Recurrent horizontal gene acquisition in the evolutionary history of ants

Lukas Schrader

schradel@uni-muenster.de

University of Münster

Janina Rinke

University of Münster https://orcid.org/0009-0006-0830-3411

Ding He

University of Copenhagen

Joel Vizueta

University of Copenhagen https://orcid.org/0000-0003-0139-3013

Rasmus Larsen

University of Copenhagen

Zijun Xiong

BGI Research https://orcid.org/0000-0003-3923-0703

Jürgen Gadau

University of Münster

Guojie Zhang

Zhejiang University https://orcid.org/0000-0001-6860-1521

Jacobus Boomsma

University of Copenhagen https://orcid.org/0000-0002-3598-1609

Article

Keywords: Horizontal gene transfer, comparative genomics, social insects, bacteria, endosymbionts

Posted Date: January 13th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-5741203/v1

Additional Declarations: There is **NO** Competing Interest.

Recurrent horizontal gene acquisition in the evolutionary history of ants

Janina L. Rinke[1], Ding He[2], Joel Vizueta[2], Rasmus Stenbak Larsen[2], Zijun Xiong[6], Jürgen Gadau[1], Guojie Zhang[3, 4, 5], Jacobus J. Boomsma[2], Lukas Schrader[1, *]

[1] Institute for Evolution and Biodiversity, University of Münster, Münster, Germany

[2] Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark

[3] Evolutionary & Organismal Biology Research Center, Zhejiang University School of Medicine, Hangzhou 310058, China.

[4] Villum Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark

[5] Women's Hospital, School of Medicine, Zhejiang University, Shangcheng District, Hangzhou 310006, China.

[6] School of Basic Medical Sciences, Jiangxi Medical College, Nanchang University, China.

***Correspondence:** Lukas.Schrader@uni-muenster.de, Institute for Evolution and Biodiversity, University of Münster, Hüfferstr. 1, DE-48149 Münster, Germany

Running title: Horizontal gene acquisition in ants

Figures: 4

Tables: 0

Abstract

Horizontal Gene transfer (HGT) from bacteria has often led to Horizontal Gene Acquisition (HGA), subsequently contributing to phenotypic innovation. Ants are interesting potential targets for HGA because they host many mutualistic associations with vertically transmitted symbionts, but the overall prevalence of HGA across the ants and other insect lineages remains virtually unexplored. Here, we systematically screened the genomes of 163 ant species and identified 497 HGA events of protein-coding genes, predominantly derived from intracellular symbionts, in 85 species belonging to eight subfamilies. Apart from convergent horizontal transfers of *Wolbachia*-derived ankyrin repeat proteins into the genomes of 45 ant species, we identified dozens of other HGAs that likely offered adaptive innovations of phenotypic functions, primarily mediating immune-system adaptations or facilitating nutritional niche expansions. We provide in-depth characterizations of multiple clade- and species-specific HGAs, some as old as 40 MY, consistent with strong evolutionary conservation. Our study is the first of its kind in ants and considerably expands our general appreciation of the evolutionary significance of HGA from bacteria to eukaryotes.

Keywords: Horizontal gene transfer, comparative genomics, social insects, bacteria, endosymbionts

Introduction

Horizontal gene transfer (HGT) between unrelated genomes is a key driver of evolutionary change provided such random transfers result in horiziontal gene acquisition (HGA) that natural selection can act on[1–4]. HGT between prokaryotic species has been extensively studied for the adaptive innovations it allowed, such as the spread of antibiotic resistance across species boundaries[5]. While prokaryote HGT events are often reciprocal, there is increasing evidence for HGT towards multicellular eukaryotes, primarily from bacteria, fungi, or viruses[6–10]. These transfers are asymmetrical because very few eukaryote genes are known to have become established in prokaryotes[6]. Owing to recent advances in genomics and molecular biology, systematic comparative analyses of HGA in multicellular eukaryotes have now become feasible[11–13]. However, in such studies it is crucial to realize that HGA events are functionally comparable to random macromutations and that they can only result in lasting phenotypic effects when they become subject to natural selection. Genome-wide screens of HGA in multicellular organisms should thus ask which of these elements are likely to have been maintained by selection against a null-hypothesis of them having survived in multicellular genomes as neutral or slightly deleterious acquisitions in finite populations subject to genetic drift.

Until recently, bacteria-to-eukaryote HGT has been controversial[14], as bacterial contaminations, ancestral genes lost from related extant lineages, or incorrect phylogenetic inferences have posed challenges for the correct identification of such putative HGA events[15]. However, many eukaryotic HGAs have now been confirmed by rigorous follow-up experiments and functional characterizations[16]. For example, more than hundred candidate HGAs have been identified in protists[17–19] and at least 1,400 genetic elements from non-metazoan donors have recently been identifed as part of insect genomes[12]. Many such functional HGAs in eukaryotes have also been shown to mediate nutritional and metabolic innovations[9,11,20], as well as novel immune-system responses[21–24] and defenses against parasitism[25–27].

Successful HGA requires integration of bacterial genetic material into the metazoan germline, which implies that vertically transmitted endosymbionts were most likely to act as donors. Such endosymbionts occur more commonly in some animal groups than in others, likely explaining why HGAs occured more often in insects[28] than in vertebrates[20,29]. Vertically transmitted endosymbionts such as *Wolbachia* or *Blochmannia* are widespread within the ants, an ecologically highly diverse and exclusively social insect family with over 15,000 described species. These obligate endosymbionts have been vertically co-transmitted with their ant

hosts for millions of years[30] which makes these intricate relationships obvious sources for HGT events resulting in HGA. However, in-depth comparative studies to illucidate the prevalence of HGA in ants are lacking.

In this study, we comprehensively searched for HGAs across 163 ant genomes that were recently subjected to general analysis[31], of which 120 were *de-novo* sequenced on a PacBio platform. This large-scale approach extends previous coverage by at least an order of magnitude, because HGAs have so far only been described in detail for two ant species, the wood ant *Formica exsecta*[32] and the heart-node ant *Cardiocondyla obscurior*[33]. In *F. exsecta*, multiple putative genes encoding ankyrin repeat domain (ANK) proteins, DNA repair proteins, and transposases have been identified as HGAs from *Wolbachia*[32] while a HGA from *Blochmannia*-like enterobacteria has been described for *C. obscurior*[33]. Another recent study across 218 insect genomes identified putative HGAs in 20 ant species, but no effort was made to describe these in any detail[12]. Here, we identify and characterize 497 HGAs across 85 ant species, covering eight of the 17 extant ant subfamilies. Focusing on the most striking cases, we further provide in-depth analyses of the potential impact of HGAs on adaptive evolution in the ants.

Results

Large-scale horizontal gene acquisition from bacteria

To systematically identify HGA from bacteria to ants, we used a conservative approach, favoring specificity (accepting false negatives) over sensitivity (avoiding false positives). We screened 163 ant genomes from 12 subfamilies, including Amblyoponinae (3), Dolichoderinae (6), Dorylinae (4), Ectatomminae (2), Formicinae (39), Leptanillinae (1), Myrmiciinae (3), Mymicinae (77), Paraponerinae (1), Ponerinae (21), Proceratiinae (4), and Pseudomyrmecinae (2). After careful filtering and manual curation (see methods: Detection, validation, and quality assessment of HGA candidates), we identified 497 high-confidence HGA events of bacterial origin across 85 ant genomes (Fig.1A, Tab.S1A) belonging to eight of the twelve investigated ant subfamilies. The highest HGA numbers were found in three *Myrmica* species (*M. scabrinoides, M. rubra*, and *M. angulata*) and we failed to find any HGAs in *Atta* (*A. cephalotes, A. colombica*) or *Acromyrmex* (*A. ameliae, A. echinatior, A. lobicornis*), despite high-quality reference genomes and previously described prevalence of endosymbionts[34,35]. We were unable to identify high-confidence bacterial HGAs in the lower-

quality genome assemblies of Leptanillinae, Proceratiinae, Paraponerinae, and Amblyoponinae, because all candidate HGA loci were excluded during filtering and manual curation (see methods). However, the number of candidate HGAs in these lower quality genomes before filtering was similar to the numbers in other more contiguous assemblies (Fig.1A), suggesting that the prevalence of HGAs is similar across the ant subfamilies.

We used PCR and Sanger Sequencing to confirm a subset of the computationally predicted HGA loci at the molecular level. Out of the 43 tested HGA candidate events, 36 could be confirmed by PCR and Sanger Sequencing, while results for seven remained inconclusive (Tab.S2). Additionally, we compared our list of identified HGAs with the few previously reported bacterial HGTs in ants[12,32,33] and found that several HGTs were re-identified in our study (Tab.S3).

The 497 identified HGAs contained coding sequences (CDS) for 1,053 bacterial proteins (Tab.S1B). Among these, genes coding for ANK proteins were most abundant, identified in 45 ant genomes from eight subfamilies and with broadly distributed sequence identity percentages relative to their respective bacterial reference proteins (Fig.1B). We further detected four clade-specific HGAs: (i) Cyclopropane-Fatty-Acyl-Synthases (CFA) and (ii) Ribosomal RNA methyltransferases (MetA) in eight Formicini (Formicinae) species (*Formica* and *Iberoformica*), (iii) Lysozymes (Lys) in 21 species belonging to two different clades in the Myrmicinae, and (iv) *N*-Acetyl-muramicacid-6-P-etherases (MurNAc) in eight Camponotini (Formicinae) species (Fig.1A,B). All clade-specific HGAs showed around 75% average sequence identity with their closest bacterial reference sequence, consistent with adaptive changes in HGA loci becoming fixed and being maintained by selection (Fig.1B), although distributions did vary (see below). We also identified seven cases of HGA shared between two or three not closely related species, of which two showed conserved synteny suggesting a single origin (Fig.S1, Tab.S4). Finally, we detected 61 single-species HGAs (Tab.S1).

CDS lengths ranged from 150 to 10,000 bp, with 58 HGAs having lengths >6,000 bp. Out of the 1,053 annotated CDS sequences, 384 were expressed (read counts >100 in RNAseq analyses), consistent with these proteins being functional. Gene Ontology (GO) term analyses highlighted enrichment in lipid biosynthesis, prokaryotic cell wall catabolism, bacterial cell wall degradation, methylation, and nucleotide metabolism (Fig.S2).

Wolbachia endosymbionts (Alphaproteobacteria) were the most frequent source of HGAs, accounting for 79% of the 497 identified loci in the ant genomes (Fig.1A,C). Gammaproteobacteria (*Blochmannia* and related genera) contributed 10% (n=49), while

156 *Spiroplasma/Mycoplasma* (Mollicutes) were donors for 37 HGAs, followed by Sphingobacteriia (n=9) and other, not further specified, bacteria (n=9, Fig.1C, Tab.S1). Ants
158 of the subfamily Myrmicinae showed high prevalences of *Wolbachia*-derived HGAs, along with lineage-specific acquisitions of *Spiroplasma* and *Cardinium* (Sphingobacteriia) in
160 *Temnothorax* and close relatives (Fig.1A). The Formicinae and Dolichoderinae subfamilies exhibited a greater diversity of bacterial HGA donors, with Enterobacteria (e.g. *Sodalis,*
162 *Yersinia,* or *Blochmannia-like* bacteria) contributing 25% across the Formicinae genomes and 51% across the Dolichoderinae genomes (Fig.1A, Tab.S1).

164 **Ankyrin repeat genes: Convergent HGAs across many species but with unclear functions**

166 245 HGAs (49% of the 497 total) encoded one or multiple ankyrin repeat (ANK) proteins, distributed across 45 ant species from eight subfamilies (Fig.1, Fig.S3). ANK HGA frequencies
168 ranged from one (in twelve ant species) to 42 and 64 in *Myrmica scabrinoidis* and *Myrmica rubra,* respectively. Out of the 418 ANK protein CDS, 249 were expressed (read counts >100)
170 across 38 species from seven subfamilies. Notably, 80 of these 249 expressed ANKs were found in the genus *Myrmica*, a significant overrepresentation of bacterial ANK repeats relative
172 to what is normal in ants (Fisher's exact test, $p<0.0001$, odds ratio=7.31).

174 All ANK HGAs originated from *Wolbachia*, except for nine uncharacterized or fragmented proteins with ANK domains that were predicted to have originated from Bacteroidetes or
176 Rickettsiales other than *Wolbachia* but were removed as putatively false positives during filtering. Integrating information from Uniprot, we identified 19 different *Wolbachia* strains as
178 potential donors with numbers derived from specific *Wolbachia* strains ranging from 1-132 ANK CDS but without a clear pattern within single *Wolbachia* species or across host ant
180 species (Fig.1, Fig.S3, Tab.S1).

ANK HGAs could be partitioned into 21 UniRef homology clusters, varying in length from 189
182 to 4751 amino acids and also in domain architecture within and between ant species. Major ANK clusters recurred broadly across host ant species and subfamilies suggesting ancient
184 origins without a discernible pattern (Fig.S3). ANK loci were often densely clustered within the host genomes, indicating recurrent independent insertions of *Wolbachia*–derived gene sets.
186 The frequent expression of ANK insertions seems incompatible with a neutral scenario in which rapid duplication and divergence would always be followed by degradation. On the other
188 hand, if ANK insertions would have straightforward adaptive roles, we would expect the extent of their structural expression to be positively correlated with lack of degradation, i.e. with the
190 proportional identity with the (supposedly ancestral) bacterial reference proteins. We did not

find such a positive correlation (Pearson correlation coefficient r=-0.029, p=0.55, Fig.S4). This would be consistent with ANKs often having adaptive functions in gene regulation, but at present this cannot be more than a conjecture.

Ancient orthologous HGAs

In-depth comparative analyses revealed four orthologous HGAs across at least four clusters of related ant species (Fig.1), all but one of which were expressed (Tab.S4, S5, S6). These orthologous HGAs showed conserved synteny, consistent with purifying selection acting to preserve these regions, which comprised: (1) a bacterial lysozyme, derived from *Wolbachia* and maintained independently in the ancestors of *Carebara* and *Temnothorax* and a number of related genera (Fig.1, Fig.2A); (2) an *N*-Acetyl-Muramic-Acid-Etherase (MurNAc) originating from *Spiroplasma* in all Camponotini (Fig.1, Fig.2B); and (3) a Cyclopropane-Fatty-Acyl-Synthase (CFA Synthase) locus, derived from Enterobacteria and maintained in *Formica* and its sister genus *Iberoformica* (Fig.1, Fig.S6).

Bacterial lysozyme HGAs were present in 21 ant species from two distinct Crematogastrini clades (Fig.1, Fig.2A, Tab.S5), and were all expressed with predicted transcripts containing a 5'-non-coding exon, consistent with the secondary emergence of gene regulatory structures (Fig.S5). Further phylogenetic analyses characterized this HGA event in both ant clades as having integrated the *Wolbachia glycosyl hydrolase muramidase* lysozyme gene (GH25) into the ancestral ant genomes. Our analyses also revealed a secondary loss of this HGA in *Pristomyrmex punctatus* (Fig.1, Fig.2A) and its independent insertion in the *Carebara* and *Temnothorax* clades (Fig.2A). These findings were substantiated by absence of any synteny between the *Carebara* and the *Temnothorax* insertions (Fig.2A), suggesting convergent horizontal acquisitions of bacterial lysozymes in these ant clades 29-39 MYA and ca. 51 MYA, respectively.

We further identified a conserved MurNAc HGA (*murQ*) originating from Mollicutes bacteria in eight species of Camponotini (Fig.1, Fig.2B, Tab.S6), which was expressed in all eight species. Synteny and phylogenetic analyses confirmed the single ancestral HGA transfer into the ancestor of Camponotini 40-57 MYA (Fig.2B).

Lastly, we identified a horizontally transferred *cfa* gene (encoding a CFA synthase) shared by eight species from the Formicini tribe. Phylogenetic analyses revealed a likely origin from *Sodalis*-like enterobacteria and diversification into 87 *cfa* HGAs encoding 177 CDS sequences with 20 full-length expressed CFA genes (Tab.S7, Fig.S6). Each species had five to ten CFA synthases with lengths varying from short fragments to a full-length CDS (Tab.S7). *Formica*

japonica showed the highest number of full-length *cfa* genes (n=5, all expressed), followed by *F. sanguinea* (n=4, two expressed), and *F. cf. japonica* (n=3, all expressed). *F. fusca, F. exsecta,* and *F. cinerea* all carried one complete and expressed *cfa* gene of 1,148 bp, while *Iberoformica subrufa,* the most basal of the eight species, had two complete and expressed CFA synthase sequences. Phylogenetic and syntenic relationships of the full-length CFA synthase HGAs suggested a single evolutionary origin ca. 33 MYA, followed by a complex evolutionary history with recurrent gene duplications, deletions, and/or translocations (Fig.S6, Fig.S7).

Other HGAs in ant host genomes

Among the 75 remaining HGAs that were neither ANK loci, fragmented, nor ancestrally conserved functional loci, we identified six HGA candidates for further investigation, because they were all expressed and encoded full-length bacterial proteins of >65% sequence identity. Five out of these six could be confirmed by PCR (Tab.S2, for *Colobopsis sp.* no DNA was available). Two of these coded for proteins related to bacterial cell wall and membrane biosynthesis functions: An enterobacterial D-alanine–D-alanine ligase (*ddl2*) (Fig.3A, Fig.S1) conserved in three Formicoxenini species (*Formicoxenus nitidulus, Harpagoxenus sublaevis, Leptothorax acervorum*), and *a Wolbachia*-derived UDP-*N*-acetyl-glucosamine-1-carboxyvinyltransferase (*murA*) in *Pheidole pallidula* (Fig.3B). Additionally, four HGAs were associated with metabolic pathways (Fig.3C-F): (i) a phenazine biosynthesis protein (PhzF) in *Liometopum microcephalum*, (ii) an Aryl-sulfate sulfotransferase (ASST) in *Colobopsis sp.*, (iii) a DNA helicase (*uvrD*) involved in DNA mismatch repair in *Kalathomyrmex emeryi* and two Ponerinae *(Hypoponera opacior*, *Euponera pilosior)* and (iv) a Xanthine-guanine-phosphoribosyltransferase in *Cardiocondyla obscurior*. Three of these were derived from *Sodalis*-like endosymbionts of the Enterobacteriaceae family (Fig.3C,D,F) while the *uvrD* HGA in *K. emeryi, E. pilosio*r, and *H. opacior* originated from *Wolbachia* (Fig.3E). For some of these HGAs (e.g. *PhzF* in *L. microcephalum* and *murA* in *P. pallidula*), gene expression patterns suggested the presence of regulatory 5' non-coding exons (Fig.3B,C).

The candidate set also included DNA helicase HGAs in two Ponerini species and in the myrmicine ant *K. emeryi* (Fig.3E), which are likely to be three independent evolutionary events because there was no synteny between these three HGA loci (Fig.S1). We found a higher degree of homology between the *E. pilosior* and *H. opacior* sequences compared to the *K. emeryi* sequence and any of the *Wolbachia* strains (Fig.3E) but, in general, this HGA shows over 30 % divergence from the closest *Wolbachia* hit. The alternative interpretation of an ancient origin of the *uvrD* HGA in the common ancestor of *E. pilosior* and *H. opacior* would

imply convergent losses in at least twelve other Ponerinae species, which would be a much less parsimonious scenario (Fig.1).

Finally, the HGA in *C. obscurior,* encoding a protein involved in the bacterial purine salvage pathway, showed a highly conserved CDS with high expression levels (Fig.3F). This HGA has already been reported in a study by Klein et al.[33] and is likely derived from the intracellular Enterobacteriaceae symbiont *Candidatus westeberhardia cardiocondylae*.

Discussion

In this study we systematically assessed bacteria-to-ant horizontal gene transfers and identified 497 HGA loci encoding 1,053 genes in 85 ant species from eight subfamilies. Despite revealing a rich functional and evolutionary diversity of HGAs, our results remain a conservative estimate of the frequency of bacteria-to-ant HGTs and are thus likely to underestimate the true prevalence of HGAs. The genomic and transcriptional signatures of the analysed HGAs suggest functional and evolutionary significance, consistent with HGA-driven adaptive innovations in ant biology. Particularly the secondary addition of 5' untranslated region (UTR) elements, upstream of the start codon, to a number of HGAs suggests their post-HGA fine-tuning by natural selection processes. Based on our present data, we can conclude that HGA has occurred regularily during ant evolution and that ensuing HGA-dependent functional innovations were most commonly associated with immune system processes or metabolic enrichments.

HGA events require intracellular host-symbiont association

It is too rarely made explicit that direct donor-recipient contact is a necessary condition for bacterial HGTs to the host germline. This implies that mostly hosts with intracellular and vertically transmitted symbionts can be expected to experience a certain frequency of symbiont-mediated HGT. This explains in turn why we expect HGAs to be variably prevalent in insects[28], but to be virtually absent in vertebrates[20,29] and that HGAs, also in the present study, were derived mostly from *Wolbachia* (Fig.1A,C), the most widely distributed maternally-inherited intracellular symbiont of insects[36]. This finding aligns with previous studies showing that *Wolbachia* sequences of considerable length have been transferred to the nuclear genome of solitary insect hosts[12,32,37,38]. The *Drosophila ananassae* genome even integrated an entire genomic copy of its *Wolbachia* symbiont in its own genome[37,38]. Bacteriophages such as the temperate phage *WO* can mediate *Wolbachia*–derived HGAs, potentially enabling incorporation of genetic material from different *Wolbachia* strains in the same host genome[39].

296 Apart from *Wolbachia*, the intimate relationships of *Blochmannia* and *Blochmannia*-like intracellular endosymbionts with e.g. *Camponotus*, *Plagiolepis, Formica* and *Cardiocondyla*
298 ants also provided opportunities for HGA. These have been documented in isolated ant lineages previously[40,41], but are now shown to likely characterize entire Formicinae clades.
300 Such HGAs from longterm coevolved endosymbiont lineages (e.g. *Blochmannia* app. 80 my) might be important for the communication between hosts and these mutualists to ensure
302 longterm cooperation between partners.

We also discovered HGAs from *Sodalis*-like endosymbionts in the Formicoxenini and the
304 genus *Liometopum* (Fig.4A,C), despite such endosymbionts not occuring in extant populations of these ants, suggesting they constitute remnants of past symbioses or that they relied on
306 other transmission routes.

308 **The functional significance of bacterial HGAs in ants**

HGAs of bacterial origin can obtain diverse functions in recipient insect genomes, e.g. affecting
310 body coloration, plant- or bacterial cell wall degradation, defensive functions, detoxification capabilites, and male courtship[12,28]. In our study, HGAs often involved genes related to
312 metabolic and cell-wall related processes in bacteria. Targeted functional studies will be necessary to determine the true function of these HGAs. For example, *uvrD* (a DNA helicase,
314 Fig.4E) and *gpt* (a Xanthine-guanine phosphoribosyltransferase of the purine salvage pathway, Fig.4F) are functionally well characterized in bacteria but their HGA significance in
316 ant genomes remains obscure.

318 Despite these limitations, our analyses suggest that biological functions of ant HGAs often relate to defenses against and resistance to pathogens, mostly via bacterial cell-wall
320 degradation (Fig.4). Key examples are the clade-specific Lysozymes in several Myrmicinae species and the MurNAc etherases in Camponotini ants (Fig.2).

322

Lysozymes can also serve as a protection from pathogens by peptidoglycan cleavage
324 between *N*-Acetylglucosamine (NAG) and *N*-Acetylmuramic acid (NAM), while MurNAc etherases act in similar ways directly on NAM[42,43]. These HGAs might thus provide
326 antibacterial defense systems to the ants, killing pathogenic bacteria by cell wall degradation (Fig.4). Remarkably, bacterial lysozyme HGAs have occurred independently also in diverse
328 fungi, plants, archaea, bivalves, and solitary insects and have led to, for example, the generation of antibiotic GH25 muramidases in plants and archaea[22,24,44,45]. Disease defenses
330 are well documented to be a pervasive threat to ant colonies that has maintained selection for

multilayer recognition and immune defense mechanisms, so the HGAs discovered here add to a much broader spectrum of individual and social immune strategies[46].

In contrast to lysozymes, the HGA of *murQ* genes has not been reported previously and is potentially unique to the Camponotini ants. HGA-encoded murQ can convert *N*-acetylmuramic acid-phosphate to *N*-acetylglucosamine-phosphate by cleavage of the lactyl residue. This can then be further degraded, used in glycolysis, or directed into peptidoglycan *de novo* synthesis and recycling[42,47,48]. The *murQ* HGA has a strong adaptive potential in enhancing the ants' immune defense by using this peptidoglycan-degrading enzyme to kill bacterial pathogens, while leaving an endosymbiotic relationship with cell-wall deficient *Spiroplasma*, the presumable HGA donor, unaffected[49].

CFA synthases, such as those acquired by two sister lineages of Formicini, catalyze the cyclopropanation of unsaturated fatty acids of bacterial membranes (Fig.5). In bacteria, this function has been associated with adaptive stress responses to changes in pH, temperature, and salinity[50–54]. CFA synthases were previously identified in eukaryotes such as plants and *Leishmania* parasites[51,53,55,56], and have presumably also been horizontally acquired from bacteria[53,57–59]. We found that most Formicini species have several expressed and likely functional *cfa* gene copies and that full-length CFA synthases were conserved because they had retained synteny. This suggests that CFA synthases emerged from a single ancient HGT to the common ancestor of *Formica* and *Iberoformica* ca. 33 MYA[60] with secondary diversification by gene duplications and rearrangements (Fig.S3) coinciding with the adaptive radiation of the genus *Formica*. Finally, the *ddl2* HGA in three Formicoxenini species and the *murA* gene in *Pheidole pallidula* might convey antibacterial functions as well, as both enzymes are involved in peptidoglycan anabolism and catabolism[2,61].

HGA-mediated reinforcement of ant-bacteria symbioses

Our study markedly expands our knowledge of HGAs across the global ants, and in fact across all insects. The most prevalent HGAs were ANK-domain proteins homologous to *Wolbachia* ANK proteins, occurring in single- or multiple copies across 45 species from eight subfamilies, frequently with high copy numbers in specific ant genomes (Fig.1A,C). ANKs consist of relatively short, tandem repeat motifs which fold into structures mediating molecular recognition via protein-protein interactions[62–65]. They are involved in a diverse set of functional host-symbiont interactions and may be employed by symbionts like *Wolbachia* to mimic or manipulate host functions following infection of eukaryotic cells[66–69]. In general, the evolutionary history of ANK HGAs is difficult to infer accurately, due to frequent expansions

and contractions of ANK domains within genes[70], secondary modifications that are likely to have affected many ANKs after their HGA in ant genomes.

The recurrent integrations of phylogenetically diverse *Wolbachia* ANK genes into ant genomes suggest that a diverse set of *Wolbachia* strains have functioned as ANK donors in ants (Fig.S3), similar to what is known from other insects[70-73]. The general prevalence of ANK HGAs in insects suggest that they may continue to serve manipulative *Wolbachia* interests. However, it has been notoriously difficult to document that *Wolbachia* symbionts express reproductively parasitic phenotypes in ants[72]), so their ANK HGAs might also extend the finetuning of mutualistic functions. The highly prevalent ANK HGAs remain enigmatic in many respects. They are not deeply conserved (Fig. 1), which suggest that many of them may be selectively neutral or even slightly deleterious for the ants acquiring them, consistent with only 49% being phenotypically expressed. However, our somewhat ambiguous general correlative results might also be consistent with ANKs often having adaptive functions in gene regulation, which would be consistent with an earlier study in Nasonia wasps[70]. The maintenance of bacterial ANKs in animal genomes thus remain enigmatic, similar to humans and other vertebrates whose genomes also harbor a plethora of ANK repeats[73].

Similar to ANKs, the Aryl-sulfate-sulfotransferase (ASST) detected in *Colobopsis sp.* (Fig.3D) is known to be involved in the regulation of prokaryote-eukaryote interactions. In bacteria, sulfurylated molecules mediate cell-cell and host-pathogen interactions, being especially upregulated in bacterial pathogens such as *Escherichia coli* during host infection. However, ASSTs are also implicated in numerous pathogenic processes, as well as in metabolic pathways for detoxifying endogenous and exogenous compounds with phenolic functional groups[74,75].

Evolutionary implications of HGAs in ants

Ants are one of the most diverse insect families worldwide. Their social family structures, colony sizes and ecological niches vary enormously, and our study indicates that regular HGAs from bacterial endosymbionts may have allowed a number of ant lineages to further finetune their fit to particular ecological niches. This perspective would be consistent with inferred HGA-mediated adaptations in other eukaryotes[76–78]. However, it is important to emphasize that most HGAs become subject to genomic degradation and pseudogenization[79], which was also the case in our present study which showed that > 65% of all HGAs were fragmented and/or not phenotypically expressed. In that light it is actually striking our study did recover rather numerous convincing cases of HGAs that likely mediate adaptive responses to environmental challenge, and often with strong signatures of evolutionary conservation and

404 secondary elaboration (e.g. the incorporation of introns and UTRs) over time. In that sense, the fate of HGA events is the same as that of any other mutation in the genome - they are
406 most likely to persist and not degrade when they convey an adaptive benefit. Only after such positive maintenance directly following HGA can secondary elaborations of new syntenic
408 sequences become part of broader gene regulatory networks that mediate complex phenotypic traits, consistent with conjectures brought forward by Li et al[12].

410

The results reported here should encourage further research, both to extend coverage across
412 the ants (as the 163 GAGA-generated genomes represent just over 1% of the total number of described ant species) and to probe HGA functionality in greater detail at the level of specific
414 tribes or genera.

416 Methods

Taxon sampling

418 The vast majority of all investigated ant genomes were sampled, sequenced and annotated by the Global Ant Genomics Alliance[31,80]. This set included 145 genomes sequenced and
420 assembled by GAGA, as well as 18 previously published genomes (Tab.S9). Our total dataset thus contained 163 species distributed across 99 genera (i.e. 29% of the 347 known genera),
422 from 12 out of the 17 extant subfamilies. These included Amblyoponinae (3), Dolichoderinae (6), Dorylinae (4), Ectatomminae (2), Formicinae (39), Leptanillinae (1), Myrmiciinae (3),
424 Mymicinae (77), Paraponerinae (1), Ponerinae (21), Proceratiinae (4), and Pseudomyrmecinae (2). Twenty of the 163 ant genomes were obtained from short-read stLFR
426 (single-tube long fragment read) data, with the corresponding assemblies showing low contiguity (light grey species names in Fig.1A). Only the remaining 143 genomes were PacBio-
428 sequenced and assembled and had sufficient contiguity to reliably identify HGA candidates. For 15 of the 143 PacBio sequenced species, it was possible to obtain chromosome-resolved
430 genomes using Hi-C sequencing to aid the assembly (Tab.S9). Further information on all genome assemblies, gene annotation, and analyses of genome assembly completeness is
432 provided in Vizueta et al[31].

Detection, validation, and quality assessment of HGA candidates

434 To systematically identify instances of bacteria-to-ant horizontal gene transfer, candidates were predicted by an automated pipeline using a homology-based approach. The sequenced

436 ant genomes were divided into sliding windows of 2 kb (with 500 bp overlap) and blasted against curated prokaryotic and insect genome databases with *mmseqs2*[81] to identify
438 homologous regions. The best blast hit was selected by evalue (-k7,7g) and bitscore (-k8,8gr) for each sliding window. Bacterial and eukaryotic rRNAs in the ant genomes were annotated
440 with *barrnap*[82]. GC content and coverage of sliding windows and complete scaffolds of target genomes were calculated with *minimap2*[83], *samtools*[84], and *bedtools*[85]. Assembled genomes
442 were screened for bacterial scaffolds using previously calculated coverage information for identifying bacterial contaminants, which were removed from the assembly before further
444 processing. The exact criteria to define and remove bacterial scaffolds, together with all other information about the automated pipeline that we used can be found at
446 https://github.com/dinhe878/GAGA-Metagenome-LGT.

448 All individual sliding windows with a High-scoring Segment Pair (HSP) against the bacterial database that did not overlap with a HSP from the insect database were identified as candidate
450 HGAs, after which any sequences less than 500 bp apart were merged into candidate loci for further analyses. The automated HGT finder pipeline resulted in 13,664 potential HGA
452 candidates across the 163 investigated ant genomes, which were used as starting point for downstream analyses. Detailed overview plots with multiple HGA-quality parameters were
454 produced for every predicted candidate locus, in addition to the standard files containing sequence information (Figs.S8–S11, giving examples of analyses of selected HGA
456 candidates). Predicted HGA candidates were filtered further using a conservative approach to systematically reduce false positives, caused by e.g., low-complexity regions,
458 missassemblies, or bacterial contaminations. Filtering was performed in R version 4.1.2 using the packages *data.table, dplyr, tidyr, tidyverse* and *stringr*.
460
To assess appropriate filters for the whole dataset, predicted HGAs from seven randomly
462 selected GAGA genomes were evaluated manually. Parameter distributions were then plotted for all candidate-HGAs to determine cut-off thresholds for removing false-positives (Fig.S12).
464 After that, filtering criteria were fixed for all candidate loci to yield an unbiased selection of high quality HGA candidates. The automated filtering described above resulted in 1,149 HGA loci,
466 containing 7,348 HGA candidates across 134 genomes, which were subject to further manual curation and prokaryotic gene annotation (Tab.S10). All these HGA candidates were
468 subsequently validated, both by inspecting alignments of raw sequencing data against the predicted candidate loci and by PCRs followed by additional Sanger Sequencing.
470
Utlimately, we relied on multiple lines of evidence to avoid false-positives. First, the set of high-
472 quality HGAs that we end up defining, all share a high sequence identity with bacterial

homologs, despite being unambiguously integrated into the assembled ant genomes. Second, the emergence of introns, clear phylogenetic synteny conservation, and RNAseq-based expression validation of several HGAs provided additional lines of evidence for the evolutionary significance of bacteria-to-ant HGT. Third, in ant-clade-specific HGAs, we found essentially the same HGA in multiple independently sampled species, providing strong evidence for the integration, evolutionary conservation, and functionality of formerly bacterial genes in ant hosts that might have allowed for the emergence of clade-specific adaptations. However, manual evaluation of HGAs was also required, which confirmed that a dual approach of combining automatic and manual filtering is likely to be the most reliable method to detect HGAs in ants.

PCR and Sanger Sequencing of HGAs

Genomic DNA of 25 available GAGA samples was extracted using a Chelex protocol to verify incorporation of detected HGAs into their respective ant genomes. PCR Primers were designed with a length of 18 – 22 bp, Tm between 58 – 62 °C and high target specificity (i.e., no off-target binding sites) for all possible HGA candidates. Primer pairs were also required to span the expected amplicon as a fragment of the predicted HGA CDS in combination with the ant DNA in both up- and downstream directions of the HGA (Tab.S2). The amplification of PCR products was verified using agarose gel electrophoresis. Correctly amplified PCR products matching expected size were then sequenced using Sanger Sequencing technology after which chromatograms were re-aligned to the reference genome to confirm HGA presence within the ant genome.

Evaluation of border regions between ant DNA and bacterial HGA

Border regions between HGA and ant sequences were investigated to identify missassemblies and chimeric bacterial-ant scaffolds. For this, available short stLFR and long PacBio reads were mapped to their respective genomes after which reads overlapping a predicted candidate region were extracted and counted. Read counts for 5' and 3' boundary regions were included as filters for all candidates within the dataset. Candidates with less than two reads overlapping the boundary between predicted ant DNA and bacterial DNA were considered missassemblies and excluded. To expand the HGA boundaries appropriately and map reads to the HGA candidates, the average read length distribution across all GAGA genomes, separated into short-read assembled (single-tube long fragment read, stLFR) and long-read assembled genomes (PacBio), was calculated. In total, three read count values were obtained: the number of reads overlapping the start of the HGA candidate, reads overlapping the end of the candidate, and reads covering full length of the HGA, including boundary expansion by 1000 bp (PacBio assembled genomes) or 25 bp (stLFR assembled genomes) in both 5' and 3'

direction. After filtering, clusters of HGA regions in close proximity were merged or divided manually after visual inspection, informed by runs of homology to bacterial sequences according to blast bitscores. Here, all HGA regions were split into single remaining HGA candidate sequences.

Our strict filtering criteria led us to exclude the low-contiguity stLFR-based genome assemblies at this point of the analysis as inspection of mapped short-read data did not allow for conclusive discrimination between assembly artefacts and properly integrated HGA events. The scripts used for filtering and obtaining HGA candidates are available in our GitHub repository (https://github.com/janina-rinke/HGT_in_ants).

Prokaryotic gene annotation and functional analyses

Protein-coding and non-coding genes were annotated for all high-quality HGA candidates, using a combination of Prodigal[86], Kraken 2[87], and DFAST[88]. All high-quality HGA candidates CDS sequences were then blasted against NR and NT databases, bacterial protein sequences included in UniProt90, and TIGRFAM and COG databases. For DFAST we required a minimum-length of 100 bp for all bacterial reference sequences while allowing the metagenome option for incomplete genomes.

Examining gene completeness and identification of fragmented HGAs

To investigate gene completeness and identify fragmented, putatively non-functional HGAs, we extracted start and stop codons of predicted coding gene sequences (CDS) from all resulting DFAST files using SeqKit[89]. Accordingly, parameters reporting query coverage (q_cov), subject coverage of the bacterial reference (s_cov), and e-value were examined to identify cases of incomplete or fragmented HGAs. By default, query sequences with a subject coverage <75% were marked as partial hits by DFAST. We additionally used Geneious Prime[90] to visually inspect open reading frames (ORFs) and completeness of selected HGA candidates. Our analyses concluded that several HGA regions had been too narrowly defined, rendering many CDS of HGAs truncated. To complete such fragmental and undersized HGA candidates resulting from our too conservative filtering, we extended all HGA loci by 1000 bp at the 5' and 3' boundary and annotated again with DFAST. All reannotated sequences were then intersected with the originally predicted CDS using *bedtools*[85] to make sure that we only extended previously obtained loci. A summary covering both the original annotation and the reannotation is provided in Tab. S1, which covers all identified HGAs.

Additionally, we integrated information from UniProt (retrieved with *UniProtR*[91]) to obtain sequences from the closest bacterial homolog from UniProt90. This included gene ontology (GO) terms, protein names and predicted bacterial reference taxa (Fig.S2, Tab.S1).

We finally used RNAseq data available for 130 of the 163 studied ant species to assess gene expression of the HGA loci. The RNAseq data covered different ant castes and developmental stages, collected by the GAGA project Vizueta et al. (2024). First, the raw RNAseq reads were aligned to the HGA genomic regions using STAR v2.7.2b[92] with stringent parameters only allowing >99% identity and >90% alignment lengths (--outFilterMismatchNoverReadLmax 0.01 --outFilterScoreMinOverLread 0.9 --outFilterMatchNminOverLread 0.9). For each ant species, we merged mapped reads from different samples using *samtools*[84] and retained only uniquely mapped reads overlapping with predicted HGA genes. We finally estimated overall gene expression for every candidate HGA and reported them as raw read counts.

Comparative genomic analysis of selected HGA candidates

We analysed in detail all remaining expressed HGAs (read count > 100), which had: i) < 80% coverage of the annotated Uniprot hit (to reduce the possibility of fragmented or wrongly annotated HGAs, while still considering different evolutionary trajectories), ii) at least 65% identity with the identified bacterial donor sequence to ensure bacterial origin, and iii) a complete ORF verified by the NCBI ORF finder. For these HGAs, we manually verified completeness of each CDS by conducting BLAST searches, comparing ORFs, using the GAGA annotations[31] and incorporating RNAseq data.

Gene models of these HGAs were manually refined in Geneious, using transcripts obtained with *StringTie* (default settings, Pertea et al., 2015) as guides. In cases where several exon-intron structures were predicted, we used parsimony to manually select a single representative model based on the RNAseq data. Synteny analyses were conducted for all candidate HGAs occurring in narrow phylogenetic clades of ants to evaluate the conservation of the HGA regions. For this purpose, every HGA locus was extended by 40 kb on each side after which all ant genes and protein sequences within this flanking region were extracted. Minimap2[83] was then used to conduct an all-vs-all alignment after which OrthoFinder[94] was used to determine orthogroups across species. The extent of synteny was plotted with the package *gggenomes*[95]. Finally, the candidate HGA sequences were blasted against all GAGA ant genomes to uncover additional HGA events that might previously have been excluded due to our strict filtering criteria (Tab.S5–S7). Using identified clade-specific HGA sequences as queries, we conducted a local blast against all GAGA genomes to uncover potential additional HGA events which were previously excluded because of our strict filtering criteria. The resulting blast hits were then again intersected with all HGA loci initially predicted by the automatic pipeline using *bedtools*[85], which showed that these additional HGA events had indeed been identified as candidate HGAs by the automated HGT finder pipeline, confirming

582 that no HGA event was missed by that pipeline and that we may have filtered candidate HGAs that were real in our aim to avoid false positives. We extracted the FASTA sequences for all
584 resulting intersected HGA candidates and ran DFAST again to annotate them. We also obtained gene expression and synteny data again for all of these selected clade-specific HGAs
586 to complete the in-depth analyses.

588 We then performed phylogenetic analyses to infer the evolutionary origins of selected HGA events. Datasets were generated from protein sequences obtained with DFAST and
590 homologous proteins (>5 best BLAST hits) were retrieved using NCBI blastp and searching against the non-redundant (nr) database. These sequences were then aligned using MAFFT
592 with default options[96] and processed further to conduct Maximum likelihood phylogenies in IQ-Tree[97] using 100 bootstraps. Phylogenies were visualized and annotated using iTOL v4[98] and
594 analyzed for rooting ambiguity. In particular, we checked whether bacterial and eukaryotic sequences could be split into separate monophyletic groups similar to the approach used by
596 Irwin et al. (2021). The phylogenies were rooted on the branches leading to *Caulobacter sp*. and *Mesorhizobium sp*. (Fig.2A) and on the branches leading to *Catenibacterium mitsuokai, Vibrio alginolyticus*, and *Xenorhabdus poinarii* (Fig.2B). Detailed HGA analyses including
598 RNAseq expression, gene models, and overview plots can be found in separate markdown
600 files, as well as in summary tables for the clade-specific and other HGAs (Tab.S4-S8).

602 # Data and Code Availability

All data and code, used for the detection and analysis of HGAs in this study are available from
604 GitHub and can be found in the repository "HGT_in_ants" (https://github.com/janina-rinke/HGT_in_ants.git) as well as under https://github.com/dinhe878/GAGA-Metagenome-
606 LGT (for the automatic HGT finder pipeline). A ReadMe file gives a detailed overview of all files and scripts included in the folders.

608

References

1. Chou, S., Daugherty, M. D., Peterson, S. B., Biboy, J., Yang, Y., Jutras, B. L., Fritz-Laylin, L. K., Ferrin, M. A., Harding, B. N., Jacobs-Wagner, C., Yang, X. F., Vollmer, W., Malik, H. S. & Mougous, J. D. Transferred interbacterial antagonism genes augment eukaryotic innate immune function. *Nature* **518**, 98–101 (2015).

2. Husnik, F., Nikoh, N., Koga, R., Ross, L., Duncan, R. P., Fujie, M., Tanaka, M., Satoh, N., Bachtrog, D., Wilson, A. C. C., von Dohlen, C. D., Fukatsu, T. & McCutcheon, J. P. Horizontal Gene Transfer from Diverse Bacteria to an Insect Genome Enables a Tripartite Nested Mealybug Symbiosis. *Cell* **153**, 1567–1578 (2013).

3. Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. Horizontal gene transfer accelerates genome innovation and evolution. *Mol. Biol. Evol.* **20**, 1598–1602 (2003).

4. Van Etten, J. & Bhattacharya, D. Horizontal Gene Transfer in Eukaryotes: Not if, but How Much? *Trends Genet.* **36**, 915–925 (2020).

5. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).

6. Dunning Hotopp, J. C. Horizontal gene transfer between bacteria and animals. *Trends Genet.* **27**, 157–163 (2011).

7. Irwin, N. A. T., Pittis, A. A., Richards, T. A. & Keeling, P. J. Systematic evaluation of horizontal gene transfer between eukaryotes and viruses. *Nat. Microbiol.* (2021) doi:10.1038/s41564-021-01026-3.

8. Liu, H., Fu, Y., Jiang, D., Li, G., Xie, J., Cheng, J., Peng, Y., Ghabrial, S. A. & Yi, X. Widespread Horizontal Gene Transfer from Double-Stranded RNA Viruses to Eukaryotic Nuclear Genomes. *J. Virol.* **84**, 11876–11887 (2010).

9. McKenna, D. D. *et al.* The evolution and genomic basis of beetle diversity. *Proc. Natl. Acad. Sci.* **116**, 24729–24737 (2019).

10. Perreau, J. & Moran, N. A. Genetic innovations in animal–microbe symbioses. *Nat. Rev. Genet.* **23**, 23–39 (2022).

11. Kirsch, R., Gramzow, L., Theißen, G., Siegfried, B. D., ffrench-Constant, R. H., Heckel, D. G. & Pauchet, Y. Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: Key events in the evolution of herbivory in beetles. *Insect Biochem. Mol. Biol.* **52**, 33–50 (2014).

12. Li, Y., Liu, Z., Liu, C., Shi, Z., Pang, L., Chen, C., Chen, Y., Pan, R., Zhou, W., Chen, X., Rokas, A., Huang, J. & Shen, X.-X. HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* S009286742200719X (2022) doi:10.1016/j.cell.2022.06.014.

13. Undheim, E. A. B. & Jenner, R. A. Phylogenetic analyses suggest centipede venom arsenals were repeatedly stocked by horizontal gene transfer. *Nat. Commun.* **12**, 818 (2021).

14. Martin, W. F. Eukaryote lateral gene transfer is Lamarckian. *Nat. Ecol. Evol.* **2**, 754–754 (2018).

15. Leger, M. M., Eme, L., Stairs, C. W. & Roger, A. J. Demystifying Eukaryote Lateral Gene Transfer (Response to Martin 2017 DOI: 10.1002/bies.201700115). *BioEssays* **40**, 1700242 (2018).

16. Wheeler, D., Redding, A. J. & Werren, J. H. Characterization of an Ancient Lepidopteran Lateral Gene Transfer. *PLoS ONE* **8**, e59262 (2013).

17. Maruyama, S., Matsuzaki, M., Misawa, K. & Nozaki, H. Cyanobacterial contribution to the genomes of the plastid-lacking protists. *BMC Evol. Biol.* **9**, 197 (2009).

18. Matriano, D. M., Alegado, R. A. & Conaco, C. Detection of horizontal gene transfer in the genome of the choanoflagellate Salpingoeca rosetta. *Sci. Rep.* **11**, 5993 (2021).

19. Yue, J., Sun, G., Hu, X. & Huang, J. The scale and evolutionary significance of horizontal gene transfer in the choanoflagellate Monosiga brevicollis. *BMC Genomics* **14**, 729 (2013).

20. Sun, B.-F., Li, T., Xiao, J.-H., Jia, L.-Y., Liu, L., Zhang, P., Murphy, R. W., He, S.-M. & Huang, D.-W. Horizontal functional gene transfer from bacteria to fishes. *Sci. Rep.* **5**, 18676 (2015).

21. Li, H.-S., Tang, X.-F., Huang, Y.-H., Xu, Z.-Y., Chen, M.-L., Du, X.-Y., Qiu, B.-Y., Chen, P.-T., Zhang, W., Ślipiński, A., Escalona, H. E., Waterhouse, R. M., Zwick, A. & Pang, H. Horizontally acquired antibacterial genes associated with adaptive radiation of ladybird beetles. *BMC Biol.* **19**, 7 (2021).

22. Metcalf, J. A., Funkhouser-Jones, L. J., Brileya, K., Reysenbach, A.-L. & Bordenstein, S. R. Antibacterial gene transfer across the tree of life. *eLife* **3**, e04266 (2014).

23. Moran, Y., Fredman, D., Szczesny, P., Grynberg, M. & Technau, U. Recurrent Horizontal Transfer of Bacterial Toxin Genes to Eukaryotes. *Mol. Biol. Evol.* **29**, 2223–2230 (2012).

24. Verster, K. I., Tarnopol, R. L., Akalu, S. M. & Whiteman, N. K. Horizontal Transfer of Microbial Toxin Genes to Gall Midge Genomes. *Genome Biol. Evol.* **13**, evab202 (2021).

25. Acuna, R., Padilla, B. E., Florez-Ramos, C. P., Rubio, J. D., Herrera, J. C., Benavides, P., Lee, S.-J., Yeats, T. H., Egan, A. N., Doyle, J. J. & Rose, J. K. C. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc. Natl. Acad. Sci.* **109**, 4197–4202 (2012).

26. Danchin, E. G. J., Rosso, M.-N., Vieira, P., de Almeida-Engler, J., Coutinho, P. M., Henrissat, B. & Abad, P. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc. Natl. Acad. Sci.* **107**, 17651–17656 (2010).

27. Lukeš, J. & Husník, F. Microsporidia: A Single Horizontal Gene Transfer Drives a Great Leap Forward. *Curr. Biol.* **28**, R712–R715 (2018).

28. Liu, C., Li, Y., Chen, Y., Chen, X., Huang, J., Rokas, A. & Shen, X. How has horizontal gene transfer shaped the evolution of insect genomes? *Environ. Microbiol.* **25**, 642–645 (2023).

29. Stanhope, M. J., Lupas, A., Italia, M. J., Koretke, K. K., Volker, C. & Brown, J. R. Phylogenetic analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature* **411**, 940–944 (2001).

30. Wernegreen, J. J. Endosymbiont evolution: predictions from theory and surprises from genomes. *Ann. N. Y. Acad. Sci.* **1360**, 16–35 (2015).

31. Vizueta, J., Xiong, Z., Ding, G., Larsen, R. S., Ran, H., Gao, Q., Stiller, J., Dai, W., Jiang, W., Zhao, J., Guo, C., Zhang, X., Zuo, D., Zhong, W., Schiøtt, M., Liu, C., Zhang, H., Dai, X., Andreu, I., Shi, Y., Tretter, S., He, D., Gautam, S., Li, Z., Hickey, G., Ivens, A., Meurville, M-P., Hita-Garcia, F., Kass, J. M., Guenard, B., Moreau, C., Paten, B., LeBoeuf, A. C., Economo, E. P., consortium, GAGA, Chapuisat, M., Shik, J. Z., Ward, P. S., Heinze, J., Schultz, T. R., Li, Q., Dunn, R. R., Sanders, N. J., Liu, W., Schrader, L., Boomsma, J. J., Zhang, G., The adaptive radiation and social evolution of the ants, *under review*, *Cell.* Available at SSRN: http://dx.doi.org/10.2139/ssrn.5055090.

32. Dhaygude, K., Nair, A., Johansson, H., Wurm, Y. & Sundström, L. The first draft genomes of the ant Formica exsecta, and its Wolbachia endosymbiont reveal extensive gene transfer from endosymbiont to host. *BMC Genomics* **20**, 301 (2019).

33. Klein, A., Schrader, L., Gil, R., Manzano-Marín, A., Flórez, L., Wheeler, D., Werren, J. H., Latorre, A., Heinze, J., Kaltenpoth, M., Moya, A. & Oettler, J. A novel intracellular mutualistic bacterium in the invasive ant Cardiocondyla obscurior. *ISME J.* **10**, 376–388 (2016).

34. Van Borm, S., Wenseleers, T., Billen, J. & Boomsma, J. J. *Wolbachia* in leafcutter ants: a widespread symbiont that may induce male killing or incompatible matings. *J. Evol. Biol.* **14**, 805–814 (2001).

35. Zhukova, M., Sapountzis, P., Schiøtt, M. & Boomsma, J. J. Diversity and Transmission of Gut Bacteria in Atta and Acromyrmex Leaf-Cutting Ants during Development. *Front. Microbiol.* **8**, 1942 (2017).

36. Werren, J. H. Biology Of *Wolbachia*. *Annu. Rev. Entomol.* **42**, 587–609 (1997).

37. Hotopp, J. C. D. *et al.* Widespread Lateral Gene Transfer from Intracellular Bacteria to Multicellular Eukaryotes. *Science* **317**, 1753–1756 (2007).

38. Nikoh, N., Tanaka, K., Shibata, F., Kondo, N., Hizume, M., Shimada, M. & Fukatsu, T. Wolbachia genome integrated in an insect chromosome: Evolution and fate of laterally transferred endosymbiont genes. *Genome Res.* **18**, 272–280 (2008).

39. Wang, G. H., Sun, B. F., Xiong, T. L., Wang, Y. K., Murfin, K. E., Xiao, J. H. & Huang, D. W. Bacteriophage WO Can Mediate Horizontal Gene Transfer in Endosymbiotic *Wolbachia* Genomes. *Front. Microbiol.* **7**, (2016).

40. Feldhaar, H., Straka, J., Krischke, M., Berthold, K., Stoll, S., Mueller, M. J. & Gross, R. Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. *BMC Biol.* **5**, 48 (2007).

41. Jackson, R., Monnin, D., Patapiou, P. A., Golding, G., Helanterä, H., Oettler, J., Heinze, J., Wurm, Y., Economou, C., Chapuisat, M. & Henry, L. Convergent Evolution of a Nutritional

720 Symbiosis in Ants. https://www.researchsquare.com/article/rs-830142/v1 (2021) doi:10.21203/rs.3.rs-830142/v1.

722 42. Jaeger, T., Arsic, M. & Mayer, C. Scission of the Lactyl Ether Bond of N-Acetylmuramic Acid by *Escherichia coli* "Etherase". *J. Biol. Chem.* **280**, 30100–30106 (2005).

724 43. Jaeger, T. & Mayer, C. N-acetylmuramic acid 6-phosphate lyases (MurNAc etherases): role in cell wall metabolism, distribution, structure, and mechanism. *Cell. Mol. Life Sci.* **65**, 928–939 726 (2008).

44. Ding, J., Wang, R., Yang, F., Zhao, L., Qin, Y., Zhang, G. & Yan, X. Identification and 728 characterization of a novel phage-type like lysozyme from Manila clam, *Ruditapes philippinarum*. *Dev. Comp. Immunol.* **47**, 81–89 (2014).

730 45. Ren, Q., Wang, C., Jin, M., Lan, J., Ye, T., Hui, K., Tan, J., Wang, Z., Wyckoff, G. J., Wang, W. & Han, G.-Z. Co-option of bacteriophage lysozyme genes by bivalve genomes. *Open Biol.* 732 **7**, 160285 (2017).

46. Cremer, S., Armitage, S. A. O. & Schmid-Hempel, P. Social Immunity. *Curr. Biol.* **17**, R693–734 R702 (2007).

47. Uehara, T., Suefuji, K., Jaeger, T., Mayer, C. & Park, J. T. MurQ Etherase Is Required by 736 *Escherichia coli* in Order To Metabolize Anhydro- *N* -Acetylmuramic Acid Obtained either from the Environment or from Its Own Cell Wall. *J. Bacteriol.* **188**, 1660–1662 (2006).

738 48. Walter, A. & Mayer, C. Peptidoglycan Structure, Biosynthesis, and Dynamics During Bacterial Growth. in *Extracellular Sugar-Based Biopolymers Matrices* (eds. Cohen, E. & Merzendorfer, 740 H.) vol. 12 237–299 (Springer International Publishing, Cham, 2019).

49. Kautz, S., Rubin, B. E. R. & Moreau, C. S. Bacterial Infections across the Ants: Frequency and 742 Prevalence of *Wolbachia, Spiroplasma* , and *Asaia*. *Psyche J. Entomol.* **2013**, 1–11 (2013).

50. Grandvalet, C., Assad-García, J. S., Chu-Ky, S., Tollot, M., Guzzo, J., Gresti, J. & Tourdot-744 Maréchal, R. Changes in membrane lipid composition in ethanol- and acid-adapted *Oenococcus oeni* cells: characterization of the cfa gene by heterologous complementation. 746 *Microbiology* **154**, 2611–2619 (2008).

51. Grogan, D. W. & Cronan, J. E. Cyclopropane ring formation in membrane lipids of bacteria. 748 *Microbiol. Mol. Biol. Rev.* **61**, 429–441 (1997).

52. Jiang, X., Duan, Y., Zhou, B., Guo, Q., Wang, H., Hang, X., Zeng, L., Jia, J. & Bi, H. The 750 Cyclopropane Fatty Acid Synthase Mediates Antibiotic Resistance and Gastric Colonization of *Helicobacter pylori*. *J. Bacteriol.* **201**, (2019).

752 53. Oyola, S. O., Evans, K. J., Smith, T. K., Smith, B. A., Hilley, J. D., Mottram, J. C., Kaye, P. M. & Smith, D. F. Functional Analysis of *Leishmania* Cyclopropane Fatty Acid Synthetase. *PLoS* 754 *ONE* **7**, e51300 (2012).

54. Yuan, Y. & Barry, C. E. A common mechanism for the biosynthesis of methoxy and cyclopropyl 756 mycolic acids in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* **93**, 12828–12833 (1996).

55. Bao, X., Thelen, J. J., Bonaventure, G. & Ohlrogge, J. B. Characterization of Cyclopropane
758 Fatty-acid Synthase from *Sterculia foetida*. *J. Biol. Chem.* **278**, 12846–12853 (2003).

56. Liu, Y., Srivilai, P., Loos, S., Aebi, M. & Kües, U. An Essential Gene for Fruiting Body Initiation
760 in the Basidiomycete *Coprinopsis cinerea* Is Homologous to Bacterial Cyclopropane Fatty Acid Synthase Genes. *Genetics* **172**, 873–884 (2006).

762 57. Beach, D. H., Holz, G. G. & Anekwe, G. E. Lipids of *Leishmania Promastigotes*. *J. Parasitol.* **65**, 203 (1979).

764 58. Peacock, C. S. *et al.* Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat. Genet.* **39**, 839–847 (2007).

766 59. Xu, W., Mukherjee, S., Ning, Y., Hsu, F.-F. & Zhang, K. Cyclopropane fatty acid synthesis affects cell shape and acid resistance in *Leishmania mexicana. Int. J. Parasitol.* **48**, 245–256
768 (2018).

60. Borowiec, M. L., Cover, S. P. & Rabeling, C. The evolution of social parasitism in *Formica* ants
770 revealed by a global phylogeny. *Proc. Natl. Acad. Sci.* **118**, e2026029118 (2021).

61. Szabó, G., Schulz, F., Toenshoff, E. R., Volland, J.-M., Finkel, O. M., Belkin, S. & Horn, M.
772 Convergent patterns in the evolution of mealybug symbioses involving different intrabacterial symbionts. *ISME J.* **11**, 715–726 (2017).

774 62. Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: Mobile modules that cross phyla horizontally? *Proteins Struct. Funct. Genet.* **17**, 363–374 (1993).

776 63. Jernigan, K. K. & Bordenstein, S. R. Ankyrin domains across the Tree of Life. *PeerJ* **2**, e264 (2014).

778 64. Li, J., Mahajan, A. & Tsai, M.-D. Ankyrin Repeat: A Unique Motif Mediating Protein−Protein Interactions. *Biochemistry* **45**, 15168–15178 (2006).

780 65. Mosavi, L. K., Cammett, T. J., Desrosiers, D. C. & Peng, Z. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13**, 1435–1448 (2004).

782 66. Jahn, M. T., Arkhipova, K., Markert, S. M., Stigloher, C., Lachnit, T., Pita, L., Kupczok, A., Ribes, M., Stengel, S. T., Rosenstiel, P., Dutilh, B. E. & Hentschel, U. A Phage Protein Aids
784 Bacterial Symbionts in Eukaryote Immune Evasion. *Cell Host Microbe* **26**, 542-550.e5 (2019).

67. Pan, X., Lührmann, A., Satoh, A., Laskowski-Arce, M. A. & Roy, C. R. Ankyrin Repeat Proteins
786 Comprise a Diverse Family of Bacterial Type IV Effectors. *Science* **320**, 1651–1654 (2008).

68. Siozios, S., Ioannidis, P., Klasson, L., Andersson, S. G. E., Braig, H. R. & Bourtzis, K. The
788 Diversity and Evolution of *Wolbachia* Ankyrin Repeat Domain Genes. *PLoS ONE* **8**, e55390 (2013).

790 69. Voronin, D. A. & Kiseleva, E. V. Functional role of proteins containing ankyrin repeats. *Cell Tissue Biol.* **2**, 1–12 (2008).

70. Pers, D. & Lynch, J. A. Ankyrin domain encoding genes from an ancient horizontal transfer are functionally integrated into *Nasonia* developmental gene regulatory networks. *Genome Biol.* **19**, 148 (2018).

71. Klasson, L., Kambris, Z., Cook, P. E., Walker, T. & Sinkins, S. P. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti. BMC Genomics* **10**, 33 (2009).

72. Andersen, S. B., Boye, M., Nash, D. R. & Boomsma, J. J. Dynamic *Wolbachia* prevalence in *Acromyrmex* leaf-cutting ants: potential for a nutritional symbiosis. *J. Evol. Biol.* **25**, 1340–1350 (2012).

73. Islam, Z., Nagampalli, R. S. K., Fatima, M. T. & Ashraf, G. M. New paradigm in ankyrin repeats: Beyond protein-protein interaction module. *Int. J. Biol. Macromol.* **109**, 1164–1173 (2018).

74. Ayuso-Fernández, I., Galmés, M. A., Bastida, A. & García-Junceda, E. Aryl Sulfotransferase from *Haliangium ochraceum*: A Versatile Tool for the Sulfation of Small Molecules. *ChemCatChem* **6**, 1059–1065 (2014).

75. Malojčić, G., Owen, R. L. & Glockshuber, R. Structural and Mechanistic Insights into the PAPS-Independent Sulfotransfer Catalyzed by Bacterial Aryl Sulfotransferase and the Role of the DsbL/DsbI System in Its Folding. *Biochemistry* **53**, 1870–1877 (2014).

76. Chen, R., Huangfu, L., Lu, Y., Fang, H., Xu, Y., Li, P., Zhou, Y., Xu, C., Huang, J. & Yang, Z. Adaptive innovation of green plants by horizontal gene transfer. *Biotechnol. Adv.* **46**, 107671 (2021).

77. Husnik, F. & McCutcheon, J. P. Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* **16**, 67–79 (2018).

78. Schönknecht, G., Chen, W.-H., Ternes, C. M., Barbier, G. G., Shrestha, R. P., Stanke, M., Bräutigam, A., Baker, B. J., Banfield, J. F., Garavito, R. M., Carr, K., Wilkerson, C., Rensing, S. A., Gagneul, D., Dickenson, N. E., Oesterhelt, C., Lercher, M. J. & Weber, A. P. M. Gene Transfer from Bacteria and Archaea Facilitated Evolution of an Extremophilic Eukaryote. *Science* **339**, 1207–1210 (2013).

79. Xing, B., Yang, L., Gulinuer, A. & Ye, G. Research progress on horizontal gene transfer and its functions in insects. *Trop. Plants* **2**, 1–12 (2023).

80. Boomsma, J. J., Brady, S. G., Dunn, R. R., Gadau, J., Heinze, J., Keller, L., Moreau, C. S., Sanders, N. J., Schrader, L. & Schultz, T. R. The global ant genomics Alliance (GAGA). *Myrmecol. News* (2017).

81. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).

82. Seemann, T. barrnap 0.9: rapid ribosomal RNA prediction. *Google Sch.* (2013).

83. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

84. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

85. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

86. Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W. & Hauser, L. J. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

87. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).

88. Tanizawa, Y., Fujisawa, T. & Nakamura, Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* **34**, 1037–1039 (2018).

89. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE* **11**, e0163962 (2016).

90. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. & Drummond, A. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

91. Soudy, M., Anwar, A. M., Ahmed, E. A., Osama, A., Ezzeldin, S., Mahgoub, S. & Magdeldin, S. UniprotR: Retrieving and visualizing protein sequence and functional information from Universal Protein Resource (UniProt knowledgebase). *J. Proteomics* **213**, 103613 (2020).

92. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

93. Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T. & Salzberg, S. L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

94. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, (2019).

95. Hackl, T. & Ankenbrand, M. J. gggenomes: a grammar of graphics for comparative genomics. *R Package Version 09* **5**, (2022).

96. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

97. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

98. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).

Acknowledgements

We thank the University of Münster PALMA Computational Service and Danish National Life Science Supercomputing Center, Computerome, for providing computational resources. We thank S. Mathiasen, N. Kortüm, and N. Vo for assisting with the laboratory work. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 502787686 to L.S. under the Priority Programme SPP 2349 and supported by the Villum Foundation (Villum Investigator Grant, grant no. 25900 to G.Z.).

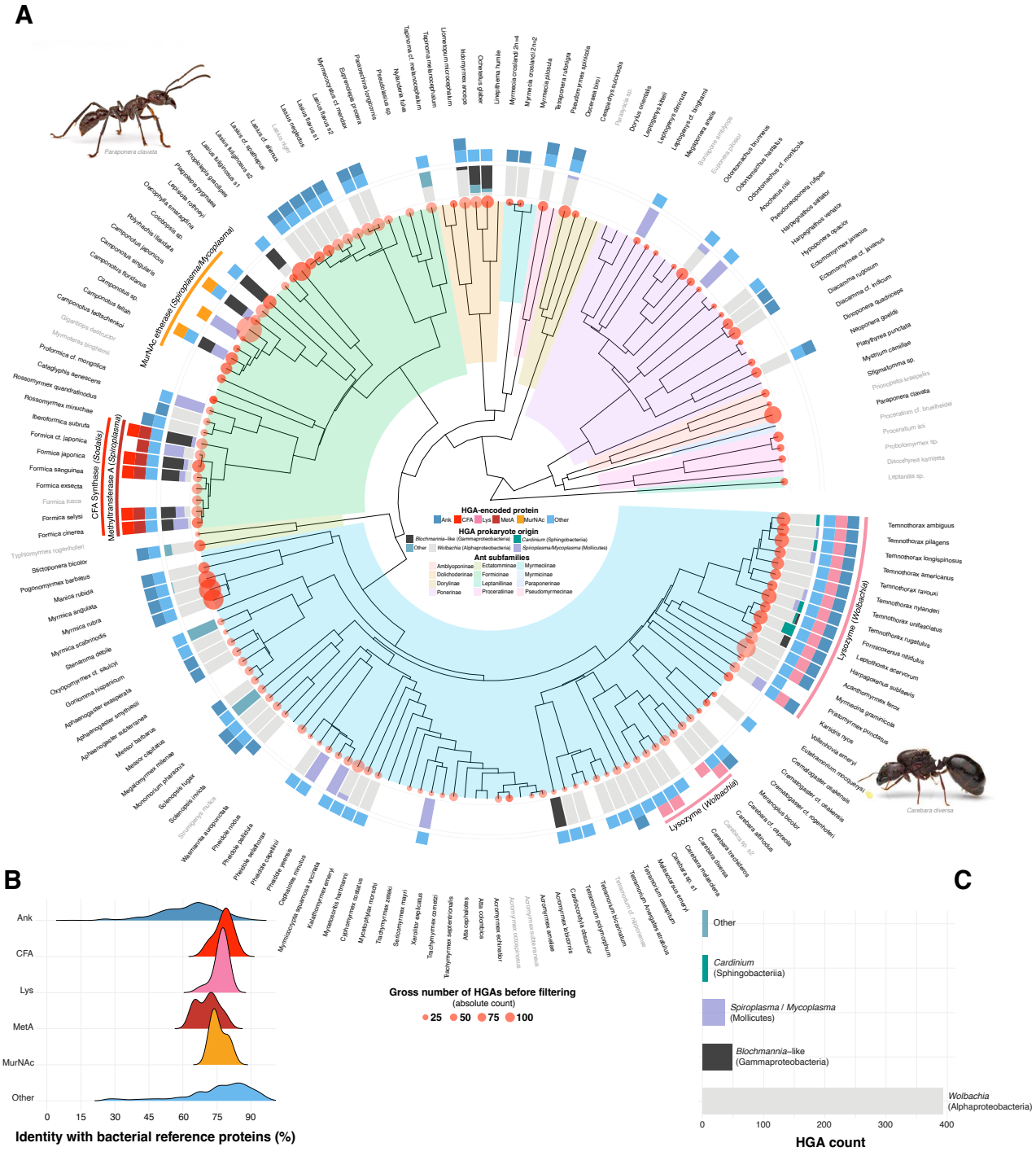
Author contributions

L.S. designed the study and discussed it with J.J.B and G.Z., after which J.R., L.S. and R.S.L. designed, supervised and performed the molecular experiments. J.R., L.S., D.H., Z.X.,J.V. developed and performed bioinformatics analyses. J.R. and L.S. drafted the manuscript together with J.G., J.J.B, and G.Z., D.H., and J.V. All authors contributed to the interpretation of the results and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Figures



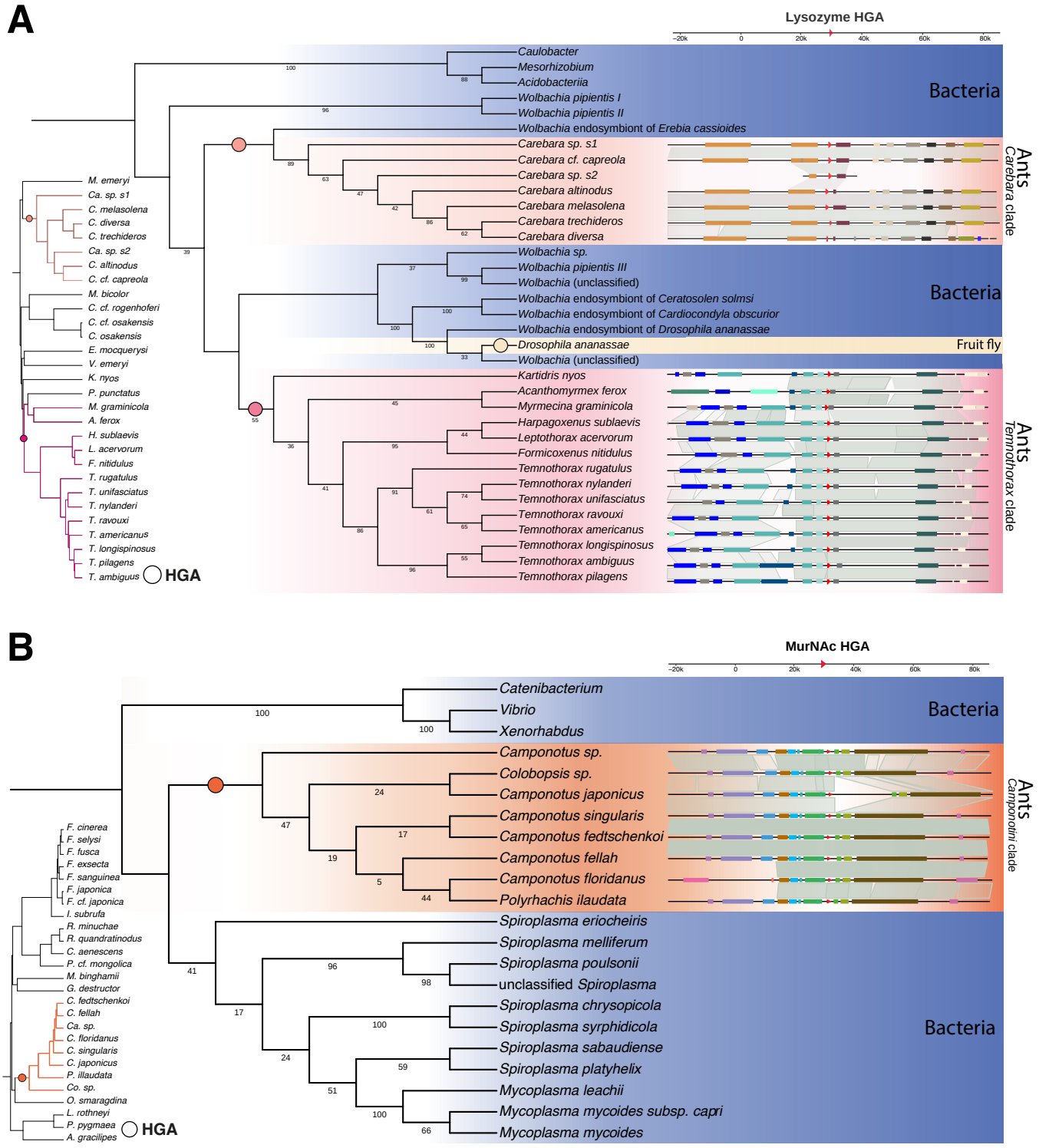
884

Figure 1. Phylogenetic distribution, prevalence, and origins of bacterial HGAs in ants.

886 **A.** Species phylogeny of the 163 analyzed ant genomes and overview information on presence/absence and origin of the bacteria-to-ant HGTs detected by the automated HGT

888 finder pipeline. Background clade colors in the phylogeny specify different ant subfamilies. The number of candidate HGA loci identified before manual curation and gene annotation (n

= 1,148 loci harboring 7,348 putative HGA events) is indicated by red points at the branch tips. Stacked bar plots next to the branch tips show the prokaryotic origin of HGAs as fractions. The most prevalent donors are *Wolbachia* (light grey), followed by *Blochmannia*-like bacteria (black), *Spiroplasma/Mycoplasma* (purple), *Cardinium* (dark green), and other bacteria (turquoise). The outer circle indicates the presence/absence of HGA-encoded proteins, of which ankyrin repeats (Ank, dark blue) were most abundant. Cyclopropane formic acid (CFA) synthases (red) are restricted to the Formicini tribe, similar to RNA methyltransferases (MetA, dark red). Lysozymes (Lys) were detected in *Carebara spp.*, as well as in *Temnothorax spp.* and closely related genera (pink), while *N*-acetyl-muramic acid etherases (MurNAc) were detected only in Camponotini ants (orange). Other identified proteins are highlighted in light blue (Tab. S1). Additional candidates, which were detected during in-depth analyses of clade-specific HGAs (CFA, Lys, MetA, MurNAc) are not highlighted in the Figure, but mentioned and described in the main text, as well as in Fig.2 and in Tab.S5-S7). Names of species with short-read stLFR genome assemblies are printed in grey. **B.** Percentage identity of HGA loci with their inferred bacterial donor proteins based on CDS sequences for the categories: Ank, CFA, Lys, MetA, MurNAc, Other (cf. panel **A**). All conserved clade specific HGAs (CFA, Lys, MetA, and MurNAc) had around 75% (range ca 60-90%) sequence identity with their inferred bacterial donor sequence while ANKs and other (unspecified) HGAs that occurred across many ant subfamilies had a broader range (20 – 100 %). **C.** Taxonomic distribution of bacterial HGA donors in ants. We inferred the bacterial origin of the 497 identified HGA loci based on their prokaryotic gene annotations. *Wolbachia* (Alphaproteobacteria) were detected as donors in 79 % of the cases (n = 393), followed by *Blochmannia*-like Gammaproteobacteria (n = 49), and *Spiroplasma/Mycoplasma* (Mollicutes, n = 37). Other bacterial donors were *Cardinium* (n = 9) and a number of not further specified bacteria (n = 9, Tab.S1). Photo credits: *Paraponera clavate* ©Global Ant Genomics Alliance; *Carebara diversa* ©Eduard Florin Niga.



916 **Figure 2. Representation of selected ancient orthologous HGAs. A.** Phylogenetic tree and synteny of lysozyme (Lys) loci incorporated in the genomes of two clades of Crematogastrini
918 ants, in *Drosophila ananassae*, and in the source bacteria. The phylogeny was rooted on the bacterial branch leading to *Caulobacter, Mesorhizobium and Acidobacteriia*. The independent
920 HGA of the fruit fly *D. ananassae* emanated from a BLAST search but was not investigated further. Different colors in the top and bottom synteny plots indicate putative independent HGA
922 events. Orthologous genes are drawn in the same color while the focal HGA is shown as a

red triangle in the synteny on the right. Homologous regions are visualized by grey bars. **B.** Phylogenetic tree and synteny of MurNAc genes incorporated in the genomes of Camponotini ants and their matching bacterial source sequences based on the best hits from NCBI BLAST. The gene tree has been calculated with the protein sequences of all respective species and was rooted on the bacterial branch leading to *Catenibacterium, Vibrio,* and *Xenorhabdus*.

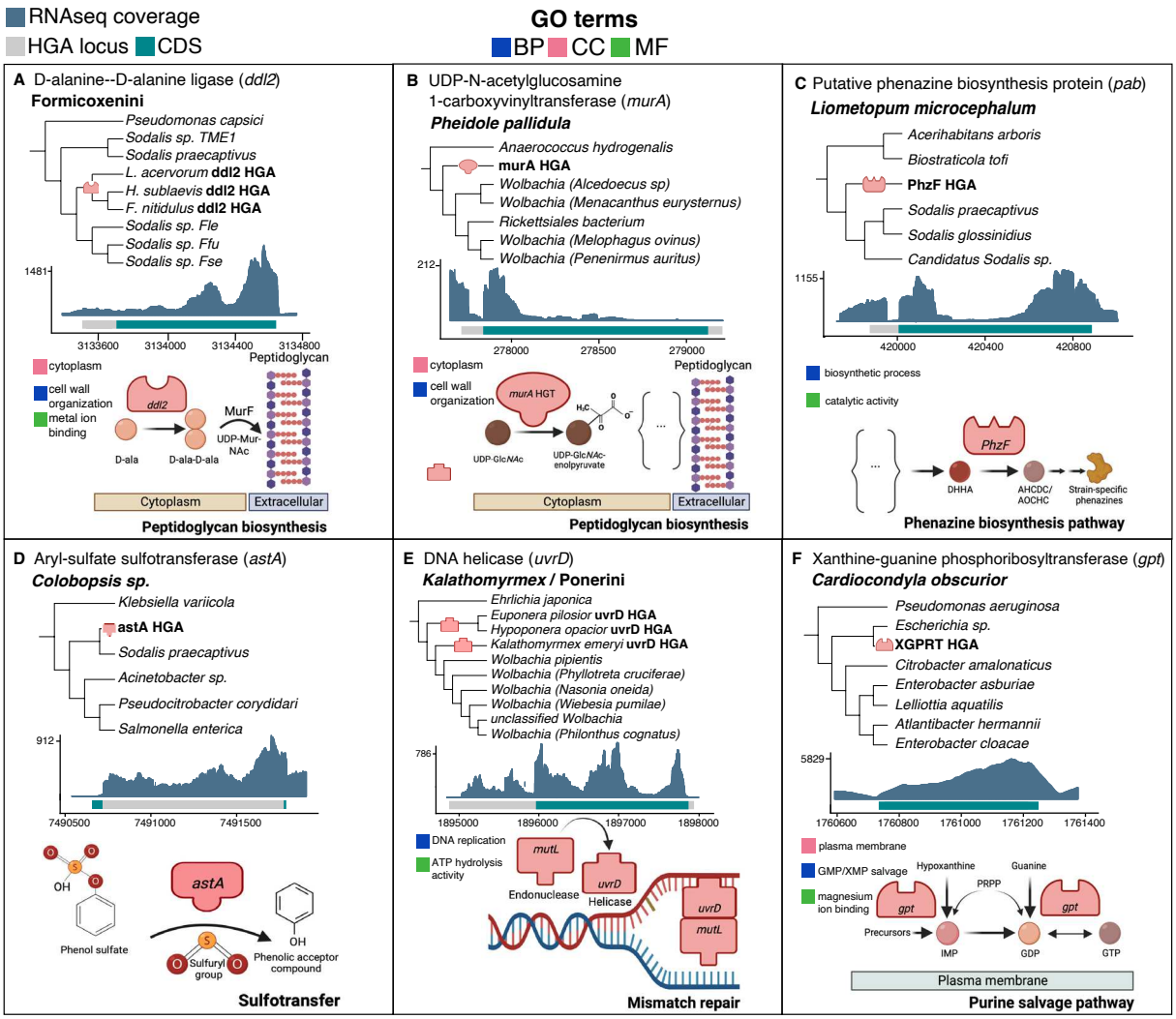


Figure 3. In-depth analysis of expressed HGAs other than those presented in Figure 2. Six expressed and full-length HGAs were analyzed and illustrated in separate panels (A-F) with a rooted phylogenetic gene tree including homologous bacterial proteins. RNAseq coverage is visualized for each HGA locus with CDS regions shown in cyan. The putative function of each HGA is illustrated at the bottom of each panel. The focal HGA proteins are drawn in red and the associated Gene Ontology (GO) terms from incorporated UniProt data are given in other colors: Biological Process (BP; blue), Cellular Component (CC; pink) and Molecular Function (MF; green).

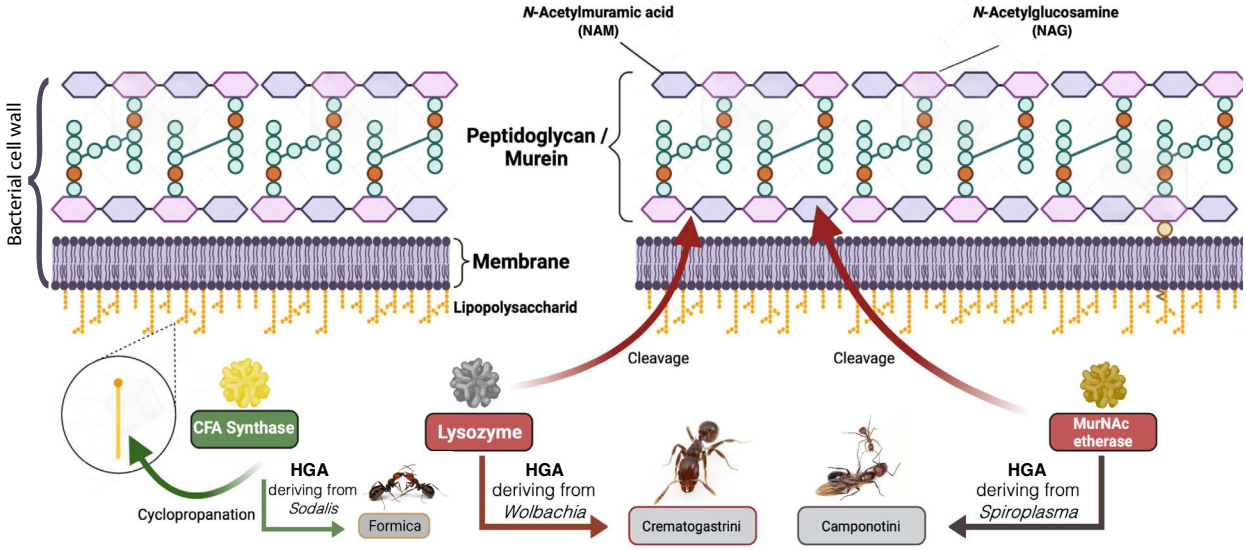


Figure 4. Schematic representation of conserved bacterial HGAs in ants with functions related to bacterial cell wall degradation. A bacterial cell wall consists of peptidoglycan/murein, accompanied by a membrane with associated lipopolysaccharides. The monosaccharide NAM occurs ubiquitously in the cell walls of gram-positive and gram-negative bacteria, forming the backbone of peptidoglycan together with *N*-acetylglucosamine (NAG). CFA Synthases are involved in the cyclopropanation of lipopolysaccharides, associated with bacterial stress responses and were detected as HGAs in *Formica* ants. Lysozymes cause a cleavage of peptidoglycan by acting on the bond between NAG and NAM (conserved in Crematogastrini ants), while *murQ* genes encode *N*-acetylmuramic acid 6-phosphate etherases (MurNAc etherases), which are bacteria-specific enzymes that can act upon NAM itself (conserved in Camponotini).

Supplementary Information

Figure S1. Synteny visualization of recurrent HGAs outside the clade-specific HGAs.

Figure S2. Gene Ontology Enrichment Analysis of all HGAs.

Figure S3. Number and variety of ANK HGA gene loci visualized across the GAGA ant phylogeny.

Figure S4. Regression of ANK expression on % identity with bacterial reference proteins.

Figure S5. Lysozyme gene models inferred from StringTie.

Figure S6. Gene Tree of full-length CFA synthases within the Formicini tribe together with expression information.

Figure S7. Synteny visualization of all CFA synthases within the Formicini tribe.

Figure S8-S11. Example overview plots for several HGA candidates.

Figure S12. Parameters used for automated filtering of HGA candidates predicted by the HGT finder pipeline.

Table S1. A. Details about the 497 distinct HGA candidate loci detected in this study, with details about Ankyrin Repeat proteins (ANKs), lysozymes, MurNAc etherases, and CFA synthases, and the species-or clade-specific HGA loci identified. **B.** Details about the 1,053 CDS identified within the 497 HGA loci.

Table S2. Information about PCRs and Sanger Sequencing performed to validate HGA candidates in ant genomes.

Table S3. Comparison of detected HGAs in this study with previously identified HGTs in past studies covering 22 ant genomes.

Table S4. Synteny information about multiple other HGAs identified in ant genomes outside the clade-specific HGAs.

Table S5. Information about clade-specific lysozyme HGAs occurring in 21 ant species.

Table S6. Information about clade-specific MurNAc HGAs occurring in Camponotini ants.

Table S7. Information about clade-specific CFA synthase HGAs identified in Formicini ants.

Table S8. Information about other HGAs used for in-depth analysis with expression information.

Table S9. The examined GAGA genomes used as dataset for this study, including GAGA ID, species information and sequencing method.

Table S10. Information about all 1,148 HGA loci containing 7,348 candidates predicted by the HGT finder pipeline after automated filtering.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1HGAsallinfo.xlsx
- TableS2PCRvalidations.xlsx
- TableS3ComparisonpreviousHGTsants.xlsx
- TableS4SyntenyotherHGAloci.xlsx
- TableS5LysozymeHGAs.xlsx
- TableS6MurNAcHGAs.xlsx
- TableS7CFAHGAs.xlsx
- TableS8IndepthOtherHGAssummary.xlsx
- TableS9GAGAspecieslist.txt
- TableS10GAGA.HGAs.afterFil.preCuration.xlsx
- FigureS1syntenyotherHGAs.pdf
- FigureS2GOEnrichmentAnalysis.png
- FigureS3AnksPhylo.pdf
- FigureS4ANKs.expression.pid.Correlation.pdf
- FigureS5LysozymesGeneModels.pdf
- FigureS6CFAGeneTree.pdf
- FigureS7CFAsynteny.pdf
- FigureS8OverviewHGAXGPRTGAGA0515.Scaffold817607781761241.pdf
- FigureS9OverviewHGALysGAGA0099.Scaffold2026581622658792.pdf
- FigureS10OverviewHGACFAGAGA0485.Scaffold654373575438799.pdf
- FigureS11OverviewHGAAnkGAGA0087.Scaffold1546696304672999.pdf
- FigureS12HGAfilteringparameters.pdf