bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Data-driven analyses of social complexity in bees reveal phenotypic diversification following a major evolutionary transition

Ohad Peled[1], Gili Greenbaum[1,*,†], Guy Bloch[1,*,†]

[1] Department of Ecology, Evolution, and Behavior, The Silberman Institute of Life Sciences, The Hebrew University of Jerusalem; Jerusalem, Israel.

[*] Equal contribution

[†] Correspondence: gil.g@mail.huji.ac.il, guy.bloch@mail.huji.ac.il

Abstract: The evolution of social complexity, from solitary to highly social species, represents one of the most enigmatic phenotypic transitions in the animal kingdom, with insects being a prime example. However, current classification methods for discrete social classes are oversimplified, limiting our ability to quantitatively study the evolutionary changes across the phenotypic space of social complexity. Here, we propose a data-driven approach that is well-suited to overcoming these limitations, allowing us to avoid constraining our inference to a narrow evolutionary trajectory, and to study the full diversity of social phenotypes without condensing social complexity into putative discrete classes. To investigate the evolution of social complexity in bees, we constructed and analyzed a comprehensive dataset encompassing 17 social traits for 77 bee species. We find that corbiculate bees — honey bees, stingless bees, and bumble bees — underwent a major evolutionary transition ~70 mya, which does not follow the commonly assumed "social ladder" route in phenotypic space. We also show that this major transition was followed by a phase of substantial phenotypic diversification of social complexity. In contrast, other bee lineages display a continuum of social complexity, ranging from solitary to simple societies. Bee evolution, therefore, provides a unique demonstration of a macroevolutionary process in which a major transition removed biological constraints and opened novel evolutionary opportunities, driving the exploration of the space of social phenotypes. Our approach can be readily applied to illuminate the evolution of social complexity in additional animal groups.

Introduction

One of the most fascinating examples of increased complexity in biological systems is the evolution of social complexity. Sociality evolved multiple times across different lineages and has played a key role in animal evolution[1–4]. Despite remarkable variation among animal

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

societies[4,5], comparative studies have typically classified levels of social complexity based on a few qualitative traits, such as group composition, reproductive skew, and parental care (e.g., in primates[6–9]; in birds[10,11]; in insects[12–15]). This approach was crucial for the development of influential theories such as 'Kin Selection'[16], 'Major Transitions in Evolution'[17] and 'Developmental Plasticity'[18]. However, focusing on a few qualitative traits forces social phenotypes to fit into a small number of coarsely defined classes (e.g., pair-living, subsocial, eusocial) that represent a set of evolutionary transitions[19–23]. This constraint limits our ability to infer the evolutionary processes leading to complex sociality phenotypes because it imposes a narrow evolutionary trajectory. In addition, oversimplified classification fails to capture the full diversity of social phenotypes within these classes, possibly masking important evolutionary processes such as phenotypic diversification.

Insects from the order Hymenoptera provide a powerful system with which to study the evolution of social complexity because species of the same lineage often show diverse social phenotypes[12,24,25]. Theories on the evolution of social complexity in insects can be divided into two main frameworks. The first emphasizes a single pivotal increase in the level of social complexity that leads to an irreversible "point-of-no-return" ("superorganismality")[13,17,21,26,27]. This framework focuses on caste differentiation processes producing a highly fertile caste (termed "queen" or "king") which functions as the reproductive entity of the superorganism, and "workers" with severely reduced reproductive potential, which are functionally analogous to somatic cells in a multicellular organism. The second framework, commonly termed "social ladder", assumes a stepwise evolutionary trajectory passing through a few key levels of sociality[12,15,28]. The less complex societies are viewed as primitive states, metaphorically comparable to lower rungs on a ladder[28–30]. Despite the long tradition and significant contribution of the social ladder framework, it is still debated whether sociality in insects evolved in similar evolutionary routes across different lineages[19,20,23,31–33].

Both theories have suffered from repeated semantic disagreements regarding their definitions[20,22,23,33], a discussion that hampered the use of comparative methods to identify factors associated with variation in social complexity[19,20,25]. For example, all ants and termites are classified as eusocial despite including species with simple colonies of a few dozen individuals to complex societies composed of millions of individuals[15,34]. Indeed, recent attempts to understand the molecular underpinnings of the evolution of social complexity have had limited success[28,35–40]. This limited success can be partially attributed to the low statistical

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

power of using few traits as proxies for social complexity[19,20], while traits contributing to complex phenotypes may be shaped by various selection pressures, ecological niches, preadaptations, pleiotropic effects, and developmental constraints[41–46].

Here, we adopt a data-driven approach to untangle the evolutionary history of social complexity, which does not assume a putative evolutionary trajectory or social classes. Such a model-free approach may be able to provide a detailed characterization of social complexity and allow analysis of complex interactions among traits in a high-dimensional space[47–50]. We argue that the increase in life history, behavioral, morphological, and molecular data for related species differing in social phenotypes, coupled with improved phylogenetics and computational methods, sets the stage to develop data-driven approaches for the study of social complexity. To investigate and reconstruct macroevolutionary processes underlying social complexity in bees, we compiled and analyzed a comprehensive dataset of life-history, morphological, genetic, and behavioral socially related features.

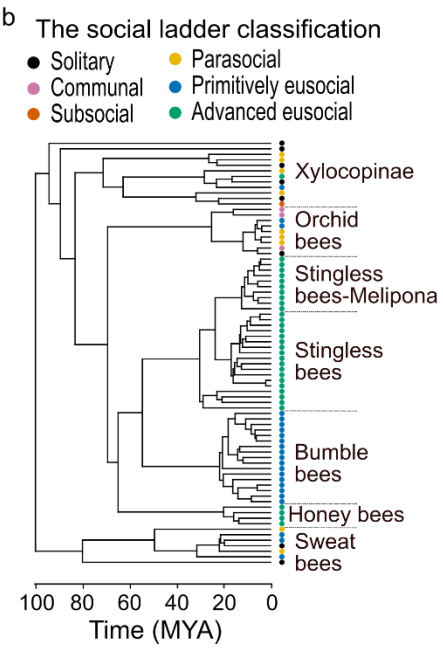
Results

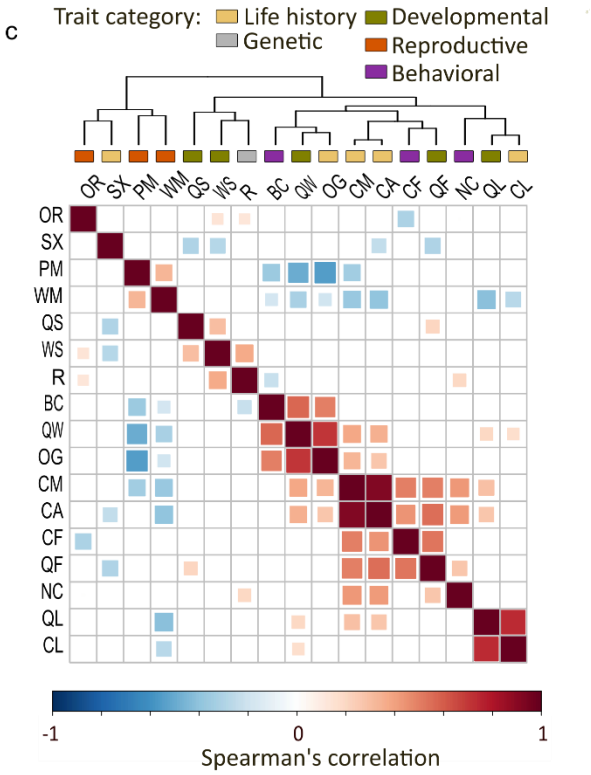
We generated a quantitative dataset composed of 17 social-complexity related traits for 77 species of bees showing diverse levels of social complexity (Fig. 1; see Fig. S1 for the list of species), through an extensive literature survey. We focused on reliable quantitative social traits that can be compared across species (Fig. 1a and Table S1). Our dataset includes colony-level traits (e.g., queen-worker reproductive skew, colony size, colony longevity, type of brood care), life-history characteristics (e.g., overlap of generations, colony founding), and morphological traits (e.g., queen-worker and worker-worker size polymorphism). We did not include traits that cannot be properly compared across species (e.g., behavioral repertoire) or that are not available for many taxa (e.g., detailed morphological and anatomical traits). The 77 bee species in our dataset were selected based on the amount and quality of published information, as well as to appropriately represent the broad spectrums of social complexity and phylogeny (Fig. 1b). Some clades are essentially characterized by a single social complexity classification (e.g., honeybees, stingless bees, bumble bees, Megachilidae solitary bees), whereas other clades include several levels of social complexity (e.g., sweat bees, carpenter bees, allodapine bees, and orchid bees; Fig. 1b).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a

abbrv	Trait
OR	Ovarioles ratio
SX	Sex ratio
PM	% Worker produced males
WM	% Workers mated
QS	Q-W size ratio
WS	Worker size variation
R	Relatedness
BC	Brood care
QW	Q-W polymorphism
OG	Overlap of Generations
CM	Colony size maximum (log)
CA	Colony size average (log)
CF	Colony foundation (log)
QF	Queen fecundity (log)
NC	Nest complexity
QL	Queen longevity
CL	Colony longevity





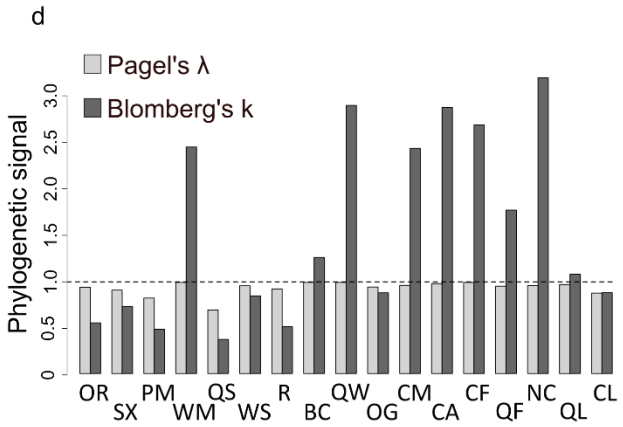


Figure 1. Overview of our dataset, which consist of 17 traits and 77 bee species. **(a)** A list of social complexity related traits in our dataset. **(b)** Time-calibrated phylogenetic tree of the species in our dataset. For species names, see Supp. Fig. S1. Colors correspond to the common social ladder classification of Michener[12]. **(c)** A heatmap summarizing the correlation between social complexity traits in our dataset. The color and size of the boxes correspond to the strength of the Spearman correlation, with positive correlations in red and negative in blue. Only statistically significant correlations are shown (p-value<0.05 after FDR correction). Hierarchical clustering is shown on top of the correlation matrix. The color bars on top of the matrix classify each trait into one of five broadly defined categories. (**d**) The phylogenetic signal for each social complexity trait in our dataset; Pagel's λ in dark gray and Blomberg's K in light gray. The dashed horizontal line delineates the value 1 indicating a neutral Brownian motion process. All traits showed a significant phylogenetic signal after multiple comparison correction (p-value<0.001 for both measures) when compared to a null model where traits evolve independently of phylogeny.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Many traits in our dataset, including colony size-related traits, queen fecundity, and nest complexity are highly correlated (Fig. 1c). Interestingly, the defining attributes of eusociality[15], brood care, overlap of generations, and queen-worker polymorphism were strongly correlated with each other, but not with many of the other social traits in our dataset. Likewise, colony and queen longevity are correlated with each other but not with other traits. Relatedness between females, a putative driver for social evolution, was not correlated with many traits in our dataset. In contrast, the percentage of mated workers was negatively correlated with many other social traits. Overall, the correlation matrix suggests that not all social traits are necessarily positively correlated and that multiple combinations of traits may contribute to an increase in social complexity.

Phylogeny significantly accounted for variation in social complexity. For both Pagel's lambda and Blomberg's K we observed a statistically significant phylogenetic signal ($p<0.001$) for all traits (Fig. 1d). The two traits with the lowest phylogenetic signal were worker-produced males (PM) and queen-worker size ratio (QS). For seven traits, Blomberg's K values were >1: colony-size-related traits (CA, CM, CF, FQ), 'queen-worker polymorphism', and the percentage of mated workers, which is consistent with higher-than-expected similarity based on phylogenetic relatedness alone. These traits are also those that were highly correlated. 'Nest complexity' (NC) showed the highest Blomberg's K value, probably due to its high occurrence in Meliponini bees, which build remarkably elaborate nests. The strong phylogenetic signals for the social traits are consistent with the occurrence of similarly high levels of social complexity in honeybees, stingless bees, and bumble bees which are represented by a relatively large number of species in our database (Fig. 1b).

The phenotypic space of social complexity in bees

We used dimensionality reduction analyses to describe the phenotypic space of social complexity using the 17 traits in our dataset. We used both principal component analysis (PCA) and uniform manifold approximation and projection (UMAP). These analyses are complementary in that PCA describes the global structure of the phenotypes of the species in our dataset, whereas UMAP emphasizes within-group variation, which can add fine resolution to clustering in the phenotypic space. Given the significant phylogenetic signals for all traits in our dataset (Fig. 1d), we applied phylogenetic correction for all downstream analyses. The clustering in UMAP was consistent over multiple iterations and parameters (Fig. S2), indicating that the overall structure of the data we obtained is robust regarding the relations

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between species. In the PCA, the 'broken stick' method[51] suggested four potentially informative PCs which together explain ~70% of the variation. PC1 and PC2 accounted for 28% and 15% of the variation, respectively (Fig. 2a; data on all four main principal components are shown in Figs. S3 and S4). Both the PCA and UMAP identify clear clusters of species occupying demarcated areas in the phenotypic space, but the composition and structure of the clusters were somewhat different across analyses (Fig. 2a).

In UMAP, clustering suggested a clear separation between (i) honey bees and stingless bees (ii) bumble bees, and (iii) the remaining lineages. In the PCA plots (Fig. 2a and S2), bumble bees and stingless bees overlapped in PC1 and PC2 but not in PC3 and PC4, whereas honeybees were clustered separately. In addition, solitary, communal, and subsocial species were clustered separately from parasocial (often termed `facultatively social') and non-*Bombus* primitively eusocial species in most of the PC combinations. Considering both PCA and UMAP, we identified four main phenotypic groups that occupy well-defined areas in the phenotypic space (hereafter 'groups'), representing different social complexity phenotypes (grey shaded areas labeled A-D in Fig. 2a-b). The main groups correspond to: (A) solitary, communal, and subsocial species; (B) parasocial and non-*Bombus* primitively eusocial species; (C) bumble bees; (D) honey bees and stingless bees. The clearest and widest separation in both the UMAP and PCA is between the monophyletic group of corbiculate bees — honey bees, stingless bees, and bumble bees (groups C and D; hereafter corbiculate bees) — to the rest of the species (groups A and B, Fig. 2a-b). This partition reflects the separation of species with obligate sociality (C and D) from species with facultative or no social lifestyle (A and B). Notably, facultative and obligate sociality was not one of the traits included in our dataset, indicating that our data-driven approach successfully captured key attributes of social complexity that were not explicitly defined.

The clusters in our phenotypic space differ from the commonly used social ladder classification (color-coded in the phenotypic spaces in Figs 2, S3, S4, S5). Bees that are assigned to different qualitative levels of sociality clustered together in our phenotypic space and increase from solitary lifestyle (group A in Figs 2a-b) to parasocial and non-*Bombus* primitively eusocial species (group B) was gradual rather than stepwise as predicted by the social ladder model. Another apparent disagreement with social ladder models is that species that are classified as primitively eusocial do not occupy a defined phenotypic region, but rather are separated into two discrete clusters: one that includes bumble bees, and the other that includes the other

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

primitively eusocial species as well as some species that are typically classified as parasocial (Figs. 2a-b, S2, S4). *Exonourella tridentata* (Allodapini, typically classified as advanced eusocial) and *Lasioglossum marginatum* (Halictidae; typically classified as primitively eusocial) were the only species in groups C and D that do not belong to the corbiculate bees. These two species were associated with different groups (clusters) across analyses (Fig. S4 and S5). Notably, *Bombus atratus* (marked in Fig. 2a) is positioned in the margins of the bumble bee cluster, attributed to its perennial colonies which can be larger than those of some of the stingless bees.

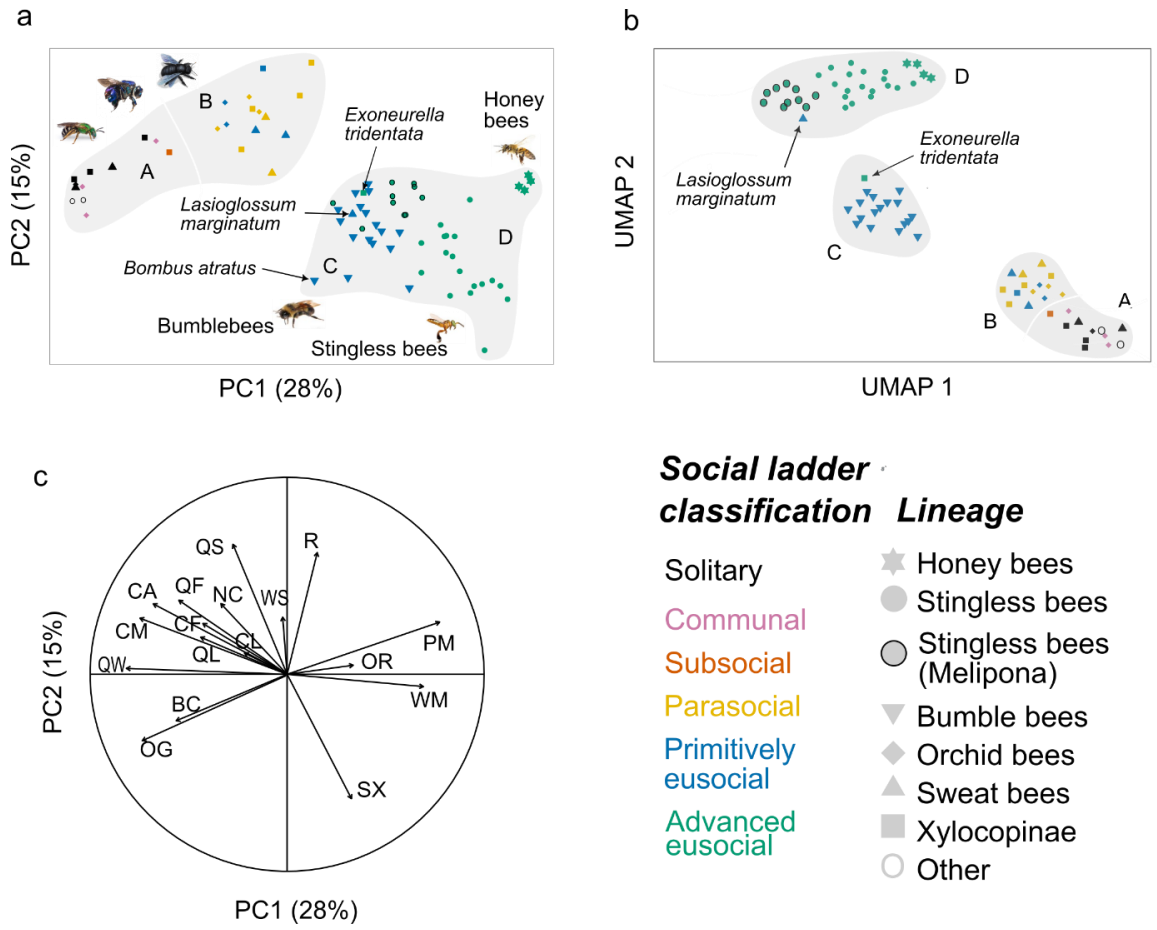


Figure 2. The phenotypic space of social complexity in bees. (a) Phylogenetically corrected Principal Component Analysis (PCA). The plot shows Principal Components 1 and 2 (PC1 and PC2), which together account for 43% of the variation in our dataset. Symbol color corresponds to the commonly used social classification and its shape to the species taxonomic clade (legend on the bottom right). *Habropoda labriosa* and *Megachile rotundata* are not included in the main clades in our database and were therefore marked as 'other'. (b) Phylogenetically corrected Uniform Manifold Approximation and Projection (UMAP). We visually identified four putative phenotypic groups, marked with capital letters A to D. (c) The variable correlation plot describes the contribution of traits to PC1 and PC2, with label abbreviations detailed in Fig. 1a. The direction of each vector corresponds to the PCA in (a) and indicates the proportion of its contribution to PC1 and PC2, and its length indicates the amount of contribution.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In the PCA, the species clusters are not tight, with each occupying a relatively large area in the phenotypic space. Notably, the within-group variation is as large as the between-group variation (i.e., the size of the region of each group in the PCA is as large as the distances between the groups), highlighting significant variation in social complexity that is entirely overlooked in qualitative classifications of social complexity. One clear example of this is the large region occupied by advanced eusocial species, suggesting that there is substantial variation among species typically assigned to this class (group D in Figs. 2, S2, and S4). We computed the convex hulls of regions in the phenotypic spaces occupied by corbiculate bees and the remaining species. For PC1 and PC2, the area occupied by sweat bees, orchid bees, and Xylocopinae bees was slightly larger than the corbiculate bees (14.1 compared to 12.8, respectively; Fig. S6). However, considering all four main PCs, the area occupied by corbiculate bees is significantly larger (t-test, p-value=0.003) with mean±s.d. of 17.8±4.5 compared to 8.9±2.9. Our analyses suggest that group D can be further partitioned into three subgroups: honey bees, *Melipona* genus stingless bees, and non-*Melipona* stingless bees (Figs. 2a and 2b). The bumble bee group (marked C in the UMAP analyses, Fig. S2) was consistently partitioned along the common distinction between mass provisioners ('pocket-making') and progressive provisioners ('pollen-storer'). Analyzing the correlation of traits with the PCs (Figs. 2c and S3) suggests that the variation characterizing the corbiculate bees is attributed to many of the traits in our dataset, whereas variation in groups A and B (e.g., sweat bees, orchid bees, and Xylocopinae bees) stemmed primarily from variability in brood care (BC), the percentage of worker-produced males (PM), overlapping generations (OG), sex ratio (SX), and relatedness (R) (Fig. 2c).

The evolutionary history of social complexity in bees

To gain insight into the evolutionary history of social complexity, we conducted an ancestral state reconstruction using principal component values as an estimate for the social complexity phenotype (Fig. 3a and S7). To visualize the evolutionary reconstruction in more than a single dimension, we projected the ancestral state reconstruction into different combinations of the four main PCs (Figs. 3b and S8). There are pronounced changes in the social complexity phenotype during evolution in species from groups A+B along PC1 and PC2 (Fig. 3b). This flexibility is less evident in PC3 and PC4 (Figs. S7 and S8). The assumed evolutionary trajectory leading to the corbiculate bees demonstrated a relatively consistent increase in social complexity values along PC1 over time (Fig. 3a). We identified a presumed "point-of-no-

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

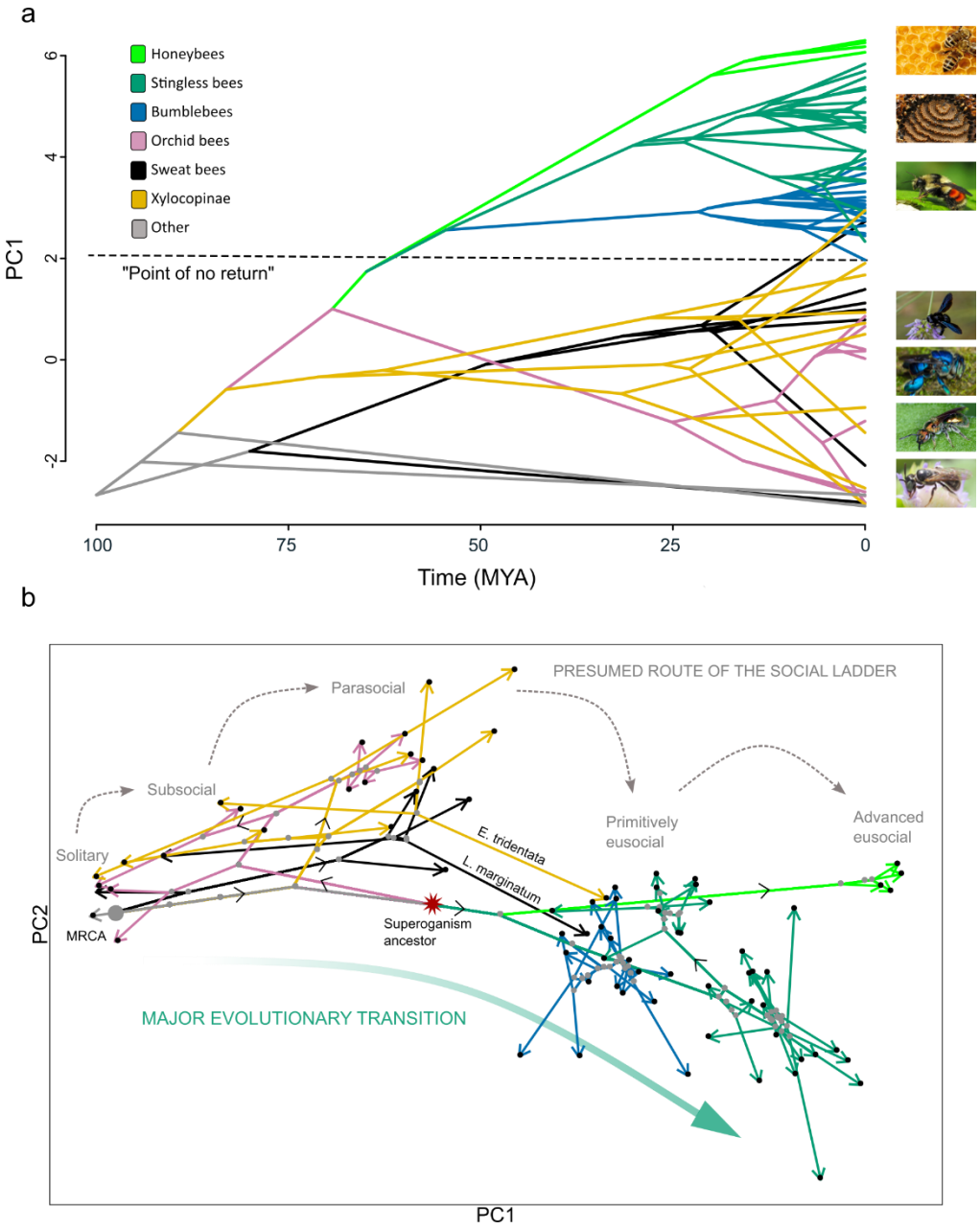


Figure 3. Ancestral state reconstruction for the evolution of social complexity. Values were obtained from the phylogenetically corrected PCA (Figure 2a). **(a)** A phenogram of the evolutionary trajectory of PC1 over time. The root was set to a minimum value based on the solitary *M. rotundata*. Colors represent the main lineages in our dataset, broadly aligning with the color scheme in Fig. 2. A dashed line indicates a suggested 'point-of-no-return,' beyond which species cannot revert to lower levels of social complexity. **(b)** Integration of the PCA phenotypic space from Figure 2a with ancestral state reconstruction analysis of PC1 and PC2. Arrows depict the inferred evolutionary route of species within the phenotypic space. The arrowheads and the black dots at their end point mark the position of extant species in our dataset (as in Fig. 2a). Gray dots denote the assumed social phenotype of putative ancestral species (i.e., internal nodes in the phylogeny). large gray dot depicts the most recent common ancestor of all species (MRCA). The red star marks the putative common ancestor of corbiculate bees. Gray dashed arrows illustrate an attempt to fit a stepwise increase in social complexity as predicted by social ladder models. The green thick arrow suggests a putative trajectory consistent with a major transition in social complexity. Similar analyses for PC1-4 are presented in Fig. S8).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

return" that is consistent with an irreversible phenotypic threshold in PC1 and was crossed by the corbiculate bees, as well as by *E. tridentata* and *L. marginatum*. The evolutionary reconstruction of PC2, with high variation contributed by the traits Q-W size ratio, sex ratio, and relatedness, also shows an early separation of the corbiculate bees from the other species. After the common ancestor of corbiculate bees showed a significant change in social complexity separating it from the other bee lineages in PC1 and PC2, the social complexity of bumble bees, stingless bees, and honey bees further diversified within the phenotypic space. For example, the Melipona stingless bees lineage diverged from the rest of the genera to occupy a distinct phenotypic space (Figs. 3b). The corbiculate bees show the largest divergence during the reconstruction of PC3 and PC4 (Fig. S7 and S8). Our evolutionary reconstruction suggests that the social complexity of *E. tridentata* and *L. marginatum* reflects a later change in social complexity through different evolutionary routes than that leading to high social complexity in corbiculate bees (Figs. 3b, S7, S8).

A shift detection analysis using the values of PC1 and PC2, revealed a single major evolutionary shift, which occurred in the ancestor of corbiculate bees (green dot in Fig. 4). An analysis incorporating actual social complexity traits, rather than inferred PC values, also identified a single shift. (Fig. S10b). Models assuming a larger number of shifts, as hypothesized by social ladder models, received significantly less support (Fig. S9). We did find additional phenotypic shifts when including all four main PCs, but these were in the specific

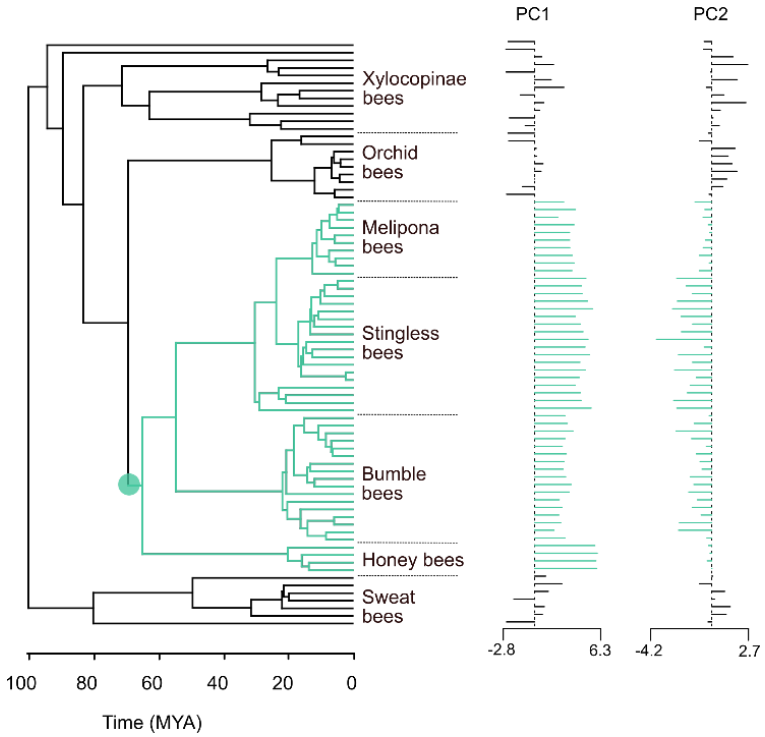


Figure 4. Shift detection analysis. On the left, a single shift detected on the phylogenetic tree. The green circle indicates the inferred position of adaptive regime shift from PCs 1 and 2, marking the ancestor of corbiculate bees. On the right, the PC1 and PC2 values for each species. Additional analysis for PCs 1-4 is shown in Fig. S10a.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

branches leading to honey bees, stingless bees, *Melipona* genus, and bumble bees (fig. S10a). This suggests that PC3 and PC4 offer a more detailed description of phenotypic changes in these lineages. Sensitivity analyses suggested that the phenotypic shift we identified is not biased by the taxonomy representation, missing data, or specific traits in our dataset. (Figs. S11-S13). We, therefore, conclude that after the divergence of corbiculate bees from the remaining species, corbiculate lineages have crossed a major evolutionary transition in social complexity and followed a different evolutionary trajectory compared with other social species that have not crossed this threshold.

Evolutionary transition leading to phenotypic diversification

To quantitatively study the directionality of the evolutionary trajectory and phenotypic diversifications, we analyzed the angle of change and the size of each step in the phenotypic space in the reconstructed trajectories (illustrated in Fig. 5a). We used the trajectories in the phenotypic space of PCs 1-4. In terms of the changes in direction in the phenotypic spaces (angels between two "steps"), we hypothesized that after the divergence of the corbiculate bees from the remaining lineages, corbiculate bees would present a more directional and constrained evolution indicated by degrees closer to π, and lower variance of the degrees, compared with non-corbiculate species (Fig. 5a).

Before the divergence of the corbiculate bees, we observe a high variance of the degree distribution for all species (Fig. 5b). However, the trajectories of corbiculate bees after the split showed a statistically significant lower variance of the degree distribution compared with the rest of the species (F-test, 7.31, $p<0.001$), which was seven times higher. We found a higher and closer-to-zero mean angle in the reconstructed evolutionary trajectory of the corbiculate bees compared with species of other lineages after their divergence (radians [mean, s.d.] = 3.01±0.2 and 2.5±0.54 for corbiculate and other lineages, respectively; t-test, $p<0.001$), and higher but not significantly compared with the mean angle of all species before divergence (2.81±0.43; $p=0.36$). These findings are consistent with the hypothesis that the increase in social complexity over time was more directional and less flexible in the corbiculate bees compared to the lineages leading to the sweat bees, orchid bees, and Xylocopinae species.

Undergoing a major transition can remove biological constraints and open novel evolutionary opportunities, generating a phase in which the phenotypic space is rapidly explored and phenotypic diversification is increased. We found that the rate of phenotypic diversification in

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the corbiculate bees indeed substantially increased ~70 mya, coinciding with the directed adaptive regime shift (green line in Fig. 5c). The phenotypic diversification of social complexity in these lineages quadrupled in rate for a short period, then decreased to about a half. In contrast, the remaining lineages show a rather constant low rate of diversification throughout their evolutionary history. The later increase at ~25 mya is attributed to the flexible and rapid changes in social complexity in these lineages occurring closer to the present, which is similar for corbiculate and non-corbiculate bees.

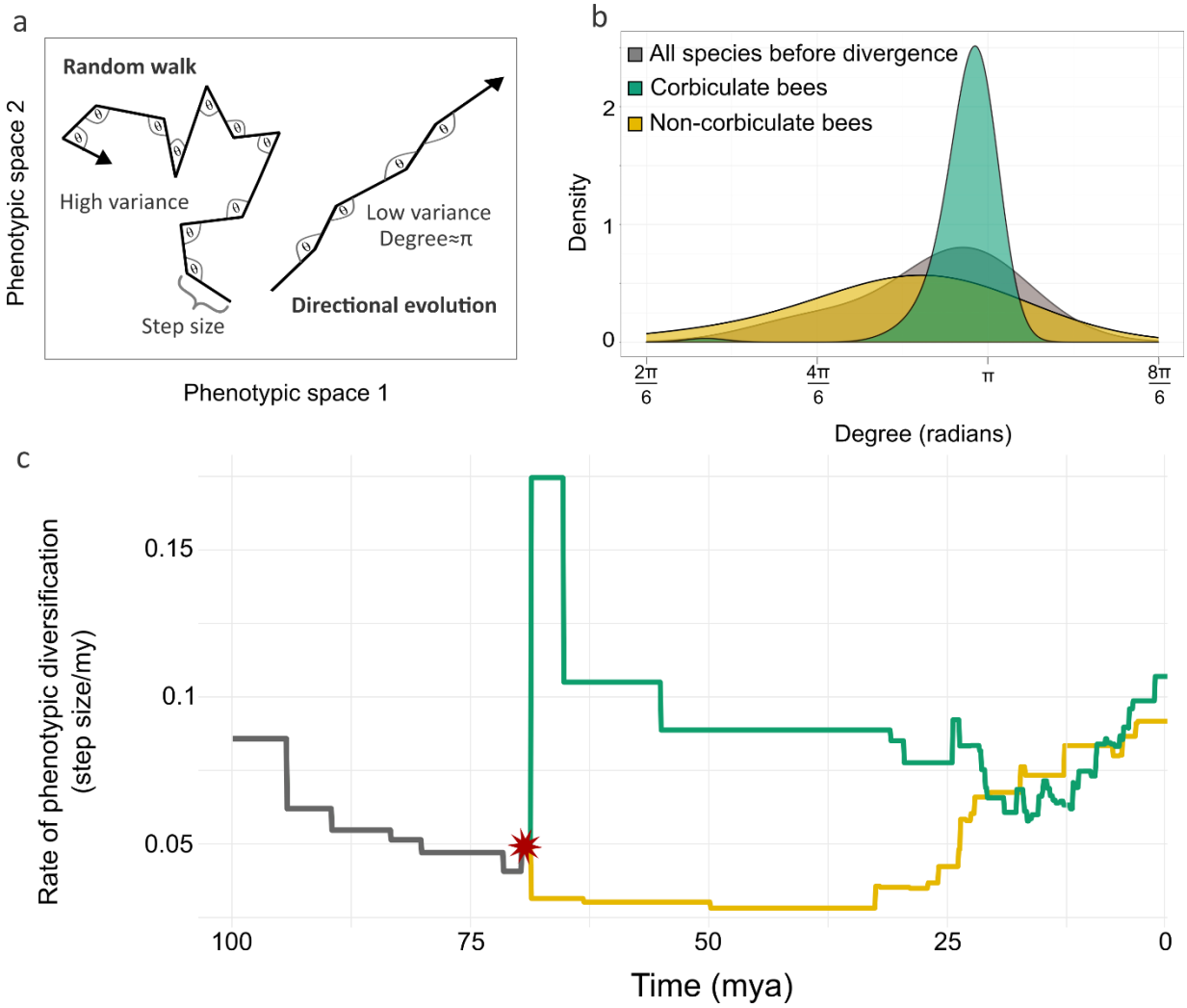


Figure 5. Analysis of evolutionary trajectories in the phenotypic space of social complexity. Analyses are based on PC 1-4 (shown in Fig S8). **(a)** Schematic representation of evolutionary routes under a directional evolution and a random walk. The mean and variance of the angles are used as measures for the directionality in the evolution of social complexity. The step size **(b)** Density plot of the turn degrees in the reconstructed evolutionary routes for the corbiculate bees in green, and the remaining lineages in yellow. **(c)** The phenotypic diversification rate across time before and after the inferred major evolutionary transition (indicated with a red star). The grey line shows the average diversification of all species before the split; after the split, the rate of corbiculate bees is shown in green and that of the remaining lineages in yellow. The rate of phenotypic diversification is computed by dividing the lengths of the "steps" in the phenotypic spaces by the phylogenetic branch length.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Testing for selection signals in socially related genes

Lastly, we studied the selection on several genes that have been repeatedly linked to the evolution of sociality in insects[35,52–55], using the values of our main PCs as proxies for social complexity. We focused on insulin pathway related genes (i.e., *IRS, InR, InR-2*), *TOR, syx1a*, and *Dunce* (Table S2). To detect signals of selection in these genes, we computed the molecular evolutionary rate (dN\dS values) for 10-18 species for which reliable genomic sequencing data was available (Table S3). We tested whether phylogeny, the social ladder categories[12,28] or the main groups (A-D) in our dimensionality reduction analysis best explain the pattern of selection (Tables S4 and S5). For *InR-2*, we found a signal consistent with purifying selection in bumble bees (group C). For IRS and InR we detected a low signal for positive selection in *E. tridentata*. We found a similar low signal for *syx1a* in species from group A and for TOR in a single subsocial species *C.calcarata*. Lastly, we identified a strong signal of positive selection signal for *Dunce in* bumble bees (group C). Generally, our findings indicate that the selection patterns in focal social complexity genes are better explained by the phenotypic regions we identified and the taxonomy of species rather than by the social ladder classification (Tables S4 and S5). Linear regression analyses testing the correlation between the dN\dS value of each gene and PC1-4 in our PCA revealed a significant correlation only for *Dunce* with PC4, and for *TOR* with PC1 (Table S6). These findings do not support the hypothesis that the selection on several key "social genes" underpinned the evolution of social complexity across multiple independent origins of sociality in bees[28].

Discussion

We adopted a bottom-up data-driven approach to analyze socially related traits, generating a phenotypic space of social complexity for 77 species of bees. The use of multiple quantitative traits enabled us to avoid semantic inconsistencies and *a priori* assumptions that are inherent to qualitative classifications, which currently dominate the research on the evolution of social complexity. We identified clear clusters of species occupying defined areas in the social complexity phenotypic space, which correspond to different social phenotypes (Figs. 2, S2-S5). Importantly, we also identified substantial overlooked variability within each of these clusters. Our ancestral state reconstructions are consistent with a single major transition of increase in social complexity, which occurred in the common ancestors of honey bees, stingless bees, and bumble bees (Figs. 3-5). This transition into complex societies was followed by considerable phenotypic diversification (Fig. 5c). Notably, bumble bees align more closely

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with honey and stingless bees in our phenotypic space, and are clearly separated from other "primitively eusocial" species. This is consistent with some of previous suggestions regarding the classification of bumble bee sociality[20,21,25,33]. Conversely, we observe that non-corbiculate "eusocial" species followed a different route towards high sociality than the corbiculate bees.

Our findings are consistent with fossil evidence suggesting that complex sociality in corbiculate bees dates back to their common ancestor, as observed in ~45 mya amber fossils[56,57]. These fossils from extinct corbiculate tribes show evidence of a clear worker caste, which had a barbed sting and reduced metasoma[56,57]. Notably, there are no other bee lineages with fossil evidence of a morphological caste system. Additionally, no extant corbiculate bee species have reverted from social to solitary lifestyles (the *Psithyrus* Cuckoo bumble bees evolved a parasitic lifestyle that is not strictly solitary). This suggests that high levels of social complexity form an evolutionary adaptive peak in the phenotypic landscape[58,59], making a reversion to a less complex lifestyle unlikely[27]. Colonies of corbiculate bees have clear castes and the point-of-no-return is assumed to be crossed when unconditional differentiation of permanently unmated castes is established[13,21]. After crossing this threshold, a positive feedback process is assumed to support further development of key social traits such as finer division of labor, elaborated communication systems, and an increase in colony survival[60–64].

The elaboration of trait interactions is manifested in the corbiculate bees and is highlighted by extensive phenotypic diversity of social complexity. Our results indicate that extant corbiculate diversity is the result of phenotypic diversification following an adaptive and directional evolutionary regime shift approximately 70 mya (Figs. 4 and 5). Based on previous conservative estimates of the fossil record, clade diversity among social corbiculate bees was higher in the past, with at least a 50% reduction in eusocial bee clades during the Eocene (55-35 mya)[56,57]. This trend suggests that a major transition has opened novel evolutionary opportunities, driving the occupation of new niches in the social complexity phenotypic landscape. Such a macroevolutionary process is likely to include the breaching of structural, developmental, or ecological barriers, similar to the burst of multicellular organisms after the development of sexual reproduction and division of labor[65,66] or the rapid speciation after the colonization of a remote island[67,68].

In sweat bees, orchid bees, and Xylocopinae bees, we observe lower levels of social complexity compared to corbiculate bees. We found evidence for repeated reversions to low levels of social complexity (from group B to A), supporting earlier suggestions based on qualitative

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

classifications[23,25,69–74]. The reversion to lower social complexity suggests that in these groups the point-of-no-return has not been traversed. This notion is independently supported by evidence for intra-species variability in social complexity with solitary and social lifestyles expressed under different environmental conditions[75–78]. The evolutionary increase in social complexity in these lineages was relatively continuous, with no evidence for stepwise incremental changes. Two species in these lineages, *E. tridentata* and *L. marginatum*, are typically classified as eusocial; however, in our analyses, it is unclear whether their social complexity is indeed comparable to that of corbiculate bees. The social phenotype of these two species was inconsistent across our dimensionality reduction analyses (Figs. 2, S2, S4, S5) and their evolutionary trajectories in our phenotypic space differ from that of the corbiculate bees, as well as from their close sister taxa. These findings most likely reflect their unusual nesting biology, as *E. tridentata* displays facultative social nesting with reproductive workers[79] and *L. marginatum* has no apparent queen-worker castes[80]. However, because the social biology of these species is understudied, additional investigations are needed to robustly reconstruct their evolutionary history[36,81,82].

The increased availability of life history and biological data for a growing number of bee species allowed us, for the first time, to describe the phenotypic space of social complexity in bees and quantitively study it. Our data-driven approach circumvents the need to rely on explicit or implicit assumptions, facilitating unconstrained evolutionary inference[19,20,23,31–33,83]. Contrary to expectations of the social ladder perspective, our findings are not consistent with a continuous[25] or stepwise progression towards higher sociality across independent bee lineages[12,28,84]. Additionally, we did not find support for the notion that key 'social genes' are selected across independent origins of sociality as would be expected in a narrow and restricted evolutionary trajectory towards sociality[28–30,85,86]. Instead, our findings shed light on an important macroevolutionary process – traversal of a substantial phenotypic distance and subsequent diversification within corbiculate bees. Our framework can be expanded to additional lineages, including primates and birds[6–8], in which research on the evolution of social complexity suffers from the same limitations as in social insects. Our approach is especially promising for studies on the molecular underpinnings of social complexity because it is more likely to capture the relationship between proxies of social complexity and quantitative molecular data.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Unless stated otherwise, analyses described below were performed using R v.4.2.1[87].

Curating a dataset of social complexity traits in bees

We conducted an extensive literature survey using "Google Scholar" for any paper that includes relevant names of species and specific search words for each trait. Table S1 details the social traits that we used, their ecological context (e.g., field, lab, etc.), the keywords used to find relevant publications, and the references supporting the significance of the trait to social complexity. Overall, we surveyed >1000 articles, of which 219 were identified as providing relevant information, and were incorporated into our dataset. Based on the biology and life history of each focal species, we designed strict and consistent quality control guidelines to minimize the risk of including erroneous data that could introduce noise into our dataset (see Table S1 for details). For example, for "average colony size" to be included in the dataset, we required that it be measured for a healthy colony sampled during favorable conditions (e.g., in the spring) when food sources are abundant and the climate is suitable for colony growth. Similarly, "colony longevity" values were included only when measured in a natural setting or a field study, without intervention (e.g., supply of artificial food, disease control, etc.). Our dataset provides a comprehensive description of the sociality phenotype of species according to their ecological and evolutionary background. In species with social polymorphism (i.e., when individuals can nest solitarily or socially), we used data for the most putatively "advanced" social form, unless stated otherwise.

We included in the dataset traits that are commonly used in the literature and which describe different aspects of species sociality. These measures are more available, allowing us to obtain consistent data across more species. When determining which traits to include, we aimed to choose quantitative, informative, and comparable traits. For example, the percentage of workers with "active ovaries" can provide quantitative information on the degree of reproductive skew in a colony; however, the degree of "ovary activation\development" is often only qualitatively described and is measured differently across species. Therefore, this trait was rejected as a comparable measure for our dataset. Furthermore, some traits are similar and complementary to one another, allowing us to obtain a comprehensive description of sociality aspects instead of assuming that one trait is indicative of another. For example, to measure the degree of reproductive skew in a colony we used the percentage of mated workers, which indicates the totipotency of females, as well as the percentage of worker-produced males, which represents a more functional measure of the reproductive skew. Several traits that cannot be

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

quantified but are important in the literature on insect sociality were coded as ordinal values. Because our correlation matrix (Fig. 1c) showed that most traits are not highly correlated, the inclusion of similar traits does not inflate their importance in downstream analyses. To verify that no single trait has a dramatic impact on our results, we conducted a sensitivity analysis in which each trait was removed from the dataset (Fig. S11).

We included in our dataset species from all bee lineages in which considerable social behavior has been reported (Apinae, Xylocopinae, Lasioglossum, Halictus, and Augochlorini). In choosing the species, we aimed to include as much variation as possible across the social complexity phenotype of bees. Therefore, our focus was more on sampling many different sociality phenotypes rather than extensively sampling the phylogeny. For example, stingless bees were found to be excellent candidates for diverse sampling and were indeed overrepresented in our dataset. On the other hand, many social species of sweat bees and Xylocopineae bees, which do not add additional information (i.e., they are similar to the species we included), were not overly sampled. Only two known species: *Hasinamelissa minuta* (Allodapine) and *Lasioglossum umbripenne* (Halictidae), might represent a different sociality type[12,88]; however, insufficient data did not allow their inclusion, and the available data suggested they follow a pattern similar to E. tridentata and L. marginatum, at best. Furthermore, our sensitivity analyses with more equal sampling (Fig. S12) only reduced the variation within the phenotypic space but did not alter our results.

For each species in our dataset, we also provide the common social complexity category using the common classification of Michener[12]. Because we wanted to minimize *a priori* assumptions, we did not include these classifications in our analyses but only used them as a reference to which we compared our findings. Overall, we gathered information on 77 Species and 17 social traits, with 15% missing data. Four indices: colony average, colony maximum, colony foundation, and queen's fecundity were log-transformed due to a high degree of variation that could skew the results (e.g., 1-80,000 for colony maximum). Additional details on the data curation procedure and the social traits and species included in our dataset are provided in Tables S1 and Dataset 1.

To impute missing data, we used methods that offer sufficient statistical power while preserving the original data distribution. For example, we avoid approaches such as omitting all species with missing data or replacing missing data with the mean of the trait value, because this may bias or distort data distribution, respectively[89]. Based on our data structure, we used

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

iterative PCA for imputation, which was found to perform well for data structures similar to ours (dataset 1) and especially for PCA analyses[89]. For missing data imputation, we used the R package 'missMDA'[90]. Each imputed data value was manually examined to verify that it was within the accepted ranges found in the literature on the relevant species. For example, we ensured that imputed values for colony longevity of known annual species are indeed of up to a year. 14 Imputed entries (1% of the entries in our dataset) that did not meet this requirement or were not legitimate (e.g., negative values) were replaced manually with an approximated value based on the literature (e.g., species with colony longevity of 'few months' without more accurate information was assigned a value of three months).

Correlations and phylogenetic patterns of traits in the dataset

We used two methods to estimate the phylogenetic signal in the dataset: (i) Pagel's λ[91], which is estimated through maximum likelihood procedures using the 'phylosig' function from the R package 'phytools'[92]. Pagel's λ value ranges from 0 (the trait evolved independently from phylogeny) to 1 (the trait's value fully depends on phylogeny and its evolution fits a neutral Brownian motion); (ii) Blomberg's K[93], which is not constrained to be between values 0 and 1, and K>1 implies that species are more similar to one another than expected under a Brownian motion process[93]. In other words, values higher than 1 indicate a constrained, lineage-specific trait. Additionally, the flexibility of Blomberg's K allows us to capture the effects of changing evolutionary rates (e.g., non-Brownian motion processes; Münkemüller et al., 2012). For both λ and K, we performed a likelihood ratio test to evaluate the statistical significance of the phylogenetic signal[92].

To test whether the different social traits are correlated to one another, as implied by the 'social ladder' model, we computed the Spearman's correlation coefficient between all trait pairs, while accounting for the shared evolutionary history between the species (i.e., non-independent data points). We first computed the phylogenetically independent contrasts (PICs) using the R package 'ape'[95], which implements the method described by Felsenstein[42]. This method computes the contrasts, or differences, in traits between each pair of sister taxa, generating a set of statistically independent values. We then computed Spearman's correlation coefficients for these contrasts, allowing us to measure the strength and direction of the monotonic relationship between two variables without the need to assume normally distributed and linearly related variables. To account for the possibility of spurious correlations, we used the Benjamini–Hochberg method, which controls the false discovery rate (FDR) using sequential

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

modified Bonferroni correction for multiple hypothesis testing[96] as implemented in the 'p.adjust' function. Only significant correlations (with a significance level of α=0.05) were considered as indicating correlated traits.

Dimensionality reduction visualization

We used two procedures of dimension reduction, PCA and UMAP, to visualize and investigate the social phenotypic space of the species in our dataset. Principal component analysis (PCA) maximizes the variation in principal components and is useful for visualizing and quantifying a phenotypic space that captures the global structure of the dataset along a few major axes. In contrast, Uniform Manifold Approximation and Projection (UMAP) emphasizes local structures (e.g., within-group variation) in the dataset, while still accounting for the global structure (i.e., the between-group variation), and is more effective in revealing tight clusters[97]. The combination of PCA and UMAP analyses provides us with a comprehensive reflection of a phenotypic space from our social-trait dataset (other dimensionality reduction techniques that focus on local structures and clusters, such as t-SNE, would be less useful for this purpose).

For PCA, we used the R packages 'FactoMineR' for visualization[98], and 'MissMDA' for performing PCA with missing data[90]. Before analysis, we normalized the traits to z-scores (i.e., mean=0 and Standard Deviation=1). To correct the PCA for phylogeny we applied pPCA using the 'phyl.pca' function of the R package 'phytools'[92], which estimates ancestral states of the traits using maximum likelihood under the λ model. This model allows variation in rates of evolution through time[91] and was preferred over the passive Brownian motion model with equal rates. Finally, we used the broken stick method, as implemented in the R package 'PCDimension'[51], to determine the number of statistically significant PCs. This is done by randomly selecting n breakpoints (n being the number of variables in the dataset) from a uniform distribution; a PC is considered significant if it explains more variance than the randomly "broken sticks". To understand the effect of traits on our phenotypic space, we examined the PCA loadings. Lastly, we computed the Euclidean space of the convex hull occupied by species in the phenotypic space of all PC combinations and compared the area of corbiculate bees to the area of the remaining lineages. Analysis of UMAP was implemented in the R package 'umap'[99]. Because phylogenetic correction has not been developed for UMAP, we established a phylogenetic procedure that uses the methods for PCA phylogenetic correction. This was achieved by taking the phylogenetically corrected PCA values from the first ten principal components and using them as input for the UMAP algorithm.

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Estimating the evolutionary routes of social complexity

Based on the common assumption that the ancestral lifestyle of bees was solitary[23,24], we investigated the phenotypic routes by which social complexity increased throughout bee evolution. We performed ancestral state reconstruction using the *ace* function in the 'ape' package in R[95] and considered the PCA values (PC1-PC4) as describing the sociality phenotype of the species, which together represent the majority of variation of social complexity found in our dataset (~70% of the variation). Because UMAP, unlike PCA, allows for data-dense regions to be "stretched out" in the representation, it is more challenging to interpret distances between species in the UMAP space, and therefore we did not use UMAP values for downstream analyses[100,101]. We set the root to the values of the solitary *Megachile rotundata*, which, unlike other solitary species in our dataset, belongs to a strictly solitary family and is more likely to represent a common phenotype of solitary bees. Fixing the root provides more statistical power because nodes deeper in the tree tend to be harder to estimate without *a priori* knowledge (e.g., fossils), and directional evolution, if exists, is difficult to capture[102].

Identifying shifts in social complexity

Maximum likelihood model fitting approaches have been widely used to quantify and compare evolutionary patterns in continuous traits (e.g., geometric morphometrics, functional morphology, biomechanical indices, etc.) based on categorical explanatory states (e.g., diet, habitat, life history, etc.[103–107]). These methods are based on the adaptive evolution process of Ornstein–Uhlenbeck (OU)[59,108]. OU models are a modified Brownian motion process where the trait is attracted toward an optimum value. If the trait value at the root of the phylogeny is different from the optimum, the mean trait value of the lineages will tend to increase or decrease over time, eventually converging to around the optimum. In other words, OU models describe processes where trait values are constrained around one or several optima that can be considered adaptive peaks in a macroevolutionary landscape. However, this model-fitting approach still relies on a priori assumptions for species classification and, more importantly, forces qualitative transitions between states, which might not always be appropriate for reversible and small fluctuations in trait values. Recently developed methods[109–112] have applied similar multi-peak OU models to detect shifts in the pattern of morphological evolution, creating an adaptive landscape of trait values that reflects Simpsons's framework of "adaptive zones" across multivariate trait spaces and multiple selective optima. This approach is particularly relevant for our framework because it does not require an a priori hypothesis of

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

shift locations (e.g., with changes in ecology or environments), nor classification of species and directionality of transitions between states (e.g., based on their social complexity level).

We used the 'PhylogeneticEM' R package – a shift detection method that identifies phenotypic shifts in social complexity along the phylogeny[112,113]. To account for trait correlations (i.e., interdependencies) 'PhylogeneticEM' uses a simplified scalar OU (scOU) model which assumes that all traits evolve at the same rate towards their optimal values. This assumption facilitates the inference of the selection strength and drift rate for traits. The evolutionary trajectory of traits is modeled on a fixed, time-calibrated phylogenetic tree, with shifts in traits interpreted as responses to developmental, environmental, or phylogenetic changes. These shifts are integrated into the model as changes in the primary optimum. To determine the most supported number of shifts in the tree, the approach involves fixing the number of shifts and finding the best solution for that number, iterating this for various values of K (the number of shifts). An Expectation-Maximization (EM) algorithm[114] is used for this process, which involves ancestral trait reconstruction and determining shift positions and magnitudes. The true number of shifts is estimated using a penalized likelihood criterion to avoid overfitting. We used two maximum likelihood methods to determine the number of shifts, both aim to balance between simple and weak models to complex overfitting models: 'LINselect' evaluates how well each model predicts new data, instead of just how well it fits the data we have[115]. 'Djump' (Jump Heuristic) looks for a "jump" in the model's complexity that leads to a significant improvement in performance[116].

Estimation of directionality and phenotypic diversification in evolution

To explore the difference between species in the main sociality phenotypes (groups A-D Fig. 2), we implemented a method from movement ecology to evaluate the directionality of the evolutionary routes of species in the phenotypic space. In the original method, the assumption is that movement is a correlated random walk, and movement directionality can be estimated by computing and analyzing the turning angle distributions of movement in a landscape[117]. Here, we used the high-dimensional phenotypic space of PC1-PC4 as the landscape and computed the turning angle between vectors along the evolutionary route of each species based on our ancestral reconstruction (Fig. S8). When the angle equals π (180°), we interpret this as no change in the direction of evolution; when the absolute value of the turning angle is higher or lower than π, evolutionary directionality is lower. We compared the distributions of angles between three groups of trajectories: all species before the divergence of corbiculate bees,

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

corbiculate bees after divergence, and the remaining species after divergence. This comparison was to test whether putative trajectories that have crossed a major evolutionary transition display more directional evolution than trajectories that have not crossed it and show more flexible evolution. We performed an F-test to compare the variances between the groups and a t-test to compare the means.

We evaluated the phenotypic change in social complexity across time to estimate the rate of phenotypic diversification of extant lineages. To compute phenotypic diversification, we again used the phenotypic space of PC1-PC4 and the evolutionary trajectories within this space. For each branch in the phylogeny, we calculated the Euclidean distance between the positions of the nodes defining the branch in the four-dimensional phenotypic space (i.e., the change in PC values; ΔPC). The distance between points represents the amount of phenotypic change along that branch. The rate of phenotypic diversification per million years is calculated by dividing the temporal length of the branch, in mya, by the ΔPC:

$$rate\ of\ phenotypic\ diversification = \frac{phenotypic\ change\ (\Delta PC)}{time\ (branch\ length)} \quad (1)$$

Phylogenetic tree

We used species from across all the major social lineages of bees: Apidae (honeybees–Apini, stingless bees–Meliponini, Bumblebees–Bombini, orchid bees–Euglossin.), Xylocopinae (Allodapini, Xylocopa and Ceratina), and Halictidae (Sweat bees; including the tribe Augochlorini and the genus Lasioglossum and Halictus). Each of these lineages is commonly considered to have evolved high social complexity (eusociality) independently[12,69]. Time-calibrated phylogenetic tree to be used in our analyses was taken from 'beetreeoflife'[118] and included all species in our dataset except *scaptorigona postica*, which was replaced with its closest sister taxa *Scaptotrigona polysticta*. The phylogeny we used is based on the most comprehensive and up-to-date analysis of bee phylogenetics, constructed from a supermatrix of molecular data of more than 4500 species of bees across all families[118].

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Tomasello, M. The adaptive origins of uniquely human sociality. *Philosophical Transactions of the Royal Society B* **375**, 20190493 (2020).
2. Nelson, A. C., Colson, K. E., Harmon, S. & Potts, W. K. Rapid adaptation to mammalian sociality via sexually selected traits. *BMC Evol Biol* **13**, 1–14 (2013).
3. Silk, J. B. The adaptive value of sociality in mammalian groups. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**, 539–559 (2007).
4. Rubenstein, D. R. & Abbot, P. The evolution of social evolution. *Comparative social evolution* 1–18 (2017).
5. Wilson, E. O. *Sociobiology: The New Synthesis*. (Harvard University Press, 2000).
6. Shultz, S., Opie, C. & Atkinson, Q. D. Stepwise evolution of stable sociality in primates. *Nature* **479**, 219–222 (2011).
7. Kappeler, P. M. & Pozzi, L. Evolutionary transitions toward pair living in nonhuman primates as stepping stones toward more complex societies. *Sci Adv* **5**, eaay1276 (2019).
8. Olivier, C.A. *et al.* Primate social organization evolved from a flexible pair-living ancestor. *Proceedings of the National Academy of Sciences* ***121***, (2024).
9. Smuts, B. B., Cheney, D. L., Seyfarth, R. M., Wrangham, R. W. & Struhsaker, T. T. Primate Societies. Chicago: Univ. Preprint at (1987).
10. Downing, P. A., Griffin, A. S. & Cornwallis, C. K. Group formation and the evolutionary pathway to complex sociality in birds. *Nat Ecol Evol* **4**, 479–486 (2020).
11. Crook, J. H. Social behaviour in birds and mammals. (1970).
12. Michener, C. D. *The Social Behavior of the Bees: A Comparative Study*. (Harvard University Press, 1974).
13. Wheeler, W. M. *The Social Insects: Their Origin and Evolution*. (Routledge, 1915).
14. Batra, S. W. T. Behavior of some social and solitary halictine bees within their nests: a comparative study (Hymenoptera: Halictidae). *J Kans Entomol Soc* 120–133 (1968).
15. Wilson, E. O. *The insect societies.* (Harvard University Press, 1971).
16. Hamilton, W. D. The genetical evolution of social behaviour. II. *J Theor Biol* **7**, 17–52 (1964).
17. Smith, J. M. & Szathmary, E. *The Major Transitions in Evolution*. (Oxford University Press, 1997).
18. West-Eberhard, M. J. *Developmental Plasticity and Evolution*. (Oxford University Press, 2003).
19. Linksvayer, T. A. & Johnson, B. R. Re-thinking the social ladder approach for elucidating the evolution and molecular basis of insect societies. *Curr Opin Insect Sci* **34**, 123–129 (2019).
20. Holland, J. G. & Bloch, G. The complexity of social complexity: A quantitative multidimensional approach for studies of social organization. *Am Nat* **196**, 525–540 (2020).
21. Boomsma, J. J. & Gawne, R. Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. *Biological Reviews* **93**, 28–54 (2018).
22. Costa, J. T. & Fitzgerald, T. D. Developments in social terminology: semantic battles in a conceptual war. *Trends Ecol Evol* **11**, 285–289 (1996).
23. Schwarz, M. P., Richards, M. H. & Danforth, B. N. Changing paradigms in insect social evolution: insights from halictine and allodapine bees. *Annu Rev Entomol* **52**, 127–150 (2007).
24. Danforth, B. N., Cardinal, S., Praz, C., Almeida, E. A. B. & Michez, D. The impact of molecular data on our understanding of bee phylogeny and evolution. *Annu Rev Entomol* **58**, 57–78 (2013).
25. Kocher, S. D. & Paxton, R. J. Comparative methods offer powerful insights into social evolution in bees. *Apidologie* **45**, 289–305 (2014).
26. Wheeler, W. M. *Ants: Their Structure, Development and Behavior*. (Columbia University Press, 1910).
27. Wilson, E. O. & Hölldobler, B. Eusociality: origin and consequences. *Proceedings of the National Academy of Sciences* **102**, 13367–13371 (2005).
28. Rehan, S. M. & Toth, A. L. Climbing the social ladder: the molecular evolution of sociality. *Trends Ecol Evol* **30**, 426–433 (2015).
29. Taylor, D., Bentley, M. A. & Sumner, S. Social wasps as models to study the major evolutionary transition to superorganismality. *Curr Opin Insect Sci* **28**, 26–32 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

30. Evans, H. E. & Eberhard, M. J. W. *The wasps*. (1973).
31. Crespi, B. J. & Yanega, D. The definition of eusociality. *Behavioral Ecology* **6**, 109–115 (1995).
32. Gadagkar, R. Why the definition of eusociality is not helpful to understand its evolution and what should we do about it. *Oikos* 485–488 (1994).
33. Richards, M. H. Social trait definitions influence evolutionary inferences: a phylogenetic approach to improving social terminology for bees. *Curr Opin Insect Sci* **34**, 97–104 (2019).
34. Anderson, C. & McShea, D. W. Individual versus social complexity, with particular reference to ant colonies. *Biological reviews* **76**, 211–237 (2001).
35. Kocher, S. D. *et al.* The genetic basis of a social polymorphism in halictid bees. *Nat Commun* **9**, 1–7 (2018).
36. Shell, W. A. *et al.* Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees. *Commun Biol* **4**, 1–9 (2021).
37. Glastad, K. M. *et al.* Variation in DNA methylation is not consistently reflected by sociality in Hymenoptera. *Genome Biol Evol* **9**, 1687–1698 (2017).
38. Berens, A. J., Hunt, J. H. & Toth, A. L. Comparative transcriptomics of convergent evolution: different genes but conserved pathways underlie caste phenotypes across lineages of eusocial insects. *Mol Biol Evol* **32**, 690–703 (2015).
39. Toth, A. L., Sumner, S. & Jeanne, R. L. Patterns of longevity across a sociality gradient in vespid wasps. *Curr Opin Insect Sci* **16**, 28–35 (2016).
40. Kapheim, K. M. *et al.* Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143 (2015).
41. Lande, R. & Arnold, S. J. The measurement of selection on correlated characters. *Evolution (N Y)* 1210–1226 (1983).
42. Felsenstein, J. Phylogenies and quantitative characters. *Annu Rev Ecol Syst* 445–471 (1988).
43. Armbruster, W. S. & Schwaegerle, K. E. Causes of covariation of phenotypic traits among populations. *J Evol Biol* **9**, 261–276 (1996).
44. Armbruster, W. S., Pélabon, C., Bolstad, G. H. & Hansen, T. F. Integrated phenotypes: understanding trait covariation in. (2014).
45. Walsh, B. Escape from flatland. *J Evol Biol* **20**, 36–38 (2007).
46. Walsh, B. & Blows, M. W. Abundant genetic variation+strong selection=multivariate genetic constraints: a geometric view of adaptation. *Annu Rev Ecol Evol Syst* **40**, 41 (2009).
47. Blows, M. W. A tale of two matrices: multivariate approaches in evolutionary biology. *J Evol Biol* **20**, 1–8 (2007).
48. Johnson, R. A. & Wichern, D. W. Applied multivariate statistical analysis. (2002).
49. Wismüller, A., Dsouza, A. M., Vosoughi, M. A. & Abidin, A. Large-scale nonlinear Granger causality for inferring directed dependence from short multivariate time-series data. *Sci Rep* **11**, 7817 (2021).
50. Yoo, K. *et al.* Multivariate approaches improve the reliability and validity of functional connectivity and prediction of individual behaviors. *Neuroimage* **197**, 212–223 (2019).
51. Coombes, K. R., Wang, M. & Coombes, M. K. R. Package 'PCDimension'. (2019).
52. Patel, A. *et al.* The making of a queen: TOR pathway is a key player in diphenic caste development. *PLoS One* **2**, e509 (2007).
53. Lu, H. & Pietrantonio, P. V. Insect insulin receptors: insights from sequence and caste expression analyses of two cloned hymenopteran insulin receptor cDNAs from the fire ant. *Insect Mol Biol* **20**, 637–649 (2011).
54. Manfredini, F. *et al.* Sociogenomics of cooperation and conflict during colony founding in the fire ant Solenopsis invicta. *PLoS Genet* **9**, e1003633 (2013).
55. Withee, J. R. & Rehan, S. M. Social aggression, experience, and brain gene expression in a subsocial bee. *Integr Comp Biol* **57**, 640–648 (2017).
56. Engel, M. S. Monophyly and extensive extinction of advanced eusocial bees: insights from an unexpected Eocene diversity. *Proceedings of the National Academy of Sciences* **98**, 1661–1664 (2001).
57. Barden, P. & Engel, M. S. Fossil social insects. *Encyclopedia of social insects* 384–403 (2021).
58. Simpson, G. G. *Tempo and Mode in Evolution*. (Columbia University Press, 1984).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

59. Hansen, T. F., Svensson, E. & Calsbeek, R. Adaptive landscapes and macroevolutionary dynamics. *The adaptive landscape in evolutionary biology* **205**, 26 (2012).
60. Johnson, B. R. & Linksvayer, T. A. Deconstructing the superorganism: social physiology, groundplans, and sociogenomics. *Q Rev Biol* **85**, 57–79 (2010).
61. Bourke, A. F. G. Colony size, social complexity and reproductive conflict in social insects. *J Evol Biol* **12**, 245–257 (1999).
62. Bourke, A. F. G. Principles of social evolution. (2011).
63. Crespi, B. J. Vicious circles: positive feedback in major evolutionary and ecological transitions. *Trends Ecol Evol* **19**, 627–633 (2004).
64. Miller, J. S. & Reeve, H. K. Feedback loops in the major evolutionary transition to eusociality: the status and potential of theoretical approaches. *Curr Opin Insect Sci* **34**, 85–90 (2019).
65. Choi, S.-W. *et al.* Ordovician origin and subsequent diversification of the brown algae. *Current Biology* (2024).
66. Ispolatov, I., Ackermann, M. & Doebeli, M. Division of labour and the evolution of multicellularity. *Proceedings of the Royal Society B: Biological Sciences* **279**, 1768–1776 (2012).
67. Reznick, D. N. & Ricklefs, R. E. Darwin's bridge between microevolution and macroevolution. *Nature* **457**, 837–842 (2009).
68. Cooney, C. R. *et al.* Mega-evolutionary dynamics of the adaptive radiation of birds. *Nature* **542**, 344–347 (2017).
69. Cardinal, S. & Danforth, B. N. The antiquity and evolutionary history of social behavior in bees. *PLoS One* **6**, e21086 (2011).
70. da Silva, J. Life History and the Transitions to Eusociality in the Hymenoptera. *Front Ecol Evol* **9**, (2021).
71. Crespi, B. J. Comparative analysis of the origins and losses of eusociality: causal mosaics and historical uniqueness. *Phylogenies and the comparative method in animal behavior* 253–287 (1996).
72. Danforth, B. N., Conway, L. & Ji, S. Phylogeny of eusocial Lasioglossum reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae). *Syst Biol* **52**, 23–36 (2003).
73. Jones, B. M. *et al.* Convergent and complementary selection shaped gains and losses of eusociality in sweat bees. *Nat Ecol Evol* **7**, 557–569 (2023).
74. Gibbs, J., Brady, S. G., Kanda, K. & Danforth, B. N. Phylogeny of halictine bees supports a shared origin of eusociality for Halictus and Lasioglossum (Apoidea: Anthophila: Halictidae). *Mol Phylogenet Evol* **65**, 926–939 (2012).
75. Davison, P. J. & Field, J. Season length, body size, and social polymorphism: size clines but not saw tooth clines in sweat bees. *Ecol Entomol* **42**, 768–776 (2017).
76. Rehan, S. M. & Richards, M. H. The Influence of Maternal Quality on Brood Sex Allocation in the Small Carpenter Bee, *Ceratina calcarata*. *Ethology* **116**, 876–887 (2010).
77. Soucy, S., Sociaux, T. G.-I. & 2003, undefined. Solitary and group nesting in the orchid *bee Euglossa hyacinthina* (Hymenoptera, Apidae). *Springer* **50**, 248–255 (2003).
78. Soucy, S. L. & Danforth, B. N. Phylogeography of the socially polymorphic sweat bee *Halictus rubicundus* (Hymenoptera: Halictidae). *Evolution* **56**, 330–341 (2002).
79. Hurst, P. S. Social biology of *Exoneurella tridentata*, an allodapine bee with morphological castes and perennial colonies. *Unpublished D. Phil. Thesis, Flinders University* (2001).
80. Plateaux-Quénu, C. Biology of *Halictus marginatus Brullé*. *J Apic Res* **1**, 41–51 (1962).
81. Dew, R. M., Rehan, S. M., Tierney, S. M., Chenoweth, L. B. & Schwarz, M. P. A single origin of large colony size in allodapine bees suggests a threshold event among 50 million years of evolutionary tinkering. *Insectes Soc* **59**, 207–214 (2012).
82. Schwarz, M. P., Bull, N. J. & Hogendoorn, K. Evolution of sociality in the allodapine bees: a review of sex allocation, ecology and evolution. *Insectes Soc* **45**, 349–368 (1998).
83. Kappeler, P. M. A framework for studying social complexity. *Behav Ecol Sociobiol* **73**, 1–14 (2019).
84. Toth, A. L. & Robinson, G. E. Evo-devo and the evolution of social behavior. *Trends in Genetics* **23**, 334–341 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

85. Evans, H. E. The evolution of social life in wasps. *Proceedings of the 10th International Congress of Entomology* (1958).
86. Dew, R. M., Shell, W. A. & Rehan, S. M. Changes in maternal investment with climate moderate social behaviour in a facultatively social bee. *Behav Ecol Sociobiol* **72**, 1–12 (2018).
87. Team, R. C. R: A language and environment for statistical computing. (2013).
88. Rehan, S. M., Leys, R. & Schwarz, M. P. A mid-cretaceous origin of sociality in xylocopine bees with only two origins of true worker castes indicates severe barriers to eusociality. *PLoS One* **7**, e34690 (2012).
89. Dray, S. & Josse, J. Principal component analysis with missing values: a comparative survey of methods. *Plant Ecol* **216**, 657–667 (2015).
90. Josse, J. & Husson, F. missMDA: a package for handling missing values in multivariate data analysis. *J Stat Softw* **70**, 1–31 (2016).
91. Pagel, M. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
92. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol* 217–223 (2012).
93. Blomberg, S. P., Garland Jr, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
94. Münkemüller, T. *et al.* How to measure and test phylogenetic signal. *Methods Ecol Evol* **3**, 743–756 (2012).
95. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
96. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289–300 (1995).
97. McInnes, L., Healy, J. & Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv:1802.03426* (2018).
98. Lê, S., Josse, J. & Husson, F. FactoMineR: an R package for multivariate analysis. *J Stat Softw* **25**, 1–18 (2008).
99. Konopka, T. & Konopka, M. T. R-package: umap. *Uniform Manifold Approximation and Projection* (2018).
100. Chari, T. & Pachter, L. The specious art of single-cell genomics. *PLoS Comput Biol* **19**, e1011288 (2023).
101. Diaz-Papkovich, A., Anderson-Trocmé, L. & Gravel, S. A review of UMAP in population genetics. *J Hum Genet* **66**, 85–91 (2021).
102. Slater, G. J., Harmon, L. J. & Alfaro, M. E. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution* **66**, 3931–3944 (2012).
103. Harmon, L. J. *et al.* Early bursts of body size and shape evolution are rare in comparative data. *Evolution* **64**, 2385–2396 (2010).
104. Mahler, D. L., Ingram, T., Revell, L. J. & Losos, J. B. Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science* **341**, 292–295 (2013).
105. O'meara, B. C., Ané, C., Sanderson, M. J. & Wainwright, P. C. Testing for different rates of continuous trait evolution using likelihood. *Evolution* **60**, 922–933 (2006).
106. Mahler, D. L., Revell, L. J., Glor, R. E. & Losos, J. B. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. *Evolution* **64**, 2731–2745 (2010).
107. Adams, D. C. Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. *Syst Biol* **62**, 181–192 (2013).
108. Hansen, T. F. & Martins, E. P. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* **50**, 1404–1417 (1996).
109. Ingram, T. & Mahler, D. L. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods Ecol Evol* **4**, 416–425 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2024.02.09.579609; this version posted February 9, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

110. Uyeda, J. C. & Harmon, L. J. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Syst Biol* **63**, 902–918 (2014).
111. Khabbazian, M., Kriebel, R., Rohe, K. & Ané, C. Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. *Methods Ecol Evol* **7**, 811–824 (2016).
112. Bastide, P., Mariadassou, M. & Robin, S. Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. *J R Stat Soc Series B Stat Methodol* **79**, 1067–1093 (2017).
113. Bastide, P., Ané, C., Robin, S. & Mariadassou, M. Inference of adaptive shifts for multivariate correlated traits. *Syst Biol* **67**, 662–680 (2018).
114. Moon, T. K. The expectation-maximization algorithm. *IEEE Signal Process Mag* **13**, 47–60 (1996).
115. Baraud, Y., Giraud, C. & Huet, S. Estimator selection in the Gaussian setting. in *Annales de l'IHP Probabilités et statistiques* vol. 50 1092–1119 (2014).
116. Birgé, L. & Massart, P. Minimal penalties for Gaussian model selection. *Probab Theory Relat Fields* **138**, 33–73 (2007).
117. Kareiva, P. M. & Shigesada, N. Analyzing insect movement as a correlated random walk. *Oecologia* **56**, 234–238 (1983).
118. Henríquez-Piskulich, P., Hugall, A. F. & Stuart-Fox, D. A supermatrix phylogeny of the world's bees (Hymenoptera: Anthophila). *Mol Phylogenet Evol* **190**, 107963 (2024).