

RESEARCH ARTICLE

Elucidation of host and symbiont contributions to peptidoglycan metabolism based on comparative genomics of eight aphid subfamilies and their *Buchnera*

Thomas E. Smith , Yiyuan Li , Julie Perreau , Nancy A. Moran 

Department of Integrative Biology, University of Texas at Austin, Austin, Texas, United States of America

* smit4227@gmail.com

Abstract

Pea aphids (*Acyrtosiphon pisum*) are insects containing genes of bacterial origin with putative functions in peptidoglycan (PGN) metabolism. Of these, *rlpA1-5*, *amiD*, and *IdcA* are highly expressed in bacteriocytes, specialized aphid cells that harbor the obligate bacterial symbiont *Buchnera aphidicola*, required for amino acid supplementation of the host's nutrient-poor diet. Despite genome reduction associated with endosymbiosis, pea aphid *Buchnera* retains genes for the synthesis of PGN while *Buchnera* of many other aphid species partially or completely lack these genes. To explore the evolution of aphid horizontally-transferred genes (HTGs) and to elucidate how host and symbiont genes contribute to PGN production, we sequenced genomes from four deeply branching lineages, such that paired aphid and *Buchnera* genomes are now available for 17 species representing eight subfamilies. We identified all host and symbiont genes putatively involved in PGN metabolism. Phylogenetic analyses indicate that each HTG family was present in the aphid shared ancestor, but that each underwent a unique pattern of gene loss or duplication in descendant lineages. While four aphid *rlpA* gene subfamilies show no relation to symbiont PGN gene repertoire, the loss of aphid *amiD* and *IdcA* HTGs coincides with the loss of symbiont PGN metabolism genes. In particular, the coincident loss of host *amiD* and symbiont *murCEF* in tribe Aphidini, in contrast to tribe Macrosiphini, suggests either 1) functional linkage between these host and symbiont genes, or 2) Aphidini has lost functional PGN synthesis and other retained PGN pathway genes are non-functional. To test these hypotheses experimentally, we used cell-wall labeling methods involving a D-alanine probe and found that both Macrosiphini and Aphidini retain *Buchnera* PGN synthesis. Our results imply that compensatory adaptations can preserve PGN synthesis despite the loss of some genes considered essential for this pathway, highlighting the importance of the cell wall in these symbioses.

OPEN ACCESS

Citation: Smith TE, Li Y, Perreau J, Moran NA (2022) Elucidation of host and symbiont contributions to peptidoglycan metabolism based on comparative genomics of eight aphid subfamilies and their *Buchnera*. PLoS Genet 18(5): e1010195. <https://doi.org/10.1371/journal.pgen.1010195>

Editor: Diarmaid Hughes, Uppsala University, SWEDEN

Received: September 24, 2021

Accepted: April 9, 2022

Published: May 6, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1010195>

Copyright: © 2022 Smith et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Sequencing data can be found at NCBI (PRJNA758084). Genome annotations and other data generated during the

current work can be found at Zenodo (<https://zenodo.org/record/5517159>). Scripts and code used in this study can be found at GitHub (https://github.com/lyy005/peptidoglycan_related_genes_in_basal_aphids). All accession numbers and URLs to data repositories can be found in the main text of the manuscript.

Funding: This work was supported by National Institutes of Health (<https://www.nih.gov/>) Awards 5F32GM126706 (to T.E.S.) and 5R35GM131738 (to N.A.M.), and National Science Foundation (<https://www.nsf.gov/>) Award 1551092 (to N.A.M.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Throughout evolution, animals have sometimes gained novel abilities by acquiring bacterial genes through horizontal gene transfer. For some insects harboring bacterial symbionts, horizontally-transferred genes may enable hosts to regulate symbiosis by influencing symbiont cell wall metabolism. While mealybug horizontally-transferred genes work collectively to synthesize the symbiont cell wall, the role of aphid horizontally-transferred genes in symbiont cell wall metabolism is unclear. We examined whether different aphid horizontally-transferred genes co-occur with symbiont genes underlying cell wall metabolism across different aphid lineages, indicative of linked function. We included 17 aphid species representing eight distantly related lineages, four of which we sequenced for this study. We found that two of the three horizontally-acquired gene families are present only when symbionts possess cell wall pathway genes, while the third shows no correlation. These results reveal that despite their putative involvement in symbiont cell wall synthesis, aphid horizontally-acquired genes operate independently from one another and likely have lineage-specific functions. Furthermore, we observed that symbiont cell wall synthesis is maintained in one aphid lineage despite loss of genes considered essential for producing the cell wall, implying that other adaptations preserve the cell wall in aphid species with incomplete cell wall synthesis pathways.

Introduction

Horizontal gene transfer (HGT) has played a pivotal role in eukaryotic evolution, granting novel functions to its recipients and providing the means to expand to new ecological environments [1–7]. HGT has been particularly relevant in the evolution of eukaryotes that live in close association with bacteria [8,9]. At the extreme, the majority of mitochondrial and plastid genes have been transferred to the host genome [10,11], affording hosts complete control of their bacterial symbionts and the benefits they provide. In contrast, transfer of symbiont genes to the host genome is absent or limited in a number of insect endosymbioses [12–15]. In these systems, host genomes may instead harbor genes anciently acquired from bacteria other than the current symbiont(s) that have obtained roles in maintaining extant symbioses through nutrient provisioning [15,16] or the metabolism of peptidoglycan (PGN), the building block of bacterial cell walls [12–14]. For example, the genome of the mealybug *Planococcus citri* contains horizontally-transferred genes (HTGs) encoding enzymes of the PGN synthesis pathway that perfectly complement the PGN gene repertoire of the symbionts [14] and contribute to the construction of the symbiont cell wall [17].

Aphids are sap-feeding insects that rely on their *Buchnera* symbionts to provide essential amino acids lacking in their diet. The pea aphid (*Acyrtosiphon pisum*) *Buchnera* genome is missing many genes considered to be essential in free-living relatives like *Escherichia coli* [18], including most PGN pathway genes. Pea aphids contain eight HTGs with putative functions in PGN metabolism: *ldcA*, *amiD*, *rlpA1-5*, and *bLys* [12,13], of which all but *bLys* show greatly elevated expression in bacteriocytes, the specialized host cells that harbor *Buchnera*, relative to other aphid tissues [13]. The putative functions of the remaining seven HTGs correspond to specific steps in bacterial PGN metabolism: *amiD* and *rlpA* putatively encode an amidase and lytic transglycosylase (LT), respectively, both of which are involved in cell wall remodeling, while *ldcA* encodes a putative L,D-carboxypeptidase of the PGN recycling pathway (Fig 1). Aphids also contain invertebrate-type (i-type) lysozymes that are highly expressed in bacteriocytes [19], especially during late adulthood when symbionts and bacteriocytes are degraded

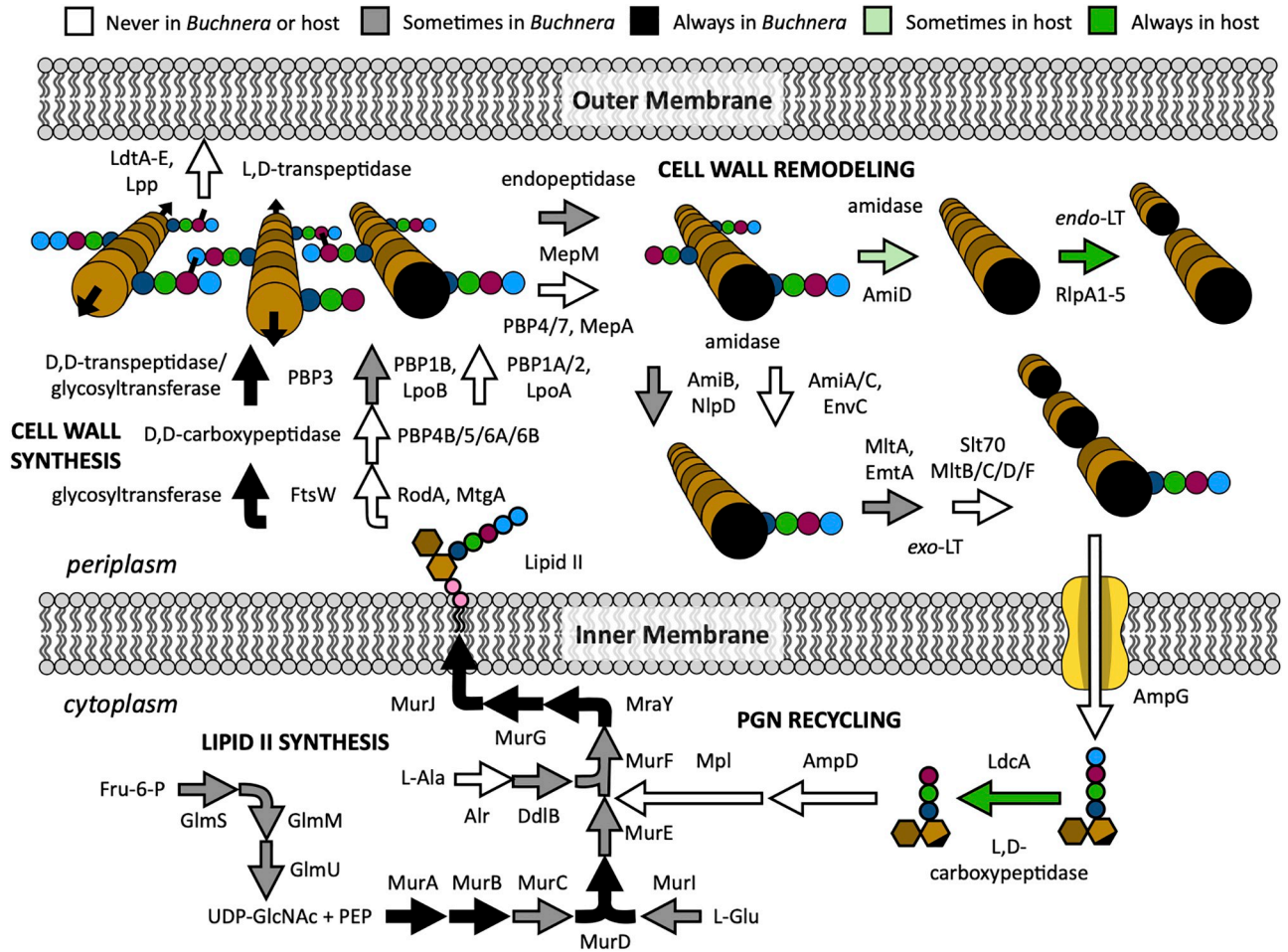


Fig 1. Hypothetical metabolic reconstruction of *Buchnera* PGN metabolism with potential involvement of aphid HTGs in the Aphidinae subfamily (including tribes Macrosiphini and Aphidini). PGN metabolism is divided into four pathways: PGN recycling and lipid II synthesis in the cytoplasm, and cell wall synthesis and remodeling in the periplasm. Arrows depict enzymatic reactions, with black, gray, and white colors indicating complete, sporadic, or no conservation of the responsible gene(s) among ten *Buchnera* genomes. Light and dark green arrows denote incomplete and complete conservation of host HTGs. For PGN cartoons, dark and light brown indicate GlcNAc and MurNAc glycans, respectively, while black-ended light brown indicates anhydro-MurNAc. Stem peptides consist of L-Ala (dark blue), iso-D-Glu (green), meso-diaminopimelic acid (*mDap*) (maroon), and D-Ala (light blue) amino acids. For lipid II, the phosphate groups of undecaprenyl-diphosphate are shown in pink.

<https://doi.org/10.1371/journal.pgen.1010195.g001>

[20]. Thus, aphid hosts may have the enzymatic potential to control *Buchnera* by influencing its cell wall structure.

Despite the apparent relationship between aphid HTGs and symbiont cell wall metabolism, it is unclear from their putative functions alone how they might contribute to this process. The *A. pisum amiD* and *rlpA1-5* genes may be functionally redundant with *Buchnera amiB* and *emtA* or *mltA*, respectively, and *Buchnera* completely lacks the PGN recycling pathway to which aphid *LdcA* putatively belongs (Fig 1) [21]. Recent evidence shows that aphid *LdcA* instead acts on PGN within or shed from the *Buchnera* cell wall [22]. One way to establish whether genes are functionally linked is to determine whether they are preserved or lost together throughout evolution using phylogenetic profiling—the correlated presence of genes within genomes of distinct evolutionary lineages can provide a signal of functional interactions [23,24]. Similarly, if host and symbiont genes interact, they should co-occur in paired host and symbiont genomes across diverse lineages, whereas genes that do not interact will be gained or

lost independently of one another. Such an approach has been used previously to detect host-symbiont collaborations among various mealybug species [25]. Gene distributions across species have also been used to correlate the lack of class A penicillin-binding proteins with an intracellular lifestyle among distantly related bacteria [26]. Within the aphid superfamily Aphidoidea, *Buchnera* symbionts vary widely in genome size and content, including PGN genes [27], providing a natural source of variation to test whether host HTGs coincide with certain symbiont PGN genes or pathways. Furthermore, in a few aphid lineages, *Buchnera* has been lost and replaced with a novel symbiont type [28,29] or has been joined by additional symbionts [30,31]. While many *Buchnera* genomes are available from diverse aphid species [27,30–34], only 13 aphid genomes have been sequenced, annotated, and published to date, and most are concentrated in a single subfamily, the Aphidinae (S1 Table).

In this work, we introduce draft genomes of aphid species from four additional aphid subfamilies: *Geopemphigus* sp. (Fordinae), *Stegophylla* sp. (Phyllaphidinae), *Chaitophorus viminalis* (Chaitophorinae), and *Pemphigus obesinymphae* (Pemphiginae). We applied phylogenetic profiling across the Aphidoidea superfamily to evaluate whether certain host and symbiont PGN genes co-occur across different aphid species, supporting a potential functional relationship. For the 17 annotated aphid genomes, we assayed the incidence (presence/absence) of host genes with putative functions in PGN metabolism, including homologs of the *A. pisum* HTGs *ldcA*, *amiD*, *rlpA*, and *bLys*, as well as PGN pathway genes from the corresponding primary symbiont of each host species. For the latter, we used a manually-curated database of primarily *E. coli* and *Buchnera* genes representing known PGN genes and members of the cell division and flagellar basal body (FBB) pathways, both of which may interact with or depend on cell wall metabolism [35–37]. We show that the PGN gene repertoire of each aphid species varies for both host and symbiont. The pattern of aphid *ldcA* and especially *amiD* gene loss correlates with the absence of PGN pathway genes in their respective symbiont(s), while *rlpA* genes, though variable in the incidence and abundance of four distinct paralogous groups, are conserved among all aphids. These observations imply that aphid HTGs are not all involved in symbiont PGN metabolism and likely function independently of each other, suggesting that the encoded enzymes do not cooperate to complement a missing multi-step pathway in symbionts as proposed for mealybug HTGs [25]. Furthermore, a clear difference in gene repertoire exists between two tribes of the Aphidinae subfamily. Relative to tribe Macrosiphini, Aphidini aphids lack *amiD* and their symbionts lack 2–3 genes central to lipid II biosynthesis (*murF*, *murC*, and/or *murE*). Thus, gene repertoires suggest that Aphidini *Buchnera* cannot synthesize cell walls at all. In contrast with this hypothesis, we provide experimental evidence for functional cell wall synthesis in an Aphidini species, *Rhopalosiphum maidis*, demonstrating that 1) novel adaptations have evolved within the PGN biosynthesis pathway to preserve the cell wall in this species, and 2) genomic analysis alone is an insufficient predictor of symbiont capabilities.

Results

Genome sequencing and annotation

To understand the evolution of genes related to PGN pathways, we sequenced the genomes of four divergent aphid lineages, including the first genome sequences for subfamilies Fordinae (*Geopemphigus* sp.), Phyllaphidinae (*Stegophylla* sp.) and Pemphiginae (*Pemphigus* sp.), and the second genome sequence for subfamily Chaitophorinae (*Chaitophorus* sp.) after *Sipha flava* (NCBI BioProject PRJNA472250). The *Geopemphigus* genome also represents the first aphid genome for which *Buchnera* has been lost and replaced with an unrelated endosymbiont [29]. For each genome, we generated 38–58 Gb of genomic data from one short insert size Illumina library (S2 Table) and 5–7 Gb RNA-seq data (S3 Table). After assembly, estimated

genome sizes range from 245 Mb—658 Mb (S4 Table). The sequenced genomes show high completeness based on the presence of single-copy Benchmarking Universal Single-Copy Orthologs (BUSCOs) for Hemiptera (n = 2,510) [38], ranging from 97%-99% for both the assemblies (S5 Table) and our gene annotation results (S6 Table). These values are similar to those for other aphid genome sequences [33]. All sequencing data can be found at NCBI under the accession number PRJNA758084.

Aphid HTG incidence varies across aphid species

In order to understand 1) the incidence of *A. pisum* HTG homologs among different aphid lineages and 2) the evolutionary history of host-encoded PGN genes, we reconstructed the aphid phylogeny using the four newly sequenced aphid genomes, 13 available aphid genomes, and six related insects in suborder Sternorrhyncha included as outgroups (S1 Table). The phylogeny was based on 829 single-copy genes found in at least 80% of the species (Fig 2). Genes with

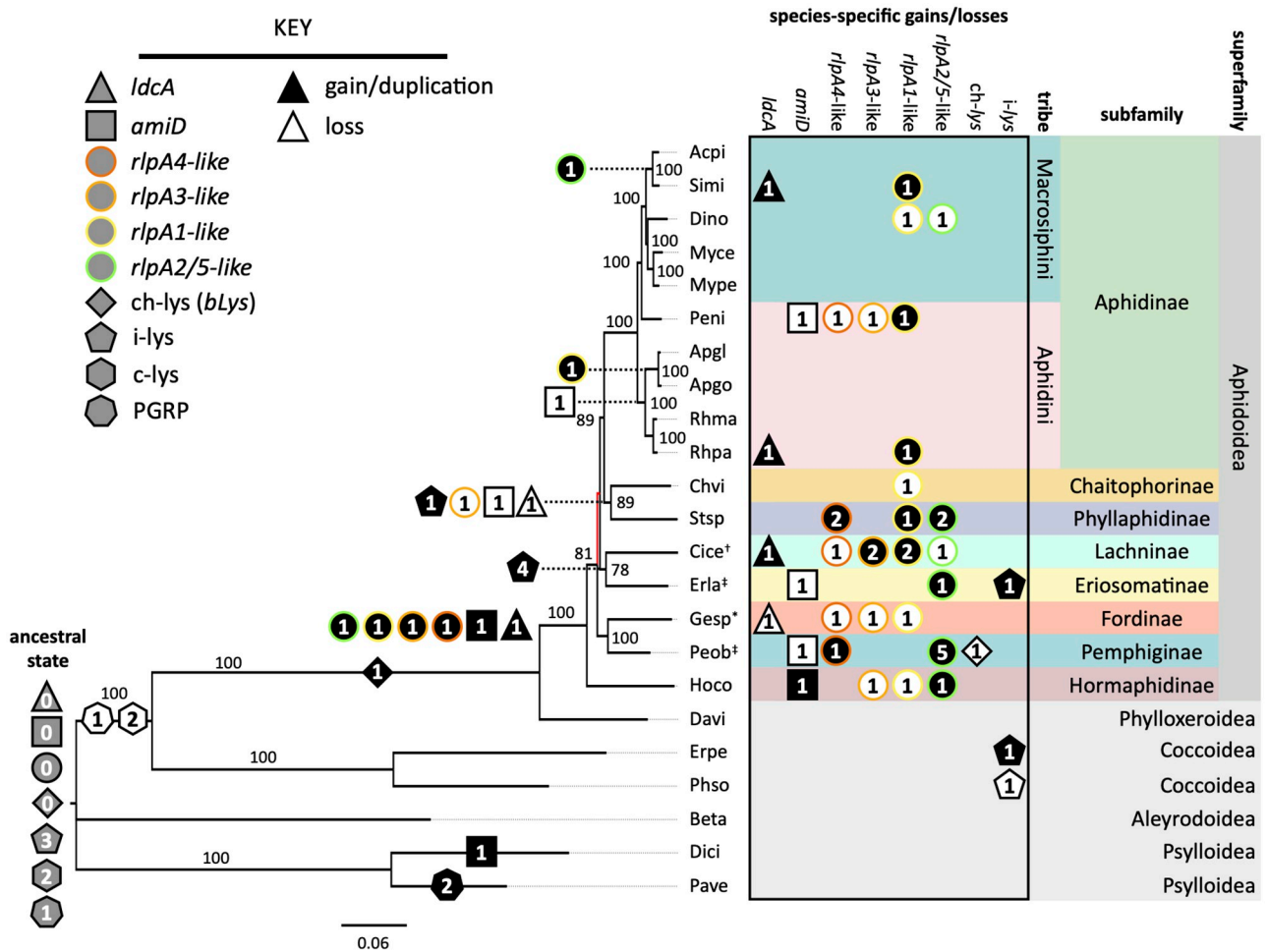


Fig 2. Ancestral reconstruction of aphid genes that putatively interact with Gram-negative bacterial PGN. The aphid phylogeny was constructed using maximum likelihood from a concatenated nucleotide alignment of 829 core genes shared by all 23 insect species. Bootstrap supports after 1,000 replicates are shown, with red branches indicating supports < 70%. All symbionts are *Buchnera* except those of *Geopemphigus* sp, indicated by an asterisk, which contains a novel Bacteroidetes symbiont. Host species containing known co-obligate or uncharacterized facultative symbionts are indicated with an obelisk (†) or a double obelisk (‡), respectively. Each species code represents the combined first two letters of the genus and species. The ancestral state (gray) and placement of gene gains (black), duplications (black), and losses (white) were determined using Count (38). Numbers in shapes indicate the number of gene copies involved in the event. Shapes indicate distinct gene families, and colored borders denote the four RlpA paralogs types, named after the *A. pisum* *rlpA* genes.

<https://doi.org/10.1371/journal.pgen.1010195.g002>

putative activities pertaining to PGN metabolism, including peptidoglycan recognition proteins (PGRPs), lysozymes, and aphid HTGs (*bLys*, *ldcA*, *amiD*, and *rlpA*), were identified by BLAST search using *A. pisum* protein sequences as queries (S7 Table). With the exception of *Daktulosphaira vitifoliae* [39], the outgroup species contain bacterial symbionts, none closely related to *Buchnera*.

We first evaluated the distribution of genes encoding enzymes that target PGN and genes that represent conserved components of insect innate immunity (S7 Table). PGRPs sense bacterial infection by binding and/or cleaving PGN and activating downstream pathways that stimulate the humoral immune response, which includes increased production of bacteriostatic lysozymes [40]. Lysozymes, characterized by their *N*-acetylmuramidase activity, include members from four different glycoside hydrolase families: chicken- and invertebrate-type (c- and i-types, respectively; GH22), goose-type (g-type; GH23), viral-type (v-type; GH24), and Chalaropsis-type (ch-type; GH25) [40]. Both PGRPs and c-type lysozymes are evolutionarily conserved in insects [40,41]. We found no genes encoding PGRPs or c-type lysozymes among aphid or phylloxerid genomes, but their presence in the scale insect, whitefly [42], and psyllid [43,44] genomes support their loss in a shared ancestor of the Aphidomorpha infraorder (Fig 2). The absence of these key immune genes may have facilitated the establishment of endosymbioses in aphids [45], though their absence from phylloxerids, which lack symbionts [39] and presence in other outgroups that harbor obligate symbionts suggest that alternative forces underlie their loss from the Aphidomorpha. We observed little variation in the number of invertebrate-type (i-type) lysozymes among most insects investigated (Fig 2), indicating a functional difference between these and c-type lysozymes. Major exceptions are *Cinara cedri* (Lachninae) and *Eriosoma lanigerum* (Eriosomatinae) which harbor eight to nine i-lysozyme paralogs compared with three or four paralogs in other species. Finally, we found *bLys* in all Aphidomorpha genomes, indicating that this gene family was acquired concomitantly with the loss of PGRPs and c-type lysozymes (Fig 2). The ch-type lysozyme domain of *bLys* is most closely related to domains in prophages of *Wolbachia*-like bacteria [13,46] and is minimally expressed within bacteriocytes, implying a role outside of symbiosis [13]. The multiple independent HGTs of ch-type lysozymes to all domains of life [46] and the unique *N,O*-diacetylmuramidase activity described for some ch-type lysozymes [47,48] suggest that aphid *bLys* is more likely involved in host defense against bacterial pathogens than in symbiont PGN metabolism.

Unique to superfamily Aphidoidea are the HTGs *rlpA*, *ldcA*, and *amiD*, all of which were acquired in a shared ancestor of this superfamily (Fig 2). With the exception of an *amiD* homolog within the *D. citri* genome, none of the bacteriocyte-expressed HTGs found in aphids were detected outside of the Aphidoidea (Fig 2). In the case of *amiD*, phylogenetic analysis revealed that the *D. citri* homolog clusters separately from the aphid copies, indicating that *D. citri amiD* was acquired from a bacterial source independently of aphid *amiD* (S1 Fig).

Aphid lineages vary in whether HTGs have been lost or duplicated (Fig 2). RlpA subfamilies are maintained to different extents among aphid lineages, with each genome containing at least one type (Fig 2). We examined the phylogenetic relationships of aphid RlpA proteins and found that each sequence falls within one of four paralogous subfamilies: RlpA4-like, RlpA3-like, RlpA1-like, and RlpA2/5-like (S2 Fig). RlpA4-like proteins are the most basally branching subfamily and RlpA2/5 is the most recent branch, supporting the initial analysis of *A. pisum rlpA* genes [13]. Ancestral reconstruction indicates that the most recent aphid common ancestor had all *rlpA* subfamilies, implying that the underlying gene duplications occurred in a common ancestor of the Aphidoidea (Fig 2). In contrast, *amiD* and *ldcA* have been lost independently five and two times, respectively (Fig 2). The repeated loss of these genes provides an opportunity to test our hypothesis—if aphid HTGs contribute to symbiont PGN

metabolism, the loss of aphid HTGs should correlate with the absence of interacting symbiont genes.

Aphid *amiD* and *ldcA* coincide with the loss of symbiont PGN

We anticipated that the symbionts of host species lacking *amiD* and/or *ldcA* genes might display greater loss of genes from PGN-related pathways due to a functional link between these host genes and symbiont cell wall metabolism. To test this idea, we looked for patterns of coincidence between host and symbiont PGN gene repertoires among all available sequenced aphid species. First, we measured the incidence of 138 bacterial genes, collectively involved in PGN metabolism or two potentially related pathways, within the primary symbiont genome of each aphid species (Fig 3 and S8 Table). As noted previously, *Buchnera* from aphids of tribe Macrosiphini (subfamily Aphidinae) harbor the most complete set of PGN metabolism genes among *Buchnera* species, maintaining a nearly complete lipid II synthesis pathway and the minimal enzyme functions necessary for cell wall synthesis (glycosyltransferase and D,D-transpeptidase) (Fig 3) [21,27]. Though Macrosiphini *Buchnera* generally maintain genes encoding two of the three essential activities for cell wall remodeling, an amidase (*amiB*) and at least one lytic transglycosylase (LT; *mltA*, *emtA*) gene, they often lack the gene typically encoding the third, endopeptidase activity (*mepM*) (Fig 1). In comparison, *Buchnera* of the tribe Aphidini (subfamily Aphidinae) display more frequent loss of *murF*, *murC*, and/or *murE* of the lipid II synthesis pathway. Most other subfamilies (Chaitophorinae, Phyllaphidinae, Lachninae, Eriosomatinae, and Pemphiginae) exhibit near-complete loss of all PGN pathway genes, while *Buchnera* from *Hormaphis cornu* (subfamily Hormaphidinae) represents an intermediate, maintaining a single cell wall synthesis gene (*mrcB*, encoding the bifunctional glycosyltransferase/D,D-transpeptidase PBP1B) and partial lipid II synthesis pathway (Fig 3). Notably, no aphid primary symbionts, including *Buchnera* of most aphids and *Candidatus Skilesia alternata* (referred to hereafter as *Skilesia*) of *Geopemphigus*, maintain any of the PGN recycling genes used in our searches (S8 Table).

Next, we considered whether the presence or absence of host PGN genes coincides with the degree of symbiont PGN gene loss. The ubiquity of aphid i-type *lys* and *bLys* genes in host genomes suggests no involvement of these genes with symbiont PGN metabolism (Figs 2 and 3). While the incidence of *rlpA* paralog types varies among aphid lineages, all species contain multiple *rlpA* genes, implying no relationship with symbiont gene repertoire. This is less surprising when the functional basis of RlpA's role in bacterial cell division is considered. Bacterial RlpA proteins participate in cell division largely via the N-terminal sporulation-related repeat (SPOR) domain (Pfam 05036) that binds "denuded" (stemless) PGN glycan chains that accumulate at the septum [49,50]. In contrast, aphid RlpA proteins contain only the catalytic double-ψ β-barrel (DPBB) domain (Pfam 03330) of RlpA [12,13]. While the RlpA DPBB domain is unique among LTs for its ability to cleave denuded PGN [51,52], the absence of SPOR domains from aphid RlpA proteins suggests that these enzymes likely do not participate in symbiont cell division. Furthermore, aphid RlpA proteins contain one or more eukaryotic inhibitor cysteine-knot (ICK)/knottin domains that are unrelated to PGN metabolism [12,13], further distancing RlpA from a role in this pathway. Collectively, these observations suggest that the DPBB domain may serve an alternative function in aphid RlpA proteins.

Aphid *ldcA* exhibits a mixed distribution among host species harboring symbionts with and without PGN. The uniform conservation of *ldcA* in the Aphidinae subfamily, where *Buchnera* maintains most PGN pathway genes (Fig 3), supports a relationship between *ldcA* and symbiont PGN metabolism, at least in this lineage. Aphid *ldcA* is also present in *H. cornu*, whose *Buchnera* harbors a partial PGN pathway (Fig 3). However, aphid *ldcA* is

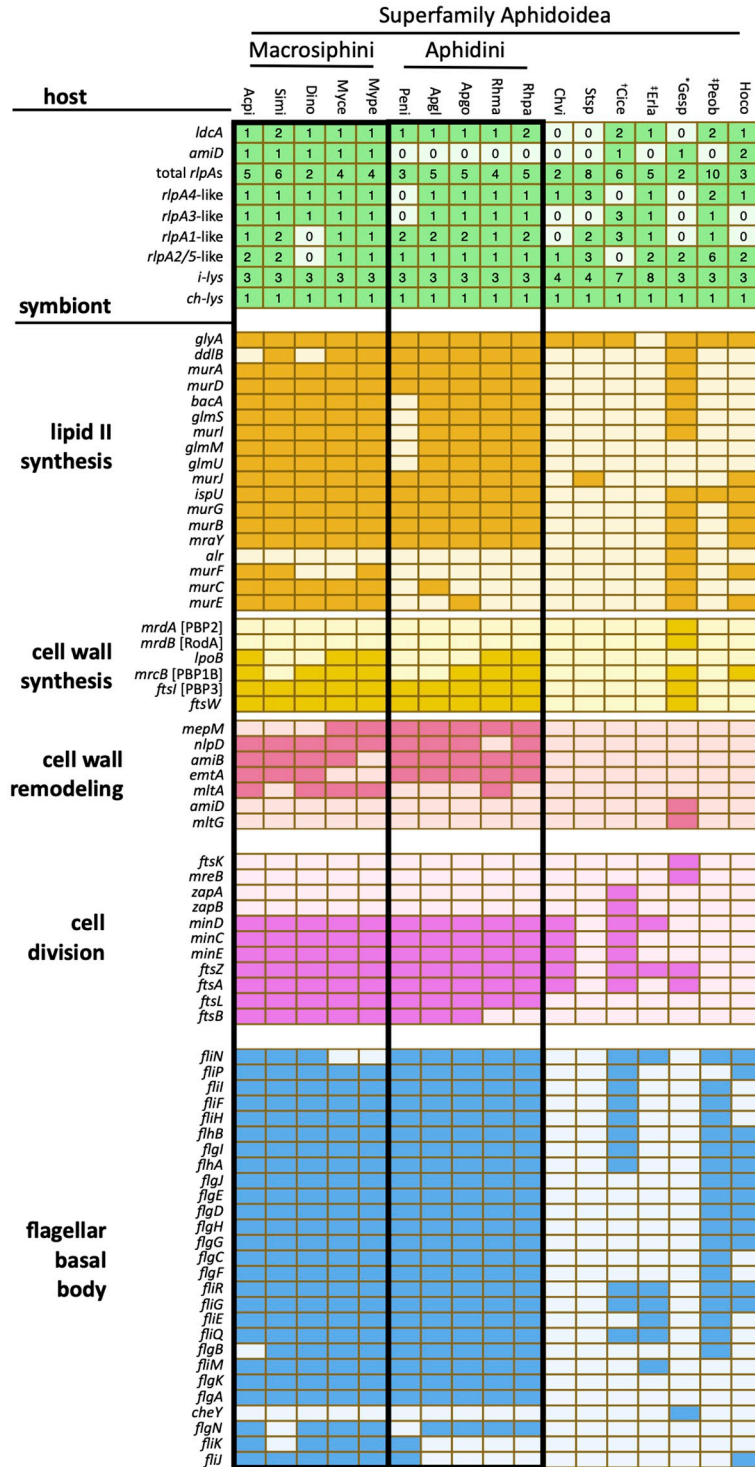


Fig 3. Incidence of genes underlying PGN metabolism and related pathways from 17 aphid host-symbiont species. Filled cells indicate gene presence, while empty cells indicate gene absence. For aphid genes, numbers indicate the number of gene paralogs identified per genome. For symbiont genes, filled cells all indicate a single copy per genome. All symbionts are *Buchnera* except those of *Geopemphigus* sp, indicated by an asterisk, which contains a novel Bacteroidetes symbiont. Host species containing known co-obligate or uncharacterized facultative symbionts are indicated with an obelisk (†) or a double obelisk (‡), respectively. Each species code represents the combined first two letters of the genus and species (S1 Table). Colored circles indicate genes with dual-membership in the pathway of the same color, while black indicates dual-membership in amino acid metabolism.

<https://doi.org/10.1371/journal.pgen.1010195.g003>

present in three aphid subfamilies with symbionts lacking PGN pathway genes (Eriosomatinae, Lachninae, and Pemphiginae) and is absent in two (Chaitophorinae, Phyllaphidinae). While the role of *ldcA* in the former lineages is unclear, *Buchnera* from each of the latter species has undergone significantly more gene loss relative to symbionts of other aphids [27]. Aphid *ldcA* is also absent in *Geopemphigus sp.* (Fordinae), where excessive genome reduction in *Buchnera* is thought to have led to its evolutionary replacement with the Bacteroidetes symbiont *Skilesia* [29]. Thus, our data suggest that the two independent losses of *ldcA* in these clades are associated with extensive loss of *Buchnera* genes, including most PGN genes (Figs 2 and 3).

In contrast, the incidence of aphid *amiD* in different lineages strongly suggests a link to symbiont cell wall metabolism. The loss of host *amiD* from most subfamilies with highly reduced *Buchnera* genomes (Chaitophorinae, Phyllaphidinae, Eriosomatinae, and Pemphiginae) or *Buchnera* loss (Fordinae) suggests that the gene is expendable in the absence of *Buchnera* PGN (Fig 3). The putative ability of AmiD to cleave stem peptides from anhydro-MurNAc suggests a dependence on the prior activity of LTs [53], which are distinct from lysozymes in their production of cyclic anhydro-MurNAc. As LTs are found only in *Buchnera* from Aphidinae aphids (*emtA* and/or *mltA*) and *Skilesia* (*mltG*), the presence of aphid *amiD* largely correlates with LT incidence, with the exception of *C. cedri* (Fig 3). In tribe Aphidini, the loss of host *amiD* coincides with that of *murF*, *murC*, and/or *murE* in symbionts—while the encoded enzymes putatively function in two distinct pathways of PGN metabolism (remodeling and lipid II synthesis, respectively), all four enzyme activities involve PGN stem peptides (Fig 1), implying the possibility of a functional relationship. Alternatively, a lineage-specific functional redundancy of host *amiD* with symbiont *amiB* could have rendered *amiD* obsolete. The presence of both aphid *amiD* and *ldcA* in *H. cornu* coincides with incomplete PGN gene loss in its symbiont (Fig 3), suggesting a stronger reliance on host HTGs for symbiont PGN remodeling. Finally, though *Buchnera* from *Cinara cedri* (Lachninae) lacks all PGN pathway genes, the host maintains both *amiD* and *ldcA*. Many Lachninae aphids harbor secondary symbionts that have taken on essential roles for host survival [54–56]—aphid PGN genes may regulate PGN metabolism in these secondary symbionts instead of *Buchnera*.

We also evaluated the symbiont gene repertoire of pathways related to but outside of PGN metabolism. Cell division is closely coordinated with cell wall synthesis and remodeling [34], while assembly of flagella basal bodies FBBs involves interaction with the cell wall via the β -N-acetylglucosaminidase activity of FlgJ [35,36]. Like the PGN pathway, loss of both cell division and FBB genes is more extensive in all subfamilies relative to the Aphidinae (Fig 3). However, there is no apparent coincidence in these losses with those of PGN pathway genes that would implicate a dependence of either pathway on the cell wall. For example, though symbionts of both *C. viminalis* (Chaitophorinae) and *Stegophylla sp.* (Phyllaphididae) lack virtually all PGN genes, the former maintains many cell division genes, similar to the tick-borne pathogens of the genus *Ehrlichia* [21], while the latter lacks all of them (Fig 3), a unique phenomenon among obligately intracellular bacteria [21,26]. Similarly, *Buchnera* of *C. cedri*, *E. lanigerum* (Eriosomatinae), and *P. obesinymphae* (Pemphiginae) lack the PGN pathway but maintain many FBB genes (Fig 3). Furthermore, cell wall synthesis and remodeling genes that are known components of the *E. coli* divisome are not preferentially maintained in aphid symbionts that lack the PGN pathway but maintain cell division genes (Fig 3) [35]. Additionally, all *Buchnera* FlgJ enzymes lack the PGN hydrolase domain, effectively demonstrating a decoupling of FBB assembly to cell wall metabolism. Thus, the *Buchnera* cell wall generally does not appear to support cell division or FBB assembly.

Aphidini *ldcA* genes are influenced by positive selection

Aphid HTGs present in species harboring symbionts undergoing gene loss in the PGN pathway or lacking it altogether may have evolved alternative functions through natural selection. Because *ldcA* is present in species with varying degrees of PGN pathway degradation, we were able to test which conditions might lead to positive selection. The most complete PGN pathways are found in tribe Macrosiphini, followed by tribe Aphidini—PGN pathways are near-absent in subfamily Hormaphidinae and entirely absent in all other subfamilies with *Buchnera*. We tested for lineage-specific diversifying selection in the aphid *ldcA* gene family using BUSTED [57], selecting different sets of test lineages based on the criteria described above. For example, we asked whether diversifying selection has influenced *ldcA* in members of the tribe Aphidini, where the loss of aphid *amiD* and *Buchnera murF*, *murC*, and/or *murE* may signal the absence or modification of stem peptides from *Buchnera* PGN in this lineage, leading to a potential shift in LdcA function, which also putatively targets stem peptides. We found evidence for positive selection only when all lineages or Aphidini-specific *ldcA* genes were selected as the test group (Table 1), suggesting that host *ldcA* is undergoing adaptive evolution in at least one member of the Aphidini. These results are especially interesting in the light of our recent biochemical characterization of *A. pisum* LdcA, which shows that the aphid enzyme is multifunctional and plays a novel role in PGN degradation relative to its homolog in *E. coli* [22]. In contrast, we did not detect evidence for positive selection in *ldcA* when test groups included: all Macrosiphini species, species with symbionts lacking PGN genes and lacking secondary symbionts (*P. obesinymphae* and *E. lanigerum*), species with intermediate PGN gene loss (*H. cornu*), and species in which *ldcA* has been duplicated (*Sitobion miscanithi*, *Rhopalosiphum padi*, *C. cedri*, and *P. obesinymphae*) (Table 1). Additionally, we were unable to detect evidence for positive selection of *amiD* when all genes or only duplicated genes were selected as test groups (Table 1). Thus, tribe Aphidini stands out in showing evidence of positive selection for *ldcA*. However, the lack of evidence for positive selection in some of our BUSTED tests may reflect lack of power due to small sample sizes.

Both Macrosiphini and Aphidini *Buchnera* build cell walls

The observation that *Buchnera* of most Aphidini aphids lack *murF*, *murC*, and/or *murE* (Fig 3), representing three critical steps in lipid II biosynthesis (Fig 1), raises the question of whether these symbionts are capable of synthesizing a cell wall at all [26]. The loss of these particular genes may be coincidental, reflecting the breakdown of cell wall synthesis altogether rather than evidence of functional linkage. To test this hypothesis, we developed a method to fluorescently label *Buchnera* PGN via incorporation of ethynyl-D-alanine (EDA), a D-alanine

Table 1. BUSTED tests for diversifying selection among aphid *ldcA* and *amiD* gene families (18 and 9 protein sequences, respectively).

gene family	test group	test group size	p-value
<i>ldcA</i>	all genes, all species	18	0.001
<i>ldcA</i>	tribe Macrosiphini	6	0.499
<i>ldcA</i>	tribe Aphidini	6	0.000
<i>ldcA</i>	subfamily Hormaphidinae	1	0.104
<i>ldcA</i>	symbionts without PGN	3	0.148
<i>ldcA</i>	genes with duplications	8	0.193
<i>amiD</i>	all genes, all species	9	0.071
<i>amiD</i>	genes with duplications	2	0.057

<https://doi.org/10.1371/journal.pgen.1010195.t001>

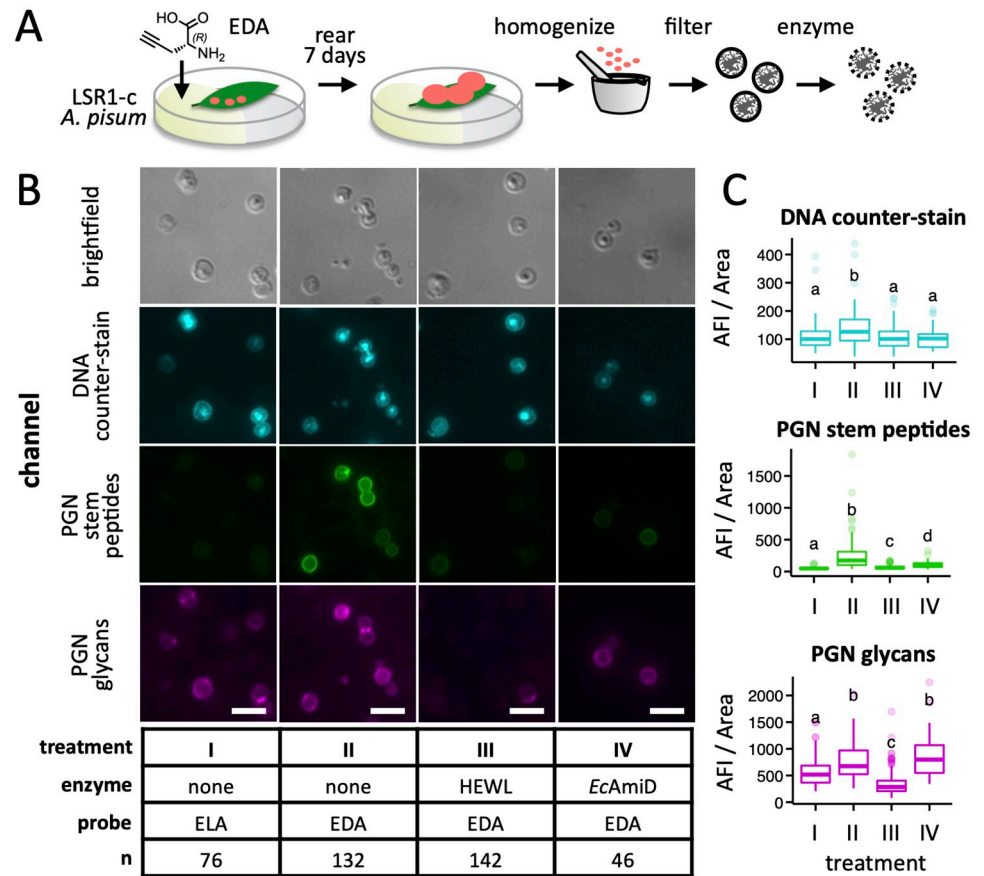


Fig 4. Labeling of *A. pisum* LSR1-c *Buchnera* cells with EDA by host feeding. A) Aphids were reared on plant leaves immobilized in agar containing ethynyl-D-Ala (EDA) or the control L enantiomer from birth until the 4th instar larval stage, after which *Buchnera* cells were purified from aphid homogenate. B) Cells were fixed and stained with azide-linked Alexa Fluor 488 to label PGN stem peptides (PGN stem peptides), wheat germ agglutinin CF-640R conjugate to label the PGN glycan backbone (PGN glycans), and the counter-stain DAPI to label DNA (DNA counter-stain). Hen eggwhite lysozyme (HEWL; PGN glycan-targeting) and *E. coli* AmiD (PGN stem peptide-targeting) were used to demonstrate incorporation of EDA within *Buchnera* PGN stem peptides. Images were falsely colored to aid visualization. The white scale bar represents 5 μ m. C) For each combination of enzyme and probe, the average fluorescence intensity (AFI) per area was measured for each of n *Buchnera* cells using ImageJ software (S9 Table). Significant differences between means are indicated by compact letter displays above each boxplot—means with no letters in common are significantly different by the Dunn's Multiple Comparison test (p-value < 0.01 following adjustment for false discovery).

<https://doi.org/10.1371/journal.pgen.1010195.g004>

analogue compatible with click chemistry [58,59], by adapting a host feeding approach previously used to label symbiont PGN in the mealybug system [17]. Aphids were reared from birth on individual leaves immobilized in agarose gel containing EDA until reaching seven days of age (Fig 4A). To validate our approach, *Buchnera* cells were isolated from pea aphids (tribe Macrosiphini) reared on EDA, fixed, and stained with azide-linked Alexa Fluor 488 (AF488) via click reaction. *Buchnera* derived from hosts fed EDA exhibited green fluorescence at the cell periphery while control aphids raised on ethynyl-L-alanine (ELA) did not (Fig 4B), indicating 1) transport of EDA from host plant xylem to phloem, and 2) selective incorporation of EDA from plant phloem into *Buchnera* cell walls. Furthermore, the observed fluorescence was definitively linked to *Buchnera* cell wall stem peptides by treating cells with PGN-degrading enzymes prior to fixation and counter-staining with the PGN backbone-specific probe wheat germ agglutinin CF-640R conjugate (WGA640). Hen eggwhite lysozyme (HEWL) and *E. coli*

AmiD target distinct PGN features: the backbone and stem peptides, respectively. We observed a decrease in fluorescence of both EDA-AF488 and WGA640 following HEWL treatment and of EDA-AF488 alone for AmiD (Fig 4B and 4C and S9 Table), firmly associating AF488 fluorescence with PGN stem peptides.

Next, we applied the same cell wall labeling strategy to *R. maidis* aphids (tribe Aphidini) and pea aphids either cured of or containing the facultative symbiont *Serratia symbiotica*. Symbiont identities were confirmed using diagnostic FISH microscopy alongside the clickable cell wall probes (Fig 5). As expected, no cell wall fluorescence was observed in symbionts fed ELA. However, *Buchnera* cell walls were labeled with EDA in both aphid species (Fig 5A), demonstrating that even species lacking aphid *amiD* and *Buchnera murC*, *murE*, and *murF* can synthesize a cell wall. Additionally, pea aphids harboring *Serratia* show only minor incorporation of EDA into *Serratia* symbionts relative to *Buchnera* (Fig 5B), suggesting that *Serratia* differs from *Buchnera* in their primary source of D-Ala.

Discussion

The HTGs *rlpA*, *ldcA*, and *amiD* of *A. pisum* encode putative PGN hydrolases that are more highly expressed in bacteriocytes than other host tissues, suggesting that they might play a role in *Buchnera* cell wall metabolism [13]. The putative functions of these HTGs collectively represent two of the three (LT and amidase) enzyme activities required for cell wall remodeling [60]. Furthermore, some LdcA homologs from intracellular bacteria are known to exhibit the third (endopeptidase) activity [61–63], implying that aphid HTGs might function cooperatively. Indeed, we recently showed that the *A. pisum* LdcA enzyme exhibits enhanced endopeptidase activity relative to its *E. coli* homolog, cleaving both a greater range and amount of PGN substrates [22]. Further, we found evidence of this activity in the composition of *Buchnera* PGN from pea aphids, demonstrating its biological relevance.

We examined the idea that aphid HTGs and *Buchnera* PGN metabolism might be functionally linked by identifying correlations between host and symbiont PGN gene repertoires based on newly sequenced genomes from basally branching aphid lineages. Aphid HTGs vary in their incidence and copy number among different aphid species, with each gene family displaying a unique distribution. While aphid *rlpA* genes are ubiquitous in aphids, the absence of aphid *ldcA* and *amiD* genes in some lineages coincides with greater loss of symbiont PGN pathway genes, suggesting functional linkage of *ldcA* and *amiD* with symbiont PGN metabolism. These results demonstrate that each aphid HTG family has an independent evolutionary trajectory, implying that they are unlikely to function cooperatively as mealybug HTGs do for symbiont PGN synthesis [17].

Despite the differences in the incidence of HTGs among aphids, the distribution of *rlpA*, *amiD*, and *ldcA* throughout the Aphidoidea superfamily and their absence from closely related Sternorrhynchan insects (Fig 2) suggests that their acquisition was a milestone in aphid evolution. The *Buchnera* symbiont of the most recent aphid ancestor is predicted to have had functional lipid II, cell wall synthesis, and cell wall remodeling pathways, despite evidence for significant genome reduction [27]. Thus, host HTGs could have provided an immediate benefit following their acquisition by influencing symbiont PGN metabolism (Fig 1). Aphid *ldcA* and *amiD* may continue to serve this purpose in species for which symbiont PGN pathways are maintained (Aphidinae, Hormaphidinae). Outside of these subfamilies, most symbiont PGN genes were lost, along with many other non-essential genes, due to the elevated genetic drift characteristic of endosymbionts (Fig 3) [64]. Following the loss of PGN metabolism in symbionts, HTGs were not needed and therefore lost (*amiD* in Chaitophorinae, Phyllaphidinae, Eriosomatinae, and Pemphiginae, and *ldcA* in the first two), repurposed for interaction

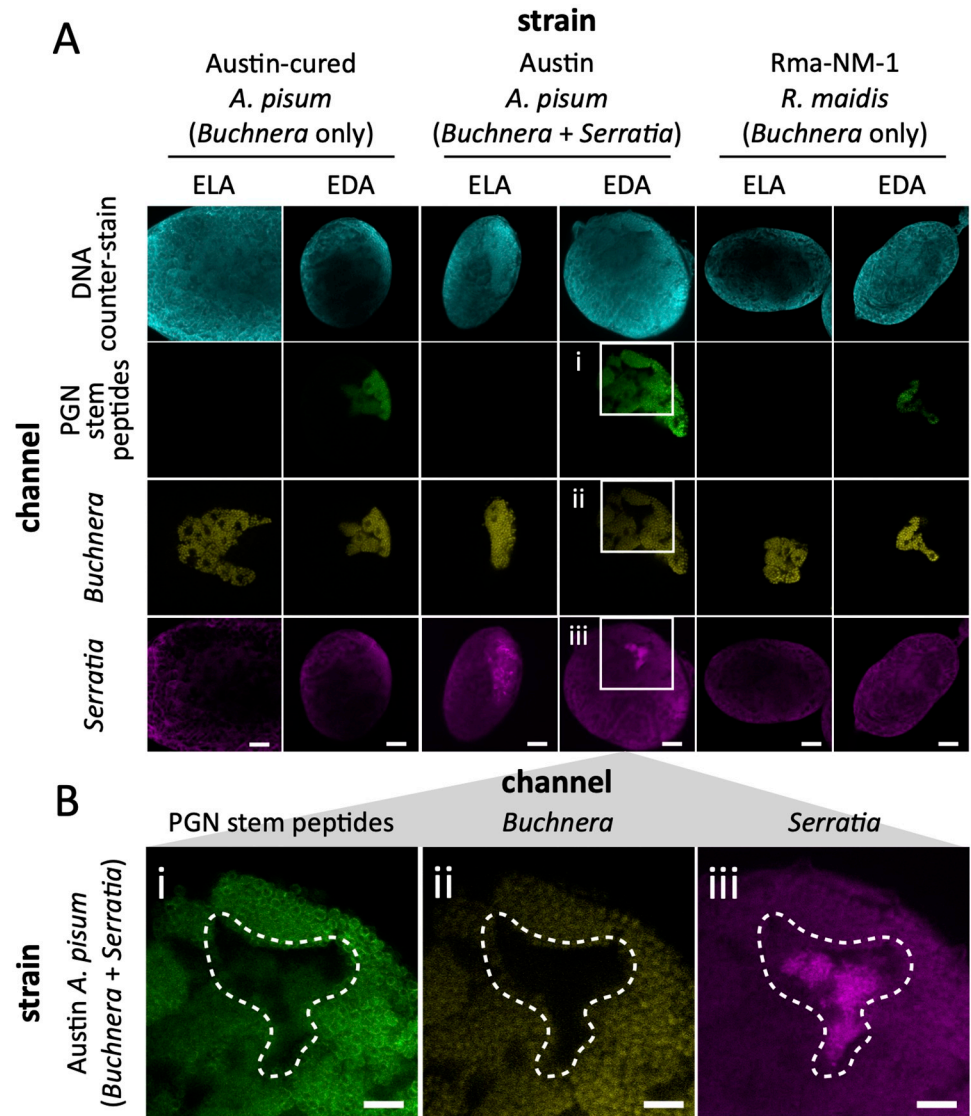


Fig 5. FISH and cell wall labeling of Macrosiphini and Aphidini aphid symbionts. A) Pea aphids, both cured (Austin-c; left column) and uncured (Austin; middle column) of secondary *Serratia* symbionts, and *R. maidis* (Rma-NM-1; right column) were fed ELA (left-side) or EDA (right-side) via their host plants as described in Fig 4A. Embryos removed from 4th instar larvae were fixed, treated with azide-linked Alexa Fluor 488 (AF488), hybridized with *Buchnera*- and *Serratia*-specific Cy5- and Cy3-labeled riboprobes, respectively (S10 Table), and counter-stained with DAPI. B) Enlarged cutout images (white boxes in A) of EDA-labeled, uncured *A. pisum* show overlapping AF488 fluorescence (i) with *Buchnera* cells (ii) and, to a minor extent, *Serratia* cells (iii). A bacteriocyte containing *Serratia* is highlighted with a dashed line in each image. The white scale bars represent 20 μ m. Images were falsely colored to aid visualization.

<https://doi.org/10.1371/journal.pgen.1010195.g005>

with non-*Buchnera* symbionts (Lachninae, Fordinae), or maintained by purifying selection (*ldcA* in Eriosomatinae and Pemphiginae). These cases are discussed further below.

In the Lachninae aphids and in *Geopemphigus* sp., aphid *ldcA* and *amiD* may contribute to PGN metabolism in symbionts other than *Buchnera*. In *Geopemphigus* sp., *Buchnera* has been completely replaced with *Skilesia* [29]. It is possible that the *Geopemphigus amiD* gene contributes to PGN metabolism in *Skilesia*, while the *ldcA* gene could have been lost concomitantly with loss of *Buchnera* by the *Geopemphigus* ancestor. In contrast, *C. cedri* harbors both

Buchnera and *Serratia* [54]. The *Buchnera* symbionts of Lachninae aphids exhibit significant genome reduction relative to other subfamilies [54–56], including the loss of PGN pathway genes [31]. The gain of a co-obligate symbiont before the complete loss of *Buchnera* peptidoglycan may have prevented the loss of either aphid *ldcA* or *amiD*. Host HTGs may contribute to unconfirmed secondary or co-obligate symbionts in other species harboring *Buchnera* strains lacking the PGN pathway (Eriosomatinae and Pemphiginae). Reads from *Serratia* and *Pantoea* bacteria were found among the *E. lanigerum* and *P. obesinymphae* sequencing data, respectively [33]. Though both genera include many insect-associated bacteria [65,66], a mutualistic relationship between host and bacteria cannot be inferred from sequencing alone. Alternatively, host HTGs in lineages that lack symbiont PGN may have evolved functions that differ from their putative original roles in symbiont PGN metabolism (Eriosomatinae and Pemphiginae). Although we did not find evidence of diversifying selection for *ldcA* in these lineages (Table 1), their presence in some species lacking symbiont PGN and absence in others suggests that they remain functional.

Gene duplication is often accompanied by selection for divergence in function, although we did not detect evidence for positive selection associated with gene duplication in *ldcA* (Table 1). Phylogenetic analysis of aphid RlpA proteins reveals four groups of paralogs, each distributed throughout all aphid lineages (S2 Fig). Each paralog group exhibits a distinct phylogenetic profile and none coincides with the loss of symbiont PGN pathway genes, implying 1) distinctive functions of each paralog group, and 2) a lack of functional interaction between RlpAs and PGN metabolism. Because each aphid *rlpA* gene has eukaryotic domains with putative functions unrelated to PGN metabolism, the aphid *rlpA* gene family may have never been involved in symbiont cell wall metabolism. Knottin domains, characterized by interwoven β -sheets stabilized by disulfide bonds, represent the conserved scaffold of a diverse range of target-selective proteins found in multiple eukaryotic kingdoms, many of which are components of venom [67]. Most knottins, including Btk1-4 in whiteflies, are encoded as small, single-domain proteins [68]. Aphids do not encode any knottin domains outside of RlpA, suggesting that the RlpA DPBB domain could have been horizontally transferred into the coding sequence of an ancestral, single-domain knottin gene. The role of knottin-RlpA proteins in the aphid-*Buchnera* symbiosis may derive from the function of their knottin domains, which are involved in immunity in whiteflies [69–71], instead of from the DPBB domain.

The question remains as to why the *Buchnera* cell wall is maintained in some species and not others. We show that cell division and FBB assembly genes, both of which interact with the cell wall in *E. coli* [35,37], are not coincident with the PGN pathway in *Buchnera* and are thus unlikely to be functionally co-dependent (Fig 3). Additionally, many intracellular environments are isotonic with the symbiont cytoplasm, eliminating the dependence on the cell wall to maintain turgor pressure [21]. Furthermore, eukaryotes typically recognize PGN via PGRPs [40], which all aphids lack (Fig 2) [45]. Our results suggest that the retention of the *Buchnera* cell wall is related to the presence of aphid *amiD* and *ldcA*—species that are able influence symbiont PGN metabolism are likely better able to control or support their symbionts, maximizing the benefits and reducing the costs associated with symbiosis. In many aphid lineages, genomic degradation in *Buchnera* has resulted in the loss of the cell wall, and this loss may eliminate the ability of hosts to regulate symbionts through the actions of AmiD and LdcA. Under this scenario, the selective forces maintaining aphid *amiD* and *ldcA* are dependent on the presence of a cell wall in symbionts, and the function of the symbiont cell wall is to enable hosts to control and/or support symbionts.

In further support of this hypothesis, we observed cell walls that fully envelope *Buchnera* cells in *R. maidis* (Fig 5) despite predictions based on gene repertoires that cell wall synthesis is not possible (Fig 3) [26]. Our experiments demonstrate the functionality of the remaining

PGN pathway genes. Thus, the concomitant loss of *Buchnera murF*, *murC*, and *murE* and host *amiD*, implies these genes are functionally related. As *Buchnera* genes underwent elimination due to drift, *AmiD* became useless and was subsequently lost. In Aphidini, *ldcA* genes may have enhanced importance for regulating symbiosis, as Aphidini *ldcA* could represent the sole host modulator of *Buchnera* PGN. The preservation of the cell wall despite loss of multiple genes related to the PGN stem peptide also implies the existence of compensatory mechanisms for stem peptide synthesis. The stem peptide composition of Aphidini *Buchnera* PGN may differ from that of typical Gram-negative bacteria, prompting evolutionary changes in *LdcA* substrate selectivity (Table 1). Furthermore, the only amino acid ligase remaining to Aphidini species is *MurF*, suggesting that this enzyme may have expanded its role to ligate a wide range of amino acids together to construct the stem peptide. Multifunctional enzymes likely represent a widespread adaptation to genome reduction in bacteria [72,73], and two such enzymes have been described in pea aphid *Buchnera* [22,74]. However, Aphidini *Buchnera* species maintain most genes necessary to synthesize or incorporate each of the canonical stem peptide amino acids: *murI* (missing only in *Pentalonia nigonervosa*) produces D-Glu, *ddlB* synthesizes D-Ala-D-Ala, *ftsI* encodes PBP3, which forms crosslinkages involving *mDap*, and *mepM* is an endopeptidase that cleaves those crosslinkages (Fig 3). Thus, it is unclear whether PGN structure differs between Aphidini *Buchnera* and other Gram-negative bacteria. Experimentation is needed to test conclusions drawn from genetics alone.

To date, hypotheses about aphid HTG function are informed by studies pertaining to a single aphid species, *A. pisum* (tribe Macrosiphini). In this system, HTGs are expressed more highly in bacteriocytes than in other tissues [13]. Further, RNAi knockdown [75] and correlations of host-symbiont gene expression [76] both suggest that *amiD* and *ldcA* expression positively influence *Buchnera* abundance. Finally, RlpA4 is localized to *Buchnera* [77]. Studies utilizing different aphid species, especially from tribe Aphidini and subfamily Hormaphidinae, which harbor symbionts with partial PGN pathways, are needed to establish whether the host-symbiont interactions observed for *A. pisum* are conserved in other species, especially when the potential for species-specific adaptations is high. Furthermore, sequencing of aphid-symbiont genomes from additional species within each subfamily would enable the use of statistical tools for detecting patterns in gene repertoire, whereas our analysis was limited because only a single representative species is available for most subfamilies. For example, our conclusions about the differences between the Macrosiphini and Aphidini, for which genomes are available from multiple species, are more robust than those pertaining to more basally branching but less deeply sampled lineages.

In conclusion, cell wall structure and HTG function appear to vary even among closely related aphid species, such that the relationship between aphid HTGs and symbiont PGN metabolism appears to be unique for each species. Furthermore, we emphasize the importance of experimental validation of genomics-based hypotheses, as comparative genomics alone is insufficient to predict the functional capabilities of aphids and their symbionts.

Materials and methods

Aphid collection

Aphids were collected from multiple locations between June, 2020 and May, 2021 (S1 Table). Aphids were removed from leaves or galls, flash frozen in liquid nitrogen, and stored at -80°C. Excess wax was removed from *Geopemphigus* aphids by dipping individuals in Dawn Professional Grade Liquid Detergent (0.1% v/v), then rinsing with deionized water. Multiple individuals regardless of developmental stage were pooled per sample, with each *Geopemphigus* and *Pemphigus* sample derived from a single gall, and thus expected to contain members of a single

all female clonal lineage. *Geopemphigus* sp. is an undescribed species corresponding to the “straight galls” found on *Pistacia texana* [29].

Sequencing and annotation

Genomic DNA was extracted from aphid samples using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturers’ protocol for DNA purification from insects using a mortar and pestle and adding RNase cocktail (5 μ l; Invitrogen) along with proteinase K. DNA concentration was measured using the Qubit dsDNA BR Assay Kit (ThermoFisher). To improve the assembly and gene annotation, we also extracted RNA from the four species. For *Geopemphigus* sp. and *Stegophylla* sp. samples, RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) following the modifications described by Smith and Moran [75]. Genomic DNA was removed from RNA samples using the RQ1 RNase-free DNase Kit (Promega) and repurified using the RNeasy MiniElute Cleanup Kit (Qiagen). For *P. obesinymphae* and *C. viminalis* samples, RNA was extracted using the Quick-RNA MiniPrep Kit (Zymo Research) according to the manufacturers’ protocol. RNA concentration was measured using the Qubit ssRNA BR Assay Kit (ThermoFisher). Whole genome DNA (350 bp insert size) and poly-A enrichment mRNA libraries were prepared and sequenced on an Illumina NovaSeq 6000 instrument (2x150) by Novogene Corporation Inc. For *Geopemphigus* sp., we downloaded the already available genome sequencing data via NCBI under NCBI SRA accession number SRR6849201 [29]. For *H. cornu*, we manually translated the available nucleotide sequences of all protein-coding genes (S1 Table).

Our bioinformatics workflow is summarized in S3 Fig. First, we removed low quality sequences and sequencing adaptors from raw reads for both DNA and RNA sequencing data using Trimmomatic version 0.38 [78]. We first assembled the genomes with reads from DNA sequencing data using SPADes version 3.14.1 [79] with kmer sizes (-k) = 21, 33, 55, 77, 99, 127. Then we further improved the genome continuity by using the paired-end information from RNA-seq data. We mapped RNA-seq data to SPADes scaffolds using HISAT2 version 2.2.1 [80] and scaffolded the SPADes assemblies using P_RNA_scaffolder [81] with default settings.

To remove potential contamination and retrieve symbiont genomes, we first aligned scaffolds to NCBI NT database (last accessed January 27, 2021) using BLASTN megablast with “-culling_limit 5 -evalue 1e-25” in BLAST+ version 2.11.0 [82]. Then we mapped DNA sequencing data to the assemblies to estimate sequencing depth using bwa version 0.7.17 [83]. We created blobplots based on the BLAST results and BAM files using BlobTools version 1.1.1 [84]. Sequences that were annotated as human or bacterial were removed from downstream analysis. Sequences that were annotated as *Buchnera*, *Bacteroidetes*, *Rhizobiales*, *Rickettsiales*, or *Pantoea* were extracted as the symbiont genomes.

Lastly, as we detected potential contaminations from *Pemphigus* and ants in the *Chaitophorus* genome assembly (reflecting the origin of these samples from field sites where these organisms coexisted), we aligned *Chaitophorus* scaffolds to *Pemphigus* scaffolds using BLAST with e-value = 1e-10. *Chaitophorus* scaffolds were removed if 1) the scaffolds were \geq 1000 bp and have at least 1000 bp matches with 95% identity to *Pemphigus* scaffolds, or 2) the scaffolds were < 1000bp and showed at least 95% identity with any length mapped to *Pemphigus* scaffolds. Scaffolds assigned as Hymenoptera by BlobTools were also removed from downstream analysis. We evaluated the completeness of genomes using BUSCO version 4.1.4 [38] with the Hemiptera gene set (n = 2,510).

For gene annotation, we first soft-masked the assemblies using Tantan [85] in Funannotate version 1.8.8 [86]. We used BRAKER2 version 2.1.5 [87,88] with both RNA evidence and

protein evidence from closely related species (—etp mode). For RNA evidence, we mapped RNA-seq data to assemblies using HISAT2. For protein evidence, we downloaded all annotated proteins from aphid genomes on RefSeq and the gene annotation of *E. lanigerum* [33], as it is the closest species available. Lastly, we evaluated the quality of gene annotations using BUSCO Hemiptera gene sets.

Host phylogenetic analysis

To construct the host phylogeny, we first extracted the longest isoform for each annotated gene across 23 species. We assigned genes to orthologous groups using OrthoFinder version 2.5.2 [89] with default parameters. Orthologous groups exist in at least 80% of the species were used for phylogeny. For each orthologous group, we aligned amino acid sequences using MAFFT version 7.407 L-INS-i algorithm [90] and removed gappy regions using Gblocks version 0.91b [91] with default parameters. Lastly, we concatenated the resulting alignments and constructed phylogenetic trees using IQ-Tree version 2.0.6 [92] with 1,000 ultrafast bootstrap replicates. *Bemisia tabaci*, *Diaphorina citri*, and *Pachypsylla venusta* were used as outgroups.

Insect PGN pathway genes were identified using BLASTP against a custom database of insect PGN genes, including aphid HTGs. To search for *bLys* genes, the ch-type lysozyme domain of *A. pisum* *bLys* was used as a query in a TBLASTN search of all insect genomes. For each search with the exception of *RlpA*, hits were aligned using MUSCLE [93]. For *RlpA*, hits from the search of each *A. pisum* paralog were pooled, manually dereplicated, aligned with MUSCLE, and then manually adjusted based largely on the position of Cys residues in the ICK domains. Alignments were then trimmed using trimAl [94]. Gene trees were constructed for *AmiD* and *RlpA* protein sequence alignments using IQ-Tree 2 [92] with Transfer Bootstrap Estimates [95], with VTR4 and VTR5 models selected for *amiD* (S1 Fig) and *rlpA* (S2 Fig) gene families using ModelFinder [96]. Gene incidence was determined by counting the number of sequences per species in each alignment following manual inspection of alignment quality and sequence integrity.

We constructed a gain-loss-duplication model to explore the origin and fate of PGN pathway genes, including HTGs, lysozymes, and PGRPs, among aphids and close relatives. Using Count [97], we first performed rate optimizations for gene gain, loss, and duplication using flexible gain-loss and duplication-loss ratios for all lineages and a Poisson distribution at the tree root. Next, we used Dollo parsimony to determine the positions of gene gains and losses within the aphid phylogeny. Fig 2 incorporates the results of the reconstruction for each gene family with the exception of *rlpA3*- and *rlpA1*-like genes—since *H. cornu* branches at the base of Aphidoidea, parsimony maps the acquisition of these genes after the divergence of *H. cornu* from other aphids, but a phylogenetic analysis of *RlpA* protein sequences places *RlpA3*- and *RlpA1*-like paralogs as basal to *RlpA2/5*-like paralogs, suggesting that *H. cornu* lost the ancestral *rlpA3*- and *rlpA1*-like genes following the acquisition of *rlpA2/5*-like genes by gene duplication (S2 Fig).

Analyses of *Buchnera* genomes

Sixty-nine bacterial genomes, including 51 *Buchnera*, 11 *S. symbiotica*, *Skilesia*, and five out-group genomes, were used to identify orthologous gene groups belonging to the PGN and related pathways (S11 Table). As the assembled *Buchnera* genome sequence of *Sitobion miscanthi* was not available, we obtained its genome sequencing data from NCBI SRA (accession number: SRX5767526) [98], cleaned reads using Trimmomatic and assembled the *Buchnera* genome using SPADES under the same parameters described above. All other genomes are available online (S11 Table). First, all genomes were annotated using DFAST version 1.2.13

[99]. Coding sequences containing internal stop codons or frameshifts, likely representing pseudogenes, were removed from annotation files using a custom bash script (available at https://github.com/lyy005/peptidoglycan_related_genes_in_basal_aphids). Second, orthologous groups of DFAST-annotated bacterial proteins were determined using OrthoFinder version 2.5.2 [89]. The inflation parameter was set at 10 to account for the high phylogenetic distance among selected bacterial species. Seven proteomes downloaded from EnsemblBacteria, including *Bacteriodes fragilis* YCH46 (ASM992v1), *Escherichia coli* MG1655 (ASM584v2), *Buchnera* from *A. pisum* (ASM960v1), *B. pistaciae* (ASM772v1), and *Schizaphis graminum* (ASM736v1), and *S. symbiotica* from *A. pisum* str. Tucson (ASM18648v1), were included to facilitate retrieval of the gene names represented by each orthogroup. A SmartTable containing all *E. coli* MG1655 gene names and synonyms was downloaded from EcoCyc.org [100] and a custom R script, written with R version 3.6.1 [101] and operated in RStudio version 1.1.463 [102], was used to 1) ensure uniform gene nomenclature between *E. coli* and *Buchnera*, 2) identify the incidence of 138 genes involved in PGN metabolism or related pathways (S8 Table), and 3) count the number of genes per orthogroup per species (available at https://github.com/lyy005/peptidoglycan_related_genes_in_basal_aphids). The heatmaps shown in Fig 3 were constructed with the pheatmap package [103] using the same R script.

Test of positive selection

To test whether certain aphid HTGs are under positive selection, we prepared codon alignments of each HTG protein family that match the protein alignments described above using PAL2NAL [104]. We then employed the BUSTED method (57) to test for positive selection through the Datamonkey web server [105].

Aphid rearing for symbiont cell wall labeling

Clonal lines of *A. pisum*, including the parental lines LSR1-c and Austin-c, previously cured of secondary symbionts (Austin-cured) [106], as well as Austin aphids uncured of secondary symbionts (Austin), were reared on broad bean (*Vicia faba*) at 20°C constant temperature with a 14L/10D daily light cycle. The clonal *R. maidis* line Rma-NM-1, collected from *Sorghum halepense* in Austin, Texas in 2017, was reared on barley (*Hordeum vulgare*) under the same temperature and light conditions as *A. pisum*. To incorporate D-alanine probes into *Buchnera* cell walls, aphids of either species were reared on leaves of their respective host plants immersed in a probe-agar solution. Individual leaves were immobilized in liquid 1.5% agar containing 35.4 mM of either EDA (D-proargylglycine; Alfa Aesar) or ELA probe (L-propargylglycine; Santa Cruz Biotechnology) within petri dishes (50 x 10 mm) and incubated for 24 hours. Two adult aphids (≥ 10 days old) were placed on each leaf and allowed to reproduce for 24 hours. Adults were removed and offspring reared for 3–4 days on the same leaves before transferring aphids to fresh probe-fed leaves and rearing for another 3–4 days. Seven-day-old aphids were used for all downstream experiments.

Purification of cell wall-labeled *Buchnera*

Seven-day-old *A. pisum* nymphs from six leaves per treatment were transferred to 1.5 ml microfuge tubes, surface-sterilized in bleach (500 μ l; 0.5%) for 1 minute, rinsed twice in deionized water (500 μ l), and homogenized in buffer A (300 μ l; 25 mM KCl, 35 mM Tris, 10 mM MgCl₂, 250 mM sucrose, 250 mM EDTA, pH = 7.5) by mortar and pestle. Homogenates were passed through Sterile Swinnex filter holders (13 mm, EMD Millipore) loaded with UV-irradiated 100 μ m nylon mesh filters (EMD Millipore) cut into 1/2 inch diameter circles using a

leather hollow punch. Mortar tubes were rinsed with additional buffer A (300 μ l) and the rinses filtered and pooled with the initial filtrates. Pooled filtrates were centrifuged for 20 min at 4000 x g, 25°C, and the pellets resuspended in buffer A and filtered again using a 5 μ m nylon mesh filters. Filtrates were centrifuged the pellets washed with enzyme reaction buffer (300 μ l; 0.1 M NaCl, 25 mM Tris, pH 7.5). Cells were centrifuged once more, the pellets resuspended in enzyme solution (50 μ l; 1 μ M enzyme, 0.1 M NaCl, 25 mM Tris, pH 7.5), and cell suspensions incubated at 37°C overnight in a thermocycler with the lid set to 60°C. Recombinant *E. coli* AmiD was recombinantly expressed (see below) while HEWL was purchased (EMD). Treated cells were centrifuged and washed twice with phosphate-buffered saline (PBS; 300 μ l).

To fix and stain *Buchnera* cells, fixative (60 μ l; 16% paraformaldehyde, 0.05% glutaraldehyde) was prepared and aliquoted to fresh tubes, to which sodium phosphate was added (12 μ l; 1 M, pH = 7.4) immediately followed by cell suspension (300 μ l). The mixture was inverted twice and incubated at room temperature for 15 min, then on ice for 30 min. Fixed cells were centrifuged at 10,000 x g for 10 min, rinsed twice with PBS (300 μ l), permeabilized in PBS + 0.5% Triton X-100 (300 μ l) for five min, and rinsed twice more with PBS. Permeabilized cells were blocked with PBS containing 3% bovine serum albumin (BSA; 300 μ l) at room temperature for 1 hour, then centrifuged. Cells were reacted with AF488 azide using the Click-iT Cell Reaction Buffer kit in 200 μ l volumes of reaction mixture according to the manufacturer's specifications (Invitrogen). After washing reacted cells with BSA, cells were suspended in PBS (300 μ l) and counter-stained with WGA640R (25 μ g/ml) and DAPI (10 μ g/ml) for 15 min in darkness. Stained cells were washed once more with PBS and suspended in 10 μ l PBS, then mounted onto glass microscope slides and coverslips sealed using clear nail polish. Microscopy images were acquired using a Nikon Eclipse TE2000-U inverted microscope equipped with a Hamamatsu ORCA-Flash4.0 V2 camera and Elements software (Nikon; version 4.50.00). Cells were visualized under brightfield (100 ms), GFP (λ_{ex} = 470 nm, λ_{em} = 535 nm, exposure time = 2 s), Cy5 (λ_{ex} = 640 nm, λ_{em} = 700 nm, exposure time = 500 ms), and DAPI (λ_{ex} = 395 nm, λ_{em} = 460 nm, exposure time = 2 s) channels. Fluorescence images were falsely colored green, magenta, and cyan for these channels, respectively. Several multi-channel images were recorded for each sample.

To quantify the effects of enzyme treatment of cellular fluorescence of each fluorescent probe, images were analyzed using ImageJ software [107]. For each image, regions of interest (ROIs) were established around individual cells visualized in the DAPI channel using the free-hand selection tool. The area and mean fluorescence across each of the four channels was calculated for each ROI using the "Multi Measure" function within the ROI Manager tool and exported as csv files. Measurements for each image were uploaded to RStudio [102] and analyzed using custom R scripts written with tidyverse package functions [108]. The mean fluorescence intensity per area was calculated for the GFP, Cy5, and DAPI channels, and statistical comparisons made for each using the dunn.test package (S9 Table) [109], which performs a Kruskal-Wallis test followed by the Dunn test to extract significant pairwise differences (p-value < 0.01 after correction for false-discovery rate). Plots were generated using ggplot2 [110] and compact letter displays added using the cldList function from the rcompanion package [111]. Custom scripts are available at GitHub (https://github.com/lyy005/peptidoglycan_related_genes_in_basal_aphids).

Cell-wall labeling and FISH microscopy in aphid embryos

FISH microscopy largely followed the methods of Koga *et al.* [112]. Embryos were dissected from several seven-day-old, probe-fed Austin *A. pisum* and *R. maidis* aphids in buffer A and

temporarily stored in 70% ethanol before fixing in Carnoy's solution (5 ml; ethanol:chloroform:acetic acid at a 6:3:1 ratio) for ≥ 30 min at room temperature. Fixed embryos were then washed once with 70% ethanol (5 ml) and thrice with 100% ethanol (5 ml) for 5 min per wash with gentle agitation, then bleached overnight in 1% H₂O₂ in ethanol. Bleached embryos were then washed thrice with 100% ethanol and stored at -20°C in ethanol. For hybridization with FISH probes, embryos were washed thrice with PBST (PBS + 0.2% Tween-20; 5 ml each), twice with hybridization buffer (20 mM Tris-HCl (pH 8.0) 0.9 M NaCl, 0.01% SDS, 30% v/v deionized formamide; 5 ml each), then incubated in hybridization buffer containing FISH probes (100 nM per Cy5-ApisP2A and Cy3-PASSisR 16S rRNA probe (S10 Table), 20 ng/ml DAPI; 3 ml) in darkness at room temperature overnight with gentle agitation. For the click reaction, hybridized embryos were washed twice with PBS + 3% BSA (1 ml), then blocked in PBS + 3% BSA (1 ml) for 60 min at room temperature with gentle agitation before replacing the solution with Click-iT reaction containing AF488 azide (500 μ l) as described previously. Reacted embryos were washed twice with PBS (1 ml) and gentle agitation, then mounted onto glass depression slides with a drop of SlowFade Diamond Antifade Mountant (Molecular Probes) and fixed with coverslips sealed with clear nail polish. Images in Fig 5 were recorded on a Zeiss LSM 710 confocal microscope.

Construction of *E. coli* AmiD protein expression vector

The *E. coli* *amiD* gene was cloned into the pET-28b bacterial expression vector as an N-terminally His-tagged coding sequence identical to that used by Uehara *et al.* [53]. *E. coli* DH5 α genomic DNA was purified using the DNeasy Blood and Tissue kit following the protocol for Gram-negative bacteria (Qiagen). PCR amplification of *amiD* was performed with Phusion DNA polymerase (NEB), primers 1 and 2 (S12 Table), 1:200 volumes of genomic DNA template per 10 μ l reaction, and the following thermocycler program: an initial denaturation (98°C, 1 min) followed by 35 cycles of denaturation (98°C, 15s), annealing (61–66°C, 15 s), and elongation (72°C, 35s) and a final elongation step (72°C, 5 min). Amplified DNA was purified using the QIAquick PCR Purification kit (Qiagen), and digested with NdeI-HF and XhoI restriction enzymes (NEB). The pET-28b vector was digested with the same restriction enzymes and dephosphorylated using Antarctic Dephosphorylase (NEB). Digested *amiD* insert was ligated into vector by incubating with T4 DNA ligase (NEB) at 4°C overnight. The ligation reaction was transformed into chemically competent DH5 α cells, plated on LB agar plates with kanamycin (50 μ g/ml), and grown at 37°C overnight. Colonies were screened for the expected insert size by colony PCR using Taq Polymerase (ThermoFisher) with primers 3 and 4 (S12 Table) and the following thermocycler program: an initial denaturation (94°C, 3 min) followed by 30 cycles of denaturation (94°C, 30s), annealing (46°C, 30s), and elongation (72°C, 80s), and a final elongation step (72°C, 7 min). Positive colonies were grown in LB liquid culture (10 ml) plus kanamycin (50 μ g/ml) at 37°C, 220 rpm overnight and plasmids purified using the QIAprep Spin Miniprep kit (Qiagen). Purified plasmids were verified by Sanger sequencing using primers 3 and 4 (S12 Table).

Production and purification of *E. coli* AmiD enzyme

Chemically competent *E. coli* Rosetta (DE3) cells were transformed with 1 μ l of purified plasmid, plated on LB agar plates with kanamycin (50 μ g/ml) and chloramphenicol (25 μ g/ml), and grown at 37°C overnight. For each strain, six colonies were used to inoculate six wells of a deep 24-well plate containing LB (6 ml) plus kanamycin (50 μ g/ml) and chloramphenicol (25 μ g/ml), and the plate shaken at 37°C, 220 rpm overnight. Wells were pooled and the mixture (5 ml) used to inoculate an LB culture (1L) containing kanamycin (50 μ g/ml) and

chloramphenicol (25 µg/ml) in an Erlenmeyer flask (2.8 L). The culture was shaken at 37°C, 200 rpm until an OD₆₀₀ of 0.4–0.6 was reached, upon which the cultures were chilled to 18°C, induced with isopropyl β-D-1-thiogalactopyranoside (IPTG, 1 mM), and shaken overnight at 18°C. Cells were harvested by centrifugation at 4000 rpm, 4°C for 30 min and the pellets frozen at -80°C until purification.

To purify proteins, cell pellets (from 500 ml culture) were thawed on ice and suspended in lysis buffer (35 ml; 0.5 M NaCl, 25 mM imidazole, 10% glycerol, pH 8.0, with freshly added 0.4 mg/ml HEWL, 1 mM *N*-methyl-*p*-toluenesulfonamide, 1 mM MgCl₂, and 10 µg/ml DNaseI). Cell suspensions were rocked at 4°C for 1 hour, vortexed, and homogenized using a Emulsi-Flex-C3 high pressure homogenizer. Lysates were centrifuged at 13,000 rpm, 4°C for 30 min and the supernatants decanted onto Ni-NTA resin (2ml; Qiagen) pre-equilibrated with lysis buffer. The lysate-resin slurry was gently mixed and the flow-through eluted. The resin was washed twice with lysis buffer (25 ml), twice with elution buffer (25 ml; 0.1 M NaCl, 5 mM HEPES, 25 mM imidazole, 10% glycerol, pH 8.0), and once with each of the following elution buffers of increasing imidazole: 50 mM, 100 mM, 200 mM, and 500 mM imidazole (5 ml). Fractions were analyzed by SDS-PAGE (S4 Fig), and those containing proteins of interest were pooled, packed into 3500 MWCO SnakeSkin Dialysis Tubing (Thermo), and dialyzed at 4°C over the course of two days with four exchanges of dialysis buffer (1 L; 0.1 M NaCl, 25 mM Tris, 10% glycerol, pH 7.5, with 5 mM 2-mercaptoethanol added to the first two exchanges). Dialyzed proteins were concentrated using 10 kDa MWCO Amicon Ultra-15 centrifugal filters (EMD Millipore). Protein concentrations were determined from UV absorbance ($\lambda = 280$ nm) measured on a NanoDrop Lite spectrophotometer (Thermo) using fully reduced extinction coefficients calculated using the ExPasy ProtParam online tool (Swiss Institute of Bioinformatics).

Supporting information

S1 Table. Insect species included in this study and their corresponding genome accession numbers, sources for genome annotations, and references.

(XLSX)

S2 Table. Sequence statistics of genome sequencing data.

(DOCX)

S3 Table. Sequence statistics of RNA-seq data.

(DOCX)

S4 Table. Statistics of genome assemblies.

(DOCX)

S5 Table. BUSCO analysis of genome assemblies based on the Hemiptera gene set (n = 2,510).

(DOCX)

S6 Table. BUSCO analysis of gene annotations based on the Hemiptera gene set (n = 2,510).

(DOCX)

S7 Table. Insect genes with potential roles in symbiont PGN metabolism. Genes were identified by BLAST search using *A. pisum* or outgroup sequences as queries. Accession numbers pertain to those of published genomes or annotations, listed in S1 Table, or those published with the present article.

(XLSX)

S8 Table. Manually-curated database of bacterial genes involved in peptidoglycan metabolism and related pathways. Genes were identified using Orthofinder on 69 bacterial genomes. Accession numbers refer to DFAST-annotated genomes available at GitHub (https://github.com/lyy005/peptidoglycan_related_genes_in_basal_aphids).
(XLSX)

S9 Table. Raw ImageJ data used as input to determine the average fluorescence intensity (AFI) per area measured for each of n *Buchnera* cells in Fig 4C.
(XLSX)

S10 Table. Fluorescent riboprobes used in FISH microscopy experiments (Fig 5).
(DOCX)

S11 Table. Bacterial species included in this study and their corresponding genome accession numbers and references. Species in bold represent those that appear in Fig 3, for which the genome of the host aphid is also available. Annotations are available at Zenodo (<https://doi.org/10.5281/zenodo.5484414>).
(XLSX)

S12 Table. Primers used to construct plasmid for recombinant *E. coli* AmiD expression. T_m values were calculated for the specific DNA polymerase used for each primer set. Sequence in bold denotes restriction enzyme recognition sites, while underlined sequence highlights the target sequence used to calculate the T_m.
(DOCX)

S1 Fig. Independent horizontal acquisition of bacterial *amiD* genes by the common ancestor of aphids and the psyllid *D. citri*. An alignment of 256 amino acid residues was used to construct a ML tree with 100 bootstraps. TBE was used to determine bootstrap supports (88). Branches with <70% support are indicated in red. The scale bar indicates the number of substitutions per site. The tree was rooted using the *E. coli* AmpD amidase protein.
(TIF)

S2 Fig. Phylogenetic relationships of aphid *rlpA* genes shows clustering of four distinct paralogous groups. An alignment of 160 amino acid residues was used to construct a ML tree with 100 bootstraps. TBE was used to determine bootstrap supports (88). Branches with <70% support are indicated in red. The scale bar indicates the number of substitutions per site. The tree was rooted using the closest BLAST hit for the *A. pisum* RlpA4 protein, as phylogenetic positioning of the *A. pisum* RlpA proteins has shown that identity of the bacterial source of aphid *rlpA* is unclear (13). Some distinct RlpA sequences were predicted to be derived from a single coding sequence—in these instances, the amino acid range is specified.
(TIF)

S3 Fig. Bioinformatic workflow of aphid genome assembly and annotation.
(TIF)

S4 Fig. SDS-PAGE of protein fractions generated during affinity purification of recombinantly expressed, 6xHis-tagged *E. coli* AmiD. Fractions analyzed include the insoluble pellet (P), soluble cell lysate (L), Ni-NTA flow-through (FT), lysis buffer wash (W1), HEPES buffer wash (W2), and HEPES-buffered elutions of increasing imidazole concentration: 50 mM (E1), 100 mM (E2), 200 mM (E3), and 500 mM (E4).
(TIF)

Acknowledgments

We thank Eli Powell for assistance with aphid RNA extractions, Jerry Maeda for assistance with aphid embryo dissections, and the University of Texas at Austin Microscopy and Flow Cytometry Facility. We also thank Michael VanNieuwenhze for providing us with EDA.

Author Contributions

Conceptualization: Thomas E. Smith, Nancy A. Moran.

Data curation: Thomas E. Smith, Yiyuan Li.

Formal analysis: Thomas E. Smith, Yiyuan Li.

Funding acquisition: Thomas E. Smith, Nancy A. Moran.

Investigation: Thomas E. Smith, Yiyuan Li, Julie Perreau.

Methodology: Thomas E. Smith, Yiyuan Li, Julie Perreau.

Project administration: Thomas E. Smith, Nancy A. Moran.

Resources: Thomas E. Smith.

Software: Thomas E. Smith, Yiyuan Li.

Supervision: Nancy A. Moran.

Validation: Thomas E. Smith.

Visualization: Thomas E. Smith, Yiyuan Li.

Writing – original draft: Thomas E. Smith.

Writing – review & editing: Thomas E. Smith, Yiyuan Li, Julie Perreau, Nancy A. Moran.

References

1. Nakashima K, Yamada L, Satou Y, Azuma J, Satoh N. The evolutionary origin of animal cellulose synthase. *Dev Genes Evol.* 2004; 214: 81–88. <https://doi.org/10.1007/s00427-003-0379-8> PMID: 14740209
2. Jackson DJ, Macis L, Reitner J, Wörheide G. A horizontal gene transfer supported the evolution of an early metazoan biomineralization strategy. *BMC Evol Biol.* 2011; 11: 238. <https://doi.org/10.1186/1471-2148-11-238> PMID: 21838889
3. Acuña R, Padilla BE, Flórez-Ramos CP, Rubio JD, Herrera JC, Benavides P, et al. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci U S A.* 2012; 109: 4197–4202. <https://doi.org/10.1073/pnas.1121190109> PMID: 22371593
4. Wu B, Novelli J, Jiang D, Dailey HA, Landmann F, Ford L, et al. Interdomain lateral gene transfer of an essential ferrochelatase gene in human parasitic nematodes. *Proc Natl Acad Sci U S A.* 2013; 110: 7748–7753. <https://doi.org/10.1073/pnas.1304049110> PMID: 23610429
5. Wybouw N, Dermauw W, Tirry L, Stevens C, Grbić M, Feyereisen R, et al. A gene horizontally transferred from bacteria protects arthropods from host plant cyanide poisoning. *eLife.* 2014; 3: e02365. <https://doi.org/10.7554/eLife.02365> PMID: 24843024
6. Shelomi M, Danchin EGJ, Heckel D, Wipfler B, Bradler S, Zhou X, et al. Horizontal gene transfer of pectinases from bacteria preceded the diversification of stick and leaf insects. *Sci Rep.* 2016; 6: 26388. <https://doi.org/10.1038/srep26388> PMID: 27210832
7. Stairs CW, Eme L, Muñoz-Gómez SA, Cohen A, Dellaire G, Shepherd JN, et al. Microbial eukaryotes have adapted to hypoxia by horizontal acquisitions of a gene involved in rhodoquinone biosynthesis. *eLife.* 2018; 7: e34292. <https://doi.org/10.7554/eLife.34292> PMID: 29697049
8. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet.* 2008; 9: 605–618. <https://doi.org/10.1038/nrg2386> PMID: 18591983
9. Husnik F, McCutcheon JP. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol.* 2018; 16: 67–79. <https://doi.org/10.1038/nrmicro.2017.137> PMID: 29176581

10. Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci U S A*. 2002; 99: 12246–12251. <https://doi.org/10.1073/pnas.182432999> PMID: 12218172
11. Adams KL, Palmer JD. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol*. 2003; 29: 380–395. [https://doi.org/10.1016/s1055-7903\(03\)00194-5](https://doi.org/10.1016/s1055-7903(03)00194-5) PMID: 14615181
12. Nikoh N, Nakabachi A. Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biol*. 2009; 7: 12. <https://doi.org/10.1186/1741-7007-7-12> PMID: 19284544
13. Nikoh N, McCutcheon JP, Kudo T, Miyagishima S, Moran NA, Nakabachi A. Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLOS Genet*. 2010; 6: e1000827. <https://doi.org/10.1371/journal.pgen.1000827> PMID: 20195500
14. Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, et al. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*. 2013; 153: 1567–1578. <https://doi.org/10.1016/j.cell.2013.05.040> PMID: 23791183
15. Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, et al. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol Biol Evol*. 2014; 31: 857–871. <https://doi.org/10.1093/molbev/msu004> PMID: 24398322
16. Luan J-B, Chen W, Hasegawa DK, Simmons AM, Wintermantel WM, Ling K-S, et al. Metabolic coevolution in the bacterial symbiosis of whiteflies and related plant sap-feeding insects. *Genome Biol Evol*. 2015 Sep; 7(9):2635–2647. <https://doi.org/10.1093/gbe/evv170> PMID: 26377567
17. Bublitz DC, Chadwick GL, Magyar JS, Sandoz KM, Brooks DM, Mesnage S, et al. Peptidoglycan production by an insect-bacterial mosaic. *Cell*. 2019; 179: 703–712. <https://doi.org/10.1016/j.cell.2019.08.054> PMID: 31587897
18. Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*. 2000; 407: 81–86. <https://doi.org/10.1038/35024074> PMID: 10993077
19. Hansen AK, Moran NA. Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci U S A*. 2011; 108: 2849–2854. <https://doi.org/10.1073/pnas.1013465108> PMID: 21282658
20. Simonet P, Gaget K, Balmand S, Lopes MR, Parisot N, Buhler K, et al. Bacteriocyte cell death in the pea aphid/*Buchnera* symbiotic system. *Proc Natl Acad Sci U S A*. 2018; 115: E1819–E1828. <https://doi.org/10.1073/pnas.1720237115> PMID: 29432146
21. Otten C, Brill M, Vollmer W, Viollier PH, Salje J. Peptidoglycan in obligate intracellular bacteria. *Mol Microbiol*. 2018; 107: 142–163. <https://doi.org/10.1111/mmi.13880> PMID: 29178391
22. Smith TE, Lee M, Person MD, Heseck D, Mobashery S, Moran NA. Horizontal-acquisition of a promiscuous enzyme enables aphids to influence symbiont cell wall metabolism. *mBio*. 2021; 12: e02636–21. <https://doi.org/10.1128/mBio.02636-21> PMID: 34933456
23. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci U S A*. 1999; 96: 4285–4288. <https://doi.org/10.1073/pnas.96.8.4285> PMID: 10200254
24. Jothi R, Przytycka TM, Aravind L. Discovering functional linkages and uncharacterized cellular pathways using phylogenetic profile comparisons: a comprehensive assessment. *BMC Bioinformatics*. 2007; 8: 173. <https://doi.org/10.1186/1471-2105-8-173> PMID: 17521444
25. Husnik F, McCutcheon JP. Repeated replacement of an intrabacterial symbiont in the tripartite nested mealybug symbiosis. *Proc Natl Acad Sci U S A*. 2016; 113: E5416–E5424. <https://doi.org/10.1073/pnas.1603910113> PMID: 27573819
26. Atwal S, Chuenklin S, Bonder EM, Flores J, Gillespie JJ, Driscoll TP, et al. Discovery of a diverse set of bacteria that build their cell walls without the canonical peptidoglycan polymerase aPBP. *mBio*. 2021; 12: e01342–21. <https://doi.org/10.1128/mBio.01342-21> PMID: 34311584
27. Chong RA, Park H, Moran NA. Genome evolution of the obligate endosymbiont *Buchnera aphidicola*. *Mol Biol Evol*. 2019; 36: 1481–1489. <https://doi.org/10.1093/molbev/msz082> PMID: 30989224
28. Fukatsu T, Aoki S, Kurosu U, Ishikawa H. Phylogeny of Cerataphidini aphids revealed by their symbiotic microorganisms and basic structure of their galls: implications for host-symbiont coevolution and evolution of sterile soldier castes. *Zool Sci*. 1994; 11: 613–623.
29. Chong RA, Moran NA. Evolutionary loss and replacement of *Buchnera*, the obligate endosymbiont of aphids. *ISME J*. 2018; 12: 898–908. <https://doi.org/10.1038/s41396-017-0024-6> PMID: 29362506
30. Meseguer AS, Manzano-Marín A, Coeur d'Acier A, Clamens A-L, Godefroid M, Jousset E. *Buchnera* has changed flatmate but the repeated replacement of co-obligate symbionts is not associated with

- the ecological expansions of their aphid hosts. *Mol Ecol*. 2017; 26: 2363–2378. <https://doi.org/10.1111/mec.13910> PMID: 27862540
31. Monnin D, Jackson R, Kiers ET, Bunker M, Ellers J, Henry LM. Parallel evolution in the integration of a co-obligate aphid symbiosis. *Curr Biol*. 2020; 30: 1949–1957. <https://doi.org/10.1016/j.cub.2020.03.011> PMID: 32243856
 32. Zhang B, Leonard SP, Li Y, Moran NA. Obligate bacterial endosymbionts limit thermal tolerance of insect host species. *Proc Natl Acad Sci U S A*. 2019; 116: 24712–24718. <https://doi.org/10.1073/pnas.1915307116> PMID: 31740601
 33. Biello R, Singh A, Godfrey CJ, Fernández FF, Mugford ST, Powell G, et al. A chromosome-level genome assembly of the woolly apple aphid, *Eriosoma lanigerum* Hausmann (Hemiptera: Aphididae). *Mol Ecol Resour*. 2021; 21: 316–326. <https://doi.org/10.1111/1755-0998.13258> PMID: 32985768
 34. Korgaonkar A, Han C, Lemire AL, Siwanowicz I, Bennouna D, Kopec RE, et al. A novel family of secreted insect proteins linked to plant gall development. *Curr Biol*. 2021; 31: 1836–1849. <https://doi.org/10.1016/j.cub.2021.01.104> PMID: 33657407
 35. Typas A, Banzhaf M, Gross CA, Vollmer W. From the regulation of peptidoglycan synthesis to bacterial growth and morphology. *Nat Rev Microbiol*. 2012; 10: 123–136.
 36. Scheurwater EM, Burrows LL. Maintaining network security: how macromolecular structures cross the peptidoglycan layer: Transit of macromolecular structures through peptidoglycan. *FEMS Microbiol Lett*. 2011; 318: 1–9. <https://doi.org/10.1111/j.1574-6968.2011.02228.x> PMID: 21276045
 37. Herlihey FA, Moynihan PJ, Clarke AJ. The essential protein for bacterial flagella formation FlgJ functions as a β -*N*-acetylglucosaminidase. *J Biol Chem*. 2014; 289: 31029–31042. <https://doi.org/10.1074/jbc.M114.603944> PMID: 25248745
 38. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31: 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> PMID: 26059717
 39. Risper C, Legeai F, Nabity PD, Fernández R, Arora AK, Baa-Puyoulet P, et al. The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. *BMC Biol*. 2020; 18: 90. <https://doi.org/10.1186/s12915-020-00820-5> PMID: 32698880
 40. Wang Q, Ren M, Liu X, Xia H, Chen K. Peptidoglycan recognition proteins in insect immunity. *Mol Immunol*. 2019; 106: 69–76. <https://doi.org/10.1016/j.molimm.2018.12.021> PMID: 30590209
 41. Van Herreweghe JM, Michiels CW. Invertebrate lysozymes: diversity and distribution, molecular mechanism and *in vivo* function. *J Biosci*. 2012; 37: 327–348. <https://doi.org/10.1007/s12038-012-9201-y> PMID: 22581338
 42. Chen W, Hasegawa DK, Kaur N, Kliot A, Pinheiro PV, Luan J, et al. The draft genome of whitefly *Bemisia tabaci* MEAM1, a global crop pest, provides novel insights into virus transmission, host adaptation, and insecticide resistance. *BMC Biol*. 2016; 14: 110. <https://doi.org/10.1186/s12915-016-0321-y> PMID: 27974049
 43. Saha S, Hosmani PS, Villalobos-Ayala K, Miller S, Shippy T, Flores M, et al. Improved annotation of the insect vector of citrus greening disease: biocuration by a diverse genomics community. *Database*. 2017; 2017:1–20. <https://doi.org/10.1093/database/bax032> PMID: 29220441
 44. Li Y, Zhang B, Moran NA. The aphid X chromosome is a dangerous place for functionally important genes: diverse evolution of hemipteran genomes based on chromosome-level assemblies. *Mol Biol Evol*. 2020; 37: 2357–2368. <https://doi.org/10.1093/molbev/msaa095> PMID: 32289166
 45. Gerardo NM, Altincicek B, Anselme C, Atamian H, Barribeau SM, de Vos M, et al. Immunity and other defenses in pea aphids, *Acyrtosiphon pisum*. *Genome Biol*. 2010; 11: R21. <https://doi.org/10.1186/gb-2010-11-2-r21> PMID: 20178569
 46. Metcalf JA, Funkhouser-Jones LJ, Brileya K, Reysenbach A-L, Bordenstein SR. Antibacterial gene transfer across the tree of life. *eLife*. 2014; 3: e04266. <https://doi.org/10.7554/eLife.04266> PMID: 25422936
 47. Rau A, Hogg T, Marquardt R, Hilgenfeld R. A new lysozyme fold. *J Biol Chem*. 2001; 276: 31994–31999. <https://doi.org/10.1074/jbc.M102591200> PMID: 11427528
 48. Anzengruber J, Courtin P, Claes IJJ, Debreczeny M, Hofbauer S, Obinger C, et al. Biochemical characterization of the major *N*-acetylmuramidase from *Lactobacillus buchneri*. *Microbiology*. 2014; 160: 1807–1819. <https://doi.org/10.1099/mic.0.078162-0> PMID: 24858286
 49. Yahashiri A, Jorgenson MA, Weiss DS. Bacterial SPOR domains are recruited to septal peptidoglycan by binding to glycan strands that lack stem peptides. *Proc Natl Acad Sci U S A*. 2015; 112: 11347–11352. <https://doi.org/10.1073/pnas.1508536112> PMID: 26305949

50. Alcorlo M, Dik DA, De Benedetti S, Mahasenan KV, Lee M, Domínguez-Gil T, et al. Structural basis of denuded glycan recognition by SPOR domains in bacterial cell division. *Nat Commun.* 2019; 10: 5567. <https://doi.org/10.1038/s41467-019-13354-4> PMID: 31804467
51. Jorgenson MA, Chen Y, Yahashiri A, Popham DL, Weiss DS. The bacterial septal ring protein RlpA is a lytic transglycosylase that contributes to rod shape and daughter cell separation in *Pseudomonas aeruginosa*. *Mol Microbiol.* 2014; 93: 113–128. <https://doi.org/10.1111/mmi.12643> PMID: 24806796
52. Lee M, Heseck D, Dik DA, Fishovitz J, Lastochkin E, Boggess B, et al. From genome to proteome to elucidation of reactions for all eleven known lytic transglycosylases from *Pseudomonas aeruginosa*. *Angew Chem Int Ed.* 2017; 56: 2735–2739. <https://doi.org/10.1002/anie.201611279> PMID: 28128504
53. Uehara T, Park JT. An anhydro-*N*-acetylmuramyl-L-alanine amidase with broad specificity tethered to the outer membrane of *Escherichia coli*. *J Bacteriol.* 2007; 189: 5634–5641. <https://doi.org/10.1128/JB.00446-07> PMID: 17526703
54. Lamelas A, Gosalbes MJ, Manzano-Marín A, Peretó J, Moya A, Latorre A. *Serratia symbiotica* from the aphid *Cinara cedri*: a missing link from facultative to obligate insect endosymbiont. *PLoS Genet.* 2011; 7: e1002357. <https://doi.org/10.1371/journal.pgen.1002357> PMID: 22102823
55. Manzano-Marín A, Simon J-C, Latorre A. Reinventing the wheel and making it round again: evolutionary convergence in *Buchnera-Serratia* symbiotic consortia between the distantly related Lachninae aphids *Tuberolachnus salignus* and *Cinara cedri*. *Genome Biol Evol.* 2016; 8: 1440–1458. <https://doi.org/10.1093/gbe/evw085> PMID: 27190007
56. Manzano-Marín A, Szabó G, Simon J, Horn M, Latorre A. Happens in the best of subfamilies: establishment and repeated replacements of co-obligate secondary endosymbionts within Lachninae aphids. *Environ Microbiol.* 2017; 19: 393–408. <https://doi.org/10.1111/1462-2920.13633> PMID: 27902872
57. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-wide identification of episodic selection. *Mol Biol Evol.* 2015; 32: 1365–1371. <https://doi.org/10.1093/molbev/msv035> PMID: 25701167
58. Kuru E, Hughes HV, Brown PJ, Hall E, Tekkam S, Cava F, et al. *In situ* probing of newly synthesized peptidoglycan in live bacteria with fluorescent D-amino acids. *Angew Chem Int Ed.* 2012; 51: 12519–12523.
59. Liechti GW, Kuru E, Hall E, Kalinda A, Brun YV, VanNieuwenhze M, et al. A new metabolic cell-wall labelling method reveals peptidoglycan in *Chlamydia trachomatis*. *Nature.* 2014; 506: 507–510. <https://doi.org/10.1038/nature12892> PMID: 24336210
60. Vollmer W, Joris B, Charlier P, Foster S. Bacterial peptidoglycan (murein) hydrolases. *FEMS Microbiol Rev.* 2008; 32: 259–286. <https://doi.org/10.1111/j.1574-6976.2007.00099.x> PMID: 18266855
61. Das D, Hervé M, Elsliger M-A, Kadam RU, Grant JC, Chiu H-J, et al. Structure and function of a novel LD-carboxypeptidase A involved in peptidoglycan recycling. *J Bacteriol.* 2013; 195: 5555–5566. <https://doi.org/10.1128/JB.00900-13> PMID: 24123814
62. Lenz JD, Hackett KT, Dillard JP. A single dual-function enzyme controls the production of inflammatory NOD agonist peptidoglycan fragments by *Neisseria gonorrhoeae*. *mBio.* 2017; 8: e01464–17. <https://doi.org/10.1128/mBio.01464-17> PMID: 29042497
63. Zellner B, Mengin-Lecreux D, Tully B, Gunning WT 3rd, Booth R, Huntley JF. A *Francisella tularensis* L,D-carboxypeptidase plays important roles in cell morphology, envelope integrity, and virulence. *Mol Microbiol.* 2021; 115: 1357–1378. <https://doi.org/10.1111/mmi.14685> PMID: 33469978
64. Moran NA. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A.* 1996; 93: 2873–2878. <https://doi.org/10.1073/pnas.93.7.2873> PMID: 8610134
65. Petersen LM, Tisa LS. Friend or foe? A review of the mechanisms that drive *Serratia* towards diverse lifestyles. *Can J Microbiol.* 2013; 59: 627–640. <https://doi.org/10.1139/cjm-2013-0343> PMID: 24011346
66. Walterson AM, Stavrinos J. Pantoea: insights into a highly versatile and diverse genus within the Enterobacteriaceae. *FEMS Microbiol Rev.* 2015; 39: 968–984. <https://doi.org/10.1093/femsre/fuv027> PMID: 26109597
67. Craik DJ, Daly NL, Waine C. The cystine knot motif in toxins and implications for drug design. *Toxicon.* 2001; 39: 43–60. [https://doi.org/10.1016/s0041-0101\(00\)00160-4](https://doi.org/10.1016/s0041-0101(00)00160-4) PMID: 10936622
68. Postic G, Gracy J, Périn C, Chiche L, Gelly J-C. KNOTTIN: the database of inhibitor cystine knot scaffold after 10 years, toward a systematic structure modeling. *Nucleic Acids Res.* 2018; 46: D454–D458. <https://doi.org/10.1093/nar/gkx1084> PMID: 29136213
69. Mahadav A, Gerling D, Gottlieb Y, Czosnek H, Ghanim M. Parasitization by the wasp *Eretmocerus mundus* induces transcription of genes related to immune response and symbiotic bacteria

- proliferation in the whitefly *Bemisia tabaci*. BMC Genomics. 2008; 9: 342. <https://doi.org/10.1186/1471-2164-9-342> PMID: 18638407
70. Zhang C-R, Zhang S, Xia J, Li F-F, Xia W-Q, Liu S-S, et al. The immune strategy and stress response of the mediterranean species of the *Bemisia tabaci* complex to an orally delivered bacterial pathogen. PLoS ONE. 2014; 9: e94477. <https://doi.org/10.1371/journal.pone.0094477> PMID: 24722540
 71. Hariton Shalev A, Sobol I, Ghanim M, Liu S-S, Czosnek H. The whitefly *Bemisia tabaci* knottin-1 gene is implicated in regulating the quantity of tomato yellow leaf curl virus ingested and transmitted by the insect. Viruses. 2016; 8: 205. <https://doi.org/10.3390/v8070205> PMID: 27455309
 72. Kelkar YD, Ochman H. Genome reduction promotes increase in protein functional complexity in bacteria. Genetics. 2013; 193: 303–307. <https://doi.org/10.1534/genetics.112.145656> PMID: 23114380
 73. Newton MS, Arcus VL, Gerth ML, Patrick WM. Enzyme evolution: innovation is easy, optimization is complicated. Curr Opin Struc Biol. 2018; 48: 110–116. <https://doi.org/10.1016/j.sbi.2017.11.007> PMID: 29207314
 74. Price DRG, Wilson ACC. A substrate ambiguous enzyme facilitates genome reduction in an intracellular symbiont. BMC Biol. 2014; 12: 110. <https://doi.org/10.1186/s12915-014-0110-4> PMID: 25527092
 75. Chung SH, Jing X, Luo Y, Douglas AE. Targeting symbiosis-related insect genes by RNAi in the pea aphid-*Buchnera* symbiosis. Insect Biochem Mol Biol. 2018; 95: 55–63. <https://doi.org/10.1016/j.ibmb.2018.02.004> PMID: 29526771
 76. Smith TE, Moran NA. Coordination of host and symbiont gene expression reveals a metabolic tug-of-war between aphids and *Buchnera*. Proc Natl Acad Sci U S A. 2020; 117: 2113–2121. <https://doi.org/10.1073/pnas.1916748117> PMID: 31964845
 77. Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima S. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. Curr Biol. 2014; 24: R640–R641. <https://doi.org/10.1016/j.cub.2014.06.038> PMID: 25050957
 78. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30: 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170> PMID: 24695404
 79. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. Curr Protoc Bioinforma. 2020; 70: e102. <https://doi.org/10.1002/cpbi.102> PMID: 32559359
 80. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015; 12: 357–360. <https://doi.org/10.1038/nmeth.3317> PMID: 25751142
 81. Zhu B-H, Xiao J, Xue W, Xu G-C, Sun M-Y, Li J-T. P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads. BMC Genomics. 2018; 19: 175. <https://doi.org/10.1186/s12864-018-4567-3> PMID: 29499650
 82. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10: 421. <https://doi.org/10.1186/1471-2105-10-421> PMID: 20003500
 83. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25: 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
 84. Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. F1000Research. 2017; 6: 1287.
 85. Frith MC. A new repeat-masking method enables specific detection of homologous sequences. Nucleic Acids Res. 2011; 39: e23. <https://doi.org/10.1093/nar/gkq1212> PMID: 21109538
 86. Palmer J, Stajich J. Funannotate v1.8.8. 2020. <https://zenodo.org/record/2604804>
 87. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2016; 32: 767–769. <https://doi.org/10.1093/bioinformatics/btv661> PMID: 26559507
 88. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. Kollmar M, editor. Methods Mol Biol. 2019; 1962: 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5 PMID: 31020555
 89. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019; 20: 238. <https://doi.org/10.1186/s13059-019-1832-y> PMID: 31727128
 90. Katoh K, Kuma K, Miyata T, Toh H. Improvement in the accuracy of multiple sequence alignment program MAFFT. Genome Inform. 2005; 16: 22–33. PMID: 16362903
 91. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000; 17: 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334> PMID: 10742046

92. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020; 37: 1530–1534. <https://doi.org/10.1093/molbev/msaa015> PMID: 32011700
93. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004; 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
94. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009; 25: 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348> PMID: 19505945
95. Lemoine F, Domelevo Entfellner J-B, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature.* 2018; 556: 452–456. <https://doi.org/10.1038/s41586-018-0043-0> PMID: 29670290
96. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017; 14: 587–589. <https://doi.org/10.1038/nmeth.4285> PMID: 28481363
97. Csuros M. Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics.* 2010; 26: 1910–1912. <https://doi.org/10.1093/bioinformatics/btq315> PMID: 20551134
98. Jiang X, Zhang Q, Qin Y, Yin H, Zhang S, Li Q, et al. A chromosome-level draft genome of the grain aphid *Sitobion miscanthi*. *GigaScience.* 2019; 8: giz101. <https://doi.org/10.1093/gigascience/giz101> PMID: 31430367
99. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Hancock J, editor. *Bioinformatics.* 2018; 34: 1037–1039. <https://doi.org/10.1093/bioinformatics/btx713> PMID: 29106469
100. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli*/K-12. *Nucleic Acids Res.* 2017; 45: D543–D550. <https://doi.org/10.1093/nar/gkw1003> PMID: 27899573
101. R Core Team. R: A language and environment for statistical computing. 2019. <https://www.R-project.org/>
102. RStudio Team. RStudio: Integrated development for R. 2020. <http://www.rstudio.com/>
103. Kolde R. pheatmap: pretty heatmaps. 2018. <https://CRAN.R-project.org/package=pheatmap>
104. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 2006; 34: W609–612. <https://doi.org/10.1093/nar/gkl315> PMID: 16845082
105. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol.* 2018; 35: 773–777. <https://doi.org/10.1093/molbev/msx335> PMID: 29301006
106. Chong RA, Moran NA. Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. *Proc Natl Acad Sci U S A.* 2016; 113: 13114–13119. <https://doi.org/10.1073/pnas.1610749113> PMID: 27799532
107. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods.* 2012; 9: 676–682. <https://doi.org/10.1038/nmeth.2019> PMID: 22743772
108. Wickham H, Averick M, Bryan J, Chang W, D'Agostino McGowan L, François R, et al. Welcome to the tidyverse. *J Open Source Software.* 2019; 4: 1686.
109. Dinno A. dunn.test: Dunn's test of multiple comparisons using rank sums; 2017 [cited 2020]. Database: r-project [Internet]. <https://cran.r-project.org/web/packages/dunn.test/index.html>
110. Wickham H. ggplot2: Elegant graphics for data analysis. 3rd ed. New York: Springer-Verlag; 2016.
111. Mangiafico S. rcompanion: Functions to support extension education program evaluation; 2022 [cited 2020]. Database: r-project [Internet]. <https://cran.r-project.org/web/packages/rcompanion/index.html>
112. Koga R, Tsuchida T, Fukatsu T. Quenching autofluorescence of insect tissues for *in situ* detection of endosymbionts. *Appl Entomol Zool.* 2009; 44: 281–291.