bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title

Genomic signatures of recent convergent transitions to social life in spiders

Authors

Chao Tong[1,2], Leticia Avilés[3], Linda S. Rayor[4], Alexander S. Mikheyev[5], Timothy A. Linksvayer[1,2]

Authors for correspondence:

Chao Tong (tongchao1990@gmail.com)

Timothy A. Linksvayer (tlinksvayer@gmail.com)

Affiliation

[1] Department of Biology, University of Pennsylvania, Philadelphia, PA, 19104, USA

[2] Department of Biological Sciences, Texas Tech University, Lubbock, TX, 79409, USA

[3] Department of Zoology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

[4] Department of Entomology, Cornell University, Ithaca, New York 14853, USA

[5] Evolutionary genomics group, Research School of Biology, Australian National University, Canberra, 0200, Australia

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

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 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 [15,17]. 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 [15-17]. 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 [18-20]. 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,

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

having dispersed from the egg sac soon after hatching [18,22]; 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 [18,19]; 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 [19]; 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 [18]. 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 [21-24]. By comparing pairs of social species and their closest nonsocial (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 [23,26-29]. These studies, together with a preliminary comparative genomic study of two social species from one genus and several solitary species [30], have begun to identify putative genomic signatures of the evolution of sociality in spiders, including elevated genome-wide rates of molecular evolution [26-30].

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 [24,31-33]. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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: Theriidiade), 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 [28,33-38] (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 (2,934 genes) in each spider species (Supplementary Table 3).

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 7,590 orthologous groups (OGs) and 3,832 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 phylogenies [24,31-33,40] (Fig. 2a). We estimated divergence times for each phylogenetic tree node (Fig. 2a).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 nonsynonymous 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 4.7a [41] (Wilcoxon rank-sum test, *p*-value = 0.013, Fig. 2b) and HyPhy 2.5 [42] (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 [43] quantifies 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 caused mainly by relaxation of both purifying and positive selection (K = 0.88, LRT, *p*-value < 0.001) (Fig. 2c).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 RERconverge [44] 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 "permulations", to assess the statistical significance of genes and enriched GO terms. This permulation 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 alone [45,46]. Importantly, permulations also correct for non-independence among genes for GO term enrichment, which is not true for the nominal parametric *p*-values reported from RERconverge [44,47].

We ran 10,000 permulations 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 permulation *p*-values were corrected for multiple comparisons by computing *q*-values [48], and we considered *q*-values < 0.15 (i.e. genes and GO terms with an estimated FDR < 0.15) to be significant. Out of 7,590 single-copy orthologs included in REconverge analysis, we identified 7 genes under convergent acceleration in social branches (permulation *q*-value < 0.15) (Fig. 3a, Supplementary Table 4) and 3 genes experiencing convergent deceleration in social branches (permulation *q*-value < 0.15) (Fig. 3a, Supplementary Table 4). The 22 GO terms significantly associated with convergent acceleration (permulation *q*-value < 0.15) were mainly related to neural function and programmed cell death (Fig. 3b), such as negative regulation of neuron death (GO:1901215), negative regulation of secretion (GO:0051048) and positive regulation of programmed cell death (GO:0043068) (Supplementary Table 5). The 7 GO terms significantly associated with convergent deceleration (permulation *q-value* < 0.15) were mainly associated with metabolism processes (Fig. 3c), such as aspartate

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

family amino acid metabolic process (GO:0009066), sulfur amino acid metabolic process (GO:0000096) and methionine metabolic process (GO:0006555) (Supplementary Table 6).

We also used RELAX [43] 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 7,590 single-copy orthologs included in the RELAX analysis, more genes showed evidence of relaxation of selection (849 genes, K < 1, *q*-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:0008033: tRNA processing; GO:0006355: 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:0007275), and metabolism (e.g. GO:0008152: metabolic process; GO:0043170: macromolecule metabolic process; GO:1901360: 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:0032652: regulation of interleukin-1 production; GO:0002443: leukocyte mediated immunity), development (e.g. GO:0032502: developmental process; GO:0048856: anatomical structure development; GO:0009792: 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-1*, *VAV1* and *Protein SYS1 homolog*, *SYS1*) out of the seven genes experiencing significant convergent acceleration were also

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 FADE [42] to identify sites experiencing directional selection towards specific amino acids in social compared to nonsocial branches [42]. After filtering sites that were also identified by FADE as experiencing directional selection in nonsocial relative to social branches, we found 1,421 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 gene Bromodomain-containing protein 4 (Fig. S3) that occured 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, genic, and site-specific changes that repeatedly occurred during recent convergent transitions to

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 spiders [23,26-29] and social insects [15,49,50], 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 genes [23,26-29]. In social spiders, a reduced effective population size should result from their inbred breeding system, reproductive skew, and female-biased sex ratios [23,51]. 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/dS [50].

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 spiders [25,52,53] and other animals [54]. 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 spiders [20,25] and insects [4]. 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 information [55,56]. The shifts in rates of molecular evolution between social and nonsocial branches that we

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 mice [57] and mutation in *AUTS2* has been associated with multiple neurological diseases, including autism in humans [58,59]. 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 mice [60,61], and mutations in *Dys* are associated with autism and intellectual disability in humans [61]. 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 worms [62,63]. The genes *Syntaxin-1A* (*STX1*), *Syntaxin-5* (*STX5*) and *Syntaxin-6* (*STX6*), which experienced relaxed selection during social transition in social spiders, regulates the secretion of neurotransmitters and neuromodulators [64]. Each of these has been implicated in the expression of social behavior in various animals [65,66], and also implicated in abnormal behavioral phenotypes in humans, including autism and schizophrenia [67,68]. The gene VAV1, which experienced acceleration and also relaxation in social spiders, regulates rhythmic behavior in worms [69]. 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 insects [70-72]. A set of genes involved in innate immunity, including *Pro-interleukin-16* (*IL16*) and *Toll-like receptor Tollo* (*TOLL8*) 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 spiders [73], 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 regulation [74].

Genes experiencing relaxed selection in social spiders were enriched for GO terms including

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [20,25]. Several candidate genes may be associated with these shifts in breeding system. For example, the gene *lilipod* (*lili*), which experienced intensified selection in social spiders, promotes self-renewal of germline stem cells during oogenesis in fly [75]. The gene *transcriptional regulator ovo* (*ovo*), which also experienced intensification in social spiders, regulates female germline development in mouse and fly [76,77 78]. 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 [9,79], and comparative genomic studies in social insects have identified metabolism-associated genes as experiencing accelerated evolution in eusocial bees [80] and ants [81]. 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 *Bromodomain-containing protein 4* (*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 [82]. 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 [83].

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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., Figs. 3b, 3c) 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 [16,30,84,85].

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 [46,86-89]. 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 [90], 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [91], 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 [92,93]. 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 [38], 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* (Theriidiidae), 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [94]. 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 Bioanalyzer 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

another two species with transcriptomes from genus *Stegodyphus*, one species with transcriptomes from genus *Theridion* species and one species with transcriptome from family Sparassidae, *Heteropoda* spp. ATS12 (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/timflutre/trimmomatic) until all reads scored above 20 at each position. We performed *de novo* transcriptome assembly using Trinity v.2.6.5 [95] or rnaSPAdes [96] with default parameters, and removed the redundant transcripts in the assemblies for each species using CD-HIT v.4.8.1 [97] with the threshold of 0.9, and further extracted the longest transcripts. Moreover, we performed the genome reassembly of *Anelosimus studiosus* using SPAdes [96], and further improved scaffolds of the draft genome using Redundans (github.com/lpryszcz/redundans), L_RNA_scaffolder [98] and PEP_scaffolder [99] with our newly sequenced *A. studiosus* transcriptome sequencing reads and annotated spider protein datasets following the analysis pipeline shown in Fig. 1b. Finally, we used BUSCO v.5.1.2 to assess the genome or transcriptome assembly 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 [100] 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 firstly identified by aligning these protein sequences to assembled transcriptomes and genomes using BLAT (github.com/djhshih/blat). We performed MAKER pipeline [101] to annotate the genome and transcriptome assemblies (Fig. 1b). Finally, we integrated the gene models predicted by both approaches using GLEAN [102] with default parameters to remove redundant genes.

Ortholog identification

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To assign the orthology among genes in above spider taxa, we used a best reciprocal hit search strategy to infer orthologous groups (OGs) [30,103]. 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.0.9.29 [104] with an E-value cutoff of $1e^{-5}$, 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 3,276 single-copy OGs. In addition, we downloaded the curated orthology map of Arachnida from OrthoDB (www.orthodb.org) which contains recorded 8,805 OGs. Of these OGs in HaMStR v.1.0 [105], 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 3,276 OGs determined with OMA, and once with the 8,805 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). For each 1:1 orthologous pair, we selected the longest transcripts associated with recorded OGs for each species as putative single-copy orthologs. Gene ontology (GO) terms were assigned to single-copy orthologs using eggNOG-mapper v2 (eggnog-mapper.embl.de/).

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 phylotranscriptomic approach based on a single-copy ortholog dataset [106] (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 [107] and trimmed gaps using trimAl v1.2 [108]. 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 [109], and built the maximum likelihood phylogenetic trees using RAxML v8.2 [110]. Statistical support for major nodes were estimated from 1,000 bootstrap replicates. Finally, we estimated the divergence time using MCMCtree in PAML [41] with the topology of the 4DTV position and four calibration time based on the Timetree database (www.timetree.org) and published literatures [32].

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [41] and HyPhy 2.5 [42] 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 [41] (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 [42] (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 [43], 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 [48]. Genes showing K-value less than 1 and *q*-value < 0.15 are considered to experience significant relaxation of selection. Similarly, genes showing K-values 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 [111], 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [44] (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 [112,113]. In brief, we prepared the amino acid alignments of 7,590 single-copy orthologs shared by up to 23 spider species. We estimated the branch lengths of each gene tree based on the inferred species tree using the R package phangorn [114], and calculated the RERs for each branch with the corresponding gene tree. We used 10,000 permulations to estimate permulation p-values for the correlation between the RER of foreground (i.e. social species) relative to the rest of the tree, and we corrected these permulation p-values for multiple comparisons by computing *q*-values [48]. We defined genes with significantly higher RERs (permulation *q*-value < 0.15) in foreground branches as experiencing convergent acceleration in social branches, and genes with significantly lower RERs in foreground branches as experiencing convergent deceleration in social branches. In addition, we performed gene set enrichment analysis for the gene lists produced from the correlation analysis using the fastwilcoxGMTall functions in RERconverge with 10,000 permulations [44]. We corrected the resulting permulation *p*-values for multiple comparisons by computing q-values, and considered *q*-values < 0.15 (i.e. GO terms with an FDR < 0.15) as significant.

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 [42] 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data availability

Raw and processed transcriptome data have been deposited in NCBI under the project PRJNA685164. We also used available spider genome and transcriptome data which were downloaded from NCBI, including *Acanthoscurria geniculata* (GCA_000661875.l), *Anelosimus studiosus* (GCA_008297655.l), *Stegodyphus mimosarum* (GCA_000611955.2), *Stegodyphus dumicola* (GCA_010614865.l), *Stegodyphus africanus* (SRR7062696), *Stegodyphus lineatus* (SRR7062695), *Theridion grallator* (SRR960715, SRR960716, SRR960718, SRR960719, SRR960611, SRR960612, SRR960614, SRR960615, SRR960616), *Heteropoda sp. ATS12* (SRR6425926). In addition, we used four calibration time points from Timetree and a recent paper [32], including *Stegodyphus mimosarum* - *Stegodyphus dumicola* (8 to 15 Mya), *Stegodyphus mimosarum* - *Stegodyphus lineatus* (13 to 23 Mya), *Anelosimus guacamayos* - *Anelosimus oritoyacu* (1.4 to 2.7 Mya), *Anelosimus eximius* to *Anelosimus domingo* (3.7 to 6.3 Mya).

Code availability

All scripts required to perform all analyses are publicly available on Github at github.com/jiyideanjiao/Social_Spider_Evolutionary_Genomics and Zenodo at doi.org/10.5281/zenodo.7222296 [115].

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 phylotranscriptomic 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 TAL.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Author Contributions

CT and TAL conceived and designed the study. LA and LSR provided spider samples and expertise about spider natural history and social evolution. ASM provided reagents and sequencing support. CT generated libraries for RNA sequencing. CT and TAL performed all analyses and created all figures. CT and TAL wrote a first draft of the manuscript and all authors contributed to the final manuscript.

Competing interests

The authors have no competing interests.

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. Amdam, G. V., Csondes, A., Fondrk, M. K. & Page, R. E., Jr. Complex social behaviour derived from maternal reproductive traits. Nature 439, 76–78 (2006).

4. Linksvayer, 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. Sociogenomics: social life in molecular terms. Nat. Rev. Genet. 6, 257–270 (2005).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. U. S. A. 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. U. S. A. 111, 17929–17934 (2014).

10. Warner, M. R., Qiu, L., Holmes, M. J., Mikheyev, A. S. & Linksvayer, 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. & Tsutsui, 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. Molecular ecology vol. 23 26–28 (2014).

14. Dogantzis, 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. Linksvayer, T. A. & Johnson, B. R. Re-thinking the social ladder approach for elucidating

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. Sociality in Spiders. in Comparative Social Evolution (eds. Rubenstein, R. & Abbot, P.) 188–223 (Cambridge University Press, 2017).

19. Yip, E. 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. The Evolution of Sociality in Spiders. 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. Sociality in theridiid spiders: repeated origins of an evolutionary dead end. Evolution 60, 2342–2351 (2006).

22. Avilés, L. Social Spiders. in Encyclopedia of Social Insects (ed. Starr, C. K.) 1–10 (Springer, 2020).

23. Agnarsson, I., Avilés, L. & 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).

24. 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).

25. Avilés, L. 23• Causes and consequences of cooperation and permanent-sociality in spiders. The evolution of social behavior in insects and arachnids 476–498 (1997).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26. Settepani, V. et al. Evolution of sociality in spiders leads to depleted genomic diversity at both population and species levels. Mol. Ecol. 26, 4197–4210 (2017).

27. Settepani, V., Bechsgaard, J. & Bilde, T. Phylogenetic analysis suggests that sociality is associated with reduced effectiveness of selection. Ecol. Evol. 6, 469–477 (2016).

28. Bechsgaard, J. et al. Evidence for Faster X Chromosome Evolution in Spiders. Mol. Biol. Evol. 36, 1281–1293 (2019).

29. Mattila, T. M., Bechsgaard, J. S., Hansen, T. T., Schierup, M. H. & Bilde, T. Orthologous genes identified by transcriptome sequencing in the spider genus Stegodyphus. BMC Genomics 13, 70 (2012).

30. Tong, C., Najm, G. M., Pinter-Wollman, N., Pruitt, J. N. & Linksvayer, T. A. Comparative Genomics Identifies Putative Signatures of Sociality in Spiders. Genome Biol. Evol. 12, 122–133 (2020).

31. 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. Scr. 36, 337–352 (2007).

32. Luo, Y. et al. Global Diversification of Anelosimus Spiders Driven by Long Distance Overwater Dispersal and Neogene Climate Oscillations. Syst. Biol. (2020) doi:10.1093/sysbio/syaa017.

33. 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).

34. Sanggaard, K. W. et al. Spider genomes provide insight into composition and evolution

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

of venom and silk. Nat. Commun. 5, 3765 (2014).

35. 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 grallator and T. californicum (Araneae: Theridiidae), with special reference to pigment genes. BMC Genomics 14, 862 (2013).

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

37. Fernández, R. et al. Phylogenomics, Diversification Dynamics, and Comparative Transcriptomics across the Spider Tree of Life. Curr. Biol. 28, 2190–2193 (2018).

38. Liu, S., Aageaard, A., Bechsgaard, J. & Bilde, T. DNA Methylation Patterns in the Social Spider, Stegodyphus dumicola. Genes 10, (2019).

39. 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).

40. 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).

41. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591 (2007).

42. Kosakovsky Pond, S. L. et al. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol. Biol. Evol. 37, 295–299 (2020).

43. Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. & Scheffler, K.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RELAX: detecting relaxed selection in a phylogenetic framework. Mol. Biol. Evol. 32, 820–832 (2015).

44. Kowalczyk, A. et al. RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics 35, 4815–4817 (2019).

45. Saputra, E., Kowalczyk, A., Cusick, L., Clark, N. & Chikina, M. Phylogenetic Permulations: a statistically rigorous approach to measure confidence in associations between phenotypes and genetic elements in a phylogenetic context. Cold Spring Harbor Laboratory 2020.10.14.338608 (2020) doi:10.1101/2020.10.14.338608.

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

47. Saputra, E., Kowalczyk, A., Cusick, L., Clark, N. & Chikina, M. Phylogenetic Permulations: A Statistically Rigorous Approach to Measure Confidence in Associations in a Phylogenetic Context. Mol. Biol. Evol. 38, 3004–3021 (2021).

48. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. 100, 9440–9445 (2003).

49. Romiguier, J. et al. Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. J. Evol. Biol. 27, 593–603 (2014).

50. Weyna, A. & Romiguier, J. Relaxation of purifying selection suggests low effective population size in eusocial Hymenoptera and pollinating bees. bioRxiv 2020.04.14.038893 (2020) doi:10.1101/2020.04.14.038893.

51. Avilés, L. Social Spiders. in Encyclopedia of Social Insects (ed. Starr, C. K.) 1–10 (Springer International Publishing, 2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

52. Viera, C. & Agnarsson, I. Parental Care and Sociality. in Behaviour and Ecology of Spiders: Contributions from the Neotropical Region (eds. Viera, C. & Gonzaga, M. O.) 351–381 (Springer International Publishing, 2017).

53. Kullmann, E. J. Evolution of social behavior in spiders (Araneae; Eresidae and Theridiidae). Am. Zool. 12, 419–426 (1972).

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

55. Wittwer, B. et al. Solitary bees reduce investment in communication compared with their social relatives. Proc. Natl. Acad. Sci. U. S. A. 114, 6569–6574 (2017).

56. Caponera, V., Avilés, L., Barrett, M. & O'Donnell, S. Behavioral Attributes of Social Groups Determine the Strength and Direction of Selection on Neural Investment. Frontiers in Ecology and Evolution 9, (2021).

57. Hori, K. et al. AUTS2 Regulation of Synapses for Proper Synaptic Inputs and Social Communication. iScience 23, 101183 (2020).

58. Oksenberg, N. & Ahituv, N. The role of AUTS2 in neurodevelopment and human evolution. Trends Genet. (2013).

59. 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).

60. Miranda, R. et al. Altered social behavior and ultrasonic communication in the dystrophin-deficient mdx mouse model of Duchenne muscular dystrophy. Mol. Autism 6, 60 (2015).

61. Daoud, F. et al. Role of mental retardation-associated dystrophin-gene product Dp71 in

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

excitatory synapse organization, synaptic plasticity and behavioral functions. PLoS One 4, e6574 (2008).

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

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

64. Bennett, M. K., Calakos, N. & Scheller, R. H. Syntaxin: a synaptic protein implicated in docking of synaptic vesicles at presynaptic active zones. Science 257, 255–259 (1992).

65. Fujiwara, T., Sanada, M., Kofuji, T. & Akagawa, K. Unusual social behavior in HPC-1/syntaxin1A knockout mice is caused by disruption of the oxytocinergic neural system. J. Neurochem. 138, 117–123 (2016).

66. Kocher, S. D. et al. The genetic basis of a social polymorphism in halictid bees. Nat. Commun. 9, 4338 (2018).

67. Braida, D. et al. Association between SNAP-25 gene polymorphisms and cognition in autism: functional consequences and potential therapeutic strategies. Transl. Psychiatry 5, e500 (2015).

68. Corradini, I., Verderio, C., Sala, M., Wilson, M. C. & Matteoli, M. SNAP-25 in neuropsychiatric disorders. Ann. N. Y. Acad. Sci. 1152, 93–99 (2009).

69. Norman, K. R. et al. The Rho/Rac-family guanine nucleotide exchange factor VAV-1 regulates rhythmic behaviors in C. elegans. Cell 123, 119–132 (2005).

70. Meusemann, K., Korb, J., Schughart, M. & Staubach, F. No evidence for single-copy

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

immune-gene specific signals of selection in termites. Front. Ecol. Evol. 8, (2020).

71. Viljakainen, L. et al. Rapid evolution of immune proteins in social insects. Mol. Biol. Evol. 26, 1791–1801 (2009).

72. Otani, S., Bos, N. & Yek, S. H. Transitional Complexity of Social Insect Immunity. Frontiers in Ecology and Evolution 4, 69 (2016).

73. Myamoto, D. T. et al. Characterization of the gene encoding component C3 of the complement system from the spider Loxosceles laeta venom glands: Phylogenetic implications. Immunobiology 221, 953–963 (2016).

74. Amdam, G. V., Page, R. E., Jr, Fondrk, M. K. & Brent, C. S. Hormone response to bidirectional selection on social behavior. Evol. Dev. 12, 428–436 (2010).

75. Dolezal, D., Liu, Z., Zhou, Q. & Pignoni, F. Fly LMBR1/LIMR-type protein Lilipod promotes germ-line stem cell self-renewal by enhancing BMP signaling. Proc. Natl. Acad. Sci. U. S. A. 112, 13928–13933 (2015).

76. Hayashi, M. et al. Conserved role of Ovo in germline development in mouse and Drosophila. Sci. Rep. 7, 40056 (2017).

77. Mével-Ninio, M., Terracol, R. & Kafatos, F. C. The ovo gene of Drosophila encodes a zinc finger protein required for female germ line development. EMBO J. 10, 2259–2266 (1991).

78. Andrews, J. et al. OVO transcription factors function antagonistically in the Drosophila female germline. Development 127, 881–892 (2000).

79. Rittschof, C. C. & Robinson, G. E. Behavioral Genetic Toolkits: Toward the Evolutionary Origins of Complex Phenotypes. Curr. Top. Dev. Biol. 119, 157–204 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

80. Woodard, S. H. et al. Genes involved in convergent evolution of eusociality in bees. Proc. Natl. Acad. Sci. U. S. A. 108, 7472–7477 (2011).

81. Roux, J. et al. Patterns of positive selection in seven ant genomes. Mol. Biol. Evol. 31, 1661–1685 (2014).

82. Wang, W. et al. Inhibiting Brd4 alleviated PTSD-like behaviors and fear memory through regulating immediate early genes expression and neuroinflammation in rats. J. Neurochem. 158, 912–927 (2021).

83. Korb, E. et al. Excess Translation of Epigenetic Regulators Contributes to Fragile X Syndrome and Is Alleviated by Brd4 Inhibition. Cell 170, 1209–1223.e20 (2017).

84. Delsuc, F. & Tilak, M.-K. Naked but not Hairless: the pitfalls of analyses of molecular adaptation based on few genome sequence comparisons. Genome Biol. Evol. 7, 768–774 (2015).

85. Enard, W. The Molecular Basis of Human Brain Evolution. Curr. Biol. 26, R1109–R1117 (2016).

86. Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M. & Andolfatto, P. Parallel molecular evolution in an herbivore community. Science 337, 1634–1637 (2012).

87. Parker, J. et al. Genome-wide signatures of convergent evolution in echolocating mammals. Nature 502, 228–231 (2013).

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

89. 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. U. S. A.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

109, 4556–4561 (2012).

90. McGlothlin, J. W. et al. Historical Contingency in a Multigene Family Facilitates Adaptive Evolution of Toxin Resistance. Curr. Biol. 26, 1616–1621 (2016).

91. Short, A. E. Z., Dikow, T. & Moreau, C. S. Entomological Collections in the Age of Big Data. Annu. Rev. Entomol. 63, 513–530 (2018).

92. Gough, H. M., Allen, J. M., Toussaint, E. F. A., Storer, C. G. & Kawahara, A. Y. Transcriptomics illuminate the phylogenetic backbone of tiger beetles. Biol. J. Linn. Soc. Lond. 129, 740–751 (2020).

93. Bazinet, A. L., Cummings, M. P., Mitter, K. T. & Mitter, C. W. Can RNA-Seq resolve the rapid radiation of advanced moths and butterflies (Hexapoda: Lepidoptera: Apoditrysia)? An exploratory study. PLoS One 8, e82615 (2013).

94. Aird, S. D. et al. Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (Ovophis okinavensis and Protobothrops flavoviridis). BMC Genomics 14, 790 (2013).

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

96. Nurk, S. et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J. Comput. Biol. 20, 714–737 (2013).

97. 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).

98. Xue, W. et al. L_RNA_scaffolder: scaffolding genomes with transcripts. BMC Genomics 14, 604 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

99. Zhu, B.-H. et al. PEP_scaffolder: using (homologous) proteins to scaffold genomes. Bioinformatics 32, 3193–3195 (2016).

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

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

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

103. Bucek, A. et al. Evolution of Termite Symbiosis Informed by Transcriptome-Based Phylogenies. Curr. Biol. 29, 3728–3734.e4 (2019).

104. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60 (2015).

105. Ebersberger, I., Strauss, S. & von Haeseler, A. HaMStR: profile hidden markov model based search for orthologs in ESTs. BMC Evol. Biol. 9, 157 (2009).

106. Oakley, T. H., Wolfe, J. M., Lindgren, A. R. & Zaharoff, A. K. Phylotranscriptomics to bring the understudied into the fold: monophyletic ostracoda, fossil placement, and pancrustacean phylogeny. Mol. Biol. Evol. 30, 215–233 (2013).

107. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539 (2011).

108. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

109. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589 (2017).

110. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).

111. Klopfenstein, D. V. et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci. Rep. 8, 10872 (2018).

112. Chikina, M., Robinson, J. D. & Clark, N. L. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. Mol. Biol. Evol. 33, 2182–2192 (2016).

113. Rubin, B. E. R., Jones, B. M., Hunt, B. G. & Kocher, S. D. Rate variation in the evolution of non-coding DNA associated with social evolution in bees. Philos. Trans. R. Soc. Lond. B Biol. Sci. 374, 20180247 (2019).

114. Schliep, K. P. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593 (2011).

115. Tong, C., Linksvayer, T. Genomic signatures of recent convergent transitions to social life in spiders. Zenodo. https://doi.org/10.5281/zenodo.7222296 (2022).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure legend

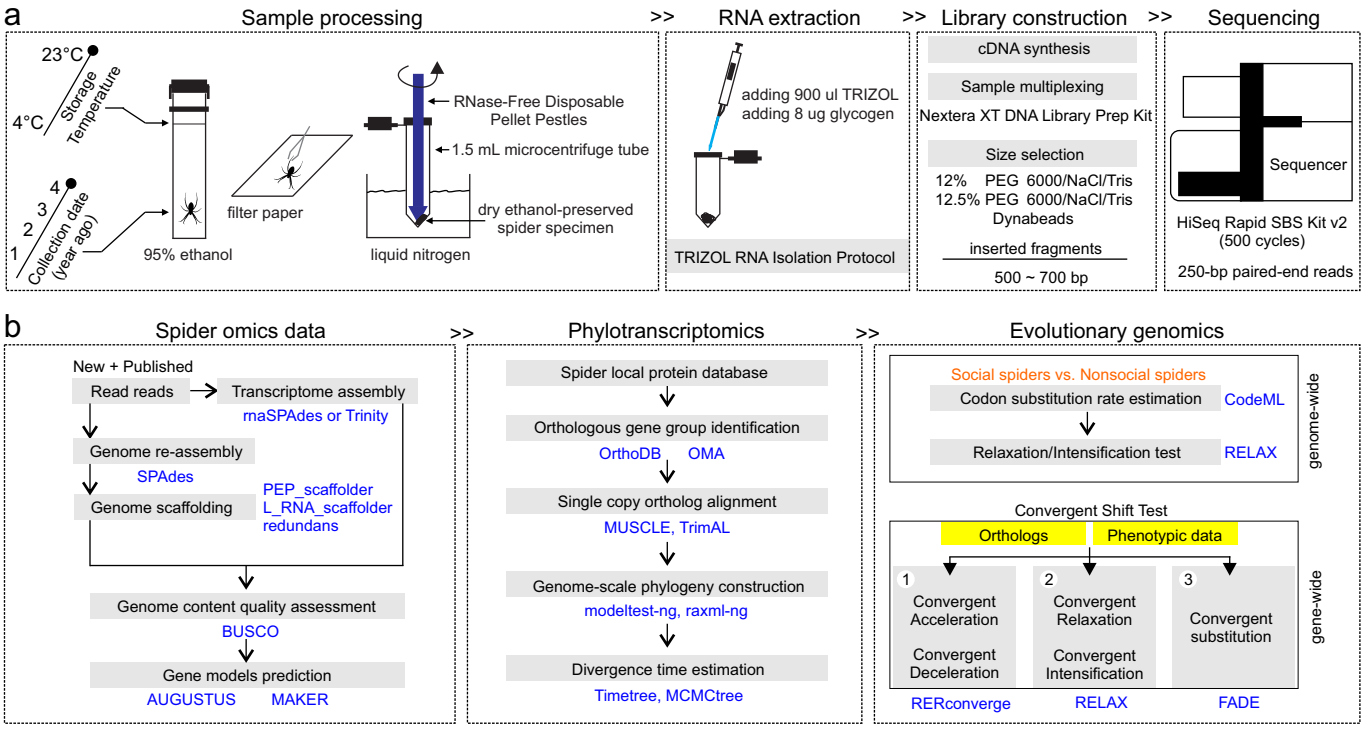
Fig 1: The flowchart represents the analysis pipeline. (**a**) sample processing and transcriptome sequencing; (**b**) spider transcriptome and genome assembly, annotation, phylotranscriptomic analyses, comparative genomic analyses and evolutionary analyses. Technical details of the workflow are provided in the Methods section.

Fig 2: **Phylogeny of study species and genome-wide pattern of molecular evolution in social and nonsocial branches.** (**a**) the maximum likelihood (ML) phylogenetic tree with estimated divergence time of the 22 spider species included in the study. The ML tree was inferred from 3,832 core shared single-copy orthologs. Bootstrap values are indicated along the branches. The divergence times at the nodes were estimated using four calibrations indicated with solid black dots. Median age estimates and 95% highest posterior densities (Mya) are shown for each node. Q and P represent the Quaternary Period and the Pliocene Epoch, respectively. Four lineages of spiders are distinguished by dark columns: genera *Anelosimus*, *Theridion*, *Stegodyphus* and the family Sparassidae. Orange dots represent social spider species, light blue dots represent nonsocial species including prolonged subsocial, subsocial and solitary spider species, the arrow tails represent the independent origins of sociality in spiders. Pictures of social spiders (from top to bottom) from *Stegodyphus dumicola* (photo credit: Noa Pinter-Wollman), *Theridion nigroannulatum* and *Anelosimus eximius* (photo credit: Leticia Avilés). (**b**) violin plot depicting the genome-wide pattern of molecular evolution (dN/dS) between social (n = 8) and nonsocial spider (n = 14) branches across the phylogeny (horizontal bars indicate 95 % CI of the means). P value was calculated by using Wilcoxon rank-sum test. Social spiders experienced convergent elevated genome-wide molecular evolution during the transition to social life. (**c**) multi-bar plot depicting the patterns of selection experienced across sites in the genomes of social and nonsocial spiders, estimated with the Partitioned Descriptive model in RELAX. P value as calculated by using Likelihood-ratio test (LRT). The distribution of dN/dS across sites in the genomes are illustrated by three categories of dN/dS for social (orange) and nonsocial (light blue) branches. The vertical dashed line at dN/dS = 1 represents neutral evolution, bars at dN/dS > 1 represent sites experiencing positive selection, and bars at dN/dS <1 represent sites experiencing purifying selection. The arrows show the direction of change in dN/dS between nonsocial and social branches, demonstrating relaxation of purifying and positive selection associated with the transition to sociality. K < 1 indicates genome-wide relaxation of selection.

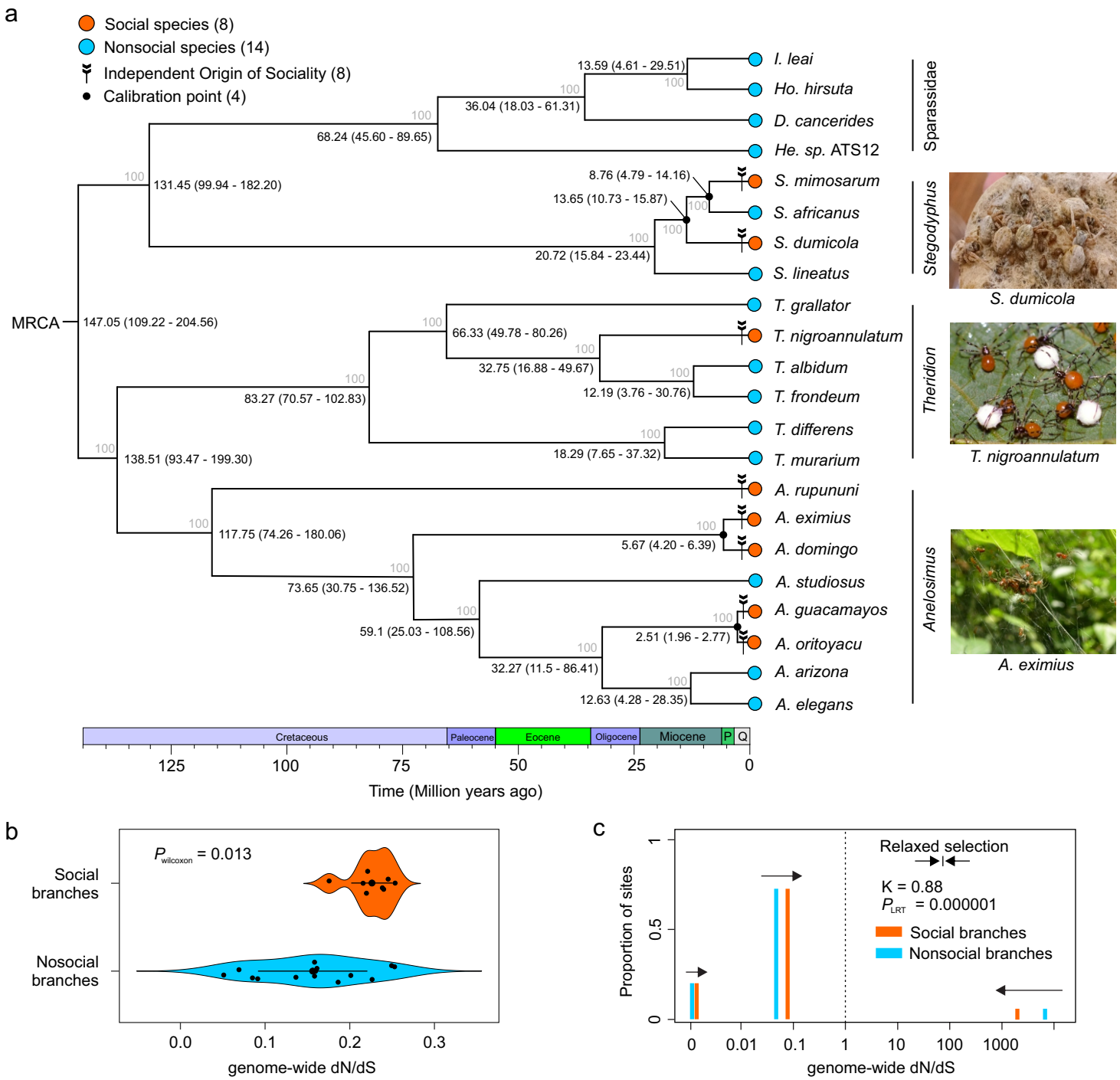
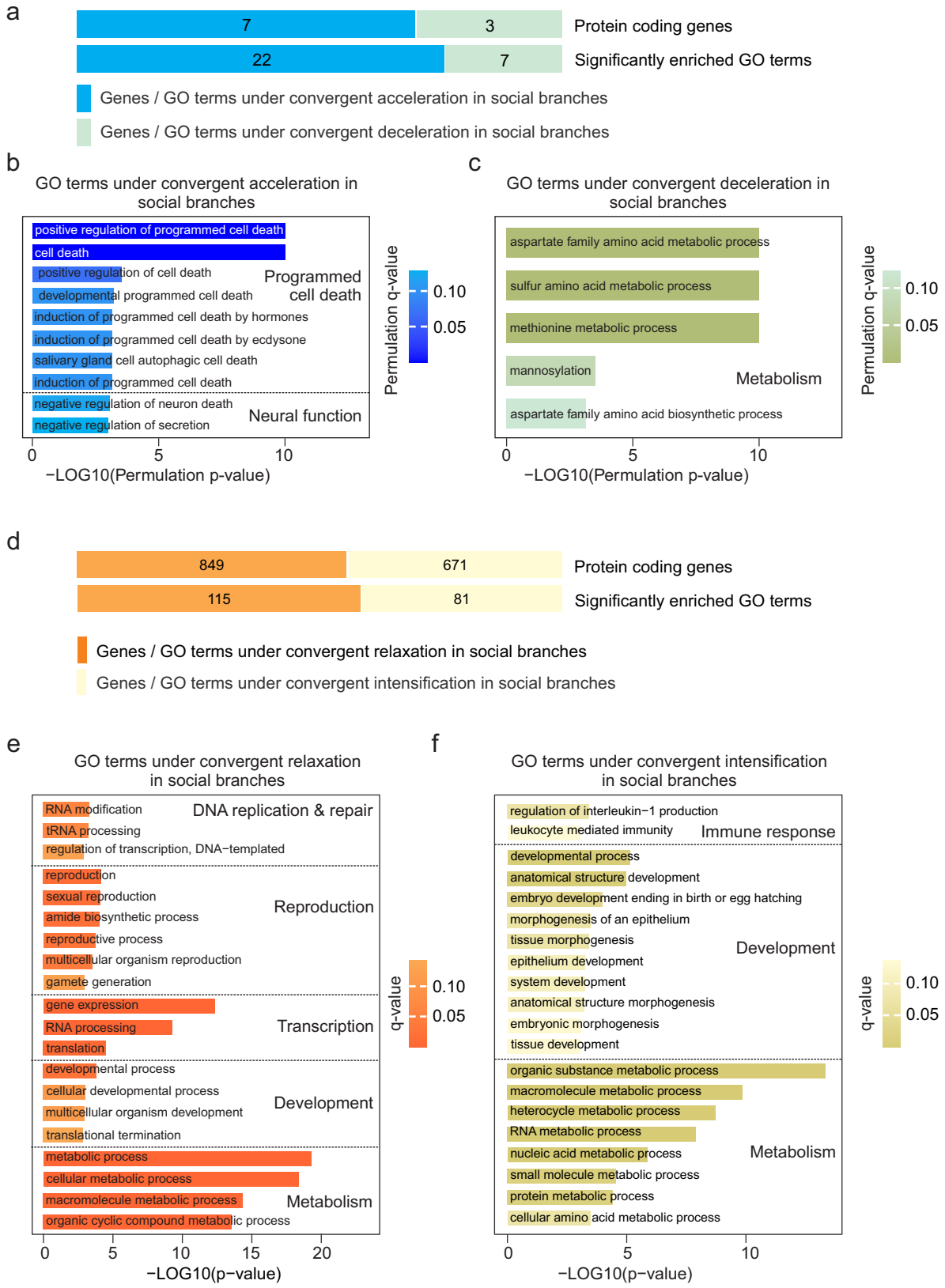
bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (permulation *q*-value < 0.15) experiencing convergent acceleration or deceleration in social branches identified by RERconverge. 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 (*q*-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 (*q*-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 (*q*-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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2021.01.27.428473; this version posted October 21, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.