Insect Systematics and Diversity, 2025, **9**(5), ixaf031
DOI: https://doi.org/10.1093/isd/ixaf031
Research

ENTOMOLOGICAL SOCIETY OF AMERICA
SHARING INSECT SCIENCE GLOBALLY

OXFORD

Down syndrome cell adhesion molecule 1 evolution and its relationship to sociality: hymenopteran Down syndrome cell adhesion molecule 1 exhibits accelerated evolution in variable exon regions

Remington R. Motte[1,2] **and Carl E. Hjelmen**[1,*]

[1]Department of Biology, Utah Valley University, Orem, UT, USA
[2]Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA

*Corresponding author. Department of Biology, Utah Valley University, 800 West University Parkway, Orem, UT 84058, USA (Email: Carl.Hjelmen@uvu.edu).

Subject Editor: Stephen Cameron

Down Syndrome Cell Adhesion Molecules (DSCAM) are transmembrane domain proteins belonging to the immunoglobulin superfamily and are expressed during brain development. While present in many other organisms, DSCAM has gone through multiple independent duplication events in arthropods and can form over 10,000 different isoforms through alternative splicing. Due to DSCAM's role in brain development and immune system functions, we investigated the relationship of DSCAM1 evolution to the evolution of sociality in arthropods. To assess structural variation, we examined phylogenetic trees derived from variable exons against the full DSCAM1 gene tree. Additionally, we performed likelihood ratio tests to identify regions undergoing evolutionary conservation or acceleration. Our findings reveal evidence of evolutionary acceleration in DSCAM1 within Hymenoptera, particularly in exons 4, 5, and 6, which correspond to the second and third immunoglobulin domains. Interestingly, this acceleration occurs regardless of social structure in bees and wasps. Accelerated evolution in these regions could have significant implications for neural circuit development in Hymenoptera, as these variable regions are responsible for generating diverse protein isoforms. This pattern of accelerated evolution potentially suggests that Hymenoptera have developed more complex neural circuits or undergone substantial changes in neuronal wiring.

Keywords: phylogenetics, comparative method, gene evolution, eusociality

Introduction

Sociality in arthropods is incredibly diverse, ranging from solitary to eusocial organization, with many other social classifications in between; eusociality is described as being one of the most complex examples of social organization throughout the animal kingdom (Nowak et al. 2010). The evolution of social complexity has been investigated through many different perspectives, including behavioral and ecological to genomic and transcriptomic (Wilson and Hölldobler 2005, Ferreira et al. 2013, Rehan 2021). Eusocial evolution is associated with the rapid evolution of gene regulation, constrained protein evolution, and a decrease in the abundance and diversity of transposable elements (Kapheim et al. 2015). Genes involved with synaptogenesis and neurogenesis, namely *derailed 2* and *frizzled*, experience constrained evolution within eusocial bees (Kapheim et al. 2015). In primitively eusocial bees, genes involved with neuronal development and differentiation were shown to have evidence of accelerated evolution (Woodard et al. 2011) and chemoreceptors experienced elevated positive selection in the transition to eusociality as well as the tendency to have expansion events in eusocial species (Terrapon et al. 2014, Zhou et al. 2015). These results suggest that genes important in neuronal development are often tied to social structure complexity.

Many interconnected gene trees contribute complex roadways to eusociality and the broader story of the evolution of eusociality. Given these complexities, the Down Syndrome Cell Adhesion Molecule (DSCAM) protein is a compelling candidate for further investigation. DSCAM are immunoglobulin proteins crucial for neuronal development; specifically, responsible for guiding neuronal cells to form the neural circuit through homophilic interactions, isoform-specific binding, chemoattraction, and self-avoidance (Agarwala et al. 2000, Garrett et al. 2012). The complexity of the neural network can be partially attributed to the extensive diversity of isoforms that DSCAM genes can generate through alternative splicing in the variable exon regions (Schmucker et al. 2000, Wojtowicz et al. 2004, Chen et al. 2006). The importance of DSCAM gene

Received: 21 January 2025; **Revised:** 30 April 2025; **Accepted:** 03 July 2025

variation is underscored by findings that link social immunity in hygienic bees to structural brain differences, which are likely mediated by differential expression of neuronal development genes such as DSCAM (Harpur et al. 2019). In addition to producing thousands of different isoforms, DSCAM has undergone multiple independent gene duplication events in arthropods, resulting in most having 4 to 6 DSCAM paralogs (Armitage et al. 2012). The presence of these DSCAM paralogs across the arthropod tree provides opportunities for the potential complex involvement in the evolution of social structure.

DSCAM can have 2 different forms: a membrane-bound protein involved with neural development and a soluble protein that plays a role in the innate immune system (Li 2021). The diversity of alternatively spliced isoforms allows DSCAM to specifically bind to pathogens and subsequently initiate phagocytosis and immune priming (Dong et al. 2006, Armitage et al. 2015, Li et al. 2018). With the loss of many immune genes in bees (Evans et al. 2006), DSCAM has undergone positive selection in the fifth immunoglobulin I-set domain before the evolution of sociality (Barribeau et al. 2015). Immune system evolution plays an important role in sociality, as many social insects have gained defenses against diseases (Cremer et al. 2007, Rosengaus et al. 2011), highlighting the importance of investigating DSCAM1's potential relationship to the evolution of eusociality.

Despite these well-documented roles in neuronal development and immune functions, research on DSCAM evolution has been predominantly focused on eusocial bees, mosquitos, and flies (Schmucker et al. 2000, Chen et al. 2006, Dong et al. 2006, Harpur et al. 2019). This narrow focus has resulted in a significant knowledge gap regarding DSCAM evolution and its relationship to sociality in other eusocial hymenopterans and different arthropod groups with complex social structures. To address this, we investigated the relationship between DSCAM1 evolution and the evolution of sociality across insects. We compared topologies and branch lengths between gene trees constructed from the DSCAM1 gene and the representative species tree using the same taxa, as well as comparing the topologies and branch lengths of the variable exons (4, 6, and 9) and the gene tree. Note, variable exon 11 was not reconstructed and compared as it only generates two protein isoforms in comparison to the dozens 4, 6, and 9 can generate (Armitage et al. 2012). We also investigate the conservation/acceleration of the DSCAM1 gene across the phylogeny between different orders and social structures. This investigation further elucidates the relationship between sociality and DSCAM1 evolution within hymenopterans.

Methods

Data Collection

For our phylogenetic tree reconstructions, we collected the DSCAM1 gene sequences in the NCBI GenBank database for 196 arthropods using the "rentrez" package in R (Winter 2017). Searches were conducted in November 2024, with the search term "DSCAM1[Gene name] AND Insecta[Organism]". Our search queries were specifically chosen to avoid the inclusion of other DSCAM paralogs. For our database, we extracted the species name, taxonomic order and family, gene ID, accession number, gene position, exon positions, and sociality. Using the entrez_fetch function in the "rentrez" package, we collected the entire DSCAM1 gene region for all 196 species into a fasta file and then trimmed out the introns using the annotated exon positions. All of the code used for collecting sequence data, fetching data, and trimming out introns is available at https://github.com/remingtonrimo/DSCAM-project and https://doi.org/10.5281/zenodo.15311036.

A species tree, with branch lengths measured in units of time, was taken from TimeTree.org (Kumar et al. 2022) and used for the comparison of node heights. Divergence estimates were collected from 2,274 studies, and hierarchical average linking was employed in order to ensure consistency across all of the studies (Hedges et al. 2015).

We categorized the sociality of each species (eusocial, social, and solitary) based on previously established definitions (Batra 1966, Gadagkar 1987, Crespi and Yanega 1995, Toth and Rehan 2017). Eusocial species were defined as those that share a common nesting site, cooperatively care for their young, exhibit reproductive castes, and display colony labor while the previous generation is still alive. For simplification, any species that lacked one or more of these characteristics (sub-social, quasi-social, or semi-social) was classified as social, and any species that lacked all of these traits were classified as solitary. Further classifications of hymenopterans for phyloP subgroup analyses included: long-tongue bees (Apidae and Megachilidae), short-tongue bees (Colletidae and Halictidae), vespids, parasitoids, sawflies, and ants.

Phylogeny Reconstruction

Using the online version of MAFFTv.7.407_1 (Katoh and Standley 2013), we generated an alignment for the DSCAM1 sequences. Preliminary maximum likelihood trees were also constructed using the FastTree2 parameters in NGPhylogeny to ensure taxonomic placement was in agreement with previous phylogenetic trees (Lemoine et al. 2019). Species that were missing large sections of the DSCAM1 gene or did not align very well and had erroneous placement compared to previous phylogenetic inferences were removed from future analyses (see *dscam.csv* in the supplementary data for a list of removed species). Some species sequences that were collected from NCBI did not have enough taxonomic information in TimeTree, so they were removed from the species tree versus gene tree comparison, which resulted in 117 species (see *dscam.csv* in the supplementary data for a list of removed species).

For tree reconstruction, we utilized BEAST v2.7.6 with the GTR+G+I substitution model, as determined using MEGA v11 (Tamura et al. 2021). The birth–death model was selected as the most appropriate for our dataset, given the extensive and diverse sampling of DSCAM1 across Insecta, while accounting for sparse lineage representation (Bouckaert et al. 2019). Additionally, we tested both the optimized relaxed clock model and the strict clock model. The species tree extracted from TimeTree.org was incorporated as a multi-monophyletic prior to ensure congruence between trees. To achieve sufficient effective sample sizes (ESS), the analyses were run for 50 million generations with the relaxed clock model and 150 million generations with the strict clock model. The assessment of our trees was conducted using Tracer.v.2.7.7, part of the BEAST package.

To compare the variable exon trees to the gene tree, we first reconstructed the DSCAM1 gene tree using maximum likelihood in IQtree using the GTR+G+I substitution model.

After constructing the initial DSCAM1 gene tree with 192 species, we used it as a prior—excluding species with minimal or no exon information—and as a constraint for exon tree reconstructions in BEAST v2.7.6. The same parameters used to reconstruct the gene tree from the previous comparison were used to reconstruct the exon trees. Species with a lack of annotation in the variable regions were removed from the analysis, resulting in 187 species in the comparison. BEAUti.v.2.7.6 and TreeAnnotator.v.2.7.6, included in the BEAST package, were used to convert the Nexus alignment file and generate a maximum credibility consensus tree with a 25% burn-in. The first 25 million trees for the exon trees were used to generate consensus trees so as not to include non-converged samples.

Comparing Phylogenies

Once the trees were reconstructed, the root nodes of each were constrained to 401 million years in accordance with an estimated divergence of Odonata and Hymenoptera when referring to TimeTree.org using the *chronos* function in the "ape" package (Paradis and Schliep 2019). By constraining the root of both trees at 401 million years, we assumed that any differences in node height after the initial splitting at the root are attributable to evolutionary change in DSCAM1. For example, if a DSCAM1 node appears deeper in time compared to the corresponding node in the TimeTree, it suggests that more evolutionary change has occurred in DSCAM1 at that node. Conversely, if the DSCAM1 node is shallower, it indicates slower evolution or conservation in that gene. To investigate the rate of evolution for DSCAM1, we compared the node heights of the species tree and DSCAM1 trees using an edited *compare.chronograms* function in "phytools" (Revell 2024). The *compare.chronogram* function was edited to include node rotation optimization from the *cophylo* function in "phytools."

We quantified the difference in estimated divergence times between the DSCAM1 and species trees by subtracting TimeTrees' branch lengths from DSCAM1 branch lengths. This was accomplished by creating a matrix of branch lengths for each phylogenetic tree using the *cophenetic* function in the "ape" package (Paradis and Schliep 2019) and subtracted the matrices [T(TimeTree)-T(DSCAM1)] in R. We followed the same procedure to compare the node heights of all 3 variable exon trees to the DSCAM1 gene maximum likelihood tree [T(gene)-T(exons)] on the family level rather than the species level. To visualize these differences in estimated divergence between the trees, we constructed a heatmap using the *heatmap.2* function in the "ggplot" package in R (Wickham 2016). Positive values indicate DSCAM1 conservation, whereas negative values indicate accelerated DSCAM1 evolution.

phyloP Analysis

We conducted a base-by-base likelihood ratio test on the DSCAM1 alignment of 192 species using the *phyloP* function available in the "rPHAST" package (Hubisz et al. 2011). First, we generated a model of neutral evolution from the DSCAM1 alignment and the maximum likelihood tree using the phyloFit program applying the GTR substitution model with four rate categories. The *tree_doctor* function was not available in rPHAST, so the interior nodes of our model file were named using the in-terminal version of PHAST. To investigate the impact of social structures on DSCAM1 evolution, we used the branch option in phyloP to partition the phylogeny according to these social structures. This allowed us to compare the conservation and acceleration of DSCAM1 across the different branches of the phylogeny. Additionally, we performed a phyloP analysis among insect orders and specifically within Hymenoptera to assess whether the rate of evolution is related to sociality or specific lineages. While social structure and eusociality occur outside of Hymenoptera, focusing within Hymenoptera also allowed us to control for confounding effects that haplodiploidy may have on the rate of evolution. We set our thresholds for acceleration and conservation at phyloP scores of –2 and 2, respectively, as these correspond to a *P* value of 0.01 (since $\log_{10}(0.01) = -2$).

Results

Most branches were estimated to be identical under both clock models (Fig. 1B). In comparison to the relaxed clock model, the strict clock produced longer branches in Hymenoptera and Hemiptera (ranges between 10 and 30 million years), but slightly shorter ones in Diptera (specifically *Drosophila*) and Lepidoptera (ranges between –10 and –20 million years)(Fig. 1B bottom). We obtained sufficient ESS values for the reconstruction under the relaxed clock model; however, despite running the strict clock model for 150 million generations, the ESS values were below 100 (see supplemental files on GitHub). We observed consistent patterns of branch length differences between the species tree and the gene tree across both clock models, with variations only in the magnitude of those differences.

When comparing the node heights of the DSCAM1 gene tree and the species tree (Fig. 1B, Supplementary Fig. S4), a clear pattern emerges: branch lengths in the species tree are generally longer than those in the DSCAM1 tree. However, notable exceptions occur in Hymenoptera, Lepidoptera, and Diptera (specifically *Drosophila*). Bees, wasps, and ants are estimated to have diverged approximately 100–150 million years ago during the Cretaceous period; the split in the DSCAM1 tree occurs earlier. The DSCAM1 tree also estimates earlier divergence times within Diptera (from the Paleogene to the upper Cretaceous) as well as Lepidoptera (from the lower Cretaceous to the Jurassic) (Supplementary Fig. S4).

The longer branch lengths in the DSCAM1 tree (regardless of clock model) compared to the species tree suggest greater sequence divergence and potentially a higher rate of evolutionary change. Conversely, in most other parts of the tree, divergence times occur later than expected, indicating that DSCAM1 is a highly conserved gene relative to the species tree. The areas showing the greatest conservation compared to the species tree are Coleoptera and Hemiptera, where branch length differences exceed 100 million years (Fig. 1). Similar patterns are observed when comparing node heights of exon- and gene family-level trees (Supplementary Fig. S2). Variable exons in Hymenoptera and Lepidoptera show longer branch lengths and greater sequence divergence, while other insect orders display shorter, more uniform branches. Hemiptera shows the least divergence, followed by the Lepidoptera–Diptera order split and Neuropteroidea–Panorpida superorder split.

The likelihood ratio test results from phyloP provide strong evidence for accelerated evolution in eusocial organisms (Fig. 2B). The lowest phyloP scores, ranging from 1300 to 2200

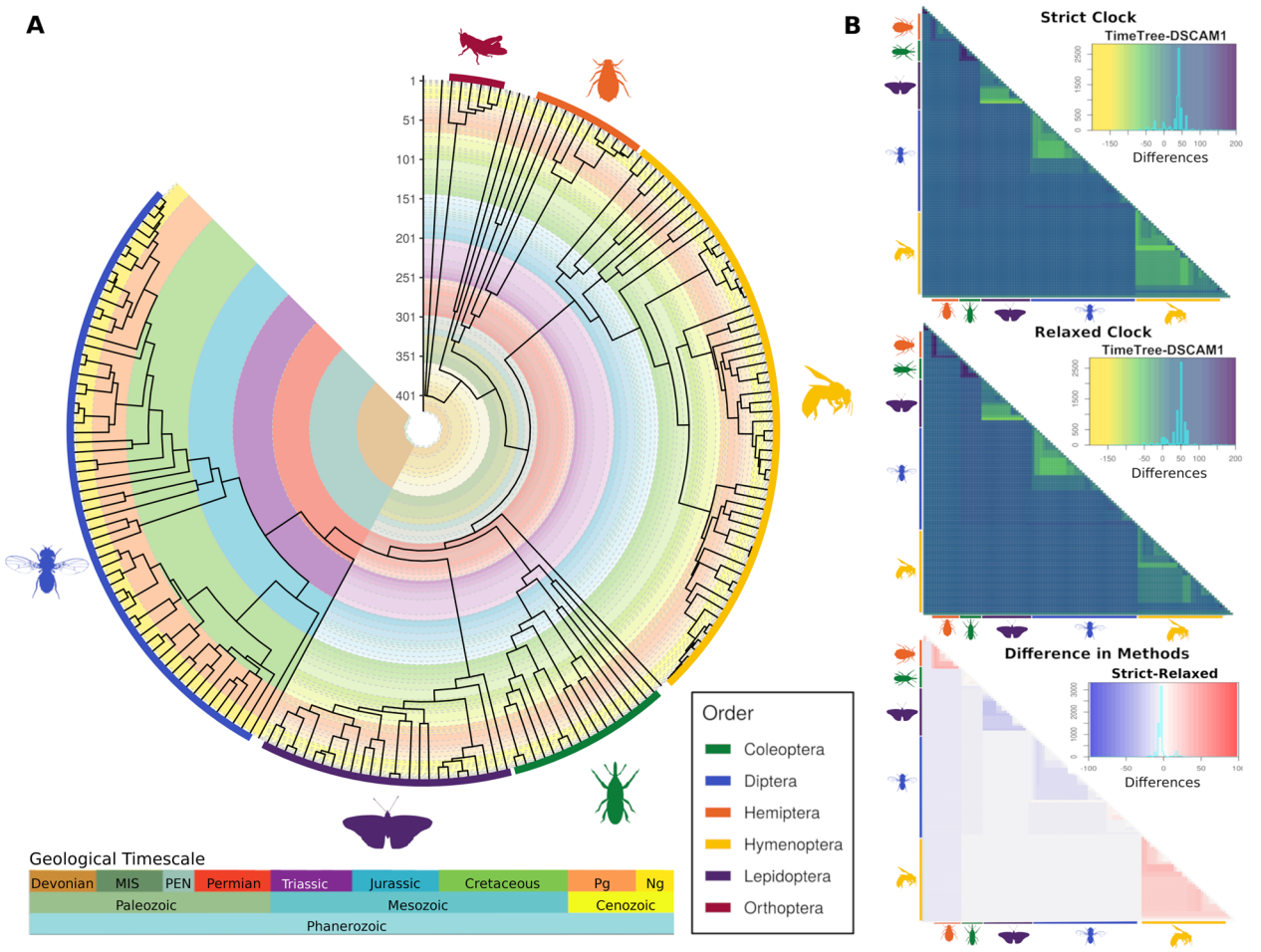

Fig. 1. Node height comparison between TimeTree and DSCAM1. A) Shows the species-level *DSCAM1* gene tree (made in IQtree) with a geological time scale plotted using *coord_geo* function in the "deeptime" and "ggtree" package (Xu et al. 2022, Gearty 2024). Orders not represented with a silhouette due to limited species sampling include Blattodea, Odonata, Phasmatodea, Siphonaptera, and Thysanoptera. B) Heatmaps for each method used, comparing the branch lengths between the TimeTree and the DSCAM1 gene tree (constrained by TimeTree topology, made in BEAST). Strict clock dating: log-Lik = –0.8339983, PHIIC = 695.67, relaxed clock dating: log-Lik = –3.550347, PHIIC = 701.1. DSCAM1 branch lengths were longer relative to the species tree in Hymenoptera, Lepidoptera and Drosophila and shorter in Hemiptera and Coleoptera. Compared to the strict clock model, the relaxed clock model estimated longer branch lengths within Hymenoptera and Hemiptera and shorter branch lengths in Lepidoptera and *Drosophila*. [public domain species silhouettes are from PhyloPic (Gearty and Jones 2023)].

nucleotides, correspond to exons 4, 5, and 6, which encode the second and third immunoglobulin domains (Fig. 3A). In solitary and social species, phyloP scores typically fall within the range of –2 to 2, with occasional outlier nucleotides. A closer examination at the nucleotide level reveals a higher proportion of nucleotides undergoing evolutionary acceleration in eusocial species compared to solitary or social species (Fig. 3C). Notably, the likelihood ratio test results remained consistent regardless of the number of rate categories specified in the phylogenetic model.

Hymenoptera shows the strongest evidence for acceleration in exons 4, 5, and 6 as well as a greater proportion of accelerated nucleotides in this region (Supplementary Fig. S1). Other orders (Coleoptera, Diptera, Hemiptera, Lepidoptera, and Orthoptera) show minimal evidence of accelerated evolution in comparison. Occasional nucleotides for these orders do show high levels of acceleration, for instance, a nucleotide in the Fibronectin Domain III for Lepidoptera (Supplementary Fig. S1A). A closer analysis of Hymenoptera reveals distinctly lower phyloP scores in Apidae and Megachilidae (long-tongue bees), alongside evidence of evolutionary acceleration in vespid wasps (Fig. 3A, Supplementary Fig. S3). In contrast, short-tongue bees, ants, parasitic wasps, and sawflies DSCAM1 remains under neutral evolution, with minimal deviation of phyloP scores from zero. For long-tongue bees, most acceleration is concentrated in exons 4 and 5, while vespids exhibit acceleration near the end of exon 4 as well as in exons 5 and 6. Additionally, long-tongue bees and vespid wasps have a higher proportion of nucleotides that are accelerated (Fig. 3B). Notably, no consistent patterns of phyloP scores were identified across all eusocial or solitary organisms.

Discussion

The initial results from the base-by-base likelihood ratio test indicate that DSCAM1 experienced accelerated evolution within eusocial species; however, taking a closer examination of Hymenoptera we do not see the same pattern. The acceleration of exons 4, 5, and 6 appears to result more from lineage-specific shifts in Apidae, Megachilidae, and Vespidae, rather than broad, repeated patterns tied to the independent origins of eusociality

Fig. 2. Base-by-base likelihood ratio test using the *phyloP* function available in the "rPHAST" package. A) Shows the 10 immunoglobulin domains and the 6 fibronectin domains in DSCAM1 for *Drosophila melanogaster* as well as the MAFFT alignment used for phylogenetic reconstruction. Shaded areas correspond to the variable exons 4, 6, and 9 (immunoglobulin domains 2, 3, and 7) from left to right. B) The phyloP scores for eusocial (top), social (middle), and solitary organisms (bottom) across the length of the DSCAM1 alignment (Wickham 2016). PhyloP scores that are positive represent constrained evolution, whereas negative phyloP scores represent evolutionary acceleration. Accelerated regions in eusocial species are highlighted in boxes, as well as the corresponding areas for social and solitary organisms. C) Corresponding phyloP scores for each exon in the 1,300–2,200 area highlighted in panel B.

(Fig. 3), as similar acceleration was not observed in ants or Halictidae. It is plausible that signs of acceleration in solitary bees were obscured by other solitary organisms across the phylogeny during the initial likelihood ratio test, leading to biased results. This inference is supported by a comparison of branch lengths between the DSCAM1 phylogeny and the species tree, which reveals longer branch lengths for bees.

Comparing the phyloP scores across each order, we do not see strong evidence of acceleration in any specific area for *Drosophila* or Lepidoptera (Supplementary Fig. S1) despite seeing differences in branch lengths between the gene tree and species tree. Several factors could explain these discrepancies. One possibility is that the inclusion of additional species in the likelihood base-by-base test may have helped normalize some of the branch lengths. Another potential explanation is the imposition of a species-tree topology on the DSCAM1 gene tree may have artificially inflated branch lengths, particularly in Lepidoptera, a group known to have a history of hybridization (Lushai et al. 2005, Kronforst et al. 2006). The branch length differences between the species tree and gene trees in *Drosophila* may also be amplified by gene tree discordance (Rosenberg and Tao 2008). This discordance tends to become more complex with increasing taxon sampling, and *Drosophila* was the most densely sampled group in our analysis.

Our results for Hymenoptera (Fig. 3) are consistent with previous research on bee immune system genes, which identified strong positive selection acting on DSCAM across the social gradient (Barribeau et al. 2015). Similar patterns in gene evolution related to social structure have been found in the expansion of chemoreceptors, where the rate and patterns of evolution were not confined to any social order in Hymenoptera (Zhou et al. 2015, Karpe et al. 2017). Together, our findings, alongside evidence from prior research, suggest that accelerated or conserved evolution in genes related to brain development is not strictly tied to social structure.

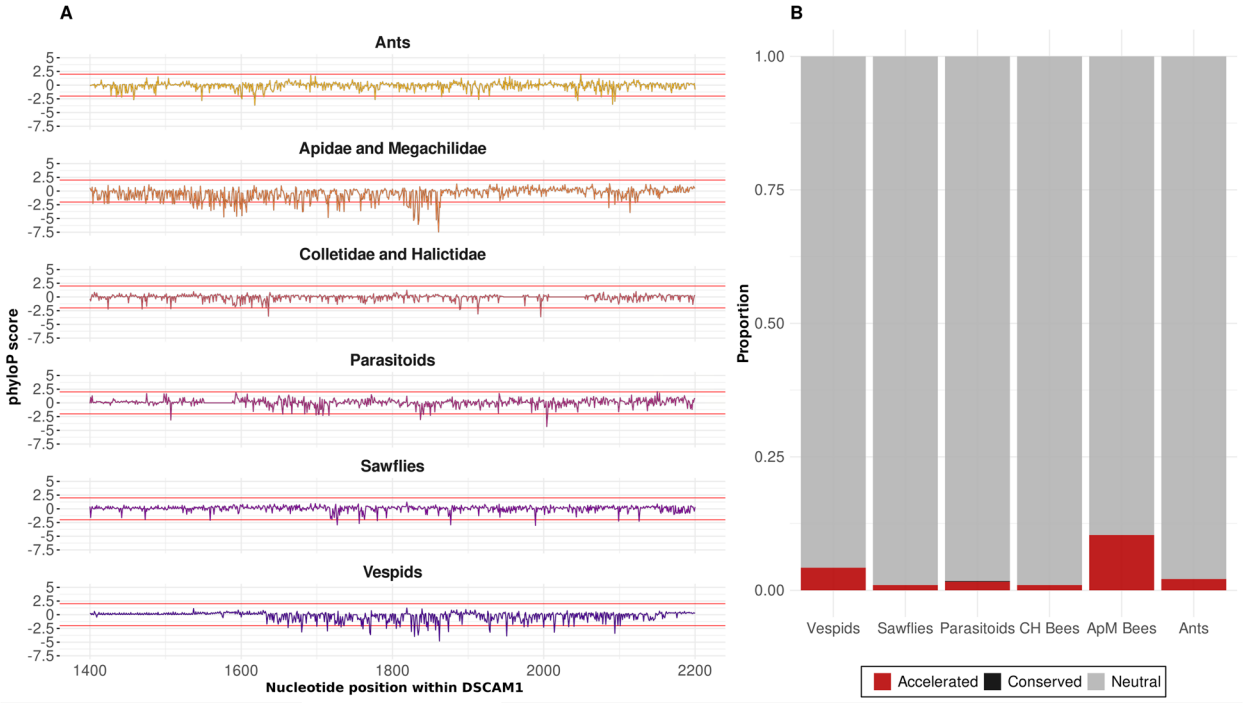

Fig. 3. Base-by-base likelihood ratio phyloP scores for ant, short-tongue bees, long-tongue bees, parasitoids, sawfly, and vespid branches. A) The phyloP scores for ants, apid and megachilid bees, colletid and halictid bees, parasitic wasps, sawflies, vespid wasps across exons 4, 5, and 6. See Supplementary Fig. S2 for the phyloP scores across the entire length of the DSCAM1 gene. B) Proportion of sites that are conserved (score > 2), neutral (–2 ≤ score ≤ 2) or accelerated (score < –2) in exons 4, 5, and 6. The cutoff was determined because phyloP = 2 corresponds to a *P* value of 0.01 (log10(0.01) = 2). 10% of the nucleotides in this region for apid and megachilid bees were accelerated, as well as 4% for vespid wasps, 2% for ants, 1% for parasitoids, and >1% for sawflies and colletids/halictid bees.

Previous research identified positive selection acting on the fifth I-set immunoglobulin domain in bees (Barribeau et al. 2015). In contrast, our study provides evidence of positive selection within the variable regions (4 and 6) of the DSCAM1 gene, corresponding to the second and third immunoglobulin domains. It is important to note that Barribeau et al. (2015) did not assess these domains due to missing data, so their absence from their findings does not rule out similar patterns. The acceleration observed in these variable regions has intriguing neurological implications. The variable exons of DSCAM1 enable the generation of tens of thousands of distinct protein isoforms, a feature critical for axonal guidance as it facilitates the precise formation of neural circuits (Schmucker et al. 2000). Accelerated evolution in these regions could significantly influence neural circuitry, potentially indicating hymenopterans have developed a more complex neural circuit or other substantial changes in neuronal wiring have occurred.

Throughout this investigation, we considered incorporating a dN/dS analysis; however, the presence of variable regions in our alignment posed significant challenges. These regions contain numerous gaps, which would introduce uncertainties and potentially yield inaccurate results in a dN/dS analysis. The gaps likely arise from incomplete annotations for some species and the possibility that not all species possess the same variable exons. One potential solution would be to remove the variable exons from the alignment, but this approach risks discarding critical evolutionary information contained within these regions. Alternatively, we could include both introns and exons in the analysis to capture the full sequence context. However, this approach would require excluding introns from the calculations to avoid overestimating synonymous mutations, which could bias the results. Additionally, including introns would not necessarily improve alignment quality and would significantly increase the complexity of the analysis.

Other studies have used the dN/dS approach in bees and on similar genes (Kapheim et al. 2015, Warner et al. 2019); however, our study differs in specificity. For instance, Kapheim et al. (2015) focused on 10 bee species, and Warner et al. (2019) concentrated on the pharaoh ant and honey bees, whereas our study included a much broader selection of species. We selected phyloP as the most suitable method for analyzing evolutionary acceleration and conservation because it enables lineage-specific comparisons and the identification of hotspot regions within DSCAM1 (Hubisz et al. 2011). Additionally, phyloP incorporates neutral substitution models, avoiding the reliance on the assumption of synonymous site neutrality (Spielman and Wilke 2015) and allows us to integrate the phylogeny into the analysis (Hubisz et al. 2011).

The scope of this study only investigated the evolution of DSCAM1 and did not include other DSCAM paralogs (2, 3, and 4). Expanding the analysis to incorporate these paralogs could provide a deeper understanding of the complexities of neuronal development across Arthropoda. For instance, while ants did not exhibit acceleration within this paralog of DSCAM, they may show such patterns in other paralogs. We excluded sequences named "Dscam2-like" in our analysis due to the uncertainty regarding their identity. As a result, many social bee species, including *Lasioglossum* and the well-known *Apis mellifera*, were likely unintentionally omitted. The effect of excluding *A. mellifera* is likely minimal due to our inclusion of

other *Apis* species. However, the exclusion of *Lasioglossum* may limit our ability to fully capture patterns of DSCAM1 evolution associated with the emergence of sociality.

This study would benefit from the inclusion of eusocial species from diverse orders, as well as more social/social species in general. Notably, termites were underrepresented, with only *Zootermopsis nevadensis* included in the analysis. Other key species to consider include ambrosia beetles, gall-dwelling aphids, thrips, and other social organisms. While the results of this study do not suggest a clear relationship between acceleration and sociality, incorporating these species could offer a more comprehensive view of DSCAM evolution and its potential connections to sociality. That being said, analyzing the evolution of DSCAM in relation to sociality does not encapsulate the complexities of eusocial evolution. The evolution of eusociality is likely the culmination of complex interactions of many genes, including other DSCAM and neurological genes. Moreover, these gene interactions may differ across species as certain genes may not be involved in every instance of eusociality.

Other promising avenues for investigating the role of neurological genes in the evolution of sociality include clustered protocadherins (PCDHs). These genes are organized into 3 major clusters (α, β, and γ) and share homologous functions with DSCAM, particularly in neurodevelopment and neurite self-avoidance (Chen and Maniatis 2013). Notable PCDHs to explore include PCDH19 (Lim et al. 2019), whose deletion has been linked to autism-like behaviors in mice; PCDH11X/Y, which has been associated with verbal language development in humans (Nardello et al. 2021); and PCDH17, which is connected to cognition and personality (Chang et al. 2018). There are likely many other PCDH genes worth investigating in this context as well.

Conclusion

Our findings suggest that accelerated evolution in DSCAM1 is not exclusive to any particular social classification but appears to be specific to long-tongue bees and wasps. This acceleration is primarily observed within the variable regions of the gene, which may reflect a more complex neural circuit or other significant changes. Further research is needed to explore the impact of accelerated evolution in these variable regions on brain and neuron morphology, as well as associated behaviors. The emergence of eusociality in arthropods is likely the result of complex interactions between many genes, and our study only begins to address these intricate gene interactions. Future research should include a broader range of eusocial species and investigate other DSCAM paralogs to determine whether these evolutionary patterns extend beyond Hymenoptera and DSCAM1.

Acknowledgements

We would like to acknowledge the members of the Hjelmen lab, Dr Jeremy Beaulieu, Christopher Wilson, and anonymous reviewers for their feedback, insightful contributions, and helpful reviews of the manuscript. Dr Jessica Cusick at Utah Valley University kindly provided helpful comments about how we should treat sociality as a trait at the beginning of the project. Finally, we extend our gratitude to the UVU College of Science and the Department of Biology for their support throughout the entirety of the project.

Author Contributions

Remington Motte (Conceptualization [equal], Data curation [equal], Formal Analysis [equal], Investigation [equal], Methodology [equal], Resources [equal], Software [equal], Validation [lead], Visualization [lead], Writing—original draft [lead], Writing—review & editing [equal]) and Carl Hjelmen (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Project administration [lead], Resources [equal], Software [equal], Supervision [lead], Writing—original draft [supporting], Writing—review & editing [equal])

Supplementary Material

Supplementary material is available at *Insect Systematics and Diversity* online.

Funding

None declared.

Conflict of Interest

None declared.

Data and Code Availability

All data, including NCBI accession numbers, log files, fasta files, alignments, tree files, mod files, and code, are available at https://doi.org/10.5281/zenodo.15311036 and https://github.com/remingtonrimo/DSCAM-project.

References

Agarwala KL, Nakamura S, Tsutsumi Y, et al. 2000. Down syndrome cell adhesion molecule DSCAM mediates homophilic intercellular adhesion. *Mol. Brain Res.* 79:118–126. https://doi.org/10.1016/S0169-328X(00)00108-X

Armitage SA, Freiburg RY, Kurtz J, et al. 2012. The evolution of DSCAM genes across the arthropods. *BMC Evol. Biol.* 12:53. https://doi.org/10.1186/1471-2148-12-53

Armitage SAO, Peuss R, Kurtz J. 2015. Dscam and pancrustacean immune memory—a review of the evidence. *Dev. Comp. Immunol.* 48:315–323. https://doi.org/10.1016/j.dci.2014.03.004

Barribeau SM, Sadd BM, du Plessis L, et al. 2015. A depauperate immune repertoire precedes evolution of sociality in bees. *Genome Biol.* 16:83. https://doi.org/10.1186/s13059-015-0628-y

Batra S. 1966. Nests and social behavior of halictine bees of India (Hymenoptera: Halictidae). *Indian J. Entomol.* 28:375–393.

Bouckaert R, Vaughan TG, Barido-Sottani J, et al. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Chang H, Hoshina N, Zhang C, et al. 2018. The protocadherin 17 gene affects cognition, personality, amygdala structure and function, synapse development and risk of major mood disorders. *Mol. Psychiatry* 23:400–412. https://www.nature.com/articles/mp2016231

Chen BE, Kondo M, Garnier A, et al. 2006. The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. *Cell* 125:607–620. https://doi.org/10.1016/j.cell.2006.03.034

Chen WV, Maniatis T. 2013. Clustered protocadherins. *Development* 140:3297–3302. https://doi.org/10.1242/dev.090621

Cremer S, Armitage SAO, Schmid-Hempel P. 2007. Social immunity. *Curr. Biol.* 17:R693–R702. https://doi.org/10.1016/j.cub.2007.06.008

Crespi BJ, Yanega D. 1995. The definition of eusociality. *Behav. Ecol.* 6:109–115. https://doi.org/10.1093/beheco/6.1.109

Dong Y, Taylor HE, Dimopoulos G. 2006. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol.* 4:e229. https://doi.org/10.1371/journal.pbio.0040229

Evans JD, Aronstein K, Chen YP, et al. 2006. Immune pathways and defence mechanisms in honey bees *Apis mellifera*. *Insect Mol. Biol.* 15:645–656. https://doi.org/10.1111/j.1365-2583.2006.00682.x

Ferreira PG, Patalano S, Chauhan R, et al. 2013. Transcriptome analyses of primitively eusocial wasps reveal novel insights into the evolution of sociality and the origin of alternative phenotypes. *Genome Biol.* 14:R20. https://doi.org/10.1186/gb-2013-14-2-r20

Gadagkar R. 1987. What are social insects? IUSSI Indian Chapter Newsl.:3–4.

Garrett AM, Tadenev ALD, Burgess RW. 2012. DSCAMs: restoring balance to developmental forces. *Front. Mol. Neurosci.* 5:86. https://doi.org/10.3389/fnmol.2012.00086

Gearty W. 2025. deeptime: an R package that facilitates highly customizable and reproducible visualizations of data over geological time intervals. *Big Earth Data*, 1–17. https://doi.org/10.1080/20964471.2025.2537516

Gearty W, Jones LA. 2023. rphylopic: An R package for fetching, transforming, and visualising PhyloPic silhouettes. *Methods Ecol. Evol.* 14112700–2708. 10.1111/2041-210X.14221

Harpur BA, Guarna MM, Huxter E, et al. 2019. Integrative genomics reveals the genetics and evolution of the Honey Bee's social immune system. *Genome Biol. Evol.* 11:937–948. https://doi.org/10.1093/gbe/evz018

Hedges SB, Marin J, Suleski M, et al. 2015. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* 32:835–845. https://doi.org/10.1093/molbev/msv037

Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* 12:41–51. https://doi.org/10.1093/bib/bbq072

Kapheim KM, Pan H, Li C, et al. 2015. Genomic signatures of evolutionary transitions from solitary to group living. *Science* 348:1139–1143. https://doi.org/10.1126/science.aaa4788

Karpe SD, Dhingra S, Brockmann A, et al. 2017. Computational genome-wide survey of odorant receptors from two solitary bees *Dufourea novaeangliae* (Hymenoptera: Halictidae) and *Habropoda laboriosa* (Hymenoptera: Apidae). *Sci. Rep.* 7:10823. https://doi.org/10.1038/s41598-017-11098-z

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–780. https://doi.org/10.1093/molbev/mst010

Kronforst MR, Young LG, Blume LM, et al. 2006. Multilocus analyses of admixture and introgression among hybridizing *Heliconius* Butterflies. *Evolution* 60:1254–1268. https://doi.org/10.1111/j.0014-3820.2006.tb01203.x

Kumar S, Suleski M, Craig JM, et al. 2022. TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* 39:msac174. https://doi.org/10.1093/molbev/msac174

Lemoine F, Correia D, Lefort V, et al. 2019. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* 47:W260–W265. https://doi.org/10.1093/nar/gkz303

Li W. 2021. Dscam in arthropod immune priming: what is known and what remains unknown. *Dev. Comp. Immunol.* 125:104231. https://doi.org/10.1016/j.dci.2021.104231

Li X-J, Yang L, Li D, et al. 2018. Pathogen-specific binding soluble Down Syndrome Cell Adhesion Molecule (Dscam) regulates phagocytosis via membrane-bound Dscam in Crab. *Front. Immunol.* 9:801. https://doi.org/10.3389/fimmu.2018.00801

Lim J, Ryu J, Kang S, et al. 2019. Autism-like behaviors in male mice with a Pcdh19 deletion. *Mol. Brain* 12:95. https://doi.org/10.1186/s13041-019-0519-3

Lushai Gugs, Allen JA, Goulson Dave, et al. 2005. The butterfly *Danaus chrysippus* (L.) in East Africa comprises polyphyletic, sympatric lineages that are, despite behavioural isolation, driven to hybridization by female-biased sex ratios. *Biol. J. Linn. Soc.* 86:117–131. https://doi.org/10.1111/j.1095-8312.2005.00526.x

Nardello R, Antona V, Mangano GD, et al. 2021. A paradigmatic autistic phenotype associated with loss of PCDH11Y and NLGN4Y genes. *BMC Med. Genomics* 14:98. https://doi.org/10.1186/s12920-021-00934-x

Nowak MA, Tarnita CE, Wilson EO. 2010. The evolution of eusociality. *Nature* 466:1057–1062. https://doi.org/10.1038/nature09205

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528. https://doi.org/10.1093/bioinformatics/bty633

Rehan SM. 2021. Genome architecture and social evolution. *Proc. Natl. Acad. Sci. U S A* 118:e2109409118. https://doi.org/10.1073/pnas.2109409118

Revell L. 2024. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things) [accessed 2024 May 14]. https://peerj.com/articles/16505/.

Rosenberg NA, Tao R. 2008. Discordance of species trees with their most likely gene trees: the case of five taxa. *Syst. Biol.* 57:131–140. https://doi.org/10.1080/10635150801905535

Rosengaus RB, Traniello JFA, Bulmer MS. 2011. Ecology, behavior and evolution of disease resistance in termites. In: Bignell DE, Roisin Y, Lo N, editors. *Biology of termites: a modern synthesis*. Dordrecht: Springer Netherlands. p. 165–191. https://doi.org/10.1007/978-90-481-3977-4_7

Schmucker D, Clemens JC, Shu H, et al. 2000. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101:671–684. https://doi.org/10.1016/S0092-8674(00)80878-8

Spielman SJ, Wilke CO. 2015. The relationship between dN/dS and scaled selection coefficients. *Mol. Biol. Evol.* 32:1097–1108. https://doi.org/10.1093/molbev/msv003

Tamura K, Stecher G, Kumar S. 2021. MEGA11: molecular evolutionary genetics analysis Version 11. *Mol. Biol. Evol.* 38:3022–3027. https://doi.org/10.1093/molbev/msab120

Terrapon N, Li C, Robertson HM, et al. 2014. Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* 5:3636. https://doi.org/10.1038/ncomms4636

Toth AL, Rehan SM. 2017. Molecular evolution of insect sociality: an eco-evo-devo perspective. *Annu. Rev. Entomol.* 62:419–442. https://doi.org/10.1146/annurev-ento-031616-035601

Warner MR, Qui L, Holmes MJ, et al. 2019. Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. Nat Commun10, 2651. https://doi.org/10.1038/s41467-019-10546-w

Wickham H. 2016. *ggplot2*. Cham: Springer International Publishing (Use R!). https://doi.org/10.1007/978-3-319-24277-4

Wilson EO, Hölldobler B. 2005. Eusociality: origin and consequences. *Proc. Natl. Acad. Sci. USA* 102:13367–13371. https://doi.org/10.1073/pnas.0505858102

Winter DJ. 2017. rentrez: an R package for the NCBI eUtils API. *R J* 9:520–526.

Wojtowicz WM, Flanagan JJ, Millard SS, et al. 2004. Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* 118:619–633. https://doi.org/10.1016/j.cell.2004.08.021

Woodard SH, Fischman BJ, Venkat A, et al. 2011. Genes involved in convergent evolution of eusociality in bees. *Proc. Natl. Acad. Sci. USA* 108:7472–7477. https://doi.org/10.1073/pnas.1103457108

Xu S, Li L, Luo X, et al. 2022. Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data. *Imeta.* 1:e56. https://doi.org/10.1002/imt2.56

Zhou X, Rokas A, Berger SL, et al. 2015. Chemoreceptor evolution in hymenoptera and its implications for the evolution of eusociality. *Genome Biol. Evol.* 7:2407–2416. https://doi.org/10.1093/gbe/evv149