminoacyl-tRNA synthetases (*23*), along with 13 bRPs (*21*). Here, using a combination of genomic, transcriptomic, and high-resolution electron cryo-microscopy (cryoEM) data, we show that *Hodgkinia* ribosome evolution parallels mitoribosome evolution despite having an independent evolutionary origin.

Preparation of ribosomes from cicada endosymbionts and structure determination. In the summer of 2025 we captured 2 male *Diceroprocta semicincta* cicadas in the Phoenix area and dissected out their bacteriomes, the specialized tissue containing *Hodgkinia* and its co-symbiont, *Candidatus* Karelsulcia muelleri (hereafter *Karelsulcia*) (**Fig. 1A**) (*24*). Cicada bacteriome tissue contains a mix of ribosomes from four sources: the cicada cytosol, the cicada mitochondria, and ribosomes from both endosymbionts, *Hodgkinia* and *Karelsulcia*. We miniaturized the RAPPL protocol (*25*) to purify ribosomes from the small amount of bacteriome tissue we were able to isolate from these two insects, which produced enough material to plunge-freeze four cryoEM grids. One of these grids was of sufficient quality in terms of ice thickness and particle distribution for data collection (**Fig. 1A**).

[1]Instituto Biofisika (UPV/EHU, CSIC), University of the Basque Country; Leioa, Spain
[2]Ikerbasque, Basque Foundation for Science; Bilbao, Spain
[3]Howard Hughes Medical Institute; Chevy Chase, MD, 20815, USA
[4]Biodesign Center for Mechanisms of Evolution and School of Life Sciences, Arizona State University; Tempe, AZ, 85287, USA
[5]John M. Cowley Center for High Resolution Electron Microscopy, Arizona State University; Tempe, Arizona 85287, USA
[6]Department of Biochemistry and Molecular Biology, University of the Basque Country (EHU/UPV), Spain
[*]Corresponding authors: israel.s.fernandez@gmail.com and john.mccutcheon@asu.edu

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

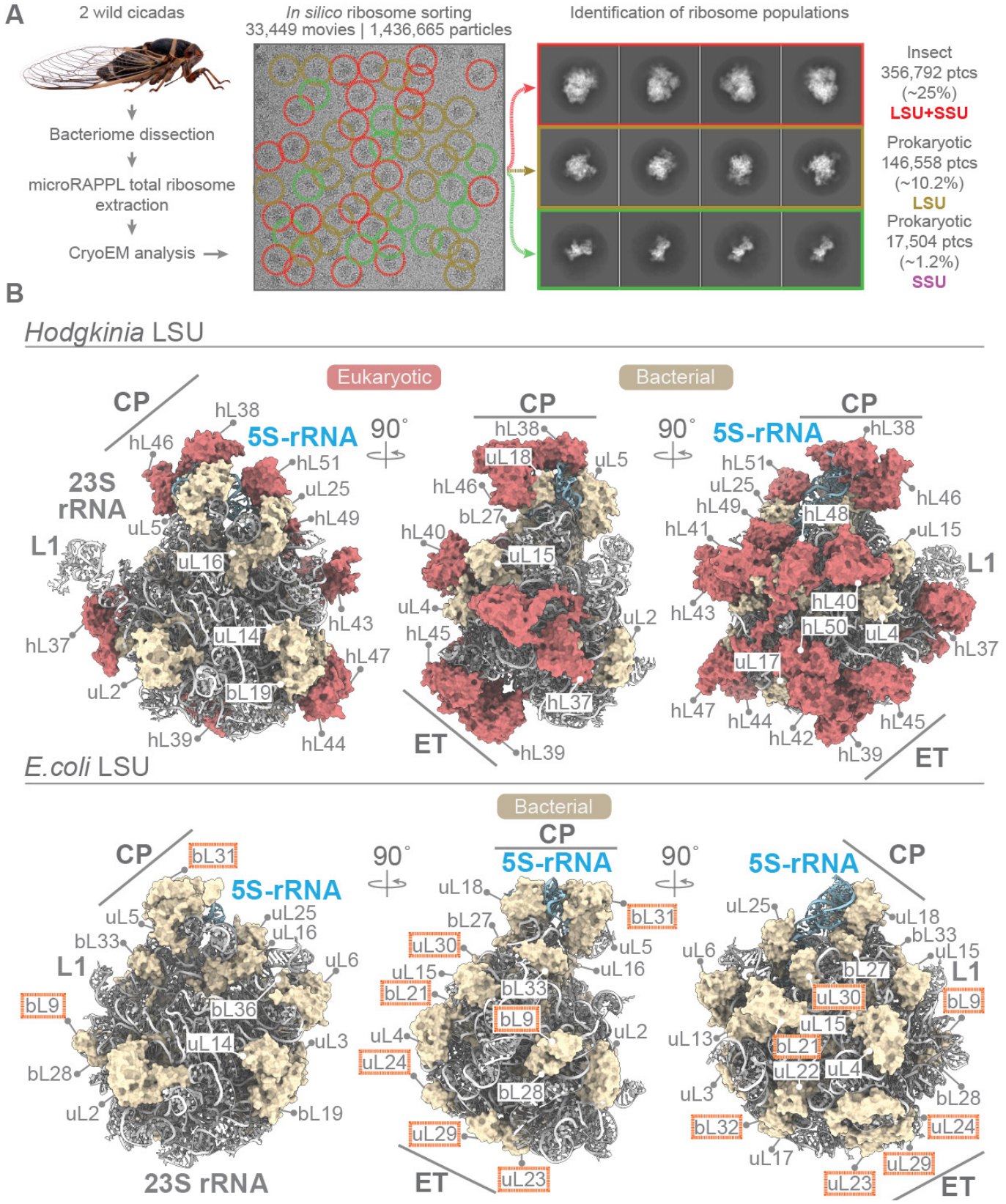


Fig. 1. Overview of the experimental strategy and the *Hodgkinia* LSU structure compared to *E. coli*. (**A**) Schematics of sample preparation (left) and cryoEM analysis showing a representative motion corrected image (center) with ribosomes labelled according to their origin, with eukaryotic ribosomes from the cicada host in red, bacterial large subunits in yellow and bacterial small subunits in green. The final reference-free 2D class averages including information from all particles on the dataset are shown on the right side of (A). (**B**) The top part of the figure shows orthogonal views of the *Hodgkinia* LSU with 23S rRNA depicted as grey cartoons, 5S rRNA as blue cartoons, ribosomal proteins of bacterial ancestry as yellow molecular surfaces, and new ribosomal proteins from the cicada genome in surface representation colored red. For comparison, the bottom part of the figure shows corresponding orthogonal views of the *E. coli* LSU (PDB ID 7k00) with the same visualization settings. *E. coli* ribosomal proteins framed in orange boxes are those missing from the *Hodgkinia* LSU and genome.

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

As expected, *in silico* ribosome sorting using empirical Bayesian methods (*26*) revealed a mixture of ribosome populations. While fully assembled host cytosolic complexes (80S ribosomes) dominated the dataset, minor populations of bacterial-like classes were visible in the 2D class-averages (**Fig. 1A**). These classes exhibited features in length, diameter and overall shape reminiscent of the bacterial large subunit (LSU, 50S in bacteria) and the ribosomal small subunit (SSU, 30S in bacteria). Fully assembled bacterial-like ribosomes (70S) were not identified in the dataset. We suspect that our miniaturization of the RAPPL protocol promoted bacterial ribosomes to disassociate into LSU and SSU but allowed the insect ribosomes, which have a larger inter-subunit surface, to remain intact.

The 2D classes corresponding to bacterial-like LSU (~146,500 particles) produced a 3D map with nominal overall resolution of ~3 Å (**Fig. 1B**, **fig. S1, and table S1**). Local resolution varied across the map, exceeding the nominal 3 Å value at the core and declining to 3.5–4 Å toward the periphery. We also identified a bacterial-like SSU approximately one order of magnitude less abundant than bacterial-like LSU (~17,500 particles). The reduced number of particles combined with severe preferential orientation on the grid limited the attainable resolution in the 3D reconstruction of the SSU, producing a reconstruction in the ~12 Å range (**fig. S2**).

Predicted losses of bRPs and the identification of the LSU map. Because the cicada species in this study has two bacterial endosymbionts, *Hodgkinia* and *Karelsulcia*, along with its own mitochondria, the reconstructed bacterial LSU particle could have originated from either bacterium, the mitochondrion, or a mixture of the three. We rejected the mixture hypothesis given the quality of the map in terms of resolution and homogeneity (**fig. S1**), and we rejected the mitochondrial hypothesis due to the size of the particle. In order to identify whether this LSU reconstruction belonged to *Hodgkinia* or *Karelsulcia*, we followed a two-pronged approach rooted in the extreme conservation of the ribosomal peptidyl-transferase center (PTC) along with the predicted differences in the sets of ribosomal proteins in the *Hodgkinia* and *Karelsulcia* genomes (*21*). We first fitted an atomic model of the LSU rRNA from *E. coli* (*27*) into our map, which clearly resolved purine from pyrimidine bases and allowed an unambiguous and precise fitting of the *E. coli* PTC. Using the *E. coli* PTC as a structural anchor, we superposed the rest of the *E. coli* LSU onto our map. Most *E. coli* ribosomal proteins fit the density of the map well, but several had no corresponding map density, even at low contour values, suggesting the loss of multiple bRPs from this LSU (**fig. S3B**).

Genome annotation for *Karelsulcia* predicts that only uL29 and uL30 are missing from its LSU compared to *E. coli*, while *Hodgkinia* is predicted to be missing 10 bRPs from the LSU (uL1, bL9, bL21, uL23, uL24, uL29, uL30, bL31, bL32, bL34) (*21*). Our preliminary structural analysis revealed weak or nonexistent density at 9 of the 10 bRPs predicted to be missing in *Hodgkinia* (**fig. S3B**). These structural patterns suggested that this LSU map was from *Hodgkinia*, not *Karelsulcia*, which we further verified using characteristic and discriminating amino acid residues in conserved bRPs. We observed robust map density corresponding to bL21 in *E. coli*, which was the sole conflict between our cryoEM and genomic data. Re-analysis of the *Hodgkinia* genome revealed that the prediction of the bL21 gene as missing in the genome (*28*) was due to an annotation error. Overall, these results confirm that this map is from *Hodgkinia*, and that its LSU is missing 9 bRPs.

The *Hodgkinia* LSU structure is a mosaic of bacterial and eukaryotic proteins. After confirming that our LSU map was from *Hodgkinia*, we built and refined the majority of the LSU 23S rRNA, the complete 5S rRNA, and the 24 *Hodgkinia* bRPs that are homologous to the canonical bacterial LSU set. After the removal of all map regions where we could find direct one-to-one mapping to homologous structures in *E. coli*, the difference map revealed substantial additional densities. These densities were located at the periphery of the LSU, with several clusters of density at the CP and ET (**fig. S3A**). In order to identify these unexpected regions of density, we used ModelAngelo (*29*) to generate structure-based profile hidden Markov models (HMMs) from our *Hodgkinia* LSU map and searched these HMMs against a predicted proteomes from the *Hodgkinia* and *Karelsulcia* genomes (*24*, *28*) along with a set of host proteins predicted from a previous cicada transcriptome study (*23*) using the hmmer (*30*) function in ModelAngelo.

Outside of the expected *Hodgkinia* bRPs, we found no other high confidence HMM hits in predicted *Hodgkinia* or *Karelsulcia* proteomes. Remarkably, we found numerous high-confidence HHM hits to proteins predicted from our cicada transcriptome (*23*). However, because this transcriptome was produced using short reads and we had no reference cicada genome, the overall quality of the transcriptome was low and we were unable to disambiguate true splice variants from transcriptome misassembly. We therefore generated a PacBio cicada genome (4.9 Gb; 1,396 contigs; N50 of 13.2 Mb; ~18X coverage) and a PacBio IsoSeq cicada transcriptome (30.5 M reads; avg. length 1,487 nts) to provide definitive references for our ModelAngelo structure-based search. These genomic and transcriptomic data confirmed our high-confidence HHM hits to predicted cicada proteins, and allowed us to unambiguously align these protein-producing transcripts onto the cicada nuclear genome (**fig. S4**). Using a combination of Alphafold (*31*) predictions and manual building and refinement, we placed 15 nuclear-encoded cicada proteins into the regions of unexplained density on the *Hodgkinia* LSU map with high confidence (**Fig. 1A, table S2, data S1**). Following the naming convention used for mitoribosomes, we have named these proteins hL37-hL51. The ‘h’ prefix reflects the presence of these proteins on the *Hodgkinia* ribosome, and the numbering starting at 37 reflects our assumption that the ribosome of *Hodgkinia*’s

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

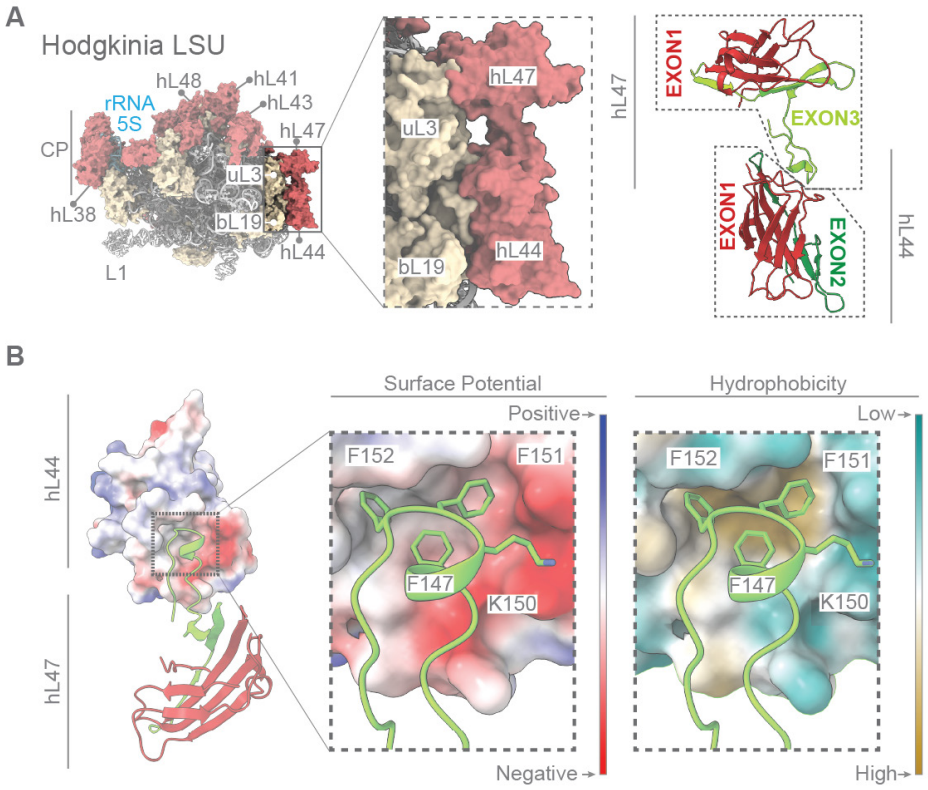


Fig. 2. Interactions between hL44 and hL47, two cicada proteins resulting from alternative splicing of a single cicada gene. (**A**) Left side shows an overall view of the *Hodgkinia* LSU with the area where h47 and h44 are located highlighted. On the middle and right shows zoomed-in views of h44/h47 in two representations: as molecular surfaces (center) and as cartoons with the secondary structure elements colored according to their genomic origin in exon 1, 2 (for hL44) or 2' (for hL47; the genomic structure of this gene is detailed in **fig. S5**). (**B**) The interactions between exon 2' of hL47 (shown as a cartoon) and hL44 (shown as surface potential or hydrophobicity surfaces). The h44/h47 interaction is driven by a cluster of hydrophobic phenylalanine residues present on exon 2' of h47 that docks into a hydrophobic cavity on h44 (right), along with an ionic interaction between a lysine in hL47 and a negatively charged region of hL44 (middle).

alphaproteobacterial ancestor had the complete set of bacterial LSU bRPs, the numbering of which ends at bL36 (*32*).

The 15 host proteins are of various fold types and predicted functions. Unlike mRPs of eukaryotic ancestry that have accumulated on mitoribosomes, which tend to be from related families of alpha-helical rich proteins such as pentatricopeptide repeat proteins (*6*, *7*), the cicada hRPs we find in the *Hodgkinia* LSU structure are a diverse mixture of fold types and have a variety of predicted functions seemingly unrelated to ribosome biology (**table S2**). The most common functional annotation is found in hL44, hL47, hL41, and hL51, all of which are members of the beta-1,3-glucan binding protein family CBM39 (*33*, *34*). All but one hRP have predicted N-terminal signal peptides (**table S2**), suggesting that they are trafficked through the cicada secretory pathway as part of their targeting to *Hodgkinia* cells.

Two of these proteins, hL44 and hL47, are alternatively spliced versions of the same gene that physically interact with each other on the *Hodgkinia* LSU (**Fig. 2 and fig. S5**). The genomic locus encoding these two splice variants is complex, with seven different splice variants all sharing the same first exon but different second exons, which apparently produce six different but related beta-1,3-glucan binding proteins (**fig. S5**). The introns of this locus are large, with the second exons of hL47 and hL48 154 kb and 158 kb downstream from their shared exon 1, respectively. hL47 interacts with bacterial ribosomal protein uL3, while hL44 interacts with both uL3 and bL19 (**Fig. 2A**). A C-terminal loop of hL47 (encoded by its unique exon 2) extends into a pocket of hL44 (encoded primarily by the shared exon 1) and makes extensive hydrophobic and ionic interactions with its neighboring protein isoform (**Fig. 2B**).

Three cicada proteins remodel the *Hodgkinia* Central Protuberance. The CP is an integral part of the LSU, directly interacting with the head of the small subunit and enabling the vectorial movement of tRNAs and mRNA to advance protein synthesis (*35*). The bacterial CP is composed of multiple bRPs along with the highly conserved 5S rRNA. The *Hodgkinia* CP is modified by the loss of bL31 and addition of hL38, hL46, and hL51, all of which interact with the 5S RNA in distinct ways. While hL38 almost exclusively interacts with the 5S rRNA via stacking interactions (**Fig. 3A, 1-2**), hL46 positions a short alpha-helix deep into the 5S rRNA, stabilizing its 5'-terminal end (**Fig. 3A, 4**). It is worth noting that the N-terminal R26 residue is exactly the predicted sequence resulting from cleavage of the hL46 signal peptide (**Fig. 3A, 4**; **table S2**), confirming the computational signal sequence prediction of at least this protein. Finally, multiple cationic resides of hL51 establishes ionic interactions to the phosphate groups of the 5S rRNA backbone (**Fig. 3A, 3**). Overall, the configuration of the *Hodgkinia* CP features a protein-rich environment compared to a normal bacterial CP, and results in a 5S rRNA almost entirely shielded from solvent (**Fig. 3B**, left). While the addition of host proteins to mitoribosomes that still retain 5S rRNA has been reported, these proteins tend to bind to bRPs, not 5S rRNA (*5*). The burying of 5S rRNA in the *Hodgkinia* ribosome is perhaps most similar to the structural changes in the CP of eukaryotic cytosolic ribosomes (**Fig. 3B**, right).

The addition of cicada proteins radically alters the outbound trajectory of the nascent peptide. The ET

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

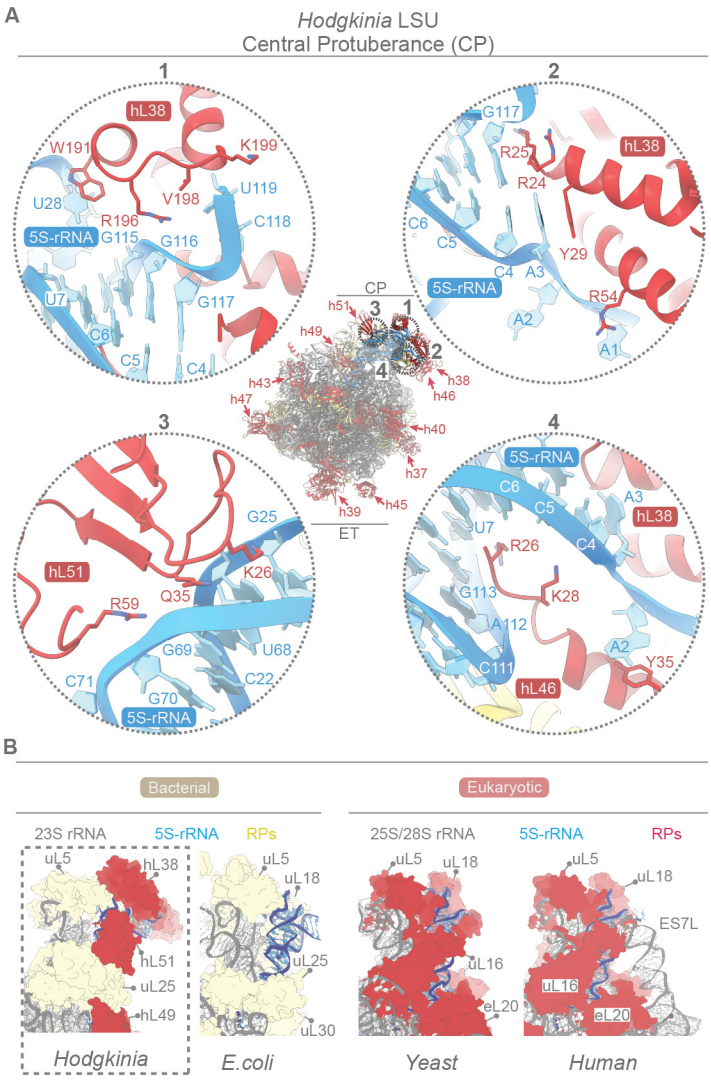


Fig. 3. The central protuberance (CP) of the *Hodgkinia* LSU is highly modified. (**A**) Center, an overview of the *Hodgkinia* LSU with cicada proteins labeled in red. The CP and the exit tunnel (ET) areas are indicated. Regions 1, 2, 3, and 4 are zoomed-in views of the CP with cicada proteins colored red, *Hodgkinia* proteins of bacterial origin yellow, and the 5S rRNA blue. hL38 interacts with the 5S rRNA mainly by stacking interactions (1, 2) while hL51 (3) and hL46 (4) establish multiple ionic interactions with the 5S rRNA. (**B**) Comparative view of the CP from *Hodgkinia* (left, boxed), *E. coli*, and the *S. cerviseae* and human cytosolic structures (right). The newly acquired cicada proteins almost fully shield the 5S rRNA from the solvent (left), in contrast to a canonical bacterial configuration, and more similar to the extensively covered 5S rRNA of eukaryotic cytosolic ribosomes (right).

guides newly synthesized peptides towards the exterior of the LSU, and is thought to prevent premature collapse of the nascent peptide into non-native fold states (*36*). These functions are accomplished by architectural constrains imposed by bRPs and LSU rRNA residues upholstering the tunnel. The first half of the tunnel is straight and narrow, while towards the end it widens into a vestibule to allow the formation of secondary structure elements. Finally, the ET opens to a wide cavity where chaperones and other protein modifying enzymes can act on the nascent peptide. In most bacteria, the ET is comprised of LSU rRNA and six RPs: uL4, uL22, uL23, uL24, uL29, and bL32 (*36*). Of these six canonical ET-related ribosomal proteins, *Hodgkinia* retains only uL4 and uL22.

The *Hodgkinia* ET is most dramatically modified, however, by the addition of three cicada proteins (hL39, hL42 and hL50) at the end of the ET, which significantly alters the canonical bacterial path of the nascent peptide towards the solvent (**Fig. 4A-C**). Cicada protein hL39 directly blocks the trajectory of the exiting nascent chain (**Fig. 4A, right**). Cicada proteins hL42 and hL50 are also located near the ET, but play a peripheral role to orient and support hL39 (**Fig. 4B**). We modeled the trajectory of the ribosome nascent chain (RNC) in our *Hodgkinia* map using a cryoEM reconstruction from *E. coli* ribosomes bound to a P-site tRNA (PDB-ID 7ZP8, **Fig. 3D-E**). This analysis revealed that the position of cicada protein h39 imposes an almost 90º turn of the RNC trajectory compared to *E. coli* (**Fig. 3D**).

The evolutionary context of the *Hodgkinia* ribosome. Comparative genomics suggests that the mitoribosome of last eukaryotic common ancestor (LECA) retained 53 of the 54 standard alphaproteobacterial bRPs, having lost only bS20 (*37*, *38*). Comparative genomics also suggests that the LECA mitoribosome had already acquired 19 new mRPs (9 on the LSU, 10 on the SSU) before the divergence of the eukaryotic lineage occurred (*38*, *39*). With 9 bRP losses and 15 hRP gains, the *Hodgkinia* LSU structure suggests that it is already more chimeric—that is, it has lost more bRPs and gained more eukaryotic hRPs—than the LSU of the inferred LECA mitoribosome.

The relative order of mitoribosome rRNA reduction and expansion in relation to bRP loss and mRP accumulation has been debated (*5*, *6*, *40*–*43*), but it now seems clear that the accumulation of mRPs preceded most of the length variation observed in mitoribosome rRNA (*5*, *43*, *44*). Our results suggest that accumulation of host protein before alterations in rRNA length is a general feature of endosymbiotic bacteria and is not an evolutionary path exclusive to mitochondria. The length of *Hodgkinia*'s LSU is 2,767 nts, only 26 nts shorter and well within a standard deviation of the average length of a non-mitochondrial alphaproteobacterial LSU in the SILVA database (mean = 2,793 nts; stdev. = 161; n = 6,955) (*45*). While the length of LSU from *Hodgkinia*'s free-living ancestor is unknown, it seems unlikely that the accumulation of 15 nuclear-encoded hRPs was an adaptive response to a trivial change in rRNA length.

Why do ribosomes from endosymbiotic bacteria lose bRPs and accumulate host proteins? We have shown that the bRPs lost from the *Hodgkinia* LSU are lost without direct functional replacement: the 15 hRPs that have become part of the *Hodgkinia* LSU ribosome do not patch holes created by missing bRPs. But what role, then—if any—might these new host proteins play? It could be that some hRPs stabilize the ribosome at positions distant to the lost bRPs in an allosteric fashion by buffering the structural instability caused by bRP protein loss. It could be that the addition of hRPs at the *Hodgkinia* ET provides some sort of chaperone-like function for the wildly divergent (and likely unstable) *Hodgkinia* proteins as they exit the ribosome. But it also could be that some of these host proteins accumulate

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

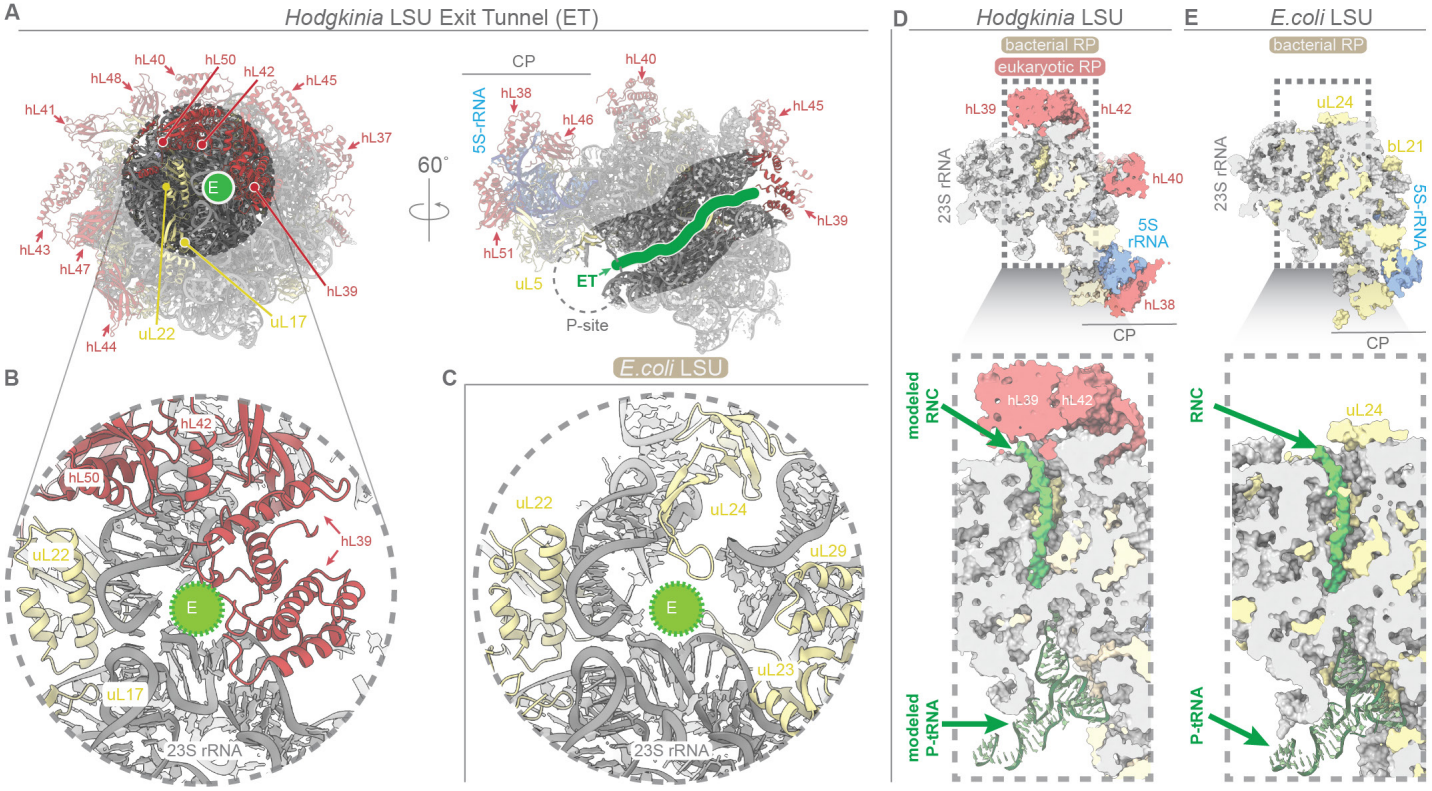


Fig. 4. **The exit tunnel (ET) of the *Hodgkinia* LSU is radically altered by a host cicada protein.** (**A**) Two views of the *Hodgkinia* LSU, with a view looking straight into the ET (left) and one cut away for a side-on view (right). Ribosomal proteins from the cicada genome are labeled and colored red and ribosomal proteins of bacterial origin are colored yellow. The trajectory of the Ribosome Nascent Chain (RNC) is highlighted in green. (**B**) Zoomed-in view of the end of the ET, where cicada proteins hL39, hL42 and hL50 are located. (**C**) Equivalent view as in (B) from *E.coli* (PDB ID 7ZP8). Comparisons of the trajectory of the RNC in *Hodgkinia* (**D**) and *E. coli* (**E**), with 23S rRNA colored grey, 5S rRNA blue, ribosomal proteins of bacterial origin yellow, ribosomal proteins from the cicada red, and the peptidyl-tRNA in green. Cicada protein hL39 imposes a severe almost 90-degree turn of the RNC to access the solvent (D) compared to a typical bacterial ET (E).

for more subtle reasons, the origin of which might not initially be related to ribosome function at all. Several authors have suggested that constructive neutral evolution (CNE)-like mechanisms (*46*) might be responsible for at least some of the changes seen in mitoribosomes, and for the accumulation of complexity in the ribosome in general (*42–44*, *47–49*). Under CNE models, hRPs accumulate through transient interactions that serve no initial purpose, but become fixed into place by mutations in other parts of the ribosome that are only allowed in the presence of the new hRP. Because these 'hidden' mutations are deleterious in the absence of the nascent hRP (*50*), the interaction then becomes required for normal ribosome function (*48*).

All extant mitoribosomes originated from the same ancestral structure and are therefore constrained to the path taken at this ancient origin. This makes inference about the relative timing, and the possible reasons, for changes to mitoribosomes difficult because few comparative examples exist. Ribosomes from chloroplasts, which is the endosymbiont next oldest compared with mitochondria, also show alterations to the ET and accumulation of some eukaryotic proteins, but in general are much bacteria-like in their structure (*51*, *52*). It should be noted, however, that recent results suggest that this conservation of bacterial structure in chloroplast ribosomes might be due in part to a lack of sampling across archaeplastidal organisms (*53*). The large diversity of insect endosymbionts in various stages of genome reduction and host integration (*12*) might provide an avenue to resolve open questions related to when, how, and why new host proteins accumulate on endosymbiont (and organelle) ribosomes, and why some structural regions are more or less tolerant of change. We also note that the presence of 15 host proteins on the *Hodgkinia* LSU strongly suggest that large numbers of host proteins may be imported into *Hodgkinia* cells, a phenomenon that was suggested by previous work in other insect systems but has only been proven in protist endosymbionts (*11*).

The mitochondrion is the oldest endosymbiont known to science, and its origin was a key event in the evolution of eukaryotic life. But its antiquity, combined with a lack of comparative examples, has made understanding which features of mitochondria have arisen because of specific mitochondrial functions, and which have arisen as byproducts of the endosymbiotic process, difficult. Our results suggest that many of the striking ribosomal features that seemed to be specific to mitoribosomes are more likely to have arisen simply because the mitochondrion used to be an endosymbiont.

MATERIALS AND METHODS

Ribosome purification. Two *Diceroprocta semicincta* cicadas were captured in Glendale, Arizona on 13 September 2025 and their abdomens were separated from their heads and thoraces. The heads and thoraces were placed in 70% ethanol at stored at -20°C for genome

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

sequencing (see below). The bacteriomes were dissected from the abdomens in sterile phosphate-buffered saline, the buffer was removed, and the bacteriome tissue was flash frozen in liquid nitrogen and stored at -70°C. The ribosomes were purified using a slightly modified RAPPL protocol (*25*): frozen bacteriomes were thawed on ice for 2 min, resuspended in 500 ml of 4°C RAPPL lysis buffer, transferred to a 2 mL T-dounce and homogenized 15 times, and the mixture was then centrifuged at 20,000 X g for 20 min at 4°C. The remaining in-solution lysate was added to 30 ml RAPPL beads prewashed in RAPPL lysis buffer, mixed by gentle swirling, and incubated for 30 min at 4°C with gentle rocking. Beads were isolated in a magnetic stand and the solution was removed. The beads were washed 3X in 500 ml RAPPL wash buffer, and the ribosomes were released into 15 ml of RAPPL elution buffer for 15 mins at 4 °C with gentle rocking.

CryoEM data acquisition. To visualize ribosomes via cryoEM, 3 µL of the RAPPL-purified sample was applied to Quantifoil 2/1, 2 nm ultrathin carbon, 200-mesh gold TEM grids (Ted Pella, Inc., Redding, CA) after they were made hydrophilic by a 15 second glow discharge. Samples were allowed to absorb on the grids for 2 minutes at 100% humidity before plunge freezing using a Vitrobot Mark IV and a liquid nitrogen cooled ethane slurry with 100% humidity at 16°C for 3 seconds blot time. Ribosome visualization was performed using a Krios G2 equipped with a Falcon4i/SelectrisX at 130,000 X nominal magnification, resulting in 27,282 EER movies with a 0.9308 Å sq./pixel size, an electron dose of 25 electrons/Å sq., and an energy filter slit width of 10 eV.

CryoEM image processing. Data processing was integrally performed in Relion-5 (*54*) with ctf estimation in Ctffind4 (*55*). Eer files were motion corrected with the Relion implementation of the MotionCorr algorithm (*56*) using whole-frame alignment and a B-factor of 150 with sums of power spectra every 4e^-/Å^2. We used a FFT window of 1024 and the exhaustive search option turned on for ctf estimation in Ctffind4. We selected a subset of 500 randomly selected images for auto-picking with the Relion LoG picker using maximum diameter of 380Å and minimal of 120Å. We tried 3 different threshold parameters (0, 0.5 and 1) and manually checked the accuracy of the picking. With a small set of particles originating from 500 images with the LoG picker, we performed 2 rounds of 2D classifications (T parameter 3 and 5) and manually selected those class averages that by shape and size looked like ribosomes. This curated small set of ~5,000 particles was used as a training set in Topaz (*57*). A first Topaz model was used to re-pick the set of initial 500 randomly selected images. After another round of two 2D classifications jobs and manual selection of class averages reminiscent of ribosome, we trained a second Topaz model that significantly outperformed the first one in terms of number of particles and alternative views. We used the second Topaz model to pick the whole dataset, which generated an initial set of around 3 million particles. We then performed 5 iterative rounds of 2D classifications and manual picking of the 2D ribosome-looking class averages, subsequently increasing the T parameter at every 2D classification round. We used T=3 for the first round, T=4 for the second, T=5 for the third, T=6 for the fourth, and T=8 for the final fifth round. We did not attempt to separate ribosome populations at this stage, even though it was evident by the shape and details of the 2D class averages that there were multiple and very different ribosome populations in the sample. The final subset of particles selected after 2D classifications was in the range of 700,000. To separate the different ribosome populations we combined the Stochastic Gradient Descent (SGD) algorithm (*58*) and masked 3D classifications without alignment within Relion-5. Using the 700K dataset we launched an Initial Model job with 6 classes and T parameter of 2. The algorithm sorted ~70% of the particles to a class which we identified as the host insect cytoplasmic eukaryotic 80S ribosome. There was a second class of ~20% of the particles that resembled a prokaryotic large subunit and a small class that look like a bacterial small subunit. We selected the particles of the class reminiscent of a bacterial large subunit and performed a 3D auto-refine job that generated a volume of good quality.

At this stage, we re-extracted the particles to a bigger box size (480 pixels) and performed another 3D refinement job that reached an overall resolution of 4 Å. We used Blush regularization for all 3D refinements (*59*). Next, after detailed analysis of the map, we performed a masked classification without alignment (4 classes, T=3). This classification job identified a homogenous subset of particles (53,191 particles) that refined to 3.5 Å resolution. A final iterative routine combining 3D auto-refinements with ctf-refinements (*60*) (defocus per-particle, astigmatism per micrograh) and Bayesian polishing improved the resolution to 3 Å (*61*). For model building we used Coot (*62*) with maps sharpened with Relion and phenix.autosharpen using the local sharpening option (*63*). The final model was sterochemically refined against the sharpened map in two steps: first, using phenix.real_space_refinement with secondary structure restrains (*64*) and after manually fixing multiple errors we performed a second refinement step in Fourier space using Refmac5 (*65*). The model statistics were analyzed with MolProbity (*66*).

Protein identification in the cryoEM map. After establishing that the map was from the *Hodgkinia* LSU, we identified the areas of unexplained density using ModelAngelo run with no input sequences (the build_no_seq flag) followed by an HMM search (the hmm_search flag) against predicted cicada proteins. During most of the course of this work, we did not have access to the cicada genome, so we used a bacteriome-enriched transcriptome from the same cicada species (*Diceroprocta semicincta*) generated in 2010 for a different project (*23*). Putative proteins from these transcriptome data were generated LongOrfs package in TD2 (*67*) with the minimum protein length set to 40 and the requirements for canonical start and stop codons relaxed (the -m 40 --alt-start --all-stopless flags). While the ModelAngelo HHM search was initially performed against the 1.4 M proteins predicted from this

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

method, and these data were used to identify the 15 cicada hRPs on the *Hodgkinia* LSU, we subsequently verified the presence and structure of these genes from a high-quality cicada genome and isoseq transcriptome generated in 2026, as described below.

Protein identifiers were extracted from the best_hits.csv file generated from the ModelAngelo HMM search against the TD2-predicted cicada proteome, and the protein sequences associated with these identifiers were used as input into ModelAngelo in an attempt to place these putative cicada proteins into the map. This process was iterative, as it was immediately clear that some of the cicada proteins were excellent fits in the map while others were less clear, especially in regions of the map that were of lower quality. AlphaFold3 was used to predict the structure of the identified proteins and, in general, the predictions featured a central core of secondary structure-rich elements along with extended N- and C-termini with unstructured tails. We roughly placed well-predicted central core regions of the AlphaFold3 models in the map, using the short but well-defined stretches built by ModelAngelo as a guide. Many of the cicada proteins showed well defined N- and C-termini in the map, usually stabilized by molecular contacts with adjacent ribosomal proteins. These fragments were built manually.

Cicada genome analysis. DNA for genome sequencing was isolated from small amounts of cicada tissue from the samples stored in ethanol at -20°C on 13 September 2025 (the same insects used to solve the *Hodgkinia* LSU structure, as described above). DNA was extracted using the Nanobind PanDNA (PacBio) high molecular weight DNA extraction kit, following the additional “Ethanol-preserved tissues” pretreatment outlined in the Nanobind PanDNA kit instructions. Extracted DNA was sent to SeqCenter (Pittsburgh, PA) for sequencing. Genomic DNA was fragmented to 40-50 kb using the Megaruptor 3 instrument from Diagenode, and sequenced on a PacBio Revio. PacBio sequencing produced 5.6 M reads with an average length of 16 kb for a total output of 88.5 Gb. These reads were assembled using hifiasm (*68*) using the default purge level (-l 2), which produced an assembly of 4.9 Gb in 1,396 contigs with an N50 of 13.2 Mb and an average coverage of ~18X. This 4.9 Gb genome size is in line with other cicada genomes on GenBank (4.45–6.52 Gb; GCA_979322825.1, GCA_980043705.1, GCA_034769895.1), and the BUSCO v.6.0.0 completeness was 98.6% when using the hemiptera_odb10 lineage dataset of 2,510 BUSCOs (*69*), suggesting that the genome assembly was of sufficient quality for gene finding and the verification of hL37-hL51 on the cicada genome.

Transcriptome sequencing and genome annotation. The bacteriomes from two cicadas caught in July 2026 were flash frozen in liquid nitrogen. Samples were thawed on ice and total RNA was extracted with TRIzol using manufacturer’s protocol. Total RNA was frozen at -70 C ans shipped shipped to SeqCenter (Pittsburgh, PA) for IsoSeq v2 sequencing on a PacBio Revio instrument. 22,768,079 reads were generated from two bacteriome libraries. To find reads encoding the 15 cicada proteins identified from the cryoEM map, the IsoSeq reads were translated using TD2.LongOrfs (minimum length set to 50aa), which yielded 78,852,357 candidate peptides. The 15 proteins were searched against these predicted peptides using BLAST+ v2.12.0: blastp -evalue 1e-10 -max_target_seqs 1000000, which identified 7,691 reads as containing the 15 proteins with ≥ 90% of that protein’s length with 5 or fewer amino acid substitutions. For each protein, a consensus sequence was built from the matching reads by taking the majority residue at each position across all read-containing ORFs, aligned in protein coordinate space. These 15 read-derived consensus sequences were aligned to the *D. semicincta* genome using miniprot v0.18: miniprot --gff -G 200k, which defined the CDS coordinates for each locus. Separately, all 22,768,079 IsoSeq reads were mapped to the cicada genome using minimap2 v2.28: minimap2 -ax splice:hq -uf -I 8G -G 500k. For each of the 15 loci, reads were counted if their alignment spanned the full CDS with no splice junctions outside the CDS model. This structural criterion, applied without reference to the protein sequence, identified 8,004 supporting reads across the 15 loci. Some additional reads were recruited in this step due to some protein predictions falling outside the 90% length cutoff described in the previous paragraph.

***Hodgkinia* LSU rRNA length analysis.** LSU rRNA sequence lengths were analyzed using the SILVA_138.2_LSURef_tax_silva.fasta file downloaded from the SILVA database (*45*) on 2 June 2026. Mean, median, and standard deviation of 6,746 alphaproteobacterial LSU sequences were computed after excluding mitochondrial sequences as well as those derived from metagenomic sampling. These analyses were performed using a custom Python script developed with Claude Code (fasta_tax_lengths.py SILVA_138.2_LSURef_tax_silva.fasta -t Alphaproteobacteria -x Mitochondria). The fasta_tax_lengths.py script is available here: doi: 10.5281/zenodo.23423998.

REFERENCES

1. J. C. Bowman, A. S. Petrov, M. Frenkel-Pinter, P. I. Penev, L. D. Williams, Root of the Tree: The Significance, Evolution, and Origins of the Ribosome. *Chem. Rev.* **120**, 4848–4878 (2020).

2. A. Amunts, A. Brown, J. Toots, S. H. W. Scheres, V. Ramakrishnan, The structure of the human mitochondrial ribosome. *Science* **348**, 95–98 (2015).

3. N. Desai, A. Brown, A. Amunts, V. Ramakrishnan, The structure of the yeast mitochondrial ribosome. *Science* **355**, 528–531 (2017).

4. V. Tobiasson, I. Berzina, A. Amunts, Structure of a mitochondrial ribosome with fragmented rRNA in complex with membrane-targeting elements. *Nat Commun* **13**, 6132 (2022).

5. F. Waltz, H. Soufari, A. Bochler, P. Giegé, Y. Hashem, Cryo-EM structure of the RNA-rich plant

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

mitochondrial ribosome. *Nat. Plants* **6**, 377–383 (2020).

6. V. Scaltsoyiannes, N. Corre, F. Waltz, P. Giegé, Types and Functions of Mitoribosome-Specific Ribosomal Proteins across Eukaryotes. *IJMS* **23**, 3474 (2022).

7. F. Waltz, T. Salinas-Giegé, R. Englmeier, H. Meichel, H. Soufari, L. Kuhn, S. Pfeffer, F. Förster, B. D. Engel, P. Giegé, L. Drouard, Y. Hashem, How to build a ribosome from RNA fragments in Chlamydomonas mitochondria. *Nat Commun* **12**, 7176 (2021).

8. A. Brown, A. Amunts, X. Bai, Y. Sugimoto, P. C. Edwards, G. Murshudov, S. H. W. Scheres, V. Ramakrishnan, Structure of the large ribosomal subunit from human mitochondria. *Science* **346**, 718–722 (2014).

9. D. J. F. Ramrath, M. Niemann, M. Leibundgut, P. Bieri, C. Prange, E. K. Horn, A. Leitner, D. Boehringer, A. Schneider, N. Ban, Evolutionary shift toward protein-based architecture in trypanosomal mitochondrial ribosomes. *Science* **362**, eaau7735 (2018).

10. A. Amunts, A. Brown, X.-C. Bai, J. L. Llácer, T. Hussain, P. Emsley, F. Long, G. Murshudov, S. H. W. Scheres, V. Ramakrishnan, Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014).

11. M. E. S. Sørensen, M. L. Stiller, L. Kröninger, E. C. M. Nowack, Protein import into bacterial endosymbionts and evolving organelles. *The FEBS Journal* **292**, 2992–3013 (2025).

12. J. P. McCutcheon, A. I. Garber, N. Spencer, J. M. Warren, How do bacterial endosymbionts work with so few genes? *PLoS Biol* **22**, e3002577 (2024).

13. A. Michalik, D. C. Franco, J. Deng, M. Prus-Frankowska, A. Stroiński, P. Łukasik, Convergent extreme reductive evolution in ancient planthopper symbioses. *Nat Commun* **17**, 2473 (2026).

14. D. C. Bublitz, G. L. Chadwick, J. S. Magyar, K. M. Sandoz, D. M. Brooks, S. Mesnage, M. S. Ladinsky, A. I. Garber, P. J. Bjorkman, V. J. Orphan, J. P. McCutcheon, Peptidoglycan Production by an Insect-Bacterial Mosaic. *Cell* **179**, 703-712.e7 (2019).

15. J. Morales, G. Ehret, G. Poschmann, T. Reinicke, A. K. Maurya, L. Kröninger, D. Zanini, R. Wolters, D. Kalyanaraman, M. Krakovka, M. Bäumers, K. Stühler, E. C. M. Nowack, Host-symbiont interactions in Angomonas deanei include the evolution of a host-derived dynamin ring around the endosymbiont division site. *Current Biology* **33**, 28-40.e7 (2023).

16. A. Nakabachi, K. Ishida, Y. Hongoh, M. Ohkuma, S.-Y. Miyagishima, Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Curr. Biol.* **24**, R640–R641 (2014).

17. A. Singer, G. Poschmann, C. Mühlich, C. Valadez-Cano, S. Hänsch, V. Hüren, S. A. Rensing, K. Stühler, E. C. M. Nowack, Massive Protein Import into the Early-Evolutionary-Stage Photosynthetic Organelle of the Amoeba Paulinella chromatophora. *Curr. Biol.* **27**, 2763-2773.e5 (2017).

18. U.-G. Maier, S. Zauner, C. Woehle, K. Bolte, F. Hempel, J. F. Allen, W. F. Martin, Massively convergent evolution for ribosomal protein gene content in plastid and mitochondrial genomes. *Genome Biol. Evol.* **5**, 2318–2329 (2013).

19. D. D. Nikolaeva, M. S. Gelfand, S. K. Garushyants, Simplification of Ribosomes in Bacteria with Tiny Genomes. *Molecular Biology and Evolution* **38**, 58–66 (2021).

20. M. Y. Galperin, Y. I. Wolf, S. K. Garushyants, R. Vera Alvarez, E. V. Koonin, Nonessential Ribosomal Proteins in Bacteria and Archaea Identified Using Clusters of Orthologous Genes. *J Bacteriol* **203** (2021).

21. J. P. McCutcheon, The bacterial essence of tiny symbiont genomes. *Curr. Opin. Microbiol.* **13**, 73–78 (2010).

22. N. A. Moran, G. M. Bennett, The Tiniest Tiny Genomes. *Annual Review of Microbiology* **68**, 195–215 (2014).

23. J. T. Van Leuven, M. Mao, D. D. Xing, G. M. Bennett, J. P. McCutcheon, Cicada Endosymbionts Have tRNAs That Are Correctly Processed Despite Having Genomes That Do Not Encode All of the tRNA Processing Machinery. *MBio* **10** (2019).

24. J. P. McCutcheon, N. A. Moran, Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biol. Evol.* **2**, 708–718 (2010).

25. J. Erath, D. Kemper, E. Mugo, A. Jacoby, E. Valenzuela, C. F. Jungers, W. L. Beatty, Y. Hashem, M. Jovanovic, S. Djuranovic, S. Pavlovic Djuranovic, A rapid, simple, and economical method for the isolation of ribosomes and translational machinery for structural and functional studies. *Nat Commun* **16**, 7185 (2025).

26. S. H. W. Scheres, "Processing of Structurally Heterogeneous Cryo-EM Data in RELION" in *Methods in Enzymology* (Elsevier, 2016; https://linkinghub.elsevier.com/retrieve/pii/S0076687916300301)vol. 579, pp. 125–157.

27. Z. L. Watson, F. R. Ward, R. Méheust, O. Ad, A. Schepartz, J. F. Banfield, J. H. Cate, Structure of the bacterial ribosome at 2 Å resolution. *eLife* **9**, e60482 (2020).

28. J. P. McCutcheon, B. R. McDonald, N. A. Moran, Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. *PLoS Genet.* (2009).

29. K. Jamali, L. Käll, R. Zhang, A. Brown, D. Kimanius, S. H. W. Scheres, Automated model building and protein identification in cryo-EM maps. *Nature* **628**, 450–457 (2024).

30. S. R. Eddy, Accelerated Profile HMM Searches. *PLoS Comput Biol* **7**, e1002195 (2011).

31. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis,

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

32. N. Ban, R. Beckmann, J. H. Cate, J. D. Dinman, F. Dragon, S. R. Ellis, D. L. Lafontaine, L. Lindahl, A. Liljas, J. M. Lipton, M. A. McAlear, P. B. Moore, H. F. Noller, J. Ortega, V. G. Panse, V. Ramakrishnan, C. M. Spahn, T. A. Steitz, M. Tchorzewski, D. Tollervey, A. J. Warren, J. R. Williamson, D. Wilson, A. Yonath, M. Yusupov, A new system for naming ribosomal proteins. *Current Opinion in Structural Biology* **24**, 165–169 (2014).

33. K. Takahasi, M. Ochiai, M. Horiuchi, H. Kumeta, K. Ogura, M. Ashida, F. Inagaki, Solution structure of the silkworm βGRP/GNBP3 N-terminal domain reveals the mechanism for β-1,3-glucan-specific recognition. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11679–11684 (2009).

34. Y. Mishima, J. Quintin, V. Aimanianda, C. Kellenberger, F. Coste, C. Clavaud, C. Hetru, J. A. Hoffmann, J.-P. Latgé, D. Ferrandon, A. Roussel, The N-terminal Domain of Drosophila Gram-negative Binding Protein 3 (GNBP3) Defines a Novel Family of Fungal Pattern Recognition Receptors. *Journal of Biological Chemistry* **284**, 28687–28697 (2009).

35. J. H. Cate, M. M. Yusupov, G. Z. Yusupova, T. N. Earnest, H. F. Noller, X-ray Crystal Structures of 70S Ribosome Functional Complexes. *Science* **285**, 2095–2104 (1999).

36. D. N. Wilson, R. Beckmann, The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Current Opinion in Structural Biology* **21**, 274–282 (2011).

37. P. Smits, J. A. M. Smeitink, L. P. Van Den Heuvel, M. A. Huynen, T. J. G. Ettema, Reconstructing the evolution of the mitochondrial ribosomal proteome. *Nucleic Acids Research* **35**, 4686–4703 (2007).

38. E. Desmond, C. Brochier-Armanet, P. Forterre, S. Gribaldo, On the last common ancestor and early evolution of eukaryotes: reconstructing the history of mitochondrial ribosomes. *Research in Microbiology* **162**, 53–70 (2011).

39. M. W. Gray, G. Burger, R. Derelle, V. Klimeš, M. M. Leger, M. Sarrasin, C. Vlček, A. J. Roger, M. Eliáš, B. F. Lang, The draft nuclear genome sequence and predicted mitochondrial proteome of Andalucia godoyi, a protist with the most gene-rich and bacteria-like mitochondrial genome. *BMC Biol* **18**, 22 (2020).

40. T. W. O'Brien, Evolution of a protein-rich mitochondrial ribosome: implications for human genetic disease. *Gene* **286**, 73–79 (2002).

41. M. R. Sharma, E. C. Koc, P. P. Datta, T. M. Booth, L. L. Spremulli, R. K. Agrawal, Structure of the Mammalian Mitochondrial Ribosome Reveals an Expanded Functional Role for Its Component Proteins. *Cell* **115**, 97–108 (2003).

42. A. S. Petrov, E. C. Wood, C. R. Bernier, A. M. Norris, A. Brown, A. Amunts, Structural Patching Fosters Divergence of Mitochondrial Ribosomes. *Mol. Biol. Evol.* **36**, 207–219 (2019).

43. M. Valach, J. A. Gonzalez Alcazar, M. Sarrasin, B. F. Lang, M. W. Gray, G. Burger, An Unexpectedly Complex Mitoribosome in *Andalucia godoyi* , a Protist with the Most Bacteria-like Mitochondrial Genome. *Molecular Biology and Evolution* **38**, 788–804 (2021).

44. E. O. Van Der Sluis, H. Bauerschmitt, T. Becker, T. Mielke, J. Frauenfeld, O. Berninghausen, W. Neupert, J. M. Herrmann, R. Beckmann, Parallel Structural Evolution of Mitochondrial Ribosomes and OXPHOS Complexes. *Genome Biology and Evolution* **7**, 1235–1251 (2015).

45. C. Quast, E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, F. O. Glöckner, The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596 (2012).

46. A. Stoltzfus, On the Possibility of Constructive Neutral Evolution. *J Mol Evol* **49**, 169–181 (1999).

47. M. W. Gray, J. Lukeš, J. M. Archibald, P. J. Keeling, W. F. Doolittle, Irremediable Complexity? *Science* **330**, 920–921 (2010).

48. J. Lukeš, J. M. Archibald, P. J. Keeling, W. F. Doolittle, M. W. Gray, How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* **63**, 528–537 (2011).

49. S. A. Muñoz-Gómez, G. Bilolikar, J. G. Wideman, K. Geiler-Samerotte, Constructive Neutral Evolution 20 Years Later. *J Mol Evol* **89**, 172–182 (2021).

50. G. K. A. Hochberg, Y. Liu, E. G. Marklund, B. P. H. Metzger, A. Laganowsky, J. W. Thornton, A hydrophobic ratchet entrenches molecular complexes. *Nature*, doi: 10.1038/s41586-020-3021-2 (2020).

51. A. Perez Boerema, S. Aibara, B. Paul, V. Tobiasson, D. Kimanius, B. O. Forsberg, K. Wallden, E. Lindahl, A. Amunts, Structure of the chloroplast ribosome with chl-RRF and hibernation-promoting factor. *Nature Plants* **4**, 212–217 (2018).

52. P. Bieri, M. Leibundgut, M. Saurer, D. Boehringer, N. Ban, The complete structure of the chloroplast 70S ribosome in complex with translation factor pY. *EMBO J* **36**, 475–486 (2017).

53. F. Waltz, P. A. Lehner, B. D. Engel, Chloroplast-encoded small subunit extensions reshape the Chlamydomonas chlororibosome. *bioRxiv*, doi: doi: https://doi.org/10.64898/2026.02.07.704542 (2026).

54. S. H. W. Scheres, A Bayesian view on cryo-EM structure determination. *J. Mol. Biol.* **415**, 406–418 (2012).

55. A. Rohou, N. Grigorieff, CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *Journal of Structural Biology* **192**, 216–221 (2015).

56. S. Q. Zheng, E. Palovcak, J.-P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331–332 (2017).

57. T. Bepler, A. Morin, M. Rapp, J. Brasch, L. Shapiro, A. J. Noble, B. Berger, Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153–1160 (2019).

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

58. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290–296 (2017).

59. D. Kimanius, K. Jamali, M. E. Wilkinson, S. Lövestam, V. Velazhahan, T. Nakane, S. H. W. Scheres, Data-driven regularization lowers the size barrier of cryo-EM structure determination. *Nat Methods* **21**, 1216–1221 (2024).

60. J. Zivanov, T. Nakane, S. H. W. Scheres, Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in *RELION* -3.1. *IUCrJ* **7**, 253–267 (2020).

61. J. Zivanov, T. Nakane, S. H. W. Scheres, A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

62. P. Emsley, K. Cowtan, *Coot* : model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).

63. T. C. Terwilliger, O. V. Sobolev, P. V. Afonine, P. D. Adams, Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr D Struct Biol* **74**, 545–559 (2018).

64. P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, Real-space refinement in *PHENIX* for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531–544 (2018).

65. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *REFMAC* 5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* **67**, 355–367 (2011).

66. V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *MolProbity* : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21 (2010).

67. A. Mao, H. J. Ji, B. J. Haas, S. L. Salzberg, M. J. Sommer, TD2: finding protein coding regions in transcripts. *bioRxiv*, doi: doi: https://doi.org/10.1101/2025.04.13.648579 (2025).

68. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).

69. F. Tegenfeldt, D. Kuznetsov, M. Manni, M. Berkeley, E. M. Zdobnov, E. V. Kriventseva, OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Research* **53**, D516–D522 (2025).

70. F. Teufel, J. J. Almagro Armenteros, A. R. Johansen, M. H. Gíslason, S. I. Pihl, K. D. Tsirigos, O. Winther, S. Brunak, G. Von Heijne, H. Nielsen, SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* **40**, 1023–1025 (2022).

71. M. Blum, E. Hobbs, L. Florentino, A. Bateman, InterProScan 6: a modern large-scale protein function annotation pipeline. *Bioinformatics Advances* **6**, vbag141 (2026).

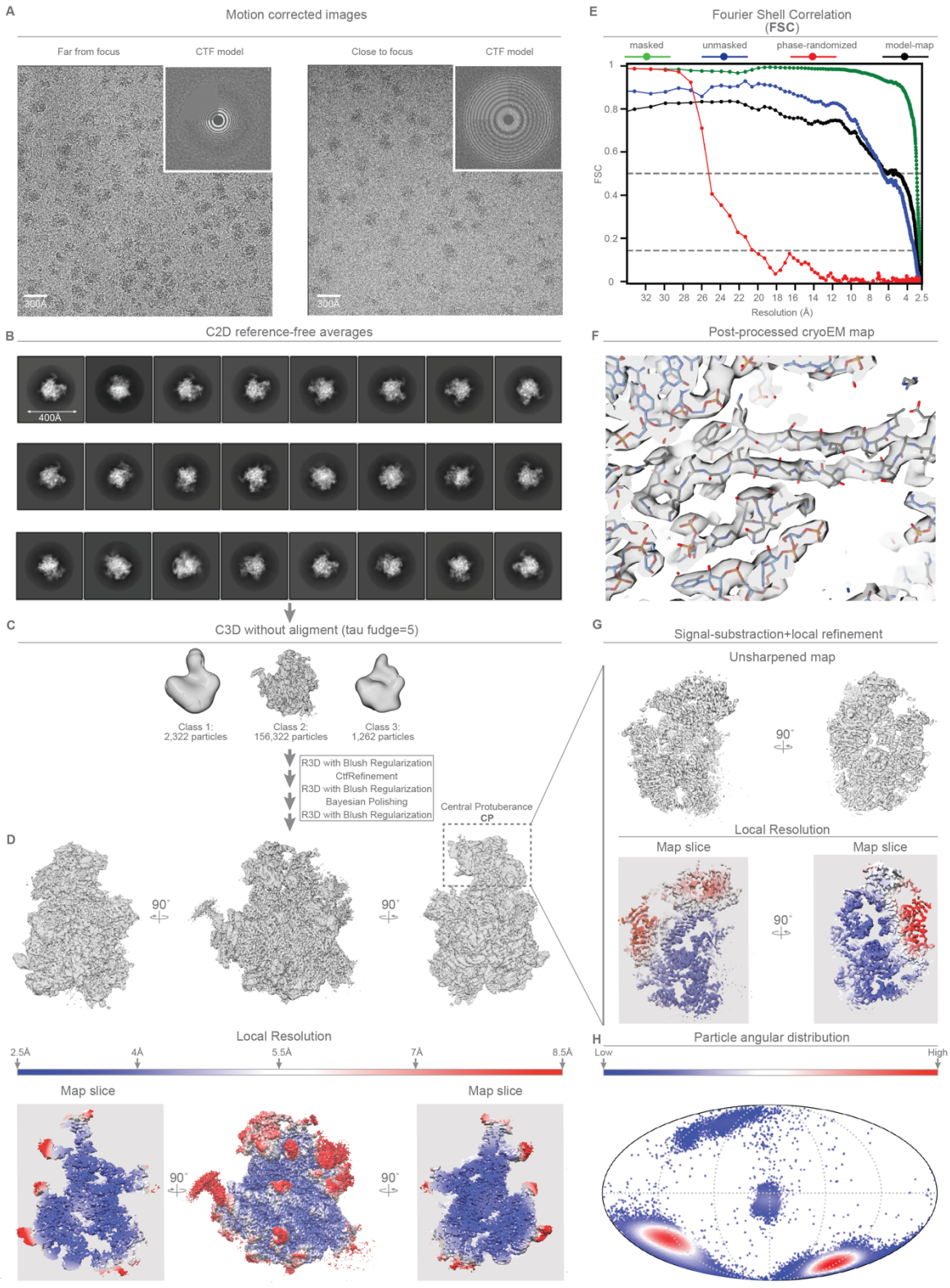
72. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

73. M. Van Kempen, S. S. Kim, C. Tumescheit, M. Mirdita, J. Lee, C. L. M. Gilchrist, J. Söding, M. Steinegger, Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* **42**, 243–246 (2024).

ACKNOWLEDGEMENTS

We thank Bil Clemons and Jodi and Beta Puglisi for introducing ISF to JPM; Juan Alfonzo for introducing JPM to Sergej Djuranovic; Sergej Djuranovic, Slavica Pavlovic-Djuranovic, and Jessey Erath for early access and assistance with the RAPPL protocol; and Jeremy Wideman for comments on the manuscript. Claude Science was used to organize the genome and transcriptome analyses for this work, and a python script developed by Claude Code was used to analyze the lengths of alphaproteobacterial LSU sequences. The cryo-EM described here was collected on a ThermoFisher Titan Krios G2 cryo-transmission electron microscope funded by a National Science Foundation Major Research Instrumentation grant (NSF-MRI 1531991) and a Selectris X imaging filter and a Falcon 4i direct electron detector funded by a National Institutes of Health Shared Instrumentation Grant (SIG; 1S10OD036204-01). **Funding:** This work was supported by the Howard Hughes Medical Institute to J.P.M and J.M.M.; the National Aeronautics and Space Administration Interdisciplinary Consortia for Astrobiology Research (ICAR) grant number 80NSSC23K1357 to J.P.M.; and the Ministerio de Ciencia, Innovación y Universidades of Spain within the Plan Estatal de Investigación Científica, Técnica y de Innovación with grant number PID2023-147463NB-I00 to I.S.F. **Author contributions:** Conceptualization: J.P.M.; Methodology: I.S.F., J.P.M., J.M.M., J.N., N.S.; Investigation: all authors; Visualization: I.S.F., J.P.M.; Funding acquisition: J.P.M., I.S.F., J.M.M.; Project administration: J.P.M., I.S.F.; Supervision: J.P.M., I.S.F.; Writing – original draft: J.P.M., I.S.F.; Writing – review & editing: all authors. **Competing interests:** Authors declare that they have no competing interests. **Data, code, and materials availability:** The assembled *Diceroprocta semicincta* cicada genome has been deposited at DDBJ/ENA/GenBank under the accession JCCJVG000000000. The PacBio reads for the genome are available from the Sequence Read Archive under the accession SRR40353839, and the PacBio reads for the IsoSeq transcriptome analysis are available from the Sequence Read Archive under the accession SRR41210425.

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Fig. S1. (**A**) Representative motion corrected images of the dataset. (**B**) 2D class averages selected for 3D refinements and classifications. (**C**) Simplified schematics of the data processing workflow followed to reach the final map. (**D**) Top, orthogonal views of the final unsharpened maps of the *Hodgkinia* LSU. The bottom shows the same map colored according to local resolution estimations. (**E**) The overall resolution was estimated to be 3Å. Regions at the core of the particle like the peptidyl transferase center (PTC) exhibited map features compatible with resolutions better than 3 Å, however the quality of the map degraded towards the periphery with local resolution estimation in the 4-5 Å range. (**F**) Post-processing using a global B-factor of -20 $Å^2$ produced a well-defined map. (**G**) The central protuberance (CP) region was independently refined using local signal subtraction and refinement, producing and enhanced map that allowed better modeling. (**H**) The angular particle distribution for the final refinement job using a Mollweide projection of the spherical values.

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

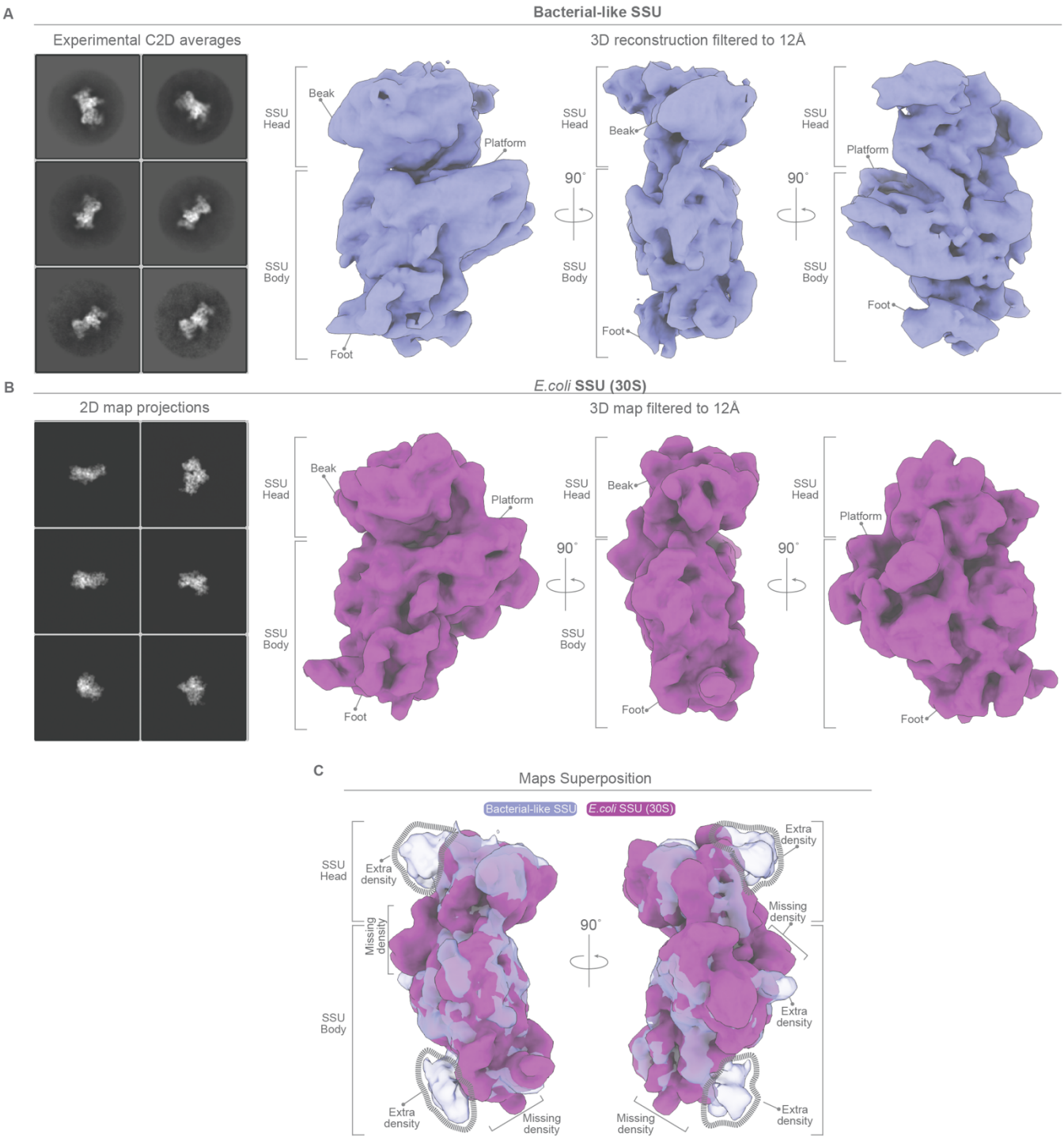


Fig S2. (**A**) 2D class averages from the dataset that resemble a prokaryotic small subunit (SSU). This set of particles was small (~2,500), however it was enough to produce a 3D map with a resolution of around 12Å. This map is shown in 3 different orthogonal views with the characteristic regions of the SSU indicated. (**B**) For comparative purposes, we filtered a SSU map from *E. coli* to 12Å and projected that map in 2D to generate a set of views in similar orientations as our experimentally identified 2D class averages. On the right, we show the *E.coli* map filtered to 12Å in similar orientations as in (A). (**C**) When we superpose these two maps, two large regions of extra density are evident at the top of the head and at the foot of the SSU. At this resolution we are unable to identify this SSU as from *Hodgkinia*, or identify proteins that make up this missing density.

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

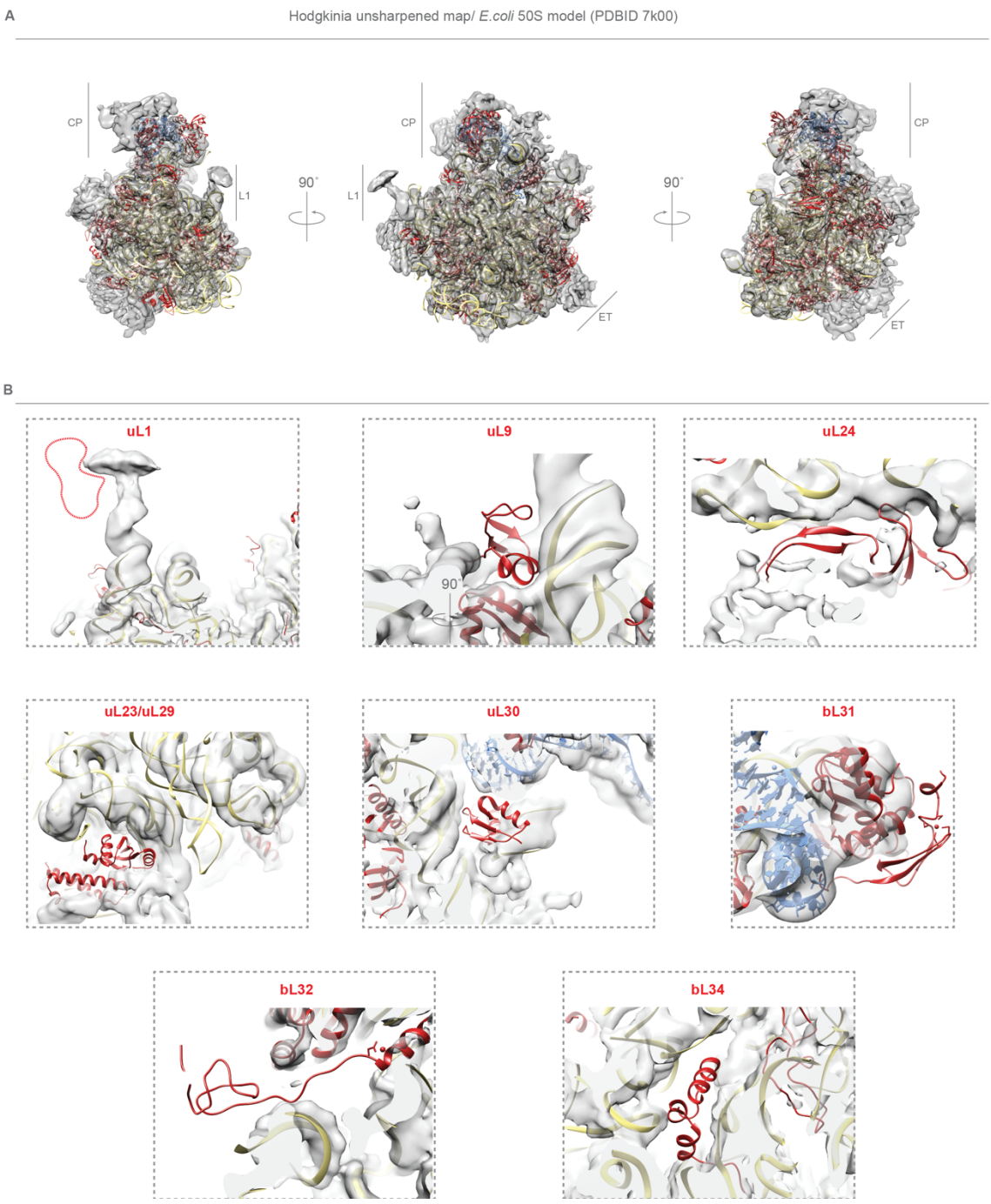


Fig. S3. (**A**) *Hodgkinia's* genome suggested that several ribosomal proteins (RPs) might be missing, as well as extra unexplained density compared to *E. coli* (gray, unfilled areas of the map). We fitted a whole *E.coli* model (PDBID 7k00) into the unsharpened map, using the highly conserve peptidyl transferase center (PTC) region of the ribosomal 23S rRNA as an initial anchor and superposed the rest of the *E.coli* LSU afterwards. (**B**) We then looked for the positions on the *E.coli* LSU of the proteins predicted to be missing in *Hodgkinia* (uL1, uL9, uL21, uL23, uL24, uL29, uL30, bL31, bL32 and bL34) and found clear evidence of missing density for all but uL21. Based on the solid cryoEM density for uL21 we went back to the genome and found an annotation error, which we confirmed and fixed.

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

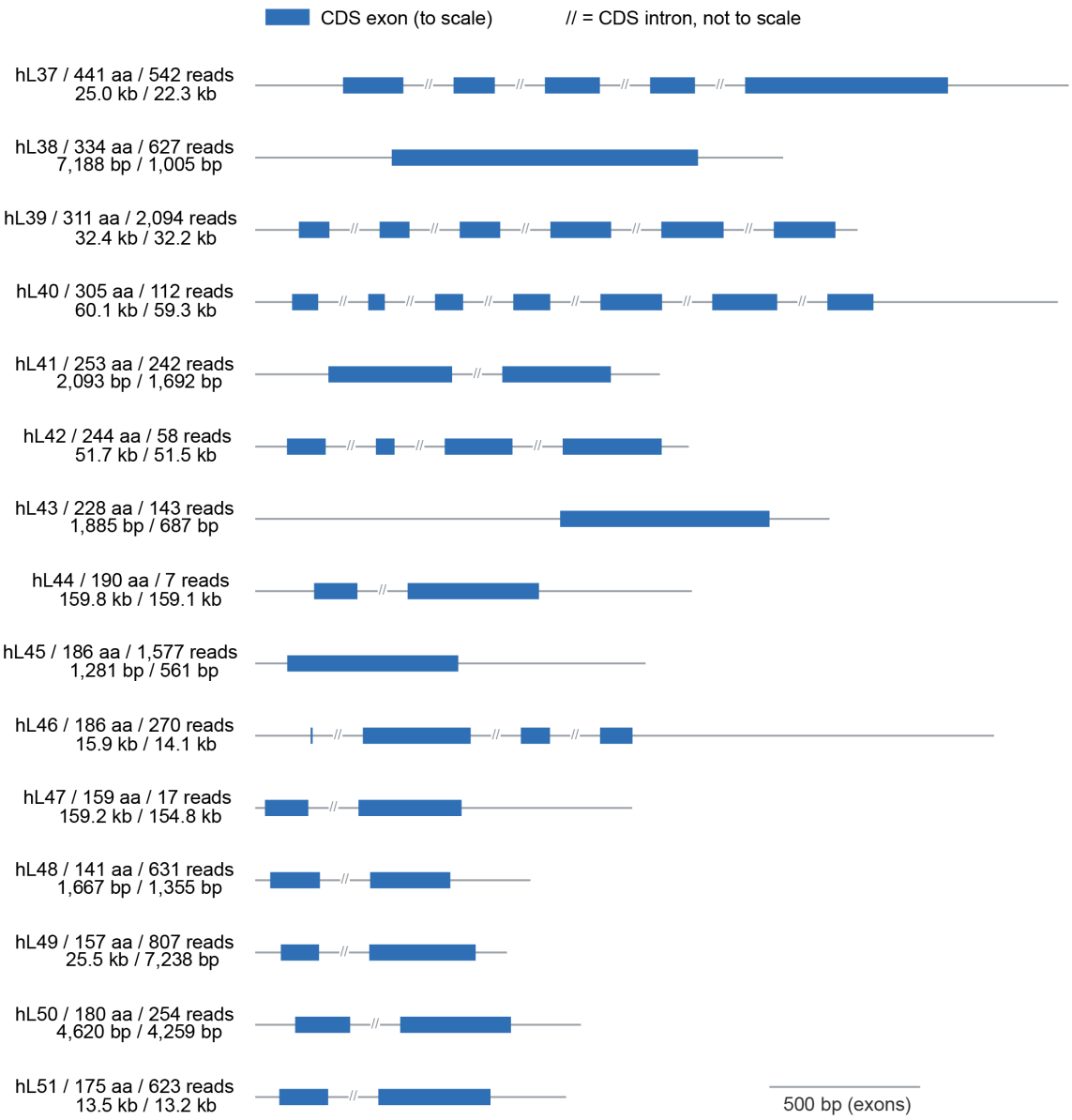


Fig. S4. Genomic structure of the genes for the 15 host proteins on the *Hodgkinia* ribosome. On the left, each gene is labeled with the protein name / length in amino acids / number of IsoSeq reads covering this exact splice form on the top line, and the length of the complete gene including the predicted UTRs / length of the coding sequence (CDS) on the bottom line. Exons are drawn to scale, while the introns are reduced in size by hash marks (//) are not not drawn to scale.

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

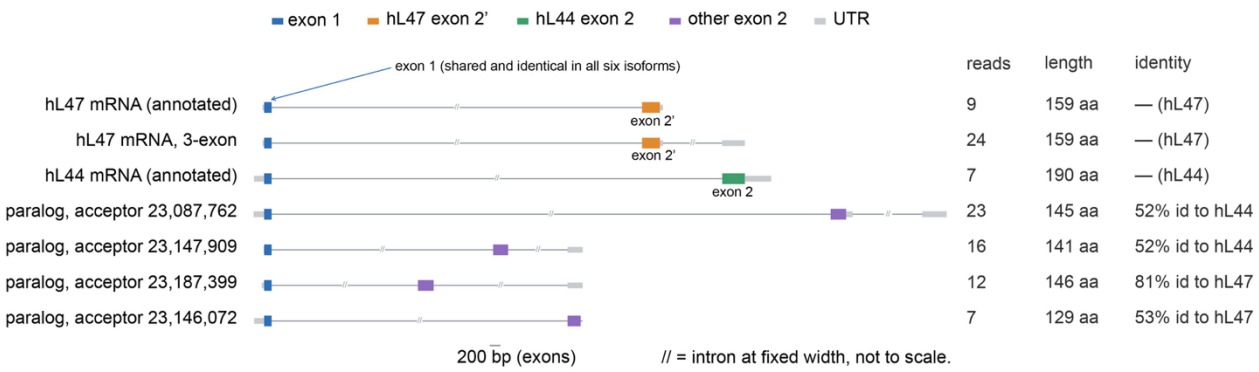


Fig. S5. Splice isoforms for the cicada gene that produces hL44 and hL47. All six predicted isoforms share the same exon 1 (blue) on contig ptg000184l (coordinates for hL44 and hL47 are in table S2). The cicada protein hL44 is produced by a unique exon 2 (green), while hL47 is produced by exon 2' (orange). Two different versions of the hL47 transcript exist (top two rows) but they produce exactly the same protein product so it is unclear which transcript produces the hL47 protein, or if both do. Exon lengths are to scale but intron lengths are not, although the relative sizes of the introns are representative. The introns for this gene are large: the distance between exon 1 and exon 2 in hL47 is 154 kb, and the distance between exon 1 and exon 2' in hL44 is 158 kb.

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Table S1. Cryo-EM data collection, refinement, and validation statistics.

	Hodgkinia LSU (EMD-XXX)
Data collection and processing	
Magnification	130,000x
Voltage (kV)	300
Electron exposure (e–/Å^2)	25,93
Defocus range (μm)	-0.8 to -2.5
Pixel size (Å)	0.93
Symmetry imposed	C1
Initial particle images (no.)	3224334
Final particle images (no.)	156322
Map resolution (Å)	3
FSC threshold	0.143
Refinement	
Initial model used (PDB code)	ModelAngelo/7k00
	Manual build
Model resolution (Å)	3.5
FSC threshold	0.5
Model resolution range (Å)	3
Map sharpening *B* factor (Å^2)	-20
Model composition	
Non-hydrogen atoms	95,574
Protein residues	4,789
Nucleotide residues	2,652
B factors (Å^2)	
Protein	94.35
Nucleotide	88.56
R.m.s. deviations	
Bond lengths (Å)	0.0091
Bond angles (°)	1.214

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Validation	
MolProbity score	2.06
Clashscore	4.75
Poor rotamers (%)	1.67
Ramachandran plot	
Favored (%)	85.9
Allowed (%)	12.51
Disallowed (%)	1.75

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Name	Contig	Strand	CDS start	CDS stop	Len. (aa)	Signal peptide	Consensus annotation	InterPro evidence	BLASTP evidence	Foldseek evidence
hL37	ptg000103l	+	2601317	2623596	433	Y, 17 and 18	Gamma-interferon-inducible lysosomal thiol reductase (GILT)	PANTHER PTHR13234 (aa 56-263, e=5.1e-29)	GILT-like protein 1 isoform X2 (acc XP_039286794.1, e=6.73e-26)	Gamma-interferon-inducible lysosomal thiol reductase-like protein (AFDB Q7PQD1, prob=1.00, e=5.00e-17)
hL38	ptg000311l	-	35669526	35668522	334	Y, 17 and 18	Uncharacterized	No hit	hypothetical protein (KAG8297639.1, e=3.05e-04, 22%)	No hit
hL39	ptg000062l	-	3944712	3912484	311	Y, 29 and 30	UMP-CMP kinase 2, mitochondrial-like	No hit	UMP-CMP kinase 2 (KAE8746981.1, e=1.77e-10)	UMP-CMP kinase 2, mitochondrial (AFDB Q5EBM0, prob=1.00, e=1.24e-9)
hL40	ptg000062l	+	3677654	3736996	305	N	UMP-CMP kinase 2, mitochondrial-like	GENE3D G3DSA:3.40.50.300 (aa 104-294, e=6.1e-08)	UMP-CMP kinase 2 (XP_021920734.1, e=3.99e-31)	UMP-CMP kinase 2, mitochondrial (AFDB Q5EBM0, prob=1.00, e=5.48e-13)
hL41	ptg000184l	+	50547705	50549396	253	Y, 27 and 28	Beta-1,3-glucan-binding protein (BGBP)	PFAM PF15886 (aa 121-231, e=1.1e-09)	hypothetical protein (acc KAJ0178108.1, e=4.37e-07)	Beta-1,3-glucan-binding protein (Fragment) (AFDB O96363, prob=1.00, e=3.56e-7)
hL42	ptg000131l	-	36660454	36608972	246	Y, 20 and 21	Magnesium-dependent phosphatase 1 (MDP-1)	PFAM PF12689 (aa 34-191, e=4.6e-31)	magnesium-dependent phosphatase 1-like (XP_049784562.1, e=8.03e-29)	Magnesium-dependent phosphatase 1 (AFDB Q86V88, prob=1.00, e=1.30e-17)
hL43	ptg000036l	+	9899134	9899820	228	Y, 42 and 43	Uncharacterized	No hit	No hits	No hit
hL44	ptg000184l	-	23277937	23118821	190	Y, 17 and 18	Beta-1,3-glucan-binding protein (BGBP)	PFAM PF15886 (aa 31-130, e=1.0e-02)	No hits	Beta-1,3-glucan-binding protein (Fragment) (AFDB O96363, prob=1.00, e=4.13e-6)
hL45	ptg000311l	+	35435798	35436358	186	Y, 24 and 25	Uncharacterized	No hit	No hits	No hit
hL46	ptg000219l	-	11203321	11189178	186	Y, 25 and 26	Reeler/ferric-chelate-reductase-like fold	No hit	No hits	Putative defense protein 3 (AFDB Q0Q028, prob=1.00, e=6.73e-7)
hL47	ptg000184l	-	23277937	23123102	159	Y, 17 and 18	Beta-1,3-glucan-binding protein (BGBP)	PFAM PF15886 (aa 31-117, e=7.2e-03)	unnamed protein product (acc CAG9861070.1, e=1.90e-02)	Beta-1,3-glucan-binding protein (Fragment) (AFDB O96363, prob=1.00, e=1.10e-5)
hL48	ptg000184l	-	28871256	28869902	141	Y, 18 and 19	Uncharacterized	No hit	No hits	No hit
hL49	ptg000311l	+	588695	595932	157	Y. 26 and 27	MBF2 / probable salivary-secreted-peptide-like	PFAM PF15868 (aa 34-125, e=2.3e-12)	probable salivary secreted peptide (XP_069694665.1, e=1.34e-04)	Probable salivary secreted peptide (AFDB D8KY57, prob=1.00, e=2.29e-9)
hL50	ptg000184l	-	30712044	30707786	180	Y, 23 and 24	Uncharacterized	No hit	No hits	No hit
hL51	ptg000184l	+	29426953	29440152	175	Y, 16 and 17	Beta-1,3-glucan-binding protein / GNBP family	GENE3D G3DSA:2.60.40.2140 (aa 33-136, e=2.2e-09); PFAM PF15886 (aa 36-120, e=1.7e-03)	No hits	Gram-negative bacteria-binding protein 3 (AFDB Q9NHA8, prob=1.00, e=4.2e-06)

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Table S2. Predicted characteristics, annotation, and genome coordinates for the fifteen cicada proteins on the *Hodgkinia* LSU. The signal peptide cleavage sites were predicted using SignalP 6.0 (*70*). The consensus annotation was made by combining the InterProScan (*71*) results with BLASTP (*72*) searches against the nr database, along with Foldseek (*73*) searches using AlphaFold 3 (*31*) structure predictions as search seeds.

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Data S1. Sequences for the 15 host proteins, hL387-hL51.

>hL37
MTPTAFVFLSLVAVAFPRLTTFRNWKVTCPRWYYEGRTSENPIGLRLLDYLLFKKFSTTEVIQVAKRV
NITVLYEALSNSSVYFMKEHVAPVFKYLHQGINLELIPYGRTTEDISEYRYEFTCPNGEKETIANKIHAC
AMYFLAKRKKHTRLVEVISCLMGHSDHKKALSTCCSKSGVQSMRIRKCAFRHEGSYLLARKGERNRQ
WISNITEFPAVVVDGVFDEAKNAMALNDLWSFVCTEVPEVCRPKPPPTTRYVRTTTKKVKRSKKERKL
VKKTLLDKYNKSLDYAMKLAKELWEYKYQKMAAMGCTTTARPSRTTTIKTQGTTSILEYAYSETESE
SEASGTDEDVSKSVSKSASESVSKSASESVSKKITKKVSKTSRKKDVHVSKSDRDIPKRRQEKSESEVID
DYEQYYDELQERRKNAKEGDYGY

>hL38
MKISLVLLMLIIKFVYSGRRHCFYMKSIDTNLTEFLDRPMKSNLEEFFRNLAAYYSALDKLLKYLNEKK
LKYMFDCKFLSQSKTPLFLTKRINISEFQYYCKNNDELFVLIKTYIRRINETYKEFLSACSENPQFLGKPE
PPEYLVDICGKALNHLFRHPIKKPDGFYQLFNNMSLYFKNLQMLMWSIRNRPVKHICEKAHQIRSESN
YDFLEKPLDFEKIKREFQLNDYEFMEFELKMERIQKRWKDFKDVYYNTFVERTTVISFATEFHVEERVE
YSGVDYEFKVTEPETTALEIENDVNKIEDVTEDLSLEEEDVTEDLSLEEEDVTENLN

>hL39
MRSNKNQGFSINCVLRTLISISLLIYASCYPKRGRKFPGSSNRIDWWSNGYYSCLRDIENSFLVKRHAKK
TCNNIIREEFTRISRQTRGRNIRYVDRDGGLIVVECMNKTAADELAKRIAANISATFIEHPTKRLKPLLDK
AKRHIQLTYAIETLSHYDAALRILDDISWKSFVLSRYWYYSAMLAIGRSTLNRATVPGRRHSIYNWPRD
LLVPDVVLFLKYPAYPITDNMTIVEKLNHIDIASRAIKALLRFVFPKVFDLYGDNLYGEIDPFERSMEIIR
NKYNYLIDTAKSWELWNETLYDVTERKYLT

>hL40
MGSIKYCLRESYKVVFFVICVSFMTLCSGFRVYPYKVRNVSILYNYRLNNLSLSEYDEIYGTYHSLQDV
IKVIESEQDRTHKVRELLNMFYAQIRTPLPKPKGRKYPYVVLEGIEHAGRKPLALHMVNTFTGKYIENP
PAIIRSYYPSLMIHTPRFRRAYYVLGNYLAAHYAKQVIHSMPVFMDRCWHSITAFSIAKTFCKNRTKLP
LPQSPIYRWPDDLLKPDIIFFLKITRRSIIKNVRDKLAGKKQSRIFRIIRLMANKEMRDVHRIVFNYYGNY
YRLEKELVTHVKTLFYSGNSTEQQMFN

>hL41
MYSCLIMANGFGLCLILVVILSESSRGERDKVVTSIIGTTTLPTTTKPSTTHVKRVHTGSFTTSRKQKRRP
SPIKRNRHLKKISKKSLRTTENPLLILARGLNYTVTGRVEINDYPEINEHVELFVGGGFRIYILDDNLKSF
VFLGNLNRPYYDVPLTLSPINVNVKHISEYLPDCDMMGVVSESVDGKWIYENRTVSLKEGDVVNYRL
LVSYVDGSYRVKIHQYFTIRKVKRTLVPEIPYKEQKDKYYVDVF

>hL42
MFNLFTFYLCSVSVCGLITPRDLVGHGIQFRNHIPKVFVFGLDRTLWPFCVDTEVAMPLFKNVTGDTTV
VTDAADQELKLYPHVRHILHTLKLKNITAVAVTRSPEVTCIGTLLKVFDMKQYFSEVVIHPGPKVWHL
VWLNNKTLIPFEDMLFFDDEIVDVRYGERLGVVSTLCPQGLYLDDFQRGLDNFQLRTTGEILPDEIAEAI
RKKTLRQFVLPVRPTKRRRKMKNIISAYIRRRQMRQFSF

>hL43
MQTILLNNTWTFHSVMFFSAHLTVSKCLLILQCLLVCISTLAFENSLYECTKPPGESYGSRIQKLEDSMV
YKLDHPDKTQWKKLLEEIEECLNMTSDFFLWLKLDDKEEIKYAQQFLKNGVPQIVTRNFTASSLLKKF

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

QFDVYQMLMVLNAIRKLNKAWEVFRDPYFQECYNNEEYYKKLIEKYEAEGYPELKITTDDETWPLNE
EDYPEGHPMRLRIAAEKKGNKVF

>hL44
MFHEVALAVCLGAAVYAITIDVAPPKPHNFQPRVALLKPTGFSVTISDASDIRQIIFFGRYNRKYDINRP
YRTCVYRYNIINKMRDYTMSFNMSTPEITFRDGDYFHYWMYVVDYQHMRHAVHDLSFHIVDMNGTL
TVQYNLYVPIRKMETTSVMHKLETSTESWKFTTTSTKQAATVTQPESFYNVFDD

>hL45
MDNLKSNGLLLTLIMAVLLNITSASKNSSQVQMFFGREIKREGIQQKVFKRKLSTINIDEVIFNLRNFSR
VIEQKVLILKTLIHEEAVTLLSDVRKYNENMHKLYRCRKVYRGKVKEMLKEMNANGEPRYLKLLNY
MFPVETNYNYSNVISLSISLKPHEIRKLVKRLGRSHDLRHKLNTITKYPL

>hL46
MKLQSSCFIFLLLVASCSLTRHVTGRGKHYAKTEYFIQTMCRNVFPRFYRNPGARRENSTPVHHAITTW
ETQDKLKITIRGNDSRKIRACVIQVRYDRLPRPRFLDYSAYIIYDCKPFKKNMIYVAPLKETTEITVFWIL
PREIPANKLWLYFSPVLDLPTYWVLHQPIAPPPLYRRPNDRYLLNY

>hL47
MFHEVALAVCLGAAVYAITIDVAPPKPHNFQPRVALLKPTGFSVTISDANNIRKVQFYGSYNKPYDMV
KTVSSCPYQANATNPTRSNIIHFEISPGPQFKYNDSFYYWMYVVDYDSVQHFVHYLRFDVERLDDDSII
AVRHREENVFLKKFFNVTKYIN

>hL48
MAKCRLLSVILLIHGIAALKYMLLKERKCVYGNSVLRPQVKFLKPRGLFFIVRGETNCWEKMELFANV
ERPVYDHTVCDLCEVLKRIYLTANRLYWYTYKDDLDLHAQQWLYFRLEATHKQGNVTVLHGKAPM
KTKLNSIW

>hL49
MKTRIFKLIFTWYLTVWSFQILLCSGRKNETVRLGRRYKEDRLLHEYRVTLNATTIKFARVQMASTYY
PGSATSRVTCIYLIDFMKDRRQGYAFIYSGGVTHRHAAIVFRSQAKVGLDMACEIYGFDPEPVTGEYES
DESGEKTENYFISSKSVDSN

>hL50
MGFTFDGTACIFILLLSEVRVKGDSIAVNTTLETLLNVTELLFPRLHFLILPNTGCIFRYSLSDTSVLVPKV
IYANLRKKFVDKYVYTRHIVETFTRRKGWYSYKYTHEVLELDDVFFGRCMFLRKDVRDRKHYVDDPF
LIKITNTTPYVKNKQFRPLRIYTPKFFGQYLEKVRRKEKRF

>hL51
MLLAGLLVFFIREVFTGHLPNCSITTVNGIRYYDIPNFRFYQIMPQGFYIGFEDVSGWRNVSIHMNTNKP
IKIKEKGQLAIPLSYRLFFKNKTEWGHINVNRKIKNDDMLYVWMNVTDYKKRIDCYYLEPFVGNKYR
HFPYIEEMTEELFISDSWLTIKRTRRPAYEELFDDFEM

bioRxiv preprint doi: https://doi.org/10.64898/2026.10.18.759926; this version posted October 22, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.