ZOOLOGICAL RESEARCH

Comparative transcriptomics highlights convergent evolution of energy metabolic pathways in group-living spiders

Han Yang[1,4], Bin Lyu[2], Hai-Qiang Yin[3,*], Shu-Qiang Li[1,4,*]

[1] *Key Laboratory of Zoological Systematics and Evolution*, *Institute of Zoology*, *Chinese Academy of Sciences*, *Beijing* 100101, *China*

[2] *Chengdu Institute of Biology*, *Chinese Academy of Sciences*, *Chengdu*, *Sichuan* 610041, *China*

[3] *College of Life Science*, *Hunan Normal University*, *Changsha*, *Hunan* 410081, *China*

[4] *College of Life Sciences*, *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*

ABSTRACT

Although widely thought to be aggressive, solitary, and potentially cannibalistic, some spider species have evolved group-living behaviors. The distinct transition provides the framework to uncover group-living evolution. Here, we conducted a comparative transcriptomic study and examined patterns of molecular evolution in two independently evolved group-living spiders and twelve solitary species. We report that positively selected genes among group-living spider lineages are significantly enriched in nutrient metabolism and autophagy pathways. We also show that nutrient-related genes of group-living spiders convergently experience amino acid substitutions and accelerated relative evolutionary rates. These results indicate adaptive convergence of nutrient metabolism that may ensure energy supply in group-living spiders. The decelerated evolutionary rate of autophagy-related genes in group-living lineages is consistent with an increased constraint on energy homeostasis as would be required in a group-living environment. Together, the results show that energy metabolic pathways play an important role in the transition to group-living in spiders.

Keywords: Autophagy; Cannibalism; Convergent; Group-living; Nutrient; Spider

INTRODUCTION

The emergence of stable social groups is a key transition in evolutionary biology (Maynard Smith & Szathmáry, 1995) and is regarded as a survival response to evolutionary stress, such as predation (Hamilton, 1971), low temperatures (Gilbert et al., 2010), and food resources (Ward & Zahavi, 1973). To study the genetic changes in the process of the transitioning from solitary to group-living, we must have a phylogenetic framework that includes an initial solitary ancestor and the subsequent emergence of group-living lineages. Group-living spiders are a biological novelty and can be used as exemplars to trace adaptations that may lead to the transition. Spiders are known to be carnivorous, solitary, and potentially cannibalistic predators: Out of 49 180 spider species (WSC, 2021), only a few dozen have evolved group-living behaviors, including permanent aggregation and suppressed cannibalism (Riechert, 1985).

Group-living spiders, including cooperative and colonial

Received: 29 December 2020; Accepted: 09 March 2021; Online: 09 March 2021

Foundation items: This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31000000) to S.Q.L

*Corresponding authors, E-mail: yinhaiqiang@hunnu.edu.cn; lisq@ioz.ac.cn

DOI: 10.24272/j.issn.2095-8137.2020.281

spiders (Whitehouse & Lubin, 2005), form stable groups in several generations. The evolution of group living is thought to be related to two alternative routes, the hypotheses of which were derived from the hymenopteran model of eusocial evolution (Michener, 1958; Wilson, 1971). Cooperative spiders are characterized by cooperative prey capture, nest construction, and brood care. In cooperative spiders, group living is thought to have evolved from solitary ancestors via maternal care (Yip & Rayor, 2014), or simply called the "sub-social route" (Whitehouse & Lubin, 2005; Wilson, 1971). Subsocial spiders are in the intermediate stage of the "sub-social route", and offspring in the colonies might aggress their siblings and disperse prior to adult. Colonial spiders maintain individual territories within the colony that may last for several generations. Group-living in colonial spiders is likely to have evolved from aggregations around resources, or the "para-social route" (Whitehouse & Lubin, 2005; Wilson, 1971). Genetic changes of spiders that occur in "sub-social route" and "para-social route" may have broader significance among other social species that have evolved along these two routes.

Animals can benefit from group living in terms of reproduction, protection, and foraging functions (Whitehouse & Lubin, 2005). Research of social invertebrates has focused on several eusocial insect species and the reproductive function, such as cooperative brood care and reproductive division of labor. The importance of these functions in particular groups is underestimated. Group-living spiders benefit mostly from the foraging function (Whitehouse & Lubin, 2005) because of foraging efficiency (Buskirk, 1981; Rypstra, 1989), capture of larger prey (Nentwig, 1985; Rypstra, 1990; Ward, 1986), and the "ricochet effect" (Uetz, 1988, 1989). Besides the dominated foraging function (Whitehouse & Lubin, 2005), group-living spiders also benefit from the reproductive function (in cooperative spiders) (Schneider, 2002) and protective function (in both cooperative and colonial species) (Evans, 1998; Lubin, 1974).

In the past few decades, much ecological and behavioral research has investigated the evolution and maintenance of group-living in spiders (Avilés & Guevara, 2017; Guevara & Avilés, 2015; Kim, 2000; Majer et al., 2013; Riechert et al., 1986; Toyama, 1999; Wise, 2006). Spiders with group-living behaviors mostly exist in tropical areas (Avilés & Guevara, 2017; Majer et al., 2013), and the existing hypotheses for this phenomenon include food resources and group defense (Avilés & Guevara, 2017; Guevara & Avilés, 2015; Majer et al., 2013; Riechert et al., 1986). Research has also focused on behavior in spider societies. Although young spiders may disperse or cannibalize fellow members of their species because of food resource competition, the initial phase of tolerance when spiders spend their first instar with siblings in the egg sac can be regarded as the basis of a longer-lasting aggregation (Shear, 1970). The critical part of forming stable groups is the reduction of competition (Korb & Heinze, 2016). Previous studies have found that maternal care (Toyama, 1999; Wise, 2006) and food abundance (Kim, 2000) can prolong the initial phase of tolerance after the yolk exhaustion. The genetic mechanism of high tolerance and reduced competition is important for adaptation to group-living.

The behavioral and ecological research of spider societies has a rich history, but little work has been carried out on the genetic mechanisms of group-living evolution in spiders (Tong et al., 2020). Recent study on genus *Stegodyphus* revealed that rapidly evolving genes in these cooperative velvet spiders were enriched for transport function, behavior, and immune response processes when compared with other solitary spiders (Tong et al., 2020). Genomic tools have answered puzzling questions about social evolution (Kapheim et al., 2015) and the division of labor (Libbrecht et al., 2018) in social insects. Studies have found that rapidly evolving genes in primitively eusocial insects are enriched in neuron differentiation, reproduction, and immunity, and those of highly eusocial insects were enriched in metabolism, nutrient, and immunity (Fischman et al., 2011; Kapheim et al., 2015; Woodard et al., 2011).

Group-living in spiders is no more complex than quasi-social in the terminology of the insect societies (Michener, 1969; Wilson, 1971), remaining at an early stage of social evolution. The evolution of group-living behavior in spiders has convergently evolved multiple times in distantly related genera (Fernández et al., 2018; Whitehouse & Lubin, 2005), providing fertile ground for examining convergent evolution of the transition to group-living. Group-living spiders benefit from foraging, protection, reproduction (Whitehouse & Lubin, 2005), and energy or water reserves (Vanthournout et al., 2016), so they have to adapt to the group-living environment and repress the cannibalistic instinct. Thus, distantly related group-living lineages may convergently show similar adaptive genetic signatures under similar social pressure. The study of group-living spiders can help to resolve the initial social transition from solitary to group-living.

To understand the transition from solitary to group-living, we conducted a comparative transcriptomic study involving fourteen spider species, twelve solitary and two group-living (Figure 1; Supplementary Table S1). Similar group-living traits, permanent aggregation and suppressed cannibalism, have convergently evolved due to similar social selection pressures in two distantly related spider species, the cooperative velvet spider *Stegodyphus dumicola* (family Eresidae) (Liu et al., 2019) and the colonial hackled-orb-weaver *Philoponella alata* (family Uloboridae) (Lin & Li, 2008). Many species of these two genera have group-living tendencies (Whitehouse & Lubin, 2005). Incipient sociality in aculeate Hymenoptera forms a simple social organization, and is predicted to be related to positive selection and protein evolution (Rehan & Toth, 2015). Considering the stable distinction of behaviors between the two group-living and twelve solitary species, the mechanisms of the initial social transition may also involve positive selection and protein evolution based on the empirical prediction. Thus, we conducted three analyses using transcriptomic data: selective pressure analysis, convergence analysis, and relative evolutionary rates analysis.

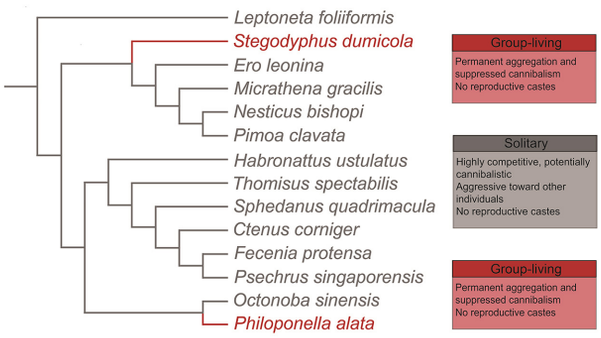

Figure 1 Topology of samples from the group-living dataset
Lineages are colored by social categories: red, group-living; grey, solitary. Traits of the social categories are listed.

MATERIALS AND METHODS

Sampling

We collected *P. alata* from Xishuangbanna Tropical Botanic Garden, Yunnan Province, China (N21.92°, E101.25°, 559 m a.s.l.) and *Octonoba sinensis* (family Uloboridae) from Olympic Forest Park, Beijing, China (N40.01°, E116.39°, 45 m a.s.l.). Before sequencing, samples were starved for 48 hours in an artificial climate incubator. Total RNA was extracted separately from the whole body of *P. alata* and *O. sinensis* using TRIzol (Invitrogen, USA) according to the manufacturer's instructions. We constructed a "group-living dataset" to explain the transition from solitary to group-living. Data from the other twelve species, including eleven transcriptomes and coding sequences of the *S. dumicola* genome, were downloaded from NCBI. To study the formation of stable group, we selected transcriptomes and genomes from distinctly solitary and group-living spiders. Subsocial species, regarded as intermediate between solitary and cooperative species, were not included in the group-living dataset (Yip & Rayor, 2014). Spider species with ambiguous social categories, such as pre-adaptation to subsociality by maternal care, were also excluded. Detailed information is listed in Supplementary Table S1. For spider transcriptomes, we only selected those from Entelegynae spiders with whole body transcriptomes that were larger than 5G bases (Supplementary Table S1). Leptonetidae, the sister group to the Entelegynae (Fernández et al., 2018), was selected as the outgroup.

Transcriptome sequencing and assembly

RNA quality and quantity of the two Uloboridae samples were assessed using TAE-agarose gels, NanoDrop 2000 (NanoDrop Technologies, USA), and the Agilent Bioanalyzer 2100 system (Agilent Technologies, USA). Libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). Subsequently, sequencing was conducted on an Illumina HiSeq 4000 platform with a paired-end read length of 150 bp. Both library construction and Illumina sequencing were performed at Novogene Bioinformatics Technology Co., Ltd., Beijing, China. The raw sequence reads of all transcriptomes were filtered by removing adapters and reads with more than 40% low-quality bases (<Q15) (Chen et al., 2018). Clean reads were evaluated using FastQC v0.11.6 (Andrews, 2010) and then assembled with Trinity v2.5.0 (Grabherr et al., 2011) using the default settings. Data are available on NCBI under BioProject No. PRJNA554940. Assembled transcriptomes with BUSCO *C*-values lower than 90% were regarded as low quality and discarded. Candidate coding regions of assembled transcripts were identified using the TransDecoder v3.0.0 (Grabherr et al., 2011) program implemented in the Trinity software. We used CD-HIT-EST v4.6.6 (Fu et al., 2012) to cluster candidate coding genes with a threshold of 0.98 and to generate representative sequences to reduce redundancy.

Ortholog identification and alignment

Translated coding sequences of the fourteen spiders were used for reciprocal best hits (Moreno-Hagelsieb & Latimer, 2008) against proteins of *Drosophila melanogaster* (Zerbino et al., 2018) (downloaded from the Ensembl database) using blastp with an *e*-value threshold of 1e–10. One-to-one orthologs were identified when sequences of the fourteen spider species matched to one reference gene. Orthologous nucleotide sequences were then retrieved using SAMtools v1.3 (Li et al., 2009), according to the ortholog entry, and used for alignment.

Alignment errors may influence subsequent analyses, so we used stringent alignment procedures. Guidance v2.02 (Penn et al., 2010) was used to weight the protein alignment generated by PRANK v150803 (Löytynoja & Goldman, 2005), and residues with a low score (≤ 0.93) were removed. Thirty bootstrap iterations were conducted instead of the default 100 for computational efficiency. Poorly aligned columns with more than 10% gaps or with a similarity score below 0.001 were trimmed by trimal v1.4.1 (Capella-Gutiérrez et al., 2009), and species with more than 5% missing data were removed from further analyses. Aligned protein sequences were then back-translated to the corresponding codon alignment based on the previous codon sequences. Finally, nucleotide sequences shorter than 150 bp were discarded.

Phylogenetic inference

Aligned nucleotide sequences were concatenated to produce a supermatrix, and 4-fold degenerate sites were extracted to conduct phylogenetic inference. A maximum-likelihood phylogenetic tree was reconstructed using RAxML v8.2.9 (Stamatakis, 2014), with a GTRGAMMA model and 500 rapid bootstrap iterations. An unrooted tree of each gene was built using the previous parameters in RAxML, and then the multi-species coalescent model in ASTRAL v4.10.2 was used to infer a species tree based on these gene trees (Mirarab et al., 2014).

Identification of positively selected genes

The branch-site model was used to detect the signature of positive selection in group-living spider species in PAML v4.9 (Yang, 2007; Zhang et al., 2005). Both group-living branches,

P. alata and *S. dumicola*, were labeled as foreground. In the null model A1, foreground sites are neutral or under purifying selection, while model A allows sites to be under positive selection. Likelihood ratio tests were performed to compare model A to the null model A1. *P*-values were calculated using a chi-square distribution with 1 degree of freedom. A false discovery rate (FDR) cutoff of 0.05 (Storey & Tibshirani, 2003) was used in multiple test correction. We further investigated the selective pressure on group-living spiders using aBSREL (Smith et al., 2015) with a priori specification of group-living branches.

Convergence analysis

We tested for molecular convergence that may be found in group-living spiders. There are two types of molecular convergence: convergent and parallel amino acid substitutions. Convergent amino acid substitutions refer to changes of different amino acids along two independent evolutionary lineages resulting in the same amino acid, while parallel substitutions start with same amino acid in two lineages. We hereafter refer to both types as convergent, unless otherwise stated. Branch model in PAML was used to reconstruct the ancestral protein sequences. The output file was modified to use as the input file of Zhang's script (Zhang & Kumar, 1997; Zou & Zhang, 2015) to calculate the random expectation of convergent amino acid substitutions. We used a Poisson distribution to compare the expected value to the observed with a false discovery rate (FDR) cutoff of 0.05 (Storey & Tibshirani, 2003). We used the result of the JTT-f_{gene} model because the JTT-f_{site} model will overestimate the expected values when only a few species are analyzed (Zou & Zhang, 2015). Genes in which convergent amino acid changes in group-living lineages significantly exceeded expectations were considered convergent candidates.

Relative evolutionary rates change

Relative evolutionary rates were obtained using the projection operator (Sato et al., 2005). The branch lengths of fourteen species were calculated by the Empirical+F model in PAML. A distance matrix was constructed with branch lengths of these fourteen species for all remaining genes and then transformed into phylogenetic vectors. The projection operator averaged the phylogenetic vectors to extract the phylogenetic information. Finally, phylogenetic relationships were excluded from the phylogenetic vectors, and the residual vectors were used to represent the relative evolutionary rates without the phylogenetic factors that affect all genes. Rates analysis was complemented by the R package RERconverge (Kowalczyk et al., 2019). For each gene, relative rates of group-living branches were compared with solitary branches using the Wilcoxon rank-sum test (P<0.05). Genes related to group-living should convergently experience rate accelerations or decelerations in the analysis.

Analysis of route dataset

We also constructed a "route dataset" to trace the genetic pattern along the "sub-social" and "para-social" routes. There were four group-living spiders in the route dataset: two cooperative species, *S. dumicola* and *S. mimosarum* (sub-social route) and two colonial species, *Metepeira labyrinthea* and *P. alata* (para-social route). Coding sequences of *S. mimosarum* (Sanggaard et al., 2014) and transcriptomes of two subsocial relatives, *S. lineatus* and *S. africanus* (Bechsgaard et al., 2019), were downloaded from NCBI. The transcriptome of the subsocial *S. tibialis* has been uploaded to NCBI. Detailed information of the four group-living species and the other eight species can be found in Supplementary Table S1. After ortholog identification and alignment, we reconstructed gene trees using the previous parameters in RAxML. The species tree was inferred by the multi-species coalescent model in ASTRAL v4.10.2 based on the gene trees. To verify the pattern of rate accelerations or decelerations in the group-living dataset, we implemented the relative evolutionary rates analysis along the four group-living branches, the sub-social route, and the para-social route, respectively.

Enrichment analysis

To analyze the function of the candidate genes, we used DAVID v6.8 (Huang et al., 2009) to perform the enrichment analysis using Fisher's exact test (P<0.05). For functional enrichment, categories of the Gene Ontology (GO_BP_Direct) and Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway were used with a count threshold of 2.

RESULTS

Positive selection in group-living spiders

We identified 3 793 raw orthologous sequences for fourteen spider species using a reciprocal best-hit method. After multiple sequence alignments, columns with gaps and low confidence were removed. Nucleotide sequences shorter than 150 bp were discarded. Genes with an overall dS>1 were inferred to be saturated and excluded from further analyses. The species tree based on the remaining 2 916 phylogenetic trees was congruent with recent phylogenomic research (Figure 1) (Cheng & Piel, 2018; Fernández et al., 2018).

The branch-site model indicated twenty genes were inferred to be under positive selection in both group-living spider species (q<0.05). GO and KEGG enrichment analyses indicated six significant terms/pathways, GO:0048102~autophagic cell death, GO:0035071~salivary gland cell autophagic cell death, GO:0051603~proteolysis involved in cellular protein catabolic process, GO:0006508~proteolysis, dme01200:Carbon metabolism, and dme01230:Biosynthesis of amino acids (Supplementary Table S2). These terms/pathways were related to energy metabolism, including autophagy and nutrient metabolism (Figure 2). The network in Figure 2 was implemented in Cytoscape (Shannon et al., 2003). To further investigate the selective pressures of group living, we examined episodic positive selection using aBSREL. Among the result of two genes, *CG12163* was related to autophagy and *Ald1* was

related to nutrient metabolism according to the functional annotation in DAVID v6.8.

Convergent amino acid substitutions involved in group-living

The observations and expectations of convergent amino acid substitutions in group-living lineages were compared using a Poisson distribution (q<0.05). Fifty-four genes with significantly higher convergent substitutions than random showed convergent signatures, and enrichment analysis discovered significant GO biological processes and KEGG Pathways involved in nutrient metabolism and autophagy (Supplementary Table S3). Furthermore, one immunity-related KEGG Pathway Biosynthesis of antibiotics (dme01130) was also significantly enriched in convergence analysis.

Adaptively convergent genes were defined as positively selected genes with significantly higher convergent substitutions than random. To investigate the conservative signatures of adaptive convergence in group-living spiders, we focused on the function of thirteen adaptively convergent genes (Table 1). We found that three genes (*CtsF*, *CtsB1*, and *Eip63F-1*) were related to autophagy, and three genes (*Cyt-b5*, *Had1*, and *Ald1*) were related to nutrient metabolism. The functional annotation of *galectin* is synaptic target recognition in nervous system.

Pattern of relative evolutionary rates

We further investigated the molecular evolutionary patterns in group-living spiders. Relative evolutionary rates analysis based on a projection method (Sato et al., 2005) was used to detect the convergent changes of the protein evolution rate in group-living lineages. We propose that convergent changes of protein evolution rate can be found when these genes are related to the convergent group-living behaviors, permanent aggregation and suppressed cannibalism. When the group-living environment selects a particular phenotype, the gene under the phenotype will experience the convergent selective pressure in two group-living lineages and will result in convergent increase of protein evolutionary rate. Alternatively, a group-living environment may also apply extra constraints on a gene because of its more important function that is required to survive in the group-living environment, resulting in a convergently slower evolutionary rate in the two group-living lineages, *P. alata* and *S. dumicola*. After excluding the phylogenetic relationships of each orthologous gene, the residual vector was used to represent the relative evolutionary

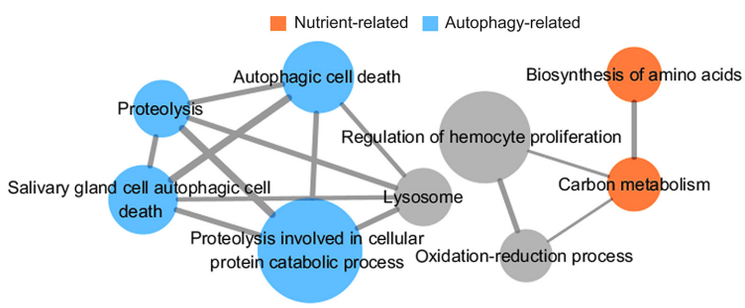

Figure 2 Network of enriched biological processes of twenty positively selected genes in group-living spiders

Nodes represent significantly enriched biological processes, and the node diameter reflects the degree of enrichment (ranging from a 3.92 to a 36.97 fold change). Edge width between nodes reflects the proportion of shared positively selected genes (ranging from 20% to 100%). The enriched results are clearly related to nutrient metabolism (orange) and autophagy (blue). Grey nodes represent other biological processes.

Table 1 Adaptively convergent genes in group-living spiders

Biological Process	Gene symbol	Description	Q-value
Autophagy	*CtsF*	Peptidase activity	3.44E–17
	CtsB1	Peptidase activity	1.96E–15
	Eip63F-1	Autophagic cell death	0.000 2
Nutrient metabolism	*Cyt-b5*	Lipid metabolism	7.64E–05
	Had1	Fatty acid metabolic process	0.000 5
	Ald1	Glycolytic process	0.044 1
Nervous system	*galectin*	Synaptic target recognition	4.35E–14
Others	*Cat*	Catalase activity	1.77E–15
	ERp60	Protein disulfide-isomerase	1.08E–09
	Achl	Gene regulation	3.47E–09
	yps	Regulation of transcription	3.12E–05
	eIF5	Translation initiation factor activity	0.001 6
	RpL27	Translation	0.002 7

rate.

To classify the function of gene groups in the evolutionary rates analysis, we identified the GO terms (direct biological process) and KEGG Pathways according to the functional annotation of all background orthologs in DAVID v6.8 (Supplementary Table S4). Considering previous hypotheses of social evolution and results of the former two analyses, four modules were chosen for classification: autophagy, immunity, nerve, and nutrient. For 73 genes, relative rates of two group-living branches convergently increased compared to the other twelve solitary branches (Wilcoxon rank-sum test P<0.05). Significantly enriched terms and pathways of these accelerated genes were regarded as accelerated terms and were mainly involved in nutrient metabolism and immunity (Figure 3; Supplementary Table S5). Accelerated terms showed significant overrepresentation in the nutrient module (6/72, Fisher's exact test P=5.73E–5) and immunity module (2/22, Fisher's exact test P=0.0212) when compared to all background orthologs (19/1813) (Figure 4; Supplementary Table S6). Two group-living branches of 86 genes evolved slowly when compared to the other twelve branches (Wilcoxon rank-sum test P<0.05). Significantly enriched terms and pathways were mainly involved in autophagy and the nervous system. (Figure 3; Supplementary Table S7). In a decelerated

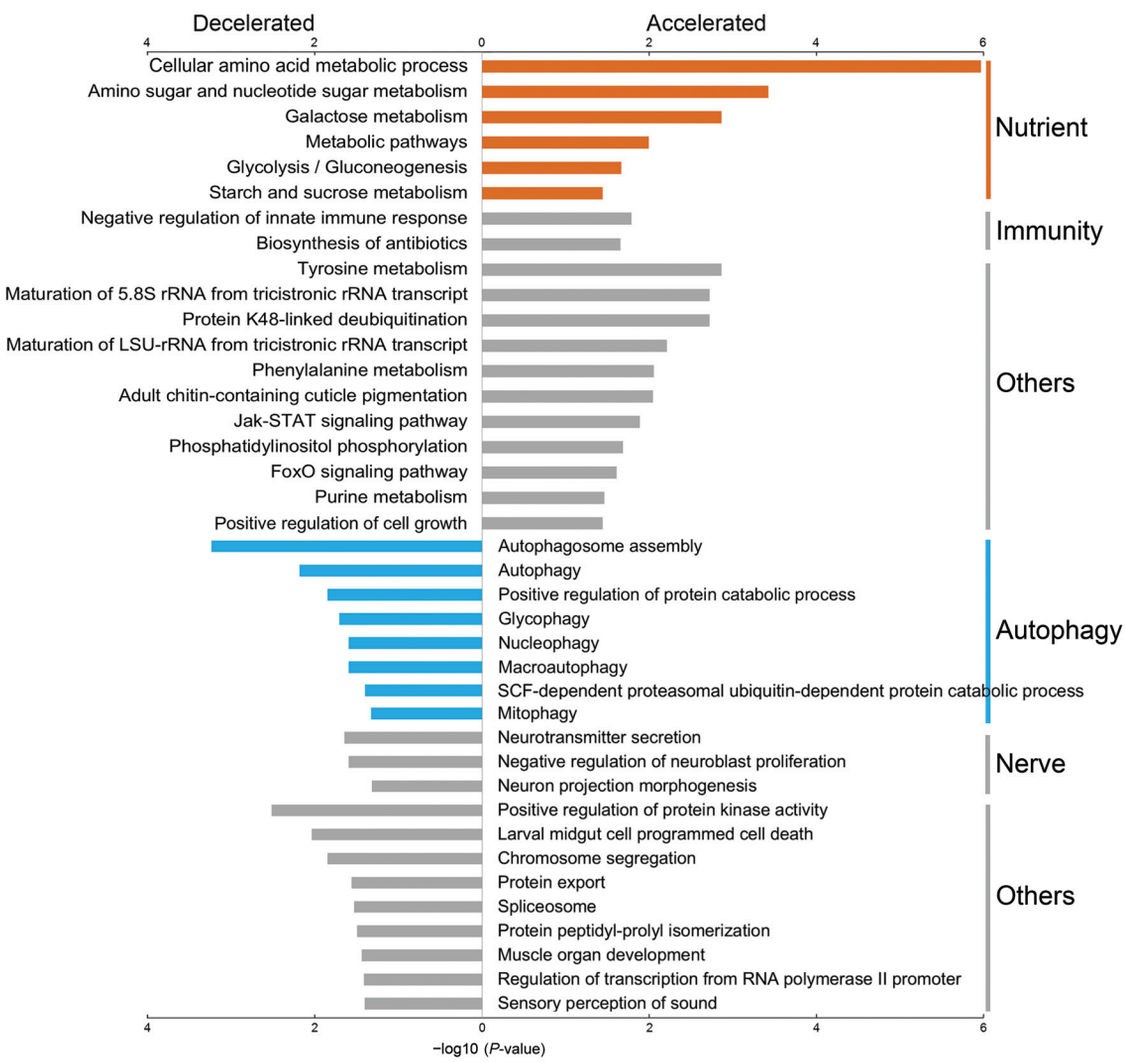

Figure 3 Enriched biological processes or KEGG Pathways of accelerated and decelerated genes in group-living spiders

The upper three categories show the enriched result of accelerated genes in group-living lineages, while the lower three categories represent decelerated terms/pathways. To differentiate them, bars of these two parts are in opposite directions. The X-axis reflects the negative of the log base 10 of the P-value calculated based on significantly enriched terms/pathways. Orange, nutrient metabolism pathways; blue, autophagy pathways; grey, other biological processes.

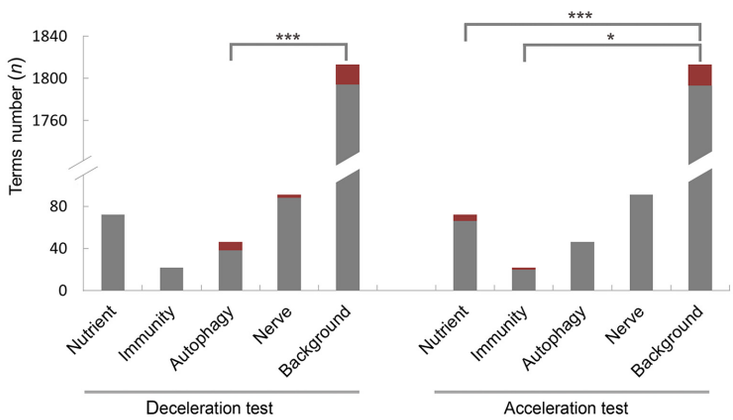

Figure 4 Overrepresentation test of four modules in deceleration/acceleration datasets when compared with the background

Red, accelerated/decelerated terms (according to the test dataset); grey, other terms in the module. *: P<0.05, **: P<0.01, ***: P<0.001.

test, terms showed significant overrepresentation in the autophagy module (8/46, Fisher's exact test P=9.19E–9) when compared to all background orthologs (20/1 813) (Figure 4; Supplementary Table S6). A similar pattern was also observed at the gene level (Supplementary Table S8).

Nutrient processes of group-living spiders are potentially under similar selective pressures in group-living environments, and the adaptive evolution of the nutrient module can ensure energy supply, increase tolerance to a nutrition deficit, and repress cannibalistic instincts. Thus, nutrient-related genes can adapt to environmental change and result in a more rapid rate of sequence evolution. The decelerated rates of autophagy-related genes accord with the important function of energy homeostasis in group-living spiders and show increased evolutionary constraint.

Result of route dataset

The species tree based on the 3 078 phylogenetic trees was in agreement with recent research (Figure 5) (Fernández et al., 2018; Settepani et al., 2016). Relative evolutionary rates in the route dataset were calculated by the projection operator. We then compared the relative rates of four group-living branches with those of solitary or subsocial species. Fourteen genes that were accelerated in four group-living branches were significantly enriched in dme04146:Peroxisome, which plays a key role in lipid metabolism and is related to nutrient metabolism (Table 2; Supplementary Tables S9, S10). Significantly enriched terms of accelerated genes in two cooperative branches (sub-social route) contained dme04146:Peroxisome, dme01212:Fatty acid metabolism, and dme00071:Fatty acid degradation. For acceleration in colonial branches (para-social route), significantly enriched terms and pathways included GO:0010906~regulation of glucose metabolic process.

Thirty-nine genes showed deceleration in four group-living branches, and the significantly enriched terms and pathways included GO:0034198~cellular response to amino acid starvation. Decelerated genes in two cooperative branches

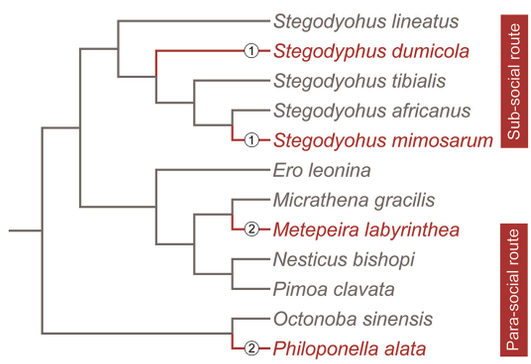

Figure 5 Topology of the route dataset

Red, group-living; grey, solitary or subsocial. Node 1, cooperative (sub-social route); node 2, colonial (para-social route).

were enriched in GO:0034198~cellular response to amino acid starvation. Significantly enriched terms of decelerated genes in two colonial branches included GO:0043162~ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway and GO:0042787~protein ubiquitination involved in ubiquitin-dependent protein catabolic process (Table 2; Supplementary Table S10). The additional route dataset also followed the pattern of the group-living dataset, accelerated nutrient metabolism and decelerated autophagy pathways in group-living branches.

DISCUSSION

Previous studies of spider social evolution have been mostly limited to behavioral and ecological aspects, and emphasize the importance of food supply (Bilde & Lubin, 2001; Rypstra, 1983, 1986; Whitehouse & Lubin, 2005) in group formation. However, the genetic changes involved in the transition from solitary to group-living are yet to be concluded (Settepani et al., 2017; Tong et al., 2020). We suggest that group-living evolution in spiders is related to nutrient metabolism and

Table 2 Accelerated and decelerated terms that are related to nutrient metabolism and autophagy pathways in the "route dataset"

Term	Count	*P*-value	*Q*-value
Accelerated in			
four group-living branches			
dme04146:Peroxisome	2	0.0107	0.0214
sub-social route			
dme04146:Peroxisome	2	0.0243	0.1945
dme01212:Fatty acid metabolism	2	0.0432	0.1729
dme00071:Fatty acid degradation	2	0.0432	0.1729
para-social route			
GO:0010906~regulation of glucose metabolic process	4	0.0080	0.0962
Decelerated in			
four group-living branches			
GO:0034198~cellular response to amino acid starvation	2	0.0065	0.0518
sub-social route			
GO:0034198~cellular response to amino acid starvation	2	0.0021	0.0148
para-social route			
GO:0043162~ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway	2	0.0101	0.1637
GO:0042787~protein ubiquitination involved in ubiquitin-dependent protein catabolic process	3	0.0277	0.2040

autophagy based on multiple lines of evolutionary evidence: positive selection, convergent substitutions, and convergent rate change. Our results indicate the importance of adequate energy in group-living emergence and prove the unification of molecular and behavioral perspectives.

Nutrient-related pathways are under adaptive evolution in a group-living environment, including positive selection, convergent substitutions, and accelerated evolution. The modification of nutrient-related genes can regulate metabolic processes and relieve cannibalistic threats or indicate a different dietary niche, to ensure the energy supply. Highly conserved in eukaryotes, autophagy pathways recycle useless and damaged cellular components and generate energy during periods of starvation (Mizushima, 2007). Studies also show that autophagy-related genes are essential to maintain energy homeostasis during the early neonatal starvation period (Kuma et al., 2004). The decreased evolutionary rate of genes related to autophagy is consistent with the increased constraint on energy homeostasis as would be required in group-living spiders. A similar pattern, accelerated nutrient metabolism and decelerated autophagy pathways, was also found in group-living branches of the additional route dataset (Table 2). The adaptative evolution of the autophagy module, indicated by selective pressure analysis and convergence analysis, is not contrary to the results of the decelerated evolutionary rates. Rather, the former two analyses focus on the convergent changes of limited amino acid sites among a sequence, while the rate analysis focuses on the whole sequence. These energy metabolic pathways decrease energy demand and facilitate tolerance for aggregation.

Besides nutrient and autophagy related genes, we also identified genes involved in immunity and the nervous system. Immunity can protect the community from pathogens and make individuals adapt well to the social environment (Cremer et al., 2018). One immunity related KEGG Pathway was enriched in the convergence analysis (Fisher's exact test *P*=0.028 0; Supplementary Table S3), and one gene involved in the nervous system showed a signature of adaptive convergence (Table 1). In the evolutionary rate analysis, accelerated terms in the immunity module showed significant overrepresentation at the term level (Fisher's exact test *P*= 0.021 2; Supplementary Table S6), but not significant at the gene level (Supplementary Table S8). Summarizing the results of the additional analysis, we found that the immunity and nerve modules were not more representative than nutrient and autophagy modules in our analyses of group-living evolution.

The first step toward sociality is the formation of stable social groups (Korb & Heinze, 2016), and this is the case in group-living spiders. The ultimate cause of group-living is increased direct fitness for individuals. The foraging function is dominant in group-living spiders (Whitehouse & Lubin, 2005), and they can benefit from increased foraging efficiency, taking full of advantage of adequate food resources (Buskirk, 1981; Powers & Avilés, 2007; Uetz, 1988). The major hurdle to the evolution of stable groups in spiders is food competition (Korb & Heinze, 2016). Because spiders are aggressive, carnivorous, and potentially cannibalistic predators, they may aggress their fellow nest members or disperse if starved, which would break the aggregation state. Group-living spiders are primarily found in tropical or subtropical regions (Avilés & Guevara, 2017; Majer et al., 2013) where there are plenty of food resources. It is well-documented that animals are more likely to be cooperative when living in harsh environments in order to cope with difficulties (Rubenstein & Lovette, 2007). However, the situation is different in carnivorous and potentially cannibalistic spiders. Habitats with adequate resources are needed to support group-living spiders, and

harsh environments will trigger aggressive behaviors and block the formation of stable social groups. Experimental manipulation of spiders has shown that sufficient food supply can effectively reduce cannibalism and promote sociality (Bilde & Lubin, 2001; Rypstra, 1986). These ecological and behavioral results indicate the importance of energy supply to the aggregation state. Spiders feed on live prey or prey that has been recently killed. Even in a habitat with adequate resources, seasonal variation or a discrete food shortage may induce cannibalism and hinder aggregation. Considering the benefits of group-living, permanent-social spiders must establish tolerance mechanisms in response to nutrient deficiencies. The internal energy metabolic pathways, rather than the external food supply, are essential for the formation of stable groups. In theory, spiders may also form stable groups when there is an unlimited energy supply (Figure 6). However, a recent study has emphasized the importance of social interaction in the route to permanent sociality, and found that reduction in the level of starvation failed to repress the aggressiveness triggered by social isolation (Chiara et al., 2019). Another study denoted that underfed colonies dispersed significantly earlier than colonies fed without restriction (Krafft et al., 1986). It's difficult to compare the importance of food supply and other factors in group-living transition. Conclusions should be made with caution when determining causal relationships in group-living evolution. Additional factors, such as developmental stage, starvation state, dispersal time, and aggressiveness, should be considered.

Although spider societies are not as successful as insect societies in terms of diversity or complexity, they, in addition to social insects, can help to understand the evolutionary trajectory of social emergence. Molecular evolutionary analyses of eusocial insects have focused on several biological processes, such as chemical signaling, brain development, immunity, reproduction, and nutrition (Fischman et al., 2011; Woodard et al., 2011). The complex result can be attributed to distinct selective forces under different processes, such as eusocial origin, maintenance, and elaboration. As a simple form of collective organization, group-living spiders are a product of social emergence without extra improvements. Unlike the diverse molecular mechanisms of insect societies, energy metabolic pathways stand out in our molecular evolutionary analyses of group-living spiders. Food storage, closely related to prolonged periods without foraging, is common in social insects, such as bees (Brodschneider & Crailsheim, 2010) and ants (Bazazi et al., 2016). Nonetheless, due to the predatory habits of spiders, food storage behavior is unlikely (Sandidge, 2003). The solution is a habitat with adequate food resources, the utilization of nutrient elements for individuals, and autophagy pathways which would occur in a harsh environment.

Group-living spiders exist mostly in tropical or subtropical regions, which introduces sampling bias and potential confounding factors, such as prey abundance and temperature. To relieve the sampling problem, we

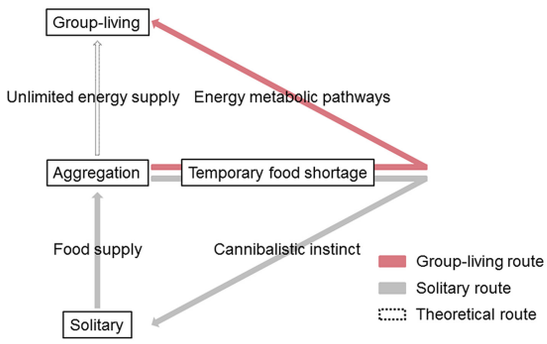

Figure 6 Hypothesis of the group-living route in spiders

implemented the Wilcoxon rank-sum test to compare the latitudes of samples (Supplementary Table S1). Group-living samples were not closer to the equator when compared with solitary samples (P=0.3896); however, we still should be cautious when interpreting the results of energy metabolic pathways. Although displaying similar group-living behaviors, permanent aggregation and suppressed cannibalism, cooperative and colonial spiders have different effective population sizes. Cooperative spiders are inbred and show depleted genomic diversity (Settepani et al., 2017), while the colonial species have outbred population structures like solitary species. Thus, in cooperative and colonial spiders, the effectiveness of selection or selection response will differ even under similar group-living selection intensities. The influence of selection response difference on the protein evolution of group-living spiders is unknown. Further study on more group-living spiders (cooperative and colonial) with high-quality sequencing data may help to reveal the genetic adaptations of group-living spiders and validate the results of the route dataset.

We propose that spider societies are promising models for studying the initial transition to sociality from solitary living. Most spiders are solitary and face fierce competition among individuals. The conflict between stable groups and cannibalistic instincts in spiders can help us to study the balance between social formation and discord. Without a caste system, spider societies are no more advanced than quasi-sociality in the terminology of insect societies (Michener, 1969; Wilson, 1971) and can be considered exemplars of non-eusocial organisms. The study of spider societies can also inspire us to think about the social strategies of other carnivores (Elbroch et al., 2017; Macdonald, 1983). Sociality in spiders has independently evolved many times (Fernández et al., 2018; Whitehouse & Lubin, 2005), and spider societies provide a promising route to determine the mechanisms underlying the first steps to sociality.

CONCLUSIONS

Our results indicate the important role of energy metabolic pathways in forming stable groups. This work outlines the genetic signatures of group-living spiders, and candidate

genes from our analyses could help to answer questions about group-living at both molecular and phenotypic levels. More research is required to understand how genetic mechanisms influence group-living behaviors, and our research provides an important starting point. The genomic era will provide broader insights to the evolutionary origin of sociality in spiders and other non-eusocial species.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

S.Q.L., H.Q.Y., and B.L. conceived and designed the study. H.Y. collected the samples and analyzed the data. H.Y. and S.Q.L. wrote the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Zhong-E Hou, Xin Zhou, Bao-Cheng Guo, and Ke-Xin Li for helpful discussion. In addition, we thank Xiao-Qing Zhang for assistance with figure preparation. Sarah Crews kindly checked English.

REFERENCES

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Avilés L, Guevara J. 2017. Sociality in spiders. *In*: Rubenstein DR and Abbot P. Comparative Social Evolution. Cambridge: Cambridge University Press, 188–223.

Bazazi S, Arganda S, Moreau M, Jeanson R, Dussutour A. 2016. Responses to nutritional challenges in ant colonies. *Animal Behaviour*, **111**: 235−249.

Bechsgaard J, Schou MF, Vanthournout B, Hendrickx F, Knudsen B, Settepani V, et al. 2019. Evidence for faster X chromosome evolution in spiders. *Molecular Biology and Evolution*, **36**(6): 1281−1293.

Bilde T, Lubin Y. 2001. Kin recognition and cannibalism in a subsocial spider. *Journal of Evolutionary Biology*, **14**(6): 959−966.

Brodschneider R, Crailsheim K. 2010. Nutrition and health in honey bees. *Apidologie*, **41**(3): 278−294.

Buskirk RE. 1981. Sociality in the arachnida. *In*: Hermann HR. Social Insects. New York: Academic Press, 281–367.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15): 1972−1973.

Chen SF, Zhou YQ, Chen YR, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884−i890.

Cheng DQ, Piel WH. 2018. The origins of the psechridae: web-building lycosoid spiders. *Molecular Phylogenetics and Evolution*, **125**: 213−219.

Chiara V, Ramon Portugal F, Jeanson R. 2019. Social intolerance is a consequence, not a cause, of dispersal in spiders. *PLoS Biology*, **17**(7): e3000319.

Cremer S, Pull CD, Fürst MA. 2018. Social immunity: emergence and evolution of colony-level disease protection. *Annual Review of Entomology*, **63**: 105−123.

Elbroch LM, Levy M, Lubell M, Quigley H, Caragiulo A. 2017. Adaptive social strategies in a solitary carnivore. *Science Advances*, **3**(10): e1701218.

Evans TA. 1998. Factors influencing the evolution of social behaviour in Australian crab spiders (Araneae: Thomisidae). *Biological Journal of the Linnean Society*, **63**(2): 205−219.

Fernández R, Kallal RJ, Dimitrov D, Ballesteros JA, Arnedo MA, Giribet G, et al. 2018. Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. *Current Biology*, **28**(9): 1489−1497.

Fischman BJ, Woodard SH, Robinson GE. 2011. Molecular evolutionary analyses of insect societies. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(S2): 10847−10854.

Fu LM, Niu BF, Zhu ZW, Wu ST, Li WZ. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**(23): 3150−3152.

Gilbert C, McCafferty D, Le Maho Y, Martrette JM, Giroud S, Blanc S, et al. 2010. One for all and all for one: the energetic benefits of huddling in endotherms. *Biological Reviews*, **85**(3): 545−569.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**(7): 644−652.

Guevara J, Avilés L. 2015. Ecological predictors of spider sociality in the Americas. *Global Ecology and Biogeography*, **24**(10): 1181−1191.

Hamilton WD. 1971. Geometry for the selfish herd. *Journal of Theoretical Biology*, **31**(2): 295−311.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics resources. *Nature Protocols*, **4**(1): 44−57.

Kapheim KM, Pan HL, Li C, Salzberg SL, Puiu D, Magoc T, et al. 2015. Genomic signatures of evolutionary transitions from solitary to group living. *Science*, **348**(6239): 1139−1143.

Kim KW. 2000. Dispersal behaviour in a subsocial spider: group conflict and the effect of food availability. *Behavioral Ecology and Sociobiology*, **48**(3): 182−187.

Korb J, Heinze J. 2016. Major hurdles for the evolution of sociality. *Annual Review of Entomology*, **61**: 297−316.

Kowalczyk A, Meyer WK, Partha R, Mao WG, Clark NL, Chikina M. 2019. RERconverge: an R package for associating evolutionary rates with convergent traits. *Bioinformatics*, **35**(22): 4815−4817.

Krafft B, Horel A, Julita JM. 1986. Influence of food supply on the duration of the gregarious phase of a maternal-social spider, *Coelotes terrestris* (Araneae, Agelenidae). *Journal of Arachnology*, **14**(2): 219−226.

Kuma A, Hatano M, Matsui M, Yamamoto A, Nakaya H, Yoshimori T, et al. 2004. The role of autophagy during the early neonatal starvation period. *Nature*, **432**(7020): 1032−1036.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078−2079.

Libbrecht R, Oxley PR, Kronauer DJC. 2018. Clonal raider ant brain transcriptomics identifies candidate molecular mechanisms for reproductive division of labor. *BMC Biology*, **16**(1): 89.

Lin YC, Li SQ. 2008. Description on a new *Philoponella* species (Araneae, Uloboridae), the first record of social spiders from China. *Acta Zootaxonomica Sinica*, **33**(2): 260−263.

Liu SL, Aagaard A, Bechsgaard J, Bilde T. 2019. DNA methylation patterns in the social spider, *Stegodyphus dumicola*. *Genes*, **10**(2): 137.

Löytynoja A, Goldman N. 2005. An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(30): 10557−10562.

Lubin YD. 1974. Adaptive advantages and the evolution of colony formation in *Cyrtophora* (Araneae: Araneidae). *Zoological Journal of the Linnean Society*, **54**(4): 321−339.

Macdonald DW. 1983. The ecology of carnivore social behaviour. *Nature*, **301**(5899): 379−384.

Majer M, Svenning JC, Bilde T. 2013. Habitat productivity constrains the distribution of social spiders across continents - case study of the genus *Stegodyphus*. *Frontiers in Zoology*, **10**(1): 9.

Maynard Smith J, Szathmáry E. 1995. The Major Transitions in Evolution. Oxford: Oxford University Press.

Michener CD. 1958. The evolution of social behavior in bees. *Proceedings of the Tenth International Congress of Entomology*.

Michener CD. 1969. Comparative social behavior of bees. *Annual Review of Entomology*, **14**: 299−342.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics*, **30**(17): i541−i548.

Mizushima N. 2007. Autophagy: process and function. *Genes & Development*, **21**(22): 2861−2873.

Moreno-Hagelsieb G, Latimer K. 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics*, **24**(3): 319−324.

Nentwig W. 1985. Social spiders catch larger prey: a study of *Anelosimus eximius* (Araneae: Theridiidae). *Behavioral Ecology and Sociobiology*, **17**(1): 79−85.

Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. 2010. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Research*, **38**(S2): W23−W28.

Powers KS, Avilés L. 2007. The role of prey size and abundance in the geographical distribution of spider sociality. *Journal of Animal Ecology*, **76**(5): 995−1003.

Rehan SM, Toth AL. 2015. Climbing the social ladder: the molecular evolution of sociality. *Trends in Ecology & Evolution*, **30**(7): 426−433.

Riechert SE, Roeloffs R, Echternacht AC. 1986. The ecology of the cooperative spider *Agelena consociata* in equatorial Africa (Araneae, Agelenidae). *The Journal of Arachnology*, **14**(2): 175−191.

Riechert SE. 1985. Why do some spiders cooperate? *Agelena consociata*, a case study. *The Florida Entomologist*, **68**(1): 105−116.

Rubenstein DR, Lovette IJ. 2007. Temporal environmental variability drives the evolution of cooperative breeding in birds. *Current Biology*, **17**(16): 1414−1419.

Rypstra AL. 1983. The importance of food and space in limiting web-spider densities; a test using field enclosures. *Oecologia*, **59**(2-3): 312−316.

Rypstra AL. 1986. High prey abundance and a reduction in cannibalism: the first step to sociality in spiders (Arachnida). *The Journal of Arachnology*, **14**(2): 193−200.

Rypstra AL. 1989. Foraging success of solitary and aggregated spiders: insights into flock formation. *Animal Behaviour*, **37**: 274−281.

Rypstra AL. 1990. Prey capture and feeding efficiency of social and solitary spiders: a comparison. *Acta Zoologica Fennica*, **190**: 339−343.

Sandidge JS. 2003. Scavenging by brown recluse spiders. *Nature*, **426**(6962): 30.

Sanggaard KW, Bechsgaard JS, Fang XD, Duan JJ, Dyrlund TF, Gupta V, et al. 2014. Spider genomes provide insight into composition and evolution of venom and silk. *Nature Communications*, **5**: 3765.

Sato T, Yamanishi Y, Kanehisa M, Toh H. 2005. The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. *Bioinformatics*, **21**(17): 3482−3489.

Schneider JM. 2002. Reproductive state and care giving in *Stegodyphus* (Araneae: Eresidae) and the implications for the evolution of sociality. *Animal Behaviour*, **63**(4): 649−658.

Settepani V, Bechsgaard J, Bilde T. 2016. Phylogenetic analysis suggests that sociality is associated with reduced effectiveness of selection. *Ecology and Evolution*, **6**(2): 469−477.

Settepani V, Schou MF, Greve M, Grinsted L, Bechsgaard J, Bilde T. 2017. Evolution of sociality in spiders leads to depleted genomic diversity at both population and species levels. *Molecular Ecology*, **26**(16): 4197−4210.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**(11): 2498−2504.

Shear WA. 1970. The evolution of social phenomena in spiders. *Bulletin of the British Arachnological Society*, **1**(5): 65−76.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, **32**(5): 1342−1353.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9): 1312−1313.

Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(16): 9440−9445.

Tong C, Najm GM, Pinter-Wollman N, Pruitt JN, Linksvayer TA. 2020. Comparative genomics identifies putative signatures of sociality in spiders. *Genome Biology and Evolution*, **12**(3): 122−133.

Toyama M. 1999. Adaptive advantages of maternal care and matriphagy in a foliage spider, *Chiracanthium japonicum* (Araneae: Coubionidae). *Journal of Ethology*, **17**(1): 33−39.

Uetz GW. 1988. Group forating in colonial web-building spiders: evidence for risk-sensitivity. *Behavioral Ecology and Sociobiology*, **22**(4): 265−270.

Uetz GW. 1989. The "ricochet effect" and prey capture in colonial spiders. *Oecologia*, **81**(2): 154−159.

Vanthournout B, Greve M, Bruun A, Bechsgaard J, Overgaard J, Bilde T. 2016. Benefits of group living include increased feeding efficiency and lower mass loss during desiccation in the social and inbreeding spider *Stegodyphus dumicola*. *Frontiers in Physiology*, **7**: 18.

Ward P, Zahavi A. 1973. The importance of certain assemblages of birds as “information-centres” for food-finding. *Ibis*, **115**(4): 517−534.

Ward PI. 1986. Prey availability increases less quickly than nest size in the social spider *Stegodyphus mimosarum*. *Behaviour*, **97**(3-4): 213−225.

Whitehouse MEA, Lubin Y. 2005. The functions of societies and the evolution of group living: spider societies as a test case. *Biological Reviews*, **80**(3): 347−361.

Wilson EO. 1971. The Insect Societies. Cambridge: Harvard University Press.

Wise DH. 2006. Cannibalism, food limitation, intraspecific competition, and the regulation of spider populations. *Annual Review of Entomology*, **51**: 441−465.

Woodard SH, Fischman BJ, Venkat A, Hudson ME, Varala K, Cameron SA, et al. 2011. Genes involved in convergent evolution of eusociality in bees. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(18): 7472−7477.

WSC. [2021-01-30]. World Spider Catalog Version 22.0. Natural History Museum Bern. http://wsc.nmbe.ch.

Yang ZG. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(8): 1586−1591.

Yip EC, Rayor LS. 2014. Maternal care and subsocial behaviour in spiders. *Biological Reviews*, **89**(2): 427−449.

Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. 2018. Ensembl 2018. *Nucleic Acids Research*, **46**(D1): D754−D761.

Zhang J, Kumar S. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. *Molecular Biology and Evolution*, **14**(5): 527−536.

Zhang JZ, Nielsen R, Yang ZH. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, **22**(12): 2472−2479.

Zou ZT, Zhang JZ. 2015. Are convergent and parallel amino acid substitutions in protein evolution more prevalent than neutral expectations?. *Molecular Biology and Evolution*, **32**(8): 2085−2096.