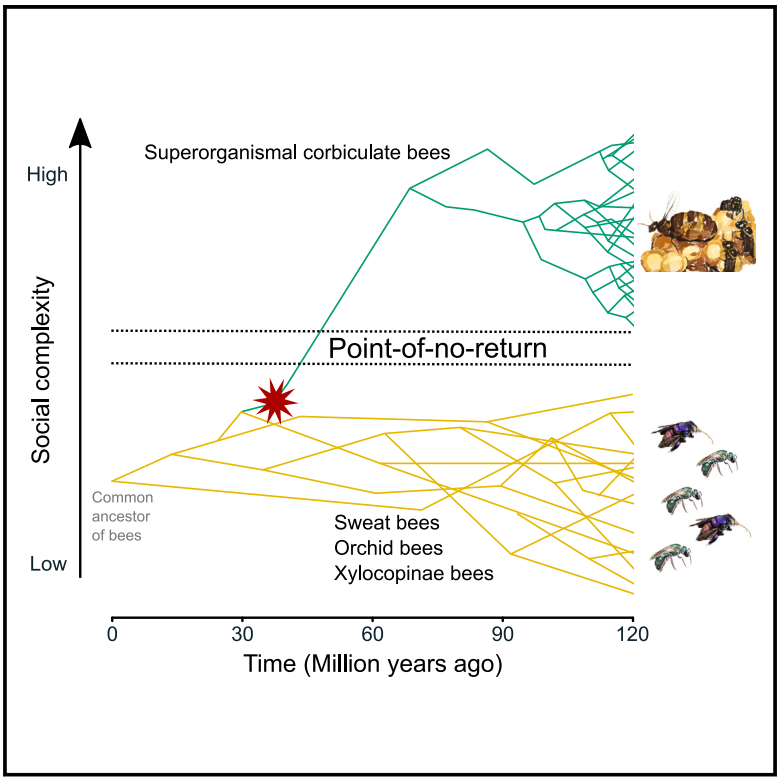
Article

Current Biology

Diversification of social complexity following a major evolutionary transition in bees

Graphical abstract



Authors

Ohad Peled, Gili Greenbaum, Guy Bloch

Correspondence

gil.g@mail.huji.ac.il (G.G.), guy.bloch@mail.huji.ac.il (G.B.)

In brief

Peled et al. develop a data-driven approach to study social phenotypes and their evolution in bees. They analyze 17 traits for 80 species and find that only the corbiculate bees overcame biological barriers to achieve highly complex social organization. This evolutionary transition led to remarkable diversifications of social phenotypes.

Highlights

- We developed a high-dimensional data-driven approach for studying social complexity
- Results support a major evolutionary transition only in corbiculate bees
- The major transition was followed by diversification of the social phenotype
- This data-driven approach can be applied to additional animal lineages

Peled et al., 2025, Current Biology *35*, 1–13
March 10, 2025 © 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.
https://doi.org/10.1016/j.cub.2025.01.009

CellPress

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

Current Biology

CellPress

Article

Diversification of social complexity following a major evolutionary transition in bees

Ohad Peled,[1] Gili Greenbaum,[1,2,*] and Guy Bloch[1,2,3,*]
[1]Department of Ecology, Evolution, and Behavior, The Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
[2]These authors contributed equally
[3]Lead contact
*Correspondence: gil.g@mail.huji.ac.il (G.G.), guy.bloch@mail.huji.ac.il (G.B.)
https://doi.org/10.1016/j.cub.2025.01.009

SUMMARY

How social complexity evolved remains a long-standing enigma. In most animal groups, social complexity is typically classified into a few discrete classes. This approach is oversimplified and constrains our inference of social evolution to a narrow trajectory consisting of transitions between classes. Such categorical classifications also limit quantitative studies on the molecular and environmental drivers of social complexity. The recent accumulation of relevant quantitative data has set the stage to overcome these limitations. Here, we propose a data-driven, high-dimensional approach for studying the full diversity of social phenotypes. We curated and analyzed a comprehensive dataset encompassing 17 social traits across 80 species and studied the evolution of social complexity in bees. We found that honey bees, stingless bees, and bumble bees underwent a major evolutionary transition ~80 mya, inconsistent with the stepwise progression of the social ladder conceptual framework. This major evolutionary transition was followed by a phase of substantial phenotypic diversification of social complexity. Other bee lineages display a continuum of social complexity, ranging from solitary to simple societies, but do not reach the levels of social complexity seen in honey bees, stingless bees, and bumble bees. Bee evolution, therefore, provides a remarkable demonstration of a macroevolutionary process in which a major transition removed biological constraints and opened novel evolutionary opportunities, driving the exploration of the landscape of social phenotypes. Our approach can be extended to incorporate additional data types and readily applied to illuminate the evolution of social complexity in other animal groups.

INTRODUCTION

One of the most fascinating examples of increased complexity in biological systems is the evolution of animal societies. Sociality has evolved multiple times across independent taxonomic groups and has played a key role in the evolution of animals.[1–4] Despite the remarkable variation among animal societies, comparative studies have typically classified levels of social complexity based on a few qualitative traits, such as group composition, reproductive skew, or parental care (e.g., in primates,[5–8] in birds,[9,10] and in insects[11–13]). This approach was crucial for the development of influential theories such as kin selection,[14] major transitions in evolution,[15] and developmental plasticity.[16] 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, and eusocial) assumed to represent a set of evolutionary transitions.[17–21] This limited set of transitions imposes a narrow trajectory for the evolution of social complexity.[17] Such qualitative classifications fail to capture the full diversity of social phenotypes within these classes. For example, most ants and termites are typically classified as "advanced eusocial" or "superorganismal," despite including species with small colonies, limited queen-worker dimorphism, and sexually reproducing workers ("gamergates").[13,22] The commonly used qualitative approach may mask 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 within the same taxonomic lineage often exhibit diverse social phenotypes.[11,23,24] Theories on the evolution of social complexity in insects can be divided into two main frameworks. The first framework emphasizes a single pivotal increase in the level of social complexity that leads to an irreversible point of no return, known as "superorganismality."[3,4,15,19,25] This framework focuses on caste differentiation processes producing highly fertile castes (queens and kings), which function 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 "the social ladder," assumes a stepwise evolutionary trajectory passing through a few key levels of sociality.[11,13,26] The less complex societies are viewed as primitive states, metaphorically comparable to lower rungs on a ladder.[26–28] Despite the long tradition and significant contribution of the social ladder framework, it remains debated whether sociality in insects evolved along

All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

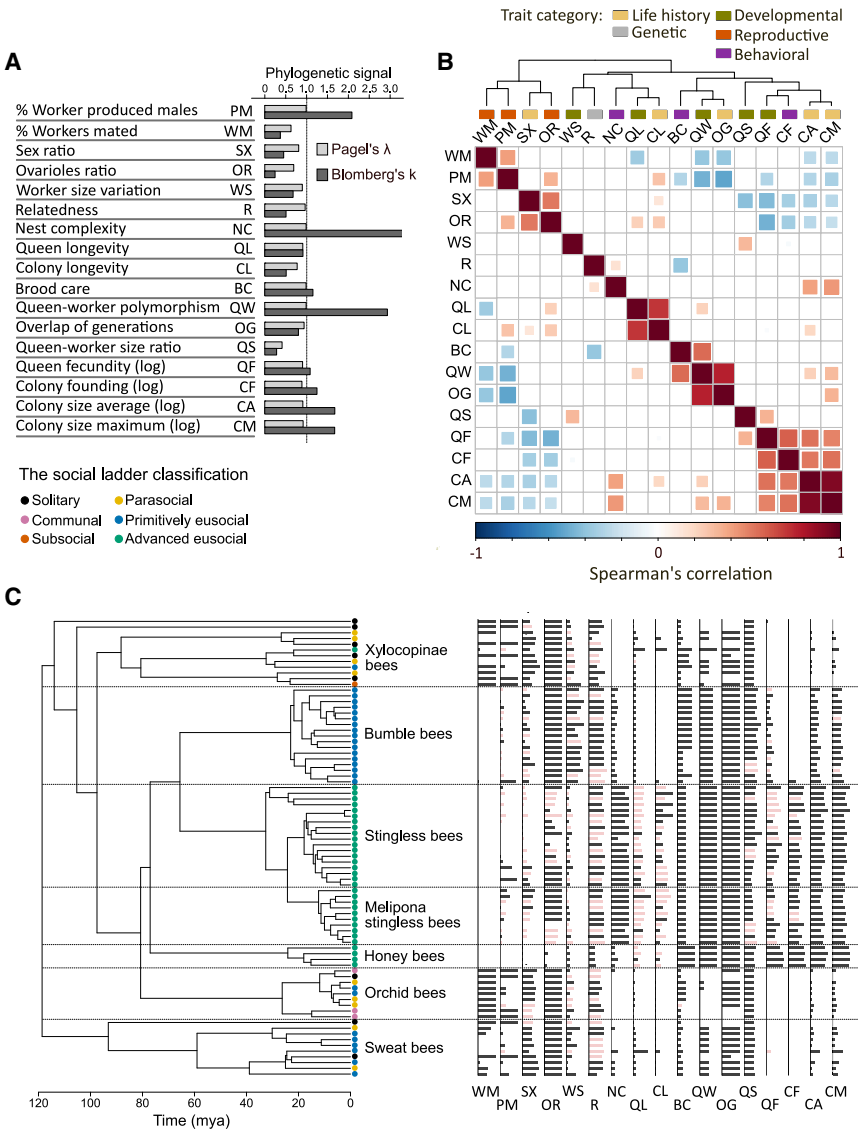


Figure 1. An overview of our dataset consisting of 17 traits and 80 species of bees

(A) The list of social complexity-related traits in our dataset and the phylogenetic signals for each trait. Pagel's λ values are in light gray, and Blomberg's K is in dark gray. The dashed horizontal line delineates a neutral evolutionary process of Brownian motion. All traits showed a significant phylogenetic signal ($p < 0.001$ for both measures after false discovery rate [FDR] correction) when compared with a null model in which traits evolve independently of phylogeny.

(B) 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 coefficient, with positive correlations in red and negative in blue. Only statistically significant correlations are shown ($p < 0.05$ 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.

(C) Time-calibrated phylogenetic tree of the species in our dataset. Colors correspond to the common social ladder classification of Michener.[11] The barplot on the right corresponds to the values of traits for each species (black bars are data extracted from literature, and pink bars are imputed data). See Figure S1 for species names and Table S1 for trait information.

similar evolutionary trajectories across different taxonomic lineages.[17,18,20,29–31]

Both theories have suffered from repeated semantic disagreements regarding their definitions,[18,20,21,31] a discussion that hampered the use of comparative methods to identify factors associated with variation in social complexity.[17,18,24] Indeed, recent attempts to understand the molecular underpinnings of the evolution of social complexity have had partial success.[26,32–36] This lack of substantial progress might be attributed to the limited inference that can be made when using a small number of traits as proxies for social complexity,[17,18] given that traits contributing to complex phenotypes may be shaped by various selection pressures, ecological niches, preadaptations, pleiotropic effects, and developmental constraints.[37–40] Previous attempts to apply quantitative indices for social complexity focused on only a few traits and were restricted by both the number of species included and the range of taxonomic sampling,[18,41–45] limiting their capacity to identify and characterize macroevolutionary processes.

The increased availability of life history, behavioral, morphological, molecular, and genomic data for related species differing in social phenotypes, coupled with well-resolved phylogenetics and appropriate computational methods, sets the stage for quantitative investigations of social complexity phenotypes.[46–51] Here, we adopt a data-driven approach to untangle the evolutionary history of social complexity in bees, which does not assume that there are specific social classes or predetermined evolutionary trajectories. Such a model-free approach enables the analysis of complex interactions among traits in a high-dimensional space,[52,53] and can potentially resolve controversies arising from a qualitative representation of social complexity.

RESULTS

We generated a quantitative dataset consisting of 17 traits related to social complexity for 80 bee species exhibiting diverse levels of social complexity (Figure 1; see Figure S1 for species names). Through an extensive literature survey, we identified reliable quantitative social traits that are comparable across species (Figure 1A; Table S1). Our dataset includes colony-level traits (e.g., colony size, colony longevity [CL], colony reproductive skew, and level of brood care [BC]), life-history traits (e.g., overlap of generations [OG] and colony founding [CF]), and morphological traits (e.g., queen-worker [QS] and

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

worker-worker [WS] 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 80 bee species in our dataset were selected based on the availability and quality of published information, aiming to accurately represent the broad spectrum of social complexity and phylogeny (Figure 1C). Some clades are characterized by a single level of social complexity (e.g., honey bees, stingless bees, and bumble bees), while other clades encompass multiple levels of social complexity (e.g., sweat bees, orchid bees, and Xylocopinae bees; Figure 1C). To allow the full potential of quantitative analysis for our dataset, missing data were imputed while accounting for correlations between traits and between species (see STAR Methods).

Phylogeny significantly accounted for variation in social complexity. There is a statistically significant phylogenetic signal ($p < 0.001$) for all traits using two complementary measures: Pagel's lambda, which quantifies the phylogenetic signal from 0 to 1, and Blomberg's K, which allows values >1 to indicate higher-than-expected phylogenetic similarity between species (Figure 1A). Both approaches measure the extent to which trait evolution follows a Brownian motion model of neutral evolution (see STAR Methods). The three traits with the lowest phylogenetic signal were ovarioles ratio (OR), percentage of workers mated (WM), and queen-worker size ratio (QS). For seven traits, Blomberg's K values were >1: colony size-related traits (CA, CM, and CF), queen fecundity (QF), queen-worker polymorphism (QW), the percentage of worker-produced males (PM), and nest complexity (NC), which is consistent with higher-than-expected similarity based on phylogenetic relatedness alone. NC showed the highest Blomberg's K value, probably due to its high score in stingless 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 honey bees, stingless bees, and bumble bees, which are represented by a relatively large number of species in our dataset (Figure 1C).

Traits can be influenced by common evolutionary history, and we therefore computed the correlations between all pairs of traits while accounting for shared ancestry (Spearman's correlation coefficients using phylogenetically independent contrasts (PICs) method to account for phylogenetic relationships). Many traits in our dataset, including colony size-related traits, QF, and NC (NC), are highly correlated (Figure 1B). This group of traits is also negatively correlated with sex ratio (SX) and traits reflecting worker reproduction. Interestingly, the defining characteristics of eusociality[13]— BC, OG, and QW—were clustered together in the hierarchical clustering analysis and were correlated with each other, but not with many of the other social traits in our dataset. Likewise, Colony and queen longevity (CL and QL) are correlated with each other but not with other traits. Relatedness (R) between females, a putative driver for social evolution, was not correlated with most traits and shows a low ($r = -0.38$) but statistically significant negative correlation with BC. This relationship is perhaps indirect, driven by honey bees, which have the lowest relatedness values but show a high degree of BC. Interestingly, the hierarchical clustering analysis does not group traits based on the broad trait categories we defined (colors in Figure 1B), reflecting the complex correlation structure of the data. Overall, the correlation matrix suggests that not all social traits are necessarily positively correlated and that multiple combinations of traits may contribute differently to an increase in social complexity. Nonetheless, the interpretation of specific correlations between traits based on this analysis should be made with caution, and we therefore focus our analysis on broad patterns that are not trait-specific.

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 employed both principal-component analysis (PCA) and uniform manifold approximation and projection (UMAP). These analyses are complementary, and 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 (Figure 1A), we applied a phylogenetic correction for all downstream analyses. This correction accounts for the shared evolutionary history among species, aiming to minimize the bias generated by the phylogenetic relatedness among species. The clustering in UMAP was consistent over a range of parameters of the analysis (Figure S2A), indicating that the overall structure of the species clustering is robust regarding the relations between species. We used the broken stick method[54] to find the number of informative principal components (PCs) in our PCA (the method determines how many PCs are informative by comparing the variance explained by each PC to a threshold of random variance, and only the PCs with variance exceeding the threshold are considered informative). We find four informative PCs, which together explain ~70% of the variation in the social phenotype space, with PC1 and PC2 accounting for 28% and 15% of the variation, respectively (Figures 2A, S2B, and S2C). Both the PCA and UMAP identify clusters of species occupying clear areas in the phenotypic space, but the composition and structure of the clusters differed to some extent across analyses (Figure 2A).

In the UMAP (Figures 2B and S2A), clustering suggested a clear separation between (1) honey bees and stingless bees, (2) bumble bees, and (3) the remaining species. In the PCA plots (Figures 2A and S2C), bumble bees and stingless bees overlapped in PC1 and PC2 but not in PC3 and PC4, whereas honey bees were clustered separately. In addition, solitary, communal, and subsocial species were clustered separately from parasocial (often termed facultatively social[55,56]) and non-*Bombus* primitively eusocial species in PC1. Considering both PCA and UMAP, we identified four main phenotypic groups that occupy well-defined areas in the phenotypic space representing different social complexity phenotypes (hereafter termed "groups"; gray shaded areas labeled groups A–D in Figures 2A and 2B). The main groups correspond to (A) solitary, communal, and subsocial species; (B) parasocial and non-*Bombus* primitively eusocial species; (C) bumble bees; and (D) honey bees and stingless bees. The clearest and widest separation in both the UMAP and PCA is between the monophyletic group within corbiculate bees—honey bees, stingless bees, and bumble bees (groups C and D)—and the rest of the species (groups A

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

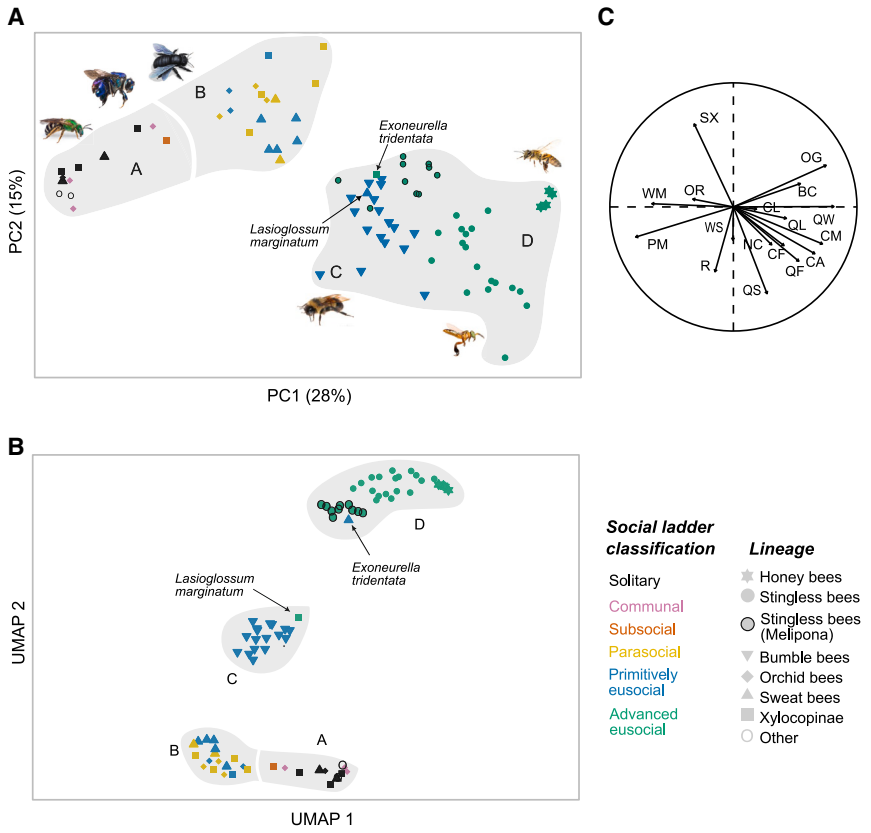


Figure 2. The phenotypic space of social complexity in bees

(A) Phylogenetically corrected PCA. The plot shows PC1 and PC2, which together account for 43% of the variation among species in our dataset. Symbol color corresponds to the commonly used social ladder 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 dataset and were therefore marked as "other."

(B) Phylogenetically corrected 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 Figure 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. See Figures S2 and S3 for additional analyses.

and B; Figures 2A and 2B). Thus, even after phylogenetic correction, social complexity is still associated with phylogeny, specifically for honey bees, stingless bees, and bumble bees. Additionally, the partition between C and D to A and B reflects the separation of species with obligate sociality (C and D) from species with mostly facultative or no social lifestyle (A and B). Notably, facultative and obligate sociality were 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 Figures 2 and S2). Bees with different classifications are clustered together within our phenotypic space. The change from solitary lifestyle (group A) to parasocial and non-*Bombus* primitively eusocial species (group B) was gradual rather than stepwise as predicted by social ladder models. Another apparent disagreement with social ladder models is that species that are classified as primitively eusocial do not occupy a defined phenotypic area but rather are separated into two different clusters: one that includes bumble bees and one that includes the other primitively eusocial species as well as some species that are typically classified as parasocial or facultatively social; however, bumble bees do overlap with species commonly classified as having lower levels of sociality in PC3 and PC4. *Exoneurella tridentata* (Allodapini, typically classified as advanced eusocial) and *Lasioglossum marginatum* (Halictidae; typically classified as primitively eusocial) were the only species that do not belong to the honey bees, stingless bees, and bumble bees, that in some analyses clustered within groups C or D. It should be noted, however, that these two species were associated with different clusters across analyses (Figures S2A, S2C, S2D, and S2F).

The species clusters in the PCA are not tight, with each occupying a large area in the phenotypic space. Notably, the within-group variation is as large as the between-group variation (i.e., the size of each group in the PCA is as large as the distances between the groups), highlighting substantial variation in social complexity that is entirely overlooked in the commonly used qualitative classifications. An important example is the large phenotypic space occupied by advanced eusocial species and bumble bees, suggesting that species typically assigned to this class vary substantially in social traits (groups C and D in Figures 2 and S2C). To compare areas within the phenotypic spaces, we calculated the convex hulls of the regions occupied by groups C + D (honey bees, stingless bees, and bumble bees) and A + B (the remaining species), using all four PCs. Our analysis revealed that the area occupied by groups C + D is 6 times larger than that of groups A + B. PC3 and PC4 accounted for much of this trend (Figure 2G), which can be attributed to the large variation in OR and queen and colony longevity (QL and CL, respectively), which is most evident in stingless bees. Group D can be further partitioned into three subgroups: honey bees, *Melipona* genus stingless bees, and non-*Melipona* stingless bees (Figures 2A and 2B). The bumble bee group was consistently partitioned along the common distinction between mass provisioners (pocket-making) and progressive provisioners (pollen-storers; shown in UMAP analysis Figure S2A). We evaluated the robustness of our phenotypic space by running a set of sensitivity analyses, in which we tested how manipulations of the data structure affect the results. Taken together, these analyses suggest that there is no substantial bias in the phenotypic space of the PCA due to taxonomy representation, missing data, phylogenetic tree construction, and inclusion of specific traits (Figure S3).

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

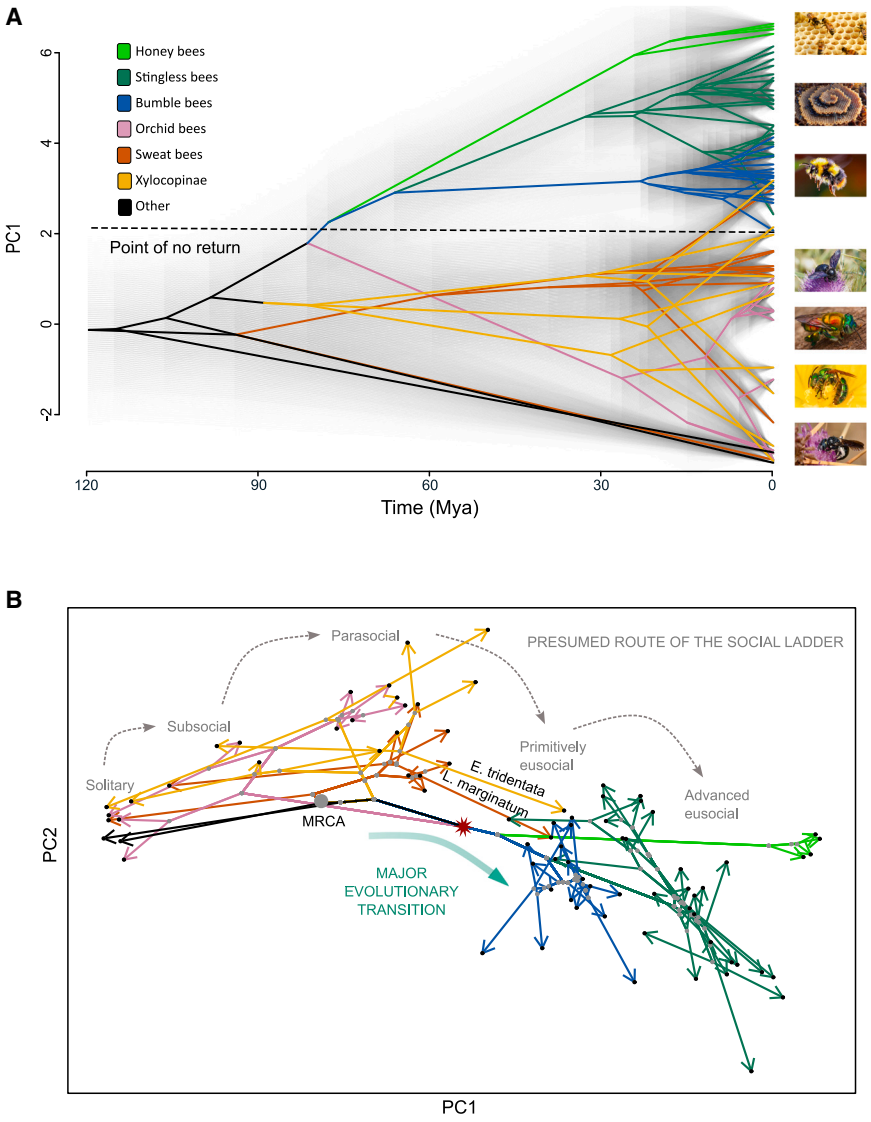


Figure 3. Ancestral state reconstruction for the evolution of social complexity

Values were obtained from the PCA (Figure 2A).

(A) A phenogram of the evolutionary trajectory of PC1 over time. Shaded gray area represents the 95% confidence interval. Colors represent the main taxonomic lineages in our dataset, broadly aligning with the color scheme of Figure 2. A dashed line indicates a putative point of no return beyond which species are not likely to revert to lower levels of social complexity.

(B) Integration of the PCA phenotypic space 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 mark the position of extant species in our dataset. 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 (MRCA) of all species. The red star marks the putative common ancestor of superorganismal corbiculate bees. Gray dashed arrows illustrate a stepwise increase in social complexity as predicted by social ladder models. The thick green arrow illustrates the major transition in social complexity. Photo credits from top to bottom: pickpik, Tim Heard, pickpik, Adobe Stock, Gil Wizen, Judy Gallagher, Gideon Pisanty. See Figures S4 and S5 for analyses on all four PCs.

The largest separation in the phenotypic space of the PCA is between clusters A + B and C + D. Assessment of the contribution of each trait suggests that colony size-related traits (CA, CM, CF, and QF), QS, QW, WM, and SX, or some combinations of these traits, likely contribute to this separation (Figures 2C and S2B). Honey bees and non-*Melipona* stingless bees have a high drone-to-gyne ratio, which may contribute to the significant effect of SX on PC2. Species in groups A + B (i.e., sweat bees, orchid bees, and Xylocopinae bees) align along an apparent diagonal line in the PCA of PC1 and PC2, an axis correlated with BC, PM, and OG (Figure 2C), but other traits may interact to further contribute to this pattern. Interestingly, these three traits that are primarily associated with the variability observed in species with the simpler societies are consistent with the definition of eusociality according to Wilson.[13]

The evolutionary history of social complexity in bees

We conducted ancestral state reconstruction using PC values as a quantitative estimate for the social complexity phenotype to gain insight into the evolutionary history of social complexity (Figures 3A and S4). 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 (Figures 3B and S5). The reconstructed evolutionary trajectory leading to the monophyletic honey bees, stingless bees, and bumble bees showed a relatively constant trend of increase in social complexity values along PC1 over time (Figure 3A). By contrast, the reconstruction of species from groups A + B showed no consistent directional changes with increases and decreases along the 4 major PCs, and particularly in PC1. We identified a putative point of no return that is consistent with an irreversible phenotypic threshold in PC1 and was crossed by all honey bees, stingless bees, and bumble bees (dashed line in Figure 3A). Hence, we refer to this group as the "superorganismal corbiculate bees." The phenotypic threshold was also crossed by *E. tridentata* and *L. marginatum*. The evolutionary reconstruction of PC2 (Figure S4), which mainly reflects the traits QW, QS, SX, and R (Figure S2B), suggests an early separation of the superorganismal corbiculate bees from most other lineages, but the overlap with species from other lineages is larger than for PC1. Taken together, these analyses suggest that both PC1 and PC2 contributed to the increase in social complexity that separates the superorganismal corbiculate bees from the other lineages (Figure 3B).

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

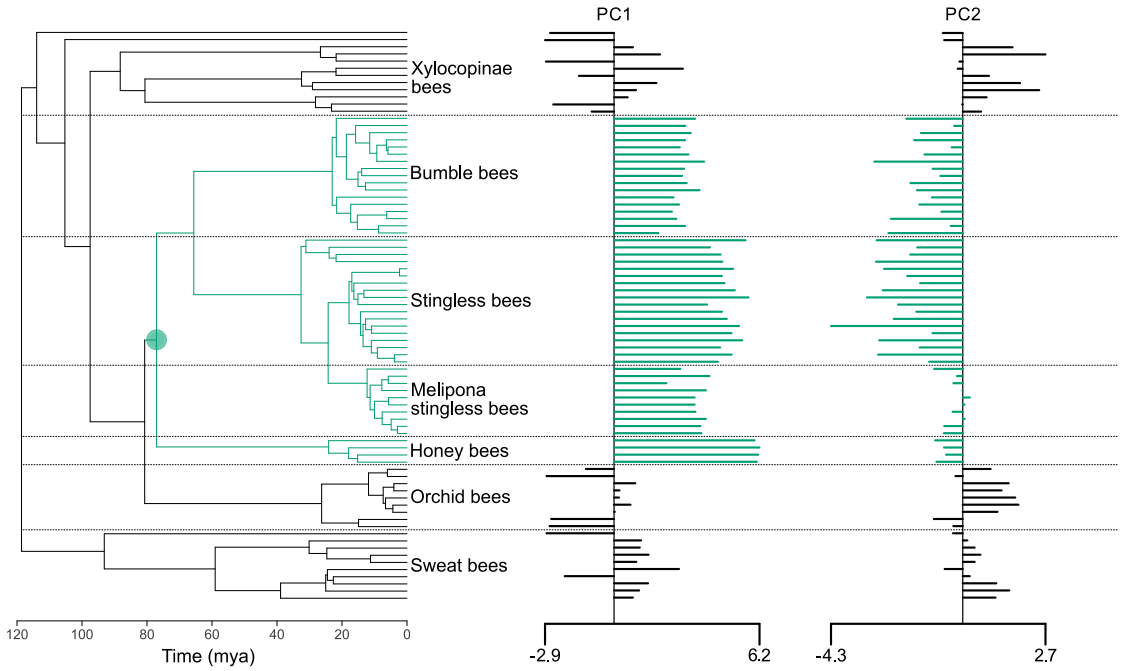


Figure 4. Adaptive evolutionary shift of social complexity

On the left, a single shift detected on the phylogenetic tree. The green circle indicates the inferred position of adaptive regime shift from PC1 and PC2, marking the ancestor of the superorganismal corbiculate bees. On the right, bars show the PC1 and PC2 values for each species. See additional results for shift-detection analysis in Figure S6.

After the split of the superorganismal corbiculate bees from the remaining lineages, the social complexity of bumble bees, stingless bees, and honey bees diversified within the phenotypic space of PC1 and PC2. For example, the *Melipona* stingless bees diverged from the rest of the stingless bees to occupy a distinct phenotypic space (Figure 3B). The superorganismal corbiculate bees show the largest phenotypic expansion during the reconstruction of PC3 and PC4, with PC3 (influenced by OR, R, CF, and QF) also separating honey bees from all other lineages (Figures S4 and S5). Our evolutionary reconstructions suggest that the high social complexity of *E. tridentata* and *L. marginatum* reflects a later increase in social complexity through different evolutionary routes than that leading to high social complexity in the superorganismal corbiculate bees (Figure S4). Due to the increase in uncertainty as we approach the root of the phylogeny, we refrain from interpreting information about the inferred social phenotype of the ancestor of bees and, more specifically, about the ancestor of the superorganismal corbiculate bees. Yet, the overall results remain consistent concerning the unique evolutionary route separating the superorganismal corbiculate bees from the rest of the species.

A shift-detection analysis using the values of PC1 and PC2 revealed a single major evolutionary shift, which occurred in the ancestor of the superorganismal corbiculate bees (green dot in Figure 4), and the same result was obtained by analyzing the specific social traits instead of PCs (see Figure 1C for the distribution of traits among species). Models assuming a larger number of shifts, for example, as may be predicted by social ladder models, received less support (Figure S6A). We did find additional phenotypic shifts when including all four main PCs, but these were in the specific branches leading to honey bees, stingless bees, *Melipona* genus, and bumble bees (Figure S6B). This suggests that PC3 and PC4 offer a more detailed description of phenotypic changes in these taxonomic lineages. Sensitivity analysis based on a dataset with less missing data supported this trend and suggested an additional shift in honey bees (Figure S6C). These analyses are consistent with the notion that, after the superorganismal corbiculate bees crossed a presumed major evolutionary transition in social complexity, they were subjected to further diversification in social complexity. This diversification may reflect selection pressures at the superorganism level rather than the individual level.

Evolutionary transition leading to phenotypic diversification

To quantitatively study the directionality of the evolutionary trajectories 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 Figure 5A). We used the trajectories in the high-dimensional phenotypic space of PCs 1–4. A directional evolutionary trajectory is expected to have a mean angle closer to π radians (equivalent to 180°), with low variance among angles (Figure 5A). Before the divergence of the superorganismal corbiculate bees from other species, the variance in the mean angle of the evolutionary trajectory of all species was higher (F test, $p < 0.001$), and the mean angle was lower (t test, $p < 0.001$) compared with after the divergence (Figure 5B). When comparing the mean angle of the evolutionary trajectory after the divergence between the superorganismal corbiculate bees and other lineages, we found that the superorganismal corbiculate bees had a mean angle closer to π (3.03 ± 0.09 radians) compared with other species

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

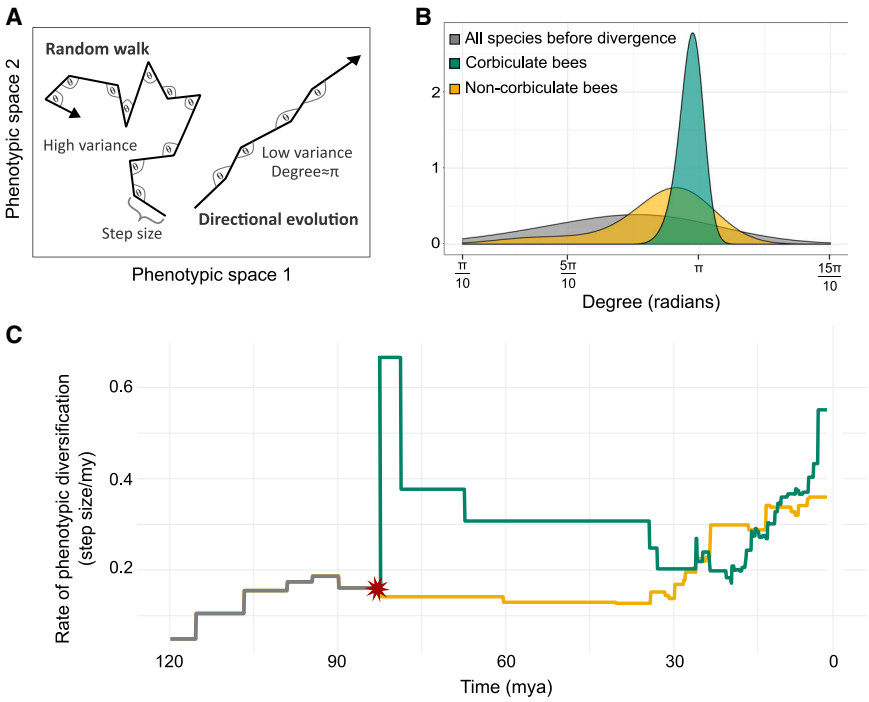


Figure 5. The directionality and rate of assumed evolutionary trajectories in the phenotypic space of social complexity

Analyses are based on PCs 1–4.

(A) Schematic representation of evolutionary trajectories 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 sizes are used to measure the degree of phenotypic change along a branch.

(B) Density plot of the turn degrees in the reconstructed evolutionary routes.

(C) The phenotypic diversification rates across time before and after the inferred major evolutionary transition (indicated with a red asterisk). The gray line shows the average diversification of all species before the split, and after the split, the rate for the superorganismal corbiculate bees is shown in green and that for the remaining taxonomic 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 (time).

(2.59 ± 0.64 radians; t test, $p < 0.001$). The variance in angles was significantly lower in the superorganismal corbiculate bees compared with the remaining species (F test, $p < 0.001$). These findings are consistent with the hypothesis that the increase in social complexity over time was more directional and less variable in the superorganismal corbiculate bees compared with the phylogenetic branches leading to social sweat bees, orchid bees, and Xylocopinae species.

Undergoing a major transition may remove biological constraints and open novel evolutionary opportunities, generating a phase in which the phenotypic space is rapidly explored, leading to a significant diversification. Our findings are consistent with this notion: The rate of phenotypic diversification in the superorganismal corbiculate bees substantially increases ~80 mya, coinciding with the directed adaptive regime shift (green line in Figure 5C). The phenotypic diversification of social complexity in the superorganismal corbiculate bees quadrupled in rate for a short period, then decreased to about half. By contrast, the remaining species show a rather constant low rate of diversification during this period. The later increase at ~30 mya may be attributed to the higher phylogenetic resolution we have closer to the present.

Testing for selection signals in socially related genes

The hypothesis that complex sociality evolved in a narrow evolutionary trajectory across lineages suggests strong selection on a limited number of genes or pathways. We examine this idea by analyzing a limited set of 12 genes, which have been linked to the evolution of sociality in insects in several studies,[57–65] and for which there is genomic data for a sufficient number of species in our dataset (Table S2). To evaluate signals of selection in these genes, we used a branch-site model to compare the molecular evolutionary rate (dN/dS values) of each branch on a phylogeny of the subset of 22 species (Table S3). We found that the dN/dS ratio is above 1 for 6 out of the 12 genes, but for none of the genes was the test statistically significant after correction for multiple comparisons (Table S4). In a complementary and less conservative approach, we used the likelihood-ratio test (LRT) values as a quantitative estimate for the evidence of positive selection acting on specific internal or terminal branches in our phylogeny. Using this approach, the evolution of superorganismal corbiculate bees did not show more directional selection signals compared with the remaining lineages (Table S5). We next used the values of our main PCs as proxies for social complexity and tested the correlation between PCs 1–4 and the LRT values of each species. We did not find a statistically significant correlation for any of the genes (Table S6).

DISCUSSION

We applied a bottom-up data-driven approach to analyze socially related traits, generating a high-dimensional phenotypic space of social complexity for 80 species of bees. By using multiple quantitative traits, we avoided the semantic inconsistencies and *a priori* assumptions inherent in qualitative classifications, which currently dominate the research on the evolution of social complexity. Quantitative data also enabled rigorous analysis of social complexity and its evolution using a broad toolkit of statistical and analytical methods that cannot be applied to qualitative data. Using this novel approach, we found that indices of social complexity are not necessarily positively correlated. These findings are consistent with recent findings showing trade-offs between social traits[56,66–68]; for example, a discrete increase in foundress adult life span in the transition from non-eusociality to eusociality is independent of colony size.[69] Together, these results are consistent with the notion that there is no

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

single or narrow evolutionary trajectory for increasing social complexity.[17,18,70] We identified clusters of species occupying defined areas in the phenotypic space of social complexity, corresponding to different social phenotypes (Figures 2 and S2). Importantly, there is substantial variability within each of these clusters that is typically overlooked in prevailing qualitative analyses. Ancestral state reconstructions suggest a single major transition toward a substantially higher level of social complexity, which occurred in the common ancestor of honey bees, stingless bees, and bumble bees (Figures 3, 4, and 5). This transition into complex societies was followed by considerable phenotypic diversification (Figure 5C). Two non-corbiculate species with relatively complex societies followed a different evolutionary trajectory. Notably, bumble bees, a group whose social complexity level is constantly debated,[18,19,24,31] align more closely with honey bees and stingless bees in our phenotypic space and are clearly separated from other species typically classified as primitively eusocial.

Amber fossil evidence from ~45 mya is consistent with our findings suggesting that complex sociality in corbiculate bees dates back to their common ancestor.[71,72] These fossils from extinct corbiculate tribes are apparently worker bees, with a barbed sting and reduced metasoma. There are no other bee groups with fossil evidence of a morphological caste system, which highlights the strong effect of phylogeny on levels of social complexity in bees. Additionally, except for social parasites, no extant superorganismal corbiculate bee species have reverted from social to genuinely solitary lifestyles. This suggests that high levels of social complexity represent an evolutionary adaptive peak in the phenotypic landscape,[73,74] making a reversion to a less complex lifestyle unlikely.[25] Social parasitism has evolved multiple times in bumble bees, most notably in the genus *Psithyrus*. Parasitic females do not rear their own workers but instead invade an already-established bumble bee colony, where they kill or dominate the resident queen. They then manipulate the workers of the colony into treating them as the new queen and caring for their offspring.[75] A parasitic lifestyle is also found in other social Hymenoptera, including *Lestrimelitta* robber stingless bees, ants, and wasps.[51,76,77] While the origins of this parasitic lifestyle remain largely obscure, it has been proposed to evolve from the hosts themselves (i.e., Emery's rule[78,79]). This adaptive lifestyle is considered an independent replicate for understanding the mechanisms of social dominance[80] and likely represents a new adaptive peak rather than a genuine reversion to a solitary lifestyle.[51,75]

The superorganismal corbiculate bees exhibit extensive phenotypic diversity in social complexity, highlighting the potential of multiple trait combinations to generate novel phenotypes. One intriguing example is the variability in sex ratio with honey bees and non-*Melipona* stingless bees showing a substantial male-biased ratio. A high proportion of male production in these lineages may have been driven by factors such as breeding structure,[81] relatedness asymmetry,[82] or colony foundation by swarming.[83] Our results suggest that extant diversity among superorganismal corbiculate bees is the result of phenotypic diversification following an adaptive and directional evolutionary regime shift estimated to occur approximately 70–80 mya (Figures 4 and 5). Conservative estimates based on fossil records suggest that clade diversity among social corbiculate bees was higher in the past, experiencing at least a 50% reduction in social bee clades during the Eocene epoch (55–35 mya).[71,72] This trend is consistent with the idea that the major transition has opened novel evolutionary opportunities, driving the occupation of new niches in the social complexity phenotypic landscape. Such a macroevolutionary process likely involves overcoming neurobiological, morphological, developmental, or ecological barriers, similar to the explosion of multicellular life following the development of sexual reproduction and division of labor,[84,85] or the rapid speciation observed after the colonization of remote islands.[86,87] Recent molecular evolution analyses showing that selection on social phenotypes acts on different genes and processes in superorganismal species compared with species that have not crossed the point of no return are in line with this notion.[88,89]

Sweat bees, orchid bees, and Xylocopinae bees exhibit lower levels of social complexity compared with superorganismal corbiculate bees, with phylogeny having a more variable influence on their social complexity. In our analyses (Figures 2 and S2), none of the clades in these lineages is restricted to a single group in our social phenotypic space. The evolutionary reconstruction of PCs 1–4 suggests both increases and decreases in these lineages, and our directionality analysis points to higher variability compared with the superorganismal lineages (Figure 5). This high variability is consistent with the notion that these lineages have not crossed the point of no return and aligns with previous suggestions for "gains and losses" of sociality in these groups.[20,24,90–95] The labile social phenotype may be evident even within the same species in which solitary and social lifestyles are expressed under different environmental conditions.[96–99] The evolutionary increases in social complexity in these taxonomic groups were relatively continuous, without evidence for stepwise (rungs on a ladder) changes. There is molecular evidence that increases in social complexity in these species are associated with incremental linear changes such as SNPs, indels, and differential gene expression.[57,100–103] By contrast, there is some evidence that the major transition to a superorganism may be associated with more substantial genome rearrangements, genome duplications, supergenes, and the emergence of novel genes.[104–108] The rapid increase in the availability of social complexity and genomics datasets sets the stage to rigorously explore the molecular evolutionary processes underlying the major transition from organismal to superorganismal social organization.[89]

Two species, the Xylocopinae *E. tridentata* and the sweat bee *L. marginatum*, are commonly classified as eusocial[93,109–111] and show a high level of social complexity in our analyses, but it is not as high as in stingless bees or honey bees. Their evolutionary trajectories toward high sociality occurred later and are different from that in the superorganismal corbiculate, questioning whether they indeed passed the point of no return. It is also notable that their social phenotype was inconsistent across our dimensionality reduction analyses (Figures 2 and S2). Their evolutionary trajectories also appear to differ from that of their close sister taxa, perhaps reflecting their unusual nesting biology. For example, *E. tridentata* displays strong queen-worker polymorphism but yet exhibits facultative social nesting, small colony size (only ~10 workers on average), and can have reproductive and mated workers.[110] Similarly, *L. marginatum* has

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

large colonies and long-lived queens,[109] but lacks apparent queen-worker castes. Additional studies are needed to better understand their social lifestyle and robustly reconstruct their evolutionary history to determine whether these species indeed show independent origins of superorganismality.[32,112,113]

The increased availability of life history, morphological, and social behavior data for an increasing number of species enabled us to describe, for the first time, the phenotypic space of social complexity in bees and to study it using rigorous quantitative methods. Our data-driven approach circumvents the need to rely on explicit or implicit assumptions, facilitating unconstrained evolutionary inference.[17,18,20,29–31] Contrary to expectations of the social ladder perspective, our findings do not support a continuous[24] or stepwise[11,26,114] progression toward the highest levels of sociality across independent bee groups. Our findings shed light on an important macroevolutionary process—traversal of a substantial phenotypic distance and subsequent diversification within superorganismal corbiculate bees. Our high-dimensional data-driven framework can be expanded to additional animal groups, including primates and birds.[10,115] 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 genomic, epigenetic, or proteomic data, as demonstrated in the limited genomic analysis we conducted here.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact Guy Bloch (guy.bloch@mail.huji.ac.il)

Materials availability

This study did not generate new, unique reagents.

Data and code availability

All data and code necessary to replicate the results and figures of this study are publicly available on GitHub at https://github.com/Greenbaum-Lab/bees.git.

This includes:

- The bee trait dataset used in this study.
- All original code used to analyze the data.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Omer Abramson for the beautiful drawings in the graphical abstract. We thank Michael Schwarz for providing us with important data on *Exoneurella tridentata* and Igor Medici de Mattos for helping us translate papers from Portuguese. We thank three anonymous reviewers for their excellent and constructive comments, which helped us improve the manuscript. This project was supported by grants from the Israel Science Foundation (ISF number 3391/20 and 2105/2 to G.B.) and the Center for Interdisciplinary Data Science Research (CIDR, number 3035000363 to G.G. and G.B.).

AUTHOR CONTRIBUTIONS

Conceptualization, O.P., G.G., and G.B.; study design, O.P., G.G., and G.B.; coding, O.P.; analysis, O.P.; data curation, O.P.; writing – original draft, O.P.; writing – review and editing, O.P., G.G., and G.B.; funding acquisition, G.G. and G.B.; supervision, G.G. and G.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Curating a dataset of social complexity traits in bees
  - Correlations and phylogenetic patterns of traits in the dataset
  - Dimensionality reduction visualization
  - Estimating the evolutionary trajectories of social complexity
  - Identifying shifts in social complexity
  - Estimation of directionality and phenotypic diversification in evolution
  - Phylogenetic tree
  - Sensitivity analysis
  - Selection in candidate genes of social complexity
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2025.01.009.

Received: July 9, 2024
Revised: October 16, 2024
Accepted: January 7, 2025
Published: February 10, 2025

REFERENCES

1. Rubenstein, D.R., and Abbot, P. (2017). The evolution of social evolution. In Comparative Social Evolution, D.R. Rubenstein, and P. Abbot, eds. (Cambridge University Press), pp. 1–18.
2. Wilson, E.O. (2000). Sociobiology: the New Synthesis (Harvard University Press).
3. Wheeler, W.M. (1915). The Social Insects: Their Origin and Evolution (Routledge).
4. Wheeler, W.M. (1910). Ants: Their Structure, Development and Behavior (Columbia University Press).
5. Shultz, S., Opie, C., and Atkinson, Q.D. (2011). Stepwise evolution of stable sociality in primates. Nature *479*, 219–222.
6. Kappeler, P.M., and Pozzi, L. (2019). Evolutionary transitions toward pair living in nonhuman primates as stepping stones toward more complex societies. Sci. Adv. *5*, eaay1276.
7. Olivier, C.-A., Martin, J.S., Pilisi, C., Agnani, P., Kauffmann, C., Hayes, L., Jaeggi, A.V., and Schradin, C. (2024). Primate social organization evolved from a flexible pair-living ancestor. Proc. Natl. Acad. Sci. USA *121*, e2215401120.
8. Smuts, B.B., Cheney, D.L., Seyfarth, R.M., Wrangham, R.W., and Struhsaker, T.T. (1987). Primate Societies (University of Chicago Press).
9. Downing, P.A., Griffin, A.S., and Cornwallis, C.K. (2020). Group formation and the evolutionary pathway to complex sociality in birds. Nat. Ecol. Evol. *4*, 479–486.
10. Crook, J.H. (1970). Social Behaviour in Birds and Mammals (Academic Press).

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

11. Michener, C.D. (1974). The Social Behavior of the Bees: a Comparative Study (Harvard University Press).
12. Batra, S.W.T. (1968). Behavior of some social and solitary halictine bees within their nests: a comparative study (Hymenoptera: Halictidae). J. Kans. Entomol. Soc. 120–133.
13. Wilson, E.O. (1971). The Insect Societies (Harvard University Press).
14. Hamilton, W.D. (1964). The genetical evolution of social behaviour. II. J. Theor. Biol. *7*, 17–52.
15. Smith, J.M., and Szathmary, E. (1997). The Major Transitions in Evolution (Oxford University Press).
16. West-Eberhard, M.J. (2003). Developmental Plasticity and Evolution (Oxford University Press).
17. Linksvayer, T.A., and Johnson, B.R. (2019). Re-thinking the social ladder approach for elucidating the evolution and molecular basis of insect societies. Curr. Opin. Insect Sci. *34*, 123–129.
18. Holland, J.G., and Bloch, G. (2020). The complexity of social complexity: A quantitative multidimensional approach for studies of social organization. Am. Nat. *196*, 525–540.
19. Boomsma, J.J., and Gawne, R. (2018). Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. Biol. Rev. Camb. Philos. Soc. *93*, 28–54.
20. Schwarz, M.P., Richards, M.H., and Danforth, B.N. (2007). Changing paradigms in insect social evolution: insights from halictine and allodapine bees. Annu. Rev. Entomol. *52*, 127–150.
21. Costa, J.T., and Fitzgerald, T.D. (1996). Developments in social terminology: semantic battles in a conceptual war. Trends Ecol. Evol. *11*, 285–289.
22. Anderson, C., and McShea, D.W. (2001). Individual versus social complexity, with particular reference to ant colonies. Biol. Rev. Camb. Philos. Soc. *76*, 211–237.
23. Danforth, B.N., Cardinal, S., Praz, C., Almeida, E.A.B., and Michez, D. (2013). The impact of molecular data on our understanding of bee phylogeny and evolution. Annu. Rev. Entomol. *58*, 57–78.
24. Kocher, S.D., and Paxton, R.J. (2014). Comparative methods offer powerful insights into social evolution in bees. Apidologie *45*, 289–305.
25. Wilson, E.O., and Hölldobler, B. (2005). Eusociality: origin and consequences. Proc. Natl. Acad. Sci. USA *102*, 13367–13371.
26. Rehan, S.M., and Toth, A.L. (2015). Climbing the social ladder: the molecular evolution of sociality. Trends Ecol. Evol. *30*, 426–433.
27. Taylor, D., Bentley, M.A., and Sumner, S. (2018). Social wasps as models to study the major evolutionary transition to superorganismality. Curr. Opin. Insect Sci. *28*, 26–32.
28. Evans, H.E., and Eberhard, M.J.W. (1973). The Wasps (University of Michigan Museum of Art).
29. Crespi, B.J., and Yanega, D. (1995). The definition of eusociality. Behav. Ecol. *6*, 109–115.
30. Gadagkar, R. (1994). Why the definition of eusociality is not helpful to understand its evolution and what should we do about it. Oikos *70*, 485–488.
31. Richards, M.H. (2019). Social trait definitions influence evolutionary inferences: a phylogenetic approach to improving social terminology for bees. Curr. Opin. Insect Sci. *34*, 97–104.
32. Shell, W.A., Steffen, M.A., Pare, H.K., Seetharam, A.S., Severin, A.J., Toth, A.L., and Rehan, S.M. (2021). Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees. Commun. Biol. *4*, 253.
33. Glastad, K.M., Arsenault, S.V., Vertacnik, K.L., Geib, S.M., Kay, S., Danforth, B.N., Rehan, S.M., Linnen, C.R., Kocher, S.D., and Hunt, B.G. (2017). Variation in DNA methylation is not consistently reflected by sociality in Hymenoptera. Genome Biol. Evol. *9*, 1687–1698.
34. Korb, J., Meusemann, K., Aumer, D., Bernadou, A., Elsner, D., Feldmeyer, B., Foitzik, S., Heinze, J., Libbrecht, R., Lin, S., et al. (2021). Comparative transcriptomic analysis of the mechanisms underpinning ageing and fecundity in social insects. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20190728.
35. Toth, A.L., Sumner, S., and Jeanne, R.L. (2016). Patterns of longevity across a sociality gradient in vespid wasps. Curr. Opin. Insect Sci. *16*, 28–35.
36. Kapheim, K.M., Pan, H., Li, C., Salzberg, S.L., Puiu, D., Magoc, T., Robertson, H.M., Hudson, M.E., Venkat, A., Fischman, B.J., et al. (2015). Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. Science *348*, 1139–1143.
37. Walsh, B., and Blows, M.W. (2009). Abundant genetic variation+ strong selection= multivariate genetic constraints: a geometric view of adaptation. Annu. Rev. Ecol. Evol. Syst. *40*, 41–59.
38. Wagner, G.P. (1988). The influence of variation and of developmental constraints on the rate of multivariate phenotypic evolution. J. Evol. Biol. *1*, 45–66.
39. Olson, E.C., and Miller, R.L. (1999). Morphological Integration (University of Chicago Press).
40. Arendt, J., and Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends Ecol. Evol. *23*, 26–32.
41. Avilés, L., Harwood, G., and Koenig, W. (2012). A quantitative index of sociality and its application to group-living spiders and other social organisms. Ethology *118*, 1219–1229.
42. Keller, L., and Perrin, N. (1995). Quantifying the level of eusociality. Proc. R. Soc. Lond. B *260*, 311–315.
43. Chak, S.T.C., Duffy, J.E., Hultgren, K.M., and Rubenstein, D.R. (2017). Evolutionary transitions towards eusociality in snapping shrimps. Nat. Ecol. Evol. *1*, 96.
44. Sherman, P.W., Lacey, E.A., Reeve, H.K., and Keller, L. (1995). Forum: The eusociality continuum. Behav. Ecol. *6*, 102–108.
45. Master of Science thesis Wyman, L.M. (2002). Intraspecific and interspecific social variation in the sweat bee Lasioglossum Malachurum and other members of the subgenus Evylaeus (Brock University).
46. Henríquez-Piskulich, P., Hugall, A.F., and Stuart-Fox, D. (2024). A supermatrix phylogeny of the world's bees (Hymenoptera: Anthophila). Mol. Phylogenet. Evol. *190*, 107963.
47. Bossert, S., Murray, E.A., Almeida, E.A.B., Brady, S.G., Blaimer, B.B., and Danforth, B.N. (2019). Combining transcriptomes and ultraconserved elements to illuminate the phylogeny of Apidae. Mol. Phylogenet. Evol. *130*, 121–131.
48. Wenseleers, T., and Ratnieks, F.L.W. (2006). Comparative analysis of worker reproduction and policing in eusocial Hymenoptera supports relatedness theory. Am. Nat. *168*, E163–E179.
49. Goulson, D. (2010). Bumblebees: Behaviour, Ecology, and Conservation (Oxford University Press).
50. Koeniger, G., Koeniger, N., and Phiancharoen, M. (2011). Comparative reproductive biology of honeybees. In Honeybees of Asia, H.R. Hepburn, and S.E. Radloff, eds. (Springer Berlin Heidelberg), pp. 159–206.
51. Grüter, C. (2020). Stingless Bees (Springer International Publishing).
52. Blows, M.W. (2007). A tale of two matrices: multivariate approaches in evolutionary biology. J. Evol. Biol. *20*, 1–8.
53. Johnson, R.A., and Wichern, D.W. (2002). Applied Multivariate Statistical Analysis (Pearson).
54. Coombes, K.R., Wang, M., and Coombes, M.K.R. (2019). Package 'PCDimension'. https://cran.r-project.org/web/packages/PCDimension/index.html.
55. Rehan, S.M., Richards, M.H., Adams, M., and Schwarz, M.P. (2014). The costs and benefits of sociality in a facultatively social bee. Anim. Behav. *97*, 77–85.
56. Séguret, A., Bernadou, A., and Paxton, R.J. (2016). Facultative social insects can provide insights into the reversal of the longevity/fecundity trade-off across the eusocial insects. Curr. Opin. Insect Sci. *16*, 95–103.

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

57. Kocher, S.D., Mallarino, R., Rubin, B.E.R., Yu, D.W., Hoekstra, H.E., and Pierce, N.E. (2018). The genetic basis of a social polymorphism in halictid bees. Nat. Commun. *9*, 4338.
58. Patel, A., Fondrk, M.K., Kaftanoglu, O., Emore, C., Hunt, G., Frederick, K., and Amdam, G.V. (2007). The making of a queen: TOR pathway is a key player in diphenic caste development. PLoS One *2*, e509.
59. de Azevedo, S.V., and Hartfelder, K. (2008). The insulin signaling pathway in honey bee (Apis mellifera) caste development—differential expression of insulin-like peptides and insulin receptors in queen and worker larvae. J. Insect Physiol. *54*, 1064–1071.
60. Manfredini, F., Riba-Grognuz, O., Wurm, Y., Keller, L., Shoemaker, D., and Grozinger, C.M. (2013). Sociogenomics of cooperation and conflict during colony founding in the fire ant Solenopsis invicta. PLoS Genet. *9*, e1003633.
61. Withee, J.R., and Rehan, S.M. (2017). Social aggression, experience, and brain gene expression in a subsocial bee. Integr. Comp. Biol. *57*, 640–648.
62. Sasaki, K., and Harano, K. (2010). Multiple regulatory roles of dopamine in behavior and reproduction of social insects. Trends Entomol. *6*, 1–13.
63. Amdam, G.V., Norberg, K., Hagen, A., and Omholt, S.W. (2003). Social exploitation of vitellogenin. Proc. Natl. Acad. Sci. USA *100*, 1799–1802.
64. Pandey, A., and Bloch, G. (2015). Juvenile hormone and ecdysteroids as major regulators of brain and behavior in bees. Curr. Opin. Insect Sci. *12*, 26–37.
65. Kocher, S., and Kingwell, C. (2024). The molecular substrates of insect eusociality. Annu. Rev. Genet. *58*, 273–295.
66. Blanchard, B.D., and Moreau, C.S. (2017). Defensive traits exhibit an evolutionary trade-off and drive diversification in ants. Evolution *71*, 315–328.
67. Powell, S., Price, S.L., and Kronauer, D.J.C. (2020). Trait evolution is reversible, repeatable, and decoupled in the soldier caste of turtle ants. Proc. Natl. Acad. Sci. USA *117*, 6608–6615.
68. Cueva del Castillo, R., Sanabria-Urbán, S., and Serrano-Meneses, M.A. (2015). Trade-offs in the evolution of bumblebee colony and body size: A comparative analysis. Ecol. Evol. *5*, 3914–3926.
69. da Silva, J. (2022). The extension of foundress life span and the evolution of eusociality in the Hymenoptera. Am. Nat. *199*, E140–E155.
70. Johnson, B.R., and Linksvayer, T.A. (2010). Deconstructing the superorganism: social physiology, groundplans, and sociogenomics. Q. Rev. Biol. *85*, 57–79.
71. Engel, M.S. (2001). Monophyly and extensive extinction of advanced eusocial bees: insights from an unexpected Eocene diversity. Proc. Natl. Acad. Sci. USA *98*, 1661–1664.
72. Barden, P., and Engel, M.S. (2021). Fossil social insects. In Encyclopedia of Social Insects (Springer), pp. 384–403.
73. Simpson, G.G. (1984). Tempo and Mode in Evolution (Columbia University Press).
74. Hansen, T.F., Svensson, E., and Calsbeek, R. (2012). Adaptive landscapes and macroevolutionary dynamics. Adapt. Landsc. Evol. Biol. *205*, 26.
75. Lhomme, P., and Hines, H.M. (2019). Ecology and evolution of cuckoo bumble bees. Ann. Entomol. Soc. Am. *112*, 122–140.
76. Buschinger, A. (2009). Social parasitism among ants: a review (Hymenoptera: Formicidae). Myrmecol. News *12*, 219–235.
77. Cervo, R., Stemmer, C., Castle, W., Queller, D., and Strassmann, J.E. (2004). Social parasitism of Polistes dominulus by Polistes nimphus (Hymenoptera, Vespidae). Insectes Soc. *51*, 101–108.
78. Lowe, R.M., Ward, S.A., and Crozier, R.H. (2002). The evolution of parasites from their hosts: intra–and interspecific parasitism and Emery's rule. Proc. Biol. Sci. *269*, 1301–1305.
79. Emery, C. (1909). Uber den Ursprung der dulotischen, parasitischen und myrekophilen Ameisen. Biol. Centralblatt *29*, 352–362.
80. Lhomme, P., and Hines, H.M. (2018). Reproductive dominance strategies in insect social parasites. J. Chem. Ecol. *44*, 838–850.
81. Fournier, D., Keller, L., Passera, L., and Aron, S. (2003). Colony sex ratios vary with breeding system but not relatedness asymmetry in the facultatively polygynous ant Pheidole pallidula. Evolution *57*, 1336–1342.
82. Sundström, L. (1994). Sex ratio bias, relatedness asymmetry and queen mating frequency in ants. Nature *367*, 266–268.
83. Bulmer, M.G. (1983). Sex ratio theory in social insects with swarming. J. Theor. Biol. *100*, 329–339.
84. Choi, S.-W., Graf, L., Choi, J.W., Jo, J., Boo, G.H., Kawai, H., Choi, C.G., Xiao, S., Knoll, A.H., Andersen, R.A., et al. (2024). Ordovician origin and subsequent diversification of the brown algae. Curr. Biol. *34*, 740–754.e4.
85. Ispolatov, I., Ackermann, M., and Doebeli, M. (2012). Division of labour and the evolution of multicellularity. Proc. Biol. Sci. *279*, 1768–1776.
86. Reznick, D.N., and Ricklefs, R.E. (2009). Darwin's bridge between microevolution and macroevolution. Nature *457*, 837–842.
87. Cooney, C.R., Bright, J.A., Capp, E.J.R., Chira, A.M., Hughes, E.C., Moody, C.J.A., Nouri, L.O., Varley, Z.K., and Thomas, G.H. (2017). Mega-evolutionary dynamics of the adaptive radiation of birds. Nature *542*, 344–347.
88. Wyatt, C.D.R., Bentley, M.A., Taylor, D., Favreau, E., Brock, R.E., Taylor, B.A., Bell, E., Leadbeater, E., and Sumner, S. (2023). Social complexity, life-history and lineage influence the molecular basis of castes in vespid wasps. Nat. Commun. *14*, 1046.
89. Sumner, S., Favreau, E., Geist, K., Toth, A.L., and Rehan, S.M. (2023). Molecular patterns and processes in evolving sociality: lessons from insects. Philos. Trans. R. Soc. Lond. B Biol. Sci. *378*, 20220076.
90. Cardinal, S., and Danforth, B.N. (2011). The antiquity and evolutionary history of social behavior in bees. PLoS One *6*, e21086.
91. da Silva, J. (2021). Life History and the Transitions to Eusociality in the Hymenoptera. Front. Ecol. Evol. *9*, 727124.
92. Crespi, B.J. (1996). Comparative analysis of the origins and losses of eusociality: causal mosaics and historical uniqueness. In Phylogenies and the comparative method in animal behavior (Oxford University Press), pp. 253–287.
93. Danforth, B.N., Conway, L., and Ji, S. (2003). Phylogeny of eusocial Lasioglossum reveals multiple losses of eusociality within a primitively eusocial clade of bees (Hymenoptera: Halictidae). Syst. Biol. *52*, 23–36.
94. Jones, B.M., Rubin, B.E.R., Dudchenko, O., Kingwell, C.J., Traniello, I.M., Wang, Z.Y., Kapheim, K.M., Wyman, E.S., Adastra, P.A., Liu, W., et al. (2023). Convergent and complementary selection shaped gains and losses of eusociality in sweat bees. Nat. Ecol. Evol. *7*, 557–569.
95. Gibbs, J., Brady, S.G., Kanda, K., and Danforth, B.N. (2012). Phylogeny of halictine bees supports a shared origin of eusociality for Halictus and Lasioglossum (Apoidea: Anthophila: Halictidae). Mol. Phylogenet. Evol. *65*, 926–939.
96. Davison, P.J., and Field, J. (2017). Season length, body size, and social polymorphism: size clines but not saw tooth clines in sweat bees. Ecol. Entomol. *42*, 768–776.
97. Rehan, S.M., and Richards, M.H. (2010). The Influence of Maternal Quality on Brood Sex Allocation in the Small Carpenter Bee, Ceratina calcarata. Ethology *116*, 876–887.
98. Soucy, S.L., Tugrul, G., and Ward, D. (2003). Solitary and group nesting in the orchid bee Euglossa hyacinthina (Hymenoptera, Apidae). Insectes Sociaux *50*, 248–250.
99. Soucy, S.L., and Danforth, B.N. (2002). Phylogeography of the socially polymorphic sweat bee Halictus rubicundus (Hymenoptera: Halictidae). Evolution *56*, 330–341.
100. Patalano, S., Vlasova, A., Wyatt, C., Ewels, P., Camara, F., Ferreira, P.G., Asher, C.L., Jurkowski, T.P., Segonds-Pichon, A., Bachman, M., et al. (2015). Molecular signatures of plastic phenotypes in two eusocial insect species with simple societies. Proc. Natl. Acad. Sci. USA *112*, 13970–13975.

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

101. Arsenault, S.V., Hunt, B.G., and Rehan, S.M. (2018). The effect of maternal care on gene expression and DNA methylation in a subsocial bee. Nat. Commun. *9*, 3468.

102. Rehan, S.M., Glastad, K.M., Steffen, M.A., Fay, C.R., Hunt, B.G., and Toth, A.L. (2018). Conserved genes underlie phenotypic plasticity in an incipiently social bee. Genome Biol. Evol. *10*, 2749–2758.

103. Kapheim, K.M. (2016). Genomic sources of phenotypic novelty in the evolution of eusociality in insects. Curr. Opin. Insect Sci. *13*, 24–32.

104. Johnson, B.R., and Tsutsui, N.D. (2011). Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. BMC Genomics *12*, 164.

105. Kent, C.F., Minaei, S., Harpur, B.A., and Zayed, A. (2012). Recombination is associated with the evolution of genome structure and worker behavior in honey bees. Proc. Natl. Acad. Sci. USA *109*, 18012–18017.

106. Schwander, T., Libbrecht, R., and Keller, L. (2014). Supergenes and complex phenotypes. Curr. Biol. *24*, R288–R294.

107. Simola, D.F., Wissler, L., Donahue, G., Waterhouse, R.M., Helmkampf, M., Roux, J., Nygaard, S., Glastad, K.M., Hagen, D.E., Viljakainen, L., et al. (2013). Social insect genomes exhibit dramatic evolution in gene composition and regulation while preserving regulatory features linked to sociality. Genome Res. *23*, 1235–1247.

108. Purcell, J., Brelsford, A., Wurm, Y., Perrin, N., and Chapuisat, M. (2014). Convergent genetic architecture underlies social organization in ants. Curr. Biol. *24*, 2728–2732.

109. Plateaux-Quénu, C. (1962). Biology of Halictus Marginatus Brullé. J. Apic. Res. *1*, 41–51.

110. Hurst, P.S. (2001). Social biology of Exoneurella tridentata, an allodapine bee with morphological castes and perennial colonies. PhD thesis (Flinders University).

111. Steen, Z., and Schwarz, M.P. (1998). Within-nest behaviour in a eusocial Australian allodapine bee Exoneura (Exoneurella) Tridentata Houston (Apidae: Xylocopinae). Trans. R. Soc. S. Aust. *122*, 55–63.

112. Dew, R.M., Rehan, S.M., Tierney, S.M., Chenoweth, L.B., and Schwarz, M.P. (2012). 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.

113. Schwarz, M.P., Bull, N.J., and Hogendoorn, K. (1998). Evolution of sociality in the allodapine bees: a review of sex allocation, ecology and evolution. Insectes Soc. *45*, 349–368.

114. Toth, A.L., and Robinson, G.E. (2007). Evo-devo and the evolution of social behavior. Trends Genet. *23*, 334–341.

115. Kappeler, P.M. (2019). A framework for studying social complexity. Behav. Ecol. Sociobiol. *73*, 1–14.

116. Munoz-Torres, M.C., Reese, J.T., Childers, C.P., Bennett, A.K., Sundaram, J.P., Childs, K.L., Anzola, J.M., Milshina, N., and Elsik, C.G. (2011). Hymenoptera Genome Database: integrated community resources for insect species of the order Hymenoptera. Nucleic Acids Res. *39*, D658–D662.

117. R Core Team (2013). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).

118. Proux, E., Studer, R.A., Moretti, S., and Robinson-Rechavi, M. (2009). Selectome: a database of positive selection. Nucleic Acids Res. *37*, D404–D407.

119. Sela, I., Ashkenazy, H., Katoh, K., and Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. *43*, W7–W14.

120. Landan, G., and Graur, D. (2008). Local reliability measures from sets of co-optimal multiple sequence alignments. Pac. Symp. Biocomput. *13*, 15–24.

121. Privman, E., Penn, O., and Pupko, T. (2012). Improving the performance of positive selection inference by filtering unreliable alignment regions. Mol. Biol. Evol. *29*, 1–5.

122. Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. *20*, 1160–1166.

123. Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., and Delsuc, F. (2018). MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol. Biol. Evol. *35*, 2582–2584.

124. Ranwez, V., Chantret, N., and Delsuc, F. (2021). Aligning protein-coding nucleotide sequences with MACSE. Multiple Sequence Alignment. In Methods in Molecular Biology (Springer), p. 51.

125. Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E.J.P. (2011). MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PLoS One *6*, e22594.

126. Rehan, S.M., Leys, R., and Schwarz, M.P. (2012). 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.

127. Dray, S., and Josse, J. (2015). Principal component analysis with missing values: a comparative survey of methods. Plant Ecol. *216*, 657–667.

128. Josse, J., and Husson, F. (2016). missMDA: a package for handling missing values in multivariate data analysis. J. Stat. Softw. *70*, 1–31.

129. Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature *401*, 877–884.

130. Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. *3*, 217–223.

131. Blomberg, S.P., Garland, T., Jr., and Ives, A.R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution *57*, 717–745.

132. Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., and Thuiller, W. (2012). How to measure and test phylogenetic signal. Methods Ecol. Evol. *3*, 743–756.

133. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. Bioinformatics *20*, 289–290.

134. Felsenstein, J. (1988). Phylogenies and quantitative characters. Annu. Rev. Ecol. Syst. *19*, 445–471.

135. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat Methodol. *57*, 289–300.

136. McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. Preprint at arXiv.

137. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. J. Stat. Softw. *25*, 1–18.

138. Konopka, T., and Konopka, M.T. (2018). R-package: umap. Uniform Manifold Approximation and Projection. https://cran.r-project.org/web/packages/umap/index.html.

139. Diaz-Papkovich, A., Anderson-Trocmé, L., and Gravel, S. (2021). A review of UMAP in population genetics. J. Hum. Genet. *66*, 85–91.

140. Chari, T., and Pachter, L. (2023). The specious art of single-cell genomics. PLoS Comput. Biol. *19*, e1011288.

141. Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W., Kozak, K.H., McPeek, M.A., Moreno-Roark, F., Near, T.J., et al. (2010). Early bursts of body size and shape evolution are rare in comparative data. Evolution *64*, 2385–2396.

142. Mahler, D.L., Ingram, T., Revell, L.J., and Losos, J.B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. Science *341*, 292–295.

143. O'Meara, B.C., Ané, C., Sanderson, M.J., and Wainwright, P.C. (2006). Testing for different rates of continuous trait evolution using likelihood. Evolution *60*, 922–933.

144. Mahler, D.L., Revell, L.J., Glor, R.E., and Losos, J.B. (2010). Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. Evolution *64*, 2731–2745.

145. Adams, D.C. (2013). Comparing evolutionary rates for different phenotypic traits on a phylogeny using likelihood. Syst. Biol. *62*, 181–192.

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

146. Hansen, T.F., and Martins, E.P. (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution *50*, 1404–1417.

147. Ingram, T., and Mahler, D.L. (2013). SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. Methods Ecol. Evol. *4*, 416–425.

148. Uyeda, J.C., and Harmon, L.J. (2014). A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. Syst. Biol. *63*, 902–918.

149. Khabbazian, M., Kriebel, R., Rohe, K., and Ané, C. (2016). Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. Methods Ecol. Evol. *7*, 811–824.

150. Bastide, P., Mariadassou, M., and Robin, S. (2017). Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. J. R. Stat. Soc. B *79*, 1067–1093.

151. Simpson, G.G. (1953). The Major Features of Evolution (Columbia University Press).

152. Bastide, P., Ané, C., Robin, S., and Mariadassou, M. (2018). Inference of adaptive shifts for multivariate correlated traits. Syst. Biol. *67*, 662–680.

153. Moon, T.K. (1996). The expectation-maximization algorithm. IEEE Signal Process. Mag. *13*, 47–60.

154. Baraud, Y., Giraud, C., and Huet, S. (2014). Estimator selection in the Gaussian setting. In Ann. Inst. H. Poincaré Probab. Statist., *50*, pp. 1092–1119.

155. Birgé, L., and Massart, P. (2007). Minimal penalties for Gaussian model selection. Probab. Theor. Relat. Fields *138*, 33–73.

156. Kareiva, P.M., and Shigesada, N. (1983). Analyzing insect movement as a correlated random walk. Oecologia *56*, 234–238.

157. Miyata, T., and Yasunaga, T. (1980). Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J. Mol. Evol. *16*, 23–36.

158. Yang, Z. (1994). Estimating the pattern of nucleotide substitution. J. Mol. Evol. *39*, 105–111.

159. Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. *15*, 568–573.

160. Bakewell, M.A., Shi, P., and Zhang, J. (2007). More genes underwent positive selection in chimpanzee evolution than in human evolution. Proc. Natl. Acad. Sci. USA *104*, 7489–7494.

161. Carneiro, M., Albert, F.W., Melo-Ferreira, J., Galtier, N., Gayral, P., Blanco-Aguiar, J.A., Villafuerte, R., Nachman, M.W., and Ferrand, N. (2012). Evidence for widespread positive and purifying selection across the European rabbit (Oryctolagus cuniculus) genome. Mol. Biol. Evol. *29*, 1837–1849.

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

163. Davydov, I.I., Salamin, N., and Robinson-Rechavi, M. (2019). Large-scale comparative analysis of codon models accounting for protein and nucleotide selection. Mol. Biol. Evol. *36*, 1316–1332.

164. Yang, Z., and Dos Reis, M. (2011). Statistical properties of the branch-site test of positive selection. Mol. Biol. Evol. *28*, 1217–1228.

165. Álvarez-Carretero, S., Kapli, P., and Yang, Z. (2023). Beginner's guide on the use of PAML to detect positive selection. Mol. Biol. Evol. *40*, msad041.

166. Jeffares, D.C., Tomiczek, B., Sojo, V., and dos Reis, M. (2015). A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. Methods Mol. Biol. *1201*, 65–90.

167. Liu, J., and Robinson-Rechavi, M. (2018). Adaptive evolution of animal proteins over development: support for the Darwin selection opportunity hypothesis of evo-devo. Mol. Biol. Evol. *35*, 2862–2872.

168. Daub, J.T., Moretti, S., Davydov, I.I., Excoffier, L., and Robinson-Rechavi, M. (2017). Detection of pathways affected by positive selection in primate lineages ancestral to humans. Mol. Biol. Evol. *34*, 1391–1402.

169. Studer, R.A., Penel, S., Duret, L., and Robinson-Rechavi, M. (2008). Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. Genome Res. *18*, 1393–1402.

170. Jaron, K.S., Parker, D.J., Anselmetti, Y., Tran Van, P., Bast, J., Dumas, Z., Figuet, E., François, C.M., Hayward, K., Rossier, V., et al. (2022). Convergent consequences of parthenogenesis on stick insect genomes. Sci. Adv. *8*, eabg3842.

171. Roux, J., Privman, E., Moretti, S., Daub, J.T., Robinson-Rechavi, M., and Keller, L. (2014). Patterns of positive selection in seven ant genomes. Mol. Biol. Evol. *31*, 1661–1685.

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw and analyzed data	This paper	Database: https://github.com/Greenbaum-Lab/bees.git; RRID:SCR_026102
Hymenoptera Genome Database	Torres et al.[116]	https://hymenopteramine.rnet.missouri.edu/hymenopteramine/begin.do
Software and algorithms		
R project	Team[117]	https://www.r-project.org/
Godon	Proux et al.[118]	https://bitbucket.org/Davydov/godon/
GUIDANCE2	Sela et al.[119]; Landan et al.,[120] Privman et al.,[121]	https://taux.evolseq.net/guidance/
MAFFT	Katoh et al.[122]	https://mafft.cbrc.jp/alignment/server/index.html
MACSE v2.0	Ranwez et al.[123–125]	https://github.com/ranwez/MACSE_V2_PIPELINES.git

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The social complexity trait data were retrieved following an extensive literature review detailed in Data S1. The coding domain sequences of social complexity related genes were taken from Hymenoptera Genome Database (HGD).

METHOD DETAILS

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.). 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 to describe various aspects of sociality, allowing us to acquire consistent data across multiple 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 information on the degree of reproductive skew in a colony; however, the degree of ovary activation or development is often only qualitatively described and is measured differently across species; therefore, the latter trait was rejected. 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 quantified but are important in the literature on insect sociality were coded as ordinal values (e.g., overlap of generations). Because our correlation matrix (Figure 1B) showed that traits are not necessarily correlated, the inclusion of similar traits does not inflate their importance in downstream analyses. To verify that no single trait biases our results, we conducted a sensitivity analysis in which we reanalyzed the data after removing every trait from the dataset (Figure S3C).

Our dataset includes species from all bee groups in which considerable information on 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

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

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 to avoid redundancy. Only two known species: *Hasinamelissa minuta* (Allodapine) and *Lasioglossum umbripenne* (Halictidae), might represent an informative sociality type[11,126]; however, insufficient data did not allow their inclusion. Furthermore, our sensitivity analyses with more equal sampling (Figure S3A) only reduced the variation within the phenotypic space but did not alter our results.

For each species in our dataset, we also provide the social complexity category using the common classification of Michener.[11] 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 80 species and 17 social traits, with 15% missing data in total. Four indices: colony average, colony maximum, colony founding, 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 Table S1 and Data S1.

To impute missing data, we selected methods that offer sufficient statistical power while preserving the original data distribution. For example, we avoided 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.[127] Based on our data structure, we used iterative PCA for imputation, which was found to perform well for data structures similar to ours, and specifically for PCA analyses.[127] For missing data imputation, we used the R package missMDA.[128] 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. Fourteen 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 reported 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,[129] which we estimated with maximum likelihood procedures using the phylosig function from the R package phytools.[130] Pagel's λ value ranges from 0 (the trait evolves independently from phylogeny) to 1 (the trait fully depends on phylogeny and its evolution fits a neutral Brownian motion); (ii) Blomberg's K,[131] 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. 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).[132] For both λ and K, we performed a likelihood-ratio test to evaluate the statistical significance of the phylogenetic signal.

To test whether the different social traits are correlated, as implied by the social ladder model, we computed 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,[133] which implements the method described by Felsenstein.[134] 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 all pairs of variables. To account for the possibility of spurious correlations, we used the Benjamini–Hochberg method,[135] which controls the false discovery rate, as implemented in the p.adjust function. Only significant correlations ($p<0.05$) were considered as indicating correlated traits. Importantly, interpreting any specific correlation as potentially causative requires additional analyses, and is beyond the scope of the current study; we therefore avoid inferring causality based on our findings.

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 visualizing tight clusters.[136] The combination of PCA and UMAP analyses provides us with a comprehensive reflection of a phenotypic space from our social complexity dataset (other dimensionality reduction techniques that focus primarily on local structures, such as t-SNE, would be less useful for this purpose).

For PCA, we used the R packages FactoMineR for visualization,[137] and MissMDA for performing PCA with missing data.[128] Before analysis, we standardized the trait values as z-scores. To correct the PCA for phylogeny, we applied pPCA using the phyl.pca function of the R package phytools,[130] which estimates the ancestral states of the traits using maximum likelihood under the λ model. This model allows variation in rates of evolution through time[129] 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,[54] to determine the number of informative PCs. This is done by randomly selecting n breakpoints (n being the number of variables in the dataset, 17 in our case) from a uniform distribution; under this method, a PC is considered informative 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. We considered the 4-dimensional volume which

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

represents the four informative PCs. Normalizing the PCs by weighting each one according to the variance it explains will not affect the results because the geometry of the convex hull changes uniformly for all species, as the same scaling factor is applied across all dimensions. We compared the convex hull volume of the superorganismal corbiculate bees to that of the remaining species.

Analysis of UMAP was implemented in the R package umap[138] with parameters set to a minimum distance of 0.2 and number of neighbors of 20. Because phylogenetic correction has not yet 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.

Estimating the evolutionary trajectories of social complexity

We investigated the phenotypic trajectories by which social complexity increased throughout bee evolution. We performed ancestral state reconstruction using the fastAnc function in the phytools package[130] and considered the PCA values (PCs 1-4) as describing the social complexity phenotype of the species, which together represent most of the 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.[139,140] Because our main goal was to provide a data-driven approach, we performed the reconstruction without including any *a priori* assumption regarding the evolution of sociality in bees. Hence, we did not fix the state of the common ancestor of bees or limit the transitions allowed between social complexity phenotypes. This approach introduces greater uncertainty in the reconstruction, especially near the root of the phylogeny. To quantify the extent of this uncertainty, we calculated 95% confidence intervals for the ancestral node estimates using the 'phenogram95' option in the 'fancyTree' function from the phytools package.

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.).[141–145] These methods are based on the adaptive evolution process of Ornstein–Uhlenbeck (OU).[74,146] 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[147–150] have applied similar multi-peak OU models to detect shifts in the pattern of trait evolution, creating an adaptive landscape of trait values that reflects Simpson's framework of adaptive zones across multivariate trait spaces and multiple selective optima.[151] This approach is particularly relevant for our framework because it does not require an *a priori* hypothesis of 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 along the phylogeny.[150,152] 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, if such optimum exists. This assumption facilitates the inference of the selection strength and drift rate for traits. The evolutionary trajectory of traits is modeled on a 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[153] 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. 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.[154] Djump (Jump Heuristic) looks for a jump in the model's complexity that leads to a significant improvement in performance.[155]

Estimation of directionality and phenotypic diversification in evolution

To explore the difference between species in the main sociality phenotypes (groups A-D; Figure 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.[156] Here, we used the high-dimensional phenotypic space of PCs 1-4 as the landscape and computed the turning angle between vectors along the evolutionary route of each species based on our ancestral reconstruction. 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 superorganismal corbiculate bees, superorganismal 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

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

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 taxonomic lineages. To compute phenotypic diversification, we again used the phenotypic space of PCs 1-4 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 ΔPC by the temporal length of the branch, in mya:

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

Phylogenetic tree

We used species from across all the major social taxonomic lineages of bees: Apidae (honey bees–*Apini*, stingless bees–*Meliponini*, bumble bees–*Bombini*, orchid bees–*Euglossini*), 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 complex sociality (eusociality) independently.[11,90] Time-calibrated phylogenetic tree to be used in our analyses was taken from 'beetreeoflife'[46] and included all species in our dataset except *Scaptorigona postica*, which was replaced with its closest sister taxa *Scaptotrigona polysticta* and *Augochlorella striata* which was replaced by its closest sister taxa *Augochlorella aurata*. 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.[46]

Sensitivity analysis

To evaluate the robustness of our results, we conducted a series of sensitivity analyses. These tests were conducted to verify whether our qualitative conclusions remained consistent under various assumptions and data structures.

Removal of traits

First, to assess whether specific traits bias our results, we reanalyzed the data after excluding each trait individually. We performed 17 separate PCA analyses and generated a single PCA plot showing the 95% confidence interval ellipse for each species, calculated based on the mean and standard deviation from all 17 analyses. To generate a single plot that remains coherent across multiple analyses, we preserved the orientation of the original PCA axes (i.e., multiplying the axes by -1) in cases of axis rotation.

Missing data threshold

Our dataset includes ~15% missing data, and the number of missing data entries ranged from 0-47% per species. To assess the influence of missing data on our conclusions, we repeated our analyses using subsets of the data that included only 10%, 8%, and 6% total missing data. Applying these thresholds resulted in the removal of 14, 24, and 32 species from the original dataset, respectively.

Taxonomic bias

Our dataset was limited by the availability of information in the literature for different species. As a result, species of honey bees, stingless bees, and bumble bees were overrepresented compared to species of orchid bees, sweat bees, and Xylocopinae bees. This unequal representation could introduce bias in our results, particularly in comparative phylogenetic analyses. To address this issue in a sensitivity analysis, we repeated our PCA analysis with an equal number of superorganismal corbiculate species (honey bees, stingless bees, and bumble bees) and non-superorganismal corbiculate species (orchid bees, sweat bees, and Xylocopinae bees).

Phylogenetic tree

The phylogenetic tree used in our main analyses was the consensus tree selected from a set of 1,000 bootstrap trees from the 'beetreeoflife' analysis.[46] Although all trees in this set are derived from the same analysis, they may differ in their topology and time calibration. To ensure our results are not affected by the values of a specific phylogenetic tree, we repeated our PCA analysis using a subset of 100 randomly chosen phylogenetic trees from the original set.

Selection in candidate genes of social complexity

To start investigating the potential association between dimensions in the social complexity phenotypic space with genomic signals, we used selection tests based on the non-synonymous to synonymous substitution ratio (dN\dS).[157,158] We focused on several candidate genes that were associated with aspects of social complexity in previous studies (e.g., regulation of reproduction, caste determination, dominance, brood care; Table S2). To avoid incorrect comparisons (e.g., caused by neofunctionalization or subfunctionalization), we only used single-copy genes (one-to-one orthologs). Orthologs for 22 species were taken from Hymenoptera Genome Database (HGD).[116] Species with no identified orthologs in certain genes were excluded only for the focal gene. Following Yang,[159] we used the coding domain sequence (CDS) for all the tested genes, and when multiple transcripts were available for a given gene due to alternative splicing, the longest transcript was kept.[160,161] We performed multiple sequence alignment of coding sequences using the MAFFT webserver,[122] and then fine-tuned the alignment using MACSE v2.0 software which accounts for frameshifts and stop codons.[123,124] This approach is suitable for dealing with protein-coding sequences because it considers the

Please cite this article in press as: Peled et al., Diversification of social complexity following a major evolutionary transition in bees, Current Biology (2025), https://doi.org/10.1016/j.cub.2025.01.009

corresponding amino acid translations during computation.[125] To further reduce alignment errors that can affect downstream inference, we filtered the data using GUIDANCE2 webserver, which provides reliability scores for the alignment of each gene.[119,120,162] To increase the likelihood of detecting signals of selection and to avoid unnecessary data loss, nucleotides with low confidence scores (below 0.93) were masked (replaced with N).[121] Overall, we consider the sequences generated with this approach conservative, i.e., more robust and error-prone but less powerful to detect signals of positive selection.

To test for a signature of positive selection we used the godon software[118] (https://bitbucket.org/Davydov/godon/, version 2020-02-17, option BS –ncat 4), which allows codon substitution rate variation along the sequence.[163] In codon analysis, the dN\dS (ω) value can be <1, =1, or >1, indicating purifying (negative) selection, neutral evolution, and positive selection, respectively.[158] We used the branch-site model, which detects positive selection acting on particular branches of the tree and particular sites in the sequence.[164] We compared the selection pattern of all branches relative to the background selection of the remaining branches. Each branch was tested iteratively, in one run per gene tree, using the species tree with branch lengths.[165,166]

Usually, tested branches (ω_2 in the analysis) that have a class of sites with a dN/dS ratio > 1 are candidates for positive selection. Although the strength of positive selection can, in principle, be estimated by the ratio ω_2 or by the proportion of sites in this class, previous studies have found the likelihood ratio (ΔlnL, also termed final D in the godon software) to be a good estimator of the evidence for positive selection.[167–171] In more detail, for each branch, the maximum likelihood of the data is estimated under two models: one that allows for positive selection (H1), and one that only allows negative selection and neutral evolution (H0), and a log-likelihood ratio statistic $\Delta\ln L = 2(\ln LH1 - \ln LH0)$ is computed. To determine their significance, we took two approaches. First, for each test, we performed the standard likelihood ratio test (LRT) to attain a *p*-value by comparing ΔlnL with a chi-square (χ^2) distribution with one degree of freedom. The resulting *p*-values were corrected for multiple comparisons using Benjamini and Hochberg's algorithm implemented in R, with α=0.05 (Table S4).

Second, to link the LRT values to our social complexity phenotypic space, we also used LRT as a quantitative estimator of positive selection, without applying any *a priori* significance cut-off. Using the non-parametric Mann-Whitney rank test, also compared the LRT values for all internal and terminal branches (i.e., species) between our two main social groups: superorganismal corbiculate bees and the remaining species (Table S5). We also tested the association between the LRT values with PCs 1-4 using the Spearman correlation test, implemented in the cor.test function in R (Table S6). Four out of the 22 species for which we had genetic sufficient data were not included in our trait analysis (*Eufriesea mexicana, Lasioglossum lativentre*, *Lasioglossum morio*, and *Lasioglossum albipes*); we therefore replaced them for this analysis with the mean values of socially and phylogenetically similar species from our dataset (Table S3).

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless stated otherwise, statistical analyses were performed in R v.4.2.1.[117] The details are presented in the results and STAR Methods sections.