Sociality in spiders is an evolutionary dead-end

Jilong Ma[1,2*], Jesper Bechsgaard[2], Anne Aagaard[2], Palle Villesen[1]. Trine Bilde[2*$], Mikkel Heide Schierup[2*$]

1. Bioinformatics Research Centre, Aarhus University, DK-8000 Aarhus C, Denmark
2. Department of Biology, Aarhus University, DK-8000 Aarhus C, Denmark

* Corresponding authors: aujilongm@birc.au.dk, trine.bilde@bio.au.dk, mheide@birc.au.dk
$ shared last authorship

Abstract

Sociality in spiders has evolved independently at least 23 times. However, their phylogenetic distribution suggests that all social species are of recent origin and evolutionarily short-lived. Here, we present genomic evidence for the evolutionary dead-end hypothesis in the three independent transitions to sociality observed in the spider genus *Stegodyphus*. We sequenced, assembled and annotated high-quality, chromosome-level reference genomes from three pairs of closely related social and subsocial *Stegodyphus* species. Genome sizes range from 2.65 Gb to 3.32 Gb with high synteny, and we identify 10,065 single-copy orthologous genes. We timed the divergence between the social and subsocial species pairs to be from 1.3 to 1.8 million years. Social evolution involves a shift from outcrossing to inbreeding and from equal to female-biased sex ratio, causing severe reductions in effective population size and decreased efficacy of selection. Based on analyses of purifying selection, we determined whether transitions to sociality co-occurred with divergence. We show that transitions to sociality only had full effect on purifying selection at 119 kya (95CI: 71 kya -169 kya), 260 kya (95CI: 231 ky - 289 kya) and 279 kya (95CI: 230 kya - 332 kya) respectively, and follow remarkably similar convergent trajectories of progressive loss of diversity and shifts to an increasingly female-biased sex ratio. This almost deterministic genomic response to sociality may explain why social species do not persist. What causes species extinction is not clear, but could be either selfish meiotic drive that eliminates the production of males, or an inability to retain genome integrity in the face of extremely reduced efficacy of selection.

Key words:

Stegodyphus, sociality, transition times, phylogenomics, evolutionary dead-end, purifying selection, relaxed selection, genome integrity, spiders

Introduction

The evolution of sociality has originated independently a large number of times in the animal kingdom [1]. Selective forces that drive transitions to sociality include ecological constraints on individual survival and reproduction favouring delayed dispersal and cooperative breeding [2–5], or foraging benefits from communal hunting or information sharing [6–9]. The evolution of sociality invariably comes with costs such as elevated competition for resources, increased risk of disease transmission, and cheating over cooperative interactions [6,10,11]. The benefits of sociality to overcome these costs must, therefore, be substantial.

Spiders are an evolutionarily old group (>300 million years) [12], and signs of social behaviour were reported from fossil evidence in 100 million-year-old amber [13]. However, social spider species are rare, constituting only ~23 of more than 50,000 extant spider species [14]. Nevertheless, sociality in spiders has evolved independently multiple times in at least 6 families. Phylogenetic analyses show that social spider species have evolved independently and are evolutionarily young, since they are found at the terminal branches of the spider phylogenetic trees, and there are no described cases of speciation following transitions to sociality [5,15]. Remarkably, the social spider species are characterised by a suite of convergently evolved traits [5], cooperative breeding and reproductive skew where only a small fraction of females lay eggs [5], a reduction in adult body size and the production of smaller broods [16,17], genetically encoded primary female-biased sex ratio [18,19], and a shift from outcrossing to inbreeding, as elimination of pre-mating dispersal results in exclusive reproduction among siblings [5]. This pattern of convergence indicates that social evolution entails similar adaptive responses in behavioural and life history traits.

In the spider genus *Stegodyphus*, which contains more than 20 species, sociality has evolved three times independently [20,21], and each social species has a closely related subsocial sister-species (see Fig 1A) suggesting that sociality evolved recently [19,21]. The combination of strict inbreeding, primary female-biased sex ratio, and reproductive skew is predicted to reduce effective population size

dramatically [19], thereby increasing the strength of drift and reducing the efficacy of natural selection in the social species [22]. This substantiates the question of whether consequences of low Ne impacts on the evolutionary persistence of social lineages. The findings of substantially reduced neutral genetic diversity [23] and elevated rates of nonsynonymous to synonymous substitution rates [19,21,24,25] in social *Stegodyphus* species are consistent with reduced effective population sizes. Constant inbreeding over evolutionary time is likely to have resulted in purging of segregating recessive deleterious alleles, since there is only evidence for outbreeding depression and not inbreeding depression [26].

We here perform whole genome assembly and comparative analyses of the three social *Stegodyphus* species and their subsocial sister-species with the aim of timing individual transitions to sociality, and determine the extent of relaxation of purifying selection. We provide evidence that the transition period from subsociality to sociality is well defined, and that the transitions were completed at different time points within the last three hundred thousand years in the three cases. This suggests that sociality is even more recent than what can be deduced from the divergence times with the closest subsocial species (1-2 million years), and begs the question of why social species are so evolutionarily short-lived .

Results and Discussion

De novo assembly and genome annotation

Six species of *Stegodyphus* (3 social, 3 subsocial) were sequenced to ~30X coverage using Pacbio HiFi long reads. Combining this with Hi-C contact maps based on 30X Illumina sequencing, we generated chromosome-level, *de novo* assemblies. (Table 1, Supplementary Figure S1).The sex chromosomes of each species were identified by the read coverage of male individual alignments. We find two sex chromosomes in all species except for the outgroup species *S. lineatus* which has three X chromosomes (Figure S8). For each species, the inferred number of linkage groups matched the number of chromosomes found by cytological visualization [27,28]. We evaluated possible assembly errors by comparing the independent assemblies by dotplots using dotPlotly (Supplementary Figure S4-S6). For the most closely related species pairs (1-3% divergent at the genome level), assemblies were essentially syntenic, suggesting that the rate of assembly error is very low and also that the synteny breaks for more divergent species pairs indicate real chromosomal rearrangements and/or inversions. Annotation using BRAKER2 [29] independently for each assembly revealed similar numbers of protein-coding genes ranging from 32,093 to 35,724, except for *S. sarasinorum* where 47,066 protein-coding genes were annotated. This recent gene number expansion in *S. sarisinorum* is not an artefact of annotation and is the focus of a separate study. BUSCO [30] analyses show high single completeness of the genomes ranging from 91.1% to 93.3%. The GC content ranges from 33.26% to 34.18%, and the proportion of the genomes constituting repeats ranges from 57.17% to 65.20%. We found 10,065 single-copy orthologous genes present in all six species using our *de novo* assemblies, synteny information and annotations (Table 1, Table S2). Mercury [31] estimation on our genomes reveal k-mer completeness ranging from 97.65% to 99.41% and a consensus quality QV ranging from 62.54 to 73.86, which is higher than the Vertebrate Genomes Project (VGP) benchmarking recommendations of QV > 40 and k-mer completeness > 95% [32]. When incorporating the *de novo* assembled transcriptome from *S. africanus* and short reads DNA sequences from *S. pacificus*, our

final dataset ends up with 2,303 and 348 reliable single copy orthologs for autosomes and X chromosomes, respectively, across all 8 species.

Species	Chrom_Num	Sex_chrom	Scaffolded Length	GC level	RepeatMasked	Gene_Num	BUSCO (S.C)
S.lineatus	23	X1X2X3O	2658234335	33.26%	58.72%	32093	94.10%
S.mimosarum	16	X1X2O	2910110412	33.92%	57.17%	35480	92.40%
S.tentoriicola	14	X1X2O	2984342755	33.42%	64.59%	35725	91.30%
S.dumicola	14	X1X2O	3146466829	34.18%	64.53%	34096	91.00%
S.bicolor	16	X1X2O	2984548140	33.78%	61.98%	35096	92.30%
S.sarasinorum	16	X1X2O	3321123892	34.06%	65.20%	47066	93.10%

Table 1 Assembly and annotation statistics. BUSCO completeness is only reported for the single complete fraction checked with protein sequences from annotated genes, complete BUSCO results reported in Table S2.

High gene synteny and few chromosomal rearrangements in the *Stegodyphus* genus

To investigate genome rearrangements across *Stegodyphus* from the accurate *de novo* assemblies, we used GENESPACE [33] to retrieve the pair-wise synteny blocks based on the gene order of orthologs found in each *de novo* annotation (Figure 1 outlines the results). In general, synteny blocks suggest few chromosomal translocations and further support that the assemblies are accurate.

We identify a number of large-scale structural rearrangements including splits and joins of chromosomes unique to certain lineages of the phylogeny. For example, the outgroup species *S. lineatus* has 23 chromosomes including three X chromosomes, while the other species contain 14 to 16 chromosomes and only two X chromosomes, suggesting that multiple chromosome splits or joins have happened in the *S. lineatus* lineage, or that chromosome joins have occurred after the split with *S. lineatus* in the common ancestor of the other species (Table 1). We find that the third X

chromosome in *S. lineatus* appears to originate from a "cut and paste" non-homologous recombination process, which joins a part of ancestral X chromosome X1 and a part of ancestral X chromosome X2 (Fig 1. D). Thus, the three X chromosomes in the *S. lineatus* lineage have almost the same gene complement as two X chromosomes in other species. In addition, we find three chromosome join events assigned to the ancestral branch of *S. dumicola* and *S. tentoriicola*, and one chromosome join event and a non-homologous recombination event assigned to the ancestral branch of *S. bicolor* and *S. sarasinorum*.

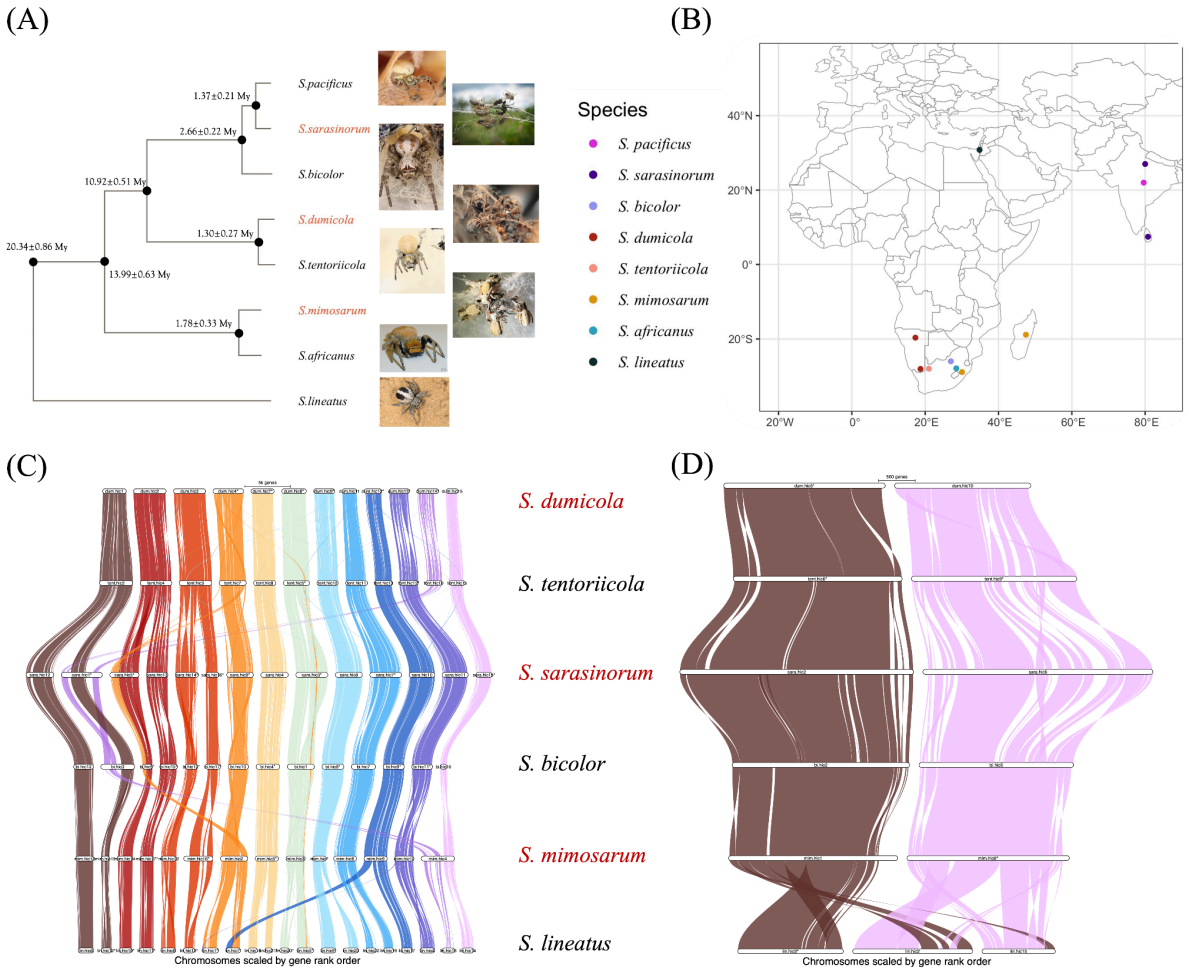


Figure 1. (A) Phylogeny of the spider genus *Stegodpyhus* with social species in red, the divergence time is estimated from dS of autosomal single-copy orthologous genes and a yearly mutation rate of 5e-09 per site per year. (B) Sampling locations of the species included in our study. The gene synteny

plots among the six *de novo* chromosome-level assemblies are shown using GENESPACE, separately for autosomes (C) and X chromosomes (D). The length of chromosome in the gene synteny plot represents the number of genes.

Estimating the beginning and end of the transitions to sociality

We first estimated divergence times at each node of the *Stegodyphus* phylogeny using the bootstrapped branch-wise dS from autosomal branches and assuming a mutation rate of 5e-09 per site per generation [34]. Among the three pairs of social and subsocial sister species, the estimated divergence times are 1.78±0.33 million years between *S. mimosarum* and *S. africanus*, 1.30±0.27 million years between *S. dumicola* and *S. tentoriicola*, and 1.37±0.21 million years between *S. sarasinorum* and *S. pacificus*.

Transitions to sociality are associated with a shift to inbreeding and reduced effective population size (Ne) causing natural selection to be less efficient.We used our genome data to estimate the time when these transitions to sociality occurred from the expected signals in purifying selection and Ne. We did this in two separate ways, first from the estimated changes in purifying selection estimated from dN/dS ratios, and second, from the inferred changes in effective population size through time estimated from coalescent analyses.

Intensity of purifying selection: We estimated the transition time to social behaviour from dN/dS ratios on the terminal branches of each social/subsocial species pair based on the following assumptions: a) dN/dS ratios of the social species was the same as for the subsocial species until social transitions on the terminal branches leading to the social species, b) after transition to sociality, the dN/dS ratio increased due to less efficient purifying selection, c) this 'social dN/dS', can be approximated by the piN/piS ratio estimated between divergent and isolated extant populations of each social species (see Methods). This dN/dS analysis suggests that the effects of a transition to sociality on purifying selection are very recent (Figure 2 A-D). For example, the dN/dS over the

entire *S. mimosarum* branch is 0.1259±0.0067, and for its subsocial sister-species *S. africanus* dN/dS is 0.0927±0.0011. The dN/dS ratio since sociality evolved is estimated at 0.3043±0.0028 comparing two *S. mimosarum* individuals from Madagascar and mainland Africa. From the divergence time estimates for this species pair, this generates an estimated time since transition to sociality of *S. mimosarum* of 279 kya (95% CI: 230 kya - 332 kya). Using the same calculation, the social dN/dS ratio of *S.dumicola* is estimated to be 0.3366±0.0057 by comparing individuals from Otavi North Namibia and Karasburg South Namibia. Combining the dN/dS ratio being 0.1776±0.0059 and 0.1615±0.0082 for the *S. dumicola* and *S. tentoriicola* branches respectively, the estimation of the transition to socialility is 119 kya (95% CI: 72 kya - 170 kya) in *S. dumicola*.

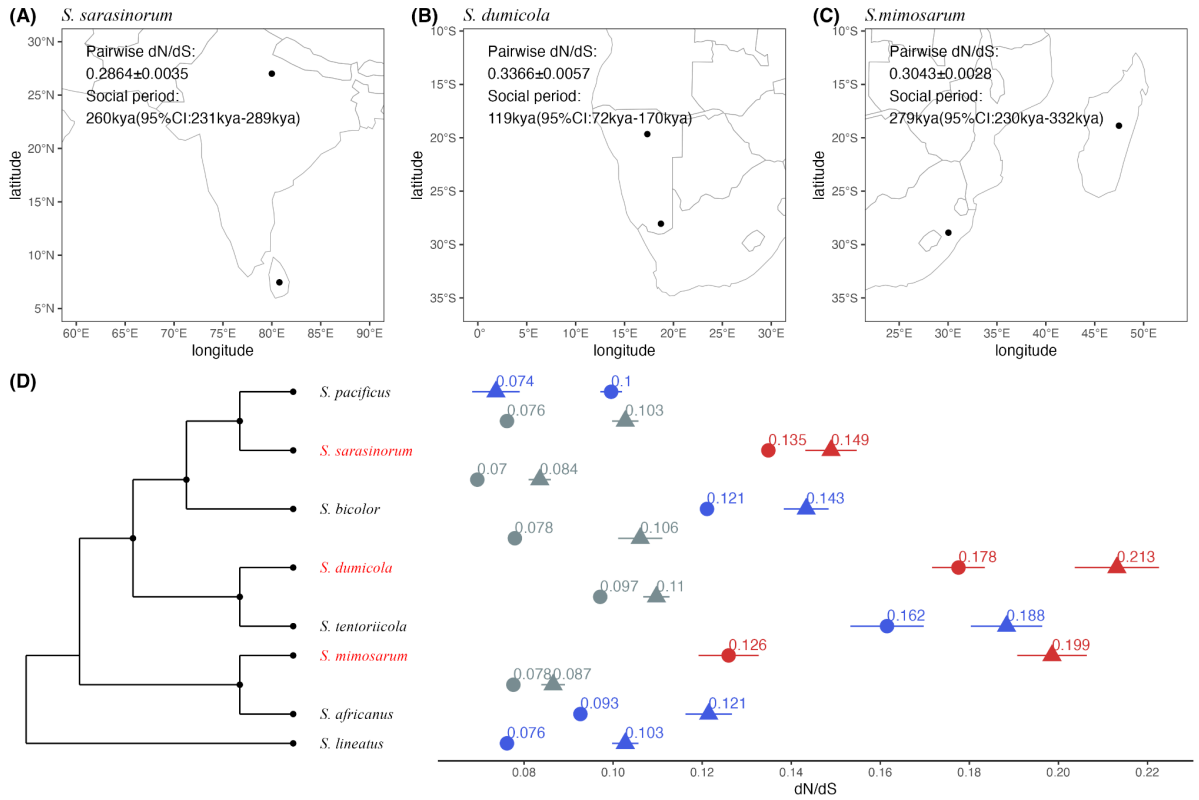


Figure 2. The intensity of purifying selection was assessed using dN/dS from the coding sequences of 2302 single-copy orthologous genes.(A) The locations of the Sri Lanka *S.sarasinorum* and Himalayas *S.sarasinorum* samples, along with their pairwise dN/dS estimations for the social period. (B) The locations of the Namibia South *S.dumicola* and Namibia North *S.dumicola* samples, with corresponding pairwise dN/dS estimations for the social period. (C) The locations of the South Africa

S.mimosarum and Madagascar *S.mimosarum* samples, and their pairwise dN/dS estimations for the social period. (D) Branch-wise dN/dS estimations using PAML, with the 99% confidence interval from bootstrapping. Red, blue, and grey denote social species terminal branches, subsocial species terminal branches, and internal branches, respectively. The dN/dS estimation, indicated by points with round and triangle denote autosomes and X chromosomes respectively, is aligned to the phylogeny that corresponds to the respective branches. The phylogeny in (D) provides an indication of the topology without branch length information.

For the social *S. sarasinorum,* we performed two separate estimations by comparing *S. sarasinorum* to the subsocial species *S. pacificus,* which diverged 1.37±0.21 million years ago, and the subsocial species *S. bicolor*, which diverged 2.67±0.29 million years ago. In *S. sarasinorum*, we estimated the social dN/dS ratio to be 0.2864±0.0035 based on individuals from Himalaya and Sri Lanka. The dN/dS ratio on the terminal branch of *S. sarasinorm* is 0.1349±0.0013. Compared to the 0.0996±0.0024 of the *S. pacificus* branch and the 0.1210±0.0011 of the *S. bicolor* branch, the final estimation for the transition to sociality of *S. sarasinorum* is 260 kya (95% CI: 231 kya - 289 kya) and 223 kya (95% CI: 199 kya - 247 kya), respectively.

We have attributed the observed higher dN/dS ratio in the terminal branches of social species to a decrease in purifying selection. However, increased adaptive selection acting on social species could have the same effect. RELAX analyses [35] suggest that the elevated dN/dS ratios in social species is almost entirely due to relaxed selection, as has previously been suggested [36]. We also show elevated dN/dS ratios on X chromosomes compared to autosomes across all branches in the phylogeny, except for the terminal branches of *S. pacificus*. Such elevated dN/dS ratios can be attributed to a decrease in purifying selection acting on X chromosomes, which is consistent with the expected outcome of reduced Ne in X chromosomes compared to autosomes. Alternatively, but not mutually exclusive, an adaptive explanation for elevated dN/dS ratios on the X chromosomes is the so-called faster-X effect [37]. Recessive variants on the X chromosomes are directly exposed to selection in males that are haploid for X chromosomes, and under certain conditions this will lead to more adaptive substitutions on X chromosomes compared to autosomes.

Additionally, we note that our implication of using piN/piS between divergent populations of the same species could potentially lead to a systematically higher "social dN/dS" due to the unfixed polymorphisms rather than reflect the actual selection strength when the selection strength acts differently on synonymous polymorphism sites and non-synonymous polymorphism sites within species. We benchmarked our strategy by comparing the site frequency spectrum of missense variants (N-sites) and synonymous variants (S-sites) in *S. dumicola* built from 9 individuals. Site frequency spectrums of the N-sites and S-sites reveal no visually significant difference in shape, indicating the similar ongoing selection pressures on N-sites and S-sites (Figure S7). Thus, we conclude that piN/piS to a good approximation can be used as a proxy for the dn/ds ratio after sociality ("social dN/dS").

Coalescent analysis: Next, we used PSMC [38] analysis to reconstruct the effective population sizes and the sex ratio in the populations over time by analysing autosomes and X chromosomes separately in each species. For the social species, which are all highly inbred, a diploid genome was created by merging two haploid genomes sampled from different divergent and isolated populations, i.e. an artificial diploid. This allows coalescent events further back in time to be inferred. The transition time to sociality is then estimated by graphically determining the time period when the effective population size of the social species became severely reduced (Figure 3). Compared to subsocial species, the population trajectories of all the social species share a gradual reduction of Ne from more than 300,000 to less than 50,000, but starting and ending at different time points within the past 900 ky. In the social species *S. dumicola*, the Ne is reduced from 300,000 around 420 kya to around 50,000 at 300 kya, and even lower today. In the social species *S. mimosarum*, the Ne is reduced from 1.4 million to 57,500 in the time interval between 900kya and 350kya. The reduction of Ne in the social species *S. sarasinorum* starts 600 kya at 800,000 to less than 30,000 around 300 kya.

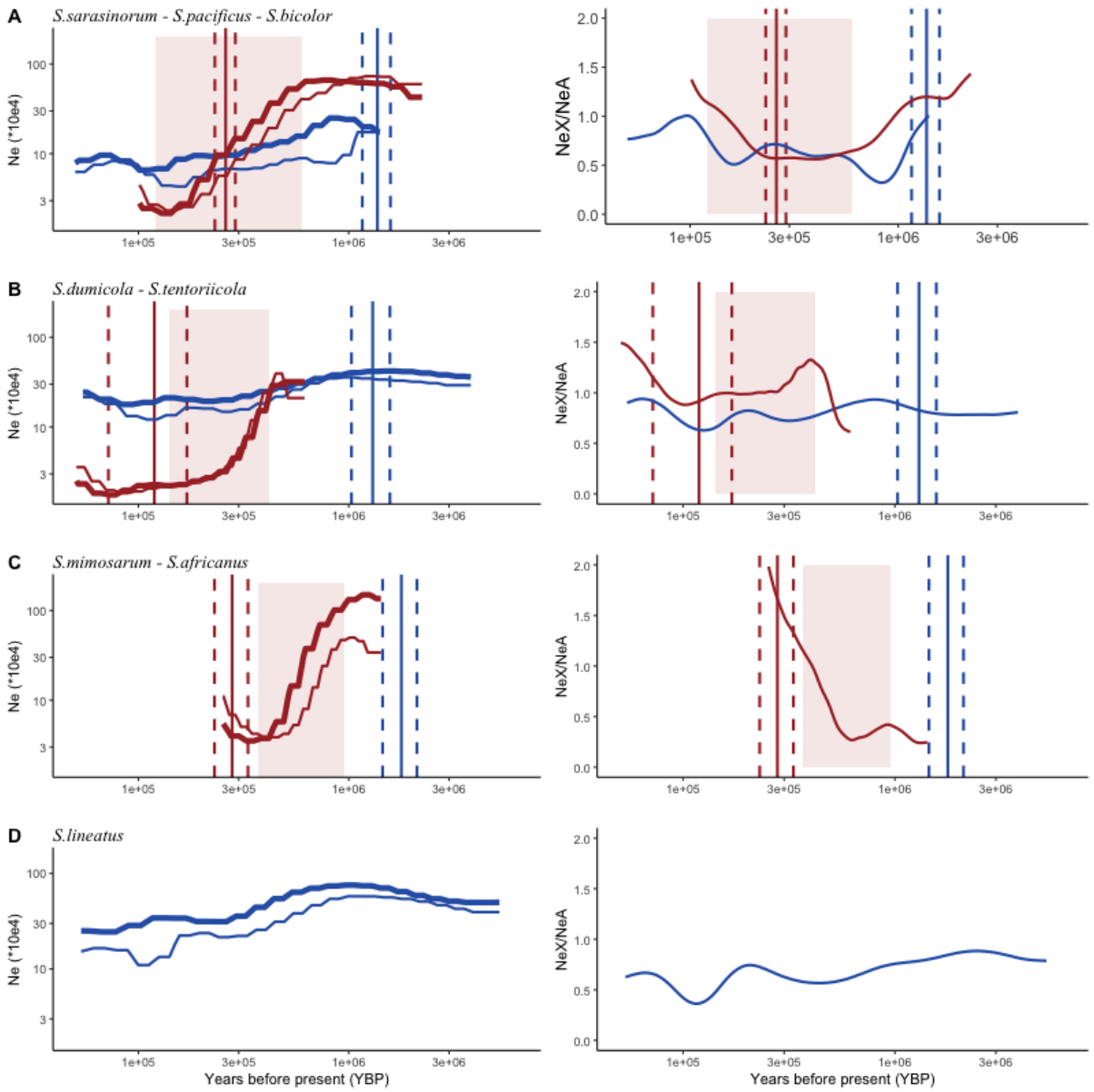


Figure 3. The historical effective population size inferred from the Pairwise Sequentially Markovian Coalescent (PSMC) model, for different *Stegodyphus* species with chromosome-level assembly. In each panel, the left figure depicts PSMC results, while the right figure displays the effective population size ratio between X chromosomes and autosomes. Each species' trajectory is shown with thicker lines for autosomes and thinner lines for X chromosomes. The type of species are distinguished by colour: red for social species and blue for subsocial species. Red shadows mark the estimated social transition period based on the decreasing effective population size in social species. Vertical lines indicate estimated time points. Red lines correspond to the estimated social transition

time, derived from relaxed selection in social species, while blue lines represent the estimated divergence time between a social species and its closest subsocial sister species. Solid lines provide point estimates, and dashed lines delineate the 95% confidence interval for red and the 99% confidence interval for blue. (A) Displays data for the social species *S.sarasinorum* and the subsocial species *S.bicolor*. Estimates for social transition and divergence time are based on a comparison between *S.sarasinorum* and its subsocial sister species, *S.pacificus*. (B) Presents findings from a comparison between the social species *S.dumicola* and its subsocial sister species *S.tentoriicola*. (C) Shows data for the social species *S.mimosarum*, with social transition and divergence time estimates based on a comparison with its subsocial sister species, *S. africanus*. Note that we do not have an effective population size estimation for *S. africanus* due to missing whole genome data. (D) Presents historical effective population size for the outgroup subsocial species *S.lineatus*.

Since the PSMC analyses of the social species are made from pseudodiploid genomes combined from separate genetically isolated populations, they can also provide an estimate of the time of separation of these populations, since coalescence can not occur when they are separate and the Ne estimate will thus be very large. We observe such an uptick in Ne in recent times in the three social species, and estimate that the two *S. mimosarum* populations from Madagascar and South Africa diverged around 200 kya, the two *S. dumicola* populations in North and South Namibia diverged 20 kya, and the two *S. sarasinorum* populations from Himalaya and Sri Lanka diverged approximately 110 ky ago.

Notably, though the pseudo-diploid of social species intended to collapsing two haploid individual as social species individual were heavily inbred with run out of heterozygosity (ROH) in 59.83%-86.27% of their genomes (Table S4), we can not rule out the fact that pseudo-diploid actually capture more than two haploid for part of the genomes outside the ROH region. However, we expect this will only increase a local coalescence rate that is randomly distributed in the genome, resulting in a lower estimation of Ne overall, without biassing the Ne reduction information retrieved from the Ne trajectories.

Evolution of sex ratio bias

The sex determination system of spiders is X0, and as there are four autosomes for each three X chromosomes in the population, the effective population size of X chromosomes is expected to be 75% of the autosomes. In subsocial species, PSMC analyses of X chromosomes and autosomes showed that although the estimated Ne for X chromosomes is indeed lower, the ratio of X chromosomes to autosome effective population size is around 50-60% instead of the expected 75%. This can be due to a lower mutation rate and/or stronger selection on the X chromosomes, as was previously suggested based on diversity estimates on X chromosomes and autosomes in *S. africanus* [19].

In contrast to the subsocial species, the social species show strongly female biased primary sex ratios [5] and this has been shown to be genetically determined [18]. When a species evolves from an unbiased to a female biased sex ratio, the effective population size of X chromosomes should approach that of autosomes. In concordance with this, the PSMC analyses reveal that X chromosome and autosome Ne are almost identical in the most recent time intervals of the PSMC, as expected when sociality has fully evolved. The evolution of a female biased sex ratio can therefore also be timed using PSMC analyses under the assumption that it occurred after the transition to sociality, by comparing the Ne of X chromosomes and autosomes back in time. We found elevated Ne(X)/Ne(autosomes) ratios in all social species associated with reductions in Ne, which links the evolution of sex-ratio bias with social transitions. This suggests that the evolution of a female biased sex ratio in the three social species occurred early in the evolutionary transition to sociality.

The evolution of female bias is predicted under strong continuous inbreeding and lack of pre-mating dispersal, characteristics of social spider species and other social arthropods [39–41], to reduce competition among brothers for fertilization of their sisters [42,43]. Furthermore, allocation to daughters as the helping sex in social species is expected by the local resource enhancement model [44]. Biased sex ratios are often caused by sex chromosome meiotic drive [45], and the driving loci are selfish genetic

elements that would usually trigger adaptations to restore even sex ratio[45]. However, selection against female bias is not expected here, because of the adaptive benefits of female biased sex allocation [18] generating predictions of rapid fixation of the sex-linked meiotic driver. If stronger female bias keeps being adaptive, selection will lead to an increasingly female biased sex ratio following transition to sociality. The results from the PSMC analyses that NeX/NeA continues to increase after transition to sociality suggest that stronger female bias continues to be adaptive. Furthermore, our estimation for the transition to sociality matches with observations of increasing female bias with age of sociality in the three species. The oldest social species *S. mimosarum* has the strongest sex ratio bias with 9% males, the younger social species *S. dumicola* and *S. sarasinorum* have a male percentage of 12% and 8.5-33%, respectively[5]. The correlation between sex ratio bias and age of the social period in the three transitions to sociality therefore indicate that the continued loss of males is genetically determined.

Speciation times and transition to social behaviour

Figure 4 summarises the estimates of species divergence times and the transition periods to sociality using the different approaches. An obvious question is whether the speciation process itself was caused by social behaviour being initiated in one of two diverging populations. The large time interval between divergence estimates (1.3-1.8 my) and the beginning of social transitions (400-900 ky) appears to argue against this. However, since the effective population sizes in subsocial species are very large (300,000-600,000), the divergence times are expected to be much more ancient than the species separation times due to polymorphism in the ancestral species. Coalescent theory predicts that it takes 2Ne generations on average for this ancestral polymorphism to find most common ancestry. With a generation time of one year, this corresponds to between 600,000 and 1.2 million years, and our best estimate of the speciation times should then be the divergence times minus almost 1 million years. This could reconcile the beginning of the social transitions with species separation, thus supporting that social behaviour emerging in a population of a species may facilitate or even be causative of separation of this population into a new species. Nevertheless, the combination of data

presented above substantiates that sociality fully evolved much more recently than the divergence estimate and is a short-lived evolutionary strategy.

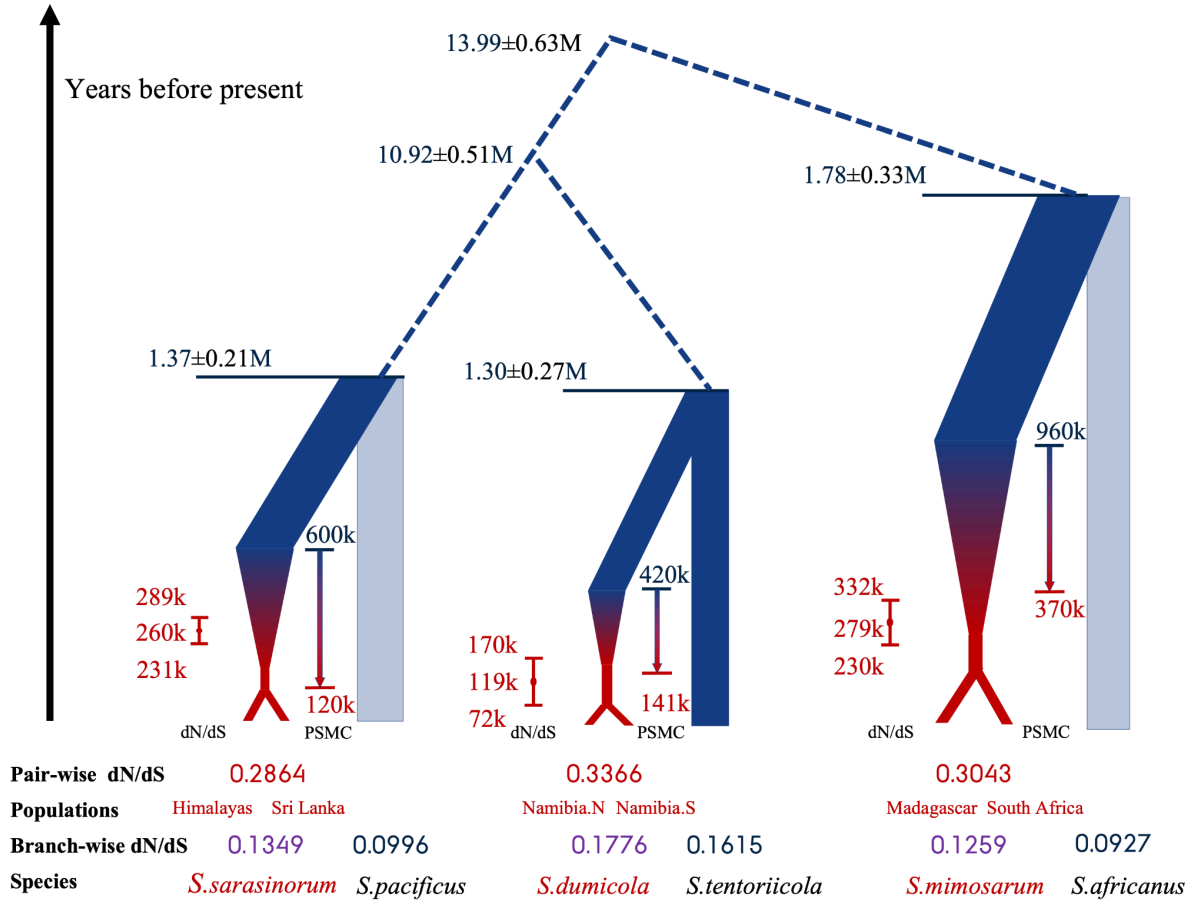


Figure 4. Estimated species divergence and social transition times in three independent social transitions within the *Stegodyphus* spider genus. The scale of the plots is symbolic and not proportional. The widths of the terminal branches represent variation in effective population size, with red symbolizing social periods and blue indicating subsocial periods. Light blue branches represent species without chromosome-level assemblies for PSMC. For each lineage where a social transition occurs, point estimates and 95% confidence intervals for social transition times are presented on the left, while social transition periods derived from population size history are displayed on the right. In PSMC estimations, the start of social transition periods corresponds to the initial decrease in effective population size, while the end coincides with Ne dropping below 50,000.

Conclusion

The genomic signatures of social species indicate that the evolution of sociality simultaneously entails a transition from outcrossing to inbreeding and the development of a female biased sex ratio. The relatively young age of accomplished sociality indicates that sociality evolves rapidly and frequently in the spider phylogeny, combined with a subsequent strong signal of decreased purifying selection caused by close inbreeding and much reduced effective population size. The recurring and rapid evolution of sociality, coupled with the absence of lineages with a long history of sociality, is strong indirect evidence that sociality in spiders represents an evolutionary impasse, unlikely to persist beyond a timespan of 1 millions years once established. Thus, it is plausible that there have been many more social spider species that emerged and went extinct throughout evolutionary history, but we can only observe the presence of social spider species recently developed.

We note that even though the three social transitions we report were initiated at different times, they have followed remarkably similar trajectories in population size reduction, increased dN/dS, and biased sex ratios that complements the strong similarities in behaviour and life history associated with transitions to sociality. This almost deterministic genomic response to sociality may also hold the key to understanding why social species in spiders do not seem to persist for very long. Several possibilities may explain the demise. These include a) loss of adaptive potential due to loss of genetic diversity, b) fixation of deleterious variation, c) loss of males, and d) loss of genome integrity. While a) and b) may definitely make sociality a doomed evolutionary strategy, we doubt that they can happen within a few hundred thousand years since the fixation of slightly deleterious variation takes a very large number of generations[46]. Loss of males could carry on until population collapse, and it is intriguing that we see fewer males in the social transition estimated to be the oldest. Finally, loss of genome integrity by, for example transposable elements, increased mutation, poorer genomic repair of damage, and double strand breaks, are all global mechanisms that could impact the future for these species. This idea is the subject of our future research.

Methods

Sample collection across the *Stegodyphus* genus

The spider genus *Stegodyphus* is distributed widely across Eurasia and Africa and we collected seven species including three social and four subsocial species in this study (Supplementary Table 1). Samples were collected from multiple regions for social species (Fig. 1 A). We collected *S. dumicola* samples from Otavi and Betta in Namibia (2017), *S. mimosarum* samples from Antanario in Madagascar and Weenen in South Africa (2011), and *S. sarasinorum* samples from the Himalaya region and Sri Lanka (2014). For subsocial species, we collected *S.lineatus* from the Negev Desert in Israel (2020), *S. bicolor* from Betta in Namibia (2021), *S. tentoriicola* from Tierberg in South Africa (2015). *Stegodyphus africanus* RNA data was previously published in Bechsgaard et al (2019).

Reference genome: DNA extraction, sequencing and assembly

DNA was extracted from single individuals using the MagAttract HWM DNA kit from Qiagen (Hilden, Germany). Sequencing was done by PacBio HiFi technology to obtain from 69 to 99 GB data.We assembled haplotype resolved contigs using Hifiasm[47] with Pacbio HiFi reads and Hi-C reads. Hi-C reads were aligned to the contigs from the longer haplotype of each species using Juicer pipeline for scaffolding chromosome-level assemblies using 3D-DNA pipeline [48]. The chromosome-level assemblies were further manually checked and curated using Juicebox Assembly Tools[49] based on consistency of Hi-C contact pattern.

Resequencing: DNA extraction and sequencing

DNA was extracted from single individuals using the Qiagen Blood and Tissue kit, Qiagen (Hilden, Germany), and the DNA was sequenced using DNBSEQ-G400.

Genome Annotation

We annotated six species with chromosome-level assemblies using BRAKER2 with RNA-seq and protein homology evidence combined. For each species, we used RepeatModeler2[50] to generate a

repeat library and combine it with the repeat library for Arthopoda from Repbase[51]. The chromosome-level assemblies are masked using Repeatmasker[52] and their corresponding repeat libraries. We used STAR[53] to align RNA-seq reads to the repeat-masked genome as transcriptome evidence for gene prediction in BRAKER2. For protein homology evidence, we used protein sequence from a previous *S.dumicola* annotation[54] and arthropoda protein from OrthoDB v10[55]. In the end, the completeness of the genome annotations is evaluated with BUSCO together with a set of 1013 *Arthropoda* genes from OrthoDB v10.

Ortholog groups finding and multiple genome alignments

We performed GENESPACE to check the gene synteny among the six hifi chromosome-level assemblies. During the GENESPACE analysis, OrthoFinder[56] was applied to identify orthologous groups across the six species. The 10065 single-copy orthologous groups out of 48143 orthologous groups identified from the six species were further used in evolutionary analysis.

From *S. africanus* we do not have the genome sequence. To include *S. africanus* in the analyses, we downloaded the transcriptome data used in and used Trinity[57] to assemble transcripts. We used DIAMOND[58] blastx mode to align the transcripts to the database of all the translated amino acid sequences from the 10065 single-copy orthologous groups of all six species. 5590 out of 10065 single-copy orthologous groups find one or more hits from the trinity transcriptome assemblies. We further check the aligned percentage of the hitted orthologous genes and filtered for a percentage of 75% to consider a transcript from *S. africanus* being a valid match to a certain single-copy ortholog groups identified, which ended up with 2649 single-copy orthologous groups.

We also only have short-read DNA-seq from *S. pacificus*. To include *S. pacificus* genes into the 2649 single-copy orthologous groups identified, we used bwa-mem2[59] to aligned a *S. pacificus* individual to the *S.sarasinorum* reference genome, which is the closest sister species of it, and used SAMtools[60] to call consensus sequence for the 2649 *S.sarasinorum* genes as the sequence of *S.pacificus.* We ignored

indels in calling consensus sequences to maintain the same coordinate systems between the two species.

Branch-wise dN/dS estimation

We used CodeML in PAML[61] to estimate dN and dS in branch-wise mode based on the 2649 single-copy orthologous groups. The 2649 orthologous groups were split up into an autosome set containing 2302 genes and a X chromosome set containing 347 genes. We randomly sample 500 genes from the 2302 autosomal genes independently 500 times to have a bootstrap estimation of dN/dS of each branch in the phylogeny with confidence intervals. For X chromosomes, we randomly sample 100 genes independently 100 times instead.

We do codon aware alignment in gene-wise for dN/dS estimation. We used alignSequence from MACSE[62] to align coding sequence of each gene from all the species excluding *S.pacificus*, which was attached to the alignment according to the *S.sarasinorum* in the end. We also examined and filtered each gene alignment for local mis-alignment. The size of each alignment block without gap should have a length for at least 300 nucleotide base pairs as well as a fraction of polymorphism sites under 15%. The passing alignment blocks of the random chosen genes were concatenated using GOalign[63] for dN/dS estimation.

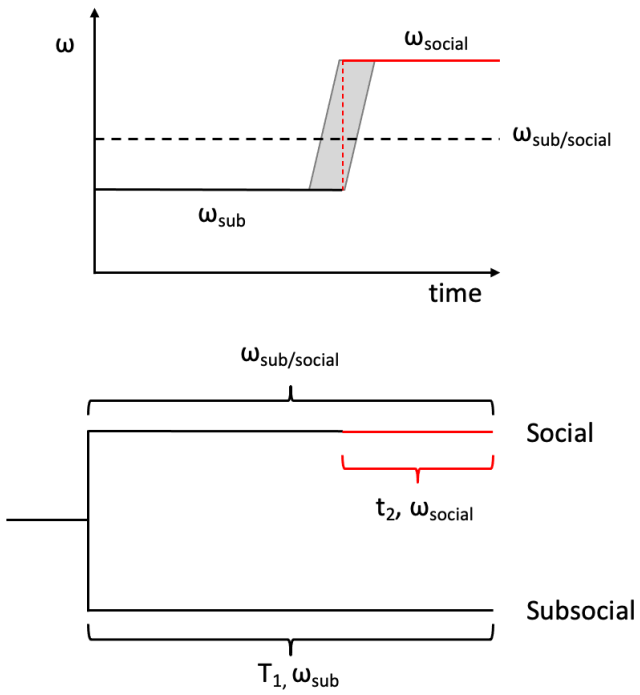
dN/dS estimation for social species

We used CodeML in PAML to estimate a pairwise dN and dS for each social species with two separate populations. We aligned one individual from different populations to the species reference separately using BWA-MEM2. After calling the SNP using platypus[64], we retrieve the alternative nucleotide to substitute the species reference as the reference genome of each sub-population. Sequences from the 2649 single-copy orthologous genes were retrieved from the sub-population references for pairwise dN/dS estimation.

dN/dS and estimation of speciation divergence and social transition

For the divergence time between lineages, we used the dS branch length, a constant mutation rate of 5e-09 per site per generation, and a generation time of one year to scale the dS length into years. We use the average dS branch length between a split as an estimation of divergence time except for the split between *S.sarasinorum* and *S.pacificus*. The exception is due to the underestimation of dS length of *S. pacificus* lineage from ignoring of indels, thus we use the dS length of *S.sarasinorum* branch alone for divergence time estimation between *S.sarasinorum* and *S.pacificus*.

Assuming the transition to sociality is an instant process. The dn/ds estimated for the social species terminal branch is an average of the subsocial period dn/ds and social period dn/ds weighted by the length of each period on that branch. The subsocial period can be substituted approximately with the dn/ds for the branch of the subsocial sister species. The social period dn/ds can be estimated by pairwise comparing two individuals from two isolated and divergent populations of the social species. The point estimation for species divergence time is estimated from the mean dS branch length of 500 sets of random selected 500 autosomal genes. Thus the 99% confidence interval of species divergence time is retrieved using the mean and three times of the standard deviation. For the time of transition to sociality, we sampled 10000 sets of observations for relevant dN/dS and species divergence time assuming normal distributions and solve for the social transition time. The point estimation is obtained from the mean of 10000 samples and the 95% confidence interval is retrieved from the empirical distribution of sampled social transition time to account for all uncertainty in dN/dS estimations and species divergence time.



$$t_2 = T_1 \times \frac{(\omega_{sub-social} - \omega_{sub})}{(\omega_{social} - \omega_{sub})}$$

Figure M1. Illustration for calculating the social period length for current social species as an instant process. t_2 denotes the length of the social period. T_1 denotes the divergence time between social species and its sister subsocial species. ω_{sub} denotes the dN/dS estimated from the subsocial species branch in branch-wise PAML estimations. $\omega_{sub-social}$ denotes the dN/dS estimated from the social species branch in branch-wise PAML estimations. ω_{social} denotes the dN/dS estimated from the pairwise PAML estimations of two separate social species populations.

PSMC estimation of social transition time

We used PSMC to reconstruct the historic effective population size for all the six species with a chromosome-level genome. The PSMC method relies on a single diploid individual, where the social species runs out of heterozygosity due to inbreeding. We combine the DNA-seq of two individuals from two separate and diverged populations to retain the loss of heterozygosity.

We benchmarked the fraction of running out of heterozygosity region in the predicted youngest social species *S. dumicola*. We used the GATK pipeline to genotype 9 individuals (5 females and 4 males) from *S. dumicola* and only kept bi-allelic SNP variants among the individuals with a genotype quality filter of 30. We did a sliding window heterozygosity with 100kb window size and 10kb step size on each individual. Windows of an individual without any variants are further classified run out of homozygosity region and summarised into ROH fraction by individual (Table S4).

We ran the PSMC method on autosomes and X chromosomes separately. The generation time is fixed to 1 year per generation. The mutation rate used for autosomes is 5e-09 per site per generation and the rate for X chromosomes is scaled to 3.8e-09 with a factor of 0.76 based on the piX/piA being 0.57 instead of 0.75 in subsocial species *S. africanus* [19]. To align the estimated historical Ne estimated between autosomes and X chromosomes, we retrieve all the time points with an estimation of Ne either for autosomal or X chromosomes, the missing value is filled with the linear interpolate according to the closest values on both sides. Then we calculated the historical NeX/NeA as an indicator for evolution of sex-ratio bias. In the end, the curve of NeX/NeA was smoothed using a spline fit in R. The period of accomplished social transition by retrieving the time interval where the decrease of Ne in social species slows down.

Testing for selection relaxation

Based on the alignment of coding sequences for all filtered autosomal genes, we use HyPhy RELAX to detect the relaxation of intensification of selection on the social species branches. We ran the RELAX test with four different 'test - reference' pair settings for the 2303 autosomal genes and 247 X chromosome genes separately. We have one pair that uses all social species terminal branches as the 'test' and all subsocial species terminal branches as the 'reference' to evaluate the overall effect on selection from the social transition. We have three pairs including each terminal branch of a social species as the 'test' and its sister subsocial species terminal branches as well as common ancestor branches as the 'reference' to test for relaxed selection in each social transition (Supplementary Figure S3).

Acknowledgement

We thank the support from Independent Research Fund Denmark (Grant number: 0135-00201B). We thank Yael Lubin, Tharina Bird, Virginia Settepani, Lena Grinsted and Clemence Rose for help in collecting samples included in this study.

References

1. Rubenstein, D. R. & Abbot, P. *Comparative Social Evolution*. (Cambridge University Press, 2017).
2. Koenig, W. D., Pitelka, F. A., Carmen, W. J., Mumme, R. L. & Stanback, M. T. The evolution of delayed dispersal in cooperative breeders. *Q. Rev. Biol.* **67**, 111–150 (1992).
3. Ross, K. G. & Keller, L. ECOLOGY AND EVOLUTION OF SOCIAL ORGANIZATION: Insights from Fire Ants and Other Highly Eusocial Insects. *Annu. Rev. Ecol. Syst.* **26**, 631–656 (1995).
4. Faulkes, C. G. *et al.* Ecological constraints drive social evolution in the African mole–rats. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **264**, 1619–1627 (1997).
5. Lubin, Y. & Bilde, T. The Evolution of Sociality in Spiders. in *Advances in the Study of Behavior* vol. 37 83–145 (Academic Press, 2007).
6. Alexander, R. D. The Evolution of Social Behavior. *Annu. Rev. Ecol. Syst.* **5**, 325–383 (1974).
7. Brown, C. R. Cliff swallow colonies as information centers. *Science* **234**, 83–85 (1986).
8. Gw, U. Colonial web-building spiders : balancing the costs and benefits of group-living. *The evolution of social behaviour in insects and arachnids* (1997).
9. Majer, M., Holm, C., Lubin, Y. & Bilde, T. Cooperative foraging expands dietary niche but does not offset intra-group competition for resources in social spiders. *Sci. Rep.* **8**, 11828 (2018).
10. Kappeler, P. M., Cremer, S. & Nunn, C. L. Sociality and health: impacts of sociality on disease susceptibility and transmission in animal and human societies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, (2015).
11. Riehl, C. & Frederickson, M. E. Cheating and punishment in cooperative animal

societies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150090 (2016).

12. Garwood, R. J. *et al.* Almost a spider: a 305-million-year-old fossil arachnid and spider origins. *Proc. Biol. Sci.* **283**, 20160125 (2016).
13. Poinar, G. & Buckley, R. Predatory behaviour of the social orb-weaver spider, Geratonephila burmanica n. gen., n. sp. (Araneae: Nephilidae) with its wasp prey, Cascoscelio incassus n. gen., n. sp. (Hymenoptera: Platygastridae) in Early Cretaceous Burmese amber. *Hist. Biol.* **24**, 519–525 (2012).
14. World Spider Catalog, W. S. World spider catalog, version 21.0. *Natural History Museum Bern.*
15. 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).
16. Bilde, T. *et al.* Survival benefits select for group living in a social spider despite reproductive costs. *J. Evol. Biol.* **20**, 2412–2426 (2007).
17. Grinsted, L., Breuker, C. J. & Bilde, T. Cooperative breeding favors maternal investment in size over number of eggs in spiders. *Evolution* **68**, 1961–1973 (2014).
18. Vanthournout, B. *et al.* Male spiders control offspring sex ratio through greater production of female-determining sperm. *Proc. Biol. Sci.* **285**, (2018).
19. Bechsgaard, J. *et al.* Evidence for Faster X Chromosome Evolution in Spiders. *Mol. Biol. Evol.* **36**, 1281–1293 (2019).
20. 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).
21. Settepani, V., Bechsgaard, J. & Bilde, T. Phylogenetic analysis suggests that sociality is associated with reduced effectiveness of selection. *Ecol. Evol.* **6**, 469–477 (2016).
22. Kimura, M., Maruyama, T. & Crow, J. F. THE MUTATION LOAD IN SMALL POPULATIONS. *Genetics* **48**, 1303–1312 (1963).
23. 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).

24. 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).
25. Bechsgaard, J., Jorgensen, T. H., Jønsson, A. K., Schou, M. & Bilde, T. Impaired immune function accompanies social evolution in spiders. *Biol. Lett.* **18**, 20220331 (2022).
26. Berger-Tal, R., Tuni, C., Lubin, Y., Smith, D. & Bilde, T. Fitness consequences of outcrossing in a social spider with an inbreeding mating system. *Evolution* **68**, 343–351 (2014).
27. Avilés, L., Varas, C. & Dyreson, E. Does the African social spider Stegodyphus dumicola control the sex of individual offspring? *Behav. Ecol. Sociobiol.* **46**, 237–243 (1999).
28. Král, J., Kořínková, T., Forman, M. & Krkavcová, L. Insights into the meiotic behavior and evolution of multiple sex chromosome systems in spiders. *Cytogenet. Genome Res.* **133**, 43–66 (2011).
29. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* **3**, lqaa108 (2021).
30. 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).
31. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
32. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
33. Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* **11**, (2022).
34. Wang, Y. & Obbard, D. J. Experimental estimates of germline mutation rate in

eukaryotes: a phylogenetic meta-analysis. *bioRxiv* 2023.01.24.525323 (2023) doi:10.1101/2023.01.24.525323.

35. Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. & Scheffler, K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).
36. Tong, C., Avilés, L., Rayor, L. S., Mikheyev, A. S. & Linksvayer, T. A. Genomic signatures of recent convergent transitions to social life in spiders. *Nat. Commun.* **13**, 6967 (2022).
37. Charlesworth, B., Coyne, J. A. & Barton, N. H. The Relative Rates of Evolution of Sex Chromosomes and Autosomes. *Am. Nat.* **130**, 113–146 (1987).
38. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
39. Johnson, P. C., Llewellyn, K. S. & Amos, W. Microsatellite loci for studying clonal mixing, population structure and inbreeding in a social aphid, Pemphigus spyrothecae (Hemiptera: pemphigidae). *Mol. Ecol.* **9**, 1445–1446 (2000).
40. Chapman, T. W., Crespi, B. J., Kranz, B. D. & Schwarz, M. P. High relatedness and inbreeding at the origin of eusociality in gall-inducing thrips. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1648–1650 (2000).
41. Keller, L., Peer, K., Bernasconi, C., Taborsky, M. & Shuker, D. M. Inbreeding and selection on sex ratio in the bark beetle Xylosandrus germanus. *BMC Evol. Biol.* **11**, 359 (2011).
42. Bulmer, M. G. & Taylor, P. D. Sex ratio under the haystack model. *J. Theor. Biol.* **86**, 83–89 (1980).
43. West, S. *Sex Allocation*. (Princeton University Press, 2009).
44. Trivers, R. L. & Willard, D. E. Natural selection of parental ability to vary the sex ratio of offspring. *Science* **179**, 90–92 (1973).
45. Lindholm, A. K. *et al.* The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends Ecol. Evol.* **31**, 315–326 (2016).

46. Lynch, M., Conery, J. & Burger, R. Mutation Accumulation and the Extinction of Small Populations. *Am. Nat.* **146**, 489–518 (1995).
47. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
48. Dudchenko, O. *et al.* De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
49. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99–101 (2016).
50. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 9451–9457 (2020).
51. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
52. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **Chapter 4**, 4.10.1–4.10.14 (2009).
53. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
54. Liu, S., Aageaard, A., Bechsgaard, J. & Bilde, T. DNA Methylation Patterns in the Social Spider, Stegodyphus dumicola. *Genes* **10**, (2019).
55. Kriventseva, E. V. *et al.* OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **47**, D807–D811 (2019).
56. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
57. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
58. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

59. Vasimuddin, M., Misra, S., Li, H. & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. in *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* 314–324 (2019).
60. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
61. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
62. Ranwez, V., Douzery, E. J. P., Cambon, C., Chantret, N. & Delsuc, F. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Mol. Biol. Evol.* **35**, 2582–2584 (2018).
63. Lemoine, F. & Gascuel, O. Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. *NAR Genom Bioinform* **3**, lqab075 (2021).
64. Rimmer, A. *et al.* Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat. Genet.* **46**, 912–918 (2014).