UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

De Novo genome assembly and annotation of *X. darwini*: exploring the evolution of venom components and sociality in the Galapagos carpenter bee

Tesis de Maestría

Verónica Yolanda Baquero Méndez

María de Lourdes Torres (PhD)

Director de Trabajo de Titulación

Trabajo de titulación de posgrado presentado como requisito

para la obtención del título de Magíster en Ecología Tropical y Conservación

Quito, 9 de Mayo 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

De Novo genome assembly and annotation of *X. darwini*: exploring the evolution of venom components and sociality in the Galapagos carpenter bee

Verónica Yolanda Baquero Méndez

Nombre del Directora del Programa:
Elisa Bonaccorso
Título académico:
PhD in Ecology and Evolutionary Biology
Director del programa de:
Maestría en Ecología Tropical y Conservación

Nombre del Decano del colegio Académico:
Carlos A. Valle Castillo
Título académico:
PhD in Ecology and Evolutionary Biology
Decano del Colegio:
Ciencias Biológicas y Ambientales

Nombre del Decano del Colegio de Posgrados:
Darío Niebieskikwiat
Título académico:
Doctor en Física

Quito, Mayo 2025

© DERECHOS DE AUTOR[1]

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombre del estudiante:	Verónica Yolanda Baquero Méndez
Código de estudiante:	00335261
C.I. o Pasaporte:	1716132400
Lugar y fecha:	Quito, 9 de Mayo de 2025.

ACLARACIÓN PARA PUBLICACIÓN

UNPUBLISHED DOCUMENT

DEDICATORIA

A Valentina, Ramón y Killa, tres almas únicas que la vida me regaló. Con su amor y su luz, hicieron de este camino un viaje más dulce, más cálido y lleno de compañía incondicional.

AGRADECIMIENTOS

To María de Lourdes Torres, thesis advisor, for her constant support and guidance throughout this process.
To Brian Hollis and Sofía Carvajal, committee members, for their invaluable guidance, thoughtful comments, and constant motivation, which greatly enriched and strengthened this work.
To Elisa Bonaccorso, master's program director, for her support, understanding, and motivation throughout the program.
To María José Pozo, for her support during various stages of the project.
To Graham McLaughlin, for his collaboration on the analyses and his generous time.
To the Plant Biotechnology Laboratory at USFQ, for providing access to their facilities and technical support.
To the Galapagos Science Center, for access to their facilities and their constant collaboration.
To the Galapagos National Park and the Ministry of Environment, Water and Ecological Transition, for granting the necessary permits to carry out this work (MAATE-DBI-CM-2022-0268).
To the USFQ Research Dean's Office, for providing funding for the completion of this project.

RESUMEN

Xylocopa darwini, la única abeja endémica de las islas Galápagos, desempeña un papel ecológico vital como polinizador supergeneralista. A pesar de su importancia, poco se sabe sobre su composición genómica o su comportamiento social. Este estudio presenta el primer ensamblaje genómico y transcriptoma de alta calidad de *X. darwini*, proporcionando información relevante para comprender su ecología evolutiva. Utilizando datos de secuenciación PacBio long-read y RNA-seq de glándulas venenosas, hemos anotado genes relacionados con el veneno y evaluado sus trayectorias evolutivas. Los análisis comparativos con himenópteros solitarios y eusociales (avispas y abejas) revelaron que varios componentes del veneno (fosfatasa ácida, alérgeno 3 del veneno, carboxilesterasa 6 y serina proteasa) presentan patrones moleculares relacionados con el comportamiento social. Las reconstrucciones filogenéticas y del estado ancestral sugieren que los perfiles del veneno de *X. darwini* están más estrechamente alineados con los taxones eusociales, a pesar de su socialidad desconocida. Estos hallazgos destacan el veneno como una firma molecular de la evolución social y sugieren que *X. darwini* puede poseer rasgos sociales latentes o facultativos moldeados por presiones ecológicas. Este trabajo contribuye a nuestra comprensión de la evolución del veneno, el comportamiento social y la genómica de la conservación en especies insulares endémicas, y sienta las bases para futuros estudios funcionales y comparativos en todo el género *Xylocopa*.

Palabras clave: Islas Galápagos, comportamiento social, abeja carpintera endémica, ensamblaje del genoma, proteínas del veneno

ABSTRACT

Xylocopa darwini, the only endemic bee of the Galapagos Islands, plays a vital ecological role as a super-generalist pollinator. Despite its importance, little is known about its genomic makeup or social behavior. This study presents the first high-quality genome assembly and transcriptome of *X. darwini*, providing relevant information for understanding its evolutionary ecology. Using PacBio long-read sequencing and RNA-seq data from venom glands, we annotated venom-related genes and assessed their evolutionary trajectories. Comparative analyses with solitary and eusocial Hymenoptera (wasps and bees) revealed that several venom components (acid phosphatase, venom allergen 3, carboxylesterase 6, and serine protease) exhibit molecular patterns linked to social behavior. Phylogenetic and ancestral state reconstructions suggest that *X. darwini* venom profiles are more closely aligned with eusocial taxa, despite its unknown sociality. These findings highlight venom as a molecular signature of social evolution and suggest that *X. darwini* may possess latent or facultative social traits shaped by ecological pressures. This work contributes to our understanding of venom evolution, social behavior, and conservation genomics in endemic island species, and lays the groundwork for future functional and comparative studies across the genus *Xylocopa*.

Key words: Galapagos Islands, social behavior, endemic carpenter bee, genome assembly, venom proteins

TABLE OF CONTENTS

INTRODUCTION

In the Galapagos Islands, the carpenter bee *X. darwini* is the only endemic bee species in the archipelago (Linsley et al., 1966; Rasmussen et al., 2012). This species is believed to have reached the islands primarily via sea dispersal from South America, utilizing tree trunks for nesting structures capable of traveling by sea or on boats as driftwood (Chamorro et al., 2012; Vargas et al., 2015). Their capacity for inter-island immigration suggests an ability to cross sea barriers between islands (Chamorro et al., 2012) and the high migratory potential of the Galapagos carpenter bee is probably facilitated by active flight by adults (Vargas et al.,2015). Indeed, carpenter bees have often been observed flying several miles offshore from boats (Vargas et al., 2015). The successful colonization of the archipelago by *X. darwini* has been supported not only by these dispersal strategies but also by its role as a super generalist within the plant–pollinator networks of the islands (Chamorro et al., 2012; Traveset et al., 2013). *X. darwini* is recognized as a super pollinator, visiting at least 60 plant species (Linsley et al., 1966; Cockerell, 1935; Rasmussen et al., 2012), predominantly native and endemic flowers such as Scalesia affinis and Centratherum punctatum (McMullen, 1989; Linsley et al., 1966; McMullen, 1999; Rasmussen et al., 2012). This ecological versatility has contributed to its presence on nine of the twelve major islands (Vargas et al., 2015).

Due to its endemic status and ecological importance as a pollinator, *X. darwini* warrants significant conservation attention in the Galapagos Islands (Linsley, 1966; Chamorro et al., 2012; Vargas et al., 2015). However, introduced bird species such as the smooth-billed ani have been reported to prey on *X. darwini*, posing a potential threat to its populations (Schluter, 1986; Chamorro et al., 2012; Cooke et al., 2020). *X. darwini* is not currently listed on the IUCN Red List (Environment NSW, n.d.), however, the Galapagos

Conservation Trust classifies it as vulnerable species (Galapagos Conservation Trust, n.d.).

Genetic research on *X. darwini* remains limited, highlighting the need for deeper studies in this field. Vargas et al. (2015) reconstructed the colonization history of *X. darwini* using the mitochondrial COII gene. Their analysis revealed 12 distinct haplotypes among 118 individuals, indicating early colonization of most islands, followed by isolation, where limited gene flow over time led to the divergence of populations. This droves the formation of two main genetic groups: one in the eastern islands and another in the central-western islands. Although informative, this study relied on a single molecular marker, underscoring the need for whole-genome data.

A high-quality genome assembly and annotation of *X. darwini* could facilitate a comprehensive characterization of the species' genetic diversity, evolutionary history, and population structure which is critical for the evaluating of conservation strategies across the Galapagos archipelago. Furthermore, a well-annotated genome would facilitate the identification and characterization of functionally genes with important biological roles, such as those responsible for producing venom components, which could also contribute to broader scientific initiatives aimed at understanding how proteins evolved in species that have been underexplored (Sunagar et al., 2016).

Venom-related proteins offer an opportunity to explore evolutionary questions related to sociality. The rationale behind this approach lies in the established links between venom composition, ecological adaptations, and evolutionary trajectories in bees, especially within the genus *Xylocopa* (Sless & Rehan, 2023: Koludarov et al., 2023). Venom composition in Hymenoptera, including carpenter bees, has been shown to co-evolve with species' social behavior (Schmidt et al., 2014; Lee et al., 2016). The subfamily Xylocopinae exhibits a broad range of social organization, from solitary to eusocial, which is associated with distinct ecological and defensive pressures. Solitary bees generally use venom for individual defense, while eusocial species rely on venom to protect the colony, including brood and resources. These differences have led to the hypothesis that venom composition reflects adaptations to these varying lifestyles (Tan et al., 2020; Zhu et al., 2022;Shi et al., 2022; Koludarov et al., 2023).

Since bee and wasp venom components are thought to reflect evolutionary adaptations to ecological pressures, particularly sociality (Qiu et al., 2011; Lee et al., 2016; Yoon et al., 2020), it is important to study proteins that exhibit dual roles in both allergenicity and defense, as it is hypothesized that these proteins may diverge in their functions between solitary and eusocial lineages (Grunwald et al., 2006;Cardinal & Danforth.,2011; Choo et al., 2012; Lee et al., 2016). Acid phosphatase, is a neurotoxic venom component, is a major allergen in honeybee venom (*A. mellifera* and *Apis cerana*), known for inducing histamine release and allergic reactions (Grunwald et al., 2006). Its role appears conserved in these species (Kim & Jin, 2016; Hossen et al., 2017). Venom allergens 3 and 5 are highly expressed in social wasps and play key roles in defense and allergenicity. Phylogenetic analyses suggest that while venom allergen 5 originated in a common ancestor of Aculeata, it is absent from the venom gland transcriptomes of solitary wasps, indicating possible functional divergence aligned with sociality (Lee et al., 2016; Yoon et al., 2020). Carboxylesterase 6 (Api m 8), another honeybee allergen, exhibits high expression in social species like *Polistes varia* and bumblebees (*Bombus consobrinus*, *B. ussurensis*), but is nearly absent in solitary wasps, suggesting a defensive function specific to social taxa (Yoon et al., 2020). In *Bombus terrestris*, venom serine protease (Bt-VSP) enhances venom diffusion by degrading fibrinogen, a function largely absent in *Apis*, which retains only trace amounts of this enzyme and relies more on pain-inducing compounds for colony defense (Qiu et al., 2011; Lima & Brochetto-Braga, 2003). These differences imply functional shifts in venom composition depending on social behavior and ecological context (Choo et al., 2012; Lee et al., 2023).

This study explores the evolutionary divergence of venom-associated proteins—acid phosphatase, venom allergen 3, venom carboxylesterase 6, and venom serine protease—in *X. darwini*, comparing them with both solitary and eusocial species. Specifically, it hypothesizes the following:

1.Venom-associated proteins in solitary bee species differ in evolutionary trajectory from those in eusocial species, offering insights into the molecular evolution of *X. darwini*.

2.Venom components such as allergen 3, acid phosphatase, venom carboxylesterase 6, and venom serine have undergone differential selection pressures linked to social behavior in bees. If these proteins exhibit distinct evolutionary signatures between eusocial and solitary species, a comparative analysis involving *X. darwini* (a species with unknown sociality) could provide valuable insights into its behavioral ecology. By aligning *X. darwini* with lineages of known social or solitary behavior, it may be possible to identify adaptive selection patterns related to venom defense strategies, shedding light on the molecular-behavioral links in bee social organization.

By integrating genomic resources with comparative analyses of venom components, this study aims to shed light on the evolutionary dynamics of venom in carpenter bees particularly in relation to social behavior. Understanding how venom composition varies across solitary and eusocial lineages will provide valuable insights into the molecular adaptations that accompany shifts in sociality. For *X. darwini*, a key pollinator in the Galapagos whose behavior and social organization remain largely undocumented, this research offers a unique opportunity to infer aspects of its ecology through molecular signatures. Ultimately, this work contributes to a broader understanding of venom evolution in Xylocopinae could contribute to conservation efforts by deepening our knowledge of a critical endemic species in one of the world's most iconic archipelagos.

METHODS

Sample collection

The sampling was carried out under the permit MAATE-DBI-CM-2022-0268. The collection of the male carpenter bee used to obtain the genome was carried out at ‘La Galapaguera’, San Cristobal Island. Using a net, the male was captured and placed in a falcon tube and immediately placed in a cooler with ice packs. The sample was then transported to the Galapagos Science Center and was kept at -20°C until DNA extraction. For the analysis of RNA from the venom glands, three females were collected in San Cristobal near the Galapagos Science Center and captured following the same procedure as mentioned before.

DNA extraction and sequencing

The DNA was obtained from 35 mg of the abdomen of a male from “La Galapaguera”. DNA was extracted using the High Molecular Weight Extraction Kit (Omega). DNA quality and concentration was quantified using NanoDrop (Thermo Scientific), and agarose gel electrophoresis was performed to determine the integrity of DNA. Finally, the sample was sent to the Plant Biotechnology Laboratory of the University of San Francisco de Quito with the permit (067-2023 DNPG) and was stored at -20°C. The DNA was then sent to the University of South Carolina for Hi-Fi PacBio sequencing.

Genome assembly of *X. darwini*

With the reads obtained from PacBio sequencing, we first used the Jellyfish v2.3.0 pipeline to obtain the estimated size of the genome (Marçais and Kingsford 2011). In addition, we performed a de novo genome assembly using several assembly algorithms, including Mecat (Xiao et al. 2017), Verkko(Rautiainen et al. 2023), HiCanu(Nurk et al.2020), Hifiasm(Cheng et al.2024), and Flye. All assemblies were generated using default parameters. For the Flye assembly, we use the PacBio-hifi option, and the genome size obtained by the Jellyfish pipeline. Comparisons of quantitative and qualitative metrics corresponding to the obtained assemblies were performed to evaluate their quality, continuity and completeness with Quast v5.2.0 (Gurevich et al., 2013). We also use the Hymenoptera library of BUSCO v5.4.7 (Mosè et al., 2021), to assess the quality and completeness of genome assemblies, annotated gene sets. BUSCO bioinformatics tool that generates a report summarizing the completeness of your genome based on the presence and status of Hymenoptera orthologs. The output includes a summary file that provides the overall completeness score (percentage of complete orthologs), as well as detailed information about each ortholog (complete, duplicated, fragmented, or missing). A SnailPlot was generated using BlobTools v1 (Challis et al., 2020) to comprehensively analyze assembly quality metrics, including N50, L50, N90, L90, the total number of contigs, the size of the largest contig, and the BUSCO completeness score. In addition, we use the AGAT tool with the agat_sp_statistics.pl script (Dainat., SF) to have a detail of the statistics of the genome.

RNA extraction of venom glands from *X. darwini*

The venom glands were extracted from three female bees using blades and tweezers under a safety cabinet, with all materials cleaned using RNAzap and UV light. The glands, located in the lower abdomen, were removed using the stinger as a guide. Each gland was placed in a 1.5 ml Eppendorf tube containing 500 µL of RNA later to prevent degradation, and samples were stored at -20°C. The samples were transported in a cooler with ice packs from San Cristóbal to Quito and stored at -20°C at the Plant Biotechnology Laboratory at USFQ until RNA extraction, which was performed two days later.

RNA was extracted using a modified Trizol protocol (Invitrogen). Samples were lysed in 1 mL Trizol, mechanically disrupted with a micropestle, and incubated for 5 minutes. After adding 500 µL chloroform, the mixture was centrifuged at 16,000 x g for 15 minutes at 4°C. The aqueous RNA-containing phase was transferred to a new tube, precipitated with 500 µL isopropanol and 2 µL Glycoblue, and incubated at 4°C for 20 minutes with periodic mixing. Following centrifugation at 16,000 x g for 45 minutes at 4°C, the RNA pellet was washed twice with 75% ethanol, centrifuging at 16,000 x g for 5 minutes each time. The pellet was air-dried for 5 minutes and quantified using Qubit RNA HS and Nanodrop One.

RNA sequencing and genome annotation

An initial genome annotation was performed using the Augustus 3.5.0 pipeline in the OmicsBox software using the default specifications (Hoff et al. 2019) and as organism model B. terrestris. Due to the absence of *X. darwini* specific protein resources, homology evidence from the Apidae family (NCBI databases) was incorporated. To improve gene prediction, RNA-seq data from *X. darwini* were sequenced on the Illumina NovaSeq platform (paired-end 150 bp) and assembled into transcripts from the three individuals using Trinity (Haas et al. 2013). This *X. darwini* transcript assembly was then used as species-specific evidence in a refined annotation step. To address the lack of species-specific protein data for *X.darwini*, RNA-seq data from the same species was integrated as a critical step. This approach aimed to improve the accuracy of gene predictions, especially for venom-associated genes. To evaluate improvement of the annotation, a BUSCO analysis (Hymenoptera lineage library) was applied to both the initial and the refined annotations. A successful result for the genome and annotated genome is a completeness higher than 90% (Feron et al. 2022).

We also assembled and annotated genomes for X. virginica and C. australensis due to the lack of venom components data in the NCBI. For X. virginica, PacBio long-read sequencing data (SRX26426276) were assembled de novo using Flye. Genome quality was assessed using BUSCO (Hymenoptera dataset) and QUAST to evaluate completeness and contiguity. The genome of C. australensis (GCA_004307685.1) was annotated using Augustus 3.5.0 from the software OmicsBox (Hoff et al. 2019).

Functional annotation of the genomes of *X. darwini*, *X. violacea*, *X. virginica* and *C. australensis*

The protein FASTA file of genes of the carpenter bees (*X. darwini*, *X. violacea*, *X. virginica* and *C. australensis*) predicted by Augustus was functionally annotated using the Diamond BLASTp tool in OmicsBox focusing on matches within the genera *Xylocopa*, *Apis*, *Bombus*, and *Ceratina* (Altschul et al., 1990; Buchfink et al., 2021), with a focus on identifying venom-related proteins through homology searches against the NCBI non-redundant database. Particular attention was given to identifying venom-related proteins. Additionally, we utilized the annotated protein file from *X. violacea* (Koludarov et al., 2023) to perform Diamond BLASTp searches for further identification of venom-associated genes. Sequences of venom acid phosphatase, venom carboxylesterase 6, venom serine and venom allergen 3 from *Apis*, *Bombus*, and *C. calcarata* were retrieved from NCBI to expand the venom dataset. Notably, *Bombus* and

Apis, *X. virginica* and *C. australensis* represent eusocial species, whereas *X. violacea* and *C. calcarata* are solitary (Sless & Rehan, 2023). To assess structural and functional similarities between *X. darwini* venom proteins and those of related taxa, homology modeling was performed using SWISS-Model (Waterhouse et al., 2018). Amino acid sequences were aligned and modeled against the closest homologs reported in hymenopteran species. Proteins with significant sequence homology (>30% identity) to hymenopteran venom components were identified and reported in this study.

Phylogenetic tree and ancestral reconstruction of sociality

Venom protein sequences (acid phosphatase, venom allergen 3, carboxylesterase 6, and serine protease) from eusocial species (*Apis*, *Bombus*, *C. australensis*, *X. virginica*) and solitary species (*C. calcarata*, *X. violacea*) and unknown sociality (*X. darwini*) were aligned using MAFFT implemented in Geneious Prime 2025.03 (Biomatters Ltd.,2025). To ensure the comparability of the data, the isoform of each venom component was selected based on two criteria: first, its availability in NCBI, and second, its identification in the functional annotation reported in this study. This was done to prioritize consistency across species. This focused approach minimizes noise from isoform-specific variation (Ye et al., 2023).

The phylogenetic trees for each venom protein were reconstructed using Bayesian inference in BEAUti (Drummond et al., 2012), applying the LG + Γ (gamma-distributed rate variation) substitution model. For the ancestral reconstruction states the reconstruct state ancestor and reconstruct change count were selected and with a length chain of 10 million. Later, Beast v1.10.4 (Drummond et al., 2012), was used to create the trees. Tree annotator (Drummond et al., 2012), was used to select the best tree using the burning number of trees 1000, target tree type of maximum clade credibility tree, node heights with mean heights and a posterior probability of 0.95.

Ancestral state estimation for sociality was conducted using the fastAnc function from the phytools R package (Revell, 2024), in combination with the ape package (Paradis & Schliep., 2019). Sociality was coded as a continuous trait with eusocial species assigned a value of 1, solitary species a value of 0, and *X. darwini* as 0.5 to reflect its unknown or intermediate status. To visualize continuous trait evolution on the venom protein trees, we used the contMap function from phytools (Revell, 2024), which allowed the identification of shifts in venom composition associated with transitions in social behavior. Moreover, the ancestral value at each branch node was added using the nodelabels() function, based on the estimates obtained from fastAnc.

Analysis of venom components and sociality in *X. darwini*

For phylogenetically informed statistical modeling, we applied phylogenetic generalized least squares (PGLS) using the gls() function from the nlme package in R (Pinheiro et al., 2025). To account for phylogenetic non-independence, we modeled the error structure using a Brownian motion model of trait evolution (corBrownian) from the ape package (Paradis & Schliep., 2019). For this analysis the phylogenetic tree of the species from this study was generated using phyluce, a software package designed for the analysis of ultraconserved elements (UCEs) (Faircloth, 2016). This UCE-based species tree provided a robust phylogenetic framework for interpreting venom evolution in a comparative context. By leveraging this approach, we were able to disentangle the effects of social behavior from underlying phylogenetic signal in the evolution of venom-associated proteins.

To assess the evolutionary relationships of *X. darwini* in the context of venom gene variation, we conducted a principal component analysis (PCA) using aligned sequences of a single isoform for each venom protein (selected based on NCBI and functional annotation, as described earlier). For consistency, we used the same sequence alignment employed in the phylogenetic tree reconstruction, ensuring direct comparability between the PCA results and the tree topology. This PCA, performed in R, reduced sequence variation into major axes and allowed us to statistically evaluate the relationship between venom protein composition and sociality. The PCs used for plotting were selected based on their significance in the PGLS analysis that is, the axes where sociality had the strongest explanatory power (lowest p-values). Using these informative components, we also calculated which species was closest to *X. darwini* in multivariate space. Dashed lines were drawn to connect *X. darwini* to the closest species and to both eusocial and solitary centroids, helping to visualize its proximity to different sociality groups. Centroids for eusocial and solitary groups were calculated as the mean of the selected PC values for each group, and the euclidean distance between *X. darwini* and each centroid was computed to assess its molecular similarity to either social strategy. The use of centroids and distance calculations quantifies this relationship and supports the idea that venom gene evolution may be shaped by social structure, beyond phylogenetic relatedness alone. Plotting centroids is particularly valuable because individual species (represented as points) can overlap in PCA space, making broader patterns difficult to interpret. Centroids provide a clear reference for the typical venom composition of each social group, removing individual variation and highlighting overall trends. The resulting PCA plot thus offers a comparative framework for evaluating *X. darwini* molecular affinity, helping to infer its potential social behavior based on venom composition and its relative position among bees with known sociality (eusocial and solitary).

RESULTS

De novo genome assembly of *X. darwini*

Flye produce the most contiguous genome assembly for our study, archiving superior contiguity metrics (highest N50 and N90 values and lowest L50 and L90 values; Table 1) and the fewest contigs (433), resulting in a 215 Mb genome (Figure 1). To visualize Flye's assembly (selected for its optimal balance of contiguity and completeness), we generate a BlobToolkit Snailplot (Figure 1), which integrates contig size distribution (N50/N90) and sequence composition with BUSCO completeness scores. The radial plot highlights Flye strong performance: a large N50, minimal fragmentation (L90 = 36) and a BUSCO completeness score of 97.5%, with only 0,3% of fragmented and 2.2% missing values indicating a highly complete genome (Figure 1).

RNA-seq sequencing of venom glands

The RNA-seq of the venom glands displays a Q30 of 94-95% and an error rate of 0.01, which indicates a high sequencing accuracy. The assembly of the reads generated from the RNA-Seq of the three venom glands shows a BUSCO completeness score of 92.3%, indicating a high-quality assembly suitable for downstream analyses such as genome annotation. Of the expected genes, 87.8% are present as single-copy BUSCOs, and 5.8% appear as duplicates. A smaller portion of the genes (2.5%) are fragmented, while 5.4% are completely missing. The high proportion of complete BUSCOs strongly supports the completeness and reliability of the assembled transcript set.

The initial genome annotation, based on publicly available bee evidence from NCBI, identified 13,182 genes with a BUSCO completeness score of 84.6% (Supplementary Figure 1). In contrast, when incorporating the RNA-Seq assembly from *X. darwini* venom glands as additional evidence, we annotated 14,471 genes, achieving a higher BUSCO protein completeness score of 91.8% (Figure 1). This substantial improvement reflects the enhanced quality and completeness of the annotation. The inclusion of *X. darwini*-specific transcriptomic data (Figure 2) allowed the Augustus algorithm to generate more accurate gene models. Additionally, functional annotation using BLAST searches against the NCBI protein database identified six venom-related homologous proteins: acid phosphatase, venom allergen 3/5, venom carboxylesterase 6, venom dipeptidyl peptidase, venom serine protease, and icarapin. These findings further support the reliability of the gene predictions and highlight the biological relevance of the annotation.

De novo **genome assembly of** ***X. virginica*** **to support comparative venom analysis**

To enable downstream functional analysis of *X.darwini*, we generated a *De novo* genome assembly for *X. virginica*, a species with limited prior genomic resources and performed its annotation. Using the Flye assembler, we obtained a total assembly size of 368 Mb comprising 4,725 contigs, with a largest contig of 1.82 Mb and an N50 of 0.18 Mb (Supplementary Table 1). The assembly shows a high level of completeness, with 97% of Benchmarking Universal Single-Copy Orthologs (BUSCOs) detected, indicating strong representation of conserved genes. Assembly contiguity and completeness are further illustrated in the Snailplot (Supplementary Figure 2), providing a visual summary of scaffold size distribution and coverage. This genome resource was critical for improving ortholog genes assignment and comparative venom analysis, particularly given the absence of annotated proteins from *X. virginica* in public databases.

Genome annotation, venom genes and isoform comparison in three Xylocopinae

species: *X. darwini, X. virginica and C. australensis*

The total number of genes annotated in *X. darwini* was 14,471. When this number is compared to other members of the subfamily Xylocopinae, we see that *X. darwini* shows a higher number of genes than the eusocial *C. australensis* (8,269) and the solitary *X. violacea* (10,152), but fewer than the eusocial *X. virginica* (22,019) (Table 2).In terms of venom gene isoforms, *X. darwini* shows the highest diversity (16 isoforms) whereas solitary and eusocial relatives, such as *X. violacea* (15 isoforms), *X. virginica* (11), and *C. australensis* (8). Notably, all species share core venom components like acid phosphatases, serine proteases, and venom allergens, but differ in the number of isoforms per component (Table 2).

Comparative analysis of venom components in different bee species

To investigate the composition and evolutionary dynamics of venom in *X. darwini*, we performed a detailed analysis of four venom-associated proteins identified through genome annotation. For each gene, we conducted functional annotation. These components were selected based on known functions in other hymenopterans and their relevance to venom activity. Additionally, comparative analyses such as phylogenetic reconstruction and principal component analysis were used to explore patterns related to social behavior in different bee species.

Venom allergen 3

To explore how venom components may be linked to the evolution of social behavior, we reconstructed the ancestral states of *venom allergen 3* across a range of solitary and eusocial bee species. Figure 3a integrates phylogenetic relationships with sociality data to assess whether sequence patterns of this protein reflect transitions in social organization. Using character mapping, the resulting phylogenetic tree reveals a gradient of sociality-associated signals, allowing us to infer how *venom allergen 3* may have evolved in relation to solitary or eusocial lifestyles. Eusocial species such as *A. mellifera* and *B. flavifrons* cluster together with high ancestral values (~0.98), indicating a strongly eusocial origin for this clade. *C. australensis*, which is considered eusocial, shows an ancestral state value of 0.74, while the solitary *X. violacea* and *C. calcarata* exhibit lower values(~0.28–0.54). *X. darwini*, a species with unknow sociality, falls in a transitional position with an inferred ancestral value of ~0.61, suggesting its venom allergen 3 sequence is more similar to those of eusocial species than to solitary ones.

The principal component analysis (PCA) of amino acid variation in venom allergen 3 (Figure 3b) reveals distinct clustering of bee species based on sociality, with clear separation between eusocial and solitary groups along PC4 (11.74%) and PC6 (2.52%). *X. darwini*, marked as an unknown in terms of social behavior, is positioned closer to the eusocial centroid than to the solitary one, suggesting that its venom allergen 3 sequence is more similar to *B. flavifrons*. Together with the PGLS models, which take into account phylogenetic relatedness, a statistically significant association between sociality and venom allergen variation is confirmed ($p < 0.05$ Supplementary Table 2).

Venom acid phosphatase

The evolutionary analysis of venom acid phosphatase (Acph-1) (Figure 4a) reveals intermediate ancestral state values across the phylogeny, with most nodes ranging from 0.57 to 0.73. Notably, *X. darwini* exhibits an ancestral reconstruction value of approximately 0.73, placing it on the higher end of the eusociality scale being clustered with *X. virginica*. The PCA plot of Acph-1 (Figure 4b) sequence variation (PC4: 17.86%, PC6: 9.46%) further supports this observation: *X. darwini* is positioned closer to the eusocial species *X. virginica* and more distant from solitary taxa like *C. calcarata* and *X. violacea*.

Venom carboxylesterase 6

The evolutionary analysis of the venom carboxylesterase 6 (Figure 5a) shows a range of intermediate ancestral state values, with most nodes centered around 0.5, indicating ambiguity in the sociality signal. *X. darwini* is reconstructed with a value of 0.5, the same as several other species, placing it in an intermediate position between solitary and eusocial tendencies. The PCA plot (Figure 5b) variation (PC4: 2.31%, PC6: 1.21%) shows a less pronounced clustering pattern compared to other venom components, but *X. darwini* remains closer to the eusocial species, particularly *A. mellifera*, than to the solitary group. Although the variation explained by PC4 and PC6 is limited, shows consistent proximity of *X. darwini* to eusocial species.

Venom serine protease

The ancestral state reconstruction for the venom serine protease component (Figure 6a) indicates that *X. darwini* has an intermediate reconstructed value of 0.61, again suggesting partial similarity to eusocial species in this venom component. The PCA plot (Figure 6b) (PC2: 19.53%, PC6: 2.01%) further supports this, showing *X. darwini* positioned near eusocial taxa like *A. mellifera*, rather than with the solitary cluster. Although PC6 explains limited variation, the strong signal in PC2 captures clear sociality-related structure in the data. These findings reinforce the pattern observed across other venom components: *X. darwini* consistently exhibits molecular features more aligned with eusocial species.

DISCUSSION

Genome assembly, annotation, and functional analysis

Both *X. darwini* and *X. virginica* were sequenced using PacBio long-read technology, but their assembly results are different. The *X. darwini* assembly (215 Mb) was more contiguous and compact (evidenced by its thick dark orange ring in (Figure 1), with high BUSCO completeness (97.5%), a low duplication rate (0.1%), and fewer contigs (Figure 1). In contrast, *X. virginica* (Supplementary Figure 2) had a substantially larger and more fragmented genome (369 Mb), with a lower N50 supported by its thinner dark orange ring and a higher duplication rate (16.4%) (Supplementary Figure 2), suggesting expanded repeat content, or structural complexity (Nash et al., 2024). While Flye is among the top-performing assemblers for long-read data capable of reconstructing even complex segmental duplications, extensive repeat expansion, may have challenged Flye repeat-collapsing capabilities (Kolmogorov et al., 2019). This comparison highlights how species-specific genome architecture and sequence complexity can shape assembly performance, even under similar sequencing and assembly protocols (Rhie et al., 2021; Meng et al., 2022).

When comparing *X. darwini* with *X. violacea*, a related species with a larger genome (1 Gb, 1301 scaffolds) (Nash et al., 2024), the differences in sequencing strategies and repetitive content become even more apparent. Unlike *X. darwini*, which relied solely on PacBio, the *X. violacea* genome was assembled using a combination of Hi-C and Iso-Seq data. Despite a similar BUSCO completeness (96.5%) and low duplication rate (0.6%), the expanded genome size of *X. violacea* underscores the role of repetitive elements in driving structural variance within *Xylocopa* (Nash et al., 2024). Over 80% of its genome consists of repetitive sequences especially unclassified and satellite repeats consistent with the presence of pseudo-acrocentric chromosomes and heterochromatin-rich regions commonly found in bees (Hoshiba & Imai, 1993; Nash et al., 2024). Together, these comparisons highlight how assembly outcomes reflect both algorithmic limitations (e.g., Flye handling of repeats) and biological factors (e.g., lineage-specific repeat expansions) (Rhie et al., 2021; Meng et al., 2022).

In this study, the genomes of *X. darwini* and *X. virginica*, which represent the second and third genomes reported for the genus *Xylocopa*, after *X. violacea* (Nash et al., 2024), were assembled and represent key resources for understanding venom evolution and sociality in carpenter bees. No genomic information or protein data was available for *X. darwini* prior to this research. Thanks to the venom gland RNA-seq data generated in this investigation improved gene prediction using Augustus, allowing accurate annotation of venom-related genes by BLASTp searches of NCBI protein databases (Table 2; Sunagar et al., 2016). In addition, genome assembly and venom gene data were also lacking for *X. virginica*, limiting comparative analyses; the genome generated here now provides a basis for studying venom components in a eusocial *Xylocopa* species (Sunagar et al., 2016; Koludarov et al., 2023).

To further support this comparative framework, we conducted functional annotation of the previously published genomes of *X. violacea* and *C. australensis* (Table 2), which had no venom genes reported in the NCBI prior to this study. The inclusion of these species, which represent solitary and eusocial lineages respectively, enabled a more refined analysis of venom gene diversity in the subfamily Xylocopinae.

As shown in Table 2, venom gene isoform counts vary between species, with *X. darwini* showing the highest number (16 isoforms), followed by *X. violacea* (15), *X. virginica* (11), and *C. australensis* (8). These differences may reflect both assembly quality and lineage-specific variation in venom repertoires (Sunagar et al., 2016; Ye et al., 2023). The consistent presence of gene families such as acid phosphatases, venom allergens, serine proteases, and dipeptidyl peptidases across species highlights conserved components, while the variation in isoform number may reflect adaptations linked to sociality or ecological niches (Koludarov et al., 2023; Ye et al., 2023).

Venom Proteins: Phylogenetic Signals and Functional Divergence

Venom composition is thought to co-evolve with social behavior in bees, as different levels of sociality impose distinct defensive demands (Shi et al., 2022; Koludarov et al., 2023). If venom genes have co-evolved with sociality, one would expect this to be reflected in their phylogeny; conversely, a discrepancy would suggest neutrality or adaptation to other ecological pressures (Sless & Rehan, 2023). This evolutionary dynamic is particularly evident in carpenter bees, a group in which social behavior has evolved multiple times independently from solitary ancestor, with several reversals back to solitary life (von Reumont et al., 2022;Koludarov et al., 2023;Sless & Rehan, 2023).Understanding how venom gene evolution tracks these shifts in sociality could reveal whether venom adaptations are shaped primarily by social behavior or other ecological factors (Drukewitz & von Reumont, 2019; Schendel et al., 2019;Koludarov et al., 2022).

The phylogenetic tree of venom allergen 3 (Figure 3a) shows a major cluster composed of two subclusters: one containing *Xylocopa* species and the other *Ceratina* species, while *Apis* and *Bombus* form a separate lineage. This topology captures the broad separation between major lineages such as *Apis*/*Bombus* versus *Xylocopa*/*Ceratina* but differs in the specific relationships among lineages, particularly within the *Xylocopa* and *Ceratina* clade. For example, the species tree (Supplementary Figure 3) places *Ceratina* more distantly related to *Xylocopa*, whereas the venom gene tree of venom allergen 3 shows *Ceratina* clustering more closely with *Xylocopa*, suggesting differences from the expected species relationships. Given the close evolutionary relationship between bees and wasps and venom allergen 5 having similar roles like antimicrobial activity or prey-specific toxicity to venom allergen 3 in bees (Brock et al., 2017; von Reumont et al., 2022), the comparison between wasp allergen 5 and venom allergen 3 in bees is relevant, where a similar discordance has been observed. Despite high amino acid conservation, the phylogenetic topology is different from the species tree (Yoon et al., 2020).

The ancestral state reconstruction of venom allergen 3 further shows that eusocial species such as *A. mellifera* and *B. flavifrons* exhibit high ancestral values (Figure 3a), reflecting strong retention of venom traits associated with social defense (Schmidt, 2014). In contrast, solitary species like *X. violacea*, *X. darwini* and *C. calcarata* display lower values (Figure 3a). While the eusocial *X. virginica* and *C. australensis* displays a moderate value (Figure 3a).These values in carpenter bees reveal that lineages retain ancestral venom features, potentially serving as molecular preadaptations for shifts in social behavior (Sless & Rehan, 2023). The distribution of values across species highlights the dynamic interplay between phylogenetic history and ecological adaptation Koludarov et al., 2022).

To further investigate the dynamic interplay between phylogenetic history and sociality in venom allergen 3 we performed a PCA using amino acid sequence variation in this component 3 and tested for associations with sociality using phylogenetic generalized least squares (PGLS), which incorporated the species tree to control the phylogenetic signal. Interestingly, the PCA (Figure 3b) showed that *X. darwini* clusters closely with *B. flavifrons*, a eusocial species, with a Euclidean distance of 0.75. This suggests that *X.*

darwini could share molecular similarities with eusocial species in some dimensions of venom allergen 3 variation (Chang et al., 2015; Yoon et al., 2020).

Additionally, the PGLS from venom allergen 3 revealed that PC4 is significantly associated with solitary species, and PC6 is significantly associated with unknown sociality of *X. darwini (Supplementary Table 2)*. These p-values indicate that variation along PC4 is consistently structured by solitary behavior across the phylogeny, while variation along PC6 specifically captures a pattern unique to *X. darwini*. This distinction is important: although *X. darwini* is close to *Bombus flavifrons* in the reduced PCA space of venom allergen 3, the significant PGLS results reveal a distinct evolutionary signal in *X. darwini* that aligns neither strictly with solitary nor eusocial patterns, but instead with a lineage-specific trajectory of this species.

These findings underscore a critical insight: while the ancestral reconstruction reflects deep phylogenetic similarity between *X. darwini* and solitary species (as indicated by the low ancestral value), the PCA focused on molecular variation reveals a more complex picture. The proximity to *Bombus flavifrons* in the PCA space may reflect convergent evolution or functional similarity, potentially hinting at a shift toward eusocial features. Thus, the combined analyses suggest that *X. darwini* may be undergoing molecular changes in venom allergen 3 that distance it from its solitary ancestry, reflecting flexibility that may facilitate rapid behavioral transitions in response to ecological pressures (Sless & Rehan, 2023).

Homology modeling using SWISS-Model (Waterhouse et al., 2018) revealed that, among *X. darwini* venom proteins, only venom allergen 3 exhibited sufficient sequence identity (>30%) with the model structures of the hymenopteran taxa (Supplementary Figure 4). Notably, its closest structural homolog was found in *Vespula vulgaris,* a well-characterized eusocial species (Harrop et al., 2020). This alignment underscores both conserved sequence motifs and structural domains suggesting potential functional retention of ancestral defensive or immune-related roles that are often associated with social behavior (Breed et al., 2004; Wan et al., 2014).

Acid phosphatase (Figure 4) provide another compelling example of how venom composition could evolve under social ecological pressures, independent of strict phylogenetic constraints. In the phylogenetic tree of this component (Figure 4a), *A. mellifera* forms a distinct branch, while *B. flavifrons* and *C. calcarata* cluster together with intermediate ancestral values (Figure 4a). From within this subcluster *X. darwini* and the eusocial *X. virginica* group closely and both exhibit higher ancestral values, whereas *C. australensis* and *X. violacea*, appear on more distantly related branches. This phylogenetic pattern aligns with the PCA results for acid phosphatase (Figure 4b) using significant principal components identified by PGLS (Supplementary Table 2). *X. darwini* is positioned very closely to *X. virginica*, with a Euclidean distance of 0.12, reinforcing its similarity in amino acid sequence with a known eusocial species. While the PGLS analysis (Supplementary Table 2) identifies significant associations in PC4 and solitary species, PC6 captures an independent trajectory for *X. darwini*. These significant p-values reflect that variation captured in these axes is influenced both by solitary ancestry and the unknown position of *X. darwini*, suggesting that while it clusters with eusocial species in PCA space, its venom profile still retains evolutionary signals linked to solitary behavior. This may suggest that acid phosphatase in *X. darwini* is evolving towards a similar profile to eusocial species (Yoon et al., 2020). Venom acid phosphatase in eusocial species such as *Bombus*, is similar to *Apis* and is capable of causing tissue damage (Schmidt, 2014). Conversely, solitary bees are known to repurpose acid phosphatases for predation or niche-specific threats (von Reumont et al., 2022). This highlights how sociality can reshape venom function in two distinct ways: convergent evolution (e.g., tissue-damaging acid phosphatase in *Apis* and *Bombus*) (Yoon et al., 2020), and through divergent evolution, where closely related species adapt their venom to different ecological roles, such as solitary bees repurposing acid phosphatases for predation or niche-specific defense (Schmidt, 2014: von Reumont et al., 2022).

Another venom component that underlines the evolutionary distinction in the context of sociality is carboxylesterase 6 (Figure 5). The phylogenetic divergence observed in Figure 5a underscores a potential link between eusociality and ancestral venom composition. The high ancestral value in the *A. mellifera* and *B. flavifrons* cluster aligns with their highly derived eusocial lifestyles (Figure 5a). In contrast, the heterogeneous clustering of carpenter bee species, particularly the intermediate ancestral value in *X. darwini* (Figure 5a) and its relatives, raises questions about transitional evolutionary states. This pattern could reflect a dynamic interplay between ancestral traits and emerging adaptations in species with less rigid social hierarchies (Koludarov et al., 2023; Sless & Rehan, 2023). Further, the PCA based on principal components (Figure 5b) significantly associated with sociality in the PGLS analysis (PC4 and PC6) (Supplementary Table 2) reveals that *X. darwini* is most closely positioned to *A. mellifera*, a eusocial species, with a Euclidean distance of 0.77 (Supplementary Table 2). The contrast between tree-based ancestral reconstruction and the PCA-PGLS analysis reveals different evolutionary path of carboxylesterase 6. While the tree places *X. darwini* in an intermediate position suggesting shared ancestry with both solitary and eusocial taxa (Koludarov et al., 2023; von Reumont et al., 2022), the PCA indicates that the carboxylesterase 6 maybe shifting toward a profile more similar to eusocial lineages (Argiolas et al., 1985; Yoon et al., 2020; Shi et al., 2022; Sless & Rehan, 2023).

Building on the patterns observed in carboxylesterase 6, venom serine protease (Figure 6) also reveals a trajectory of divergence shaped by social behavior. The phylogenetic tree (Figure 6a) places *X. darwini* on a distinct branch within a cluster of mainly eusocial taxa (*A. mellifera*, *B. flavifrons* and *X. virginica*), with an ancestral value (0.64). In particular, *X. darwini* diverges from *A. mellifera* and *Bombus* (Figure 6a), although it falls within the broader clade that includes these eusocial lineages, supporting potential historical links to social traits (von Reumont et al., 2022;Koludarov et al., 2023; Sless & Rehan, 2023). In the PGLS, PC2 shows a significant association with the 'unknown' category corresponding to *X. darwini* while PC6 is associated with solitary species (Supplementary Table 2), again highlighting a dual signal of eusocial and solitary traits in this species. The PCA (Figure 6b) places *X. darwini* closest to *A. mellifera*, with a Euclidean distance of 2.24. This could suggest some molecular convergence, as functional differences in serine proteases in *B.terrestris* have been shown to degrade fibrinogen, which prevents clot formation, enhancing venom spread in vertebrates (Qiu et al., 2011) while in *Apis* serine proteases are present in trace amounts and lack fibrinolytic activity possibly favoring pain-inducing amines for colony defense (Lima & Brochetto-Braga, 2003; Qiu et al., 2011; Choo et al., 2012;Lee et al., 2023). However, *Apis* serine protease exhibit antimicrobial activity (Wan et al., 2014; Kim et al., 2013b), a characteristic of eusocialiality (Dashevsky et al., 2023). Thus, *X. darwini* molecular similarity to *A. mellifera* and its phylogenetic position may reflect adaptative shifts tied to eusocial behavior. This again supports the idea that venom evolution can lead to social traits, consistent with the hypothesis that venom serves as a flexible toolkit in the evolution of sociality (Choo et al., 2012; Schendel et al., 2019).

Despite revealing intriguing patterns of molecular convergence and sociality-linked divergence, these analyses of venom components face some limitations due to lack of studies on the social behavior of *X. darwini* which complicates the interpretation of its molecular proximity to eusocial taxa. Additionally, PCA and PGLS rely on reduced dimensions and statistical associations that may overlook complex interactions between genes, behavior, and ecology (Yoon et al., 2020; Guido-Patiño & Plisson, 2022).

Phylogenetic reconstructions provide valuable insights into ancestral states but may mask more recent adaptive changes (Sless & Rehan, 2023). Moreover, many functional annotations rely on homologous proteins characterized in well studied taxa like *Bombus* and *Apis*, while sequences for these venom components are lacking in databases for species such as *C. australensis*, *X. virginica*, *X. violacea*, and *X. darwini*. As a result, the venom proteins studied here only partially reflect the phylogenetic relationships of these species. This is in part because other venom proteins that could provide additional evolutionary insights were not included in the analysis due to limited sequence availability or incomplete annotations in these less well-characterized taxa. Moreover, the modeling of venom proteins via SWISS-model (Waterhouse et al., 2018) for comparative analysis is limited due to the lack of sufficient structurally characterized templates from solitary or less-studied lineages. The gap in hymenopteran structural databases, underscore the importance of integrating ecological, behavioral, and molecular approaches to strengthen comparative interpretations (Sunagar et al., 2016).

Tracing sociality in *X. darwini*

Defense is a central driver of caste evolution and collective behavior in social insects (Breed et al., 2004; Baracchi et al., 2011). Survival strategies reflect coordinated nest and territory protection, with defenders employing threat-responsive tactics (Qiu et al., 2011; Choo et al., 2012;Lee et al., 2023). However, defense is energetically costly and risky, leading species to adopt context-dependent strategies (Sunagar et al., 2016; Schendel et al., 2019; Abbot., 2022). This cost-benefit balance likely served as a critical ecological precondition for the transition from solitary to group living. Ancestral insects that separated offspring in fortified nests, providing progressive provisioning and extended parental care, established foundational traits for eusociality. Nest security, sustained resource investment, and intergenerational care thus emerged as precursors to complex social systems (Buchmann & Minckley., 2019;Abbot., 2022).

Carpenter bees (*Xylocopa*) illustrate how ecological pressures shape social flexibility (Sless & Rehan, 2023; Koludarov et al., 2023). Their labor-intensive excavation of nests in hard wood creates durable but vulnerable resources, targeted by predators such as ants, woodpeckers, and parasitoids (Buchmann & Minckley., 2019). Foundresses exhibit remarkable longevity (up to 3 years), facilitating prolonged interactions with offspring a trait that, along with sibling assemblies, acts as a preadaptation buffering the costs of social experimentation (Michener, 1990; Buchmann & Minckley., 2019). Nests consist of linear tunnels partitioned into larval cells, each provisioned with pollen and nectar and protected by large, resilient eggs. Nest durability, influenced by wood type, further modulates opportunities for social dynamics (Buchmann & Minckley., 2019; Sless & Rehan., 2023).

The transitional signals in *X. darwini* venom components intermediate ancestral values yet molecular proximity to eusocial taxa reinforce the dynamic origins of sociality proposed by Sless & Rehan (2023). While phylogenetically constrained in some components (venom allergen 3 and carboxylesterase 6), its venom shows molecular proximity to eusocial taxa in all venom components suggesting evolution patterns probably driven by ecological roles like defense and nest sanitation rather than by deep ancestry (Schendel et al., 2019; Sunagar et al., 2016). This molecular plasticity aligns with facultative sociality, enabling repeated shifts between solitary and cooperative strategies. For *X. darwini*, such traits may reveal ancestral retention or convergent evolution under colony-defense pressures (Schmidt, 2014; Yoon et al., 2020), reflecting the behavioral flexibility seen in other carpenter bees (Buchmann & Minckley., 2019).

Notably, while obligate eusocial groups such as *Apis* may cross a "point of no return", facultative taxa such as *Xylocopa* retain adaptive versatility (Wilson & Hölldobler, 2005; Buchmann & Minckley., 2019; Sless & Rehan., 2023). The venom profiles and social behaviors of these taxa highlight a dynamic interplay: ecological pressures (e.g. predation, nest site scarcity) select for molecular and behavioral traits, which in turn modulate the costs and benefits of group living. This dynamic interplay provides a potential explanation for the coexistence of both solitary and social nesting behaviors within a population, as individuals evaluate and trade-off factors such as resource investment, kinship, and predation risk (Buchmann & Minckley, 2019).

Xylocopa genus exemplifies how sociality can emerge from the dynamic interplay between phylogenetic legacy and ecological opportunity (Buchmann & Minckley, 2019;Sless & Rehan., 2023. In the Galapagos Islands, the endemic carpenter bee, *X. darwini,* the archipelago's sole bee species faces unique biogeographical pressures (Linsley et al., 1966; Rasmussen et al., 2012; Chamorro et al., 2012; Vargas et al., 2015), and the inter-island migration of this carpenter bee may reflect local adaptation in the islands (Chamorro et al., 2012; Vargas et al., 2015). To understand whether the observed venom signatures in *X. darwini* reflect convergent evolution due to shared ecological pressures or independent evolutionary trajectories, future work should incorporate functional assays of enzymatic activity tests to assess the physiological roles of specific venom components with behavioral ecology studies. Additionally, analyzing specific amino acid sequences and peptides in venom evolution is crucial for understanding how ecological pressures (e.g., eusociality vs. solitary lifestyles) drive biochemical adaptations (Lee et al., 2016). For example, novel peptides in carpenter bee venom may reflect solitary-specific innovations, while a dominant toxin that diverges between eusocial (honeybees, bumblebees) and solitary bees, may suggst sociality-driven selective pressures (Lee et al., 2016; von Reumont et al., 2022).

Functional differences in venom composition, such as toxin synergy and structural peptide optimization, underscore how subtle molecular variations enhance ecological efficacy (Lee et al., 2016; Schendel et al., 2019). However, disparities in isoform diversity seen, for example, in the understudied carpenter bee's unique peptides compared to the well-documented toxins of honeybees may reflect biases in research focus, uneven data quality, or limited proteomic/genomic resources for non-model species (Shi et al., 2022; Ye et al., 2023). This gap highlights the critical need to prioritize understudied species, as their venoms may hold unrecognized evolutionary innovations. By integrating proteomic, genomic, and functional analyses, we can unravel how specific peptide regions and interactions drive venom evolution, ultimately connecting sequence-level adaptations to broader ecological and behavioral shifts such as those differentiating solitary and eusocial lifestyles (Sunagar et al., 2016).This is particularly critical given *X. darwini* role as a super-generalist pollinator, visiting over 60 plant species a niche range that may favor flexible social strategies to exploit dispersed floral resources (Linsley et al., 1966; Rasmussen et al., 2012; Chamorro et al., 2012).

Additionally, comparative genomic analyses across both solitary and social *Xylocopa* species will be crucial for identifying lineage-specific adaptations versus conserved molecular features (Koludarov et al., 2023; Sless & Rehan, 2023). For *X. darwini*, such studies could clarify how island isolation and novel predation pressures (e.g., from invasive smooth-billed anis) interact with preadaptations like longevity and nest-site fidelity to drive sociality (Cooke et al., 2020). This may connect molecular function to colony-level strategies, reinforcing that social evolution is not a linear process, but a context-dependent mosaic of adaptations (Buchmann & Minckley, 2019; Sless & Rehan, 2023).

As a keystone pollinator endemic to the Galapagos Islands, *X. darwini* plays a vital role in maintaining the archipelago's unique ecosystems, which are shaped by extreme isolation and adaptive radiation (Linsley et al., 1966; Cockerell, 1935; Rasmussen et al., 2012). Despite its ecological importance, *X. darwini* remains understudied, with limited genomic or behavioral data (Linsley et al., 1966; Cockerell, 1935; Rasmussen et al., 2012;Vargas et al.2015). The genome assembly reported in this study fills a critical gap, enabling researchers to investigate conservation genetics, given the vulnerability of island species to climate change and invasive competitors (Cockerell, 1926; McMullen.,1989;Galapagos Conservation Trust, s.f). The combined use of phylogenetic reconstruction and PCA-PGLS reveals how different analytical approaches uncover complementary aspects of venom evolution, emphasizing the importance of integrating molecular, functional, and ecological data to interpret adaptive shifts in social behavior (Sunagar et al., 2016). In this context, the case of *X. darwini* illustrates how social plasticity may reflect evolutionary responses to resource-limited island environments (Cockerell, 1926; Chamorro et al., 2012; Vargas et al., 2015; Chang et al., 2015). As such, the *X. darwini* genome not only provides a foundation to investigate the molecular basis of venom evolution but also bridges the gap between evolutionary genomics and conservation efforts in one of the world's most iconic ecosystems.

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to the Plant Biotechnology Laboratory at USFQ for providing access to their facilities and offering invaluable technical support throughout this project. We also thank the Galapagos Science Center for their continuous collaboration and the use of their resources. Our deep appreciation goes to the Galapagos National Park and the Ministry of Environment, Water and Ecological Transition for granting the necessary permits (MAATE-DBI-CM-2022-0268) to carry out this work. Additionally, we are grateful to the USFQ Research Dean’s Office for their financial support, which made the completion of this project possible.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Argiolas, A., & Pisano, J. J. (1985). Bombolitins, a new class of mast cell degranulating peptides from the venom of the bumblebee *Megabombus pennsylvanicus*. *Journal of Biological Chemistry, 260*(3), 1437–1444.

Biomatters Ltd. (2024). *Geneious Prime 2025.0* [Computer software]. Auckland, New Zealand. https://www.geneious.com

Branstetter, M. G., Childers, A. K., Cox-Foster, D., Hopper, K. R., Kapheim, K. M., Toth, A. L., Worley, K. C., 2018. Genomes of the Hymenoptera. Current Opinion in Insect Science.25, 65-75. https://doi.org/10.1016/j.cois.2017.11.008

Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution, 57*(4), 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Buchfink, B., Reuter, K. & Drost, HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods 18, 366–368 (2021). https://doi.org/10.1038/s41592-021-01101-x

Buchmann, S. L., & Minckley, R. L. (2019). Large Carpenter Bees (*Xylocopa*). In C. Starr (Ed.), *Encyclopedia of Social Insects* (pp. 1–4). Springer International Publishing. https://doi.org/10.1007/978-3-319-90306-4_70-1

Cardinal S, Danforth BN (2011) The Antiquity and Evolutionary History of Social Behavior in Bees. PLoS ONE 6(6): e21086. https://doi.org/10.1371/journal.pone.0021086

Chamorro, S., Heleno, R., Olesen, J. M., McMullen, C. K., & Traveset, A., 2012. Pollination patterns and plant breeding systems in the Galapagos: a review. Annals of Botany, 110(7), 1489–1501. https://doi.org/10.1093/aob/mcs132

Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit: Interactive quality assessment of genome assemblies. G3: Genes, Genomes, Genetics, 10(4), 1361–1374. https://doi.org/10.1534/g3.119.400908

Chang, D., Olenzek, A. M., & Duda, T. F. (2015). Effects of geographical heterogeneity in species interactions on the evolution of venom genes. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1805), 20141984. https://doi.org/10.1098/rspb.2014.1984

Charif D, Lobry J (2007). “SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis.” In Bastolla U, Porto M, Roman H, Vendruscolo M (eds.), Structural approaches to sequence evolution: Molecules, networks, populations, series Biological and Medical Physics, Biomedical Engineering, 207-232. Springer Verlag, New York. ISBN : 978-3-540-35305-8.

Cheng, H., Asri, M., Lucas, J., Koren, S., Li, H. (2024) Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. Nat Methods, 21:967-970. https://doi.org/10.1038/s41592-024-02269-8

Choo, Y. M., Lee, K. S., Yoon, H. J., Qiu, Y., Wan, H., Sohn, M. R., Sohn, H. D., & Jin, B. R. (2012). Antifibrinolytic Role of a Bee Venom Serine Protease Inhibitor That Acts as a Plasmin Inhibitor. *PLOS ONE*, *7*(2), e32269. https://doi.org/10.1371/journal.pone.0032269

Cocolin, L., & Rantsiou, K. (2014). Molecular biology | Transcriptomics. In C. A. Batt & M. L. Tortorello (Eds.), *Encyclopedia of food microbiology* (2nd ed., pp. 803-807). Academic Press. https://doi.org/10.1016/B978-0-12-384730-0.00436-5

Cooke, S. C., Anchundia, D., Caton, E., Haskell, L. E., Jäger, H., Kalki, Y., Mollá, Ó., Rodríguez, J., Schramer, T. D., Walentowitz, A., & Fessl, B., 2020. Endemic species predation by the introduced smooth-billed ani in Galapagos. Biological Invasions, 22(7), 2113-2120. https://doi.org/10.1007/s10530-020-02251-3

Cockerell, T. D. A. 1926. Descriptions and records of bees.- CXI. Ann. Mag. Nat. Hist, (9) 17:657-665

Condamine, F. L., & Hines, H. M. (2015). Historical species losses in bumblebee evolution. *Biology Letters*, *11*(3), 20141049. https://doi.org/10.1098/rsbl.2014.1049

Crowley, L. M., Sivell, O., & Sivell, D. (2023). The genome sequence of the Buff-tailed Bumblebee, *Bombus terrestris* (Linnaeus, 1758). *Wellcome Open Research*, *8*, 161. https://doi.org/10.12688/wellcomeopenres.19248.1

Dainat, J. (2019). *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format* (Version v0.7.0) [Software]. Zenodo. https://doi.org/10.5281/zenodo.3552717

Dashevsky, D., Baumann, K., Undheim, E. A. B., Nouwens, A., Ikonomopoulou, M. P., Schmidt, J. O., Ge, L., Kwok, H. F., Rodriguez, J., & Fry, B. G. (2023). Functional and Proteomic Insights into Aculeata Venoms. *Toxins*, *15*(3), Article 3. https://doi.org/10.3390/toxins15030224

Drukewitz, S. H., & von Reumont, B. M. (2019). The Significance of Comparative Genomics in Modern Evolutionary Venomics. *Frontiers in Ecology and Evolution*, *7*. https://www.frontiersin.org/articles/10.3389/fevo.2019.00163

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution, 29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Environment NSW. (n.d.). Data deficient species not listed. NSW Government. Retrieved from https://www.environment.nsw.gov.au/topics/animals-and-plants/threatened-species/nsw-threatened-species-scientific-committee/nsw-threatened-species-scientific-committee-publications/assessment-reports/data-deficient-species-not-listed

Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics, 32*(5), 786–788. https://doi.org/10.1093/bioinformatics/btv646

Galapagos Conservation Trust. (s.f.). *Galapagos carpenter bee*. https://galapagosconservation.org.uk/species/galapagos-carpenter-bee/

Geneious Prime 2025.03. (2025). Biomatters Ltd. https://www.geneious.com

Grunwald, T., Bockisch, B., Spillner, E., Ring, J., Bredehorst, R., & Ollert, M. W. (2006). Molecular cloning and expression in insect cells of honeybee venom allergen acid phosphatase (Api m 3). *Journal of Allergy and Clinical Immunology*, *117*(4), 848–854. https://doi.org/10.1016/j.jaci.2005.12.1331

Guido-Patiño, J. C., & Plisson, F. (2022). Profiling hymenopteran venom toxins: Protein families, structural landscape, biological activities, and pharmacological benefits. *Toxicon: X*, *14*, 100119. https://doi.org/10.1016/j.toxcx.2022.100119

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., Macmanes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., Leduc, R. D., Friedman, N., & Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols, 8(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Harrop, T. W. R., Guhlin, J., McLaughlin, G. M., Permina, E., Stockwell, P., Gilligan, J., Le Lec, M. F., Gruber, M. A. M., Quinn, O., Lovegrove, M., Duncan, E. J., Remnant, E. J., Van Eeckhoven, J., Graham, B., Knapp, R. A., Langford, K. W., Kronenberg, Z., Press, M. O., Eacker, S. M., … Dearden, P. K. (2020). High-Quality Assemblies for Three Invasive Social Wasps from the Vespula Genus. *G3: Genes|Genomes|Genetics*, *10*(10), 3479–3488. https://doi.org/10.1534/g3.120.401579

Hoff KJ. and Stanke M. (2019). Predicting Genes in Single Genomes with AUGUSTUS. Current protocols in bioinformatics, 65(1), e57.

Hoshiba H, Imai H (1993) Chromosome evolution of bees and wasps (Hymenoptera, apocrita) on the basis of C-banding pattern analyses. Jpn J Entomol 61:465–492

Hossen, M. S., Shapla, U. M., Gan, S. H., & Khalil, M. I. (2017). Impact of Bee Venom Enzymes on Diseases and Immune Responses. *Molecules*, *22*(1), Article 1. https://doi.org/10.3390/molecules22010025

Kassambara, A. and Mundt, F. (2020) Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7. https://CRAN.R-project.org/package=factoextra

Katoh, Standley 2013 (Molecular Biology and Evolution30:772-780) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. (outlines version 7)

Kim, B. Y., & Jin, B. R. (2016). Molecular characterization of a venom acid phosphatase from the Asiatic honeybee *Apis cerana*. Journal of Asia-Pacific Entomology, 19(3), 793–797. https://doi.org/10.1016/j.aspen.2016.07.013

Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, *37*(5), 540–546. https://doi.org/10.1038/s41587-019-0072-8

Koludarov, I., Velasque, M., Senoner, T. et al. Prevalent bee venom genes evolved before the aculeate stinger and eusociality. BMC Biol 21, 229 (2023). https://doi.org/10.1186/s12915-023-01656-5

Lee, S. H., Baek, J. H., & Yoon, K. A., 2016. Differential Properties of Venom Peptides and Proteins in Solitary vs. Social Hunting Wasps. Toxins, 8(2). https://doi.org/10.3390/toxins8020032

Lee, K. S., Kim, B. Y., Yoon, H. J., & Jin, B.-R. (2023). Proteases and Protease Inhibitors in Bee Venoms. *Journal of Apiculture*, *38*(4), 391–396. https://doi.org/10.17519/apiculture.2023.11.38.4.391

Lima, P. R. de, & Brochetto-Braga, M. R. (2003). Hymenoptera venom review focusing on Apis mellifera. *Journal of Venomous Animals and Toxins Including Tropical Diseases*, *9*, 149–162. https://doi.org/10.1590/S1678-91992003000200002

Linsley EG, 1966. Pollinating insects of the Galapagos Islands. Berkeley.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution, 38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

McMullen, C. K.,1989. The Galapagos carpenter bee, just how important is it?. Noticias de Galapagos. 48.

McMullen, C. K,1999. Flowering Plants of the Galapagos. Cornell University Press.

Meng, Y., Lei, Y., Gao, J., Liu, Y., Ma, E., Ding, Y., Bian, Y., Zu, H., Dong, Y., & Zhu, X. (2022). Genome sequence assembly algorithms and misassembly identification methods. *Molecular Biology Reports*, *49*(11), 11133–11148. https://doi.org/10.1007/s11033-022-07919-8

Michener, C. D. (1990). Castes in xylocopine bees. In W. Engels (Ed.), *Social insects*.Springer.

Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., Miga, K. H., Eichler, E. E., Phillippy, A. M., & Koren, S. (2020). HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Research, 30(9), 1291–1305. https://doi.org/10.1101/gr.263566.120

Oksanen J, Simpson G, Blanchet F, Kindt R, Legendre. et al (2024). vegan: Community Ecology Package. R package version 2.7-0, https://github.com/vegandevs/vegan, https://vegandevs.github.io/vegan/.

Oxford Academic, SF. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, SF. https://academic.oup.com/bioinformatics/article/31/19/3210/211866

Nash, W. J., Man, A., McTaggart, S., Baker, K., Barker, T., Catchpole, L., Durrant, A., Gharbi, K., Irish, N., Kaithakottil, G., Ku, D., Providence, A., Shaw, F., Swarbreck, D., Watkins, C., McCartney, A. M., Formenti, G., Mouton, A., Vella, N., von Reumont, B. M., Vella, A., & Haerty, W. (2024). *The genome sequence of the Violet Carpenter Bee, Xylocopa violacea (Linnaeus, 1785): a hymenopteran species undergoing range expansion*. bioRxiv. https://doi.org/10.1101/2024.04.03.587942

Paradis, E., Claude, J., Strimmer, K., 2004. APE: Analyses of Phylogenetics and Evolution in R. Bioinformatics, 20(15), 289-290. https://doi.org/10.1093/bioinformatics/btg412

Pinheiro, J., Bates, D., & R Core Team. (2025). nlme: Linear and nonlinear mixed effects models (Version 3.1-167) [R package]. https://CRAN.R-project.org/package=nlme

Qiu, Y., Choo, Y. M., Yoon, H. J., Jia, J., Cui, Z., Wang, D., Kim, D. H., Sohn, H. D., & Jin, B. R. (2011). Fibrin(ogen)olytic activity of bumblebee venom serine protease. *Toxicology and Applied Pharmacology*, *255*(2), 207–213. https://doi.org/10.1016/j.taap.2011.06.020

Rasmussen, C., Carríon, A. L., Castro-Urgal, R., Chamorro, S., Gonzalez, V. H., Griswold, T. L., Herrera, H. W., McMullen, C. K., Olesen, J. M., Traveset, A., 2012. Megachile timberlakei Cockerell (Hymenoptera: Megachilidae): Yet another adventive bee species to the Galapagos Archipelago. The Pan-Pacific Entomologist, 88(1), 98-102. https://doi.org/10.3956/2012-04.1

Rautiainen, M., Nurk, S., Walenz, B. P., Logsdon, G. A., Porubsky, D., Rhie, A., Eichler, E. E., Phillippy, A. M., & Koren, S. (2023). Telomere-to-telomere assembly of diploid chromosomes with Verkko. Nature Biotechnology. https://doi.org/10.1038/s41587-023-01662-6

Rehan, S. M., Berens, A. J.,Toth, A. L., 2014. At the brink of eusociality: Transcriptomic correlates of worker behaviour in a small carpenter bee. BMC Evolutionary Biology, 14(1), 260. https://doi.org/10.1186/s12862-014-0260-6

Rehan, S. M., & Richards, M. H. (2013). Reproductive aggression and nestmate recognition in a subsocial bee. *Animal Behaviour*, *85*(4), 733–741. https://doi.org/10.1016/j.anbehav.2013.01.010

Rehan, S. M., Glastad, K. M., Lawson, S. P., & Hunt, B. G. (2016). The Genome and Methylome of a Subsocial Small Carpenter Bee, *Ceratina calcarata*. *Genome Biology and Evolution*, *8*(5), 1401–1410. https://doi.org/10.1093/gbe/evw079

Revell, L. J. (2024). phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ, 12*, e16505. https://doi.org/10.7717/peerj.16505

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., … Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Ronquist, F., Huelsenbeck, J. P., Van der Mark, P., 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Systematic Biology, 61(3), 539-542. https://doi.org/10.1093/sysbio/sys029

Salvatore Vicidomini (1996) Biology *of Xylocopa violacea* (Hymenoptera): In- nest ethology, Italian Journal of Zoology, 63:3, 237-242. https://doi.org/10.1080/11250009609356139

Shi, N., Szanto, T.G., He, J. et al. Venom composition and pain-causing toxins of the Australian great carpenter bee *Xylocopa aruana*. Sci Rep 12, 22168 (2022). https://doi.org/10.1038/s41598-022-26867-8

Schendel, V., Rash, L. D., Jenner, R. A., & Undheim, E. A. B. (2019). The Diversity of Venom: The Importance of Behavior and Venom System Morphology in Understanding Its Ecology and Evolution. *Toxins*, *11*(11), Article 11. https://doi.org/10.3390/toxins11110666

Schmidt, J. O., 2014. Evolutionary responses of solitary and social Hymenoptera to predation by primates and overwhelmingly powerful vertebrate predators. Journal of Human Evolution, 71, 12-19. https://doi.org/10.1016/j.jhevol.2013.07.018

Schluter, D., 1986. Character Displacement between Distantly Related Taxa? Finches and Bees in the Galapagos. The American Naturalist, 127 (1): 95–102, doi:10.1086/284470, S2CID 83906633

Shell, W.A., Steffen, M.A., Pare, H.K., 2021. Sociality sculpts similar patterns of molecular evolution in two independently evolved lineages of eusocial bees. Communications Biology, 4, 253. https://doi.org/10.1038/s42003-021-01770-6

Sless, T., & Rehan, S. (2023). Phylogeny of the carpenter bees (Apidae: Xylocopinae) highlights repeated evolution of sociality. *Biology Letters*, *19*(8), 20230252. https://doi.org/10.1098/rsbl.2023.0252

Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics. Bioinformatics, 30 (9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Sunagar, K., Morgenstern, D., Reitzel, A. M., Moran, Y., 2016. Ecological venomics: How genomics, transcriptomics and proteomics can shed new light on the ecology and evolution of venom. Journal of Proteomics, 135, 62-72. https://doi.org/10.1016/j.jprot.2015.09.015

Tan, J., Wang, W., Wu, F. et al. Transcriptome profiling of venom gland from wasp species: de novo assembly, functional annotation, and discovery of molecular markers. BMC Genomics 21, 427 (2020). https://doi.org/10.1186/s12864-020-06851-0

Traveset, A., Heleno, R., Chamorro, S., Vargas, P., McMullen, C. K., Castro-Urgal, R., Nogales, M., Herrera, H. W., & Olesen, J. M., 2013. Invaders of pollination networks in the Galapagos Islands: Emergence of novel communities. Proceedings of the Royal Society B: Biological Sciences, 280(1758), 20123040. https://doi.org/10.1098/rspb.2012.3040

Vargas,P., Rumeu, B., Heleno, RH., Traveset, A., Nogales M, 2015. Historical Isolation of the Galapagos Carpenter Bee (*X. darwini*) despite Strong Flight Capability and Ecological Amplitude. PLOS ONE 10(3): e0120597. https://doi.org/10.1371/journal.pone.0120597

von Reumont, B. M., Dutertre, S., Koludarov, I., 2022. Venom profile of the European carpenter bee *Xylocopa violacea*: Evolutionary and applied considerations on its toxin components. Toxicon: X, 14. https://doi.org/10.1016/j.toxcx.2022.100117

Wan, H., Kim, B. Y., Lee, K. S., Yoon, H. J., Lee, K. Y., & Jin, B. R. (2014). A bumblebee (*Bombus ignitus*) venom serine protease inhibitor that acts as a microbial serine protease inhibitor. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *167*, 59–64. https://doi.org/10.1016/j.cbpb.2013.10.002

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Research, 46(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Wilson, E. O., & Hölldobler, B. (2005). Eusociality: Origin and consequences. *Proceedings of the National Academy of Sciences of the United States of America, 102*(38), 13367–13371. https://doi.org/10.1073/pnas.0505858102

Ye, X., He, C., Yang, Y., Sun, Y. H., Xiong, S., Chan, K. C., Si, Y., Xiao, S., Zhao, X., Lin, H., Mei, Y., Yao, Y., Ye, G., Wu, F., & Fang, Q. (2023). Comprehensive isoform-level analysis reveals the contribution of alternative isoforms to venom evolution and repertoire diversity. Genome Research, 33(9), 1554–1567. https://doi.org/10.1101/gr.277707.123

Yoon, K. A., Kim, K., Kim, W. -J., Bang, W. Y., Ahn, N. -H., Bae, C. -H., Yeo, J. -H., & Lee, S. H. (2020). Characterization of Venom Components and Their Phylogenetic Properties in Some Aculeate Bumblebees and Wasps. Toxins, 12(1), 47. https://doi.org/10.3390/toxins12010047

Zhu, B., Jin, P., Hou, Z., Li, J., Wei, S., & Li, S. (2022). Chromosomal-level genome of a sheet-web spider provides insight into the composition and evolution of venom. Molecular Ecology Resources, 22(6), 2333–2348. https://doi.org/10.1111/1755-0998.13601

Tables

Table 1. Comparative metrics of genome assemblies for *X. darwini*

	Verkko	hicanu	hifiasm	Mecat	Flye
Number of contigs	992	1615	1430	420	433
Total length (Mb)	226	274	283	199	215
Largest contig (Mb)	20.69	20.69	20.69	9.8	20.68
N50 (Mb)	11.97	10.43	10.47	4.26	11.66
N90 (Mb)	0.10	0.04	0.05	0.49	0.27
L50	8	11	11	16	8
L90	72	510	503	60	36
BUSCO %	>95%	>95%	>95%	95%	97.5%

Table 2. Number of genes and venom gene isoforms in assembled genomes of the Xylocopine bees

Species	Sociality	Gene number (Source)	Total venom isoforms number	Venom component	Isoforms count
C. australensis	Eusocial	8,269 (This study)	8	Acid phosphatase	2
				Allergen 3/5	1
				Carboxylesterase 6	1
				Serine protease	2
				Icarapin	1
				Dipeptidyl peptidase	1
X. violacea	Solitary	10,152 (Nash et al. 2024)	15	Acid phosphatase	4
				Allergen 3/5	3
				Carboxylesterase 6	1
				Serine protease	2
				Icarapin	1
				Dipeptidyl peptidase	4
X. darwini	Unknown	14,471 (This study)	16	Acid phosphatase	5
				Allergen 3/5	3
				Carboxylesterase 6	2
				Serine protease	2
				Icarapin	1
				Dipeptidyl peptidase	3
X. virginica	Eusocial	22,019 (This study)	11	Acid phosphatase	4
				Allergen 3/5	2
				Carboxylesterase 6	1
				Serine protease	2
				Icarapin	1
				Dipeptidyl peptidase	1

Figures

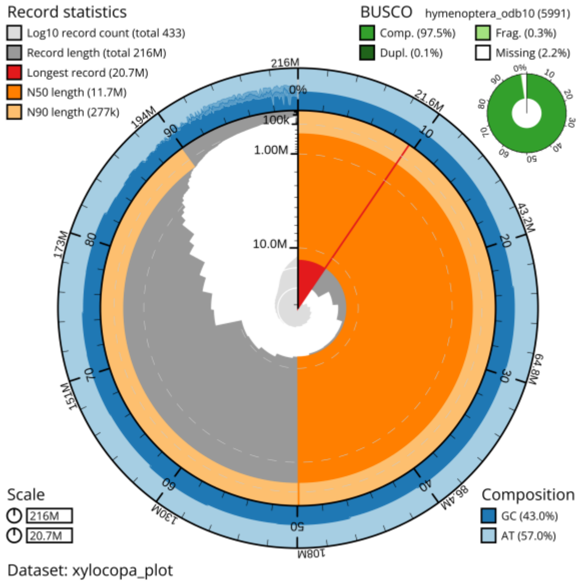


Figure 1. Snail plot of genome assembly for *X. darwini*.

The BlobToolKit Snailplot shows N50 metrics and BUSCO gene completeness. The main plot is divided into 1,000 size-ordered bins around the circumference with each bin representing 0.1% of the 215 Mb assembly. The distribution of contigs lengths is shown in dark grey with the plot radius scaled to the longest contig present in the assembly (20,7Mb, shown in red). Orange and pale-orange arcs show the N50 and N90 scaffold lengths (11.97Mb and 277 k (0.277Mb) respectively). The pale grey spiral shows the cumulative contig count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT. A summary of complete, fragmented, duplicated and missing BUSCO genes in the hymenoptera_odb10 set is shown in the top right.

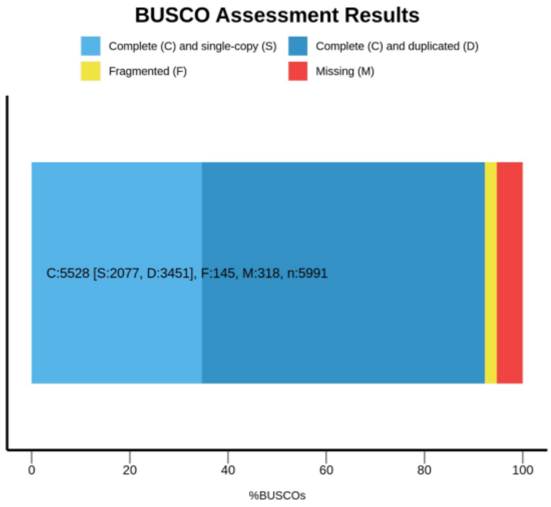


Figure 2. BUSCO completeness assessment of *X. darwini* RNA-seq read assembly. The BUSCO plot displays the completeness of the transcriptome assembly using four color-coded categories. Pale blue represents the proportion of complete single-copy BUSCO genes, while dark blue indicates complete duplicated BUSCOs. Fragmented BUSCOs are shown in yellow, and missing BUSCOs are highlighted in red. Together, these categories illustrate the overall quality and completeness of the assembled transcript sequences.

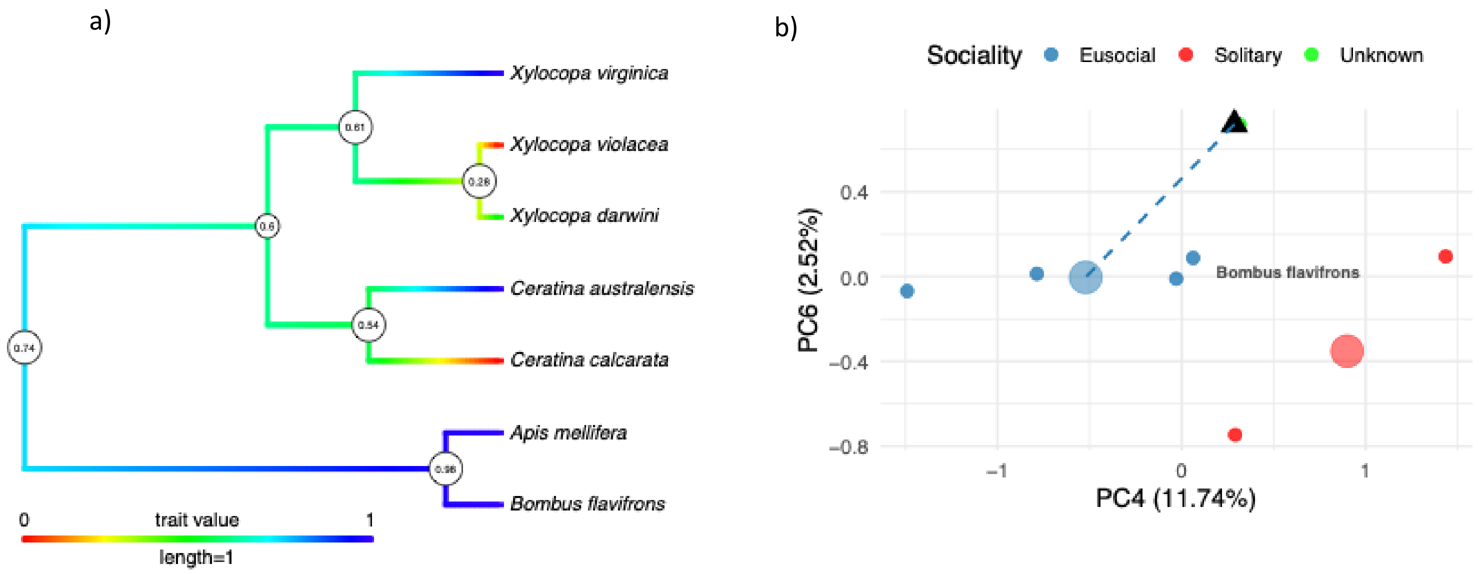


Figure 3. Reconstruction of the ancestral state of venom allergen 3 and PCA to trace sociality in bees.

a) A maximum-likelihood phylogenetic tree was constructed to examine the clustering patterns of venom allergen 3 across bee species. Node colors and values indicate sociality, ranging from solitary (0, red) to eusocial (1, blue). The tree includes species from the genera *Xylocopa*, *Ceratina*, *Apis*, and *Bombus*, capturing both solitary and eusocial lineages. b) The PCA plot visualizes the relationship between *X. darwini* and other solitary and eusocial species taking into account the significant PC4 and PC6. Each point represents a species, colored by sociality classification (Eusocial = blue, Solitary = red, Unknown = green). Larger, semi-transparent points mark the centroids of eusocial and solitary species. *X. darwini* is highlighted as a black triangle, with dashed lines representing its Euclidean distance to both centroids. The closest species to *X. darwini* is labeled.

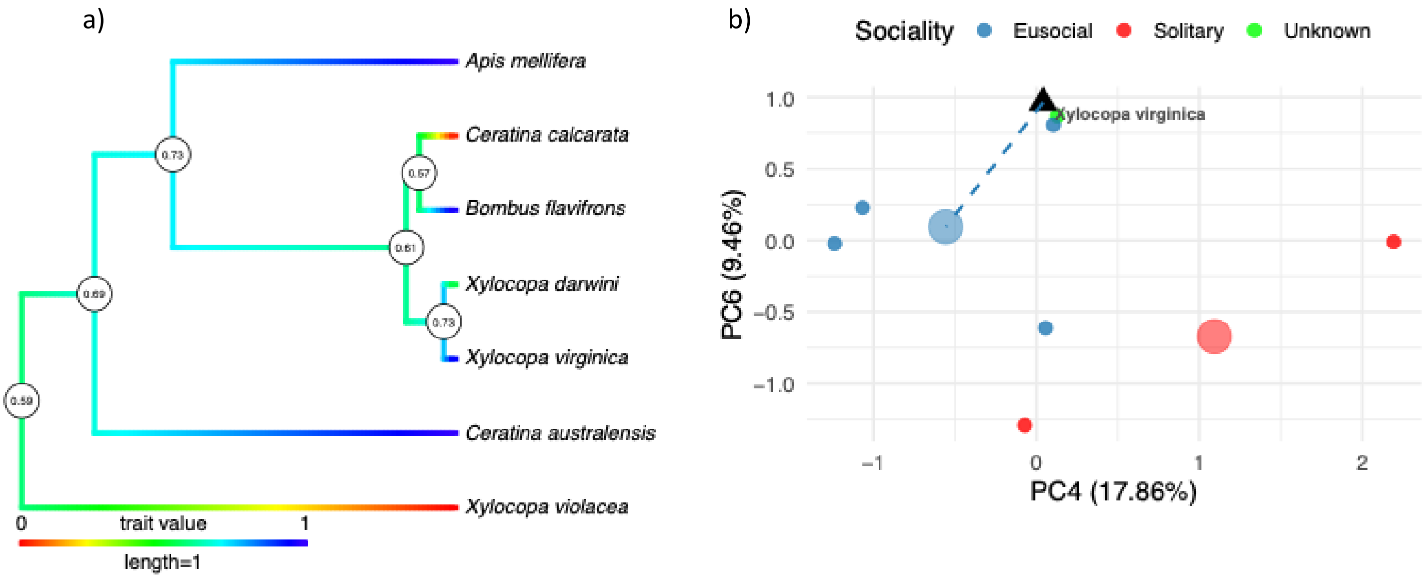


Figure 4. Reconstruction of the ancestral state of venom acid phosphatase and PCA to trace sociality in bees.

a) A maximum-likelihood phylogenetic tree was constructed to examine the clustering patterns of venom acid phosphatase across bee species. Node colors and values indicate sociality, ranging from solitary (0, red) to eusocial (1, blue). The tree includes species from the genera *Xylocopa*, *Ceratina*, *Apis*, and *Bombus*, capturing both solitary and eusocial lineages. b) The PCA plot visualizes the relationship between *X. darwini* and other solitary and eusocial species using the significant PC4 and PC6. Each point represents a species, colored by sociality classification (Eusocial = blue, Solitary = red, Unknown = green). Larger, semi-transparent points mark the centroids of eusocial and solitary species. *X. darwini* is highlighted as a black triangle, with dashed lines representing its Euclidean distance to both centroids. The closest species to *X. darwini* is labeled

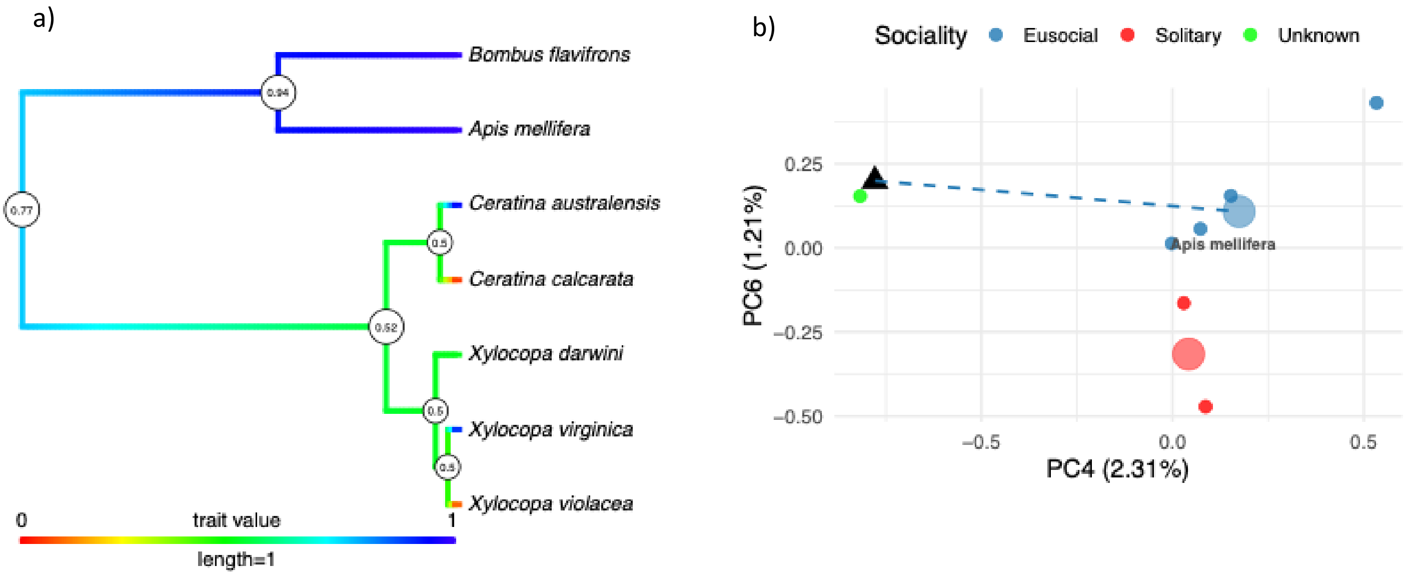


Figure 5. Reconstruction of the ancestral state of venom carboxylesterase 6 and PCA to trace sociality in bees.

a) A maximum-likelihood phylogenetic tree was constructed to examine the clustering patterns of venom carboxylesterase 6 across bee species. Node colors and values indicate reconstructed ancestral states of sociality, ranging from solitary (0, red) to eusocial (1, blue). The tree includes species from the genera *Xylocopa*, *Ceratina*, *Apis*, and *Bombus*, representing both solitary and eusocial lineages. b) The PCA plot visualizes the relationship between *X. darwini* with solitary and eusocial species taking into account the significant PC4 and PC6 from the PGLS analysis. Each point represents a species, colored by sociality classification (Eusocial = blue, Solitary = red, Unknown = green). Larger, semi-transparent points mark the centroids of eusocial and solitary species. *X. darwini* is highlighted as a black triangle, with dashed lines indicating its Euclidean distance to each centroid. The closest species to *X. darwini* is labeled.

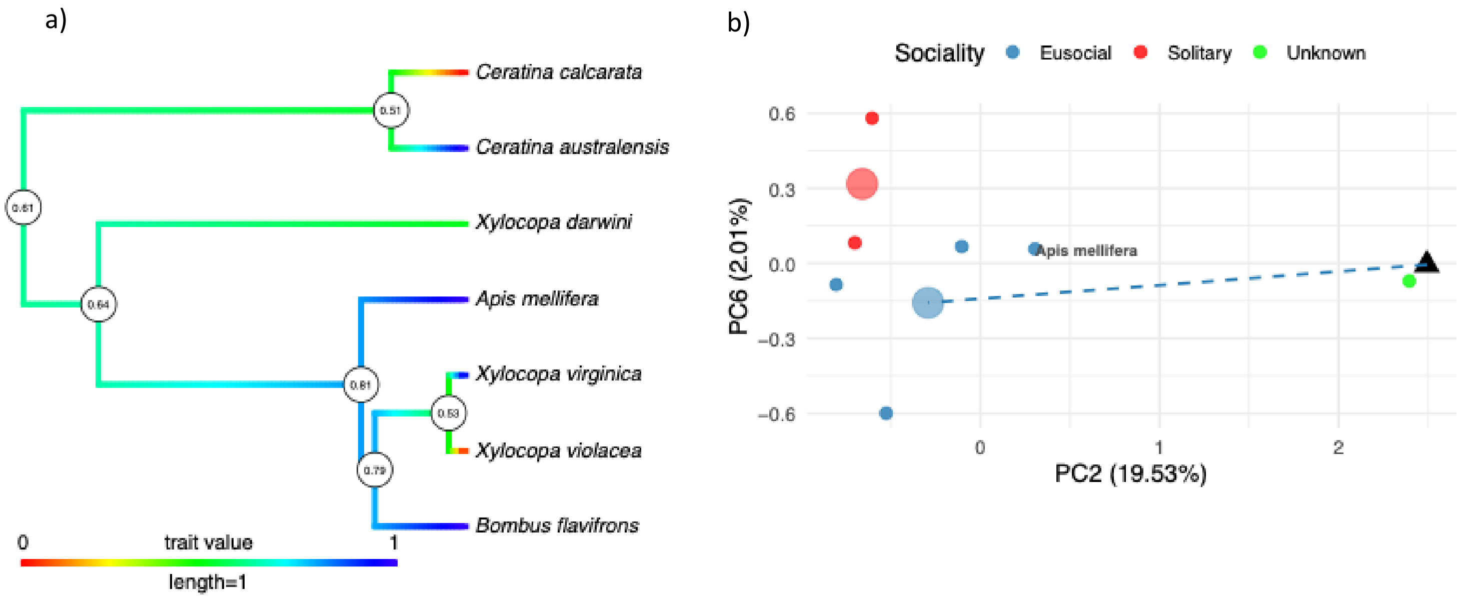
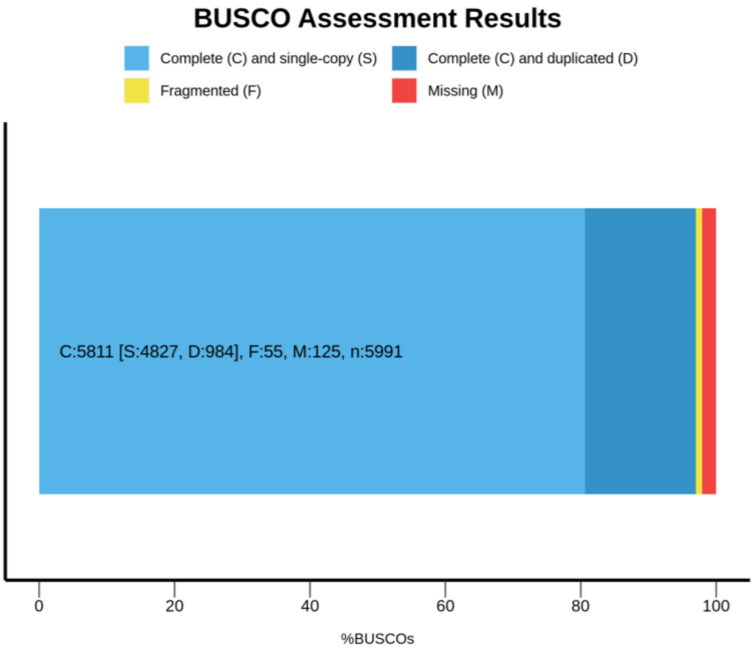


Figure 6. Reconstruction of the ancestral state of venom serine and PCA to trace sociality in bees.

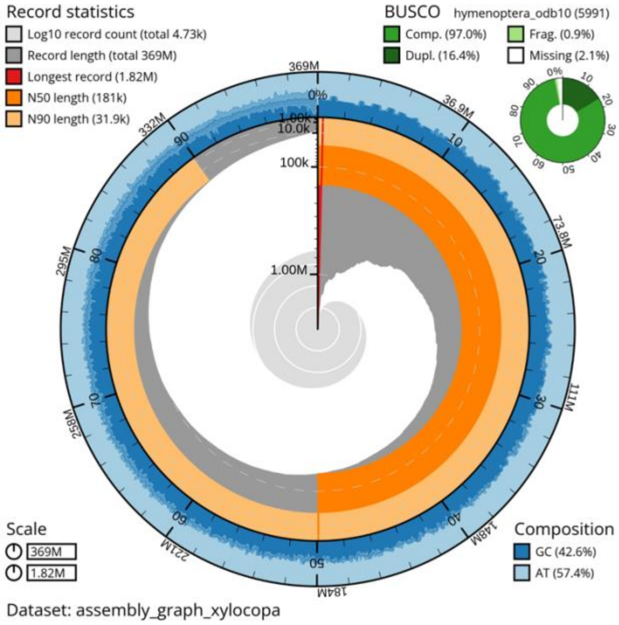
a) A maximum-likelihood phylogenetic tree was constructed to examine the clustering patterns of venom serine across bee species. Node colors and values represent reconstructed ancestral sociality values, ranging from solitary (0, red) to eusocial (1, blue). The tree includes representatives from the genera *Xylocopa*, *Ceratina*, *Apis*, and *Bombus*, reflecting a range of social behaviors. b) PCA plot visualizes the relationship between *X. darwini* and other bee species considering the significant PC2 and PC6. Each point represents a species, colored according to its sociality classification (Eusocial = blue, Solitary = red, Unknown = green). Larger, semi-transparent points mark the centroids of eusocial and solitary species. *X. darwini* is shown as a black triangle, with dashed lines indicating its Euclidean distance to each centroid. The closest species to *X. darwini* is labeled.

Supplementary Information



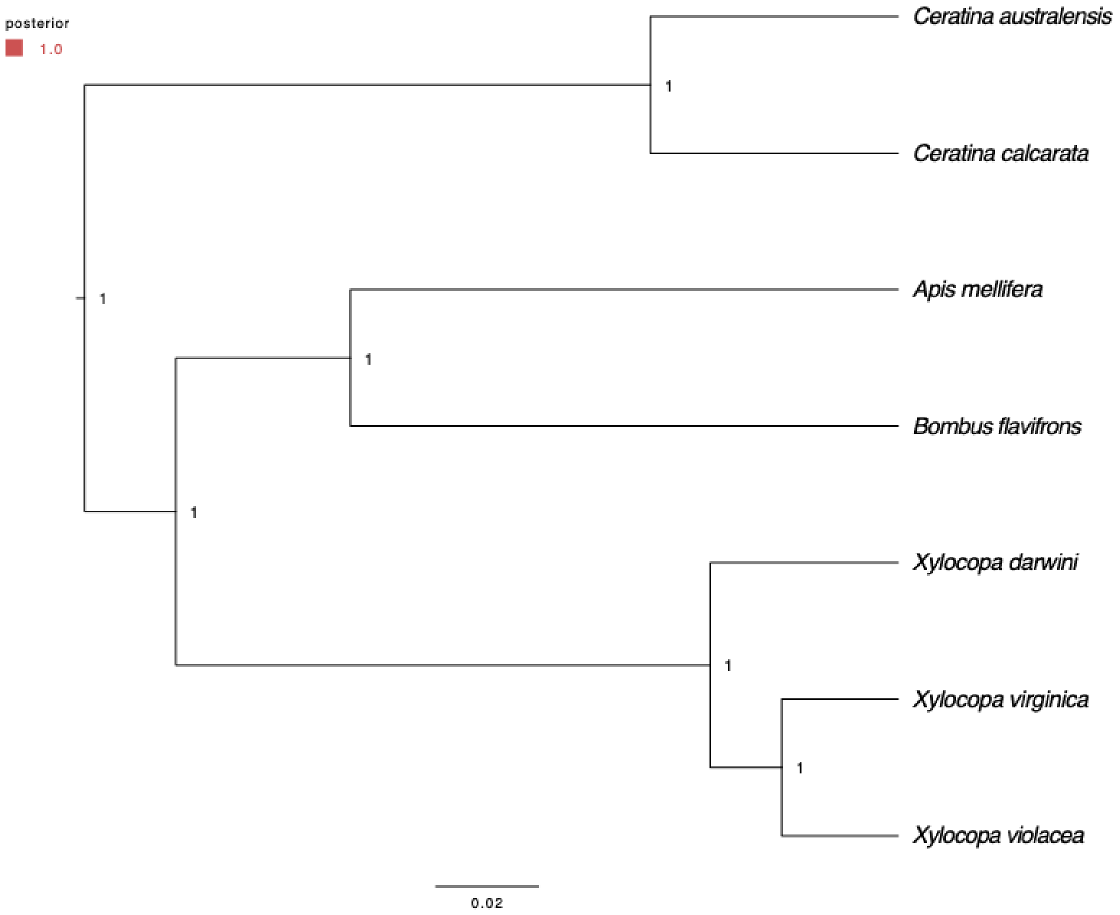
Supplementary Figure 1. BUSCO completeness of *X. darwini* genome assembly annotation using evidence available in NCBI for the Apidae family.

Out of 5,811 BUSCO groups searched, 4,827 were identified as complete and single-copy, and 984 as complete and duplicated. The figure displays the distribution of these categories as part of the overall assessment of annotation completeness using the Hymenoptera lineage dataset.



Supplementary Figure 2. Snail plot of genome assembly for *X. virginica*.

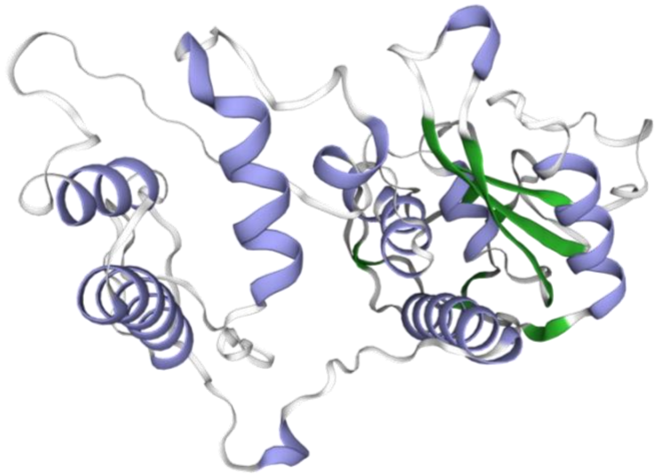
The snail plot generated by BlobTools represents the genome assembly with a total length of 369 Mb, with the longest contig reaching 1.82 Mb. The N50 length is 181 k (0.18 Mb), while the N90 length is 31.9 k (0.03Mb), indicating the distribution of the assembly with the plot radius scaled to the longest contig present in the assembly (shown in red). Orange and pale-orange arcs show the N50 and N90 contigs lengths. The blue and pale-blue area around the outside of the plot shows the distribution of GC, AT. A summary of complete, fragmented, duplicated and missing BUSCO genes in the hymenoptera_odb10 set is shown in the top right.



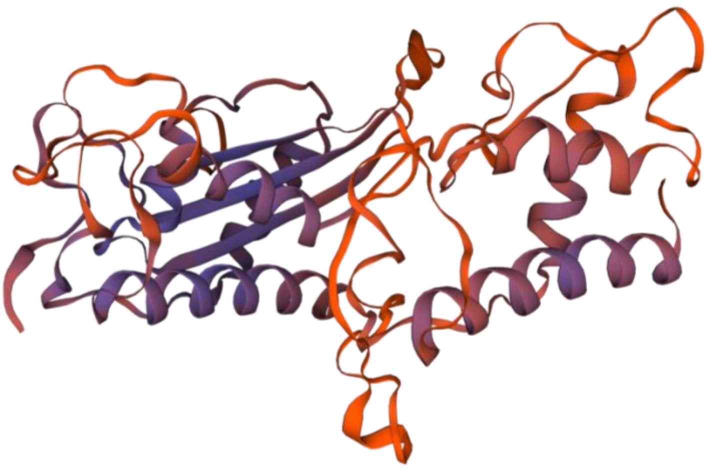
Supplementary Figure 3. Phylogenetic relationships among eusocial and solitary bee species based on UCE data.

This tree illustrates the evolutionary relationships of species used in this study, inferred from ultra-conserved element (UCE) sequences. Branch labels indicate posterior probabilities, with a value of 1.0 representing maximal statistical support.

a) Model allergen 3 protein *X. darwini*



b) Model Ves v 5 an allergen protein from *Vespula vulgaris* venom



Supplementary Figure 4. Predicted 3D structures of allergen proteins from *X. darwini* and *Vespula vulgaris.*

a) Structural model of Allergen 3 protein from *X. darwini*, showing a mix of α-helices (purple) and β-sheets (green), characteristic of venom allergen proteins. b) Structural model of Ves v 5, an allergen protein from *Vespula vulgaris* venom, displaying a complex fold with abundant α-helices (purple) and extended loop regions (orange). These models highlight conserved structural motifs potentially linked to allergenic function across different hymenopteran venoms.

Supplementary Table 1. Assembly statistics and BUSCO completeness of *X. virginica* genome using Flye assembler

	Flye
Number of contigs	4725
Total length(Mb)	368
Largest contig (Mb)	1.82
N50 (Mb)	0.18
N90 (Mb)	0.03
L50	527
L90	2504
BUSCO %	97%

Supplementary Table 2: P-values for generalized least squares fits of venom components across principal components, including distance from the closest species to *X. darwini*

Venom Component	PC	Intercept	SocialitySolitary	SocialityUnknown	Euclidean distance from *X. darwini* to closest species
Acid phosphatase	PC4	0.8623	0.0374**	0.6668	0.12
	PC6	0.4078	0.0762	0.0129**	
Allergen 3/5	PC4	0.4530	0.0302**	0.1721	0.75
	PC6	0.8364	0.0577	0.0410**	
Carboxylesterase 6	PC4	0.2303	0.0448**	0.0018**	0.77
	PC6	0.5294	0.0278**	0.9494	
Serine protease	PC2	0.1909	0.7826	0.0004**	2.24
	PC6	0.5669	0.0564	0.4604	