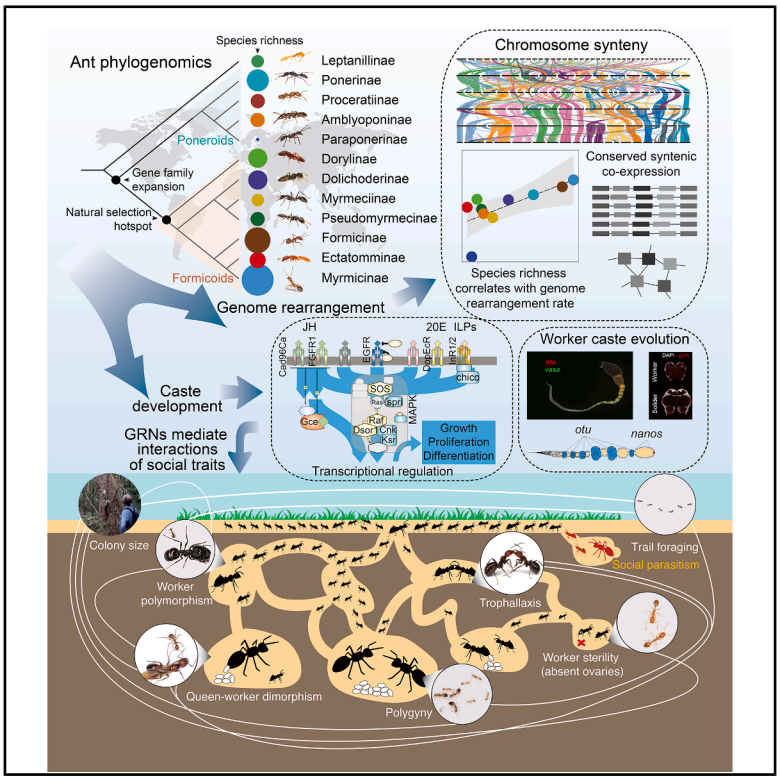
Cell

Article

Adaptive radiation and social evolution of the ants

Graphical abstract



Highlights

- Comparative analyses of 163 genomes uncover major adaptive changes in ant evolution
- Caste-associated genes maintained synteny despite high rates of macrosynteny loss
- Juvenile hormone, insulin, and MAPK pathways regulate queen-worker caste differentiation
- Co-evolving social traits left coherent signatures of selection on overlapping gene sets

Authors

Joel Vizueta, Zijun Xiong, Guo Ding, ..., Lukas Schrader, Jacobus J. Boomsma, Guojie Zhang

Correspondence

liuweiwei@mail.kiz.ac.cn (W.L.), lukas.schrader@uni-muenster.de (L.S.), jjboomsma@bio.ku.dk (J.J.B.), guojiezhang@zju.edu.cn (G.Z.)

In brief

Comparative analyses of 163 ant genomes reveal extensive genome rearrangements, context-specific gene family expansion patterns, and selection on conserved pathways that together underpin the evolution and diversification of social traits and caste differentiation in ants.

Vizueta et al., 2025, Cell *188*, 4828–4848
September 4, 2025 © 2025 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2025.05.030

CellPress

CellPress
OPEN ACCESS

Cell

Article

Adaptive radiation and social evolution of the ants

Joel Vizueta,[1,31] Zijun Xiong,[2,3,31,32] Guo Ding,[4,31] Rasmus S. Larsen,[1] Hao Ran,[4,5] Qionghua Gao,[6] Josefin Stiller,[1] Wei Dai,[3] Wei Jiang,[7] Jie Zhao,[2] Chunxue Guo,[3] Xiafang Zhang,[2] Dashuang Zuo,[2] Wenjiang Zhong,[2] Morten Schiøtt,[8] Chengyuan Liu,[4] Hailin Zhang,[2] Xueqin Dai,[2] Ignasi Andreu,[1] Yue Shi,[1] Sandra Tretter,[9,10] Ding He,[11] Shubham Gautam,[12] Zelin Li,[1] Glenn Hickey,[13] Aniek B.F. Ivens,[14] Marie-Pierre Meurville,[15] Francisco Hita-Garcia,[12,16]

(Author list continued on next page)

[1]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark
[2]State Key Laboratory of Genetic Evolution & Animal Models, Kunming Institute of Zoology, Chinese Academy of Sciences, 650201 Kunming, China
[3]BGI Research, Wuhan, China
[4]Center for Evolutionary & Organismal Biology, Women's Hospital & Liangzhu Laboratory, School of Medicine, Zhejiang University, Hangzhou, China
[5]Biological Education and Research Laboratory, Mancheng High School of Hebei Province, Baoding, China
[6]Guangxi Key Laboratory of Agro-Environmental and Agric-Products Safety/National Demonstration Center for Experimental Plant Science Education, College of Agriculture, Guangxi University, 530004 Nanning, Guangxi, China
[7]State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, 518083 Shenzhen, China
[8]Section for Protein and Enzyme Technology, Department of Biotechnology and Biomedicine, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[9]Max Planck Institute for Chemical Ecology, Jena, Germany
[10]Institute for Evolution & Biodiversity, University of Münster, Münster, Germany
[11]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[12]Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan
[13]UC Santa Cruz Genomics Institute, University of California, Santa Cruz, Santa Cruz, CA, USA
[14]Ecology & Evolution Section, A-Life, Vrije Universiteit, De Boelelaan 1108, 1081 HV Amsterdam, the Netherlands
[15]Department of Biology, University of Fribourg, Fribourg, Switzerland
[16]Center for Integrative Biodiversity Discovery, Museum für Naturkunde, Berlin, Germany
[17]Macroecology Lab, Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi, Japan
[18]School of Biological Sciences, The University of Hong Kong, Hong Kong SAR, China
[19]Department of Entomology and Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA
[20]Department of Zoology, University of Cambridge, Cambridge, UK
[21]Department of Entomology, University of Maryland, College Park, MD, USA

(Affiliations continued on next page)

SUMMARY

Ants originated over 150 million years ago through an irreversible transition to superorganismal colony life. Comparative analyses of 163 ant genomes, including newly generated whole-genome sequences of 145 ant species, reveal extensive genome rearrangements correlated with speciation rates. Meanwhile, conserved syntenic blocks are enriched with co-expressed genes involved in basal metabolism and caste differentiation. Gene families related to digestion, endocrine signaling, cuticular hydrocarbon synthesis, and chemoreception expanded in the ant ancestor, while many caste-associated genes underwent positive selection in the formicoid ancestor. Elaborations and reductions of queen-worker dimorphism and other social traits left convergent signatures of intensified or relaxed selection in conserved signaling and metabolic pathways, suggesting that a core gene set was used to diversify organizational complexity. Previously uncharacterized genetic regulators of caste development were confirmed by functional experiments. This study reconstructs the genetic underpinning of social traits and their integration within gene-regulatory networks shaping caste phenotypes.

INTRODUCTION

The ants are among the world's most successful animal lineages. Their caste-differentiated colonies originated over 150 million years ago (mya)[1–3] and became dominant parts of terrestrial ecosystems in numbers and biomass.[4–6] This success is rooted in division of labor comparable to what somatic cells in a metazoan body achieve. For example, workers typically do

 © 2025 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

CellPress
OPEN ACCESS

Jamie M. Kass,[12,17] Benoit Guénard,[18] Corrie Moreau,[19] Benedict Paten,[13] Adria C. LeBoeuf,[20] Evan P. Economo,[12,21] GAGA Consortium, Michel Chapuisat,[22] Jonathan Z. Shik,[1,23] Philip S. Ward,[24] Jürgen Heinze,[25] Ted R. Schultz,[26] Qiye Li,[3,7,27] Robert R. Dunn,[28] Nathan J. Sanders,[29] Weiwei Liu,[2,*] Lukas Schrader,[10,*] Jacobus J. Boomsma,[1,*] and Guojie Zhang[4,30,33,*]

[22]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland
[23]Smithsonian Tropical Research Institute, Apartado Postal, Balboa, Ancon, 0843-03092 Panama City, Panama
[24]Department of Entomology and Nematology, University of California, Davis, Davis, CA, USA
[25]Zoology/Evolutionary Biology, University of Regensburg, Regensburg 93040, Germany
[26]Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA
[27]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China
[28]Department of Applied Ecology, North Carolina State University, Raleigh, NC, USA
[29]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA
[30]Villum Center for Biodiversity Genomics, Department of Biology, University of Copenhagen, Copenhagen 2100, Denmark
[31]These authors contributed equally
[32]Present address: School of Basic Medical Sciences, Jiangxi Medical College, Nanchang University, Nanchang, China
[33]Lead contact
*Correspondence: liuweiwei@mail.kiz.ac.cn (W.L.), lukas.schrader@uni-muenster.de (L.S.), jjboomsma@bio.ku.dk (J.J.B.), guojiezhang@zju.edu.cn (G.Z.)
https://doi.org/10.1016/j.cell.2025.05.030

not mate and reproduce sexually, even though they rarely lost their ovaries[7] and remain capable of laying unfertilized, male-destined eggs. The ants have adaptively radiated into over 15,000 extant species belonging to 343 genera and 16 subfamilies,[8] facilitated by an extraordinary evolutionary potential for innovative resource acquisition. While termites remained decomposers, social wasps continued as insect predators, and social bees retained dependence on pollen and nectar, ants diversified trophically. Extant ant species range from herbivores to specialized predators.[5] Their colony sizes vary from approximately 10 to more than 10 million individuals. In concert with ergonomic advantages related to winglessness in workers,[9] an unparalleled functional diversity of queen and worker castes evolved[10] and the ensuing morphological and behavioral polymorphism[5] often led to large differences in lifespan among castes.[11]

In the *Origin of Species*, Darwin wrote extensively about ants and bees.[12] His profound knowledge of their natural history and colony-level adaptations allowed him to formulate the first coherent hypothesis for how natural selection had produced what he called "neuter caste" individuals. He argued that family-level selection on the combined fertile parents of a colony must have been crucial. Darwin's insight was elaborated by Weismann[13] and Wheeler,[14–16] who conjectured that ant colonies exhibit superorganismal organization with higher-level ontogenetic development (Table 1) where germline-like queens and soma-like workers develop in complete mutual dependence, analogous to cell lineages in metazoan bodies.[19,20] Much later, genetic marker studies confirmed that the ancestral ant colony had two lifetime monogamous parents,[21] similar to a metazoan body that originates from the fusion of two lifetime-committed parental gametes.[22] This analogy of origin and organization is consistent with an inclusive fitness[23] perspective where full-siblings and offspring are equally effective vectors for passing on genes to future generations. Any slight but lasting advantage gained from raising siblings over offspring could thus achieve an irreversible transition to superorganismal colony life, provided that the parents remained strictly monogamous until that transition was irreversibly established.[24]

The sequencing of ant genomes started around 2010 and allowed a completely new understanding of the mechanisms that have shaped ant social adaptations.[25,26] However, initial comparative analyses[1,27,28] also highlighted that further progress in understanding how natural selection shaped social organization requires reference genomes of high quality and extensive coverage of extant taxonomic diversity.[18,29–33] In 2017, we established the *Global Ant Genomics Alliance* (GAGA) to obtain large comparative datasets of high-quality reference genomes to address fundamental questions about the evolution of social traits.[34] Obtaining and analyzing high-quality genomes across ant lineages indeed allows new insights into the genetic underpinnings of social traits. In particular, we uncovered signatures of selection related to a suite of morphological and functional aspects of queen-worker caste differentiation. We also obtained large-scale comparative understanding of interactions among genes associated with social traits, their gene-regulatory network expression profiles, and their roles during pre-imaginal caste differentiation. We show that high rates of genome rearrangements are correlated with speciation rates across ant subfamilies. Finally, our analyses reveal that conserved microsyntenic regions are often co-expressed and associated with specific caste phenotypes and that correlated social traits experienced parallel patterns of intensified or relaxed selection.

RESULTS

The phylogenomics of ancestry, adaptive transitions, and key social traits

We generated whole-genome sequences for 145 ant species, 125 of them using a combination of PacBio and MGI-seq technologies and an additional 20 species based on stLFR (single-tube long fragment reads), a restriction imposed by insufficient available biomass (Figure S1A; Table S1A). In combination with 18 publicly available ant genomes of sufficient assembly quality (scaffold N50 > 500 kb; accessed in 2021), our dataset includes 163 assemblies covering 12 of the world's 16 ant subfamilies (75%) and 97 of the 343 extant genera (28%) (Figure 1), with a mean scaffold N50 contiguity of 5.8 Mb (Table S1A;

Table 1. Glossary of specific terminology in ant biology

Superorganismality	a level of biological organization that is irreversibly more complex than that of a metazoan body or the societies formed by such bodies; this concept applies to certain social insects, including all ants, honeybees, stingless bees, bumblebees, vespine wasps, and termites with true workers; these lineages represent four convergent major transitions in evolution (MTEs), which are all characterized by germlinized queens, somaticized workers, and canalized caste development
Gyne	the primary reproductive female caste of social insects destined to become a queen after mating
Gamergate	an inseminated worker that becomes a replacement germ line of an ant colony
Ergatoid	a permanently wingless "worker-like" (=ergatoid) gyne
Trophallaxis	the sharing of fluid mouth-to-mouth among adult colony members
Polydomy	a single ant colony occupies multiple, spatially separated nests that stay interconnected
Polygyny (pleometrosis)	colonies with more than one reproducing queen
Polyandry	insemination of a queen by more than a single male
Social parasite	a species whose queens lost the ability to found colonies independently and evolved traits to exploit or take over unrelated heterospecific host colonies
Inquiline	an obligate socially parasitic ant that infiltrates, and remains fully dependent on, a heterospecific host colony; in most cases, they lost their worker caste
Formicoid and poneroid ants	the two major ant clades (Figure 1); the formicoids comprise approximately 90%[8] of the extant ant species worldwide and exhibit the most extreme social elaborations; the poneroids make up 8%–9%

See also Table S1K.

Figure S1A). For 12 species, we used Hi-C libraries to obtain chromosome-level assemblies, doubling the number of existing ant genomes at full chromosome resolution.[18,29,31]

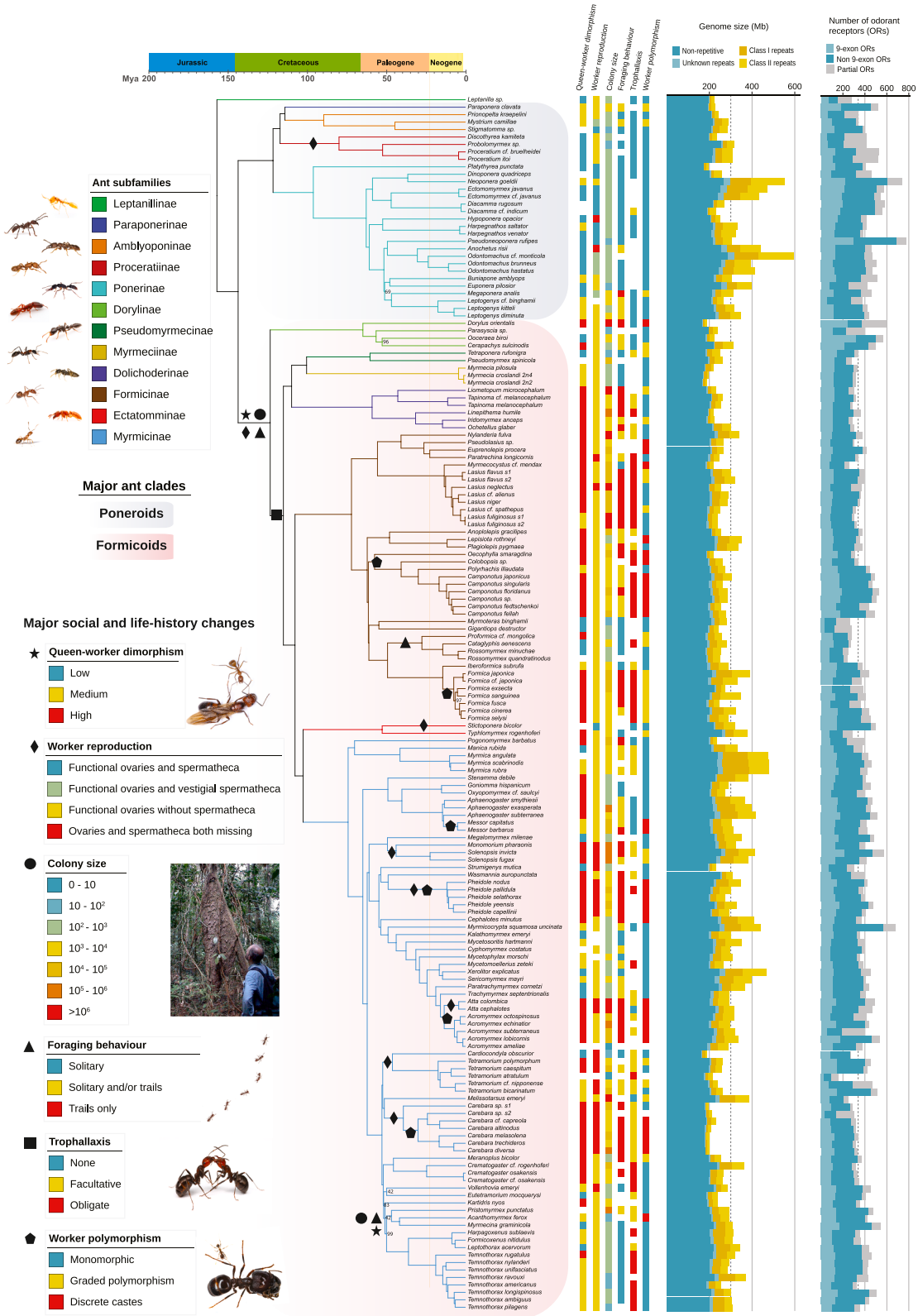
Genome sizes ranged from 188.6 Mb in *Cardiocondyla obscurior* to 594.5 Mb in *Odontomachus cf. monticola*, with an average of 300 Mb (Figure 1), a range typical for Hymenoptera but narrow relative to the variation across insects in general (0.1–21.5 Gb).[35,36] Size differences across ant genomes can be explained mostly by variable repeat content and lineage specific transposable element (TE) expansions or contractions (Figure 1). To minimize methodological biases in downstream comparative analyses, we annotated the protein-coding genes in all genomes using a standardized pipeline combining homology-based, RNA sequencing (RNA-seq), and *de novo* annotations. We also re-annotated specific genes and gene families of particular interest in ant biology (Table S1E). All high-quality genomes were predicted to contain approximately 11,000–15,500 protein-coding genes (average 13,200). Odorant receptors (ORs) were one of the largest gene families across ant genomes, ranging from 687 complete copies in the free-living predatory ant *Pseudoneoponera rufipes* to 90 in the obligately inquiline (Table 1) social parasite *Tetramorium atratulum* (cross-species average 338; Figure 1; Table S1F).

Whole-genome alignment for all 163 genomes using Cactus[37] showed that approximately 120 Mb (40% of the mean genome size) was aligned in at least 75% of the ant species (Figure S1D). We extracted 920 intergenic regions present in at least 161 species (99%) to obtain a concatenated matrix for the main maximum likelihood phylogenomic tree (Figures 1 and S1A).[38] This highly supported tree based on non-coding regions reduces systematic biases known to affect protein-coding sequences and ultraconserved elements (UCEs).[39–41] This tree supported the placement of the Paraponerinae as the sister lineage to the Amblyoponinae (100% bootstrap), in contrast to its previous placement as sister clade to the Proceratiinae (and Agroecomyrmecinae; not covered in our dataset).[1,42] Phylogenomic analyses also revealed some conflicts emanating from analyses of different genome parts, such as protein-coding regions and UCEs, and clarified the placement of some tribes and genera within subfamilies (Figure S1B). We used a fixed topology based on whole-genome alignments to time-calibrate the tree with 38 fossil records (Tables S2A–S2C; Figure S1C) using an independent rates clock inference. This analysis estimated the emergence of the ants to have occurred in the late Jurassic, 157 mya (mean, 95% highest posterior density 137–173 mya) (Figures 1 and S1C). All ant subfamilies covered in our analyses emerged during the Cretaceous, consistent with recent studies.[1,42]

To explore the evolution of social organization across ant lineages, we collected comparative data for up to 26 phenotypic traits, producing a total of 16 categorizations, including colony size, foraging mode and trophallaxis, morphological and functional aspects of caste polymorphism, diet, habitat, and biome (Figure 1; Table S3A).[17] Ancestral state reconstructions showed that major elaborations of already existing social traits emerged in the ancestral lineage from which the formicoid ants originated ca. 125–145 mya, including increased queen-worker dimorphism, larger colony size, trail foraging, and the almost complete loss of a functional spermatheca in workers.[43] We also reconstructed recurrent losses of worker ovaries, gains in worker polymorphism across ant lineages, and secondary reductions in other social traits. Distinctly polymorphic worker castes evolved several times independently but almost exclusively within the formicoid ants, suggesting that this relatively labile trait is conditional on some other traits that evolved in the formicoid ancestor.

Extensive genome rearrangement and conserved modular co-expression of caste-related genes

Ant genomes vary in chromosome number,[44] ranging from $n = 1$ in *Myrmecia croslandi*[45,46] to $n = 60$ in *Dinoponera lucida*,[44]



(legend on next page)

variation that our study covered to a large extent (Figures 2A and S2A). To explore the genomics of karyotype diversity, we compared synteny across 17 species with chromosome-level assemblies, representing the four most species-rich subfamilies (Ponerinae, Dorylinae, Formicinae, and Myrmicinae; Figure S2A). These genomes varied 2.1-fold in size (196–414 Mb) while their chromosome numbers, ranging from 9 to 42, varied 4- to 5-fold. We detected multiple forms of chromosomal rearrangements, including fusions and fissions, large insertions and deletions, and inversions and duplications that were particularly pronounced between subfamilies (Figures 2A and S2A). Further analysis of rearrangement dynamics across 135 high-quality genomes (N50 > 1 Mb) confirmed that high rates of synteny loss at lower taxonomic levels are not restricted to the attine fungus-growing ants[47] but also occur in the species-rich genera *Camponotus*, *Pheidole*, and *Temnothorax* (Figure S2B). We found a remarkable positive correlation ($p < 0.002$) between genome breakpoint rate and species-richness across the entire Formicidae family (Figures 2B and S2C), suggesting that genome rearrangement rates were associated with diversification of ant subfamilies.[48,49] The ant genome-wide rearrangement rates that we studied were similar to those in bees but approximately 1.5–4.5 times higher (significant at $p < 0.001$) than those known from vertebrate clades (Figure S2D).

Despite high rates of genome rearrangement and macrosynteny loss, we detected 970 conserved orthologous microsynteny clusters that remained unchanged in over 80% of the ant genomes (Table S4E). These clusters represent a shared history of 157 million years across an average of 3,696 genes per genome, suggesting that strong purifying selection has maintained these exceptionally high levels of conserved microsynteny, despite low macrosynteny conservation across ants, and that disruption of these gene clusters has significant fitness costs.[50–52] We further determined whether genes within microsynteny blocks were more likely to be co-expressed within the same gene-regulatory network (GRN),[26,53] by performing a weighted co-expression network analysis (WGCNA) using developmental transcriptomes of the pharaoh ant (*Monomorium pharaonis*).[20] This analysis showed that genes located within conserved microsynteny blocks have significantly higher overall co-expression coefficients than other genes (Figure S2E). In particular, we discovered four co-expression modules significantly enriched in highly conserved microsyntenic blocks (Figures 2C and S2F). The largest, module 1, comprised 527 conserved syntenic genes (out of a total of 1,383 genes) enriched for basic cell respiratory and metabolic functions, including lipid metabolism. Module 2 (482 syntenic genes) has overall gyne-biased co-expression, particularly during the pupal and adult stages, and was enriched with metabolic genes related to mRNA biogenesis, peptide biosynthesis, and ribosome biogenesis. Module 3 (170 syntenic genes) is associated with organ morphogenesis and development, while module 4 (53 syntenic genes) has worker-biased co-expression across developmental stages following the second larval instar (Table S4A).

In one of these GRNs, which we named the "worker-biased module 4 network," we identified several key genes associated with the regulation of neuropeptide hormone levels, insulin secretion and transport (e.g., *Syt*, *7b2*, *Sifa*, *Unc-104*, and *Amon*), of which the first three are co-expressed with adjacent genes that exhibit highly conserved microsynteny (Figure 2D; Table S4B). Within module 1, we discovered a conserved microsyntenic GRN involving fatty acid synthase (FASN), stearoyl coenzyme A (CoA) desaturase (SCD) and elongase (ELOVL) genes, with several of these genes showing differential expression (DEG) between gynes and workers during *M. pharaonis* development (Figure 2E). This module also includes two vitellogenin genes encoding major egg yolk protein precursors that have maintained adjacent positions in the genomes of 148 ant species surrounded by orthologs to CG8745, an ethanolamine-phosphate phospho-lyase, and CG18063, an ortholog to human SPATA17 (spermatogenesis associated 17; Figures 2F and S2G). These co-expression signatures within conserved microsyntenic blocks likely represent ancestral GRNs conserved throughout ant evolution because they are instrumental for metazoan metabolism and development (modules 1 and 3) and/or for derived functions that originated during the major ancestral transition to permanent superorganismal caste differentiation (modules 2 and 4).

Adaptive genome evolution, emergence of the formicoid ants, and social information processing

To explore adaptive genome evolution, we investigated gene family dynamics and signatures of positive selection on protein-coding genes across the 163 ant genomes and 8 Apoidea outgroups (Figure 3A). We identified 8,876 single-copy orthogroups and 617 multi-gene families present in more than 80% of the ant species (Table S5J). Gene family expansions were enriched in some recent lineages and in the common ancestor of all ants except Leptanillinae (for which a high-quality genome was unavailable; branch with black dot in Figure 3A), with 65 gene families significantly expanded in this node (Table S5E; Figure S3B). Most of these gene families are associated with chemoperception (22 families), cuticular hydrocarbon biosynthesis (17 families, including 8 cytochrome P450s),

Figure 1. Time-calibrated whole-genome phylogeny of the ants and their main social traits

Maximum likelihood phylogeny (see also Figure S1) inferred from a supermatrix of intergenic regions retrieved from whole-genome alignments and calibrated with 38 dated fossils (Tables S2A–S2C); branch support is 100% unless indicated otherwise. Subfamilies (different branch colors; upper left) are nested within background colors specifying the poneroid and formicoid ants with *Leptanilla* as outgroup. Variation in six social and life-history traits are in the central vertical bars and explained toward the bottom left, with changes inferred from ancestral state reconstructions indicated with black symbols in the tree (see Table S3A and full ancestral state reconstructions).[17] Toward the right, genomes are broken down in non-repetitive sequences, class 1 and class 2 repeats, and unknown repeats (see Table S1A); average genome size is 300 Mb (dashed line). The final column has cross-species representations of the annotated odorant receptor (OR) gene family, known to have been vastly expanded in Hymenoptera and particularly in the ants,[18] specified for complete 9-exon ORs, complete ORs of other subfamilies and partial ORs (mean 338, dashed line). Photo credits in Table S1D.

See also Figure S1.

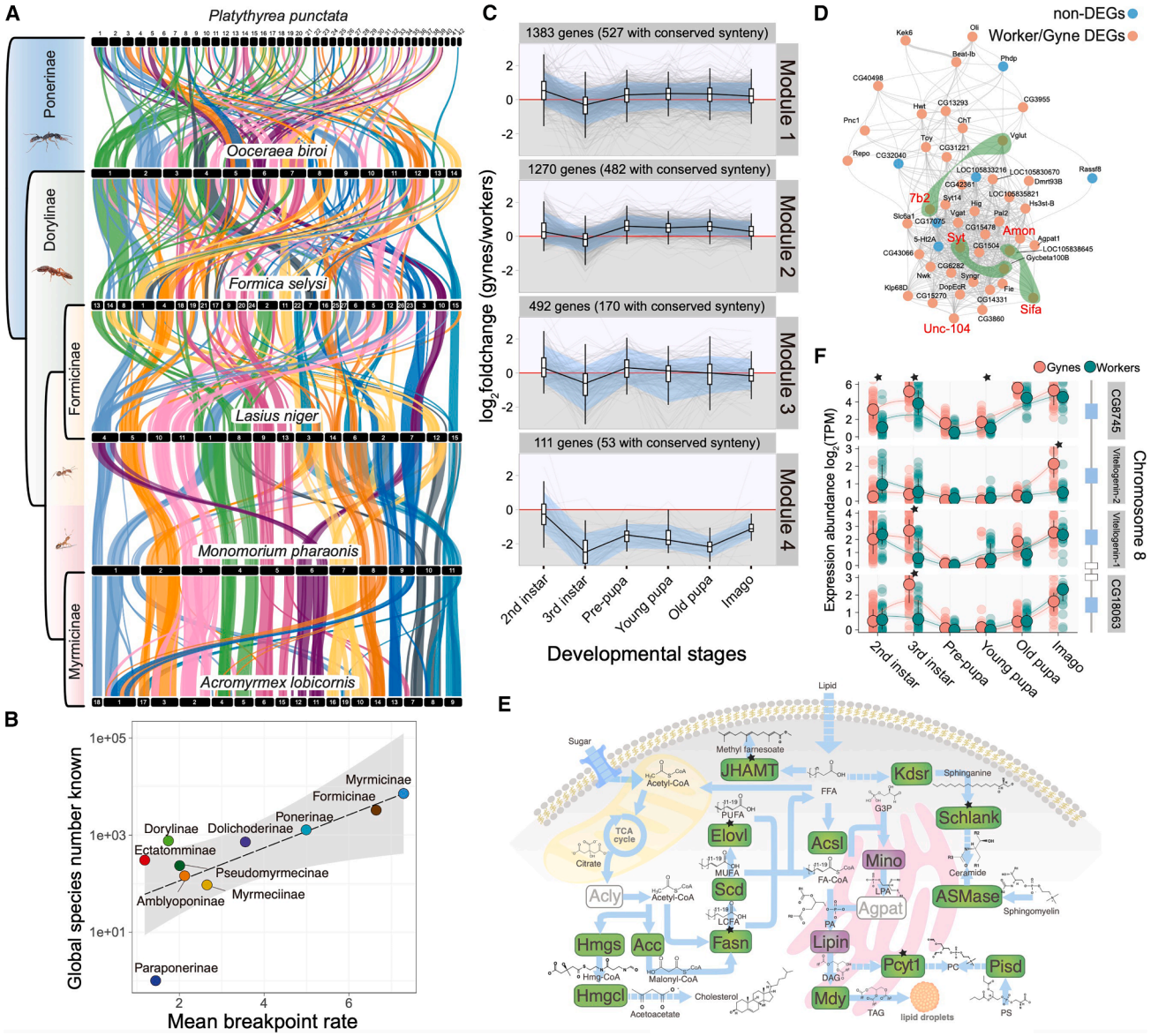


Figure 2. Genome-wide synteny and syntenically conserved co-expression

(A) Ribbon diagram illustrating chromosome-level synteny across six species from four subfamilies (see Figure S2A for all 17 chromosome-level assemblies). Horizontal bars (normalized for genome size) represent chromosomes. Vertical ribbons are syntenic blocks exceeding 500 kb, color-coded to match *M. pharaonis* chromosomes (cladogram simplified from Figure 1).

(B) Global extant species number[8] as function of mean chromosomal breakpoint rate across ten subfamilies (adjusted $r^2 = 0.67$; $p = 0.002$, based on phylogenetic generalized least squares [PGLS]; 95% confidence intervals in gray).

(C) Gene co-expression network analysis using caste developmental transcriptomes from *M. pharaonis*.[20] Red horizontal lines indicate equivalent expression between gynes and workers, and bars (median, interquartile range, and 1.5× interquartile range) represent observed expression levels across developmental stages (blue ribbons connect 90% data ranges). Four modules are enriched for genes with highly conserved synteny across more than 80% of the species (hypergeometric test, false discovery rate [FDR] < 0.05; Figure S2F).

(D) The module 4 gene co-expression network, with gray lines connecting genes with highly correlated expression adjusted for false discovery rates (FDR < 0.01). Nearly all genes are overexpressed in the *M. pharaonis* worker caste. Genes regulating insulin secretion and neuropeptide hormones are highlighted with red acronyms; green connecting shades represent syntenic co-expressed genes (Table S4B).

(E) Schematic representation of conserved syntenic genes involved in lipid biosynthesis, producing fatty acids, diacylglycerol, triacylglycerol, sphingolipids, phospholipids, and cholesterol. Convertases co-expressed in module 1 are highlighted in green and those in other modules in purple; genes with white

(legend continued on next page)

endocrine signaling (6 families), digestion (5 families), and sting venoms (2 families) (Figure 3B; Table S5E). This indicates that the ancestral ants had already realized key molecular innovations for these physiological and sensory functions, consistent with the fossil record.[55] Notably, the number of ORs underwent a major ancestral increase and continued to change dynamically during ant diversification (see also Figure 1). These changes were reflected in elaborations and secondary reductions of glomeruli number in the antennal lobes (Figure 3C), consistent with earlier studies based on smaller samples of ant species.[54]

After reconciliation with the species tree, selection assessments across single-copy genes revealed hotspots of positive selection in the ancestor of the highly diverse formicoid ants that make up approximately 90% of all extant ants, consistent with Romiguier et al.[1] We also detected enriched positive selection in the species-poor Pseudomyrmecinae, which are specialized large-eyed arboreal ants (Figures 3A and S3A),[56–58] illustrating that ancestral adaptive evolution does not necessarily induce diversification. This signature was not recovered by Romiguier et al.,[1] who analyzed conserved genes from fragmented genomes. Positive selection primarily acting on more dynamically evolving genes in the Pseudomyrmecinae is consistent with 69 of the 202 genes (34%) under positive selection lacking orthologs in *Drosophila melanogaster*.

Genes under positive selection in the formicoid ancestor were functionally enriched for many general aspects of multicellular life including development, longevity, autophagy and immunity, transcription, and chromatin organization (Figures 3D and S3C; Table S5B). Consistent with these positively selected genes having driven adaptive changes associated with caste differentiation, we found them to be significantly enriched for differential gyne-worker expression in *M. pharaonis* (33%; 76 out of 231 genes; $p = 0.0005$; Fisher's exact test). In addition, we found stronger overall caste-specific expression bias in formicoid ants compared with poneroids (Figure 3E), consistent with the more pronounced phenotypic queen/worker dimorphism in formicoid ants (Figure 1). These genes with extreme queen/worker bias in formicoids (highlighted in Figure 3E) are associated with several caste-differentiated functions, such as reproduction (*ovarian tumor* [*otu*]) ecdysone signaling (*rig*), lipid metabolism (*PLCXD1*), and brain development (*mud* and *Rab3*).

Deeply conserved signaling pathways shaped the evolution of reproductive division of labor

Ancestrally, ant colonies have been inferred to contain a single monogamous queen[21] who dispersed during a mating flight to found a colony consisting of a small number of workers specializing in non-reproductive colony functions. Over time, the ants diversified through adaptive radiation into more than 15,000 species, sometimes leading to extreme queen-worker dimorphism, polygyny (multiple queens per colony), polyandry (gyne insemination by multiple males), and secondary reductions such as ergatoid (wingless) queens or reproduction by inseminated (gamergate) or parthenogenetic workers (Figure 1; Table 1).[59] These later elaborations, which often occurred independently across ant genera, may have either deepened developmental caste-canalization processes or relaxed their functional constraints. We therefore assessed signatures of convergent positive selection on genes underpinning social traits while also exploring whether selective constraints (maintained by either positive or purifying selection) became intensified or relaxed in response to phenotypic changes in colony organization. We found that gene-level selection regimes associated with queen-caste-related traits were often enriched for deeply conserved signaling and metabolic pathways, including juvenile hormone (JH) biosynthesis and the mitogen-activated protein kinase (MAPK) signaling cascade (Figure 4), among other functions (Tables S6H–S6K).

JHs are sesquiterpenoid hormones primarily synthesized and secreted by the neuroendocrine corpora allata. Particularly, JH3 is a key regulator of growth and metamorphosis in arthropods and has a conserved role in mediating caste differentiation in ants and other social insects.[60–68] We detected intensified selection in lineages with high queen-worker dimorphism for multiple paralogs of juvenile hormone acid methyltransferases (JHAMTs) and juvenile hormone esterases (JHEs). There were also consistent signatures of positive and intensified selection associated with secondary transitions to reduced queen functionality (usually lower fertility) for aldehyde dehydrogenase (ALDH) across polygynous species and for Cyp15a1/Cyp18a1 across lineages with ergatoid queens (Figure 4A). Consistent with JHAMTs shaping caste phenotypes during immature development, queen-worker dimorphism was positively correlated with JHAMT gene family size, explaining 21% of the variation (Figure 4B). JHAMT gene copy number was also higher in species with monogynous colonies, trail foraging and the absence of a spermatheca in workers, suggesting JHAMTs were co-opted for regulating multiple social and life-history traits (Table S6D).

The MAPK cascade is a deeply conserved pathway underlying eukaryote cell differentiation and developmental patterning. Signals from transmembrane receptors converge into multiple transcriptional responses,[69,70] integrating cell surface signals including G-protein-coupled receptors and receptor tyrosine kinases (Figure 4C).[71] Components of this cascade regulate insect reproduction and longevity[72,73] and MAPK activation downstream of insulin signaling induces ovary maturation in *Harpegnathos* (Ponerinae) when a worker transitions into a gamergate.[74] Across our dataset, evolutionary transitions in queen phenotypes were correlated with selection signatures across several core proteins of the MAPK pathway and its associated receptors. Proteins involved in insulin signaling underwent intensified (InR1 and Chico) or positive (InR2) selection across lineages with high queen-worker dimorphism and relaxed selection in

background are outside syntenically conserved blocks. Solid blue arrows represent direct effects and dashed arrows indirect effects (see also Table S4C). Nonnuclear endoplasmic reticulum and mitochondria are pink and yellow, respectively.

(F) A microsyntenic block from module 1 conserved in 148 ant species, consisting of two vitellogenins, CG8745 and CG18063 (blue rectangles), and their correlated gyne-worker differential expression throughout *M. pharaonis* development (Wald test, asterisks: $p < 0.05$; see also Figure S2G).[20]

See also Figure S2.

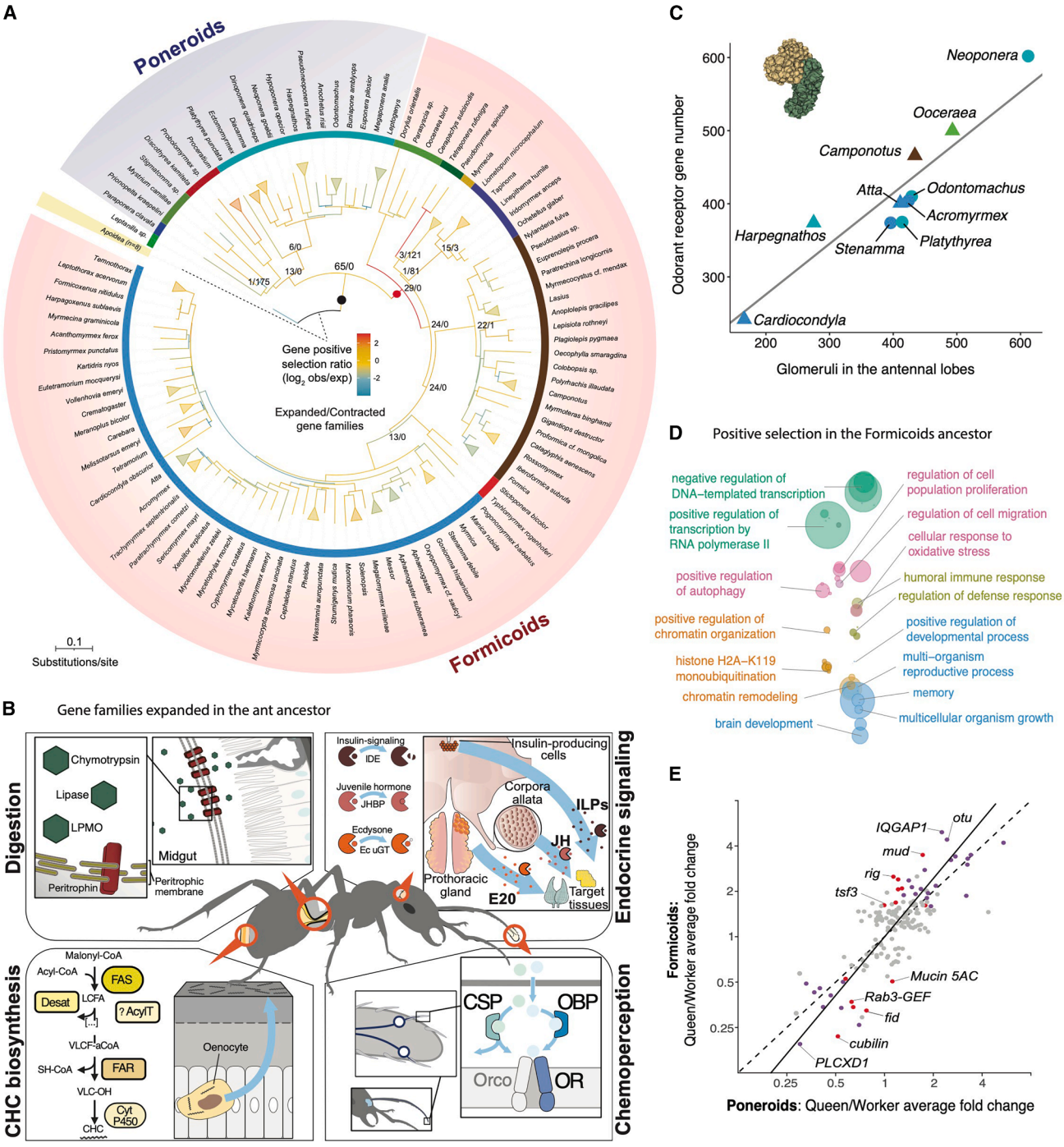


Figure 3. Positive selection and gene family evolution in the ancestor of extant ants and formicoid ants

(A) Maximum likelihood tree showing expanded/contracted gene families and gene-level positive selection ratios (observed/expected; branch-specific heatmap colors in the center; see also Figure S3). The outer circle specifies ant subfamilies (cf. Figure 1) and the Apoidea outgroup. The central black dot represents the ancestor of all extant ants beyond the *Leptanilla* branch while the red dot refers to the common formicoid ancestor.

(B) Gene families expanded in the ancestor of extant ants beyond *Leptanilla* could be assigned to four major functions: digestion, endocrine signaling, cuticular hydrocarbon synthesis and chemoperception (see Table S5G for gene family acronyms).

(C) Correlation between antennal lobe glomeruli number and odorant receptor genes (R^2 = 0.86; $p < 0.001$) across representatives from 10 genera in four subfamilies (colors as in A) fitted with a PGLS correlation axis (triangles refer to earlier studies). The top left shows the glomerular structure of a *Stenamma debile* worker antennal lobe (green: T6 cluster; yellow: rest of the antennal lobe).[54]

(legend continued on next page)

species with ergatoid queens or no queens (obligate gamergates or parthenogenetic) (Figure 4C). Two putative JH transmembrane receptors (Cad96Ca and FGFR1) experienced relaxed selection in species with high queen-worker dimorphism but Gce, the sole intracellular JH receptor in ants, has been positively selected, suggesting modifications in insulin and JH signal transduction in species with high queen-worker dimorphism. Several MAPK core proteins (Sos, spri, Raf, Dsor1, Cnk, and Ksr) showed convergent signatures of intensified selection related to secondary evolution of ergatoid and queenless reproduction, and relaxed selection associated with increasing queen-worker dimorphism and higher degrees of polygyny. Finally, several MAPK and MAPK-associated genes showed significant overexpression in worker larvae, consistent with this pathway's importance for multiple aspects of caste differentiation (Figure 4C; Table S6E).

We used *M. pharaonis* to validate the inferred significance of the MAPK pathway by pharmacological treatment using trametinib, which can inhibit the phosphorylation of the core MAPK cascade component Rolled by Dsor1, increasing lifespan in *Drosophila*.[73] Both Rolled and Dsor1 have worker-biased expression during *M. pharaonis* development (Figure 4C; Table S6E). This treatment induced significantly increased body size in workers (Figure 4D), consistent with reduced caste dimorphism when MAPK signaling is suppressed. However, the inferred signatures of intensified selection in MAPK genes in species with low queen-worker dimorphism suggest that this pathway is most important in lineages where caste developmental canalization is not yet deeply entrenched in the early developmental stages.[74]

Convergent selection on social traits and elaborations of the worker caste

Our taxon sampling covered not only variation in social organization related to queen traits as addressed above and in Figure 4 but also in worker traits related to colony size, foraging mode, and the reproductive ability and polymorphic differentiation of this caste. In addition, we evaluated the possible effects of behavioral traits (e.g., trophallaxis), ecological habits (e.g., trophobiotic tending of aphids and/or coccids), and biogeographical variables (e.g., distribution range and climate). Moreover, our dataset included several species with parasitic lifestyles, including temporary social parasites, dulotic (slave-making) species and permanent inquilines, where the worker caste has been reduced or lost.[75] This allowed us to track convergent signatures of selection associated with these trait syndromes as they originated independently across ant subfamilies and genera. We uncovered few instances of convergent positive selection on genes associated with these traits (Figure 4; Tables S6A and S6B). However, gene-level signatures of intensification or relaxation of selection were commonly observed when organizational complexity increased relative to inferred ancestral traits such as monogyny, monoandry, small colony size, low queen-worker dimorphism, a single worker caste, no trail foraging, no trophallaxis, no trophobiosis, functional ovaries in unmated workers, and independent colony founding by a single queen (Figure 5A; Table S6A).

Overall, more than twice as many genes were under relaxed selection in species with large colonies relative to intensified selection. Enrichment analyses revealed intensified selection for genes involved in somatic homeostasis and spermatogenic or meiotic functions, consistent with queen fertility being under stronger selection in large colonies (Table S6L). Conversely, relaxed selection often occurred in metabolic and signaling pathway genes, and in genes associated with responses to external stimuli (Table S6L), as expected if large colonies are better buffered against environmental disturbances.[78] Relaxed selection signatures were also predominant for traits such as polyandry, worker polymorphism, queen-worker dimorphism, trophobiosis, trophallaxis and, particularly, social parasitism. However, non-parasitic reductions in individual queen functionality (fertility), such as polygyny (multiple queens per colony reduces demands on individual queen fecundity), reproduction by inseminated (gamergate) workers, or loss of worker ovaries, induced higher prevalences of intensified selection (Figure 5A). These patterns are intriguing, but their more detailed interpretation remains challenging due to the strong correlations between many of these traits (see next section), and the possible effects of demographic changes associated with each of these social elaborations and reductions.

Ant worker castes can be uniform, continuously polymorphic, or discretely polymorphic, the latter being mostly restricted to the crown-group formicoid ants (Figure 1) where it can lead to within-worker-caste differentiation with specialized behavior and allometric morphology.[79] Consistent with differential gene expression underlying worker caste polymorphism,[20,80] intensified selection was associated with enrichment of genes involved in chromatin and transcriptional regulation across species with discretely polymorphic workers (Figure 5B; Table S6M). In addition, 57 of the 192 genes that experienced intensified selection (30%) showed differential expression between large and small workers in at least one developmental stage of *Acromyrmex echinatior* (Table S6F).[20] Figures 5B and S4A highlight sets of genes that appear to be differentially involved in worker caste polymorphism between the two largest (Formicinae and Myrmicinae) subfamilies of formicoid ants. For the Myrmicinae, genes associated with worker polymorphism are involved in JH regulation (*takeout* and *Jhe*) and brain development (*Awh* and *Sina*),

(D) Functional GO-term enrichment for the 231 positively selected genes in the formicoid ancestor (see also Figure S3C and Table S5B). Circle size represents gene count per GO term and its associated functional terms for development (blue), autophagy and longevity (pink), immunity (yellow), gene expression (green), and chromatin organization (orange).

(E) Queen-worker expression bias in poneroids (*x* axis, $n = 10$ species) and formicoids (*y* axis, $n = 60$ species) for positively selected genes in the formicoid ancestor (red circle in Figure 3A): formicoid-specific DEGs in red, shared formicoid and poneroid DEGs in purple and non-DEGs in gray; the correlation axis ($R^2 = 0.89$) is steeper than the (dashed) equal expression diagonal ($p = 0.0001$); acronyms highlight ten genes with extreme positions and with higher caste bias in formicoids (see also text and Table S5A).

See also Figure S3.

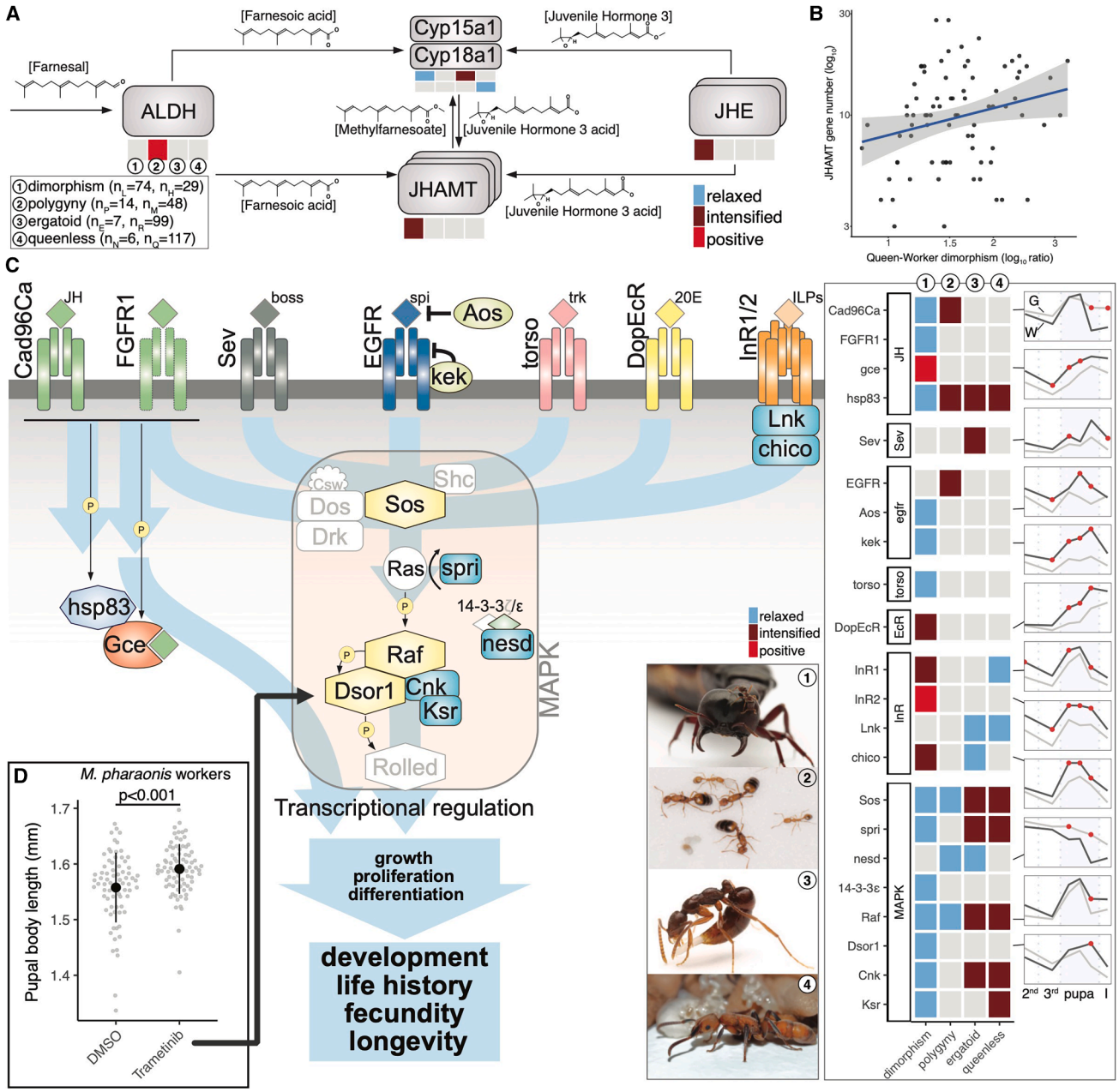


Figure 4. Genomic signatures of phenotypic caste elaboration in queen (gyne) ants

Increasing queen-worker dimorphism and reductive secondary shifts to polygyny, ergatoid queens, or reproduction via gamergate or parthenogenetic workers (①–④ in A and C) are associated with convergent evolutionary changes in key metabolic and signaling pathways.

(A) Selection signatures associated with queen caste elaborations in the insect hormone biosynthesis pathway mediating JH production.

(B) JHAMT gene family size as function of queen-worker dimorphism across ant species (PGLS $p = 0.005$; Weber's length queen-worker dimorphism ratios available for 80 species with high-quality genome assemblies; values along the axis are original before $\log_{10}$ transformation; Table S6D).

(C) Core proteins of the mitogen-activated protein kinase (MAPK) pathway and its associated transmembrane receptor tyrosine kinases evolved convergently in association with elaborations of queen caste functionality, involving 22 key genes (right-hand heatmap). The receptors include four hormone receptor homologs: intracellular JH receptor Gce, putative transmembrane JH-receptors Cad96Ca and FGFR1 (closest homolog to *heartless*; no direct ortholog known) and ecdysone (20E) receptor DopEcR; two insulin-like peptide receptor homologs (InR1 and InR2), and three transmembrane growth factor receptors: Sevenless (Sev), epidermal growth factor receptor (EGFR), and torso. Diamond symbols represent ligand types, and yellow P-circles indicate phosphorylation. Proteins with white background (e.g., Csw, Dos, and Drk) remained unaffected by changes in any of the four elaborations or reductions of queen caste functionality. The 12 plots toward the right show expression patterns across developmental transcriptomes for gyne-destined *M. pharaonis* individuals (gray lines, G) and worker-

(legend continued on next page)

whereas worker caste-biased expression in both subfamilies involves *takeout* and genes associated with brain development and function (*gcm*, *mAChR-A*, *GPCR*). The *mAChR-A* gene is particularly interesting as it is consistently overexpressed in small workers of *A. echinatior* (Figure 5B) and across the small workers of nine other ant species from both subfamilies (Table S6F). This gene is specifically expressed in the head of *Carebara diversa* workers but not in soldiers (Figure 5B) and is also differentially expressed between foragers and nurses of *Camponotus floridanus*.[81] *mAChR-A* underwent intensified selection in lineages with polymorphic workers, large colony size and trail foraging, but relaxed selection in species with gamergates where colonies effectively consist of only a single worker caste (Table S6F). Also, the *Gcm* gene underwent intensified selection in lineages with polymorphic workers and large colonies (Table S6F). *Gcm* encodes a transcription factor essential for brain gliogenesis[82] and is overexpressed in small workers across 8 ant species, including the genera *Pheidole* and *Carebara*, known to have extreme polymorphism between small workers and soldiers (Figure 5B; Table S6F). This gene was predominantly localized at the neuropil surface of *C. diversa* with higher expression in worker brains relative to the much larger soldiers (Figure 5B), suggesting it might regulate differentiated glia functions for behavioral specialization across unrelated species with a distinct subcaste of small workers.

The loss of reproductive potential in workers is most pronounced in the formicoid ants where a functional spermatheca is generally absent (*Stictoponera bicolor* is one of the few exceptions[83]) although workers can usually still lay unfertilized eggs that may develop into males. Complete sterility via the loss of worker ovaries is much rarer but is known to have evolved convergently several times, particularly in the subfamily Myrmicinae (Figure 1). When worker ovaries are maintained, molecular mechanisms to suppress worker ovary development can be expected when colonies have one or more inseminated queens.[84] However, the convergent loss of worker ovaries should have made these suppression mechanisms obsolete, consistent with relaxed selection being enriched in genes related to ovary development and oocyte meiosis, among other functions (Table S6N). In line with this reasoning, these genes under relaxed selection showed a more pronounced queen-worker bias in species with sterile workers relative to species with fertile workers (Fisher's exact test; $p < 0.00017$; Table S6G). Six of them are primarily expressed in the germ cells of *M. pharaonis* gynes (Figure 5C; steeper slopes in four out of six cases produced significant caste × sterility interaction terms in two-way ANOVAs, with gyne/worker caste and worker sterility/fertility as factors). These same genes exhibit variable expression across the oogenesis stages, from early germarium until the late vitellogenic oocyte, suggesting that they regulate distinct aspects of oogenesis (bottom ovary drawing in Figures 5C and S4B). For example, *otu* regulates germ cell proliferation and differentiation during fruit fly oogenesis.[85] *Otu* is expressed in the oocytes of both gynes and workers of the harvester ant *Messor barbarus*, but with lower expression in workers (Figure S4C), consistent with their notably reduced but functional ovaries.[86] Remarkably, *otu* evolved under positive selection in the common formicoid and poneroid ancestor but queen-worker expression biases are generally stronger in the formicoids than in the poneroids (Figure 3E), consistent with an adaptive role in modulating caste differences in fertility, which are more pronounced in the formicoids.

Some forms of social parasitism should select for secondary reductions in worker traits that are crucial for free-living ants. The excess of signatures of relaxed selection in social parasites (Figure 5A) was therefore not surprising because small effective population sizes should make these invariably rare species more liable to genetic drift, reducing the strength of positive and intensified selection.[87] Social parasites also showed convergent signatures of genome-wide erosion reflected in higher chromosomal breakpoint rates and considerable gene family losses (Figure 5D). Particularly, we found significantly higher genomic rearrangement rates in inquilines and temporary social parasites, but not in slave-making ants (whose workers carry out more of the free-living ant tasks). In addition, the ancestrally expanded chemoreceptor gene families (Figure 3A), instrumental in coordinated foraging and social communication of free-living ants, became contracted in all investigated socially parasitic species, as reported previously for inquiline and slave-making ants.[87,88] These losses were most extreme in the morphologically highly modified workerless inquiline *Tetramorium atratulum*,[89] which retained only 90 complete ORs, 27% of the 338 average ORs in ants.

Genomic basis of trait interactions: Toward an integrative view of ant social evolution

The organizational complexity of ant colonies emanates from interactions among various phenotypic and life-history traits,[90] but our analyses of trait correlations showed that these traits are unequally connected. Increases in colony size and queen-worker dimorphism (Figure 6A; Table S3D) were significantly correlated with one another and with most other traits estimated with sufficient sample size (13 and 11 links, respectively). Colony size as causal driver was suggested many years ago[91,92] and recently confirmed in a comparative study by Bell-Roberts et al.,[93] whereas higher queen-worker dimorphism was shown to correlate with colony size, worker polymorphism, and loss of worker reproductive potential by Matte and LeBoeuf.[94] Our genomic results are thus consistent with an explanatory model in which increases in colony size and queen-worker dimorphism set the stage for subsequent evolution of other social traits, including continuous worker polymorphism, polygyny, worker sterility, polyandry, trophallaxis, and discrete worker castes. This social trait syndrome also appears to have facilitated the evolution of novel ecology-related traits such as polydomy, trophobiosis, and trail foraging, which were all significantly correlated with

destined individuals (black lines, W) for a selection of these genes. Developmental stages with significant upregulation are indicated with red dots (full plots in the supplemental information[17]).

(D) Injection of trametinib, a specific inhibitor of Dsor1 (black arrow), which blocks MAPK signaling, increased body length (means ± SD across replicate experiments) of *M. pharaonis* workers compared with DMSO controls (two-tailed unpaired t test, $p < 0.001$).

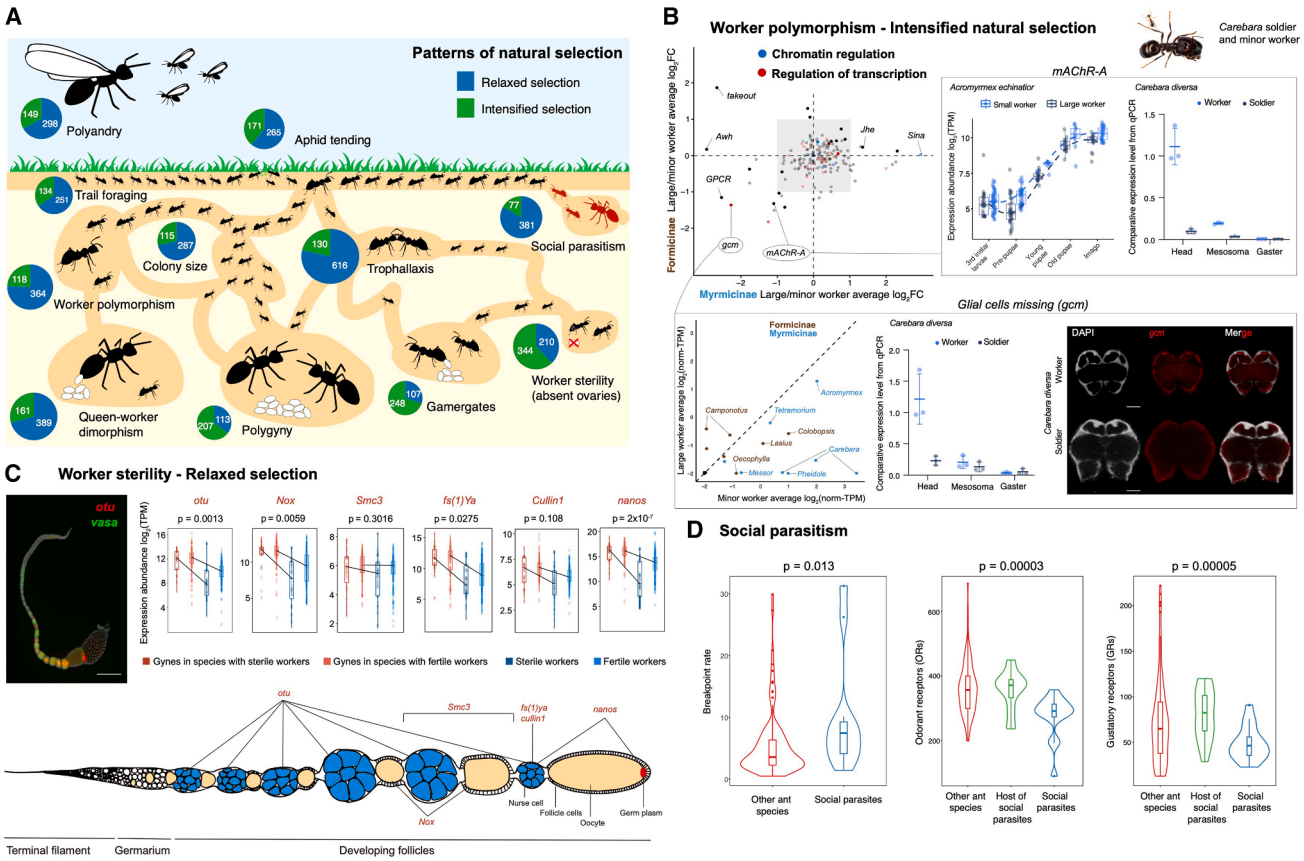


Figure 5. Convergent signatures of selection related to secondary elaborations or losses of worker caste functions

(A) Eleven social traits that evolved repeatedly while leaving consistent signatures of intensified (green) and relaxed (blue) selection (gene numbers in pies).

(B) Genes under intensified selection across species that secondarily evolved polymorphic workers, showing fold change (FC) in adult transcriptomes of large workers/soldiers versus minor workers in the Formicinae and Myrmicinae subfamilies (see also Figure S4A). Higher color-intensity dots represent genes with differential expression. Genes mediating chromatin regulation (blue) and transcription regulation (red) were often enriched (Table S6M). Two genes (ellipses) with consistent minor-worker-biased expression in both ant subfamilies are highlighted in adjacent panels: glial cells missing (*gcm*; bottom) and muscarinic acetylcholine receptor (*mAChR-A*; rightward); see main text for details. The *gcm* right-plot shows an *in situ* hybridization of *Carebara diversa* worker caste brains (scale bar, 100 μm).

(C) Genes that underwent relaxed selection across ant species with fully sterile workers are enriched in sexual reproduction (Table S6N), illustrated here by *otu* gene expression in the ovariole of *M. pharaonis* gynes from a hybridization chain reaction (scale bar, 200 μm; *vasa* is stained to show germ cells) and five additional genes in the bottom drawing of an ovariole (see also Figure S4B). Expression of these genes was generally more gyne-biased in species with fully sterile workers compared with species where workers retained functional ovaries (based on evaluating the caste × sterility interaction term in two-way ANOVAs).

(D) Forms of social parasitism, including inquilinism, dulosis (slave raiding), and temporary social parasitism (Table 1), show higher chromosomal breakpoint rates and contractions of OR and GR gene families (PGLS *p* values above each plot). As inquiline social parasites normally evolve sympatrically from within their hosts (Emery's rule[76,77]), we plotted OR and GR numbers relative to both closely related congener species and to free-living non-social parasites in general.

See also Figure S4.

colony size and/or queen-worker dimorphism. Notably, all these correlations with colony size and queen-worker dimorphism were positive, reinforcing the general consensus of which social traits are evolutionarily derived or ancestral (outer circle in Figure 6A).

In contrast, social parasitism was negatively correlated with both colony size and queen-worker dimorphism, consistent with these behaviors representing secondary reductions of organizational complexity,[95] similar to reproduction via ergatoid queens implying (by definition) a reduction in queen-worker dimorphism. Additional negative correlations were found between discrete caste polymorphism and polygyny, and between polydomy and polyandry. Some derived traits such as ergatoid queens, gamergates and parthenogenesis were poorly connected to other traits (Figure 6A), likely due to low statistical power for identifying correlations between the few species possessing these traits, and to these secondary reductions not having enabled much further diversification.

Our final question was whether and to what extent overarching aspects of progressive and reductive social evolution left consistent signatures of intensified or relaxed selection on specific combinations of genes. Combining genome-wide signatures of selection with variation in social and life-history traits in a phylogenetic framework allowed us to explore the potential of

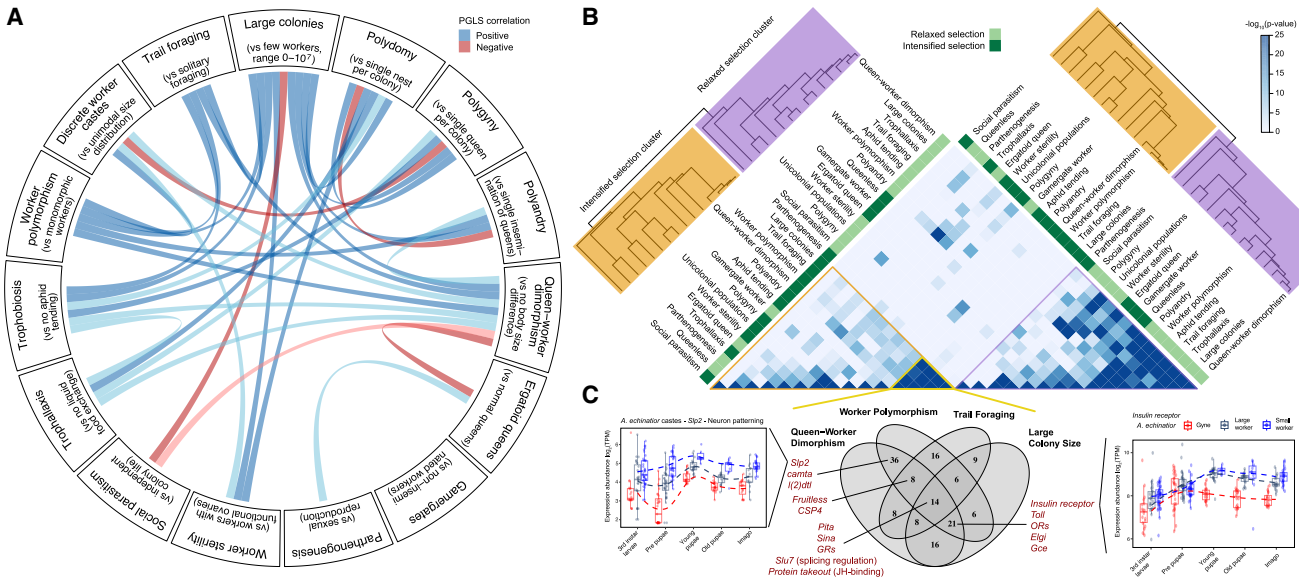


Figure 6. Evolutionarily derived elaborations of social colony life and associated genome-wide signatures of combined intensified and relaxed selection

(A) Circos plot representing 15 key traits of ants and their positive (blue) or negative (red) correlations based on PGLS adjusted for phylogenetic confounding (see Table S3D for full data matrix). Directionality of trait changes is specified by text in brackets representing the putative ancestral states. Connecting ribbons in lighter blue and red represent cases where tests were not always significant across alternative categorizations.

(B) Heatmap showing the negative logarithms of Bonferroni-corrected *p* values of hypergeometric tests, comparing the number of shared genes across pairs of phenotypic traits with joint signatures of intensified or relaxed selection. Using Pearson correlation coefficients as distance metric (shades of blue; the $\log_{10}$ *p* value legend was capped at 25) and average linkage between pairwise comparisons as clustering criterion, the (identical) left and right dendrograms show the hierarchical clustering of co-evolving traits, primarily as triangles adjacent to the baseline, in orange for intensified selection and in purple for relaxed selection.

(C) Venn diagram showing the distribution of the 148 shared genes under intensified selection across four phenotypic traits (yellow triangle; Table S7A). Caste-biased expression is illustrated with concrete *Acromyrmex echinatior* data[20] showing differential expression across developmental stages for *InR* (*insulin receptor*, right) and *Slp2* (left) genes. See main text for details.

combined selection having driven key social innovations in ants. We uncovered overlapping signatures of selection for genes clustered in two major groups (Figure 6B): one representing combined signatures of intensified selection (orange dendrogram and heatmap triangle) and another representing combined signatures of relaxed selection (purple dendrogram and heatmap triangle). The orange triangle highlights correlated instances of intensified selection associated with increasing levels of colonial organization (dark green bars). Within this triangle, overlapping gene sets mediating aspects of lower colonial organization, such as reproduction via ergatoid queens, gamergate workers, or queenless and parthenogenetic workers were subject to relaxed selection (light green bars).

Conversely, the purple heatmap triangle captures gene sets under relaxed selection (light green bars) when organizational complexity increased, while showing intensified selection (dark green bars) during the evolution of reduced organizational complexity in the form of ergatoid queens, gamergate workers and queenless reproduction. A particularly clear combination of correlated relaxed selection signatures on highly overlapping gene sets occurred across seven traits: polyandry, worker polymorphism, trophobiosis, trail foraging, trophallaxis, large colonies, and high queen-worker dimorphism (heatmap bottom right in Figure 6B). The overall results indicate that secondary reductions in social complexity often imposed intensified selection on the same genes that were under relaxed selection in social lineages where organizational complexity increased over evolutionary time. This suggests that sets of genes under intensified selection in species with ancestral traits (i.e., colonies with few monomorphic workers, low caste dimorphism, a single monandrous queen, solitary foraging, and no trophallaxis) were no longer favored in evolutionarily derived lineages with more elaborate social organization. Meanwhile, rare secondary changes toward lower organizational complexity tended to maintain some of these deeply ancestral genes via intensified selection (see lists of genes in Tables S6B and S7A–S7H).

One intensified selection cluster includes a highly correlated node (yellow triangle) containing 148 genes (numbers in Venn diagram) encompassing high queen-worker dimorphism, polymorphic workers, large colony sizes, and trail foraging, which corresponds reasonably well with the phenotypic trait analyses (Figure 6A) and with the colony-size-driven correlates of organizational complexity identified by Bell-Roberts et al.[93] and the causal effects of queen-worker dimorphism retrieved by Matte and LeBoeuf[94] (Figure 6C; Table S7A). More than half (79) of these genes have been under selection associated with queen-worker dimorphism and worker polymorphism, and 28 of them showed caste-biased expression between gynes and workers or between large- and small-worker castes (Table S7A, e.g., *Slp2* in Figure 6C). These 79 genes have

important caste-specific functions such as chemoreception (ORs, gustatory receptors [GRs], and chemosensory proteins [CSPs]) and gene expression regulation (chromatin structure, transcription factors, or protein ubiquitination: *Pita*, *Slp2*, and *Sina*), as illustrated by the concrete example of caste-biased expression of the *Slp2* gene throughout development in *A. echinatior* (Figure 6C). Similarly, genes relevant for elaborated queen phenotypes (Figure 4), such as JH-associated *gce* and *takeout* or the insulin receptor *InR* (which also shows caste-biased expression during ontogenetic development in *A. echinatior*; Figure 6C) appear to be implicated in worker polymorphism and also evolved under constrained intensified selection in large colony species.

DISCUSSION

The analyses presented here offer deeper insights into the genetic underpinnings of ant social traits, from the level of individual genes and gene families to entire pathways, microsynteny blocks, and GRNs. We also provide better understanding of the conserved genetic signatures related to increasing colony size and caste differentiation. These were particularly pronounced in the formicoid ancestor where we inferred that changes in these traits preceded the evolution of multiple further social elaborations, consistent with this node having been a hotspot of positive selection as discovered in a previous study of 65 short-read ant genomes,[1] results that we corroborate and extend here. We provide evidence for positive selection by assessing all orthologous genes instead of only conserved orthogroups and offer many detailed inferences that are only feasible with more complete and contiguous genomes across a higher species diversity and a wider spectrum of social traits. This enhanced resolution allowed analyzing key aspects of genome evolution (Figure 2), ancestral gene family evolution (Figure 3), and signatures of selection related to both elaborative and reductive social traits (Figures 4 and 5), along with their gene-level correlations in co-opted modules (Figure 6). Depending on the specific analyses, 10%–50% of the focal genes exhibited caste-biased expression, underlining the profound impact of transitioning to superorganismal organization with fully complementary caste phenotypes when the ants evolved. The results presented here represent only a fraction of the large genomic datasets that we generated (see supplemental information), which we expect will continue to provide a valuable resource for future studies of the evolutionary and ecological diversification of ants.

Similar to a recent analysis of phenotypic traits across 794 ant species,[93] our trait correlations indicate that colony size was a key driver in ant evolution and the same applies to queen-worker dimorphism, a trait not considered by Bell-Roberts et al.[93] but addressed by Matte and LeBoeuf.[94] These phenotypic correlations were recurrent across the genomic signatures of intensified and relaxed selection that affected interacting social traits, some of which were mediated by conserved drivers such as JH, insulin or MAPK signaling pathways (Figure 4). Rather many changes in gene modules that responded to selection in partly predictable ways (Figures 2C–2F and 6B, and 6C) were apparently shaped by opposing pressures of relaxed and intensified selection for clusters of elaborative or reductive social traits. In fact, the results of Figures 6B and 6C suggest that bundles of phenotypic social traits may ultimately relate to clusters of co-expressed metabolic and developmental pathways embedded in the GRNs that drive social evolution.

The evolutionary history of the ants represents an adaptive radiation potential unrivaled by other lineages of social animals. During the irreversible major transition toward superorganismal organization (Table 1), the ants went through a comprehensive genomic rewiring that enabled the evolution of unusually robust GRNs for caste differentiation.[19,20] It remains to be determined whether this deeply conserved genomic robustness is unique to ants or whether similar developmental mechanisms evolved convergently in the evolutionarily younger superorganismal clades of corbiculate bees, vespine wasps, and higher termites, as expected from Wheeler's superorganism concept.[14,15,24] In particular, it will be interesting to see future work in bees, wasps, and termites addressing whether (1) their major transitions toward superorganismal organization have induced similar correlations between chromosomal rearrangements and species diversification (comparative data suggest this might also apply in bees,[96] Figure S2D), (2) their adaptive radiations were shaped by independent but similarly adaptive co-expressed genes with conserved microsynteny, and (3) similar patterns of positive selection and gene family expansion also occurred in their early ancestors to be followed by mostly intensified and relaxed selection associated with further social diversification.

In conclusion, our study illustrates the significance of genomic rearrangement at different scales and their possible implications for caste development and speciation rates. Our high-quality comparative datasets proved invaluable for uncovering the role of positive selection and gene family evolution in the early adaptive radiation of the extant ants, particularly in the formicoid ancestor, and for reconstructing the driving role of increasing colony size and queen-worker dimorphism for the subsequent evolution of many secondary social traits. Most of the prevalent social traits of ants are correlated, and these phenotypic correlations extend into the adaptive changes in gene clusters that often evolved in a coordinated manner subjected to intensified or relaxed selection. Our comparative study offers general confirmation of the pivotal importance of conserved signaling pathways centered around JH and MAPK signaling, and the molecular evolutionary history of these widely conserved GRNs that established and maintained the caste systems of ants over evolutionary time.

Limitations of the study

We endeavored to obtain considerably more high-quality PacBio genomes than the 125 that we report here, but retrieving enough high-quality DNA from limited samples of small and rare ants was often impossible. We were therefore unable to cover four of the 16 ant subfamilies including the elusive Martialinae, and Leptanillinae are represented by a single short-read genome, which limited the inferences we could make about the common ancestor of all extant ants. Specific sampling efforts will thus be needed to obtain these missing high-quality genomes and

to retrieve chromosome-level assemblies for more species. In our functional interpretations, we were handicapped by immature caste developmental transcriptomes being only available for *M. pharaonis* and *A. echinatior* (both Myrmicinae), and by our large-scale transcriptomic assessments being restricted to adult caste phenotypes from 70 species. These constraints meant that functional experiments could be successfully developed in only a few taxa. Future studies will undoubtedly address these omissions, extending taxonomic coverage, exploring more deeply how GRNs orchestrate the development and evolution of caste-related social traits, and analyzing evolutionary changes at lower taxonomic levels, such as in attine fungus-growing ants, in convergent clades of seed harvester ants, or in speciose genera with subclades that occupy different ecological niches.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Guojie Zhang (guojiezhang@zju.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All genome assemblies, the raw genome long- and short-read sequencing data and the RNA-seq have been deposited at the NCBI Assembly Database: https://www.ncbi.nlm.nih.gov/assembly/ and the Sequence Read Archive Database: https://www.ncbi.nlm.nih.gov/sra/ under BioProject accession codes PRJNA1172379 and PRJNA1159026.
- The genome assembly and annotation files have been deposited at https://doi.org/10.17894/ucph.79f0417a-4e15-44de-91db-3f150e59a5fc (Vizueta et al.[17]) which also contains all supplemental information mentioned in the manuscript.
- All original code has been deposited at https://github.com/schraderL/GAGA.

CONSORTIA

The members of the GAGA consortium are Serge Aron, Mauricio Bacci, Abel Bernadou, Martin Bollazzi, Raphael Boulay (deceased), Sylvia Cremer, Stephane De Greef, Alain Debec, Ofer Feinerman, Heike Feldhaar, Brian L. Fisher, Susanne Foitzik, Erik T. Frank, Jürgen Gadau, Daniele Giannetti, Donato Grasso, Heikki Helanterä, Quentin Helleu, Ana Ješovnik, Michael Lang, Fredrick J. Larabee, Bálint Markó, Mathieu Molet, David Nash, Jerome Orivel, Jes Søe Pedersen, Christian Peeters (deceased), Romain Peronnet, Eyal Privman, Christian Rabeling, Jonathan Romiguier, Francisca Ruano, Enrico Schifani, Garett Slater, Jeffrey Sosa-Calvo, András Tartally, Alberto Tinaut, Axel Touchard, Gema Trigos-Peral, Kazuki Tsuji, Lumi Viljakainen, Irene Villalta, Herbert C. Wagner, Sze Huei Yek.

ACKNOWLEDGMENTS

We thank the many ant researchers worldwide who collected samples for the GAGA project but from which we could not extract enough DNA of sufficient quality to be sequenced and included in our final dataset; they are listed in Table S1I. Major collectors of samples generating genomes became co-authors, while collectors of few sequenced species became consortium authors.

We thank Bernhard Seifert for checking the species identity of the *Lasius flavus* samples and Fuminori Ito, Jacques Delabie, Xim Cerdá, Adam Khalife, E. Jordan Smith, Akane Hubbard, and Brock Harpur for clarifying literature records or providing specific data points about social or ecological ant traits. We thank Dr. Zhilin Chen from the School of Life Sciences, Guangxi Normal University, for his assistance in examining specimens from China. We thank Sean K. McKenzie for assistance in the annotation of the Odorant Receptor gene family. We are grateful to the Science Faculty at the University of Copenhagen for access to the Computerome 2.0 computational resources. This work was supported by the National Natural Science Foundation of China (grant no. 32388102 to G.Z.), the Villum Foundation (Villum Investigator Grant, grant no. 25900 to G.Z.), the New Cornerstone Science Foundation through the XPLORER PRIZE to G.Z., the European Research Council (ERC Advanced grant 323085 to J.J.B.), the Swiss National Science Foundation (31003A-173189 and 310030-207642 to M.C. for the *F. cinerea* genome), the US National Science Foundation (DEB 1927161 to T.R.S.), and the Joint Smithsonian Institute for Biodiversity Genomics and Global Genome Initiative (T.R.S., J.J.B. N.J.S., G.Z., acknowledging grant coauthors Seán G. Brady and William T. Wcislo).

AUTHOR CONTRIBUTIONS

Conceptualization, L.S., J.J.B., and G.Z.; methodology, J.V., Z.X., R.S.L., and L.S.; formal analysis, J.V., Z.X., J.S., I.A., Y.S., D.H., Z.L., G.H., and B.P.; investigation, J.V., G.D., R.S.L., H.R., Q.G., W.D., W.J., J.Z., C.G., X.Z., D.Z., W.Z., M.S., C.L., H.Z., X.D., S.T., S.G., A.B.F.I., M.-P.M., F.H.-G., J.M.K., B.G., C.M., A.C.L., E.P.E., M.C., J.Z.S., P.S.W., J.H., Q.L., R.R.D., N.J.S., W.L., L.S. J.J.B., and G.Z.; resources, H.R., C.L., M.C., J.Z.S, J.H., T.R.S., R.R.D., L.S., J.J.B., and G.Z.; writing – original draft, J.V., Z.X., L.S., J.J.B., and G.Z.; writing – review and editing, all authors; visualization, J.V., Z.X., G.D., L.S., J.J.B., and G.Z.; supervision, J.V., W.L., L.S., J.J.B., and G.Z.; funding acquisition, M.C., T.R.S., J.J.B., and G.Z.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ant samples
- METHOD DETAILS
  - Sample collection
  - DNA extraction and sequencing
  - RNA extraction and sequencing
  - Genome assembly
  - Chromosome level assemblies
  - Publicly available genomes
  - Contamination screening
  - Confirmation of species identity and genome quality assessment
  - Repeat annotation
  - Genome annotation
  - Gene and gene family re-annotation
  - Functional annotation
  - Orthology assessment
  - GO and KEGG functional enrichments
  - Whole genome alignments (WGA)
  - Phylogenomic analyses
  - Phylogeny dating
  - Macro and microsynteny analyses
  - Caste transcriptomes and differential gene expression
  - Phenotypic trait data collection
  - Ancestral state reconstructions
  - Phylogenetic correlations

- Gene family evolution
- Glomeruli counts and their correlation with the number of OR genes
- Selective constraint analyses
- Functional experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.05.030.

Received: December 4, 2024
Revised: March 11, 2025
Accepted: May 22, 2025
Published: June 16, 2025

REFERENCES

1. Romiguier, J., Borowiec, M.L., Weyna, A., Helleu, Q., Loire, E., La Mendola, C.L., Rabeling, C., Fisher, B.L., Ward, P.S., and Keller, L. (2022). Ant phylogenomics reveals a natural selection hotspot preceding the origin of complex eusociality. Curr. Biol. *32*, 2942–2947.e4. https://doi.org/10.1016/j.cub.2022.05.001.
2. Moreau, C.S., Bell, C.D., Vila, R., Archibald, S.B., and Pierce, N.E. (2006). Phylogeny of the Ants: Diversification in the Age of Angiosperms. Science *312*, 101–104. https://doi.org/10.1126/science.1124891.
3. Ward, P.S. (2014). The Phylogeny and Evolution of Ants. Annu. Rev. Ecol. Evol. Syst. *45*, 23–43. https://doi.org/10.1146/annurev-ecolsys-120213-091824.
4. Schultheiss, P., Nooten, S.S., Wang, R., Wong, M.K.L., Brassard, F., and Guénard, B. (2022). The abundance, biomass, and distribution of ants on Earth. Proc. Natl. Acad. Sci. USA *119*, e2201550119. https://doi.org/10.1073/pnas.2201550119.
5. Hölldobler, B., and Wilson, E.O. (1990). The Ants (Harvard University Press).
6. Andersen, A.N. (2021). Diversity, Biogeography and Community Ecology of Ants: Introduction to the Special Issue. Diversity *13*, 625. https://doi.org/10.3390/d13120625.
7. Gotoh, A., Billen, J., Hashim, R., and Ito, F. (2016). Degeneration patterns of the worker spermatheca during morphogenesis in ants (Hymenoptera: Formicidae). Evol. Dev. *18*, 96–104. https://doi.org/10.1111/ede.12182.
8. AntCat. (2024). An Online Catalog of the Ants of the World. https://antcat.org/.
9. Peeters, C., and Ito, F. (2015). Wingless and dwarf workers underlie the ecological success of ants (Hymenoptera: Formicidae). Myrmecol. News *21*, 117–130. https://doi.org/10.5281/zenodo.845751.
10. Burchill, A.T., and Moreau, C.S. (2016). Colony size evolution in ants: macroevolutionary trends. Insect Soc. *63*, 291–298. https://doi.org/10.1007/s00040-016-0465-3.
11. Keller, L., and Genoud, M. (1997). Extraordinary lifespans in ants: a test of evolutionary theories of ageing. Nature *389*, 958–960. https://doi.org/10.1038/40130.
12. Darwin, C. (1859). On the Origin of Species (Oxford University Press).
13. Weismann, A. (1893). The All-Sufficiency of Natural Selection: A Reply to Herbert Spencer. Contemporary Review *64*, 309–338.
14. Wheeler, W.M. (1911). The ant-colony as an organism. J. Morphol. *22*, 307–325. https://doi.org/10.1002/jmor.1050220206.
15. Wheeler, W.M. (1910). Ants: Their Structure, Development and Behavior (Columbia University Press).
16. Wheeler, W.M. (1923). Social Life Among the Insects (Harcourt, Brace).
17. Vizueta, J., Xiong, Z., Schrader, L., Boomsma, J.J., and Zhang, G. (2025). Data repository for Vizueta et al. Cell 2025, Electronic Research Data Archive. https://doi.org/10.17894/ucph.79f0417a-4e15-44de-91db-3f150e59a5fc.
18. McKenzie, S.K., and Kronauer, D.J.C. (2018). The genomic architecture and molecular evolution of ant odorant receptors. Genome Res. *28*, 1757–1765. https://doi.org/10.1101/gr.237123.118.
19. Li, R., Dai, X., Zheng, J., Larsen, R.S., Qi, Y., Zhang, X., Vizueta, J., Boomsma, J.J., Liu, W., and Zhang, G. (2024). Juvenile hormone as a key regulator for asymmetric caste differentiation in ants. Proc. Natl. Acad. Sci. USA *121*, e2406999121. https://doi.org/10.1073/pnas.2406999121.
20. Qiu, B., Dai, X., Li, P., Larsen, R.S., Li, R., Price, A.L., Ding, G., Texada, M. J., Zhang, X., Zuo, D., et al. (2022). Canalized gene expression during development mediates caste differentiation in ants. Nat. Ecol. Evol. *6*, 1753–1765. https://doi.org/10.1038/s41559-022-01884-y.
21. Hughes, W.O.H., Oldroyd, B.P., Beekman, M., and Ratnieks, F.L.W. (2008). Ancestral Monogamy Shows Kin Selection Is Key to the Evolution of Eusociality. Science *320*, 1213–1216. https://doi.org/10.1126/science.1156108.
22. Fisher, R.M., Shik, J.Z., and Boomsma, J.J. (2020). The evolution of multicellular complexity: the role of relatedness and environmental constraints. Proc. Biol. Sci. *287*, 20192963. https://doi.org/10.1098/rspb.2019.2963.
23. Hamilton, W.D. (1963). The Evolution of Altruistic Behavior. Am. Nat. *97*, 354–356. https://doi.org/10.1086/497114.
24. Boomsma, J.J. (2022). Domains and Major Transitions of Social Evolution (Oxford University Press) https://doi.org/10.1093/oso/9780198746171.001.0001.
25. Gadau, J., Helmkampf, M., Nygaard, S., Roux, J., Simola, D.F., Smith, C. R., Suen, G., Wurm, Y., and Smith, C.D. (2012). The genomic impact of 100 million years of social evolution in seven ant species. Trends Genet. *28*, 14–21. https://doi.org/10.1016/j.tig.2011.08.005.
26. Simola, D.F., Wissler, L., Donahue, G., Waterhouse, R.M., Helmkampf, M., Roux, J., Nygaard, S., Glastad, K.M., Hagen, D.E., Viljakainen, L., et al. (2013). Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. Genome Res. *23*, 1235–1247. https://doi.org/10.1101/gr.155408.113.
27. Barkdull, M., and Moreau, C.S. (2023). Worker Reproduction and Caste Polymorphism Impact Genome Evolution and Social Genes Across the Ants. Genome Biol. Evol. *15*, evad095. https://doi.org/10.1093/gbe/evad095.
28. Roux, J., Privman, E., Moretti, S., Daub, J.T., Robinson-Rechavi, M., and Keller, L. (2014). Patterns of Positive Selection in Seven Ant Genomes. Mol. Biol. Evol. *31*, 1661–1685. https://doi.org/10.1093/molbev/msu141.
29. Gao, Q., Xiong, Z., Larsen, R.S., Zhou, L., Zhao, J., Ding, G., Zhao, R., Liu, C., Ran, H., and Zhang, G. (2020). High-quality chromosome-level genome assembly and full-length transcriptome analysis of the pharaoh ant *Monomorium pharaonis*. GigaScience *9*, giaa143. https://doi.org/10.1093/gigascience/giaa143.
30. Yan, Z., Martin, S.H., Gotzek, D., Arsenault, S.V., Duchen, P., Helleu, Q., Riba-Grognuz, O., Hunt, B.G., Salamin, N., Shoemaker, D., et al. (2020). Evolution of a supergene that regulates a trans-species social polymorphism. Nat. Ecol. Evol. *4*, 240–249. https://doi.org/10.1038/s41559-019-1081-1.
31. Wurm, Y., Wang, J., Riba-Grognuz, O., Corona, M., Nygaard, S., Hunt, B. G., Ingram, K.K., Falquet, L., Nipitwattanaphon, M., Gotzek, D., et al. (2011). The genome of the fire ant *Solenopsis invicta*. Proc. Natl. Acad. Sci. USA *108*, 5679–5684. https://doi.org/10.1073/pnas.1009690108.
32. Shields, E.J., Sheng, L., Weiner, A.K., Garcia, B.A., and Bonasio, R. (2018). High-Quality Genome Assemblies Reveal Long Non-coding RNAs Expressed in Ant Brains. Cell Rep. *23*, 3078–3090. https://doi.org/10.1016/j.celrep.2018.05.014.

33. Brelsford, A., Purcell, J., Avril, A., Tran Van, P.T., Zhang, J., Brütsch, T., Sundström, L., Helanterä, H., and Chapuisat, M. (2020). An Ancient and Eroded Social Supergene Is Widespread across *Formica* Ants. Curr. Biol. *30*, 304–311.e4. https://doi.org/10.1016/j.cub.2019.11.032.

34. 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., et al. (2017). The Global Ant Genomics Alliance (GAGA). Myrmecol. News *25*, 61–66.

35. Hawlitschek, O., Sadílek, D., Dey, L.-S., Buchholz, K., Noori, S., Baez, I. L., Wehrt, T., Brozio, J., Trávníček, P., Seidel, M., et al. (2023). New estimates of genome size in Orthoptera and their evolutionary implications. PLoS One *18*, e0275551. https://doi.org/10.1371/journal.pone.0275551.

36. Kelley, J.L., Peyton, J.T., Fiston-Lavier, A.-S., Teets, N.M., Yee, M.-C., Johnston, J.S., Bustamante, C.D., Lee, R.E., and Denlinger, D.L. (2014). Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. Nat. Commun. *5*, 4611. https://doi.org/10.1038/ncomms5611.

37. Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I.T., Novak, A.M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., et al. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature *587*, 246–251. https://doi.org/10.1038/s41586-020-2871-y.

38. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

39. Chen, M.-Y., Liang, D., and Zhang, P. (2017). Phylogenomic Resolution of the Phylogeny of Laurasiatherian Mammals: Exploring Phylogenetic Signals within Coding and Noncoding Sequences. Genome Biol. Evol. *9*, 1998–2012. https://doi.org/10.1093/gbe/evx147.

40. Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., et al. (2015). Phylogenomic analyses data of the avian phylogenomics project. GigaScience *4*. s13742-014-0038-1. https://doi.org/10.1186/s13742-014-0038-1.

41. Stiller, J., Feng, S., Chowdhury, A.-A., Rivas-González, I., Duchêne, D.A., Fang, Q., Deng, Y., Kozlov, A., Stamatakis, A., Claramunt, S., et al. (2024). Complexity of avian evolution revealed by family-level genomes. Nature *629*, 851–860. https://doi.org/10.1038/s41586-024-07323-1.

42. Borowiec, M.L., Zhang, Y.M., Neves, K., Ramalho, M.O., Fisher, B.L., Lucky, A., and Moreau, C.S. (2025). Evaluating UCE Data Adequacy and Integrating Uncertainty in a Comprehensive Phylogeny of Ants. Published online January 8, 2025. Syst. Biol. syaf001. https://doi.org/10.1093/sysbio/syaf001.

43. Ward, P.S. (2007). Phylogeny, classification, and species-level taxonomy of ants (Hymenoptera: Formicidae). Zootaxa *1668*, 549–563. https://doi.org/10.11646/zootaxa.1668.1.26.

44. Lorite, P., and Palomeque, T. (2010). Karyotype evolution in ants (Hymenoptera: Formicidae), with a review of the known ant chromosome numbers. Myrmecol. News *13*, 89–102.

45. Crosland, M.W.J., and Crozier, R.H. (1986). *Myrmecia pilosula*, an Ant with Only One Pair of Chromosomes. Science *231*, 1278. https://doi.org/10.1126/science.231.4743.1278.

46. Debec, A., Peronnet, R., Lang, M., and Molet, M. (2024). Primary cell cultures from the single-chromosome ant *Myrmecia croslandi*. Chromosome Res. *32*, 10. https://doi.org/10.1007/s10577-024-09755-x.

47. Nygaard, S., Hu, H., Li, C., Schiøtt, M., Chen, Z., Yang, Z., Xie, Q., Ma, C., Deng, Y., Dikow, R.B., et al. (2016). Reciprocal genomic evolution in the ant–fungus agricultural symbiosis. Nat. Commun. *7*, 12233. https://doi.org/10.1038/ncomms12233.

48. Augustijnen, H., Bätscher, L., Cesanek, M., Chkhartishvili, T., Dincă, V., Iankoshvili, G., Ogawa, K., Vila, R., Klopfstein, S., de Vos, J.M., et al. (2024). A macroevolutionary role for chromosomal fusion and fission in *Erebia* butterflies. Sci. Adv. *10*, eadl0989. https://doi.org/10.1126/sciadv.adl0989.

49. Ayala, F.J., and Coluzzi, M. (2005). Chromosome speciation: humans, *Drosophila*, and mosquitoes. Proc. Natl. Acad. Sci. USA *102*, 6535–6542. https://doi.org/10.1073/pnas.0501847102.

50. Li, Y., Liu, H., Steenwyk, J.L., LaBella, A.L., Harrison, M.-C., Groenewald, M., Zhou, X., Shen, X.-X., Zhao, T., Hittinger, C.T., et al. (2022). Contrasting modes of macro and microsynteny evolution in a eukaryotic subphylum. Curr. Biol. *32*, 5335–5343.e4. https://doi.org/10.1016/j.cub.2022.10.025.

51. Lovell, S.C., Li, X., Weerasinghe, N.R., and Hentges, K.E. (2009). Correlation of microsynteny conservation and disease gene distribution in mammalian genomes. BMC Genomics *10*, 521. https://doi.org/10.1186/1471-2164-10-521.

52. Engström, P.G., Ho Sui, S.J., Drivenes, O., Becker, T.S., and Lenhard, B. (2007). Genomic regulatory blocks underlie extensive microsynteny conservation in insects. Genome Res. *17*, 1898–1908. https://doi.org/10.1101/gr.6669607.

53. Kapheim, K.M., Pan, H., Li, C., Salzberg, S.L., Puiu, D., Magoc, T., Robertson, H.M., Hudson, M.E., Venkat, A., Fischman, B.J., et al. (2015). Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. Science *348*, 1139–1143. https://doi.org/10.1126/science.aaa4788.

54. McKenzie, S.K., Fetter-Pruneda, I., Ruta, V., and Kronauer, D.J.C. (2016). Transcriptomics and neuroanatomy of the clonal raider ant implicate an expanded clade of odorant receptors in chemical communication. Proc. Natl. Acad. Sci. USA *113*, 14091–14096. https://doi.org/10.1073/pnas.1610800113.

55. Lepeco, A., Meira, O.M., Matielo, D.M., Brandão, C.R.F., and Camacho, G.P. (2025). A hell ant from the Lower Cretaceous of Brazil. Curr. Biol. *35*, 2146–2153.e2. https://doi.org/10.1016/j.cub.2025.03.023.

56. Chomicki, G., Ward, P.S., and Renner, S.S. (2015). Macroevolutionary assembly of ant/plant symbioses: *Pseudomyrmex* ants and their ant-housing plants in the Neotropics. Proc. Biol. Sci. *282*, 20152200. https://doi.org/10.1098/rspb.2015.2200.

57. Ward, P.S. (1990). The ant subfamily Pseudomyrmecinae (Hymenoptera: Formicidae): generic revision and relationship to other formicids. Syst. Entomol. *15*, 449–489. https://doi.org/10.1111/j.1365-3113.1990.tb00077.x.

58. Ward, P.S., and Downie, D.A. (2005). The ant subfamily Pseudomyrmecinae (Hymenoptera: Formicidae): phylogeny and evolution of big-eyed arboreal ants. Syst. Entomol. *30*, 310–335. https://doi.org/10.1111/j.1365-3113.2004.00281.x.

59. Peeters, C., and Ito, F. (2001). Colony dispersal and the evolution of queen morphology in social Hymenoptera. Annu. Rev. Entomol. *46*, 601–630. https://doi.org/10.1146/annurev.ento.46.1.601.

60. Trible, W., and Kronauer, D.J.C. (2017). Caste development and evolution in ants: it's all about size. J. Exp. Biol. *220*, 53–62. https://doi.org/10.1242/jeb.145292.

61. Nijhout, H.F., and Wheeler, D.E. (1996). Growth Models of Complex Allometries in Holometabolous Insects. Am. Nat. *148*, 40–56. https://doi.org/10.1086/285910.

62. Penick, C.A., Prager, S.S., and Liebig, J. (2012). Juvenile hormone induces queen development in late-stage larvae of the ant *Harpegnathos saltator*. J. Insect Physiol. *58*, 1643–1649. https://doi.org/10.1016/j.jinsphys.2012.10.004.

63. Dolezal, A.G. (2019). Caste Determination in Arthropods. In Encyclopedia of Animal Behavior, Second Edition, J.C. Choe, ed. (Academic Press), pp. 691–698. https://doi.org/10.1016/B978-0-12-809633-8.20815-7.

64. Cnaani, J., Robinson, G.E., and Hefetz, A. (2000). The critical period for caste determination in *Bombus terrestris* and its juvenile hormone correlates. J. Comp. Physiol. A *186*, 1089–1094. https://doi.org/10.1007/s003590000163.

65. Korb, J. (2015). Juvenile Hormone. In Advances in Insect Physiology Genomics, Physiology and Behaviour of Social Insects, A. Zayed and C.F.

Kent, eds. (Academic Press), pp. 131–161. https://doi.org/10.1016/bs.aiip.2014.12.004.

66. Mutti, N.S., Dolezal, A.G., Wolschin, F., Mutti, J.S., Gill, K.S., and Amdam, G.V. (2011). IRS and TOR nutrient-signaling pathways act *via* juvenile hormone to influence honey bee caste fate. J. Exp. Biol. *214*, 3977–3984. https://doi.org/10.1242/jeb.061499.

67. Gospocic, J., Glastad, K.M., Sheng, L., Shields, E.J., Berger, S.L., and Bonasio, R. (2021). Kr-h1 maintains distinct caste-specific neurotranscriptomes in response to socially regulated hormones. Cell *184*, 5807–5823.e14. https://doi.org/10.1016/j.cell.2021.10.006.

68. Bloch, G., Wheeler, D.E., and Robinson, G.E. (2002). Endocrine Influences on the Organization of Insect Societies. In Hormones, Brain and Behavior (Elsevier), pp. 195–235. https://doi.org/10.1016/B978-012532104-4/50042-1.

69. Cargnello, M., and Roux, P.P. (2011). Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. Microbiol. Mol. Biol. Rev. *75*, 50–83. https://doi.org/10.1128/MMBR.00031-10.

70. Bar-Sagi, D. (1994). The Sos (Son of sevenless) protein. Trends Endocrinol. Metab. *5*, 165–169. https://doi.org/10.1016/1043-2760(94)90014-0.

71. Shilo, B.-Z. (2014). The regulation and functions of MAPK pathways in *Drosophila*. Methods *68*, 151–159. https://doi.org/10.1016/j.ymeth.2014.01.020.

72. Bradshaw, T., Simmons, C., Ott, R.K., and Armstrong, A.R. (2024). Ras/MAPK signaling mediates adipose tissue control of ovarian germline survival and ovulation in *Drosophila melanogaster*. Dev. Biol. *510*, 17–28. https://doi.org/10.1016/j.ydbio.2024.02.009.

73. Slack, C., Alic, N., Foley, A., Cabecinha, M., Hoddinott, M.P., and Partridge, L. (2015). The Ras-Erk-ETS-Signaling Pathway Is a Drug Target for Longevity. Cell *162*, 72–83. https://doi.org/10.1016/j.cell.2015.06.023.

74. Yan, H., Opachaloemphan, C., Carmona-Aldana, F., Mancini, G., Mlejnek, J., Descostes, N., Sieriebriennikov, B., Leibholz, A., Zhou, X., Ding, L., et al. (2022). Insulin signaling in the long-lived reproductive caste of ants. Science *377*, 1092–1099. https://doi.org/10.1126/science.abm8767.

75. Buschinger, A. (2009). Social parasitism among ants: a review (Hymenoptera: Formicidae). Myrmecol. News *12*, 219–235.

76. Rabeling, C., Schultz, T.R., Pierce, N.E., and Bacci, M. (2014). A Social Parasite Evolved Reproductive Isolation from Its Fungus-Growing Ant Host in Sympatry. Curr. Biol. *24*, 2047–2052. https://doi.org/10.1016/j.cub.2014.07.048.

77. Dahan, R.A., and Rabeling, C. (2022). Multi-queen breeding is associated with the origin of inquiline social parasitism in ants. Sci. Rep. *12*, 14680. https://doi.org/10.1038/s41598-022-17595-0.

78. Kaspari, M., and Vargo, E.L. (1995). Colony Size as a Buffer Against Seasonality: Bergmann's Rule in Social Insects. Am. Nat. *145*, 610–632. https://doi.org/10.1086/285758.

79. Oster, G.F., and Wilson, E.O. (1978). Caste and Ecology in the Social Insects (Princeton University Press).

80. Simola, D.F., Graham, R.J., Brady, C.M., Enzmann, B.L., Desplan, C., Ray, A., Zwiebel, L.J., Bonasio, R., Reinberg, D., Liebig, J., et al. (2016). Epigenetic (re)programming of caste-specific behavior in the ant *Camponotus floridanus*. Science *351*, aac6633. https://doi.org/10.1126/science.aac6633.

81. Das, B., and de Bekker, C. (2022). Time-course RNASeq of *Camponotus floridanus* forager and nurse ant brains indicate links between plasticity in the biological clock and behavioral division of labor. BMC Genomics *23*, 57. https://doi.org/10.1186/s12864-021-08282-x.

82. Jones, B.W., Fetter, R.D., Tear, G., and Goodman, C.S. (1995). glial cells missing: a genetic switch that controls glial versus neuronal fate. Cell *82*, 1013–1023. https://doi.org/10.1016/0092-8674(95)90280-5.

83. Ramsay, C., Lasko, P., and Abouheif, E. (2020). Evo-Devo Lessons from the Reproductive Division of Labor in Eusocial Hymenoptera. In Evolutionary Developmental Biology: A Reference Guide, L, L. Nuno de la Rosa and G. Müller, eds. (Springer International Publishing), pp. 1–14. https://doi.org/10.1007/978-3-319-33038-9_173-1.

84. Ratnieks, F.L.W., Foster, K.R., and Wenseleers, T. (2006). Conflict resolution in insect societies. Annu. Rev. Entomol. *51*, 581–608. https://doi.org/10.1146/annurev.ento.51.110104.151003.

85. Glenn, L.E., and Searles, L.L. (2001). Distinct domains mediate the early and late functions of the *Drosophila ovarian tumor* proteins. Mech. Dev. *102*, 181–191. https://doi.org/10.1016/S0925-4773(01)00314-8.

86. Romiguier, J., Fournier, A., Yek, S.H., and Keller, L. (2017). Convergent evolution of social hybridogenesis in *Messor* harvester ants. Mol. Ecol. *26*, 1108–1117. https://doi.org/10.1111/mec.13899.

87. Schrader, L., Pan, H., Bollazzi, M., Schiøtt, M., Larabee, F.J., Bi, X., Deng, Y., Zhang, G., Boomsma, J.J., and Rabeling, C. (2021). Relaxed selection underlies genome erosion in socially parasitic ant species. Nat. Commun. *12*, 2918. https://doi.org/10.1038/s41467-021-23178-w.

88. Feldmeyer, B., Gstöttl, C., Wallner, J., Jongepier, E., Séguret, A., Grasso, D.A., Bornberg-Bauer, E., Foitzik, S., and Heinze, J. (2022). Evidence for a conserved queen-worker genetic toolkit across slave-making ants and their ant hosts. Mol. Ecol. *31*, 4991–5004. https://doi.org/10.1111/mec.16639.

89. Purkart, A., Wagner, H.C., Goffová, K., Selnekovič, D., and Holecová, M. (2022). Laboratory observations on *Anergates atratulus* (Schenck, 1852): mating behaviour, incorporation into host colonies, and competition with *Strongylognathus testaceus* (Schenck, 1852). Biologia *77*, 125–135. https://doi.org/10.1007/s11756-021-00901-y.

90. Anderson, C., and McShea, D.W. (2001). Individual versus social complexity, with particular reference to ant colonies. Biol. Rev. Camb. Philos. Soc. *76*, 211–237. https://doi.org/10.1017/S1464793101005656.

91. Bonner, J. (1993). Dividing the labour in cells and societies. Curr. Sci. *64*, 459–466.

92. Bourke. (1999). Colony size, social complexity and reproductive conflict in social insects. J. Evol. Biol. *12*, 245–257. https://doi.org/10.1046/j.1420-9101.1999.00028.x.

93. Bell-Roberts, L., Turner, J.F.R., Werner, G.D.A., Downing, P.A., Ross, L., and West, S.A. (2024). Larger colony sizes favoured the evolution of more worker castes in ants. Nat. Ecol. Evol. *8*, 1959–1971. https://doi.org/10.1038/s41559-024-02512-7.

94. Matte, A., and LeBoeuf, A.C. (2025). Innovation in ant larval feeding facilitated queen–worker divergence and social complexity. Proc. Natl. Acad. Sci. USA *122*, e2413742122. https://doi.org/10.1073/pnas.2413742122.

95. Rabeling, C. (2021). Social Parasitism. In Encyclopedia of Social Insects, C.K. Starr, ed. (Springer International Publishing), pp. 836–858. https://doi.org/10.1007/978-3-030-28102-1_175.

96. Jones, B.M., Rubin, B.E.R., Dudchenko, O., Kingwell, C.J., Traniello, I.M., Wang, Z.Y., Kapheim, K.M., Wyman, E.S., Adastra, P.A., Liu, W., et al. (2023). Convergent and complementary selection shaped gains and losses of eusociality in sweat bees. Nat. Ecol. Evol. *7*, 557–569. https://doi.org/10.1038/s41559-023-02001-3.

97. Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nat. Methods *17*, 155–158. https://doi.org/10.1038/s41592-019-0669-3.

98. Boetzer, M., and Pirovano, W. (2014). SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. BMC Bioinformatics *15*, 211. https://doi.org/10.1186/1471-2105-15-211.

99. Xu, G.-C., Xu, T.-J., Zhu, R., Zhang, Y., Li, S.-Q., Wang, H.-W., and Li, J.-T. (2019). LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. GigaScience *8*, giy157. https://doi.org/10.1093/gigascience/giy157.

100. Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics *36*, 2253–2255. https://doi.org/10.1093/bioinformatics/btz891.

101. Guo, L., Xu, M., Wang, W., Gu, S., Zhao, X., Chen, F., Wang, O., Xu, X., Seim, I., Fan, G., et al. (2021). SLR-superscaffolder: a de novo scaffolding tool for synthetic long reads using a top-to-bottom scheme. BMC Bioinformatics *22*, 158. https://doi.org/10.1186/s12859-021-04081-z.

102. Zimin, A.V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S.L., and Yorke, J.A. (2013). The MaSuRCA genome assembler. Bioinformatics *29*, 2669–2677. https://doi.org/10.1093/bioinformatics/btt476.

103. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95. https://doi.org/10.1126/science.aal3327.

104. Wang, Y.-B., and Zhang, X. (2022). Chromosome Scaffolding of Diploid Genomes Using ALLHiC. Bio. Protoc. *12*, e4503. https://doi.org/10.21769/BioProtoc.4503.

105. Ghurye, J., Rhie, A., Walenz, B.P., Schmitt, A., Selvaraj, S., Pop, M., Phillippy, A.M., and Koren, S. (2019). Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput. Biol. *15*, e1007273. https://doi.org/10.1371/journal.pcbi.1007273.

106. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst. *3*, 99–101. https://doi.org/10.1016/j.cels.2015.07.012.

107. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

108. Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

109. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

110. Vizueta, J., Sánchez-Gracia, A., and Rozas, J. (2020). BITACORA: A comprehensive tool for the identification and annotation of gene families in genome assemblies. Mol. Ecol. Resour. *20*, 1445–1452. https://doi.org/10.1111/1755-0998.13202.

111. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

112. Huang, N., and Li, H. (2023). compleasm: a faster and more accurate reimplementation of BUSCO. Bioinformatics *39*, btad595. https://doi.org/10.1093/bioinformatics/btad595.

113. Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. *21*, 245. https://doi.org/10.1186/s13059-020-02134-9.

114. Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics *36*, 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

115. Smit, A., Hubley, R., and Green, P.. RepeatMasker Open-4.0.

116. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

117. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580. https://doi.org/10.1093/nar/27.2.573.

118. Orozco-Arias, S., Sierra, P., Durbin, R., and González, J. (2024). MCHelper automatically curates transposable element libraries across eukaryotic species. Genome Res. *34*, 2256–2268. https://doi.org/10.1101/gr.278821.123.

119. Abrusán, G., Grundmann, N., DeMester, L., and Makalowski, W. (2009). TEclass—a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics *25*, 1329–1330. https://doi.org/10.1093/bioinformatics/btp084.

120. Challis, R., Richards, E., Rajan, J., Cochrane, G., and Blaxter, M. (2020). BlobToolKit – Interactive Quality Assessment of Genome Assemblies. G3 Genes. G3 (Bethesda) *10*, 1361–1374. https://doi.org/10.1534/g3.119.400908.

121. Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. *35*, 1026–1028. https://doi.org/10.1038/nbt.3988.

122. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

123. Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. *44*, e89. https://doi.org/10.1093/nar/gkw092.

124. Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S.O., and Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics *19*, 189. https://doi.org/10.1186/s12859-018-2203-5.

125. Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. *20*, 278. https://doi.org/10.1186/s13059-019-1910-1.

126. Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. *34*, W435–W439. https://doi.org/10.1093/nar/gkl200.

127. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

128. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput. Biol. *7*, e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

129. Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. *30*, 772–780. https://doi.org/10.1093/molbev/mst010.

130. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

131. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol. Biol. Evol. *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293.

132. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

133. Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257–258. https://doi.org/10.1093/bioinformatics/btl567.

134. Alexa, A., and Rahnenfuhrer, J. (2024). topGO: Enrichment Analysis for Gene Ontology. R package version 2.56.0. Bioconductor. http://bioconductor.org/packages/topGO/.

135. Sayols, S. (2023). rrvgo: a Bioconductor package for interpreting lists of Gene Ontology terms. MicroPubl. Biol. *2023*. https://doi.org/10.17912/micropub.biology.000811.

136. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation. *2*, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

137. Luo, W., and Brouwer, C. (2013). Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics *29*, 1830–1831. https://doi.org/10.1093/bioinformatics/btt285.

138. Harris, R.S. (2007). Improved Pairwise Alignment of Genomic DNA (The Pennsylvania State University).

139. Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R. C., Haussler, D., and Miller, W. (2003). Human-mouse alignments with BLASTZ. Genome Res. *13*, 103–107. https://doi.org/10.1101/gr.809403.

140. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. *28*, 1747–1756. https://doi.org/10.1101/gr.239244.118.

141. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19*, 153. https://doi.org/10.1186/s12859-018-2129-y.

142. Löytynoja, A., and Goldman, N. (2010). webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics *11*, 579. https://doi.org/10.1186/1471-2105-11-579.

143. Mongiardino Koch, N. (2021). Phylogenomic Subsampling and the Search for Phylogenetically Reliable Loci. Mol. Biol. Evol. *38*, 4025–4038. https://doi.org/10.1093/molbev/msab151.

144. Faircloth, B.C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics *32*, 786–788. https://doi.org/10.1093/bioinformatics/btv646.

145. Puttick, M.N. (2019). MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics *35*, 5321–5322. https://doi.org/10.1093/bioinformatics/btz554.

146. Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. *24*, 1586–1591. https://doi.org/10.1093/molbev/msm088.

147. Rambaut, A., Drummond, A.J., Xie, D., Baele, G., and Suchard, M.A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. *67*, 901–904. https://doi.org/10.1093/sysbio/syy032.

148. Osipova, E., Hecker, N., and Hiller, M. (2019). RepeatFiller newly identifies megabases of aligning repetitive sequences and improves annotations of conserved non-exonic elements. GigaScience *8*, giz132. https://doi.org/10.1093/gigascience/giz132.

149. Hao, Z., Lv, D., Ge, Y., Shi, J., Weijers, D., Yu, G., and Chen, J. (2020). RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput. Sci. *6*, e251. https://doi.org/10.7717/peerj-cs.251.

150. Muffato, M., Louis, A., Nguyen, N.T.T., Lucas, J., Berthelot, C., and Roest Crollius, H. (2023). Reconstruction of hundreds of reference ancestral genomes across the eukaryotic kingdom. Nat. Ecol. Evol. *7*, 355–366. https://doi.org/10.1038/s41559-022-01956-z.

151. Lucas, J.M., and Roest Crollius, H.R. (2017). High precision detection of conserved segments from synteny blocks. PLoS One *12*, e0180198. https://doi.org/10.1371/journal.pone.0180198.

152. Robert, N.S.M., Sarigol, F., Zieger, E., and Simakov, O. (2022). SYNPHONI: scale-free and phylogeny-aware reconstruction of synteny conservation and transformation across animal genomes. Bioinformatics *38*, 5434–5436. https://doi.org/10.1093/bioinformatics/btac695.

153. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

154. Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., Li, Y., Ye, J., Yu, C., Li, Z., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. GigaScience *7*, 1–6. https://doi.org/10.1093/gigascience/gix120.

155. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

156. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.

157. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528. https://doi.org/10.1093/bioinformatics/bty633.

158. Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. *8*, 28–36. https://doi.org/10.1111/2041-210X.12628.

159. Wang, L.-G., Lam, T.T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C.W., Jones, B.R., Bradley, T., et al. (2020). Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. Mol. Biol. Evol. *37*, 599–603. https://doi.org/10.1093/molbev/msz240.

160. Mendes, F.K., Vanderpool, D., Fulton, B., and Hahn, M.W. (2021). CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics *36*, 5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022.

161. Wu, M., Chatterji, S., and Eisen, J.A. (2012). Accounting For Alignment Uncertainty in Phylogenomics. PLoS One *7*, e30288. https://doi.org/10.1371/journal.pone.0030288.

162. Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., and Frost, S.D.W. (2006). Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. Mol. Biol. Evol. *23*, 1891–1901. https://doi.org/10.1093/molbev/msl051.

163. Di Franco, A., Poujol, R., Baurain, D., and Philippe, H. (2019). Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. BMC Evol. Biol. *19*, 21. https://doi.org/10.1186/s12862-019-1350-2.

164. Stalling, D., Westerhoff, M., and Hege, H.-C. (2005). amira: A Highly Interactive System for Visual Data Analysis. In Visualization Handbook (Elsevier), pp. 749–767. https://doi.org/10.1016/B978-012387582-2/50040-X.

165. Wang, O., Chin, R., Cheng, X., Wu, M.K.Y., Mao, Q., Tang, J., Sun, Y., Anderson, E., Lam, H.K., Chen, D., et al. (2019). Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. Genome Res. *29*, 798–808. https://doi.org/10.1101/gr.245126.118.

166. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I. D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

167. Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569. https://doi.org/10.1038/nmeth.2474.

168. Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722–736. https://doi.org/10.1101/gr.215087.116.

169. Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M., and Jaffe, D.B. (2017). Direct determination of diploid genome sequences. Genome Res. *27*, 757–767. https://doi.org/10.1101/gr.214874.116.

170. Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J. L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., et al. (2014). PATRIC,

the bacterial bioinformatics database and analysis resource. Nucleic Acids Res. *42*, D581–D591. https://doi.org/10.1093/nar/gkt1099.

171. Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA *6*, 11. https://doi.org/10.1186/s13100-015-0041-9.
172. Hoskins, R.A., Carlson, J.W., Wan, K.H., Park, S., Mendez, I., Galle, S.E., Booth, B.W., Pfeiffer, B.D., George, R.A., Svirskas, R., et al. (2015). The Release 6 reference sequence of the Drosophila melanogaster genome. Genome Res. *25*, 445–458. https://doi.org/10.1101/gr.185579.114.
173. Kim, H.S., Murphy, T., Xia, J., Caragea, D., Park, Y., Beeman, R.W., Lorenzen, M.D., Butcher, S., Manak, J.R., and Brown, S.J. (2010). BeetleBase in 2010: revisions to provide comprehensive genomic information for *Tribolium castaneum*. Nucleic Acids Res. *38*, D437–D442. https://doi.org/10.1093/nar/gkp807.
174. Richards, S., Gibbs, R.A., Weinstock, G.M., Brown, S.J., Denell, R., Beeman, R.W., Gibbs, R., Beeman, R.W., Brown, S.J., et al.; Tribolium Genome Sequencing Consortium (2008). The genome of the model beetle and pest *Tribolium castaneum*. Nature *452*, 949–955. https://doi.org/10.1038/nature06784.
175. Werren, J.H., Richards, S., Desjardins, C.A., Niehuis, O., Gadau, J., Colbourne, J.K., Werren, J.H., Richards, S., Desjardins, C.A., et al.; Nasonia Genome Working Group (2010). Functional and Evolutionary Insights from the Genomes of Three Parasitoid *Nasonia* Species. Science *327*, 343–348. https://doi.org/10.1126/science.1178028.
176. Honeybee Genome Sequencing Consortium (2006). Insights into social insects from the genome of the honeybee *Apis mellifera*. Nature *443*, 931–949. https://doi.org/10.1038/nature05260.
177. Joseph, R.M., and Carlson, J.R. (2015). *Drosophila* chemoreceptors: A molecular interface between the chemical world and the brain. Trends Genet. *31*, 683–695. https://doi.org/10.1016/j.tig.2015.09.005.
178. Benton, R., Dessimoz, C., and Moi, D. (2020). A putative origin of the insect chemosensory receptor superfamily in the last common eukaryotic ancestor. eLife *9*, e62507. https://doi.org/10.7554/eLife.62507.
179. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314. https://doi.org/10.1093/nar/gky1085.
180. Hickey, G., Paten, B., Earl, D., Zerbino, D., and Haussler, D. (2013). HAL: a hierarchical format for storing and analyzing multiple genome alignments. Bioinformatics *29*, 1341–1342. https://doi.org/10.1093/bioinformatics/btt128.
181. Branstetter, M.G., Longino, J.T., Ward, P.S., and Faircloth, B.C. (2017). Enriching the ant tree of life: enhanced UCE bait set for genome-scale phylogenetics of ants and other Hymenoptera. Methods Ecol. Evol. *8*, 768–776. https://doi.org/10.1111/2041-210X.12742.
182. Hedman, M.M. (2010). Constraints on clade ages from fossil outgroups. Paleobiology *36*, 16–31. https://doi.org/10.1666/0094-8373-36.1.16.
183. Friedman, M., Keck, B.P., Dornburg, A., Eytan, R.I., Martin, C.H., Hulsey, C.D., Wainwright, P.C., and Near, T.J. (2013). Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proc. Biol. Sci. *280*, 20131733. https://doi.org/10.1098/rspb.2013.1733.
184. Borowiec, M.L., Rabeling, C., Brady, S.G., Fisher, B.L., Schultz, T.R., and Ward, P.S. (2019). Compositional heterogeneity and outgroup choice influence the internal phylogeny of the ants. Mol. Phylogenet. Evol. *134*, 111–121. https://doi.org/10.1016/j.ympev.2019.01.024.
185. Brady, S.G., Fisher, B.L., Schultz, T.R., and Ward, P.S. (2014). The rise of army ants and their relatives: diversification of specialized predatory doryline ants. BMC Evol. Biol. *14*, 93. https://doi.org/10.1186/1471-2148-14-93.
186. Economo, E.P., Narula, N., Friedman, N.R., Weiser, M.D., and Guénard, B. (2018). Macroecology and macroevolution of the latitudinal diversity gradient in ants. Nat. Commun. *9*, 1778. https://doi.org/10.1038/s41467-018-04218-4.
187. Moreau, C.S., and Bell, C.D. (2013). Testing the museum versus cradle tropical biological diversity hypothesis: Phylogeny, diversification, and ancestral biogeographic range evolution of the ants. Evolution *67*, 2240–2257. https://doi.org/10.1111/evo.12105.
188. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.
189. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25. https://doi.org/10.1186/gb-2010-11-3-r25.
190. Helanterä, H. (2022). Supercolonies of ants (Hymenoptera: Formicidae): ecological patterns, behavioural processes and their implications for social evolution. Myrmecol. News *32*, 1–22.
191. Boomsma, J.J., and Ratnieks, F.L.W. (1996). Paternity in eusocial Hymenoptera. Phil. Trans. R. Soc. Lond. B *351*, 947–975. https://doi.org/10.1098/rstb.1996.0087.
192. Boomsma, J.J., Kronauer, D.J.C., and Pedersen, J.S. (2009). The Evolution of Social Insect Mating Systems. In Organization of Insect Societies: From Genome to Sociocomplexity, J. Gadau and J. Fewell, eds. (Harvard University Press), pp. 3–25. https://doi.org/10.2307/j.ctv228vr0t.5.
193. Peeters, C. (2012). Convergent evolution of wingless reproductives across all subfamilies of ants, and sporadic loss of winged queens (Hymenoptera: Formicidae). Myrmecol. News *16*, 75–91.
194. Greer, J.A., and Moreau, C.S. (2021). Phylogenetic analysis and trait evolution of ant cocoons. Insect Syst. Evol. *53*, 60–77. https://doi.org/10.1163/1876312X-bja10008.
195. Guenard, B.S., Weiser, M.D., Gomez, K., Narula, N., and Economo, E.P. (2017). The Global Ant Biodiversity Informatics (GABI) database: synthesizing data on the geographic distribution of ant species (Hymenoptera: Formicidae). Myrmecol. News *24*, 83–89. https://doi.org/10.25849/myrmecol.news_024:083.
196. Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., Kozlov, A., Podsiadlowski, L., Petersen, M., Lanfear, R., et al. (2017). Evolutionary History of the Hymenoptera. Curr. Biol. *27*, 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027.
197. Jordan, G., and Goldman, N. (2012). The Effects of Alignment Error and Alignment Filtering on the Sitewise Detection of Positive Selection. Mol. Biol. Evol. *29*, 1125–1139. https://doi.org/10.1093/molbev/msr272.
198. Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S.L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol. Biol. Evol. *32*, 1342–1353. https://doi.org/10.1093/molbev/msv022.
199. Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., and Scheffler, K. (2015). RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. Mol. Biol. Evol. *32*, 820–832. https://doi.org/10.1093/molbev/msu400.
200. Touchard, A., Barassé, V., Malgouyre, J.-M., Treilhou, M., Klopp, C., and Bonnafé, E. (2024). The genome of the ant *Tetramorium bicarinatum* reveals a tandem organization of venom peptides genes allowing the prediction of their regulatory and evolutionary profiles. BMC Genomics *25*, 84. https://doi.org/10.1186/s12864-024-10012-y.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Ants, see Table S1A	Collected in the wild	N/A
Chemicals, peptides, and recombinant proteins		
Trametinib DMSO solvate	APExBIO	Cat# A3887
DMSO	Solarbio	Cat# D8371
TRIzol Reagent	Invitrogen	Cat# 15596018CN
TWEEN 20	Servicebio	Cat# GC204002
ISH Fixative Solution (Animal)	Servicebio	Cat# G1113
20×SSC Buffer	Servicebio	Cat# G3015
20×PBS buffer, DEPC Treated	Sangon Biotech	Cat# B540627
Critical commercial assays		
2×EasyTaq PCR SuperMix (+dye)	TransGen Biotech	Cat# AS111-11
MinElute Gel Extraction Kit	Qiagen	Cat# 28604
PrimeScript RT reagent Kit with gDNA Eraser (Perfect Real Time)	Takara	Cat# RR047
TB Green Premix Ex Taq II	Takara	Cat# RR820
HCR Probe Hybridization Buffer	Molecular Instruments	Cat# BPH01920
HCR Probe Wash Buffer	Molecular Instruments	Cat# BPW02121
HCR Amplification Hybridization Buffer	Molecular Instruments	Cat# BAM02221
HCR Amplifier B1-h1	Molecular Instruments	Cat# S023221
HCR Amplifier B1-h2	Molecular Instruments	Cat# S023321
HCR Amplifier B2-h1	Molecular Instruments	Cat# S015221
HCR Amplifier B2-h2	Molecular Instruments	Cat# S015321
HCR Amplifier B4-h1	Molecular Instruments	Cat# S021721
HCR Amplifier B4-h2	Molecular Instruments	Cat# S021821
DNA and RNA extraction		
NucleoSpin® RNA Set for NucleoZOL	Macherey-Nagel	740406.50
Tris-Cl pH 8	ThermoFisher	15567027
Sodium chloride (NaCl)	ThermoFisher	AM9760G
Ethylenediaminetetraacetic acid (EDTA)	ThermoFisher	AM9260G
Sodium Dodecyl Sulfate (SDS)	ThGeyer	A0676-1L
Phenol/TRIS saturated	FisherScientific	10431212
Chloroform	VWR	1024451000
Isoamylalcohol	Merck	1.00979.1000
RNase A	ThermoFisher	12091021
RNAse T1	ThermoFisher	EN0541
Isopropanol	ThGeyer	45629.01
Ethanol absolute	VWR	20821.310
Experion RNA Analysis Kits	Bio-Rad	700-7103/7104
Qubit RNA quantification assays	ThermoFisher	Q32852/Q10210
Qubit DNA quantification assays	ThermoFisher	Q33230/Q33265
Tapestation Genomic DNA Screen Tape	Agilent	5067-5366/5365

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BGIseq Hi-C Protocol		
Listed below if not mentioned above		N/A
Formaldehyde	VWR	28908
Glycine	Sigma	227.77
Phosphate-buffered saline (PBS)	ThermoFisher	10010031
Igepal CA630	Sigma	I8896
cOmplete Protease Inhibitor	Sigma	5892988001
Triton X-100	Sigma	T8787
NEB buffer 2	BioNordika	B7002
MboI restriction enzyme	BioNordika	R0147M
biotin-14-dCTP	LifeTechnologies	19518018
dATP	Roche	S06416
dTTP	Roche	S03110
dGTP	Roche	S01702
DNA polymerase I, Large (Klenow) fragment	BioNordika	M0210
NEB T4 ligase buffer	BioNordika	B0202
Bovine serum albumin (BSA)	ThermoFisher	R0971
T4 ligase	BioNordika	M0202
Proteinase K	ThermoFisher	EO0491
Linear polyacrylamide	FisherScientific	J67830
AMPure XP beads Beckman Coulter	Ramcon	A63881
Dynabeads MyOne Streptavidin T1 beads	ThermoFisher	65602
Tween 20	ThGeyer	P1379
dNTP mix	ThermoFisher	R1122
NEB T4 PNK	BioNordika	M0201
NEB T4 DNA polymerase I	BioNordika	M0203
NEB Quick Ligation Buffer	BioNordika	B2200
DNA Quick Ligase	BioNordika	M2200
KAPA HotStart ReadyMix PCR	Roche	07958927001
Deposited data		
Ant genome assemblies	This study; NCBI GenBank and SRA	PRJNA1172379; Table S1A
RNA sequencing data	This study; NCBI SRA database	PRJNA1172379 Table S1H
Publicly available genomes	NCBI GenBank	Described in Table S1A
Publicly available RNA-seq	NCBI SRA database	Described in Table S1G
Experimental Models: Organisms/Strains		
Monomorium pharaonis (HCR, Pharmacological manipulation)	Lab-reared from field collected colonies	N/A
Carebara diversa (HCR)	Lab-reared from field collected colonies	N/A
Messor barbarus (HCR)	Lab-reared from field collected colonies	N/A
Oligonucleotides		
Primers and HCR probe sequences		See Table S8
Software and algorithms		
Wtdbg2 v2.5	Ruan and Li[97]	https://github.com/ruanjue/wtdbg2
PacBio SMRT-Analysis package	N/A	https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SSPACE-LongRead v1.1	Boetzer and Pirovano[98]	https://github.com/Runsheng/sspace_longread
LR_Gapcloser	Xu et al.[99]	https://github.com/CAFS-bioinformatics/LR_Gapcloser
NextPolish v1.3.0	Hu et al.[100]	https://github.com/Nextomics/NextPolish
SLR-superscaffolder pipeline	Guo et al.[101]	https://github.com/BGI-Qingdao/SLR-superscaffolder
Funannotate pipeline v1.8.3	Jon Palmer	https://github.com/nextgenusfs/funannotate
MaSuRCA v3.3.0	Zimin et al.[102]	https://github.com/alekseyzimin/masurca
3D-DNA v180922	Dudchenko et al.[103]	https://github.com/aidenlab/3d-dna/tree/phasing
AllHiC v0.9.8	Zhang et al.[104]	https://github.com/tangerzhang/ALLHiC
SALSA2 v2.2	Ghurye et al.[105]	https://github.com/marbl/SALSA
JuiceBox v1.11.08	Durand et al.[106]	https://aidenlab.org/juicebox/
barrnap	N/A	https://github.com/tseemann/barrnap
bedtools v2.28.0	Quinlan and Hall[107]	https://bedtools.readthedocs.io/en/latest/
minimap2 v2.17r941	Li[108]	https://github.com/lh3/minimap2/releases/tag/v2.17
samtools v1.9	Li et al.[109]	https://www.htslib.org/download/
BITACORA v1.3	Vizueta et al.[110]	https://github.com/molevol-ub/bitacora/releases/tag/BITACORA-v1.3
BUSCO V5.1.2	Simão et al.[111]	https://busco.ezlab.org/
compleasm v0.2.2	Huang and Li[112]	https://github.com/huangnengCSU/compleasm/releases/tag/v0.2.2
Merqury v1.3	Rhie et al.[113]	https://github.com/marbl/merqury
purge_dups v1.0.1	Guan[114]	https://github.com/dfguan/purge_dups
RepeatMasker v4.1.4	Smit et al.[115]	https://www.repeatmasker.org/
RepeatModeler v2.0.2	Smit and Hubley[116]	https://github.com/Dfam-consortium/RepeatModeler/releases/tag/2.0.2a
TRF v4.10.0	Benson[117]	https://tandem.bu.edu/trf/trf.html
MCHelper v1.6.6.0	Orozco-Arias et al.[118]	https://github.com/GonzalezLab/MCHelper/releases/tag/v1.6.6.0
Teclass v2.1.3c	Abrusán et al.[119]	http://www.compgen.uni-muenster.de/tools/teclass//index.hbi
Blobtools2 v2.2	Challis et al.[120]	https://blobtoolkit.genomehubs.org/blobtools2/
Stflr2supernova	N/A	https://github.com/BGI-Qingdao/stlfr2supernova_pipeline
RepeatProteinMask v4.1.2	Smit et al.[115]	https://www.repeatmasker.org/RepeatMasker/
HAPpy-ABCENTH v1.0	N/A	https://github.com/biorover/HAPpy-ABCENTH
Caper R package v1.0.3	Orme et al.	https://cran.r-project.org/web/packages/caper/index.html
mmseqs2	Steinegger and Söding[121]	https://github.com/soedinglab/MMseqs2
STAR v2.7.2b	Dobin et al.[122]	https://github.com/alexdobin/STAR/releases/tag/2.7.2b
GeMoMa v1.7.1	Keilwagen et al.[123,124]	https://www.jstacs.de/index.php/GeMoMa
Stringtie v2.1.5	Kovaka et al.[125]	https://ccb.jhu.edu/software/stringtie
Augustus v2.1.5	Stanke et al.[126]	https://github.com/Gaius-Augustus/Augustus
BLAST v2.2.31	Altschul et al.[127]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
HMMER v2.3.1	Eddy[128]	https://github.com/EddyRivasLab/hmmer

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
mafft v7.453	Katoh and Standley[129]	https://mafft.cbrc.jp/alignment/software/
InterProScan v5.52-86.0	Jones et al.[130]	https://www.ebi.ac.uk/interpro/search/sequence/
EggNOG-mapper v2.1.6	Huerta-Cepas et al.[131]	https://github.com/eggnogdb/eggnog-mapper/releases/tag/2.1.6
OrthoFinder v.2.5.4	Emms and Kelly[132]	https://github.com/davidemms/OrthoFinder/releases/tag/2.5.4
GOstats R package v2.64	Falcon and Gentleman[133]	http://bioconductor.org/packages/release/bioc/html/GOstats.html
TopGO R package v2.50	Alexa and Rahnenfuhrer[134]	https://bioconductor.org/packages/topGO/
rrvgo R package v1.15.2	Sayols and Elmeligy[135]	https://bioconductor.org/packages/release/bioc/html/rrvgo.html
clusterProfiler R package v4.6.2	Wu et al.[136]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
pathview R package v1.38	Luo and Brouwer[137]	https://bioconductor.org/packages/release/bioc/html/pathview.html
Cactus v2.0.2	Armstrong et al.[37]	https://github.com/ComparativeGenomicsToolkit/cactus/
LastZ v1.03.73	Harris[138]	https://github.com/lastz/lastz
MultiZ v11.2	Blanchette et al.[139]	https://github.com/multiz/multiz
mafTools	Mayakonda et al.[140]	https://github.com/dentearl/mafTools/
iq-tree v2.1.3	Minh et al.[38]	https://github.com/iqtree/iqtree2
ASTRAL v5.7.1	Zhang et al.[141]	https://github.com/smirarab/ASTRAL/releases/tag/v5.7.1
PRANK v170427	Löytynoja[142]	http://wasabiapp.org/software/prank/
genesortR	Mongiardino Koch[143]	https://github.com/mongiardino/genesortR
phyluce pipeline v1.7.0	Faircloth et al.[144]	https://phyluce.readthedocs.io/en/latest/
MCMCtreeR v1.1	Puttick[145]	https://github.com/PuttickMacroevolution/MCMCtreeR
MCMCtree v4.10.0	Yang[146]	https://github.com/abacus-gene/paml
Tracer v1.7.2	Rambaut et al.[147]	https://github.com/beast-dev/tracer/releases
Genome Alignment Tools v1.0	Hiller and Osipova	https://github.com/hillerlab/GenomeAlignmentTools
RepeatFiller v1.0	Osipova et al.[148]	https://github.com/hillerlab/GenomeAlignmentTools
ideogram R package	Weitz[149]	https://eweitz.github.io/ideogram/
AGORA v3.1	Muffato et al.[150]	https://github.com/DyogenIBENS/Agora
PhylDiag	Lucas and Roest Crollius[151]	https://github.com/DyogenIBENS/PhylDiag
SYNPHONI	Robert et a.l.[152]	https://github.com/nsmro/SYNPHONI
WGCNA v1.72	Langflder and Horvath[153]	https://cran.r-project.org/web/packages/WGCNA/index.html
seqtk v1.4	N/A	https://github.com/lh3/seqtk
SOAPnuke v2.1.5	Chen et al.[154]	https://github.com/BGI-flexlab/SOAPnuke
edgeR v4.0.16	Robinson et al.[155]	http://www.bioconductor.org/packages/release/bioc/html/edgeR.html
phytools v1.2-0	Revell[156]	https://cran.r-project.org/web/packages/phytools/index.html
ape package v5.7-1	Paradis and Schliep[157]	https://cran.r-project.org/web/packages/ape/index.html
ggtree v3.6.1	Yu et al.[158]	https://github.com/YuLab-SMU/ggtree

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
treeio v1.22 R package	Yu et al.[159]	https://www.bioconductor.org/packages/release/bioc/html/treeio.html
CAFE v5.0	Mendes et al.[160]	https://github.com/hahnlab/CAFE5
Dragonfly v2022.2	Comet	https://dev.theobjects.com/dragonfly_2022_2_release/contents.html
Zorro	Wu et al.[161]	https://sourceforge.net/projects/probmask/
HyPhy v2.5.38	Kosakovsky Pond et al.[162]	https://github.com/veg/hyphy
HmmCleaner	Di Franco et al.[163]	https://metacpan.org/release/Bio-MUST-Apps-HmmCleaner
Zeiss Scout-and-Scan Control System Reconstructor	N/A	https://www.zeiss.com/microscopy/us/l/campaigns/scout-and-scan.html
Amira	Stalling et al.[164]	https://www.thermofisher.cn/cn/zh/home/electron-microscopy/products/software-em-3d-vis/amira-software.html
Other		
Ultrasonicator	Covaris	M220
Qubit 4.0 fluorometer	ThermoFisher	Cat# Q33238
Experion Automated Electrophoresis System	Bio-Rad	N/A
Nanodrop spectrophotometer	ThermoFisher	ND-1000
4200 TapeStation Automated Electrophoresis	Agilent	G2991BA
T100 Thermal Cycler	Bio-Rad	1861096

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Ant samples

Ant colonies were collected from the field (see method details), and full details of the colony origin, collector and the sequenced samples are given in Table S1A.

METHOD DETAILS

Sample collection

We aimed to collect mature colonies for each species targeted for genome and transcriptome sequencing. In addition to dozens to several thousands of workers (depending on species), such mature colonies ideally contained one or more reproductively active queens, eggs, larvae of different instars, pupae, alate gynes (winged virgin future queens) and males. In most cases, colonies were collected live and brought back to the laboratory for taking RNAlater or liquid nitrogen samples before storage at -80°C. Voucher samples were collected simultaneously and stored in ethanol. Genomic and transcriptomic samples were then transferred to the Section for Ecology and Evolution (E&E), University of Copenhagen, or to the Kunming Institute of Zoology (KIZ), Chinese Academy of Sciences, for DNA and RNA extraction.

Voucher samples were stored (except for a few cases where we could not afford sequestering vouchers because we had very little biomass) at E&E or KIZ while others remained with collectors or were sent to the Science and Technology Graduate University (OIST) 1919-1 Tancha, Onna, Okinawa, Japan 904-0495 for CT scanning.

DNA extraction and sequencing

PacBio and stLFR or WGS sequencing libraries

High molecular weight DNA was extracted for each ant species using a phenol/chloroform phase separation DNA extraction protocol for the E&E samples. Ants were grinded with a plastic pestle in lysis buffer (20 mM Tris-Cl pH 8, 400 mM NaCl, 2 mM EDTA pH 8, 1 % SDS) after which NaCl (5 M) was added (except when using RNAlater material as the extra salt would make the phases shift). The phenol phase and chloroform/isoamylalcohol (24:1) phase wash steps were done separately and repeated twice. RNAse treatment (RNAse A and RNAse T1) was done after the first phenol phase chloroform/isoamylalcohol wash step. Half volumes of isopropanol were used to make long DNA fragments precipitate, which were then washed with 70 % ethanol and eluted in a 10mM Tris buffer. DNA purity, quality and quantity checks were done using a Nanodrop spectrophotometer (ND-1000), a 4200 Agilent TapeStation System or by Pipin Pulse pulsed field gel electrophoresis, and Qubit v2.0 or 4.0. Specific castes and numbers of individuals used for each species, as well as the sequencing technology used, are detailed in Table S1A. For KIZ samples, the specimens were sent to

Novogene for DNA extractions, for which they used mostly pools of workers. The high-quality genomic DNA obtained via both types of extraction methods was used to construct PacBio CLR or HiFi libraries with an average size of 30 Kb using SMRTbell Template Prep Kits according to the manufacturer's instructions, before sequencing on a PacBio Sequel platform at Novogene Co., Ltd, Tianjin, China. HiFi reads were generated using the Circular Consensus Sequencing (CCS) method. In addition, single-tube long fragment read (stLFR) libraries[165] were prepared using MGIEasy stLFR Library Prep Kits before sequencing on a MGIseq platform with read length 100 bp. Accounting for both long and short reads, we generated data with more than 100-fold sequence depth for most species.

Hi-C sequencing

To obtain chromosome-resolved genomes, Hi-C libraries were constructed for 18 ant species following the protocol described by Rao et al.[166] Reaction volumes were adjusted to small tissue amounts and glycogen was substituted with linear polyacrylamide as a precipitation carrier. The Hi-C libraries were sequenced on a BGI DNBseq platform with the 100 bp paired-end method, using BGI adapters and index primers (see Table S8D).

RNA extraction and sequencing

For the KIZ samples, total RNA was extracted using the RNeasy Kit (Qiagen) according to the manufacturer's instructions while the RNA was dissolved in nuclease-free water. For the Copenhagen (E&E) samples, total RNA was extracted using the NucleoSpin RNA Set for NucleoZOL. RNA concentrations were determined using a Qubit 2.0 or 4.0 fluorometer with the RNA HS assay kit (ThermoFisher). Quality of the extracted RNA was assessed with an Agilent 2100 Bioanalyzer or a Bio-Rad Experion™ Automated Electrophoresis System. Stranded RNA-seq libraries were prepared with more than 1 μg of RNA using the MGIEasy RNA Library Prep Kit (BGI). The libraries were then sequenced in paired-end 2x100 nt mode on an MGIseq platform. To eliminate batch effects, RNA-seq experiments for samples from the same species were performed simultaneously. We generated more than 15 Gb of RNA-seq data for each sample.

In addition, we downloaded publicly available RNA-seq data for the species included in our dataset to assist in gene annotation and to conduct caste-biased expression analyses as detailed below. Sample descriptions and SRA accession identifiers are specified in Table S1G.

Genome assembly

We aimed to generate both long-read PacBio and single-tube long fragment read (stLFR) sequencing data for all ant genomes. However, stringent DNA-quality requirements and limited biomass availability were sometimes challenging, precluding the generation of PacBio reads in 20 species for which we could only obtain stLFR short-reads (in one case with 10X genomics). Some species were sequenced by subprojects under the main GAGA umbrella, and here WGS Illumina reads were generated instead of stLFR. The sequencing data used for each species are provided in Table S1A. Depending on the sequencing type, we subsequently used three pipelines for genome assembly:

Genomes generated with PacBio long-reads and stLFR short-reads (n = 103)

The PacBio sequencing reads were cleaned using the PacBio SMRT-Analysis package (smrtlink-release_9.0.0.92188) to remove sequencing adapters and filter reads of low quality and short length (parameters: minSubReadLength:500). The raw stLFR reads were cleaned by filtering the adaptors, PCR duplicates and the barcode IDs assigned in the read names using the stlfr2supernova_pipeline (https://github.com/BGI-Qingdao/stlfr2supernova_pipeline). The clean Pacbio reads were then assembled using Wtdbg2 v2.5 ("-t 16 -x sq -g 300m" for PacBio CLR; or "-t 16 -x ccs -g 300m" for HiFi)[97] and scaffolded using SSPACE-LongRead v1-1.[98] An additional round of gap filling was conducted to eliminate gaps within scaffolds using LR_Gapcloser with the PacBio subreads.[99] To further improve the accuracy of the genome assemblies, we applied two rounds of polishing on each initial assembly. Arrow software was used to map the PacBio reads to the genome assembly and correct small indels and substitutions, obtaining consensus genome sequences.[167,168] Because of the high error rates of PacBio CLR reads, we conducted an additional genome polishing step using the high quality short stLFR reads. The consensus genome sequences were then further polished using NextPolish v1.3.0,[100] and subsequently scaffolded using the barcoding information from stLFR reads with the SLR-superscaffolder pipeline.[101] Putative duplicated scaffolds were identified and filtered using Funannotate "clean" pipeline v1.8.3 (https://github.com/nextgenusfs/funannotate).

Genomes generated with PacBio long-reads and WGS Illumina short-reads (n = 18)

For some genomes with Pacbio long reads and WGS illumina short reads, the assembly procedure was the same as detailed above except for the last step in which the stLFR reads were used for a final scaffolding round. Because the short WGS reads cannot be used for scaffolding, we then used the polished scaffolds directly to identify putative duplications when applying the Funnannotate "clean" pipeline.

Genomes generated with PacBio long-reads only (n = 4)

For three genomes, we only generated Pacbio HiFi long reads using the same assembly procedure as detailed above except for the second round of polishing where we used short reads, given that HiFi reads have higher quality than CLR. For a single genome, *Melissotarsus emeryi*, Pacbio CLR was sequenced but the short reads were contaminated, so we did two rounds of polishing using the long reads because no more sample material was available. We used the polished scaffolds to identify putative duplications by applying the Funnannotate "clean" pipeline.

Genomes generated with stLFR short-reads only (n = 20)

The clean stLFR reads were assembled using MaSuRCA v3.3.0 ("JF_SIZE = 2500000000").[102] The resulting assembly was further scaffolded using the barcoding information from the stLFR reads to assemble contigs with the SLR-superscaffolder pipeline (https://github.com/BGI-Qingdao/SLR-superscaffolder). Putative duplicated scaffolds were then identified and filtered using Funannotate "clean" pipeline.

A single genome, *Acromyrmex octospinosus* (GAGA-0220) was sequenced using 10X short linked-reads and assembled using the Supernova v2.0.1 assembler, which retrieved an assembly of similar quality to the assemblies from stLFR reads (10X Genomics).[169]

Chromosome level assemblies

We generated Hi-C libraries for 18 ant species, which we further used to curate, scaffold and ultimately resolve the expected number of chromosomes in these genomes. First, the genome assemblies and Hi-C data were used as input in the 3D-DNA pipeline v180922[103] to generate chromosome length scaffolds. However, only seven of the Hi-C datasets were successful in linking the scaffolds using 3D-DNA (including the previously published *Monomorium pharaonis* data[29]). Therefore we also used the AllHiC v0.9.8 pipeline[104] and SALSA2 v2.2,[105] which allowed us to improve the scaffolding to chromosome level for six additional genomes, a procedure that also improved two of the above mentioned 3D-DNA based assemblies. The quality of the assemblies was assessed by visualizing the Hi-C contact maps and we manually curated the final chromosome-level assemblies using JuiceBox.[106] Five of the total of 18 Hi-C libraries failed to link the scaffolds, resulting in the 13 chromosome level assemblies detailed in Table S1A.

Publicly available genomes

In addition to our 145 *de novo* generated genomes, we included genome assemblies for 18 ant species that were publicly available by February 2021 in our analyses (Table S1A) provided they had an N50 higher than 500 Kb. We did not reassemble these genomes, but included them in our pipelines for all the steps after genome assembly to minimize putative biases in the downstream comparative analyses that could otherwise have been induced by variation in filtering procedures or the annotation pipeline. In some of the analyses, we also included the genome sequences of eight outgroup species belonging to the Apoidea, including both bees and wasps (Table S1A).

Contamination screening

The final 163 genome assemblies (including the publicly available ant genomes) were screened for contaminating scaffolds, i.e. scaffolds that are likely of bacterial (or human) origin. For this, we compiled three different databases containing: (i) 1,908 complete bacterial genome sequences from PATRIC[170]; (ii) 43 "Chromosome"-level or "Complete Genome"-level insect genome assemblies from NCBI including 4 ant species, that were further filtered by blobtools2 v2.2[120] (using https://blobtoolkit.genomehubs.org/view/Insecta) to remove putative contaminations with bacterial sequences in these reference assemblies; and (iii) three files containing corresponding bacterial or insect (with or without ants) CDS sequences. We next created 2000 bp sliding windows (overlap 500 bp) for each target ant genome and searched the different insect and bacterial databases using mmseqs2 search release_12-113e3,[121] with '–start-sens 1 –sens-steps 2 -s 7 –search-type 3' followed by 'mmseqs convertalis'. Similarly, we searched all sliding windows against a database containing the human genome sequence (release GRCh38). We filtered the single best blast hit through sorting by evalue (-k7,7g) and then by bitscore (-k8,8gr) for each sliding window using 'sort -k1,1 -k7,7g -k8,8gr | sort -u -k1,1 –merge'. Highly conserved rRNA genes in the ant genomes could spuriously be identified as having bacterial origin so we annotated bacterial and eukaryotic rRNAs with barrnap v0.7 (https://github.com/tseemann/barrnap) and identified overlapping sliding windows with bedtools intersect v2.28.0.[107] We then used infoseq (from emboss 6.6.0) with '-nocolumn -delimiter "\t" -auto -only -name -length -pgc' to calculate the GC contents of each sliding window and of the complete scaffolds of each target genome. We finally calculated coverage from long or short read genomic data across each sliding window using minimap2 v2.17r941,[108] samtools v1.9[109] and bedtools v2.28.0.[107] These coverage data were used as further evidence to identify contaminants when coverage deviated substantially from the mean coverage across the rest of the genome.

Scaffolds were flagged as bacterial contaminants if one of the following was true: (i) all sliding windows of a scaffold had better blast hits against the bacterial than the eukaryotic database (based on bitscores); (ii) >50% of the sliding windows had a better prokaryotic hit while the difference between total bacterial and insect BLASTN v2.2.31[127] bitscores was >200; and (iii) >50% of the sliding windows had a better prokaryotic hit while the difference between total bacterial and insect BLASTN bitscores was >0 and the difference between total bacterial and insect BLASTX bitscores was >50 and the scaffold-wide GC content and sequencing coverage did not lie within the 95% confidence interval of the respective distribution of other eukaryotic scaffolds of the assembly. Scaffolds flagged as bacterial were removed from the assembly before further processing and their numbers are specified in Table S1A. Using the same approach, we also screened for contaminations with human genomic sequences. The scripts used in our pipeline are described in https://github.com/dinhe878/GAGA-Metagenome-LGT and the accessions used for compiling the different bacterial and eukaryotic databases are provided in our supplemental information.[17]

Confirmation of species identity and genome quality assessment

Mitochondrial (*CO1* and *CytB*) and autosomal gene markers (*Wingless*, *LwRh*, *AbdA* and *ArgK*) were annotated in the genome assembly after each assembly step using BITACORA,[110] to confirm the species identity of all sequencing data used in each of the

genome assembly steps. In this phase we identified some contamination in the raw data, which led us to reassign the sequencing data to the correct sample and species names (listed in Table S1B). Barcoding of sequences also allowed us to verify the assignment of biological species identifications. For dubious cases where molecular markers could not confirm species identity or gave similar identities with different species in the same genus, we reconfirmed species names with the collectors or we sent specimens to Francisco Hita-Garcia for taxonomic validation (see comments in Table S1B). In cases where species-level identification remained uncertain, we used *cf.* or *sp.* instead of a species name.

Genome assembly quality was evaluated after each assembly step by using contiguity metrics, gene completeness with BUSCO v5.1.2[111] and compleasm v0.2.2[112] using Hymenoptera_odb10, and consensus quality value (QV) and k-mer completeness using Merqury v1.3.[113] Assemblies showing high percentages of duplication based on gene completeness evaluations were further processed using purge_dups v1.0.1[114] to remove haplotigs and contig overlaps (https://github.com/dfguan/purge_dups). The metrics for the final assembly are given in Table S1A.

Repeat annotation

Prior to protein-coding gene annotation, homology-based and *de novo* methods were conducted in combination to identify and mask transposable elements (TEs) in the assembled genomes. To this end, genome sequences were aligned against the Repbase TE library v25.03[171] and the TE protein database using RepeatMasker and RepeatProteinMask v4.1.2.[115] In addition, RepeatModeler v2.0.2[116] was used to build *de novo* repeat libraries, which were subsequently used to annotate repeats using RepeatMasker. TRF v4.10.0[117] was then used to find tandem repeats with parameters: "Match = 2, Mismatch = 7, Delta = 7, PM = 80, PI = 10, Minscore = 50". Finally, we combined all evidence to generate the summary tables and annotation files containing the repeat contents.

For a more focused analysis of TE content in ants, we developed an optimized workflow, again combining *de novo* and homology-based methods, for identifying and annotating TEs. For this, raw repeat RepeatModeler libraries of the 163 ant and 8 outgroup Apoidea genomes were curated with MCHelper v1.6.6.0 (in fully automatic mode, -a F).[118] Curated libraries were decontaminated, to remove false-positive repeat models from regular protein-coding genes present in high copy numbers. Retained unclassified repeat models were further classified with RepeatClassifier (included in RepeatModeler v2.0.3)[116] and TEclass v2.1.3c.[119] The 171 curated and decontaminated repeat libraries were then merged with curated Arthropoda repeat models included in dfam3.7 by clustering with mmseqs2 release_14-7e284.[121] The final unified library was then used to annotate TEs in the 171 genomes using RepeatMasker v4.1.4.

Genome annotation

General genome annotations were conducted by combining gene models from several sources using a pipeline optimized for the ant genomes generated by the GAGA project. First, the RNA-seq data were aligned to the reference repeat soft-masked genome assembly using STAR v2.7.2b[122] with default parameters. In addition, we retrieved the publicly available gene annotations from the fruit fly *Drosophila melanogaster*, the red flour beetle *Tribolium castaneum*, the parasitoid wasp *Nasonia vitripennis*, the honeybee *Apis mellifera*, the clonal raider ant *Ooceraea biroi* and the Florida carpenter ant *Camponotus floridanus*.[18,32,172–176] The annotations from these insect species were used to conduct homology-based gene predictions using GeMoMa v1.7.1,[123,124] which also incorporated the RNA-seq evidence for splice site prediction. Second, the independent RNA-seq alignments were merged to create a consensus GTF (gene transfer format) using Stringtie v2.1.5,[125] after which BestORF (Molquest package, Softberry) was used to identify open reading frames (ORFs) in the transcript sequences. Transcripts with incomplete ORFs were filtered out. Third, we randomly selected ~1,000 high–quality genes from GeMoMa prediction to train Augustus v3.2.2.[126] *De novo* gene prediction was then performed using Augustus with the repeat-masked genome, filtering out genes shorter than 150 bp or with incomplete ORFs. Finally, gene annotations from the three sources of evidence were combined by merging them into a single GFF using an in-house perl script, and transposon-related proteins were identified and filtered using a BLASTP v2.2.31[127] search against the Swissprot database and the transposable element protein database from RepeatMasker.

For the 28 species, out of the total 163, for which we did not have RNA-seq data, gene annotation was similar as described above but excluded the Stringtie step as no evidence from RNA transcripts was available. Instead of using RNA-seq in GeMoMa, we added gene annotations from the closest species with available RNA data to allow including RNA-based gene models while minimizing putative bias due to the absence of species-specific RNA-seq data.

Finally, gene annotations were evaluated by using the following metrics: gene number, isoform number, gene N50, isoform N50 and gene set completeness using BUSCO (protein mode). The final metrics of each genome annotation are detailed in Table S1A.

Gene and gene family re-annotation

After finishing the general gene annotations, we thoroughly evaluated the annotation of specific genes and gene families of interest in ant biology. The full list of assessed genes can be found in Table S1E. Specifically, we developed a pipeline to evaluate our initial annotations, annotate previously missed genes, and label putative erroneous annotations to be manually reviewed and fixed if necessary using an Apollo web browser. The pipeline scripts and input sequences can be found in our github repository (https://github.com/schraderL/GAGA).

First, the pipeline uses the protein domains and sequences from each of the gene and gene families detailed in Table S1E to conduct a HMMER v2.3.1 (https://github.com/EddyRivasLab/hmmer) and BLASTP v2.2.31 search, respectively, against all

annotated proteins in each genome. The identified protein sequences were then included in the database of each gene and gene family, creating specific hidden markov models (HMM) using HMMER after aligning the protein sequences with MAFFT v7.453.[129] Next, the protein sequences and their HMM profiles were used as input in BITACORA v1.3,[110] a comprehensive tool for the identification and annotation of genes and gene families in genome assemblies. This bioinformatics tool identifies the proteins from each gene family, and also curates existing annotations or annotates putative missing genes in our initial general annotation.

We used the final gene annotations from BITACORA and the generated alignments comparing the HMM profiles and BLAST-based searches to evaluate aspects of the annotations obtained: (i) The number of genes identified in the similarity searches was compared to the expected number of genes (column F in Table S1E); if numbers do not match, a warning is given to manually validate instances of gene absence in cases where no copy is found, or to check whether there are more copies than expected (putative gene expansion); (ii) The length of the alignments was also compared with the full expected gene length to identify: (a) fragmented gene annotations (shorter length than expected) which are then labeled for manual inspection to confirm whether these represent pseudogenes or incorrect annotations that can be manually fixed, or cases that cannot be fixed due to genome fragmentation (i.e. the scaffold is not contiguous and does not encompass the full gene); (b) putative chimeric genes or incorrect gene structures (longer gene sequence than expected). These cases were also labeled to manually explore whether the inferred gene structure is correct or the gene should terminate earlier than what the program predicted. The cases labeled for manual review occurred in average frequencies of only 2–3 genes per genome, out of the hundreds of genes that we included in the gene re-annotation steps, highlighting the precision and usefulness of our pipeline to validate and correct annotations of genes relevant to ant biology.

With regard to the Odorant (OR) and Gustatory receptors (GR), these genes belong to a superfamily of seven transmembrane domain proteins which represents one of the largest families of ligand-gated ion channels in insects.[18,177,178] These gene families, specially the ORs, are expanded in ants with hundreds of gene members.[18] Most of these genes originated through tandem duplication explaining that they are arranged in clusters, which is a challenge for annotating them when using general pipelines, which tend to produce many incomplete or chimeric gene models. We therefore used a pipeline specifically designed and optimized for the annotation of these gene families in ants, HAPpy-ABCENTH v1.0 (https://github.com/biorover/HAPpy-ABCENTH). The ABCENTH (Annotation Based on Conserved Exons Noticed Through Homology) program is a gene finder specifically devised for multigene families with extremely high sequence divergence but highly conserved exon structure. It is also designed to avoid gene fusion in tandem arrays. ABCENTH works by taking candidate exons from a table of homology search hits and extending the candidate region by finding splice sites that recapitulate the expected exon length and phase.

Odorant receptors (OR)

HMM profiles for each OR exon were built from manually curated OR models in five ant genomes (*Harpegnathos saltator* [NCBI-0009], *Ooceraea biroi* [NCBI-0001], *Camponotus floridanus* [NCBI-0005], *Solenopsis invicta* [NCBI-0002] and *Stigmatomma sp.* -[GAGA-0391]). These profiles were used in ABCENTH to annotate the OR genes in all ant genomes. Next, the OR annotations were converted to GFF3 format and evaluated using homology-based searches. We used insect OR sequences from *Drosophila melanogaster* and the above mentioned ants to identify the OR co-receptor (ORCO), rename the OR genes as ORNNN, and finally classify them as: (i) complete sequences, with the expected exon structure and full length, (ii) pseudogenes with in-frame stop codons, and (iii) partial or fragment annotations with incomplete exon structures, which can be caused by pseudogenization, genome fragmentation or mis-assembly.

Gustatory receptors (GR)

Also the annotation of the gustatory receptors was conducted using the HAPpy-ABCENTH pipeline. However, here we conducted both HAPpy and ABCENTH methods and combined the annotations from both methods to generate the final gene models. First, we used high-quality GR gene models to create a dataset that was used to conduct the GR annotations. This dataset comprised GRs from *Drosophila melanogaster*, *Daphnia pulex*, *Strigamia maritima*, *Ixodes scapularis* and the following ant genomes: *Neoponera goeldii* [GAGA-0063], *Paratrechina longicornis* [GAGA-0198], *Anoplolepis gracilipes* [GAGA-0199], *Monomorium pharaonis* [GAGA-0245], *Stictoponera bicolor* [GAGA-0306], *Tapinoma cf. melanocephalum* [GAGA-0340], *Tetraponera rufonigra* [GAGA-0365], *Carebara altinodus* [GAGA-0378], *Stigmatomma sp.* [GAGA-0391], *Leptanilla sp.* [GAGA-0392], *Myrmecia croslandi* [GAGA-0521], *Dorylus orientalis* [GAGA-0534], *Paraponera clavata* [GAGA-0552], *Probolomyrmex sp.* [GAGA-0580], *Ooceraea biroi* [NCBI-0001] and *Solenopsis invicta* [NCBI-0002]. These sequences were then used as input to generate GR annotations using HAPpy. The ant GR annotations were also used to create the exon HMM profiles used in ABCENTH gene model predictions. Finally, the annotations were combined to generate the most complete models and were then used to identify conserved GR members and to classify genes as complete, pseudogenes, partial or fragment models.

The GFF3 files generated for each gene and gene family were then combined with the general gene annotation files, adding new gene models or replacing the existing annotations with re-annotations based on the curated gene models. We also generated a GFF3 file including only one representative isoform per gene, selected based on the following priority criteria: first, re-annotated gene models were preferred over other existing isoforms for a gene; second, homology-based predictions took precedence over RNA-seq based or *de novo* annotations; third, RNA-seq models were preferred over *de novo* based genes, which were only selected as representative when a gene was identified solely via *de novo* annotation. If a gene had several isoforms from one source we kept the longest isoform as the representative one. Partial or pseudogene gene family copies (e.g. OR and GRs) were excluded from this file.

Functional annotation

We combined several similarity-based searches to conduct the functional annotations of the annotated protein-coding genes across the ant genomes. We first used BLASTP to retrieve the 5 best hits against the following databases: Swissprot, Trembl, KEGG and KOG. In addition, we searched for the specific domain signatures in protein-coding gene sequences using InterProScan v5.52-86.0.[130] EggNOG-mapper v2.1.6[131] was also used to retrieve functional annotations from precomputed orthologous groups and phylogenies from the eggNOG database.[179] Finally, all these annotations were combined to generate an overall functional annotation for each ant genome that included the best significant hits from each database, as well as the Gene ontology (GO) and KEGG (Kyoto Encyclopedia of Genes and Genomes) terms transferred from all significant hits.

Orthology assessment

We used OrthoFinder v2.5.4[132] to conduct the orthology assessment across the annotated proteins from our 163 genomes, using one representative isoform per gene. Our comprehensive phylogeny, inferred using non-coding regions retrieved from the whole-genome alignments of the 163 ant genomes using Cactus v2.0.2[37] (see WGA section below), was used in OrthoFinder to increase accuracy by retrieving orthogroups at each hierarchical taxonomic level. The resulting orthologous groups were then classified into: (i) single copy: orthogroups with one single gene in at least 80% of the species; (ii) multiple copy or gene family: those with one or more copies in at least 80% of the species that are not single copy orthogroups; (iii) clade/species specific: genes that are only present in specific clades or species; and (iv) others. The percentage of species included in these classifications was based on the quality of the genomes (20 of the 144 GAGA genomes, and 15 of the 19 publicly available genomes were based on short reads) and on variation around the average completeness of 96% across all genomes based on compleasm (Table S1A), which allowed for putative missing genes in the orthogroups due to incomplete genomes. We also evaluated different percentages of representativeness. For instance, only 274 orthogroups were present as single copy in all of the 163 species (100%), while 8876 orthogroups were present as single copy in more than 131 species (80%; see Table S5J).

Finally, we also included the annotations from the eight outgroup genomes belonging to the Apoidea (see Table S1A and Figure S3B) in the OrthoFinder analyses. The orthologous groups were also further classified into types and used in the gene family evolution analysis described in the next sections. In addition, we retrieved the consensus functional annotation for each orthogroup by using the functional annotation of each ant genome and counting the number of genes with the same annotations in each orthogroup. We also transferred the GO and KEGG terms that were annotated in at least 1/3 of the genes in an orthogroup. All orthology tables and annotations are provided in our supplemental information.[17]

GO and KEGG functional enrichments

We conducted functional enrichments for lists of candidate genes based on orthogroup IDs, using the functional annotations for each orthogroup as described above. We used both GOstats v2.64[133] and TopGO v2.50[134] R packages to conduct the GO enrichments of Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) terms using Fisher's exact tests. The resulting enrichments were then visualized using rrvgo v2.15.2 in R,[135] as shown in Figure S3C. Enzyme enrichments based on KEGG annotations were conducted using the *enricher* function from the clusterProfiler v4.6.2 R package.[136] Enrichments of KEGG enzymatic pathways were retrieved using *enrichKEGG*, a function also included in clusterProfiler, using the fruit fly *Drosophila melanogaster* (pathways with dme as prefix in the KEGG database) or *Monomorium pharaonis* (mpha prefix in KEGG) as reference organisms. Significantly enriched pathways were visualized using *pathview* v1.38 in R.[137]

Whole genome alignments (WGA)

The 163 genome sequences were aligned using Progressive Cactus v2.0.2,[37] guiding the alignments with the ant phylogeny as described in the supplemental information.[17] In addition, pairwise whole genome alignments were generated with LastZ v1.03.73[138] using the repeat soft-masked assemblies, and *Monomorium pharaonis* as reference. The pairwise alignments were then integrated into a multiple genome alignment following the phylogenetic order using MultiZ v11.2[139] for comparative purposes. The 163-way MultiZ and Cactus alignments were compared (Figure S1D) using mafCoverage in the mafTools package[140] after converting the Cactus WGA from HAL to MAF format.[180]

Phylogenomic analyses

Cactus whole genome alignment phylogeny

We extracted intergenic regions from the Cactus-based WGA alignment, retrieving 10710 contiguous regions that were longer than 2 Kb, after filtering coding sequences, repeats and ultraconserved elements (UCEs) and after dropping sequences from the alignment that consisted of more than 25% gaps. From the loci retained, we selected 920 regions that were present in at least 161 of the 163 ant species (99% completeness, 6695204 bp, 5133116 parsimony-informative sites). Our main tree, presented in the main Figures (Figures 1, 3A, and S1A), was derived from the concatenated matrix with maximum likelihood using iq-tree v2.1.3[38] under selection of the appropriate substitution model using ModelFinder and 1000 ultrafast bootstrap replicates.

In addition, we conducted phylogenetic analyses using alternative methods and datasets that are presented in Figure S1B heatmap, and we provide all obtained trees in the supplemental information.[17] We also used all 10710 intergenic regions to infer gene trees (using maximum likelihood in iq-tree as above) and reconciled these gene trees with the species tree using ASTRAL v5.7.1.[141]

Phylogeny of orthologous genes

We retrieved 9063 single copy orthologs (see orthology assessment above) that were present in at least 75% of the species and aligned the protein and cds sequences using, respectively, MAFFT v7.453[129] and PRANK v170427.[142] We then used iq-tree to infer a gene tree for each ortholog as detailed above, and selected a subset of 1000 genes using genesortR,[143] which takes into account four sources of systematic bias (average pairwise patristic distance, compositional heterogeneity, level of saturation, and root-to-tip variance), two proxies for phylogenetic signal (Robinson-Foulds similarity to a target topology and average bootstrap support) as well as the proportion of variable sites.[143] The 1000 selected gene codon alignments were then concatenated after which we used a maximum likelihood approach to reconstruct the orthologous gene-based phylogeny. We also used the 9063 ortholog gene trees in ASTRAL to reconstruct the species tree (see Figure S1B). Additional trees using the protein alignments and alternative matrix occupancies are provided in the supplemental information.[17]

BUSCO Hymenoptera conserved genes phylogeny

We used BUSCO v5.1.2 and the Hymenoptera odb10 dataset[111] to annotate the 5991 genes that are conserved across the Hymenoptera (in genome mode). We aligned the protein sequences using MAFFT and inferred the maximum likelihood genes trees. We then retrieved 1000 genes using genesortR (see above) which we concatenated in a matrix before using iq-tree to retrieve the phylogeny represented in Figure S1B heatmap. We also conducted additional matrix and ASTRAL analyses, not shown here because of lower support or redundancy, but these are provided in the supplemental information[17]

Ultra conserved elements (UCEs) phylogeny

We used the phyluce pipeline v1.7.0[144] to annotate UCEs in our ant genome assemblies from hym-v2,[181] which comprised 2590 targeted hymenopteran loci. The UCE loci were aligned using MAFFT, after which a total of 1123 loci present in 90% of the species were concatenated and used for a maximum likelihood phylogenetic analysis in iq-tree. Additional analyses of different matrix occupancies were congruent overall and are provided in the supplemental information.[17]

Phylogeny dating

Based on a well-established literature on geological events and fossil records for ants, we identified 39 fossils that could be assigned to crown groups of clades in our phylogeny. Two of these fossil records calibrated the same node in the phylogeny, resulting in 38 fossil calibrations (Table S2A). We assigned the age of each calibration based on the deposit and then used the Hedman (2010) method,[182] as implemented in the R script by Friedman et al.,[183] to estimate the fossil prior distribution using fossil records from Formicidae outgroup species (see Table S2C for deposit ages, and Table S2B for the Hedman inferences, commands and calibrations used). The Hedman inferences were modeled as skew-t distributions and each fossil calibration was inserted into the input tree using the *estimateSkewT* function implemented in the MCMCtreeR v1.1 package (https://github.com/PuttickMacroevolution/MCMCtreeR), since we used MCMCtree program v4.10.0[146] from the PAML package for dating. MCMCtree requires a root age constraint, for which we used a uniform prior (105-170 Mya) to reflect the range of age estimates of crown Formicidae in recent studies.[184–187] The input alignment used in the divergence dating (205,913 bp) was extracted from the intergenic regions retrieved from Cactus WGA, retaining regions present in all species and filtering gaps with a custom perl script. The REV nucleotide substitution model and the independent-rates clock model were then employed for molecular dating, implemented in the MCMCtree program with approximate likelihood calculation. Two independent MCMC runs were performed, each covering 15 million generations and sampling every 500 generations after a burn-in of 5 million iterations. Convergence was assessed in Tracer v1.7.2[147] by comparing posterior parameter estimates across both runs and ensuring that the effective sample sizes (ESS) exceeded 200 with unimodal posterior distributions.

Macro and microsynteny analyses

Whole genome synteny

Pairwise whole genome alignments (WGAs) were constructed using the soft-masked chromosome-level ant genome assemblies. We employed the LastZ program v1.03.73 with parameters (“M=0 K=2200 L=6000 Y=3400 E=30 H=2000 O=400 T=1 Q=HoxD55.q”) tailored to the phylogenetic distances within the ant phylogeny. The Genome Alignment Tools v1.0 package (https://github.com/hillerlab/GenomeAlignmentTools) was subsequently utilized to refine these alignments. The output from LastZ was processed using axtChain to link the alignment blocks. Highly sensitive local pairwise alignments for loci flanked by these alignment blocks were then re-aligned using patchChain.pl, after which RepeatFiller v1.0[148] was used to detect and incorporate new alignments overlapping with repetitive regions. To enhance the specificity of the alignments, paralogous and random local alignments that overlapped with orthologous loci were removed using chainCleaner. Syntenic nets were then generated by chainNet to hierarchically collect alignment chains. To visualize whole-genome synteny, we filtered the syntenic nets to include only those greater than 500 kb. The ideogram R package v0.2.2[149] was then used to generate macrosynteny plots as shown in Figures 2A and S2A.

Genome rearrangement comparisons

We used AGORA v3.1[150] to estimate the content and order of genes in all ancestral nodes across our ant phylogeny, and to reconstruct large Contiguous Ancestral Regions (CARs). This tool computed highly contiguous, near-exhaustive reconstructions at every bifurcation of the ant phylogeny, providing evidence of genome rearrangements by comparing consecutive genomes across all nodes of the tree. To this end, we used the orthology assessment across ants (as described above) and retained the 135 high-quality

genomes with scaffold N50 above 1Mb. We then identified syntenic gene blocks between successive ancestral genomes in internal branches, and between ancestral genomes and their extant descendants in terminal branches.

Syntenic blocks were identified using PhylDiag[151] between all pairs of successive genomes in the ant phylogeny (either ancestor-ancestor in internal branches, or ancestor-extant in terminal branches). We then characterized rearrangement breakpoints located at the edges of synteny blocks. Breakpoints located at the extremes of scaffolds or chromosomes, or within small contiguous ancestral regions (containing less than 10 genes), could be false positives and were therefore identified and filtered using a custom Perl script. The breakpoint rate at each node of the phylogeny was then calculated by dividing the number of breakpoints by branch length in million years.

To investigate the relationship between rearrangement breakpoint rates and species richness, we performed Phylogenetic Generalized Least Squares (PGLS) analyses to examine the phylogenetically corrected correlation between the number of species in each subfamily and their mean breakpoint rate. We found a significant positive correlation between breakpoint rates and species richness (Figure 2B). To examine whether this pattern was also present at lower taxonomic levels, we also compared the breakpoint rates across different ant genera. To reduce the potential impact of different sample sizes across genera, we filtered out species with branch lengths shorter than 5 Ma. We also found a positive correlation across all 69 examined genera between species richness and the average breakpoint rates (Figure S2C). Finally, we identified breakpoint rates across 26 bee genomes[96] with comparable quality using the same methods and compared the breakpoint rates in ants and bees with those observed in several vertebrate clades (Figure S2D).[150]

Inference of microsynteny blocks across ants

We used the ant orthogroups in SYNPHONI[152] to identify clusters of genes that are syntenically conserved across the ants. Specifically, we considered microsyntenic blocks when the number of genes included in such a cluster comprised at least 3 ortholog groups present in more than 131 species (accounting for 80% of the total species), with the additional requirement that these ortholog groups were not separated by more than one intervening gene. The order of the orthogroups within a microsynteny block could be variable, i.e. co-linearity was not required due to high divergence times having induced genome rearrangements across all analyzed genomes so that gene expansions or translocations within microsyntenic blocks are allowed. An example of a conserved microsyntenic block is given in Figure S2G.

Weighted Gene Correlation Network Analysis (WGCNA)

The "WGCNA" R package v1.72[153] was applied to identify modules of co-expressed genes associated with gyne and worker castes in developmental transcriptomes of *Monomorium pharaonis* (source: Qiu et al.[20]). Non-expressed genes were removed, and the expression data were $\log_2(n+1)$ transformed prior to WGCNA analysis. For each developmental stage, the average TPM for each caste and gene was calculated, resulting in a TPM matrix with two castes and six developmental stages. To distinguish modules with different expression patterns throughout development, a soft-thresholding power of 18 was chosen, assuming a signed network, as recommended by WGCNA for less than 20 samples. Co-expression networks were constructed based on the similarity of gene expression patterns, allowing these genes to be grouped into modules. The minimal gene module size was set at 30 and the threshold to merge similar modules was set at 0.25. The relationships between each module and caste-biased expression (gyne or worker bias) were calculated using Pearson correlations between module eigengenes and castes. Correlation p-values were calculated with one-tailed tests using the corPvalueStudent function in R. Two out of 15 modules showed consistently biased expression in either the gyne or worker caste ($p \leq 0.05$), which we interpreted to imply that entire caste-biased modules can be associated with either gyne or worker development (Figure S2F).

We tested for enrichment of conserved syntenic genes (as described above) in each module applying hypergeometric tests with false discovery rate (FDR) correction, using *M. pharaonis* gene IDs. Gene ontology over-representation analysis was performed using the enrichGO function of clusterProfiler for the conserved syntenic genes in *M. pharaonis*, covering modules 1 to 4, with the following parameters: pvalueCutoff = 0.05, pAdjustMethod = "BH" (Benjamini and Hochberg), and qvalueCutoff = 0.05.

Caste transcriptomes and differential gene expression

Adult caste transcriptomes

We selected adult caste transcriptomes from whole-body samples sequenced by GAGA or publicly available (see Tables S1G and S1H) across ant species belonging to seven ant subfamilies. Since the RNA sequencing data for some species lacked biological replicates, we generated three pseudo-replicates for species with only a single sample per caste using seqtk v1.4 (https://github.com/lh3/seqtk) in order to improve the robustness of downstream statistical analyses. Quality control of the raw sequencing data was performed with SOAPnuke v2.1.5,[154] in which the adaptor sequences were trimmed and the low-quality reads were removed. The filtered reads for each species were mapped to their annotated genomes using STAR v2.7.2b[122] and the expression abundance was quantified with stringtie v2.1.5.[188] The gene counts and TPMs from stringtie were built into gene abundance matrices for each species. Based on the ant orthogroups (described in the Orthology assessment section above), we generated an abundance matrix for all orthologous groups where the abundance of orthologous genes in the same species was added up by our in-house scripts. This orthogroup count matrix was then used as input for the differential expression (DE) analyses, while the TPM matrix was used for generating gene expression plots (see Figures 3 and 5).

Differential expression analyses were conducted separately for each species using edgeR v4.0.16.[155] To identify genes with caste-biased expression between gynes and workers, we analyzed 70 species for which we had transcriptome data from both castes.

Low-count genes were filtered based on the count-per-million and sample library size using the default setting in edgeR. To adjust for potential skew due to differences in gene expression that are affected by overall library size, we normalized for library size.[189] More specifically, the genes with caste-biased expression were identified using the general linear model:

$$\text{Gene count} \sim \text{Colony} + \text{Caste},$$

where Colony was not considered when samples were collected from the same colony, and workers were treated as the baseline caste. Thus, genes showing a positive logarithm of fold change (logFC) had upregulated expression in gynes while negative scores meant upregulation in workers. Genes with an absolute logFC greater than 1 and a Benjamini-Hochberg adjusted $p < 0.05$ were identified as caste-biased genes. Similarly, we performed differential expression analyses between major worker and minor worker castes for 24 species with worker polymorphism to identify DEGs associated with differentiation within the somatic worker caste.

For across species comparisons, we categorized the orthogroups based on the number of caste-biased genes across all ants, the formicoid ants, the poneroid ants and each ant subfamily by combining the species-specific DEG. Orthologous groups with more than 75% of the species having differentially expressed genes towards the same caste were categorized with a "75pc" label. In addition, we defined "50pc_strict" orthogroups as those where more than 50% of the species had differentially expressed genes and the difference between caste-specific percentages (gyne/worker or minor-worker/large-worker or soldier) was higher than 50%; orthogroups meeting only the first requirement of having the same caste bias in more than 50% of the species were classified as "50pc"; finally, we labeled as "chisq" orthogroups that were significantly enriched in caste bias after chi-squared tests (Bonferroni corrected p-value < 0.05). Details are provided in the Tables S1, S2, S3, S4, S5, S6, S7, and S8 and our supplemental information.[17]

Caste developmental transcriptomes

We used the ant developmental transcriptomes described in Qiu et al.[20] to detect DEGs between gynes and workers in *Monomorium pharaonis*, as well as between major and minor workers in *Acromyrmex echinatior* across ontogenetic development of these two myrmicine ants. We then assigned the correspondence between our GAGA annotations and the gene sets in Qiu et al.[20] based on reciprocal best hit analyses with BLASTP (E-value < 1e-5) (see Qiu et al.[20] for the specific details). The individual gene expression plots and DEG tables are provided in our supplemental information.[17]

Phenotypic trait data collection

We collected data for social and ecological traits across the 163 ant species for which we had genomic data and organized them into categorizations as described below and in Table S3A. We started by retrieving data from experts in ant biology across the globe, which we supplemented with data from comparative studies, major reviews, or books (see highlighted column headings in blue in Table S3A). We then conducted a literature search for specific species and traits (including Antwiki), and contacted the collectors for validation and for providing missing data in our table where possible (referred to as "Collector" in Table S3A references when the information was provided by a collector's personal communication). References to "Expert Opinion" indicate that information was retrieved via a survey of experts in ant biology coordinated by Nathan Sanders and Rob Dunn, including a discussion hosted by Evan Economo in early 2020 that a number of ant biologists attended, including Rob Dunn, Benoit Guenard, Heloise Gibb, Brian Fisher, David Lubertazzi, Catherine Parr, Nathan Sanders, Steve Shattuck and Jonathan Shik. Data were not always complete, but Table S3A likely represents the best possible state of the art in spite of a number of missing values. The following categorizations were compiled:

- *Colony size and structure*. We retrieved the average or maximum number of workers per colony where possible, or inferred the approximate number of workers per colony based on the closest related species or used genus-specific estimates when these were otherwise consistent across species belonging to the same genus. Colony size average and maximum estimates were then $\log_{10}$ transformed so colony size could be analyzed as a continuous variable, or be grouped in discrete categories: small [<100 workers], medium [100-1000 workers], large [>1000 workers] or unicolonial, i.e. much larger but without discrete nest boundaries). Unicoloniality information was mostly retrieved from Helanterä 2022.[190] We believe that these data are accurate because log-transformation made them robust against unavoidable errors of estimation and inference. Using log-transformed data continues a decade-long tradition in ant research,[21,191,192] so that comparisons with previous studies were unambiguous. We further categorized key aspects of colony structure by including monodomy (a single nest per colony) or polydomy (multiple interconnected nests per colony), although data for this trait were only available for 31 species.
- *Colony structure and elaborations of queen reproductive function*. We assigned each species into one of the following categorizations of queen reproductive function: (1) Monogynous colonies founded by a single queen after dispersal on the wing (and without secondary gamergates or parthenogenesis), assumed to be the ancestral state[21]; (2) Facultatively polygynous colonies producing winged gynes (and without gamergates or parthenogenesis); (3) Partially obligate polygyny (with some strictly polygynous populations having many queens per nest while other populations remain predominantly monogynous, as for example in some *Solenopsis* and *Formica* species), producing winged queens (no gamergates or parthenogenesis); (4) Obligate polygyny in colonies producing alate queens, although wings are functionally vestigial and colony budding is the predominant way of reproduction (but still without transitions to reduced queen functionality via gamergates or parthenogenesis); (5a) Both regular (winged) and ergatoid/brachypterous queens coexist in the same colony; (5b) Only ergatoid queens are present because the ancestral winged gynes are no longer produced; (6a) Both regular queens and gamergate (inseminated) workers coexist in

the same colony; (6b) Gamergate workers are the sole remaining reproductives, which essentially means that developmentally canalized caste differentiation has ceased to exist as gamergate and worker status are determined only by adult reproductive role differentiation similar to society-forming social insects; it is important to note, however, that gamergates are a reductive secondary development and not a reversal of superorganismal colony structure because gamergates have no possibility to disperse without their worker siblings; (6c) Gamergates that also evolved parthenogenesis and where reproductive modes vary across populations, as for example in *Platythyrea punctata*; (7a) Facultatively parthenogenetic colonies in which morphological and behavioral caste differentiation persists; (7b) Obligate parthenogenetic colonies in which morphological caste differentiation has ceased to exist, as for example in the clonal raider ant, where larval signal triggers behavioral differentiation.

This detailed categorization is based on earlier distinctions between facultative and obligate[21,191] traits and between different kinds of secondary reductions in queen functionality (usually lower fertility),[193] but in some analyses we pooled subcategories into Polygyny, Ergatoid queens, Gamergates, Parthenogenesis or Queenless colonies.

- *Queen mating system*. We started retrieving estimates of the effective mate number for queens based on genetic marker studies analyzed with state-of-the art statistics (reviewed by Boomsma and Ratnieks[191]) but found that such studies were too scattered to give meaningful comparative data. We thus ended up using a more general discrete categorization based on distinguishing between Monoandry (single mating), Obligate polyandry (consistent multiple insemination) and Facultative polyandry (some queens are singly inseminated and others multiply inseminated).[21,191,192] For some species we inferred records from published data about other species in the same genus, which is reasonable because mating systems are usually conserved at the genus or subgenus level (Boomsma and Ratnieks[191]). For both polyandry and polygyny, our discrete categorizations are the same as used by Hughes et al.,[21] Boomsma et al.,[192] and Bell-Roberts et al.[93]
- *Gyne and worker body size and queen-worker caste-dimorphism*. We measured head width and Weber's length from gyne and worker images available in AntWeb or AntWiki using ImageJ (https://www.antwiki.org/wiki/Morphological_Measurements), and we reported averages when several pictures were available for each of the castes. For ant species collected in China such online images were not available so we used individuals from the collected colonies to obtain the same measurements. We then quantified gyne/worker dimorphism for 82 species for which gyne and worker Weber's length measurements (in mm) were available. We also used a discrete categorization of gyne/worker dimorphism based on body size differences (Low, Medium or High) from literature records and collector's or expert communications, which allowed us to obtain more approximate data on caste dimorphism for as many as 150 species.
- *Elaboration of worker reproductive function*. We categorized each species according to the reproductive abilities of their workers. These records were mostly retrieved from Gotoh et al.[7] and included the following variables: (1) Functional ovaries and intact spermatheca present (i.e. workers can mate and produce both haploid and diploid offspring in the absence of queens - this ability defines the gamergates as mentioned above); (2) Functional ovaries present but the spermatheca is vestigial (i.e. workers retain a small spermatheca with a lining of flattened cells but they never mate); (3) Functional ovaries without a spermatheca; (4) Both ovaries and spermatheca have been lost. We also analyzed the presence or absence of functional ovarioles and a spermatheca separately.
- *Worker polymorphism*. Each species was assigned to one of three discrete categories according to the size distribution and number of worker castes: (1) Monomorphic workers (i.e. a single worker caste of uniform body size); (2) Continuous polymorphism (i.e. a single worker caste with a continuous size distribution tending towards bimodality and with potentially allometric relations between some body parts); (3) Discrete worker polymorphism (i.e. the presence of two or more distinct castes, including minor and major workers and/or soldiers). In cases of discrete worker polymorphism, we characterized and analyzed separately the presence of one specialized worker caste, for example when major workers exhibited specialized morphology and/or behavior such as aggressive nest defense, passive nest guarding, or specialized foraging adaptations.
- *Diet*. We used a broad dietary classification including Predator species, Omnivores, Herbivores, Granivores (seed-eaters) and Fungus-growers. In addition, we assessed whether ant species interact with Sternorryncha (mostly aphids and coccids) to obtain honeydew, defined as Trophobiosis, or better known as aphid tending (the term used in the main text). The categorizations used were: No trophobiosis, the ants do not interact with Sternorryncha; Facultative trophobiosis, the ants attend Sternorryncha to some extent and feed on their honeydew but there are also nests without (access to) trophobionts; Obligate trophobiosis, the ants are always found in associations with at least some (and often many) Sternorryncha, retrieving the majority of their nutrients from honeydew.
- *Trophallaxis*. We characterized the absence (ancestral) or presence of trophallaxis (exchange of social fluids), for which we also distinguished between facultative or obligate. For some species, it is known that trophallaxis occurs but unclear whether it is facultative or obligate, so we used two alternate scenarios, one with a conservative approach excluding these ambiguous species, and a second inclusive classification in which these species were classified as having facultative trophallaxis.
- *Foraging stratum*. We collected data on the worker foraging stratum including the following discrete variables: Hypogaeic (workers forage below ground, often within runners or tunnels adjacent to their nests in the soil); Epigaeic (foraging on the soil surface or in the leaf litter); Arboreal (workers forage in low or high vegetation, possibly extending all the way into the forest canopy).

- *Foraging activity and behavior*. Further information about worker foraging activity was collected to obtain discrete categorizations such as: Solitary or tandem foragers; Facultative group foraging (both solitary and along trails); Obligate group foraging *en masse* or via trails. Foraging workers were also characterized as being Diurnal, Nocturnal or Both.
- *Social parasitism*. Forms of social parasitism including Inquilinism, Dulosis (slave making) and Temporary (nest-usurpation) social parasitism were annotated together with the ant host species of social parasites.
- *Sting presence in workers*. We assessed the presence or absence of a sting in the worker caste; many of these cases were inferred from genus-level information, which is justified because of the conserved nature of this trait.
- *Eye presence in workers*. We surveyed the presence of Developed, Regular, Reduced or Absent compound eyes in the worker caste. In addition, we used genus-level inferences collected by Greer and Moreau,[194] specifying the number of ommatidia in worker eyes in four discrete categories: 0 (eyes absent); 1 (between 1 and 10 ommatidia); 2 (11 to 100 ommatidia); and 3 (more than 100 ommatidia).
- *Geographic and environmental data*. We surveyed the GABI database (antmaps.org)[195] for data on species distribution, latitudinal bounds, range area, and long-term annual temperature and precipitation estimates (from WorldClim 2.1). We also used the number of known native or exotic records to categorize ant species as exotic or not, and the Global Invasive Species Database (GISD) to annotate invasive species.

Towards the end of this data collection process we communicated with researchers at Oxford University who had collected a set of phenotypic data for 794 ant species to obtain a global non-genomic database approximately 5 times larger than our own, to make sure that the 163 GAGA species were a subset contained in their larger scale path analyses.[93] This coordination of raw baseline data implies that the GAGA dataset can be considered to be a representative sample of their more encompassing data and thus of global ant biodiversity.

Ancestral state reconstructions

We used ancestral state reconstruction (ASR) to estimate the evolution of social and ecological traits (described in the previous paragraph and in Table S3A) across the ant phylogeny using the main dated tree presented in Figure 1. First, we assessed the phylogenetic signal of each trait categorization by measuring Pagel's lambda (λ) using the *phylosig* function in phytools v1.2-0 R package,[156] and we plotted traits in heatmaps next to the phylogeny when they did not have a phylogenetic signal. For discrete categorizations, ASR was conducted using the *ace* method in the R package ape v5.7-1,[157] applying maximum likelihood estimation which allows missing data (uncertain characters are coded as *NA*). The final specific best fitting model was obtained by comparing the Akaike Information Criterion (AIC) scores for the following predetermined models: equal rates (ER), symmetrical rates (SYM) and all-rates-different (ARD), and for custom biologically-relevant models. The ancestral state likelihoods were visualized as pie charts in the phylogeny together with the heatmapped trait categories for each species, using the R package ggtree v3.6.1.[158] For continuous variables, we used *fastAnc* in phytools to obtain maximum likelihood estimations of ancestral states. Since this method does not allow missing data, we used genus-level estimates or we pruned the phylogeny to exclude tips with missing data using the *drop.tip* function in the R package treeio v1.22.[159] The phylogenies with the trait categories used for each species and the ancestral reconstructions at each node are provided in our supplemental information.[17]

Phylogenetic correlations

We conducted a Phylogenetic Generalized Least Squares (PGLS) analysis using the *pgls* function in the R package caper v1.0.3 (https://CRAN.R-project.org/package=caper) using the dated ant phylogeny of Figure 1. PGLS was used for fitting regression models of correlated evolution between traits, estimating lambda with maximum likelihood to adjust for phylogenetic confounding. We assessed the robustness of significant correlations by using alternative categorizations (e.g. including or excluding data inferred from genus-level information; see Table S3A). In addition, we used PGLS to fit regression models between traits and genomic data, including gene family numbers. To avoid bias due to genome incompleteness or fragmentation in the analyses focused on genome-based features, we excluded the 20 species with lower quality genome assemblies (short-read based with N50 < 500Kb) by using *drop.tip* in the R package treeio to prune the species tree.

Gene family evolution

We inferred expansions and contractions of gene families using Computational Analysis of Family Evolution (CAFE v5.0)[160] across the ant orthogroups including the eight Apoidea outgroups (see Orthology assessment). Specifically, we used the gene count per species for each orthologous group and the time-calibrated species tree.[196] To minimize the impact of variable gene counts caused by potentially incomplete gene annotations from the 20 short-read (stLFR) ant assemblies (N50 < 500Kb), we excluded these species from the analysis (Table S1A). Given the high number of species in our dataset and the long divergence times (up to more than 150 Mya; Figure S3B), our data included several large and highly variable (in terms of gene count) gene families that impeded the convergence of CAFE's maximum likelihood chain for estimating lambda (gene family evolutionary rates). We therefore divided the CAFE analysis in two steps. First, we filtered out gene families for which the difference between minimum and maximum gene count was higher than 20 (319 out of a total of 32,033 orthogroups), and we excluded gene families containing TE-related genes (based on the functional annotations) that were not filtered during general gene annotation. We then used this set of gene families to estimate lambda and

compare two scenarios, one with the same (global) lambda for the whole phylogeny, and another with a two lambda model with different rates for the ants and the Apoidea outgroups accounting the high divergence time and number of species in these two clades (note that more complex scenarios or using multiple gamma rate categories, allowing for among family rate variation, did not converge). Each scenario was run 50 times independently to ensure convergence, and the two-lambda model showed a better overall fit based on a likelihood ratio test. Next, we ran CAFE with the whole set of gene families by fixing the retrieved lambda values (Ants-lambda = 0.00261; Apoidea-lambda = 0.00056) to identify significant expansions and contractions ($p < 0.05$) at each evolutionary node. The input and output from these CAFE analyses, including individual plots for gene families that have undergone significant changes, are included in the supplemental information.[17]

Glomeruli counts and their correlation with the number of OR genes

For assessing the number of glomeruli in the antennal lobes of *Acromyrmex echinatior*, *Neoponera verenae*, *Odontomachus hastatus*, *Platythyrea punctata*, and *Stenamma debile*, we decapitated the specimens and stained entire heads with a 0.5% phosphotungstic acid (PTA) solution for at least 14 days to enhance contrast in the antennal lobes. Each head was then placed into a pipette tip filled with 99% ethanol, serving as a specimen holder inside the scanner. Micro-CT imaging was then conducted using a Zeiss Xradia 510 Versa 3D X-ray microscope, operated with Zeiss Scout-and-Scan Control System software v 11.1.6411.17883. 3D-reconstructions of the resulting scan projection data were prepared with the Zeiss Scout-and-Scan Control System Reconstructor v11.1.6411.17883 and saved in TXM file format. Calibration of image acquisition parameters such as voltage, power and exposure time were performed independently for each specimen to ensure optimal resolution and maximal possible contrast for subsequent analyses.

Glomeruli were manually segmented and counted in Dragonfly v2022.2 (Comet Technologies Canada Inc.) or Amira 5.6.[164] The beginning and end of each glomerulus was determined by scrolling forward and backward through the slices to ensure segmentation of individual glomeruli was captured completely. For densely packed glomeruli, 3D rendering was used to distinguish between individual glomeruli.

The association between OR gene number and glomeruli count was assessed using a PGLS model (see above), which also included a number of species with glomeruli counts obtained from the literature (see Figure 3C and Table S5K).

Selective constraint analyses

General positive selection patterns in the ant phylogeny

The coding genes were aligned using PRANK v170427 codon aligner ("-codon -F" parameters),[142] a method that has a lower false positive rate when analyzing positive selection than codon alignments retrieved from amino acid alignments.[197] The multiple sequence alignment (MSA) quality was further evaluated with Zorro,[161] and orthogroups with average quality below 4 in more than 75% of the unaligned sequence lengths were filtered for the selection analyses. The Genetic Algorithm for Recombination Detection (GARD) was used in HyPhy v2.5.38 to screen the alignments for recombination breakpoints and to split them into separate partitions.[162] The alignments for each partition (or entire orthogroup sequences when no recombination was identified) were further evaluated using HmmCleaner, a segment-filtering software that detects putative errors in MSAs.[163] The identified segments were then masked as gaps, and blocks with gappy regions in more than 50% of the species were removed, similar to sequences with fewer than 15 unmasked codons. A maximum likelihood tree was then reconstructed for each MSA partition using iq-tree.

Next, the adaptive branch-site random effects likelihood (aBSREL) model[198] implemented in the HyPhy package was used to detect hallmarks of positive selection (PS) across all branches in each partitioned-orthogroup gene tree, using the high-confidence codon alignments. P-values were corrected for FDR in all branches tested for each orthogroup, and branches under positive selection were retained only when FDR<0.001 to avoid false positives. The nodes in the gene trees were mapped to the species tree by conducting a gene-tree versus species-tree reconciliation using an in-house perl script. For this analysis, we used the 8,384 single-copy orthogroups with high-quality codon alignments in $\geq$ 80% of the sequenced ants. Each node in a gene tree was assigned to the corresponding node in the species tree that included all species present in a gene-tree clade. To avoid incorrectly assigning internal nodes in the gene tree when their evolutionary history deviated from the species tree (e.g. because of incomplete lineage sorting), reconciliations between gene tree nodes and species tree nodes were retained only when at least 60% of the species descending from the species tree node were represented in the gene tree node. The resulting reconciliation allowed us to retrieve the number of genes under positive selection across all nodes of the species tree, and the proportion of positive selection events among them. The expected ratio of genes under selection across partitions was then estimated against the age of each node (branch length in Mya) using a linear regression model. The $\log_2$ ratios of observed/expected positive selection frequencies were finally mapped on the phylogeny to produce Figures 3A and S3A (Table S5D).

The 231 genes under positive selection in the formicoid ancestor were explored for functional enrichments (see GO and KEGG enrichment text above). In addition, we used the average gyne and worker expression bias across species to plot genes with the same caste bias in >60% of the 71 species for which caste-specific transcriptomes were available, which produced Figure 3E. Differential expression was accounted for as described above and specified for each positively selected gene in Table S5A.

Identifying signatures of positive selection associated with phenotypic traits

We explored 39 categorizations of phenotypic traits expressed at the individual or colony level to uncover whether they had associated signatures of positive selection. For each trait, we defined two states as being either "test" or "reference", where "test" was often the evolutionarily derived state, and "reference" the ancestral state, and we then assigned the species according to their

observed traits (see Table S6A; the list of species used for each trait analysis is provided in our supplemental information).[17] We also identified clades in the phylogeny where derived traits evolved convergently as inferred from the ancestral state reconstructions (see STAR Methods above, and plots available in our repository).[17] Using the above described assessment method for detecting positive selection (aBSREL), we then assigned the total number of species with significant positive selection (FDR<0.05) for each orthogroup, after retrieving genes under positive selection in terminal branches and in all higher taxonomic nodes of the gene tree (i.e. if an internal node that includes genes from species A and B is under positive selection, we assigned both species A and B as having experienced positive selection). We then counted the number of species with "test" and "reference" traits under positive selection, and we assessed the overall frequency of convergent positive selection across clades expressing a focal phenotypic trait. The resulting table distinguished between "test" and "reference" and was used to assess phenotype-associated positive selection in two different ways:

(i) Signatures of convergent positive selection: More than 40% of the clades and species for a particular trait showed positive selection in the terminal species branches and/or internal nodes at higher taxonomic levels, and there was a difference ≧ 50% in the proportion of independently evolved "test" and "reference" clades. For example, if there were 10 independent origins of a trait in the phylogeny, we use these 10 clades to test whether positive selection had occurred convergently using the 50% difference criterion.
(ii) Signatures of enriched positive selection: Specific genes showed signatures of enriched positive selection for a "test" or "reference" trait, but we did not find positive selection across most clades that convergently evolved the focal phenotypic trait. We therefore inferred that signatures of enriched positive selection associated with a phenotypic trait were meaningful when more than 40% of the clades and species assigned as being "test" or "reference" cases showed positive selection and there was also a difference ≧ 50% in the proportion of independently evolved "test" and "reference" species. We also considered enriched positive selection associated with traits for genes that were rarely under positive selection in species with a trait phenotype (i.e. <10% of the species assigned as "test" or "reference"), but enriched in the alternative "reference" or "trait" state, provided more than 40% of the "trait" or "reference" species had experienced positive selection in a particular orthogroup gene tree. The difference between (i) convergent positive selection and (ii) enriched positive selection, is that in (i) we required positive selection to have occurred in the different clades where the trait has evolved independently, whereas for (ii) we only checked enrichment at the species level but not across clades.

Identifying signatures of relaxed or intensified selection associated with phenotypic traits

To assess shifts in the strength of selection on genes associated with phenotypic traits, we used the RELAX model[199] implemented in HyPhy v2.5.38. Specifically, we labeled tips belonging to species expressing a "test" or "reference" trait in the orthogroup gene trees (using separate partitions when recombination was identified as described above) using a custom python script (Note that it makes no difference for the p-values obtained below when test- and reference-state are swapped). Given the high computational costs of running RELAX for all orthogroups across 163 ant species, and the requirement of running it independently for each trait, we selected a subset of 17 most relevant social traits (see Table S6A and the graphic representation in Figure 5A). The RELAX analyses were run using the cleaned and partitioned codon alignments (see above) and their labeled gene trees for each trait. This allowed us to simultaneously determine changes in the intensity of purifying or positive selection when comparing "test" and "reference" branches in the phylogeny. These changes were then captured in the selection intensity parameter (k), which indicates whether and to what extent selection was intensified (k>1) or relaxed (k<1) across the foreground "test" branches relative to the background "reference" for each orthogroup. The reported p-values of likelihood ratio tests were corrected for FDR due to multiple sampling, such that genes with significant intensification or relaxation of selection were always identified with an FDR<0.01 cutoff. A summary table with all genes showing specific signatures of selection associated with the phenotypic traits considered is provided in Table S6B; the full tables for each analysis are available in our supplemental information.[17]

Shared signatures of selection across traits

We also explored shared patterns of intensified and relaxed selection across traits by checking overlapping genes across the lists of candidate genes. Specifically, we used the total number of genes tested in RELAX and the genes with significant intensified and relaxed selection to assess whether a higher number of genes were found across pairwise trait comparisons than expected by chance, using a hypergeometric test (*phyper* function in R) while correcting for multiple comparisons (*p.adjust* with method="bonferroni" in R). The corrected p-values across these pairwise trait comparisons were $\log_{10}$ transformed, capped at 25 to allow better visualization of highly significant comparisons, and arranged in a matrix. The results were then visualized as a heatmap (see Figure 6B) using the *heatmap.2* function in the gplots R package, applying Pearson correlation coefficients as distance metric and average linkage as clustering criterion (arguments "distfun = function(x) as.dist(1 - cor(t(x), method = "pearson"))" and "hclustfun = function(x) hclust(x, method = "average")").

Functional experiments

RNA extraction and RT-qPCR

To determine the tissue specific expression of selected genes (*gcm* and *mAChR-A* in *C. diversa*), total RNA was extracted using TRIzol reagent following the manufacturer's instructions. cDNA was then generated from 1 μg of total RNA using the PrimeScript

RT reagent Kit with gDNA Eraser (Takara). Following a 5-fold dilution with ddH_2O, 2 μL of cDNA was used for RT-qPCR reactions in a total volume of 20 μL with TB Green Premix Ex Taq (Takara), using the following settings: a preincubation step at 95°C for 5 min followed by 40 cycles of 3 step amplification for 20 s at 95°C, 20 s at 55°C, and 20 s at 72°C. Relative transcript abundance was then quantified using the ΔΔCt method and normalized to the abundance of the housekeeping gene *EF1A*. The primer sequences used for RT-qPCR are specified in Table S8.

Hybridization Chain Reaction (HCR)

We investigated the spatial expression pattern of genes under selection via expressed social traits with *in situ* hybridization. The genes under relaxed selection in species with sterile workers include the ovary-related genes *otu*, *Nox*, *Smc3*, *fs(1)Ya*, *Cullin1* and *nanos*, and their localizations were examined in the ovaries of *M. pharaonis*. We also compared the expression of *otu* between ovaries of workers and gynes of *M. barbarus*. The genes under intensified selection, *gcm* and *mAChR-A*, were also examined for their expression in the brains of workers and soldiers of *C.diversa*. To do these tests, ovaries and brains were dissected in PBSTw (PBS, 0.1% Tween20) and fixed in 4% PFA for 20 min. The fixed tissues were then permeabilized in Detergent solution (1% SDS, 0.5% Tween, 50mM pH7.5 Tris-HCl, 1mM pH8.0 EDTA, 150mM NaCl) for 20 min. After washing three times with PBSTw (for 10 min, 5 min, and 5 min), tissues were incubated in 200 μL of probe hybridization buffer for 30 min at 37°C. Subsequently, 1 μL of probe was added to the solution, and the samples were incubated at 37°C overnight. The following steps were carried out according to the standard RNA FISH protocol for *Drosophila* embryos available on the MI website (https://www.molecularinstruments.com/). Image capture was performed using Nikon A1MP+ confocal microscopes.

Pharmacological manipulation

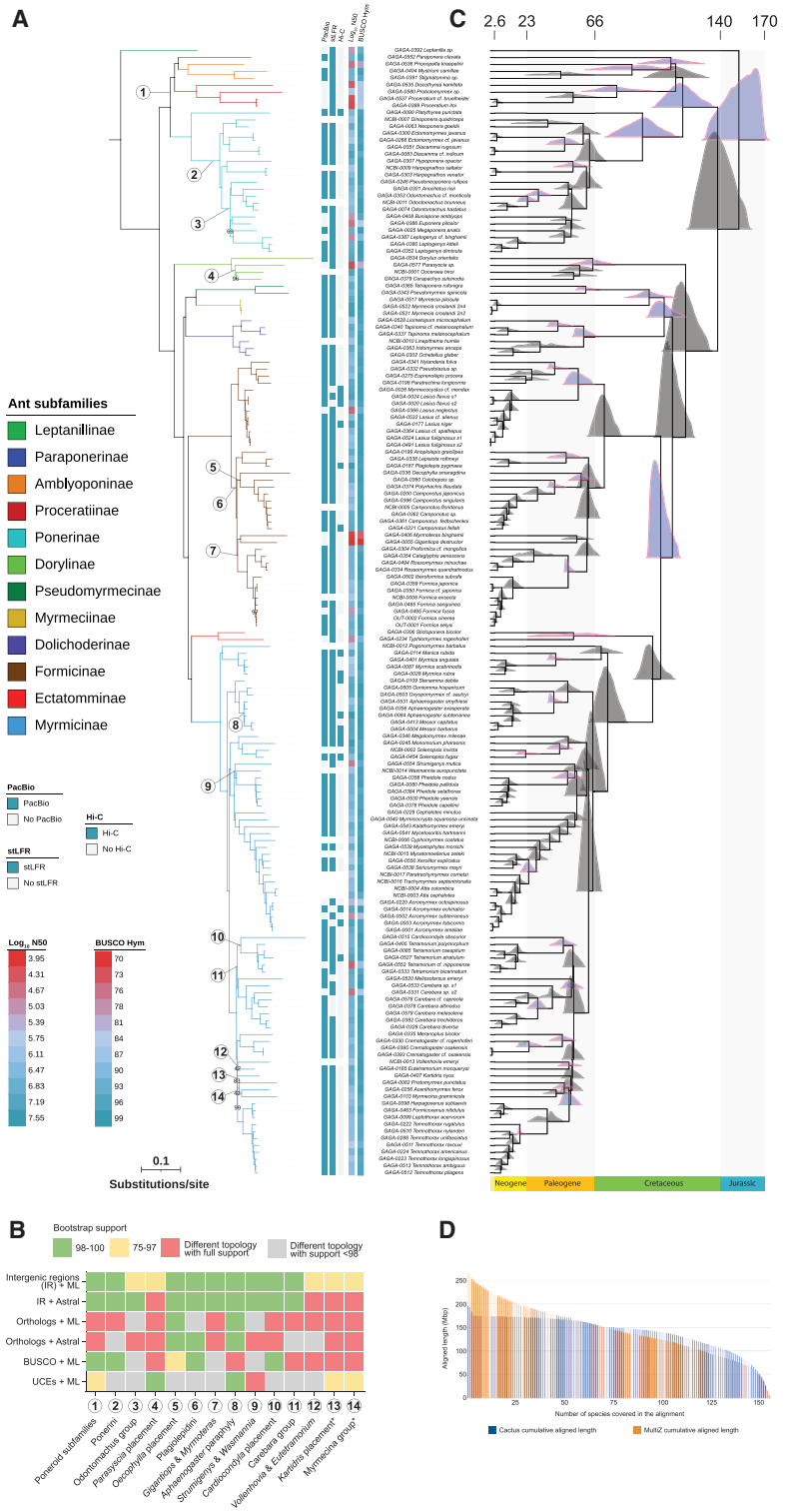
We utilized the inhibitor Trametinib to investigate the significance of *Dsor1*, a core gene in the MAPK signaling pathway, regulating morphological caste differentiation in *M. pharaonis*. Trametinib (10mM; in DMSO) was diluted 1:1000 in nuclease-free water while DMSO was diluted 1:1000 in nuclease-free water to be used as control. Late 3rd instar worker larvae were injected at the cephalothoracic junction. Body length was measured in the young pupal stage by laterally positioning worker pupae on a microscope slide so body length could be measured from the tip of the head to the end of the gaster.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical tests, significance thresholds, and sample sizes are indicated in the figure legends. For all significance tests involving multiple comparisons, FDR or Bonferroni correction was applied. Specific details of the analyses performed in this study are provided in method details.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

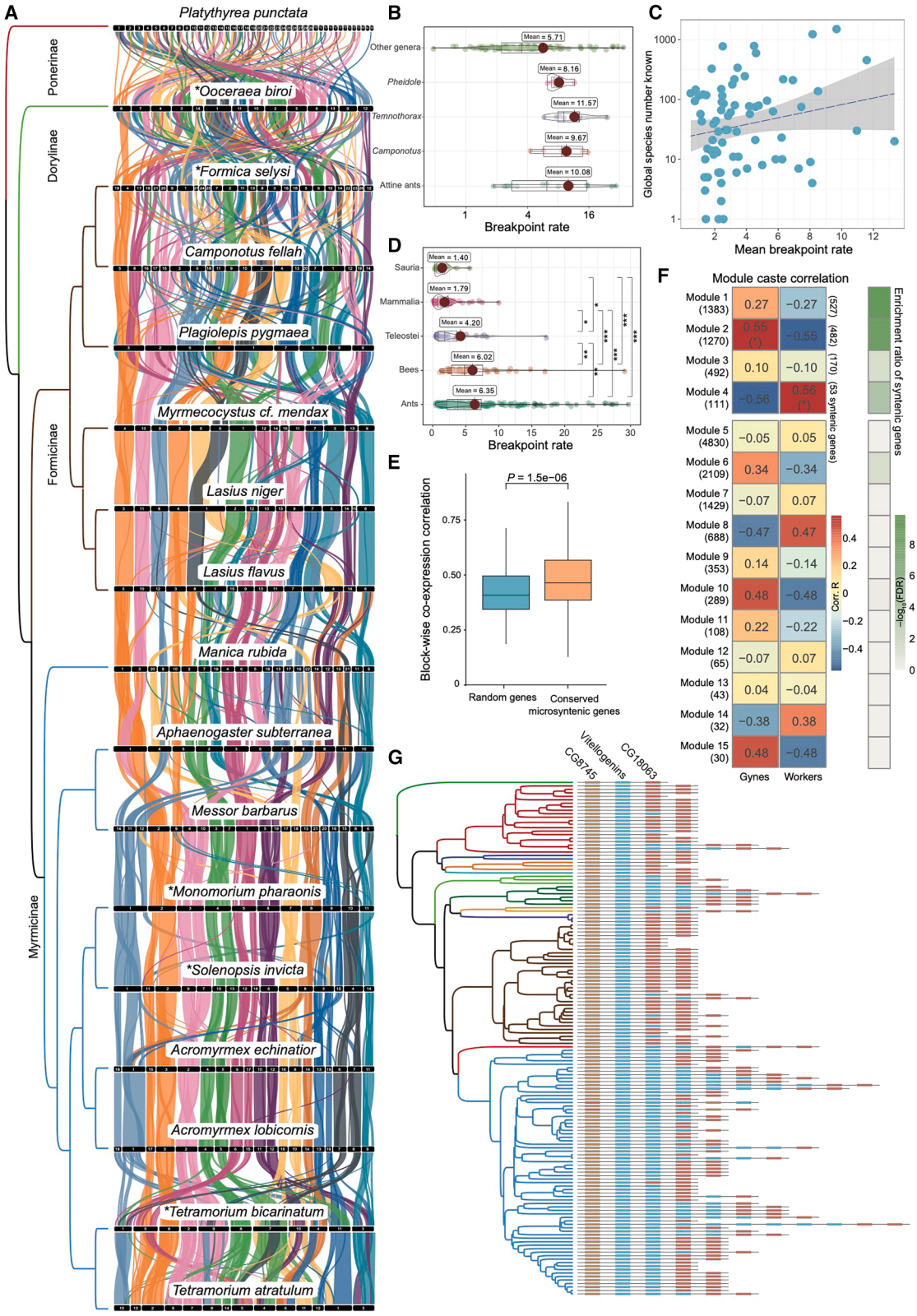
Figure S1. Genome quality, phylogeny reconstruction, and dating, related to Figure 1

(A) The main species tree presented in Figure 1 with branch length as substitutions per site based on maximum likelihood estimation of 920 intergenic regions retrieved from Cactus whole-genome alignment. Branch support from 1,000 ultrafast bootstraps is 100% unless indicated by lower numbers. The right heatmap presents the sequencing data retrieved for each genome assembly, in addition to the assembly scaffold N50 and completeness based on the BUSCO Hymenoptera dataset using Compleasm.[112] Note that the Hi-C data did not improve the assembly scaffolding in some species (see chromosome-level assemblies in Table S1A).

(B) Heatmap showing the support for challenging branches across six selected species trees (all tree files are accessible in the extended dataset[17]), showing that our present approach based on intergenic regions (IRs) + maximum likelihood offered superior resolution. Asterisks mark two cases (13,14) where all four trees in two nodes had different topologies with full support each.

(C) Dated phylogeny with posterior distributions for estimated ages at each node. Distributions in blue represent nodes that were calibrated by fossil records (Table S2).

(D) Cumulative length of the whole-genome alignments based on Cactus (used for the main tree in Figure 1) and MultiZ.



(legend on next page)

Figure S2. Genome-wide synteny in chromosome-level assemblies, breakpoint rate correlations, and syntenically conserved co-expression, related to Figure 2

(A) An extended version of the ribbon diagram of Figure 2A illustrating chromosome-level synteny between 17 ant species across the four subfamilies (Ponerinae, Dorylinae, Formicinae, and Myrmicinae) with haploid chromosome numbers ranging from 9 to 42 and genome sizes varying from 196 to 414 Mb. The five genomes not generated in the present study are marked with *.[18,29,31,33,200] Note that the recently published chromosome-level assembly for *Tetramorium bicarinatum*[200] was not included in our comparative analyses but inserted to illustrate loss of synteny between an inquiline social parasite (*T. atratulum*) and a free-living species of the same genus.

(B) Comparison of rearrangement breakpoints per million years between attine fungus-growing ants, three species-rich genera (*Camponotus*, *Temnothorax*, and *Pheidole*) and all other ant genera included in our study.

(C) Global extant species numbers[8] as function of mean chromosomal breakpoint rate across the 69 ant genera covered in this study (adjusted $r^2 = 0.10$; $p = 0.003$ based on PGLS; 95% confidence intervals are shown in gray, and dots represent the raw data per each genera).

(D) Comparison of rearrangement breakpoint rates between ants, bees, and various vertebrate taxa. *p* values refer to one-way Kruskal-Wallis tests followed by Dunn's multiple-comparison tests after FDR adjustment (*$p < 0.05$, **$p < 0.001$, ***$p < 0.0001$).

(E) The correlated expression of conserved microsyntenic blocks exceeds the correlated expression across same-size blocks of randomized genes in the *M. pharaonis* developmental transcriptomes.[20] *p* value refers to Wilcoxon rank-sum tests.

(F) The symmetric caste-specific deviations in gene co-expression profiles throughout *M. pharaonis* development for all 15 conserved co-expression modules with the number of genes in brackets. Each cell in the heatmap displays symmetrical Pearson correlation coefficients characterizing the extent of combined expression. Significant correlations were observed for module 2 in gynes and for module 4 in workers (*$p < 0.05$; see also Figure 2B). Four developmental co-expression modules are enriched for highly conserved syntenic genes (specified in brackets at the right and also in Figure 2C), as determined by hypergeometric tests with FDR correction (*$p < 0.05$) represented in the green-colored heatmap to the right.

(G) The conserved microsynteny cluster including two vitellogenin genes found to be present in 148 ant species across 12 Formicidae subfamilies (Figure 2E). Branch colors represent the different ant subfamilies and are identical to those used in Figure 1A.

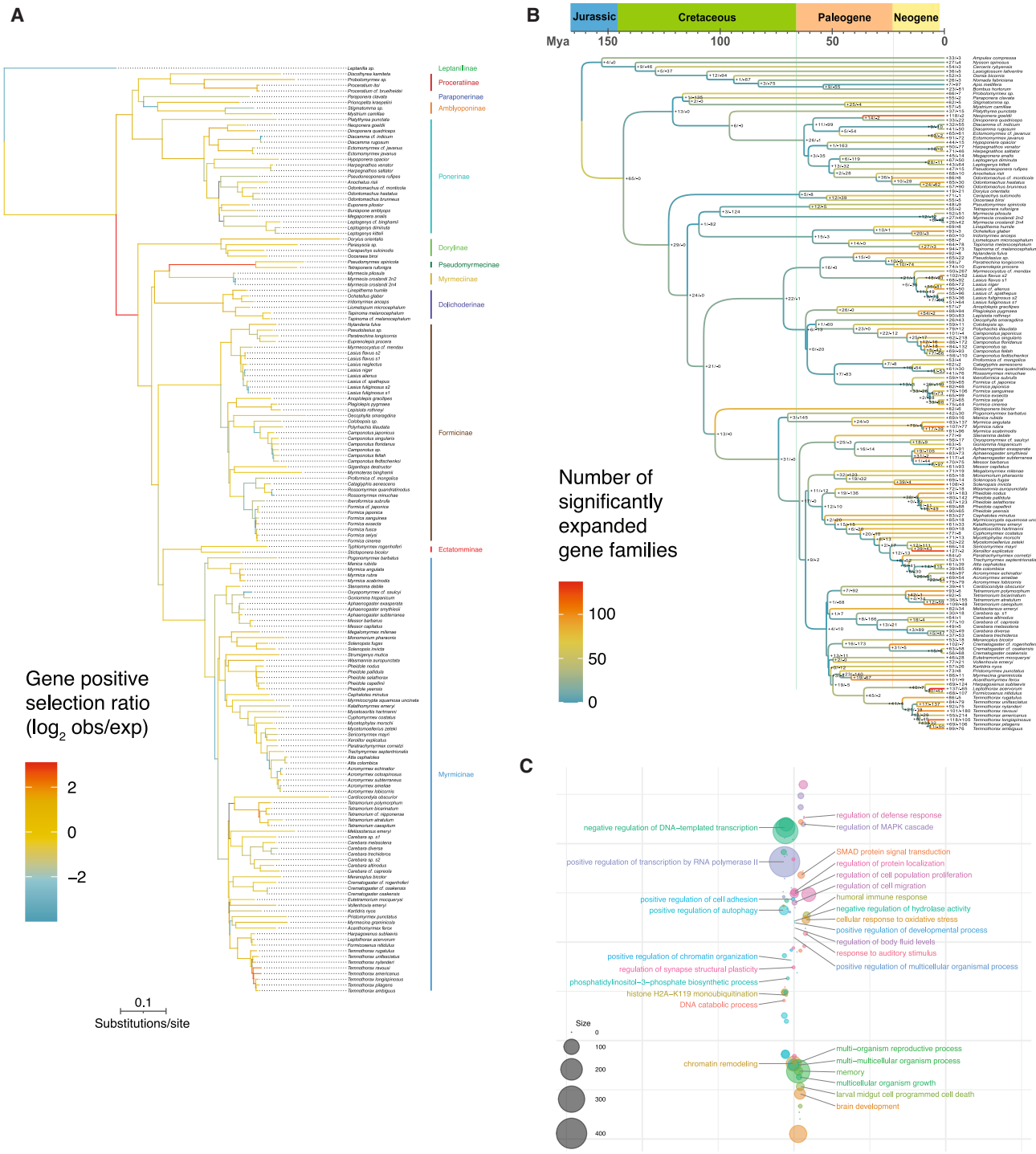
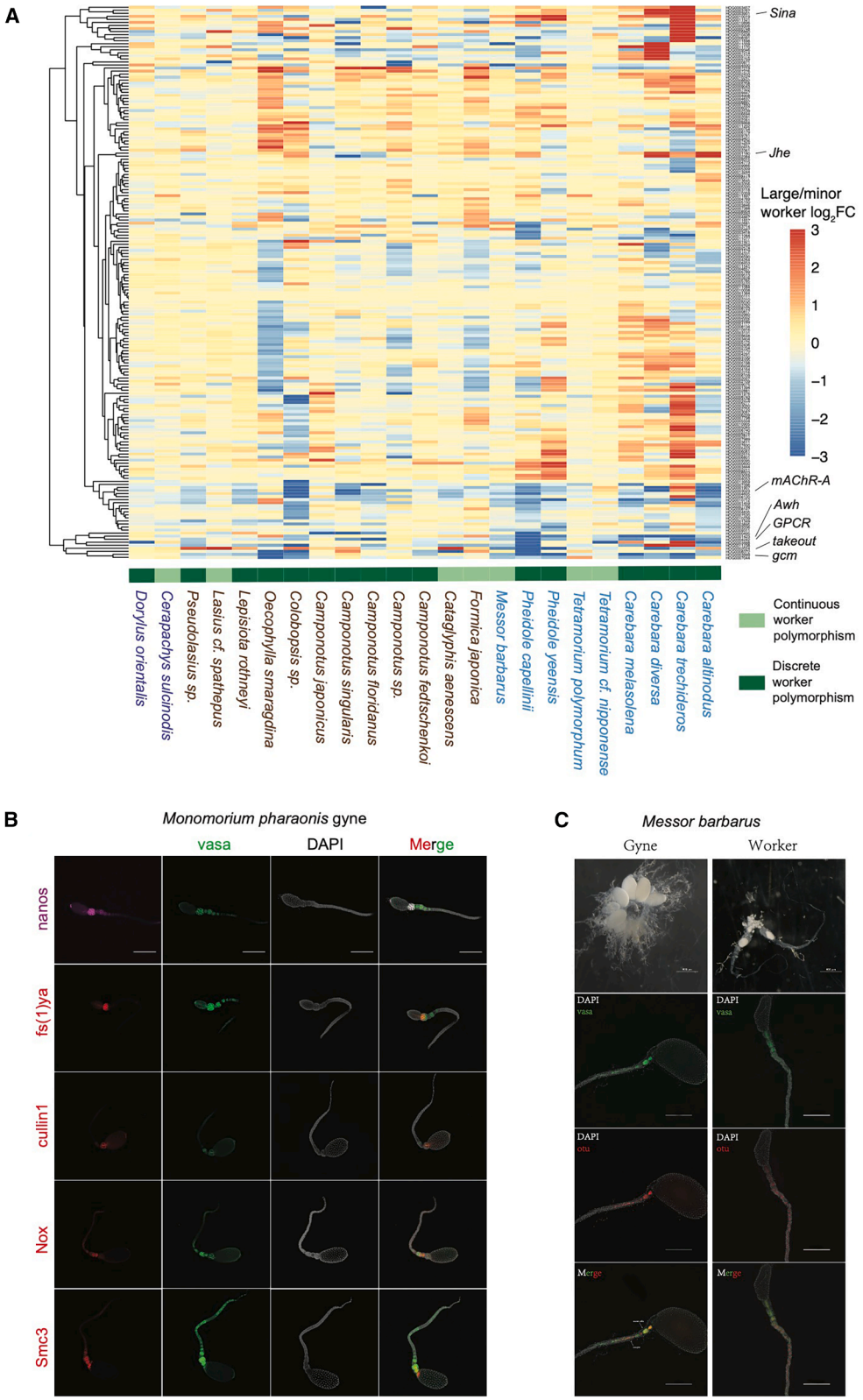


Figure S3. Positive selection and gene family evolution across the ants, related to Figure 3

(A) Main tree showing gene-level positive selection ratios, as in Figure 3A but now with all branches displayed, inferred from the 8,384 single-copy orthogroups with high-quality codon alignments in ≥80% of the sequenced ants.

(B) Calibrated phylogeny including the number of gene family expansions (+) and contractions (–) associated with each node across all orthogroups. Branches are colored according to the number of significantly expanded gene families.

(C) The unedited version of the functional GO-term enrichment diagram of Figure 3D, retrieved with the rrvgo R package, and including all enriched GOs for the 231 genes that showed signatures of positive selection in the ancestor of the formicoid ants. Circle sizes represent the numerical presence of each GO and associated term in the total gene set.



(legend on next page)

Figure S4. Gene-level selection associated with secondary elaborations of the worker caste, related to Figure 5

(A) Expression heatmap for the genes that underwent intensified selection in species with continuous (light green) and discrete (dark green) worker polymorphism, showing the $\log_2$ FC between large and small (minor) workers (color scale toward the right). On the *x* axis, the Dorylinae species have purple text, while the Formicinae and Myrmicinae subfamilies are given in brown and blue text, respectively. The dendrogram toward the left reflects hierarchical clustering of gene expression patterns; the genes highlighted in Figure 5B are also labeled in this heatmap.

(B) Expression intensities of five genes that underwent relaxed selection after workers became completely sterile, corresponding to the ovary drawing in Figure 5C. The images show hybridization chain reaction (HCR) results using *M. pharaonis* gyne ovaries (scale bars, 200 μm, *vasa* is stained to show the germ cells).

(C) Expression intensities of the *otu* gene in *Messor barbarus* where workers have greatly reduced but fully functional ovaries. The first row shows microscopy photos of gyne and worker ovaries, confirming that workers have fewer ovarioles and smaller oocytes. The next three rows are fluorescence microscopy pictures showing the expression locations of *otu* and *vasa* (scale bars, 500 μm for the normal images and 250 μm for the fluorescence images).