Genomic signatures of recent convergent transitions to social life in spiders

The Texas Tech community has made this publication openly available. Please share how this access benefits you. Your story matters to us.

| Citation       | Tong, C., Avilés, L., Rayor, L.S. et al. Genomic signatures of recent convergent transitions to social life in spiders. *Nat Commun* **13**, 6967 (2022). [https://doi.org/10.1038/s41467-022-34446-8](https://doi.org/10.1038/s41467-022-34446-8) |
|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Citable Link   | [https://hdl.handle.net/2346/91043](https://hdl.handle.net/2346/91043)                                                                                                                                 |
| Terms of Use   | [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)                                                                                                                                              |

Title page template design credit to [Harvard DASH](https://harvarddash.github.io/).
Genomic signatures of recent convergent transitions to social life in spiders

Chao Tong \(^1,2\), Leticia Avilés \(^3\), Linda S. Rayor \(^4\), Alexander S. Mikheyev \(^5\) & Timothy A. Linksvayer \(^1,2\)

The transition from solitary to social life is a major phenotypic innovation, but its genetic underpinnings are largely unknown. To identify genomic changes associated with this transition, we compare the genomes of 22 spider species representing eight recent and independent origins of sociality. Hundreds of genes tend to experience shifts in selection during the repeated transition to social life. These genes are associated with several key functions, such as neurogenesis, behavior, and metabolism, and include genes that previously have been implicated in animal social behavior and human behavioral disorders. In addition, social species have elevated genome-wide rates of molecular evolution associated with relaxed selection caused by reduced effective population size. Altogether, our study provides unprecedented insights into the genomic signatures of social evolution and the specific genetic changes that repeatedly underpin the evolution of sociality. Our study also highlights the heretofore unappreciated potential of transcriptomics using ethanol-preserved specimens for comparative genomics and phylotranscriptomics.

The evolution of sociality is a phenotypic innovation that occurred repeatedly and sporadically across vertebrates, insects, spiders, and crustaceans\(^1\). Sociality and social behavior, more generally, are hypothesized to evolve via changes in a set of deeply conserved genes\(^2\)–\(^6\). Many transcriptomic studies have emphasized overlapping sets of genes or molecular pathways underlying the expression of key social phenotypes in different lineages\(^7\)–\(^10\); others emphasize the importance of lineage-specific genes, or genes with lineage-specific expression\(^11\)–\(^13\). Studies searching for patterns of molecular evolution associated with the convergent evolution of sociality have also yielded differing results. For example, a population genomic study of two bee and one wasp species, altogether representing two origins of sociality, found significant overlap in genes experiencing positive selection among all three species\(^14\). In contrast, the largest comparative genomic study of sociality to date, which included ten bee species representing two independent origins and two independent elaborations of eusociality, found low and insignificant overlap for rapidly evolving genes associated with social evolution in each lineage\(^15\). Overall, the degree to which convergent social evolution in diverse lineages involves consistent changes in the same or similar genes remains largely unclear.

One difficulty in identifying genomic signatures of sociality is the relatively few and often ancient origins of this trait\(^16\). Social hymenopteran insects (wasps, bees, and ants), and bees in particular, for instance, are often described as ideal study systems because they include the full range of social complexity and multiple origins of sociality\(^17\). However, there are only an estimated 2–3 independent origins of sociality within the bees, and 7–8 across all hymenopterans, with these origins occurring between 20 and 150 million years ago\(^18\). Thus, other lineages with many independent and more recent origins of sociality can complement studies in social insects and help to further identify common genomic signatures of social evolution.
Like the social insects, spiders exhibit a wide range of social complexity and multiple independent origins of sociality\(^{16-20}\). Sociality evolved independently in spiders an estimated 15–16 times\(^{21,22}\), with each independent origin thought to be relatively recent, at most a few million years ago\(^{23,24}\). Spiders are classified into several categories of social organization, including solitary, subsocial, prolonged subsocial, and social: solitary species live in individual nests, having dispersed from the egg sac soon after hatching\(^{25,26}\); subsocial spiders have nests that contain a single mother and up to a few dozen offspring, which may remain together for several instars before dispersing to initiate their own nests\(^{27,28}\); prolonged subsocial spiders form colonies containing a single mother and multiple cohorts of offspring, which remain in their natal nest until late adolescence or sexual maturity before dispersing to independently found their own colony\(^2\); and social spiders (i.e., non-territorial permanent-social species) form colonies that contain multiple adult females and offspring that remain in the natal nest through maturity, mating with each other to produce new generations that reoccupy the natal nest\(^3\). In social and subsocial species, individuals cooperate in building and maintaining the communal nest, capture prey cooperatively, and share their food. In social species, individuals also exhibit communal brood care. Social spider species have additional distinct features, including colonies that grow through internal recruitment over multiple generations, in some species, reaching sizes of tens of thousands of individuals, female-biased sex ratios, high reproductive skew, and high rates of inbreeding\(^{20,25}\).

Spider sociality has a twiggy phylogenetic distribution, where the closest relatives of most social species are not social\(^{16-20}\). By comparing pairs of social species and their closest non-social (i.e., subsocial) relatives within a single genus, recent studies have begun to identify the genetic consequences of spider sociality and associated shifts from outbreeding to inbreeding\(^{14,15,21-24}\). These studies, together with a preliminary comparative genomic study of two social species from one genus and several solitary species\(^1\), have begun to identify putative genomic signatures of the evolution of sociality in spiders, including elevated genome-wide rates of molecular evolution\(^{26-28}\).

Here, we use comparative genomic approaches to determine whether there are statistically supported common genomic signatures associated with the repeated transition to sociality in spiders (Fig. 1). We use 22 spider species of a range of social systems and representing eight independent origins of sociality\(^{4,13-31}\). We test whether the convergent evolution of sociality across these lineages is associated with (1) convergent genome-wide patterns of molecular evolution; (2) convergent shifts in evolutionary rates for specific genes; and (3) convergent amino acid substitutions in specific genes. We show that genome-wide, gene-wide, and site-specific changes repeatedly occurred during recent convergent transitions to sociality in spiders. Altogether, our study provides unprecedented insights into the genomic signatures of social evolution and the specific genetic changes that repeatedly underpin the evolution of sociality.

### Results

**Spider-omics assembly, annotation, and phylogeny**

We sequenced and assembled transcriptomes of 16 species in the genera *Theridion*, *Anelosimus* (family: Theridiidae), and members of the family *Sparassidae* (Fig. 1, Figs. S1 and S2, Supplementary Table 1). We complemented these new transcriptomes with six existing transcriptomes from closely related species\(^{22,32-38}\) (Supplementary Table 2). In addition, we included two available genomes of social spider species, *Stegodyphus dumicola* and *Stegodyphus mimosarum*. For analyses that required an outgroup (i.e., the RERConverge analysis), we also included a solitary outgroup species, *Acanthoscurria geniculata* (family: Theraphosidae) (Supplementary Table 2). Moreover, we reassembled and improved the subsocial spider *Anelosimus studiosus* genome (GCA_008297655.1) with our newly sequenced *A. studiosus* transcriptome sequencing reads and annotated spider protein datasets (Supplementary Table 2). Finally, we assessed the completeness of the genome or transcriptome assembly using the BUSCO pipeline\(^{39}\) (Fig. 1b), detecting an average of 91.24% arachnid conserved genes (2934 genes) in each spider species (Supplementary Table 3).

---

**Fig. 1** The flowchart represents the analysis pipeline. **a** Sample processing and transcriptome sequencing; **b** Spider transcriptome and genome assembly, annotation, phylotranscriptomic analyses, comparative genomic analyses and evolutionary analyses. Technical details of the workflow are provided in the “Methods” section.
We employed a combination of de novo and homology-based approaches to annotate the gene models of the assembled transcriptomes and improved reassembled genome, and obtained annotated gene models ranging from 19,224 to 96,176 per species (Supplementary Table 3). Then we used a reciprocal hit search strategy with OMA (www.omabrowser.org) and OrthoDB (www.orthodb.org), and identified a total of 7590 orthologous groups (OGs) and 3832 core single-copy orthologs that included all species (Supplementary Table 3).

Based on the concatenated single-copy ortholog data, we reconstructed the phylogeny for our study species (Fig. 2a), which was...
strongly supported (bootstrap value = 100) for all nodes, and was mostly consistent with previously-published phylogenies7,31-33 (Fig. 2a). We estimated divergence times for each phylogenetic tree node (Fig. 2a).

Notably, the 22 spider species included in all of our analyses spanned the full range of spider sociality: eight social species, one prolonged subsocial species, five subsocial species, and eight solitary species. Since we were interested in identifying genomic changes associated with sociality, we compared the eight social species (hereafter labeled as "social") with the 14 species in the remaining categories (hereafter all grouped together as "nonsocial") (Fig. 2a).

Social spiders exhibit convergent genome-wide patterns of accelerated molecular evolution that is caused mainly by the relaxation of selection
As a first test of signatures of convergent molecular evolution, we asked whether social branches had different rates of genome-wide molecular evolution (i.e., dN/dS) compared to nonsocial branches. We estimated the genome-wide ratio of non-synonymous to synonymous substitution (dN/dS) across the spider phylogeny with the concatenated 3,832 single-copy ortholog dataset (Fig. 2b). We found that social spiders experienced higher genome-wide rates of molecular evolution compared to their nonsocial relatives, estimated by both CodeML in PAML 4.7a11 (Wilcoxon rank-sum test, p-value = 0.013 Fig. 2b) and Hy Phy 2.51 (Likelihood Ratio Test, LRT, p-value < 0.00001).

Elevated dN/dS can be caused by increased positive selection, relaxed purifying selection, or a combination of both. RELAX quantify the degree to which shifts in the distribution of dN/dS across individual genes or whole genomes are caused by overall relaxation of selection (i.e., weakening of both purifying selection and positive selection, towards neutrality) versus overall intensification of selection (i.e., strengthening of both purifying selection and positive selection, away from neutrality). Specifically, RELAX models the distribution of three categories of dN/dS—positive selection, neutral evolution, purifying selection—across a phylogeny, comparing foreground (i.e., social) to background (i.e., nonsocial) branches and estimating a parameter K that indicates overall relaxation (K < 1) or intensification (K > 1). We found that the shifts in genome-wide dN/dS between social and nonsocial branches were mainly caused by relaxation of both purifying and positive selection (K = 0.88, LRT, p-value < 0.001) (Fig. 2c).

Hundreds of genes experience convergent shifts in gene-wide rates of molecular evolution in social spiders
Next we asked if there was evidence for convergent molecular evolution at the gene level. First, we applied the relative evolutionary rate (RER) test implemented in RERconverge13 that is specifically designed to test for signatures of convergent molecular evolution underlying convergent phenotypic evolution (Fig. 1b). This test examines the rate of protein sequence evolution for each gene on every branch across the phylogeny, standardized to the distribution of rates across all genes. Genes for which these standardized rates (RER) are consistently either higher or lower in foreground branches (i.e., social branches) compared to background branches (i.e., all remaining branches) are identified as having experienced accelerated or decelerated molecular evolution in social species. We used a phylogenetically restricted permutation strategy, dubbed "permutations", to assess the statistical significance of genes and enriched GO terms. This permutation approach combines phylogenetic simulations and permutations to construct null expectations for p-values, and has been shown to be unbiased and more conservative than other approaches such as permutations alone14-16. Importantly, permutations also correct for non-independence among genes for GO term enrichment, which is not true for the nominal parametric p-values reported from RERconverge14,16,17.

We ran 10,000 permutations and identified genes and enriched GO terms that experienced significant convergent shifts in RERs in social compared to nonsocial branches. The resulting gene and GO term permutation p-values were corrected for multiple comparisons by computing q-values18, and we considered q-values <0.15 (i.e., genes and GO terms with an estimated FDR < 0.15) to be significant. Out of 7590 single-copy orthologs included in RERconverge analysis, we identified 7 genes under convergent acceleration in social branches (permutation q-value <0.15) (Fig. 3a, Supplementary Table 4) and 3 genes experiencing convergent deceleration in social branches (permutation q-value <0.15) (Fig. 3a, Supplementary Table 4). The 22 GO terms significantly associated with convergent acceleration (permutation q-value <0.15) were mainly related to neural function and programmed cell death (Fig. 3b), such as negative regulation of neuron death (GO:01901215), negative regulation of secretion (GO:00501048), and positive regulation of programmed cell death (GO:0043068) (Supplementary Table 5). The 7 GO terms significantly associated with convergent deceleration (permutation q-value <0.15) were mainly associated with metabolism processes (Fig. 3c), such as amino acid metabolic process (GO:0009066), sulfur amino acid metabolic process (GO:00009069) and methionine metabolic process (GO:0006555) (Supplementary Table 6).

We also used RELAX19 to identify genes that experienced relaxed or intensified selection in all social branches relative to all nonsocial branches. We corrected the p-values reported by RELAX for multiple comparisons by computing q-values as described above. Out of 7590 single-copy orthologs included in the RELAX analysis, more genes showed evidence of relaxation of selection (849 genes, K < 1, p-value <0.15). Fig. 3d, Supplementary Data 1) compared to intensification of selection (671 genes, K > 1, q-value <0.15. Fig. 3d, Supplementary Data 2). Genes experiencing significant relaxation of selection in social branches were enriched for 115 GO terms (q-value <0.15), including four main categories (Fig. 3e, Supplementary Data 3): DNA replication and repair (e.g., GO:0009451: RNA modification; GO:0008303: RNA processing (GO:0005352: regulation of transcription, DNA-templated), reproduction (e.g., GO:0000003: reproduction; GO:0019953: sexual reproduction; GO:0032504: multicellular organism reproduction, GO:0007276: gamete generation), transcription (e.g., GO:0010467: gene expression; GO:0006396: RNA processing), development (developmental process: GO:0032502: multicellular organism development; GO:0007273), and metabolism (e.g., GO:0008152: metabolic process; GO:0043170: macromolecule metabolic process; GO:0190360: organic cyclic compound metabolic process). In addition, genes experiencing intensification of selection in social branches were enriched for 81 GO terms (Fig. 3f, Supplementary Data 4), including three main categories: immune response (e.g., GO:0032635: regulation of interleukin-1 production; GO:0024433: leukocyte mediated immunity), development (e.g., GO:0032502: developmental process; GO:0048586: anatomical structure development; GO:0007992: embryo development ending in birth or egg hatching; GO:0048729: tissue morphogenesis), and metabolism (e.g., GO:0071704: organic substance metabolic process; GO:0043170: macromolecule metabolic process; GO:0019538: protein metabolic process).

To clarify further what patterns of selection contribute to the acceleration and deceleration of protein evolution in social branches, we examined the overlap of lists of genes identified by RERconverge and RELAX. Two genes (Protein vav-I, VAV1, and Protein SYSI homolog, SYSI) out of the seven genes experiencing significant convergent acceleration were also identified as experiencing significant relaxation, indicating that elevated rates of protein evolution for these genes in social branches is caused by relaxation of purifying selection.

Social spiders harbor convergent amino acid substitutions
We finally asked if there were specific amino acid substitutions consistently associated with the convergent evolution of sociality in spiders. We used FADE20 to identify sites experiencing directional selection towards specific amino acids in social compared to nonsocial branches. After filtering sites that were also identified by FADE as
experiencing directional selection in nonsocial relative to social branches, we found 1421 sites in 123 genes with evidence (Bayes Factor > 100) for directional selection in social branches representing at least three independent origins of sociality. These genes were significantly enriched in 6 GO terms, such as positive regulation of neuromuscular junction development (GO:1904398), positive regulation of synaptic assembly at neuromuscular junction (GO:0045887), positive regulation of secretion (GO:0051047) and sterol homeostasis (GO:0055092) (Supplementary Table 7). Since we were most interested in identifying sites that showed a repeated pattern of substitution across the social branches, we further focused on sites that were inferred by FADE to have experienced the same substitution at least four separate times in social branches, resulting in a list of 357 sites in 27 genes (Supplementary Data 5). We further inspected protein alignments for these candidate genes and found that there were no substitutions that occurred in all social branches and no nonsocial branches. Substitutions that showed the strongest association with sociality (Supplementary Data 5) included substitutions in the GO terms under convergent intensification in social branches.

---

**Genes / GO terms under convergent acceleration in social branches**

| Genes / GO terms | Protein coding genes |
|------------------|----------------------|
| GO terms         | Significantly enriched GO terms |
| Neuronal function| 849                  |
| Positive regulation of programmed cell death | 671 |
| Positive regulation of cell death | 115 |
| Positive regulation of programmed cell death by hormones | 81 |

**Genes / GO terms under convergent deceleration in social branches**

| Genes / GO terms | Protein coding genes |
|------------------|----------------------|
| GO terms         | Significantly enriched GO terms |
| Neural function  | 7                    |
| Programmed cell death | 3 |
| Positive regulation of secretion | 22 |

---

**GO terms under convergent acceleration in social branches**

- Positive regulation of programmed cell death
- Positive regulation of cell death
- Developmental programmed cell death
- Positive regulation of programmed cell death by hormones
- Positive regulation of programmed cell death by ecdysone
- Salivary gland cell autophagic cell death
- Positive regulation of secretion
- Negative regulation of neuron death

**GO terms under convergent deceleration in social branches**

- Aspartate family amino acid metabolic process
- Sulfur amino acid metabolic process
- Methionine metabolic process
- Mannosylation
- Aspartate family amino acid biosynthetic process

**GO terms under convergent acceleration in social branches**

- Regulation of interleukin-1 production
- Leukocyte mediated immunity
- Developmental process
- Embryonic development ending in birth or egg hatching
- Morphogenesis of an epithelium
- Tissue morphogenesis
- Epithelium development
- System development
- Anatomical structure morphogenesis
- Embryonic morphogenesis
- Tissue development

**GO terms under convergent deceleration in social branches**

- Organic substance metabolic process
- Macromolecule metabolic process
- Heterocycle metabolic process
- RNA metabolic process
- Nucleic acid metabolic process
- Small molecule metabolic process
- Protein metabolic process
- Cellular amino acid metabolic process

---

**Article**

https://doi.org/10.1038/s41467-022-34446-8
Fig. 3 | Gene-level molecular evolution signatures of the convergent evolution of sociality in spiders. a Genes that experienced convergent shifts in gene-wide rate of molecular evolution during the transition to sociality. The number of genes and significantly enriched GO terms (permuation p-value <0.15) experiencing convergent acceleration or deceleration in social branches identified by RERegression. Genes or significantly enriched GO terms are depicted in light blue and light green, respectively. b GO terms with genes experiencing convergent accelerated evolution in social branches included categories such as neural function and programmed cell death. 10 out of 22 total significantly enriched GO terms are shown (Supplementary Table 5). c GO terms with genes experiencing convergent decelerated evolution concentratedly associated with metabolism functions. 5 out of 7 are shown (Supplementary Table 6). d A substantial set of genes that experienced convergent relaxation or intensification of selection during the transition to sociality. The number of genes and significantly enriched GO terms (p-value <0.15) experiencing convergent relaxation or intensification of selection identified by RELAX. e Enriched GO terms for genes experiencing convergent relaxation of selection were associated with DNA replication and repair, reproduction, transcription, development, and metabolism functions (p-value <0.15). P-value assigned to each enriched GO term was calculated by using Fisher’s exact test and shown in Supplementary Table 5, the corresponding q value was calculated with Benjamini-Hochberg multiple test correction and shown in Supplementary Table 5. 20 out of 115 significantly enriched GO terms are shown. f Enriched GO terms for genes experiencing convergent intensification of selection were associated with immune response, development, and metabolism functions (p-value <0.15). P-value assigned to each enriched GO term was calculated by using Fisher’s exact test and shown in Supplementary Table 5, the corresponding q value was calculated with Benjamini-Hochberg multiple test correction and shown in Supplementary Table 6.

gene Bromodomains-containing protein 4 (Fig. S3) that occurred in five social species, but also in the prolonged subsocial D. cancerides and the subsocial A. arizona. This gene also experienced other substitutions associated with social branches, providing some evidence for convergent evolution at the site-level in social spiders.

Discussion
We used a comparative genomic approach with 22 spider species representing eight independent and recent origins of sociality to answer a longstanding question: what genomic changes underpin the convergent evolution of sociality? Overall, we identified genome-wide, generic, and site-specific changes that repeatedly occurred during recent convergent transitions to sociality in spiders. Our study shows that while the precise genetic changes vary across independent origins of sociality, the repeated evolution of sociality in spiders predictably leaves genome-wide signatures and involves genes with conserved functions that may often be involved in the evolution of social behavior.

Consistent with several previous studies in social spiders23,26–29 and social insects15,49,50, we found that social branches tend to experience elevated genome-wide rates of molecular evolution (i.e., dN/dS) compared to nonsocial branches. Such elevated genome-wide rates of molecular evolution are hypothesized to be caused in general by reduced effective population size in social species, which results in relaxation of purifying and positive selection experienced by genes23,26–29. In social spiders, a reduced effective population size should result from their inbred breeding system, reproductive skew, and female-biased sex ratios23,23. Indeed, we found that the genome-wide pattern was primarily driven by the relaxation of both purifying and positive selection (Fig. 2c), so that genes in social branches tend to experience more relaxed, neutral evolution when compared to genes in nonsocial branches. This genome-wide pattern is likely to be a longer-term consequence of the switch from outbreeding to inbreeding and from unbiased to female-biased sex ratios. There was also variation for genome-wide dN/dS that did not depend on sociality, which might be explained by lineage-specific differences in effective population size or other factors. Similarly, in a previous study in hymenopteran insects, social branches experienced elevated genome-wide dN/dS compared to nonsocial branches, but most variation existed between lineages, with bees—regardless social organization—having the highest dN/dS23,23.

A fundamental phenotypic difference between social and nonsocial species is the extent of social interactions with conspecifics, which would have required increased behavioral tolerance towards conspecifics as an important preliminary change necessary for the evolution of cooperative group living in social spiders23,23,23, and other animals15,49,50. Furthermore, shifts in behavioral development and the timing of social behaviors, reproductive behavior, and dispersal behaviors, are also hypothesized to be critical in the evolution of sociality in spiders23,23,23 and insects2. Finally, changes in the size of certain brain regions and sensory organs are hypothesized to be important for detecting and processing social signals and social information15,49,50. The shifts in rates of molecular evolution between social and nonsocial branches that we detected in genes associated with neural function, neurogenesis, behavior, and reproduction may underlie changes in these phenotypes. Several candidate genes stand out. For example, the gene Autism susceptibility candidate 2 (AUTS2), which experienced intensified selection in social spiders, regulates neuronal and synaptic development and influences social behaviors in mice56 and mutation in AUTS2 has been associated with multiple neurological diseases, including autism in humans57,58. The gene Dystrophin, isoforms A/C/F/G/H (Dys), which experienced intensified selection in social spiders, affects social behavior, communication, and synaptic plasticity in mice56,56, and mutations in Dys are associated with autism and intellectual disability in humans57,58. The gene Synaptogenesis protein syg-2 (syg-2), which experienced relaxed selection in social spiders, determines synapse formation, and mutations in syg-2 are associated with locomotor behaviors in worms56,56. The genes Syntaxin-1A (STXI), Syntaxin-5 (STX5), and Syntaxin-6 (STX6), which experienced relaxed selection during social transition in social spiders, regulates the secretion of neurotransmitters and neuromodulators56,56. Each of these has been implicated in the expression of social behavior in various animals56,56, and also implicated in abnormal behavioral phenotypes in humans, including autism and schizophrenia56,56. The gene VAV1, which experienced acceleration and also relaxation in social spiders, regulates rhythmic behavior in worms56. Further studies will be necessary to validate the function and mechanism of these candidates influencing spider social behavior.

Besides critical changes in neural and behavioral phenotypes, transitions to social living are also thought to be associated with phenotypic changes in disease pressure, hormone regulation, reproduction, and metabolism. Social living is hypothesized to be associated with increased disease pressure, although researchers have found mixed support for genetic changes in immune-related genes in social insects compared to solitary insects27,27. A set of genes involved in innate immunity, including Pro-interleukin-16 (IL16) and Toll-like receptor Tollo (TOLLS) experienced relaxed selection in social spiders relative to nonsocial spiders. In addition, the gene component C3 (C3), a key gene in the complement system of invertebrates, including spiders27, experienced intensified selection in social spiders. These results are consistent with social spiders experiencing shifts in disease pressure, expected to be especially relevant given their inbred social system. Like social insects, the transition to social living in spiders has also been hypothesized to be associated with shifts in hormone regulation15,49,50. Genes experiencing relaxed selection in social spiders were enriched for GO terms including hormone-mediated signaling pathway and cellular response to steroid hormone stimulus, which could underlie these phenotypic changes. As emphasized above in terms of the genome-wide results, another phenotypic consequence of transitions to sociality in spiders, which may occur over longer time
scales, is the change of breeding system from outbreeding to inbreeding and the evolution of female-biased sex ratios. Several candidate genes may be associated with these shifts in breeding system. For example, the gene *Brd4* (Bromodomain-containing protein 4), which experienced intensified selection in social spiders, promotes self-renewal of germline stem cells during oogenesis in fly. The gene transcriptional regulator *ovo* (ovo), which also experienced intensified selection in social spiders, regulates female germline development in mouse and fly. Moreover, genes under relaxation of selection were significantly enriched for GO terms associated with reproduction and development, which could be tightly associated with the shifts in breeding system and sociality in social spiders. Metabolism-related genes and gene functions (GO terms) have also frequently been implicated in the expression of various types of social behavior in many animals, and comparative genomic studies in social insects have identified metabolism-associated genes as experiencing accelerated evolution in eusocial bees and ants. Consistent with these previous studies, we found that genes experiencing convergent shifts in selection in spiders were also enriched for metabolism-related functions.

In addition to gene-wide convergent signatures, although it is notoriously difficult to search for convergent amino acid substitutions, especially across a large number of genomes, we also identified some signatures of site-specific convergent molecular evolution. Genes with such convergent signatures of specific substitutions were enriched for several key GO terms, which were consistent with the functional categories that we found enriched at the gene level. However, inspection of the amino acid alignments for these genes showed that amino acid substitutions that occurred repeatedly in social branches also tended to be found in one or more nonsocial branches. For example, the gene *Brd4*, which has a set of specific convergent substitutions in at least five social spider species, affects regulating of inflammatory responses and learning and memory in rats. The inhibition of *Brd4* can alleviate transcriptional dysfunction and Fragile X syndrome, a neurodevelopmental disorder that causes intellectual disability, behavioral deficits, and is a leading genetic cause of autism spectrum disorder in humans. As described above, we identified genes and specific sites that tended to experience different patterns of molecular evolution in the replicate social branches compared to background nonsocial branches, with these genes being enriched for certain functions. However, we did not identify genes or sites within genes that always experienced evolutionary shifts in each of the eight branches that independently evolved sociality. For example, inspection of patterns of genic relative evolutionary rates (e.g., Fig. 3b, c) shows that genes with a relatively strong association between sociality and evolutionary rate still showed a lot of variation within each category. This means that we identified certain genes and functions that tended to experience shifts in evolutionary rates, but the exact details differed somewhat between the eight independent origins of sociality. Thus, specific substitutions or changes in specific genes are not required for the evolution of spider sociality, but changes do tend to occur in certain sets of genes or molecular pathways. Put another way, the strong statistical signatures across eight independent origins of sociality indicate that social spiders tend to travel similar evolutionary paths in terms of the functional genetic changes and genome-wide consequences, but the precise genetic details do vary. In practice, these results also emphasize that for complex phenotypes such as sociality, datasets representing many independent origins may be necessary to have the statistical power to detect these signatures. When samples representing multiple independent origins are not available, lineage-specific processes such as lineage-specific adaptations or historical contingency can be confounded with genetic changes that are actually associated with the phenotype of interest.

Our study adds to the growing list of studies that use comparative genomic approaches to elucidate the molecular underpinnings of convergent evolution and the degree to which similar phenotypic outcomes result from the same or similar genetic changes. Collectively, these studies show that some convergent phenotypes, especially those involving relatively simple physiological adaptation, such as the evolution of antibiotic resistance, pesticide resistance, or resistance to neurotoxins, often predictably involve the same substitutions or changes in one family of genes. Other phenotypic changes that are relatively more complex, such as the evolution of sociality, still have identifiable genomic signatures across independent origins, but these signatures are not as highly predictable in the sense that they always involve the same substitution or changes in the same genes, but instead involve changes in genes with similar functions.

Finally, our study illustrates the potential for ethanol-preserved specimens to be used in transcriptome-based comparative genomics analysis. Increasingly, ethanol-preserved specimens are being used for DNA-based phylogenomic and comparative genomic studies, but RNA-based studies have usually been restricted to samples collected in liquid nitrogen or RNA later, and stored at ~80°C. Our study builds on previous studies that successfully extracted RNA from ethanol-preserved samples and subsequently conducted transcriptome-based phylogenomics. We extracted RNA from spider specimens that had been preserved in ethanol at room temperature for years (Supplementary Table 1) and generated relatively high-quality RNA sequencing data (Supplementary Table 2). Collectively, our study highlights the heretofore underappreciated potential of RNA sequencing using ethanol-preserved specimens for comparative genomic and phylogenomic analyses, especially in species with large genomes, such as spiders, where sequencing and assembling whole genomes remains challenging.

**Methods**

**Sample collection**

We collected spider specimens of 16 species that range in social organization, including social, prolonged subsocial, subsocial and solitary, and stored the samples in 95% ethanol (Decon Labs, USA) (Fig. 1a). Specifically, this set of spider specimens includes five social and three subsocial species from the genus *Anelosimus* (Theridiidae), one social and four solitary species from the genus *Theridion* (Theridiidae), and one prolonged subsocial and two solitary species from the Sparassidae (Supplementary Table 1).

**RNA extraction with ethanol-preserved spider specimens**

First, we selected one individual for each spider species, and dried the spider specimen using filter paper (Fig. 1a). Second, we transferred the dried specimen into a new 1.5 ml microcentrifuge tube (Thermo Fisher Scientific, USA), and grinded the sample into a fine powder using RNase-Free Disposable Pellet Pestles (Thermo Fisher Scientific, USA) on liquid nitrogen. Third, we added 900 μl Trizol (Invitrogen, USA) and 8 μl glycogen (Thermo Fisher Scientific, USA) to maximize the amount of pure RNA (Fig. 1a). Finally, we extracted total RNA from all selected individuals of 16 spider species following the Trizol manufacturer’s protocol (Fig. 1a).

**Library preparation and RNA sequencing**

We assessed the RNA quality with Agilent Bioanalyzer 2100 (Agilent Technologies, USA) using an RNA 6000 Pico Kit (Fisher Scientific, USA), and detected RNA quality with Nanodrop 1000 (Nanodrop Technologies, USA). Only RNAs with high quality were used for cDNA synthesis and amplification following the protocol as previously described. Then, the primitive sequencing libraries were prepared and individually barcoded using the Nextera XT DNA Library Prep Kit (Illumina, USA) following the manufacturer’s protocol. We further did size selection for the constructed libraries with approximately 550 bp inserted fragments. At the first step of size selection, 100 μl of 12% PEG-6000/NaCl/Tris and 10 μl prepared Dynabeads were added to the 20 μl...
primitive library mixture and resuspended. The mixture was incubated for 5 min and placed on a magnetic stand for 5 min. The supernatant (150 μl) was transferred to a new tube and the beads discarded. At the second step to select fragments between 500 and 700 bp, 100 μl of 12.5% PEG-6000/NaCl/Tris and 10 μl prepared Dynabeads were added to the supernatant and mixed (Fig. 1a). The mixture was incubated for 5 min followed by bead separation on a magnetic stand. This time, the supernatant was discarded, and the beads were collected. The beads were washed twice with 70% ethanol (with 10 mM Tris, pH 6) and dried for 5 min. The tubes were then taken off the magnetic stand, and DNA was eluted from the beads by resuspending them in 15 μl EB. Finally, we measured the concentration of libraries using Qubit (Invitrogen, USA), and assessed the length distribution of libraries with Agilent Bioanalyser 2100 (Agilent Technologies, USA). All libraries were sequenced on a single lane of an Illumina HiSeq 2500 platform (rapid run mode) in the DNA Sequencing Section at the Okinawa Institute of Science and Technology Graduate University using HiSeq Rapid SBS Kit v2 kit (500 cycles, Illumina, USA) yielding 250-bp paired-end reads (Fig. 1a).

### Additional omics data collection
Besides the new transcriptomic data from our 16 spider species, we downloaded available genome or transcriptome data of additional six closely related spider species from NCBI (Fig. 1b, Supplementary Table 2). Specifically, we included two species with whole genomes and another two species with transcripts from genus Stegodyphus, one species with transcripts from genus Theridion species and one species with transcriptome from family Sparassidae, Heteropoda spp. ATSI2 (Supplementary Table 2).

### Assembly and annotation
We used FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) to assess sequencing read quality. We trimmed adapters and low-quality ends using Trimmomatic v.0.39 (github.com/timfr-dat/trimmomatic) until all reads scored above 20 at each position. We performed de novo transcriptome assembly using Trinity v.2.6.5[89] or maSPades[89] with default parameters, and removed the redundant transcripts in the assemblies for each species using CD-HIT v.4.8.1[90] with the threshold of 0.9, and further extracted the longest transcripts. Moreover, we performed the genome reassembly of *Anelosimus studiosus* using SPAdes[91], and further improved scaffolds of the draft genome using Redundans (github.com/lpryszcz/redundans), L_RNA_scaffolder[92] and PEP_scaffold[93] with our newly sequenced *A. studiosus* transcriptome sequencing reads and annotated spider protein databases following the analysis pipeline shown in Fig. 1b. Finally, we used BUSCO v.1.3.2 to assess the genome or transcriptome completeness based on the arachnId_oDB9 single-copy orthologous gene set from OrthoDB (www.orthodb.org) for each spider species (Fig. 1b).

We performed gene annotation by combining homology-based and de novo prediction approaches. First, we utilized AUGUSTUS[96] with spider *Parasteatoda tepidariorum* (GCF_000365465.3) as the training set, for de novo annotation (Fig. 1b). As for homolog-based prediction, we download protein sequences of published spider genomes from the NCBI database (Supplementary Table 2). The candidate genes were first identified by aligning these protein sequences to assembled transcriptomes and genomes using BLAT (github.com/djhshih/blat). We performed MAKER pipeline[97] to annotate the genome and transcriptome assemblies (Fig. 1b). Finally, we integrated the gene models predicted by both approaches using CLEAN[98] with default parameters to remove redundant genes.

### Ortholog identification
To assign the orthology among genes in above spider taxa, we used a best reciprocal hit search strategy to infer orthologous groups (OGs)[99,100]. In brief, we included the predicted complete proteomes from the latest spider genomes (Supplementary Table 2). We pooled them into a local protein database and conducted a self-to-self BLAST search using DIAMOND v.0.9.23[101] with an E-value cutoff of 1e-3, and removed hits with identity <30% and coverage <30%. We identified the OGs from the BLAST results using OMA (www.omabrowser.org) with default settings, and finally inferred 3276 single-copy OGs. In addition, we downloaded the curated orthology map of Arachnida from OrthoDB (www.orthodb.org) which contains recorded 8805 OGs. Of these OGs in HaMStR v.1.0[102] we identified the putative OGs in each spider species with threshold E-values of less than 10-20. We repeated this analysis pipeline twice, once with the 3276 OGs determined with OMA, and once with the 8805 OGs from OrthoDB. Among these putative OGs, we further identified one-to-one, one-to-many, and many-to-many orthologs among these spider taxa (Fig. 1b).

### Genome-scale phylogeny construction and divergence time estimation
To estimate the topology of the phylogeny including the 22 spider species and an outgroup *A. geniculata*, we used a phylogenotranscriptomic approach based on a single-copy ortholog dataset[103] (Fig. 1b). In brief, we prepared a dataset including all amino acid sequences of single-copy orthologs shared by the 23 spider species. Then, we performed sequence alignment using clustalo[104] and trimmed gaps using trimAl v1.2[105]. Moreover, we filtered each single-copy ortholog with strict constraints, including length (minimum 200 aa), sequence alignment (maximum missing data 50% in alignments). We prepared a concatenated dataset including core-shared single-copy orthologs and detected the best-fit model of sequence evolution using ModelFinder[106], and built the maximum likelihood phylogenetic trees using RAxML v8.2[107]. Statistical support for major nodes were estimated from 1000 bootstrap replicates. Finally, we estimated the divergence time using MCMCtree in PAML[108] with the topology of the 4DTV position and four calibration time based on the TimeTree database (www.timetree.org) and published literatures[109].

### Analysis of codon substitution rate associated with sociality
To test whether the rate of molecular evolution was associated with sociality, we prepared the codon alignments of single-copy orthologs shared by 22 spider species, derived from amino acid sequence alignments and the corresponding DNA sequences using PAL2NAL v.14 (www.bork.embl.de/pal2nal/). We constructed a “supergene” dataset that used the concatenated codon sequences of all core-shared orthologs of 22 species, and another coalescent gene dataset that included all shared orthologs. We used PAML 4.7a[110] and HyPhy 2.5[111] to estimate the codon substitution rates across the spider phylogeny. First, we applied the free-ratio model (“several ω ratio”) to calculate the ratio of non-synonymous to synonymous rate (dN/dS, ω) separately for each species with the supergene dataset using the package CodeML in PAML 4.7a[110] (Fig. 1b). To further characterize the patterns of molecular evolution associated with social organization (Fig. 1b), we estimated two discrete categories of dN/dS for social and nonsocial spiders taxa with the concatenated ortholog dataset (genome-wide) using HyPhy 2.5[111] (Fig. 1b). Finally, we employed a likelihood ratio test for the comparison of genome-wide dN/dS.

### Analysis of intensification or relaxation of selection associated with sociality
To elucidate the pattern of selection acting during the evolutionary transition to sociality, we used RELAX[112], which estimates variable dN/dS ratios across sites with three discrete categories. We compared...
social spiders with nonsocial spiders at both genome-wide and gene-wide scales, and then employed a likelihood ratio test (LRT) (Fig. 1b). We corrected the p-values reported by RELAX for multiple comparisons by estimating q-values as an estimate of the false discovery rate (FDR) for each gene. Genes showing K-value less than 1 and q-value <0.15 are considered to experience significant relaxation of selection. Similarly, genes showing K-value more than 1 and q-value <0.15 are considered to experience significant intensification of selection.

Finally, we performed gene ontology (GO) enrichment analysis for these genes experiencing significant relaxation or intensification using GOATools, and we corrected the reported p-values for multiple comparisons by computing q-values, and considered those GO terms (Biological Processes, BP) < 0.15 to be shown significant enrichment.

Analysis of convergence shifts in relative evolutionary rates associated with sociality

To determine if particular orthologs experience convergent shifts in selective pressure, including acceleration or deceleration in social branches across the phylogeny, we estimated the protein evolutionary rate (relative evolutionary rate, RER) using the R package RERconverge (Fig. 1b). RERconverge calculates relative branch lengths by normalizing branches for focal branches (i.e., social branches) to the distribution of branch lengths across all genes. This enables the identification of convergent changes in evolutionary rates across foreground relative to background branches while accounting for differences in phylogenetic divergence and in baseline rates of evolution across taxa. RERconverge compares rates of change in focal foreground branches and the rest of the tree and identifies genes that have a significant correlation between RERs and a phenotype of interest (e.g., sociality), as previously described. In brief, we prepared the amino acid alignments of 7590 single-copy orthologs shared by up to 23 spider species. We estimated the branch orthologs shared by up to 23 spider species. We estimated the branch evolutionary rate (relative evolutionary rate, RER) using the R package phangorn, and calculated the RERs for each branch with corrected these permulations. We identified sites experiencing directional selection towards specific amino acids in foreground (i.e., social) relative to background (i.e., nonsocial) branches. We performed GO enrichment for genes with convergent sites using GOATools as described above.

Analysis of convergent amino acid substitutions associated with sociality

To determine if convergent amino acid substitutions at specific sites of genes in social spider branches compared to nonsocial spiders across the phylogeny, we identified convergent substitution sites using FADE (FUBAR Approach to Directional Evolution) in HyPhy 2.5 with core-shared single-copy orthologs dataset (Fig. 1b). FADE identifies sites experiencing directional selection towards specific amino acids in foreground (i.e., social) relative to background (i.e., nonsocial) branches. We performed GO enrichment for genes with convergent sites using GOATools as described above.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw and processed transcriptome data have been deposited in NCBI under the project PRJNA658164. We also used available spider genome and transcriptome data which were downloaded from NCBI, including Acanthoscurria geniculata (GCA_000661875.3), Anelosimus studiosus (GCA_000829765.3), Stegodyphus minutus (GCA_000611955.2), Stegodyphus dumicola (GCA_010614865.1), Stegodyphus africanus (SRR7062696), Stegodyphus lineatus (SRR7062695), Theridion grallator (SRR960715, SRR960716, SRR960718, SRR960719, SRR960611, SRR960612, SRR960614, SRR960615, SRR960616), Heteropoda sp. ATS12 (SRR6425926).

Code availability

All scripts required to perform all analyses are publicly available on Github at https://github.com/jiyideanjiao/Social_Spider_Evolutionary_Genomics and Zenodo at https://doi.org/10.5281/zenodo.7222296.

References

1. Rubenstein, D. R. & Abbot, P. Comparative Social Evolution (Cambridge University Press, 2017).
2. Toth, A. L. & Robinson, G. E. Evo-devo and the evolution of social behavior. Trends Genet. 23, 334–341 (2007).
3. Amundsen, G. V., Csondes, A., Fondrk, M. K. & Page, R. E. Jr. Complex social behavior derived from maternal reproductive traits. Nature 439, 76–78 (2006).
4. Links, T. A. & Wade, M. J. The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. Q. Rev. Biol. 80, 317–336 (2005).
5. O’Connell, L. A. & Hofmann, H. A. The vertebrate mesolimbic reward system and social behavior network: a comparative synthesis. J. Comp. Neurol. 519, 3599–3639 (2011).
6. Robinson, G. E., Grozinger, C. M. & Whitfield, C. W. Socio-genomics: social life in molecular terms. Nat. Rev. Genet. 6, 257–270 (2005).
7. Berens, A. J., Hunt, J. H. & Toth, A. L. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. Mol. Biol. Evol. 32, 690–703 (2015).
8. Young, R. L. et al. Conserved transcriptomic profiles underpin monogamy across vertebrates. Proc. Natl Acad. Sci. USA 116, 1331–1336 (2019).
9. Rittschof, C. C. et al. Neuromolecular responses to social challenge: common mechanisms across mouse, stickleback fish, and honey bee. Proc. Natl Acad. Sci. USA 111, 17929–17934 (2014).
10. Warner, M. R., Qiu, L., Holmes, M. J., Mikheyev, A. S. & Links, T. A. Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. Nat. Commun. 10, 2651 (2019).
11. Johnson, B. R. & Tsutui, N. D. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. BMC Genomics 12, 164 (2011).
12. Jasper, W. C. et al. Large-scale coding sequence change underlies the evolution of postdevelopmental novelty in honey bees. Mol. Biol. Evol. 33, 1379 (2016).
13. Sumner, S. The importance of genomic novelty in social evolution. Mol. Ecol. 23, 26–28 (2014).
14. Dogantiz, K. A. et al. Insects with similar social complexity show convergent patterns of adaptive molecular evolution. Sci. Rep. 8, 10388 (2018).
15. Kapheim, K. M. et al. Genomic signatures of evolutionary transitions from solitary to group living. Science 348, 1139–1143 (2015).
16. Links, T. A. & Johnson, B. R. Re-thinking the social ladder approach for elucidating the evolution and molecular basis of insect societies. Curr. Opin. Insect Sci. 34, 123–129 (2019).
17. Kocher, S. D. & Paxton, R. J. Comparative methods offer powerful insights into social evolution in bees. *Apidologie* **45**, 289–305 (2014).

18. Avilés, L. & Guevara, J. in *Comparative Social Evolution* (eds. Rubenstein, R., & Abbot, P.) 188–223 (Cambridge University Press, 2017).

19. Yang, C. & Rayor, L. S. Maternal care and subsocial behaviour in spiders. *Biol. Rev. Camb. Philos. Soc.* **89**, 427–449 (2014).

20. Lubin, Y. & Bilde, T. in *Advances in the Study of Behavior* Vol. **37**, 83–145 (Academic Press, 2007).

21. Agnarsson, I., Avilés, L., Coddington, J. A. & Maddison, W. P. Loss of genetic variability in social spiders: genetic and phylogenetic consequences of population subdivision and inbreeding. *J. Evol. Biol.* **26**, 27–37 (2013).

22. Johannesen, J., Lubin, Y., Smith, D. R., Bilde, T. & Schneider, J. M. The age and evolution of sociality in Stegodyphus spiders: a molecular phylogenetic perspective. *Proc. Biol. Sci.* **274**, 231–237 (2007).

23. Agnarsson, I., Avilés, L. & Maddison, W. P. Orthologous genes identified by transcriptome sequencing in the spider genus Stegodyphus. *BMC Genomics* **13**, 70 (2012).

24. Tong, C., Najm, G. M., Pinter-Wollman, N., Pruitt, J. N. & Linskaya, T. A. Comparative genomics identifies putative signatures of sociality in spiders. *Genome Biol. Evol.* **12**, 122–133 (2020).

25. Arnedo, M. A., Agnarsson, I. & Gillespie, R. G. Molecular insights into the phylogenetic structure of the spider genus Theridion (Araneae, Theridiidae) and the origin of the Hawaiian Theridion-like fauna. *Zool. Syst.* **36**, 337–352 (2007).

26. Luo, Y. et al. Global diversification of anelosimus spiders driven by long-distance overwater dispersal and neogene climate oscillations. *Syst. Biol.* https://doi.org/10.1093/sysbio/sya017 (2020).

27. Agnarsson, I. & Rayor, L. S. A molecular phylogeny of the Australian huntsman spiders (Sparassidae, Deleninae): implications for taxonomy and social behaviour. *Mol. Phylogenet. Evol.* **69**, 895–905 (2013).

28. Sanggaard, K. W. et al. Spider genomes provide insight into composition and evolution of venom and silk. *Nat. Commun.* **5**, 3765 (2014).

29. Croucher, P. J. P., Brewer, M. S., Winchell, C. J., Oxford, G. S. & Gillespie, R. G. De novo characterization of the gene-rich transcriptomes of two color-polymorphic spiders, Theridion gracilator and T. californicum (Araneae: Theridiidae), with special reference to pigment genes. *BMC Genomics* **14**, 862 (2013).

30. Shao, L. & Li, S. Early Cretaceous greenhouse promoted higher taxa diversification in spiders. *Mol. Phylogenet. Evol.* **127**, 146–155 (2018).

31. Fernández, R. et al. Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. *Curr. Biol.* **28**, 2190–2193 (2018).

32. Liu, S., Aagegaard, A., Bæksgaard, J. & Bilde, T. DNA methylation patterns in the social spider, *Stegodyphus dumicola*. *Genes* **10**, 137 (2019).

33. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

34. Agnarsson, I., Maddison, W. P. & Avilés, L. The phylogeny of the social Anelosimus spiders (Araneae: Theridiidae) inferred from six molecular loci and morphology. *Mol. Phylogenet. Evol.* **43**, 833–851 (2007).

35. Simões, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

36. Agnarsson, I., Maddison, W. P. & Avilés, L. The phylogeny of the social Anelosimus spiders (Araneae: Theridiidae) inferred from six molecular loci and morphology. *Mol. Phylogenet. Evol.* **43**, 833–851 (2007).

37. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1596–1597 (2007).

38. Settepani, V. et al. Evidence for faster X chromosome evolution in spiders. *Phylogenetic analysis of molecular constraints underlying extended lifespan*. *Elife* **9**, e51089 (2020).

39. Saputra, E., Kowalczyk, A., Cusick, L., Clark, N. & Chikina, M. Phylogenetic permutations: a statistically rigorous approach to measure confidence in associations between phenotypes and genetic elements in a phylogenetic context. *Mol. Biol. Evol.* **38**, 3004–3021 (2021).

40. Kowalczyk, A., Partha, R., Clark, N. L. & Chikina, M. Pan-mammalian analysis of molecular constraints underlying extended lifespan. *Elife* **9**, e51089 (2020).

41. Saputra, E., Kowalczyk, A., Cusick, L., Clark, N. & Chikina, M. Phylogenetic permutations: a statistically rigorous approach to measure confidence in associations between phenotypes and genetic elements in a phylogenetic context. *Mol. Biol. Evol.* **38**, 3004–3021 (2021).

42. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).

43. Romiguier, J. et al. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J. Evol. Biol.* **29**, 593–603 (2014).

44. Weyna, A. & Romiguier, J. Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and pollinating bees. *Peer Community Journal* 1 (2021).

45. Viera, C. & Agnarsson, I. In *Behaviour and Ecology of Spiders: Contributions from the Neotropical Region* (eds. Viera, C. & Gonzaga, M. O.) 351–381 (Springer International Publishing, 2017).

46. Lin, N. & Michener, C. D. Evolution of Sociality in Insects. *Q. Rev. Biol.* **47**, 131–159 (1972).

47. Wittwer, B. et al. Solitary bees reduce investment in communica- tion dynamics, and *Proc. Natl Acad. Sci. USA* **114**, 6569–6574 (2017).

48. Caponera, V., Avilés, L., Barrett, M. & O’Donnell, S. Behavioral attributes of social groups determine the strength and direction of selection on neural investment. *Front. Ecol. Evol.* **9**, https://doi.org/10.3389/fevo.2021.733228 (2021).

49. Hori, K. et al. AUTS2 regulation of synapses for proper synaptic inputs and social communication. *iScience* **23**, 101183 (2020).

50. Oksenberg, N. & Ahituv, N. The role of AUTS2 in neurodevelopment and human evolution. *Trends Genet.* **29**, 600–608 (2013).

51. Oksenberg, N., Stevison, L., Wall, J. D. & Ahituv, N. Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genet.* **9**, e1003221 (2013).
78. Rittschof, C. C. & Robinson, G. E. Behavioral genetic toolkits: a dictionary of behavioral motifs reveals clusters of genes affecting Caenohabditis elegans locomotion. Proc. Natl Acad. Sci. USA 110, 791–796 (2013).

79. Woodard, S. H. et al. Genes involved in convergent evolution of plasticity and behavioral functions. PLoS ONE 4, e6574 (2008).

80. Shen, K., Fetter, R. D. & Bargmann, C. I. Sympatic specificity is generated by the synaptic guidance protein SYG-2 and its receptor, SYG-1. Cell 116, 869–881 (2004).

81. Brown, A. E. X., Yemini, E. I., Grundy, L. J., Jucikas, T. & Schafe, W. R. A dictionary of behavioral motifs reveals clusters of genes affecting Caenohabditis elegans locomotion. Proc. Natl Acad. Sci. USA 110, 791–796 (2013).

82. Bennett, M. K., Calakos, N. & Scharf, R. H. Syntaxin: an synaptic response to bidirectional selection on social behavior. Proc. Natl Acad. Sci. USA 110, 791–796 (2013).

83. Viljakainen, L. et al. Rapid evolution of immune proteins in social insects. Mol. Biol. Evol. 117, 597–607 (2016).

84. Partha, R. et al. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. Elife 6, e25884 (2017).

85. Feldman, C. R., Brodie, E. D. Jr., Brodie, E. D. 3rd & Pfrender, M. E. Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. Proc. Natl Acad. Sci. USA 109, 4556–4561 (2012).

86. McGlothlin, J. W. et al. Historical contingency in a multigene family facilitates adaptive evolution of toxin resistance. Curr. Biol. 26, 1616–1621 (2016).

87. Short, A. E. Z., Dikow, T. & Moreau, C. S. Entomological collections in the age of big data. Annu. Rev. Entomol. 63, 513–530 (2018).

88. Gough, H. M., Allen, J. M., Toussaint, E. F. A., Storer, C. G. & Allen, J. M. Analyses of molecular adaptation based on few genome sequence data. Proc. Natl Acad. Sci. USA 115, 1637–1642 (2018).

89. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652 (2011).

90. Nunk, S. et al. Assembling single-cell genomes and minigenomes from chimeric MDA products. J. Comput. Biol. 20, 714–737 (2013).

91. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152 (2012).

92. Xue, W. et al. L RNA scaffold: scaffolding genomes with transcripts. BMC Genomics 14, 604 (2013).

93. Zhu, B.-H. et al. PE, saddle: homologous) proteins to scaffold genomes. Bioinformatics 32, 3193–3195 (2016).

94. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34, W435–9 (2006).

95. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18, 188–196 (2008).

96. Elsik, C. G. et al. Creating a honey bee consensus gene set. Genome Biol. 8, R13 (2007).

97. Bucek, A. et al. Evolution of termite symbiosis informed by transcriptome-based phylogenies. Curr. Biol. 29, 3728–3734.e4 (2019).
Acknowledgements

We would like to thank Dr. Marc Milne from University of Indianapolis for kindly contributing ethanol-preserved spider specimens. We would like to thank Dr. Yi-yong Zhao from Fudan University for suggestion in phylogenetic analysis and Drs. Endo Tatsuya and Lijun Qiu from Okinawa Institute of Science and Technology Graduate University for expertise in RNA-seq library construction. This work was funded by the National Institutes of Health Grant GM115509 to T.A.L.

Author contributions

C.T. and T.A.L. conceived and designed the study. L.A. and L.S.R. provided spider samples and expertise about spider natural history and social evolution. A.S.M. provided reagents and sequencing support. C.T. generated libraries for RNA sequencing. C.T. and T.A.L. performed all analyses and created all figures. C.T. and T.A.L. wrote a first draft of the manuscript and all authors contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-34446-8.

Correspondence and requests for materials should be addressed to Chao Tong or Timothy A. Linksvayer.

Peer review information Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022