# Social behavior, ovary size, and population of origin influence cuticular hydrocarbons in the orchid bee, *Euglossa dilemma*

Nicholas W. Saleh[1], Kirsten Hodgson[1], Tamara Pokorny[2], Aaron Mullins[3], Thomas Chouvenc[3], Thomas Eltz[4], Santiago R. Ramírez[1]

[1] Center for Population Biology, University of California, Davis, CA, USA

[2] Institute of Zoology, University of Regensburg, Regensburg, Germany

[3] Entomology and Nematology Department, Ft. Lauderdale Research and Education Center, Institute of Food and Agricultural Sciences, University of Florida, FL, USA

[4] Department of Animal Ecology, Evolution and Biodiversity, Ruhr University Bochum, Bochum, Germany

Keywords: fertility signals, cuticular hydrocarbons, bees, social signaling, chemical communication

Manuscript type: E-article

Dryad Digital Repository: https://doi.org/10.25338/B8FH0P

**Abstract**

Cuticular hydrocarbons (CHCs) are waxy compounds on the surface of insects that prevent desiccation and frequently serve as chemical signals mediating social and mating behaviors. Although their function in eusocial species has been heavily investigated, little is known about the evolution of CHC-based communication in species with simpler forms of social organization lacking specialized castes. Here, we investigate factors shaping CHC variation in the orchid bee *Euglossa dilemma*, which forms casteless social groups of 2-3 individuals. We first assess geographic variation, examining CHC profiles of males and females from three populations. We also consider CHC variation in the sister species, *Euglossa viridissima*, which occurs sympatrically with one population of *E. dilemma*. Next, we consider variation associated with female behavioral phases, to test the hypothesis that CHCs reflect ovary size and social dominance. We uncover a striking CHC polymorphism in *E. dilemma* spanning populations. In addition, we identify a separate set of CHCs that correlate with ovary size, social dominance, and expression of genes associated with social behavior, suggesting that CHCs convey reproductive and social information in *E. dilemma*. Together, our results reveal complex patterns of variation in which a subset of CHCs reflect the social and reproductive status of nestmates.

## Introduction

Social insects exhibit some of the most sophisticated forms of social organization among animals, with some species displaying elaborate mechanisms of communication. In most species of social insects, chemical signals play a central role in mediating communication within and between colonies (Wilson, 1965). Cuticular hydrocarbons (CHCs), which form a wax layer on the cuticular surface of all insects, are the most widespread signals that facilitate social insect communication and organization (Gibbs and Pomonis, 1995; Oi et al., 2015).

The complete set of CHCs expressed on an insect, the CHC profile, may consist of multiple classes of linear and branched alkanes and alkenes, and their qualitative and quantitative variation can be highly dynamic and sensitive to a range of environmental, genetic, or physiological factors (Walsh et al., 2020; Bonduriansky et al., 2015). The expression of some CHC compounds is directly linked to pathways that regulate reproduction in insects therefore enabling CHCs to potentially reflect honest information about individual physiological condition (Kuo et al., 2012; Makki et al., 2014). In contrast, the expression of some CHCs may vary little with physiology and instead co-vary primarily with environmental variables, population of origin, or geographic distance between colonies (Bonelli and Lorenzi, 2014; Dronnet et al., 2006; Otte et al., 2018).

Because CHC profiles are complex mixtures of compounds, different CHCs components can convey different types of information, including social role, nestmate identity, mate quality, or species identity. In the primitively eusocial wasp, *Polistes dominula,* for instance, CHCs are implicated in mate choice, social fertility signaling, and nestmate recognition (Cappa et al., 2013; Izzo et al., 2010; Bruschini et al., 2011). In the termite *Reticulitermes virginicus*, the relative abundance of specific CHCs encode information about social caste, while the presence or

absence of specific CHCs mediate species recognition behaviors with sympatric congeners (Howard et al., 1982).

In highly eusocial Hymenoptera, a robust association exists between ovary activation in reproductive queens and the correlated expression of CHCs that contrasts sharply with ovary inactivation in sterile workers (Smith and Liebig, 2017). However, whether these molecules primarily function as fertility signals, communicating the queen's presence, or instead function to directly inhibit worker ovary development remains an open question (Smith and Liebig, 2017; Holman, 2018). These functions can be difficult to disentangle as they both lead to inactive worker ovaries in the presence of the queen (Oldroyd, 2018). Although there is support for the hypothesis that queen associated CHCs have evolved from fertility linked CHCs in the ancestors of eusocial species, a lack of data from solitary or facultatively social species has hindered comparative analysis (Oliveira et al., 2015; Oi et al., 2015).

Because most studies investigating how CHCs regulate reproductive and social behaviors have been restricted to highly eusocial insect species, such as honey bees, ants, or Vespine wasps (Oystaeyen et al., 2014; Holman, 2018), little is known about the types of information that CHCs convey in species with small social groups without specialized castes (Leonhardt et al., 2016). A better understanding of such species has the potential to provide unique insight into the reproductive and social factors influencing chemical variation (Steiger and Stökl, 2018). The lack of fixed worker castes or the complex communication needs of large colonies, for example, provides a simpler background against which to assess the roles of behavior and physiology in structuring CHC variation.

Euglossine bees (>250 species) are the earliest branching group within the monophyletic corbiculate bees and the only lineage within this clade that lacks obligate eusocial behavior,

instead exhibiting solitary, communal, or facultatively social nesting strategies (Soucy et al., 2003; Solano-Brenes et al., 2018; Friedel et al, 2020). Some species, such as *Euglossa dilemma*, have distinct solitary and social phases of their life cycle, with differences in ovary size occurring across these phases, though all individuals in social nests are reproductive (Andrade-Silva and Nascimento, 2012; Saleh and Ramírez, 2019). In addition to these life-history characteristics, *E. dilemma* is geographically widespread, with multiple native and non-native populations (Brand et al., 2020).

Nests in *E. dilemma* are initiated by a single foundress that provisions an initial brood batch. After completing approximately four to ten brood cells, the foundress ceases foraging and reproduction to guard her developing brood (Saleh and Ramírez, 2019). When a foundress enters this "guard" phase, her ooctyes reduce in size. Once offspring emerge, the foundress's oocytes increase in size and she becomes the social dominant, while 1-2 of her female offspring may remain as subordinate helpers. Other female offspring disperse to begin their own nests. Both dominant and subordinate females are mated and reproductive. However, the dominant female is responsible for all the direct reproductive output of the nest, due to oophagy and replacement of all subordinate eggs. The dominant bee is usually the original foundress and thus the mother of the subordinate bees, although subordinate bees can take over the dominant position if the original foundress bee dies. In some species of *Euglossa,* females may also form social affiliations with unrelated bees (Andrade-Silva et al., 2016). Male bees do not participate in any nest activities and leave the nest shortly after emergence.

In this study, by sampling multiple populations across the geographic range of *E. dilemma*, we investigate patterns of CHC variation in relation to species identity and population of origin. We also investigate whether CHCs covary with social behavior, for which we obtained

data from females sampled while performing different solitary and social behaviors. We begin by documenting CHC variation in *E. dilemma* males and females from Florida (non-native), Costa Rica (native), and Mexico (native). Next, we consider how this relates to CHC variation in the recently diverged sister species, *Euglossa viridissima*, which occurs sympatrically with *E. dilemma* in Mexico (Eltz et al., 2011). Following this, we examine how social behavior, relatedness among nestmates, and ovary size may be reflected in CHC variation and investigate links between socially associated genes and CHC expression. Using these sources of data, we test the hypothesis that female CHCs honestly reflect social dominance and reproductive status, encoding information that may mediate nestmate interactions in *E. dilemma's* casteless social groups.

## Methods

### Florida *Euglossa dilemma* collection

*Euglossa dilemma* females were trap-nested year-round between 2015 and 2019 using small wooden boxes (3.5" x 2.5"x 1.4") placed on the eaves of buildings in Ft. Lauderdale, Florida, where they were accidentally introduced and first detected in 2003 (Skov and Wiley, 2005). Entrance holes (5/16" diameter) were drilled into the front of each box. Following colonization of the nest boxes, transparent red plexiglass lids were placed on the boxes to facilitate observation. Females were individually marked using numbered plastic discs glued to the thorax. Following daily nest observations that occurred for at least an hour per day (approximately 2-4 weeks per nest), we classified individuals into four distinct behavioral groups, sampling individuals in these groups. Two of these four groups (foundress and guard) represent solitary behaviors that a female bee will transition through before the nest becomes

social, containing dominant and subordinate bees. We assigned females to the foundress category (n = 21) if they were observed alone constructing a nest in a box that did not have prior bee activity. A solitary foundress next progresses to the "guard" phase once brood provisioning is complete. Guard bees (n = 20) showed little to no foraging, no brood cell construction, and spent most of the day inside the nest with capped brood cells and a resin seal over the nest entrance during typical daytime foraging hours. After offspring emergence, nests could become social if emerged offspring remained in the nest. We assigned bees to dominant (n = 32) and subordinate (n = 33) behavioral groups based on differences in foraging behavior and oophagy in multifemale nests. Subordinate individuals performed regular foraging trips for pollen and resin and did not perform oophagy. Dominant individuals were not observed foraging for pollen and displayed consistent oophagy of subordinate eggs. Multifemale nests never had more than one dominant bee. After observation was complete, females were anesthetized by placing the entire nest box on dry ice and then females were removed and frozen in liquid nitrogen and/or a -80 freezer until analysis. Our sampling and observations were done on existing nests and thus likely represent a variety of ages and genetic relationships among dominants and subordinates, including matrifilial and sororal affiliations as well as possible affiliations between non-relatives, which are known to happen in *Euglossa* species (Andrade-Silva et al, 2016). Newly emerged females were sampled within 24hrs of emergence from harvested trap-nests maintained in a mesh enclosure in ambient conditions during the Summer and Fall of 2019 in Ft. Lauderdale, FL (n= 17). Male FL *E. dilemma* (n = 22) were collected at chemical baits around Ft. Lauderdale, FL and stored in a -80°C freezer until further analysis.

**Costa Rica *E. dilemma* collection**

*Euglossa dilemma* females were trap-nested at Palo Verde Biological Station in Guanacaste, CR in the Spring of 2018 using small wooden boxes (as described above) placed around the station. Females were also caught from flowers found around the station. Due to logistical constraints, Costa Rica *E. dilemma* females (n = 13) were not assigned into the four behavioral groups described above. *Euglossa dilemma* males (n = 20) were collected at chemical baits placed at Palo Verde Biological Station. Males and females were frozen after collection until analysis of CHCs. Collection permits were obtained through SINAC (National System of Conservation Areas; permit number: M-P-SINAC-PNI-ACAT-073-2019).

**Mexico E. dilemma and E. viridissima collection**

CHC data on Mexican *E. dilemma* and *E. viridissima,* collected from the Yucatán peninsula, were provided by the authors of Pokorny et al., 2015 and used for comparison to our datasets from FL and Costa Rica. In addition to the samples from Pokorny et al., 2015, data from several unpublished *E dilemma* and *E. viridissima* females from Mexico was provided. We discuss integration of our datasets with these published and previously unpublished datasets below.

**CHC extractions**

We extracted CHCs from one set of forewings and hindwings adapting the protocol in Martin et al., 2009. Briefly, we placed the wings in a glass scintillation vial with 100µl of hexane, swirled the wings and then waited 10 minutes before moving the wing extract to GC-MS

vials, which were left overnight to evaporate in a closed fume hood. The next day, 30µl of hexane was added to the sample, which was then run on the GC-MS. We extracted CHCs from the wings to keep our methods consistent with the wing extractions from Pokorny et al., 2015, used in this study. To evaluate the degree to which wings extractions provide an accurate representation of CHCs on the cuticular surface, we also compared extracts from the wings, abdomen, and the whole body from 16 individuals, with details provided in Appendix 1.

Extracts were run using a 1µl splitless injection on a GC-MS (Agilent 7890B GC, 5977A MS, HP-5MS Ultra Inert 30m, 0.25mm, 0.25µm column), modifying a program from Choe et al., 2012 that began at 100°C for 1 minute, increasing 15°C per minute until 300°C was reached, after which the program held at 300°C for three minutes. Helium was used as the carrier gas.

**CHC data analysis**

Chromatograms generated using the GC-MS were integrated to include all peaks with an area of least 0.1% of the largest peak. Compound identification was accomplished by comparison to previously published information for *E. dilemma* (Pokorny et al., 2014; Pokorny et al., 2015) and comparison to available mass spectral libraries and known mass indices. We excluded peaks from analysis that were not consistently identified as CHCs (linear and branched alkenes and alkanes) across individuals. Following the removal of non-CHC peaks, we determined the relative abundance of each CHC peak per sample to generate proportional data.

Visualization of chemical datasets was done on this proportional data using NMDS plots based on Bray-Curtis dissimilarity implemented in the Vegan v2.5-4 R package (Oksanen et al, 2019; R Core Team, 2020). Data and code underlying all figures is deposited in the Dryad Digital Repository: https://doi.org/10.25338/B8FH0P. In addition, to find the relative

contribution to dissimilarity of specific CHCs, we used SIMPER analysis, as implemented in the Vegan package.

To find potential modules of covarying CHC peaks, we used the Corrplot R package (Wei and Simko, 2017) to perform hierarchical clustering of the peaks using the Ward.D2 clustering method, based on Spearman correlation coefficients among peaks.

We used one-way ANOVAs to test for differences in the relative abundances of specific CHCs across social behaviors. We adjusted for multiple comparisons using Tukey's HSD tests. We also used a Levene's test and a Shapiro-Wilk test to verify ANOVA assumptions of homogeneity of variances and normally distributed data, respectively. If the data violated either of these assumptions, the data were square root transformed. If transformed data still failed ANOVA assumptions, we used Kruskal-Wallis tests with Steel-Dwass tests for multiple comparisons (Douglas and Michael, 1991). We plot untransformed data for clarity. To identify and test for correlations between specific CHC peaks and ovary size, we used linear regression with the Pearson coefficient when datasets were normally distributed and the Spearman coefficient when data deviated from normality.

One bee that appeared as an extreme outlier in NMDS analysis (Fig. S1) was removed from this and all subsequent statistical analysis but was included in the genotype dataset (described below). This bee was observed entering a nest and usurping the dominant position. Data for this individual (N28_W85) are included in the supplemental materials.

**Combining data matrices among populations**

Because Mexican *E. dilemma* and *E. viridissima* were sampled separately and run on a separate GC-MS from the Florida and Costa Rica samples, detection of low abundance CHCs

may vary in a batch specific manner. Consequently, to avoid including CHCs that could inflate group separation due to technical differences, in our combined population analysis, we only included CHCs detected in >85% of samples across the entire dataset. In addition, we removed four of the 40 samples from the Mexican *E. dilemma* and *E. viridisimma* population (1 *E. viridissima* female and 3 *E dilemma* females) due to missing or ambiguous peaks that could not be reconciled with our Costa Rica and Florida datasets. Primarily, this appears to be due to possible sample contamination or low quantities of extracted sample. We include these individuals in our supplemental data (labeled as "removed outliers"). In total, 36 individuals were included from the Mexico population, which includes data from Pokorny et al., 2015 as well as additional unpublished samples (*E. viridissima* females, n = 8; *E viridissima* males, n = 10; *E. dilemma* females, n = 8; *E. dilemma* males, n = 10).

**Ovary size measurement**

An ovary size index was calculated on a subset of individuals that were available for dissection (n = 67), by taking the length of the longest basal oocyte divided by the inter-tegular distance to incorporate body size variation among individuals (Cane, 1987). We use this metric to represent ovary size in *E. dilemma*.

**Transcriptomic correlates with CHCs**

We examined the expression of genes previously found to be associated with social behavior in *E. dilemma* to assess possible links among social behavior, reproductive physiology, and CHC expression. Specifically, we quantified the correlation between the relative abundance

of specific CHCs and differentially expressed genes (DEGs) previously identified between dominants and subordinates (Saleh and Ramírez, 2019). Using ovary transcriptomic data (n = 27 individuals), we examined correlations between CHCs of interest and the 10 genes found to be differentially expressed between the ovaries of dominants and subordinates. From brain transcriptomic data (n = 31 individuals), we examined correlations between CHCs of interest and the 204 genes found to be differentially expressed between the brains of dominant and subordinates. The pre-existing transcriptomic data used here were generated and processed as detailed in Saleh and Ramírez, 2019 (data available at Bioproject accession prjna523381) along with CHCs taken from the same individual bees. In addition to examining correlations among genes and CHCs, we also assessed relationships between the expression of these genes and ovary size, which was not previously explored in this data set. We note that the brain and ovary tissues used here do not directly express CHCs, and instead are secreted by the oenocytes located on the cuticular surface (Makki et al., 2014). However, assessing correlations among CHCs and expression data from these tissues may provide insight into the pathways or genes that could mediate links among CHC expression, social behavior, and reproductive physiology. One foundress individual from the original transcriptomic data set failed CHC extraction and was not included in analysis. Expression data consists of the log2 normalized counts per million transcripts, relative to the average value across the entire set of individuals. Correlations between CHCs of interest were performed using spearman correlation tests. FDR p-value adjustment (<0.05) was used independently on the brain (204 genes) and ovary (10 genes) datasets to correct for the large number of correlation tests conducted.

**DNA extractions for relatedness genotyping**

We genotyped individual bees from a subset of nests to investigate the association between relatedness and CHC variation. DNA extractions were carried out using the Qiagen DNeasy Blood and Tissue Kit (Product ID: 69504), according to the manufacturer's protocol, using half of the thorax, ground and incubated with proteinase K. DNA concentration was assessed using a Qubit prior to library preparation. Genotype-by-sequencing (GBS) libraries were prepared from 55 individuals from 21 nests using 50ng of DNA per individual to generate single nucleotide polymorphisms (SNPs) following the protocol published in Elshire et al., 2011. We sequenced all adult females found in our collected nests as well as emerging offspring from several controls (discussed further in appendix 2). We used the enzyme ApeKI as previous experience indicated its high efficiency in *E. dilemma*. During enzyme digestion, we incubated the samples for two hours at 37°C, followed by 20 minutes at 65°C. Finally, we optimized the PCR protocol to include 20 cycles. The GBS library was then sent to the Vincent J. Coates Genomics Laboratory in Berkeley, CA, for one lane of 100 bp single-end sequencing on an Illumina HiSeq 4000. Raw, demultiplexed reads are available for all individuals at NCBI Bioproject accession prjna623571.

**SNP calling and relatedness estimation**

We used the program STACKS (Catchen et al., 2013) to demultiplex pooled sequences and BWA-MEM (Li, 2013) to align reads to the *E. dilemma* genome (Brand et al., 2017) before running STACKS again to call SNPs. We varied STACKS parameters so that the relationships of control individuals were optimized (for more information, see appendix 2). For these controls, we had four nests composed of individuals of known relationships (three nests with a mother and

daughter(s) and a nest with two unrelated individuals). We also included two samples of the same individual as an additional control. Ultimately, we found that setting strict parameters in STACKS gave us the best resolution for resolving relationships among our controls. We only counted SNPs found across all individuals in the dataset and set maximum heterozygosity levels at 0.7 with the minor allele frequency set at 0.1. Using this strict filtering approach, we generated 669 SNPs found across all individuals that were used to estimate relatedness.

We used the program COANCESTRY (Jinliang Wang, 2011) to generate relatedness values based on our SNP set. COANCESTRY estimates relatedness between pairs of individuals (or dyads) relative to background allele frequencies calculated from the provided samples. We used the Wang estimator statistic which should be less biased by relatively small sample sizes (tens instead of hundreds of individuals), such as in our study (Wang, 2017). We note that our final relatedness estimates are much lower than typical values for Hymenoptera (Fig. S5 and Table S1). However, underestimation of relatedness values is expected when calculating relatedness estimates with SNPs from genomic datasets that have a relatively low sample size and contain inbred individuals (Wang, 2014; Wang, 2017). Despite this, we expect our dataset to provide internally consistent estimates that can be used for relative comparisons among our samples. We used our control samples of known relationship to calibrate the approximate relatedness values between mothers, daughters, sisters, and non-kin (Fig. S5, appendix 2).

## Results

### *Euglossa dilemma* shows CHC dimorphism across populations

We identified 11 CHC compounds that were present across all populations and were present in >85% of all individuals in the dataset (see Fig. S2 for example chromatograms and

Table S3 for summary statistics). These CHCs captured most of the CHC variation present across samples. In Mexican *E. dilemma* and *E. viridissima*, these 11 CHCs made up an average of 95.4% of the total peak area of all CHCs detected. In Florida and Costa Rica individuals, which were processed together, these 11 CHCs made up an average of 97.6% of the total peak area of all CHCs detected. Plotting the three populations of *E. dilemma* and one population of the sister species *E. viridissima* using NMDS revealed a discrete polymorphism in the CHCs from *E. dilemma*, with individuals falling into two clusters, which we hereafter refer to as chemotype-A and chemotype-B (Fig. 1; Fig S2). *Euglossa dilemma* females from all three populations appear to fall into both chemotypes, while male *E. dilemma* appear to occupy one chemotype or the other depending on the population of origin. SIMPER analysis revealed that these chemotypes are overwhelmingly driven by differences in the relative abundances of two alkenes, 9-heptacosene (9-C27:1) and 9-pentacosene (9-C25:1), which account for 80.75% of the overall dissimilarity between the chemotypes. We note that the Mexican population of *E. dilemma* and sister species *E. viridissima*, which is the only sampled region where they are sympatric, shows consistent differences between males of each species. This pattern was first detected in Mexico by Pokorny et al., 2014, where it was suggested that these differences could be a useful tool in discerning species identity between males of *E. dilemma* and *E. viridissima* in this population.

**Social behaviors are not correlated with discrete observed chemotype variation**

We visualized CHC variation across social behaviors within the Florida population of *E. dilemma* using NMDS of all 17 detected CHCs in females in this population. We found representatives from all the behavioral groups (foundresses, guards, dominants, subordinates, and newly emerged females) in each of the two chemotypes (Fig. 2). In chemotype-A individuals are

loosely clustered according to behavior, though this is less clear in chemotype-B. In addition, we found that individuals within a nest can occupy either the same or different chemotypes, with seven of 21 genotyped nests containing individuals of both chemotypes (Table S1).

Using our genotype dataset, we specifically sought to assess if kin within a nest can occupy the two different chemotypes. Alternatively, mixed chemotype nests may correspond to unrelated individuals, as *Euglossa* species are known to regularly form associations with non-kin (Andrade-Silva and Nascimento, 2016). The presence of mixed chemotypes among kin nestmates would suggests that chemotype is probably not determined by early social interaction, homogenization of CHCs within a nest, or other environmental influences after eclosion. Furthermore, since *Euglossa* species have been shown to discriminate kin and non-kin (Andrade-Silva and Nascimento, 2016), this may provide insight into whether chemotype variation encodes information about nestmate relatedness.

Of the 21 nests genotyped, seven showed mixed chemotypes among nestmates and 14 showed the same chemotype among all nestmates (12 nests were all chemotype-A and two nests were all chemotype-B). Of the seven mixed chemotype nests, three of these nests show high relatedness values among nestmates expressing opposite chemotypes, indicating that they are likely kin (mothers, daughters, or sisters). Two of these seven nests show low relatedness values indicating non-kin, and two of these seven nests show intermediate relatedness values and could not be confidently assigned to kin/non kin categories (appendix 2, Table S1). These results suggest that the kin/non-kin composition of nests does not consistently correlate with the presence of mixed chemotypes within a nest. Overall, kin relationships were clearly most common among all sampled individuals, with only 2/21 nests showing relatedness values

indicative of non-kin nesting together. More detailed results and discussion are found in appendix 2.

**Female CHCs form three covarying modules**

Since the two major chemotypes that we identified do not appear to be correlated with social behavior (varying even within a nest or among kin), it is unlikely that the CHC variation associated with chemotypes A and B reliably reflects ovary size or social hierarchy. Consequently, we sought to identify additional variation among CHCs that could potentially contain information on social status and/or reproductive physiology in multifemale nests. CHC profiles are often composed of groups of correlated peaks, that may encode different sets of information (Martin and Drijfhout, 2009). Therefore, we used hierarchical clustering of all CHCs among the four female behavioral groups sampled from naturally founded trap-nests (foundress, guard, dominant, subordinate) to identify groups of covarying peaks. Newly emerged females (<24 hours post emergence) were excluded from this and further analysis, due to differences in their CHC profiles that were not representative of females sampled from established nests (appendix 3, Fig. S6). With our hierarchical clustering approach, we identified three groups of covarying peaks (Fig. 3) that we hereafter refer to as CHC modules. The first two modules are composed of CHCs that correlate with the two chemotypes (A and B), with individuals in chemotype-A showing a higher relative abundance of module-one peaks than chemotype-B individuals and vice versa. The first module is made up of alkenes and alkanes between 21 and 25 carbons long (diagnostic of chemotype-A). The second module consists of alkenes between 26 and 33 carbons long (diagnostic of chemotype-B). The third module is composed only of alkanes varying from 25 to 31 carbons. We further investigated correlations among behavior and

physiology with this third set of peaks, hereafter referred to as "module-three," to evaluate if it could reliably reflect social information not represented in the discrete variation of the chemotype A and B peaks.

**Module-three CHCs are correlated with social dominance and ovary size**

To assess if the CHCs contained in module-three reflect social dominance and ovary size, we first calculated the total relative abundance of all the CHCs in this module and tested for differences among the four behavioral groups. Dominant bees showed a higher relative abundance of these CHCs compared to bees from other behavioral groups: foundresses, guards, and subordinates ($F_{3,102} = 27.25$, $P < 0.001$; Fig. 4). We also inspected the relative abundance of this CHC module between dominants and subordinates taken from the same nest, as pairwise interactions within a nest may be particularly important for communicating social dominance. In 22/24 nests where we had data for both dominant and subordinate females, dominants showed higher relative expression of this CHC module when compared to the subordinate(s) in that same nest. Furthermore, the module was significantly correlated with ovary size ($r = 0.47$, $P <0.001$; Fig. 5).

After performing these tests on the CHC module, we re-ran the same tests on the individual component CHCs, to assess whether the overall patterns were driven by a subset of the CHCs found in the module. Full results can be found in table S2. In summary, the individual CHCs in the module show consistent patterns, in which ovary size correlates with CHCs and behavioral differences in C25, C27, C29 and C31, with a non-significant relationship for C26, though patterns are in the same direction. We also assessed these patterns in CHC modules one and two (chemotype associated peaks) to test if the module-three peaks are uniquely correlated

with ovary size and dominance status (Cotton et al., 2004). We find that the chemotype associated modules do not show significant correlations with ovary size and they are not overrepresented in dominant individuals (details and figures in appendix 4).

**Correlations between gene expression and module-three CHCs**

Next, we assessed relationships between the relative abundance of the module-three CHCs and the expression of known DEGs between dominants and subordinates. In addition, we also assessed ovary size correlations with these DEGs to identify overlap in DEGs correlated with both ovary size and CHC module-three. These analyses may provide additional evidence for links (or lack thereof) between CHCs, reproductive physiology, and gene expression. This can allow for a better understanding of the possible information content of the module-three CHCs, though correlation analysis cannot demonstrate whether any identified genes directly influence CHC expression or ovary size. Full results of the correlation analysis for the brain and ovary data with gene annotations, original p-values, and FDR adjusted p-values can be found in Tables S4 and S5. In the ovaries, five of the 10 previously identified DEGs between dominants and subordinates were significantly correlated with CHC module-three following FDR correction. Of these five genes, three were also significantly correlated with ovary size. These three genes are facilitated trehalose transporter Tret 1-like (Fig. 6), inositol oxygenase (Fig. S7), and UDP-glucuronosyltransferase 1-8 (Fig. S8), the expression levels of which are highly correlated (Fig. S9). Consequently, we show the correlation between module-three CHCs and facilitated trehalose transporter Tret-1 expression (Fig. 6) as a representative example with the other plots and correlations found in the supplemental material (Fig. S7, Fig. S8, Table S2). In the brain, three of the previously identified 204 DEGs were significantly correlated with CHC module-

three peaks following FDR correction. None of these genes show correlated expression with ovary size and none of these three genes have functional annotation information available (see Table S5 for *E. dilemma* Gene IDs and correlation results).

## Discussion

In this study, we examine CHC variation across species, populations, and social behaviors with emphasis on *E. dilemma*, finding that population and ovary size/dominance status correlate most strongly with separate components of the CHC profile. We identify a CHC polymorphism that exists across at least three populations, showing sex-specific patterns of expression that are distinct when *E. dilemma* is in sympatry with *E. viridissima*. In addition, we identify a subset of CHCs not primarily associated with this polymorphism that reflect social dominance and ovary size. These CHCs have higher relative abundance in dominant bees compared to subordinate bees, they are correlated with ovary size, and show higher relative abundance in dominant bees when compared to other behavioral groups. Furthermore, these CHCs are also correlated with the expression of several DEGs previously identified between dominants and subordinates. Taken together, these data provide an example of how a complex chemical phenotype may be parsed into multiple components, each potentially containing different types of information useful for communication. Furthermore, these results highlight the ability of CHCs to reflect social and reproductive information in an insect species with small social groups lacking obligate reproductive division of labor.

**Possible mechanisms underlying chemotype variation in *E. dilemma***

One possibility is that the discrete chemotype variation corresponds to a genetic polymorphism as opposed to an ontogenetic shift over an individual's lifetime. This scenario is supported by the finding that all behavioral groups, closely related females, and newly emerged females can exhibit both chemotypes (A and B). Furthermore, these chemotypes differ primarily in the relative abundance of two alkenes that are part of a homologous series (C27:1 and C29:1), so it is possible that the phenotypic expression of these chemotypes is the result of a relatively simple genetic polymorphism of a desaturase, elongase, or other fatty acid synthesis gene (Luo et al., 2019; Coyne et al., 1999). Alternatively, it could also be that differences in development primarily affect the expression of chemotype. In addition, the sex-specific expression pattern across populations that we see, where all female *E. dilemma* populations occupy both chemotypes, with males seeming to occupy either one or the other, does not lend itself to a clear interpretation of how chemotype is transmitted across generations.

**Chemotypes could be involved in close range interactions in sympatric populations**

It has been previously suggested that orchid bee CHCs may be involved in close-range communication between males and females in mate choice and/or species recognition contexts (Pokorny et al, 2015). Although males of *E. dilemma* from Costa Rica and Florida occupy the same chemotype as *E. viridissima* males, we find that sympatric males of *E. dilemma* and *E. viridissima* from southern Mexico occur in opposite chemotypes. This could facilitate consistent identification of species by females, which, in most euglossine bees, typically approach males that are engaged in a sex-specific display behavior for mating opportunities (Pokorny et al.,

2017). Future experimental manipulation of CHCs is a necessary next step to determine if CHCs play a role in close range mating behavior and species recognition and reproductive isolation.

The pattern of differentiation that we observed among *E. dilemma* males in Mexico also aligns with population genetic data supporting two distinct lineages of *E. dilemma*, with a northern lineage reaching up through the Yucatán peninsula and a southern lineage extending through Costa Rica and into Panama (Brand et al., 2020). The Florida population is believed to be derived from the southern *E. dilemma* lineage, which is consistent with our clustering of males from Costa Rica and Florida in the same chemotype.

**Module-three CHCs reflect dominance status and reproductive physiology**

Our data are consistent with the hypothesis that module-three CHCs in *E. dilemma* could communicate dominance and reproductive information in social groups, as these CHCs show significant associations with both. In contrast, the CHCs involved primarily in the differentiation of the two chemotypes are not correlated with ovary size and are not overrepresented in dominant bees.

These results are also consistent with the interpretation that module-three CHC variation is driven by reproductive physiology and behavior as opposed to age, though age does generally influence CHC abundances (Gershman and Rundle, 2016). Dominant bees and guard bees are, on average, closer together in age than dominant bees and subordinate bees or foundress bees, which are generally younger. Despite the similarities in age between dominant and guard bees, however, we see that ovary size differences between dominant and guard bees explains a substantial amount of variation in module-three CHC relative abundances (Fig. 5).

**Transcriptomic data reveals links between CHCs, sociality, and reproductive physiology**

Correlations among DEGs and CHCs in module-three, particularly from the ovary gene expression data, suggest robust links among social behavior, reproductive physiology, and CHC expression. While relatively few genes (10) were previously found to be differentially expressed between dominant and subordinate ovaries, five of these showed significant correlation with the CHC module-three peaks and three of these five genes were also significantly correlated with ovary size. All three of these genes are associated with either transporting or metabolizing sugars and show highly correlated expression (Fig. S9). The first, facilitated trehalose transporter Tret1-like, is thought to regulate the tissue-specific uptake of trehalose, the primary "blood sugar" circulating in insect hemolymph (Shukla et al., 2015). The second gene, inositol oxygenase, is involved in metabolizing the sugar inositol (Parker et al., 2015). The expression of inositol oxygenase is strongly associated with caste-behavior in *Polistes* wasps (Jandt et al., 2017) and is also associated with larval caste-development and adult behavioral variation in honeybee workers (Hunt et al., 2010). Finally, UDP-glucoronosyltransferase 1-8 is an enzyme belonging to a gene family primarily known for detoxification, which is accomplished by catalysis of glycosidic reactions that increase the solubility of xenobiotics for excretion (Li et al., 2017). Nutrient sensing pathways are broadly implicated in caste-associated physiology across social insects (Kapheim, 2016). We note that our dataset cannot address whether the expression of these genes has any direct influence on CHC expression. Especially given that behavior, ovary size, and CHCs are interrelated it is not necessarily surprising that we also see correlated gene expression. Future work examining connections among social behavior, nutrition, and reproductive physiology will be necessary to understand how the subtle differences between dominant and subordinate physiology emerge.

**Comparison to other species**

The linear alkanes that compose the module-three CHCs in *E. dilemma* include several CHCs that are associated with the queen caste of multiple eusocial insects, including species of ants, wasps, and bees (Steitz et al., 2018; Van Oystaeyen et al., 2014). This is consistent with the idea that linear alkanes may be ancestrally linked to reproduction and have been coopted for social communication independently across Hymenoptera, including in the corbiculate bees (Oliveira et al., 2015; Oi et al, 2015). The CHC with the highest relative abundance among the module-three CHCs that we identify in *E. dilemma,* pentacosane (C25), shows evidence of queen signaling in multiple species of bumblebee and multiple species of stingless bees (Amsalem et al., 2015; Nunes et al., 2014; Oliveira et al., 2015). This may be an especially promising candidate for further investigation for its role in reproductive signaling in the evolutionary history of corbiculate bees as it has now been identified as associated with reproductive physiology in species of three of the four tribes forming the corbiculate bees.

Within orchid bees, our findings both overlap and contrast in several ways with the only other orchid bee species where CHCs and social status have been investigated, *Euglossa melanotricha* (Andrade-Silva and Nascimento, 2015). Four of the five linear alkanes we identify among the module-three CHCs are also significantly differentiated between dominants and subordinates of *E. melanotricha* (C25, C27, C29, C31), suggesting that their association with social behavior could be conserved in *Euglossa*.

Besides these chemical similarities, however, our results linking ovary size and CHCs in *E. dilemma* contrast with those from *E. melanotricha*, where no link between CHCs and ovary size was found. This could be due to differences in the biology of the two species,

methodological differences between the studies, or some combination of the two. *Euglossa melanotricha* is distinct from *E. dilemma* in showing high levels of aggression between dominants and subordinates (Andrade-Silva and Nascimento, 2012), while *E. dilemma* (and sister species *E. viridissima*) females show little aggression among nestmates (Cocom Pech et al., 2008; Saleh and Ramírez 2019). Given this, *E. dilemma* and *E. melanotricha* may have evolved different approaches to managing within-nest conflict and cooperation. Dominant bees in *Euglossa melanotricha*, for instance, may rely on egg policing and aggression to reinforce chemical dominance and *E. dilemma* dominant bees may rely on honest signaling of ovary size to encourage subordinate cooperation or retention in the nest. Alternatively, it is possible that our inclusion of the solitary behavioral phases, which show the largest shifts in ovary size, increased our power to detect CHC and ovary size correlations. Examination of additional orchid bee species is necessary to determine how widespread the link between ovary size and CHC expression is across the phylogeny of orchid bees.

**Are module-three CHCs social signals in *E. dilemma's* casteless social groups?**

We identified CHCs that are correlated with social behavior and reproductive physiology in *E. dilemma*, though it is unknown whether these CHCs function as social signals. In some eusocial insects, exposure to queen signals can lead to the inactivation of worker ovaries (Smith and Liebig, 2017). In addition, fertility signals can direct aggression towards egg-laying workers (Smith et al., 2009). In some primitively eusocial species, however, CHCs or other compounds may correlate with dominance and ovary size but lack the ability to regulate reproduction (Oi et al., 2019). In these cases, CHCs can still be relevant to social interactions, but they may only function within specific chemical and behavioral contexts (Mora-Kepfer, 2014; Smith, et al,

2015). In *E. dilemma*, dominants and subordinates both have activated ovaries and aggression is rarely observed among nestmates. Furthermore, subordinate ovary size is equivalent to foundress ovary size and dominant ovary size is slightly larger than either (Saleh and Ramírez, 2019). This observation suggests that, though dominant bees show elevated ovary size, it is unlikely that subordinate bees are suppressed physiologically via the production of pheromones emitted by dominant bees. In addition, oophagy of subordinate eggs by dominants is not selective in *E. dilemma*, with dominant bees appearing to eat all subordinate laid eggs (Saleh and Ramírez, 2019).

Given these life-history characteristics, if CHCs are indeed serving as social signals, one possibility is that they function more to advertise dominant quality than to control subordinate behavior. It is unknown what factors determine whether a newly emerged *E. dilemma* female will stay in the nest or disperse. CHCs could be used by newly emerged females to assess the dominant individual as a part of a disperse/stay decision making process. Newly emerged *Polistes* wasps, for example, incorporate information about their nesting environment to inform dispersal decisions (Tibbetts, 2007). If dominant bees have the largest ovary size on average, there could be a greater fitness advantage for newly emerged females to remain in the nest and raise sisters laid by a highly fecund mother rather than expending energy on nest initiation. Alternatively, aggression may be rare in *E. dilemma* social groups precisely because dominant females honestly signal larger ovary size. This appears to be the case in the primitively eusocial wasp *Ropalida marginata*, for instance, where a reduction in the quantity of queen-specific pheromones stimulates aggression between otherwise non-aggressive queens and workers (Saha et al., 2012). Manipulation of CHC profiles in dominants and subordinates could therefore reveal latent aggression. Further experiments are needed to evaluate these hypotheses.

Finally, it is also possible that the CHCs in module-three do not function as signals, but instead serve as cues that, while linked to behavior and physiology, do not elicit a behavioral response shaped by natural selection. If this is the case, it would strongly support a scenario in which social signals, particularly in the corbiculate bees, began as fertility-linked cues already sensitive to physiology and social environment that subsequently evolved into signaling molecules along with a more elaborate social organization.

## Conclusions

In addition to documenting a CHC polymorphism across populations, we identify a set of CHCs largely independent of this polymorphism that correlate with ovary size and dominance status in *E. dilemma* social groups. Experimental manipulation of these CHCs in social settings is a necessary next step in assessing the role of CHCs in mediating conflict and cooperation in *E. dilemma* nests. However, these data show that, even in small casteless social groups of two or three individuals, CHCs can reflect ovary size and dominance status in a similar fashion to the information encoded in highly eusocial Hymenoptera that display large social colonies with specialized castes. This is especially interesting considering the phylogenetic position of orchid bees as sister to the rest of the eusocial corbiculate bees. These data further motivate increased sampling of CHCs among orchid bees and the other corbiculate bees, to better understand the role of CHCs in the evolution of communication in social insects.

## Acknowledgements

We thank the members of the Ramírez lab and Cheryl Dean for their help and advice throughout the project. We thank Nikki Hochberg, Ron Phenix, and all the staff at Fern Forest Nature Center for their support of this project. We also thank Joe Patt and Aleena Tarshish Moreno for their help. We also thank the editors and reviewers for their insightful comments on earlier drafts. The project was supported by the David and Lucile Packard Foundation (to S.R.R.), the NSF (to S.R.R., DEB-1457753), and the Daphne and Ted Pengelley Award (UC Davis to N.W.S.).

## Statement of Authorship

NWS and SRR initially designed the project. All authors participated in sample collection. NWS and KH performed sample processing and data analysis. All authors participated in writing and revising the manuscript.

## Data and Code Accessibility

Data and R code required to reproduce all figures and analyses in the manuscript are available on Dryad (https://doi.org/10.25338/B8FH0P). RNA sequence data is available at NCBI Bioproject accession prjna523381 and DNA data for genotyping is available at NCBI Bioproject accession prjna623571.

## Literature Cited

Amsalem, E., Orlova, M., and Grozinger, C. M. 2015. A conserved class of queen pheromones? Re-evaluating the evidence in bumblebees (*Bombus impatiens*). Proceedings of the Royal Society of London B, 282(1817), 20151800-. https://doi.org/10.1098/rspb.2015.1800

Andrade-Silva, A. C. R., and Nascimento, F. S. 2015. Reproductive regulation in an orchid bee: social context, fertility and chemical signalling. Animal Behaviour, 106, 43–49. https://doi.org/10.1016/j.anbehav.2015.05.004

Andrade-Silva, A.C.R, and Nascimento, F. S. 2012. Multifemale nests and social behavior in Euglossa melanotricha (Hymenoptera, Apidae, Euglossini). Journal of Hymenoptera Research, 26(0), 1. https://doi.org/10.3897/jhr.26.1957

Andrade-Silva, A. C. R., Miranda, E. A., Del Lama, M. A., and Nascimento, F. S. 2016. Reproductive concessions between related and unrelated members promote eusociality in bees. Scientific Reports, 6, 26635. https://doi.org/10.1038/srep26635

Bonduriansky, R., Mallet, M. A., Arbuthnott, D., Pawlosky-Glahn, V., Egozcue, J. J., and Rundle, H. D. 2015. Differential effects of genetic vs . environmental quality in *Drosophila melanogaster* suggest multiple forms of condition dependence. Ecology Letters, (18), 317–326. https://doi.org/10.1111/ele.12412

Bonelli, M., and Lorenzi, M. C. 2014. Population Diversity in Cuticular Hydrocarbons and mtDNA in a Mountain Population Diversity in Cuticular Hydrocarbons and mtDNA in a Mountain Social Wasp. Journal of Chemical Ecology, 41:22-31. https://doi.org/10.1007/s10886-014-0531-0

Brand, P., Hinojosa-Díaz, I. A., Ayala, R., Daigle, M., Yurrita Obiols, C. L., Eltz, T., and Ramírez, S. R. 2020. The evolution of sexual signaling is linked to odorant receptor tuning in perfume-collecting orchid bees. Nature Communications, 11(1). https://doi.org/10.1038/s41467-019-14162-6

Brand, P., Saleh, N., Pan, H., Li, C., Kapheim, K. M., and Ramírez, S. R. 2017. The Nuclear and Mitochondrial Genomes of the Facultatively Eusocial Orchid Bee Euglossa dilemma. G3:

Genes, Genomes,Genetics, 7, 2891-2898, g3.117.043687. https://doi.org/10.1534/g3.117.043687

Bruschini, C., Cervo, R., Cini, A., Pieraccini, G., Pontieri, L., Signorotti, L., Pieraccini, V. 2011. Cuticular Hydrocarbons Rather Than Peptides Are Responsible for Nestmate Recognition in Polistes dominulus. Chemical Senses, 36, 715–723. https://doi.org/10.1093/chemse/bjr042

Cane, J. H., Journal, S., Entomological, K., Jan, N., and Cane, J. H. 2020. Estimation of Bee Size Using Intertegular Span (Apoidea). Kansas Entomological Society, 60(1), 145–147.

Cappa, F., Bruschini, C., Cervo, R., Turillazzi, S., and Beani, L. 2013. Males do not like the working class : male sexual preference and recognition of functional castes in a primitively eusocial wasp. Animal Behaviour, 86, 801–810. https://doi.org/https://doi.org/10.1016/j.anbehav.2013.07.020

Catchen, J., Hohenlohe, P. A., Bassham, S., and Amores, A. 2013. Stacks : an analysis tool set for population genomics, Molecular Ecology, 22(11) 3124–3140. https://doi.org/10.1111/mec.12354

Choe, D. H., Ramírez, S. R., and Tsutsui, N. D. 2012. A Silica Gel Based Method for Extracting Insect Surface Hydrocarbons. Journal of Chemical Ecology, 38(2), 176–187. https://doi.org/10.1007/s10886-012-0074-1

Cocom Pech, M. E., May-Itzá, W. D. J., Medina Medina, L. A., and Quezada-Euán, J. J. G. 2008. Sociality in *Euglossa* (*Euglossa*) *viridissima* Friese (Hymenoptera, Apidae, Euglossini). Insectes Sociaux, 55(4), 428–433. https://doi.org/10.1007/s00040-008-1023-4

Cotton, S., Fowler, K., and Pomiankowski, A. 2004. Do sexual ornaments demonstrate heightened condition-dependent expression as predicted by the handicap hypothesis? Proceedings of the Royal Society B, 271(1541), 771–783. https://doi.org/10.1098/rspb.2004.2688

Coyne, J.A., Wicker-Thomas, C., Jallon, J. 1999. A gene responsible for a cuticular hydrocarbon polymorphism in *Drosophila melanogaster*. Genetics Research, 73, 189–203. https://doi.org/10.1017/S0016672398003723

Douglas, C. E., and Michael, F. A. 1991. On distribution-free multiple comparisons in the one-way analysis of variance. Communications in Statistics, Theory and Methods, 20(1), 127–139. https://doi.org/https://doi.org/10.1080/03610929108830487

Dronnet, S., Lohou, C., Christides, J., and Bagneres, A.G. 2006. Cuticular Hydrocarbon Composition Reflects Genetic Relationship Among Colonies of the Introduced Termite *Reticulitermes santonensis* Feytaud. Journal of Chemical Ecology, (32), 1027–1042. https://doi.org/10.1007/s10886-006-9043-x

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. 2011. A Robust , Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species, PLOS One, 6(5), 1–10. https://doi.org/10.1371/journal.pone.0019379

Friedel, A., Lattorff, H. M. G., Quezada-Euán, J.J.G., and Boff, S. 2020. Shared reproduction and sex ratio adjustment to clutch size in a socially polymorphic orchid bee. Ethology, 88–96. https://doi.org/10.1111/eth.12963

Gershman, S. N., and Rundle, H. D. 2016. Level up : the expression of male sexually selected cuticular hydrocarbons is mediated by sexual experience. Animal Behaviour, 112, 169–177. https://doi.org/10.1016/j.anbehav.2015.11.025

Gibbs, A., and Pomonis, J. G. 1995. Physical properties of insect cuticular hydrocarbons: The effects of chain length, methyl-branching and unsaturation. Comparative Biochemistry and Physiology Part B, 112(2), 243–249. https://doi.org/10.1016/0305-0491(95)00081-X

Holman, L. 2018. Queen pheromones and reproductive division of labor : a meta-analysis. Behavioral Ecology, 29, 1199–1209. https://doi.org/10.1093/beheco/ary023

Howard, R. W., McDaniel, C. A., Nelson, D. R., Blomquist, G. J., Gelbaum, L. T., and Zalkow, L. H. 1982. Cuticular hydrocarbons of *Reticulitermes virginicus* and their role as potential species and caste recognition cues. Journal of Chemical Ecology, 8(9), 1227–1238.

Hunt, J. H., Wolschin, F., Henshaw, M. T., Newman, T. C., Toth, A. L., and Amdam, G.V. 2010. Differential Gene Expression and Protein Abundance Evince Ontogenetic Bias toward Castes in a Primitively Eusocial Wasp. PLOS One, 5(5). https://doi.org/10.1371/journal.pone.0010674

Izzo, A., Wells, M., and Huang, Z. 2010. Cuticular hydrocarbons correlate with fertility , not dominance , in a paper wasp , Polistes dominulus. Behavioral Ecology and Sociobiology, 857–864. https://doi.org/10.1007/s00265-010-0902-7

Jandt, J. M., Suryanarayanan, S., Hermanson, J. C., Jeanne, R. L., and Toth, A. L. 2017. Maternal and nourishment factors interact to influence offspring developmental trajectories in social wasps. Proceedings of the Royal Society of London B, 284 (1857). https://doi.org/10.1098/rspb.2017.0651

Kapheim, K. M. 2016. Nutritional, endocrine, and social influences on reproductive physiology at the origins of social behavior Reproductive. Current Opinion in Insect Science, 22, 62–70. https://doi.org/10.1016/j.cois.2017.05.018

Kuo, T., Fedina, T. Y., Hansen, I., Dreisewerd, K., Dierick, H. A., Yew, J. Y., and Pletcher, S. D. 2012. Insulin Signaling Mediates Sexual Attractiveness in *Drosophila*. PLOS Genetics, 8(4). https://doi.org/10.1371/journal.pgen.1002684

Leonhardt, S. D., Menzel, F., Nehring, V., and Schmitt, T. 2016. Ecology and Evolution of Communication in Social Insects. Cell, 164(6), 1277–1287. https://doi.org/10.1016/j.cell.2016.01.035

Li, H. 2013. Aligning sequence reads , clone sequences and assembly contigs with BWA-MEM, arXiv:1303.3997.

Li, X., Shi, H., Gao, X., and Liang, P. 2017. Characterization of UDP-glucuronosyltransferase genes and their possible roles in multi-insecticide resistance in *Plutella xylostella* ( L .),. Pest Management Science, 74(3). https://doi.org/10.1002/ps.4765

Luo, Y., Zhang, Y., Farine, J., Ferveur, J., Ramírez, S., and Kopp, A. 2019. Evolution of sexually dimorphic pheromone profiles coincides with increased number of male-specific chemosensory organs in *Drosophila prolongata*. Ecology and Evolution, 13608–13618. https://doi.org/10.1002/ece3.5819

Makki, R., Cinnamon, E., and Gould, A. P. 2014. The Development and Functions of Oenocytes. Annual Review of Entomology, 59, 405–425. https://doi.org/10.1146/annurev-ento-011613-162056

Martin, S. J., and Drijfhout, F. P. 2009. How Reliable is the Analysis of Complex Cuticular Hydrocarbon Profiles by Multivariate Statistical Methods? Journal of Chemical Ecology, (35), 375–382. https://doi.org/10.1007/s10886-009-9610-z

Martin, S. J., Zhong, W., and Drijfhout, F. P. 2009. Long-term stability of hornet cuticular hydrocarbons facilitates chemotaxonomy using museum specimens. Biological Journal of the Linnean Society, 96(4), 732–737. https://doi.org/10.1111/j.1095-8312.2008.01158.x

Mora-Kepfer, F. 2014. Context-dependent acceptance of non-nestmates in a primitively eusocial insect. Behavioral Ecology and Sociobiology, 68(3), 363–371. https://doi.org/10.1007/s00265-013-1650-2

Nunes, T. M., Mateus, S., Favaris, A. P., Amaral, M. F. Z. J., von Zuben, L. G., Clososki, G. C., Lopes, N. P. 2014. Queen signals in a stingless bee: Suppression of worker ovary activation and spatial distribution of active compounds. Scientific Reports, 4, 7449. https://doi.org/10.1038/srep07449

Oi, C. A., Oliveira, R. C., van Zweden, J. S., Mateus, S., Millar, J. G., Nascimento, F. S., and Wenseleers, T. 2019. Do primitively eusocial wasps use queen pheromones to regulate reproduction? A case study of the paper wasp *Polistes satan*. Frontiers in Ecology and Evolution, 7(JUN), 1–10. https://doi.org/10.3389/fevo.2019.00199

Oi, C. A., van Zweden, J. S., Oliveira, R. C., Van Oystaeyen, A., Nascimento, F. S., and Wenseleers, T. 2015. The origin and evolution of social insect queen pheromones: Novel hypotheses and outstanding problems. BioEssays, 37(7), 808–821. https://doi.org/10.1002/bies.201400180

Oldroyd, B. P. 2018. Queen pheromone: contraceptive or a queen presence signal?—A comment on Holman. Behavioral Ecology, (1954), 1213–1214. https://doi.org/doi:10.1093/beheco/ary048

Oliveira, R.C., Oi, C. A., do Nascimento, M. M. C., Vollet-Neto, A., Alves, D. A., Campos, M. C., Wenseleers, T. 2015. The origin and evolution of queen and fertility signals in Corbiculate bees. BMC Evolutionary Biology, 15(1), 254. https://doi.org/10.1186/s12862-015-0509-8

Otte, T., Hilker, M., and Geiselhardt, S. 2018. Phenotypic Plasticity of Cuticular Hydrocarbon Profiles in Insects. Journal of Chemical Ecology, 44, 235–247. https://doi.org/https://doi.org/10.1007/s10886-018-0934-4

Oystaeyen, A. Van, Oliveira, R. C., Holman, L., Van Zweden, J. S., Romero, C., Oi, C. A., … Wenseleers, T. 2014. Conserved Class of Queen Pheromones. Science, 287, 287–291. https://doi.org/10.1126/science.1244899

Parker, D. J., Vesala, L., Ritchie, M. G., Laiho, A., Hoikkala, A., and Kankare, M. 2015. How consistent are the transcriptome changes associated with cold acclimation in two species of the *Drosophila virilis* group? Heredity, 115(1), 13–21. https://doi.org/10.1038/hdy.2015.6

Pokorny, T., Vogler, I., Losch, R., Schlutting, P., Juarez, P., Bissantz, N., Eltz, T. 2017. Blown by the wind : the ecology of male courtship display behavior in orchid bees. Ecology, 98(4), 1140–1152. https://doi.org/10.1002/ecy.1755

Pokorny, T., Ramírez, S. R., Weber, M. G., and Eltz, T. 2015. Cuticular Hydrocarbons as Potential Close Range Recognition Cues in Orchid Bees. Journal of Chemical Ecology. https://doi.org/10.1007/s10886-015-0647-x

Pokorny, T., Lunau, K., Quezada-Euán, J. J. G., and Eltz, T. 2014. Original article Cuticular hydrocarbons distinguish cryptic sibling species in *Euglossa* orchid bees. Apidologie, 45, 276–283. https://doi.org/10.1007/s13592-013-0250-5

Saha, P., Balasubramaniam, K. N., Kalyani, J. N., Supriya, K., Padmanabhan, A., and Gadagkar, R. 2012. Clinging to royalty: *Ropalidia marginata* queens can employ both pheromone and aggression. Insectes Sociaux, 59(1), 41–44. https://doi.org/10.1007/s00040-011-0185-7

Saleh, N. W., and Ramírez, S. R. 2019.. Sociality emerges from solitary behaviours and reproductive plasticity in the orchid bee *Euglossa dilemma*. Proceedings of the Royal Society of London B, 286(1906). https://doi.org/10.1098/rspb.2019.0588

Saleh, N.W., Hodgson, K., Pokorny, T., Mullins, A., Chouvenc, T., Eltz, T., Ramírez, S. R. 2021. Data from: Social behavior, ovary size, and population of origin influence cuticular hydrocarbons in the orchid bee Euglossa dilemma. American Naturalist, Dryad digital repository, https://doi.org/10.25338/B8FH0P.

Shukla, E., Thorat, L. J., Nath, B. B., and Gaikwad, S. M. 2015. Insect trehalase : Physiological significance and potential applications. Glycobiology, 25(4), 357–367. https://doi.org/10.1093/glycob/cwu125

Skov, C., and Wiley, J. 2005. Establishment of the Neotropical Orchid Bee *Euglossa Viridissima* (Hymenoptera: Apidae) in Florida. Florida Entomologist, 88(2), 225–227. https://doi.org/10.1653/0015-4040(2005)088[0225:EOTNOB]2.0.CO;2

Smith, A. A., Hölldober, B., and Liebig, J. 2009. Cuticular Hydrocarbons Reliably Identify Cheaters and Allow Enforcement of Altruism in a Social Insect. Current Biology, 19(1), 78–81. https://doi.org/10.1016/j.cub.2008.11.059

Smith, A. A., and Liebig, J. 2017. The evolution of cuticular fertility signals in eusocial insects. Current Opinion in Insect Science. https://doi.org/10.1016/j.cois.2017.05.017

Smith, A. A., Millar, J. G., and Suarez, A. V. 2015. A social insect fertility signal is dependent on chemical context. Biology Letters, 11(1), 3–6. https://doi.org/10.1098/rsbl.2014.0947

Solano-Brenes, D., Otarola, M. F., and Hanson, P. E. 2018. Nest initiation by multiple females in an aerial-nesting orchid bee, *Euglossa cybelia* (Apidae: Euglossini). Apidologie, (49), 807–816. https://doi.org/10.1007/s13592-018-0605-z

Soucy, S. L., Giray, T., and Roubik, D. W. 2003. Solitary and group nesting in the orchid bee *Euglossa hyacinthina* (Hymenoptera, Apidae). Insectes Sociaux, 50, 248–255. https://doi.org/10.1007/s00040-003-0670-8

Steiger, S., and Stökl, J. 2018. Pheromones Regulating Reproduction in Subsocial Beetles : Insights with References to Eusocial Insects. Journal of Chemical Ecology, (44), 785–795.

Steitz, I., Kingwell, C. J., Paxton, R. J., and Ayasse, M. 2018. Evolution of Caste-Specific Chemical Profiles in Halictid Bees. Journal of Chemical Ecology, (44), 827–837. https://doi.org/https://doi.org/10.1007/s10886-018-0991-8

Tibbetts, E. A. 2007. Dispersal decisions and predispersal behavior in *Polistes* paper wasp "workers." Behavioral Ecology and Sociobiology, 61(12), 1877–1883. https://doi.org/10.1007/s00265-007-0427-x

Walsh, J., Pontieri, L., d'Ettorre, P., and Linksvayer, T. A. 2020. Ant cuticular hydrocarbons are heritable and associated with variation in colony productivity. Proceedings of the Royal Society of London B, 287(1928), 20201029. https://doi.org/10.1098/rspb.2020.1029

Wang, J. 2014. Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. Journal of Evolutionary Biology, 27, 518–530. https://doi.org/10.1111/jeb.12315

Wang, J. 2017. Estimating pairwise relatedness in a small sample of individuals. Heredity, 119(5), 302–313. https://doi.org/10.1038/hdy.2017.52

Wang, Jinliang. 2011. COANCESTRY : a program for simulating, estimating and analysing relatedness and inbreeding coefficients. Molecular Ecology Resources, 141–145. https://doi.org/10.1111/j.1755-0998.2010.02885.x

Wilson, E. O. 1965. Chemical Communication in the Social Insects. Science, 149(3688), 1064–1071.

**References Cited Only in the Online Enhancements**

Arriaga-Osnaya, B. J., Contreras-Garduño, J., Espinosa-García, F. J., García-Rodríguez, Y. M., Moreno-García, M., Lanz-Mendoza, H., Cueva del Castillo, R. 2017. Are body size and volatile blends honest signals in orchid bees? Ecology and Evolution, 7(9), 3037–3045. https://doi.org/10.1002/ece3.2903

Cuvillier-Hot, V., Cobb, M., Malosse, C., & Peeters, C. 2001. Sex, age and ovarian activity affect cuticular hydrocarbons in *Diacamma ceylonense*, a queenless ant. Journal of Insect Physiology, 47(4–5), 485–493. https://doi.org/10.1016/S0022-1910(00)00137-2

Kelstrup, H. C., Hartfelder, K., Lopes, T. F., & Wossler, T. C. 2018. The Behavior and Reproductive Physiology of a Solitary Progressive Provisioning Vespid Wasp : Evidence for a Solitary-Cycle Origin of Reproductive Castes, 191(2). https://doi.org/10.1086/695336

Lacey, E. S., Ginzel, M. D., Millar, J. G., & Hanks, L. M. 2008. 7-methylheptacosane is a major component of the contact sex pheromone of the cerambycid beetle *Neoclytus acuminatus*

*acuminatus*. Physiological Entomology, 33(3), 209–216. https://doi.org/10.1111/j.1365-3032.2008.00624.x

Monnin, T., Malusse, C., & Peeters, C. 1998. Solid-phase microextraction and cuticular hydrocarbon differences related to reproductive activity in queenless ant *Dinoponera quadriceps*. Journal of Chemical Ecology, 24(3), 473–490. https://doi.org/10.1023/A:1022360718870

Turillazzi, S., Sledge, M. F., & Moneti, G. 1998. Use of a simple method for sampling cuticular hydrocarbons from live social wasps. Ethology Ecology & Evolution, 10(3), 293–297. https://doi.org/10.1080/08927014.1998.9522859

## Figure Legends

**Figure 1.** NMDS plot of CHC variation among 214 individuals across populations and sexes of *E. dilemma* and *E. viridissima*. Each color/symbol combination represents individuals of one sex from one population. Individuals fall mostly into clusters labeled "Chemotype-A" and "Chemotype-B." Stress value for NMDS configuration = 0.052.

**Figure 2.** NMDS plot of CHC variation across *E. dilemma* female behavioral groups in the Florida population. Unique color and symbol combinations represent the different behavioral groups. Individuals fall into clusters labeled "Chemotype-A" and "Chemotype-B." Stress value for NMDS configuration = 0.042.

**Figure 3.** Correlation structure among CHCs in *E. dilemma* FL females. The size and color of the circle indicate the strength and direction of the correlation between two CHCs, respectively. Larger circles indicate a higher Spearman's rho value and dark blue represents a positive correlation and dark red represents a negative correlation. CHCs have been ordered into three modules (black rectangles) using hierarchical clustering based on the Ward.D2 clustering method. CHCs labeled with a colon are alkenes and those with an "a" or "b" indicate that they are alkenes of the same carbon chain length with a different position for the double bond.

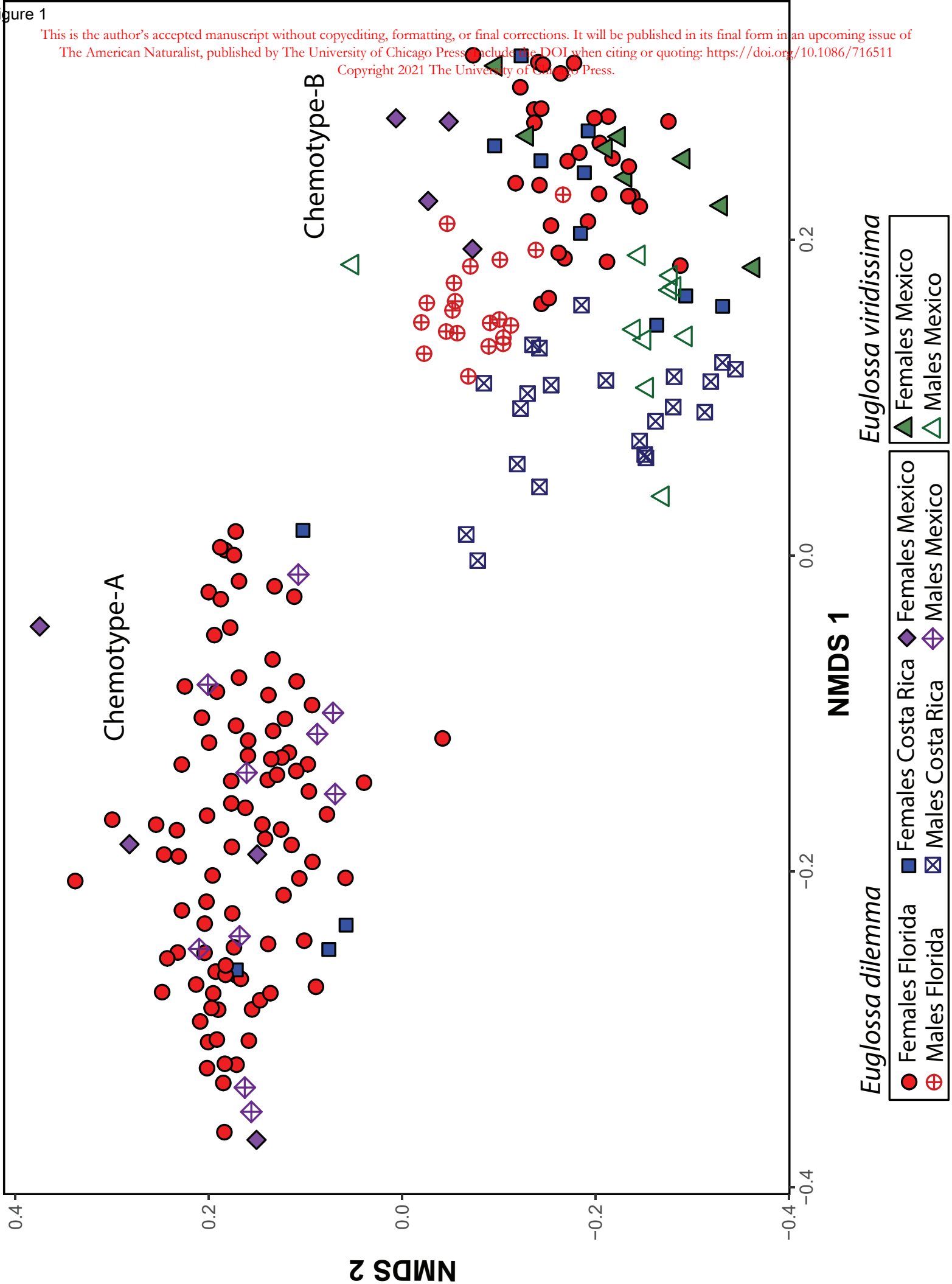
**Figure 4.** Relative abundance of module-three CHCs (C25, C26, C27, C29, C31) across behaviors. Letters denote statistical groupings determined by a Tukey HSD test and the p-value is calculated using a one-way ANOVA. The box plots show the mean value in each group.

**Figure 5.** Correlation between ovary size and the module-three (C25, C26, C27, C29, C31) CHC relative abundances across behaviors. Unique shape and color combinations correspond to the four sampled behavioral groups. The correlation coefficient and p-value were calculated with Pearson's r.

**Figure 6.** Correlation between ovary expression of facilitated transporter Tret 1-like and the relative abundances of the module-three (C25, C26, C27, C29, C31) CHCs across behavioral groups. The x-axis shows log2 expression levels for each individual relative to the mean value for that gene across individuals. Unique shape and color combinations correspond to the four sampled behavioral groups. The correlation coefficient and FDR adjusted p-value were calculated with Spearman's rho.

Figure 1





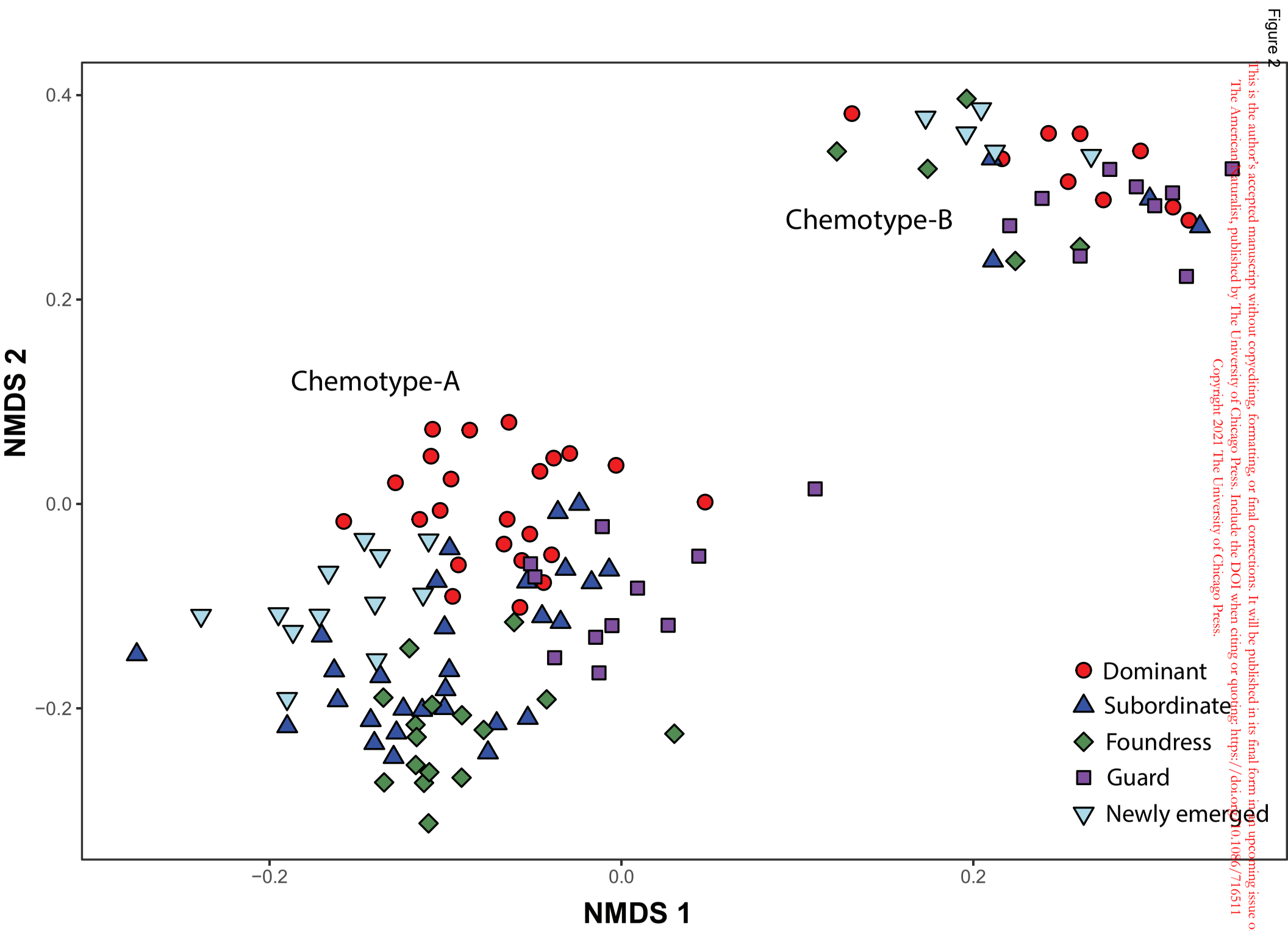



Figure 2

Figure 3



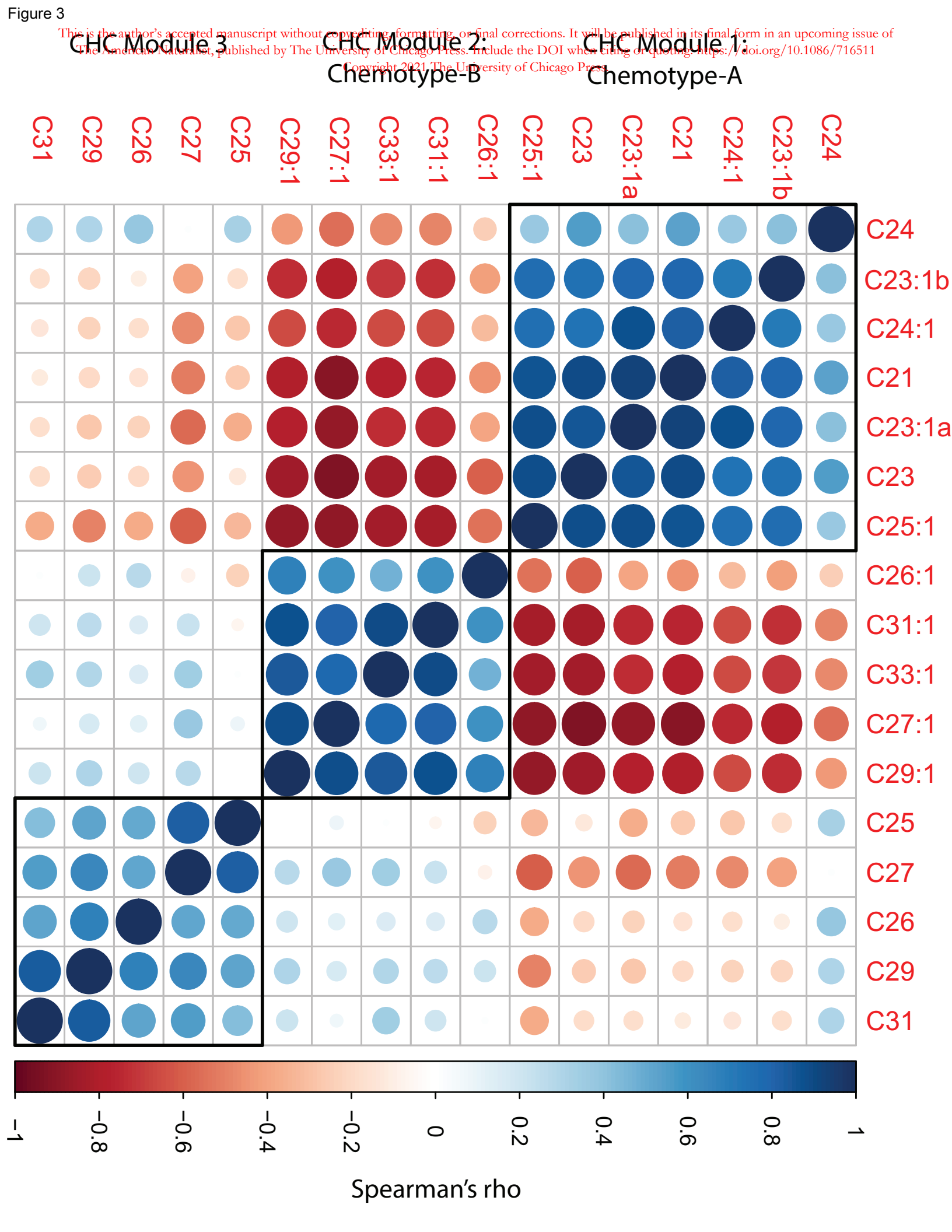


Figure 4



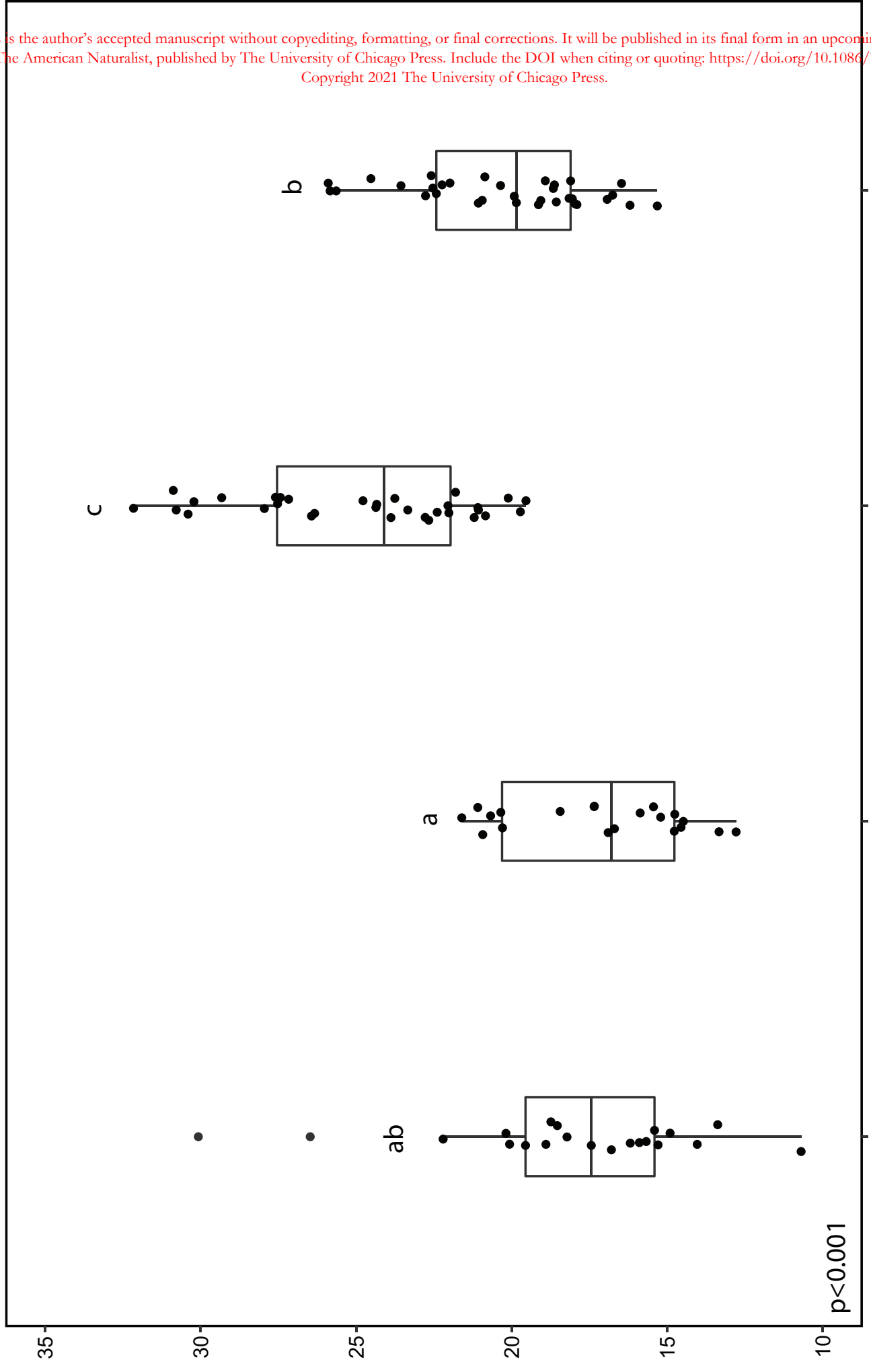


Figure 5



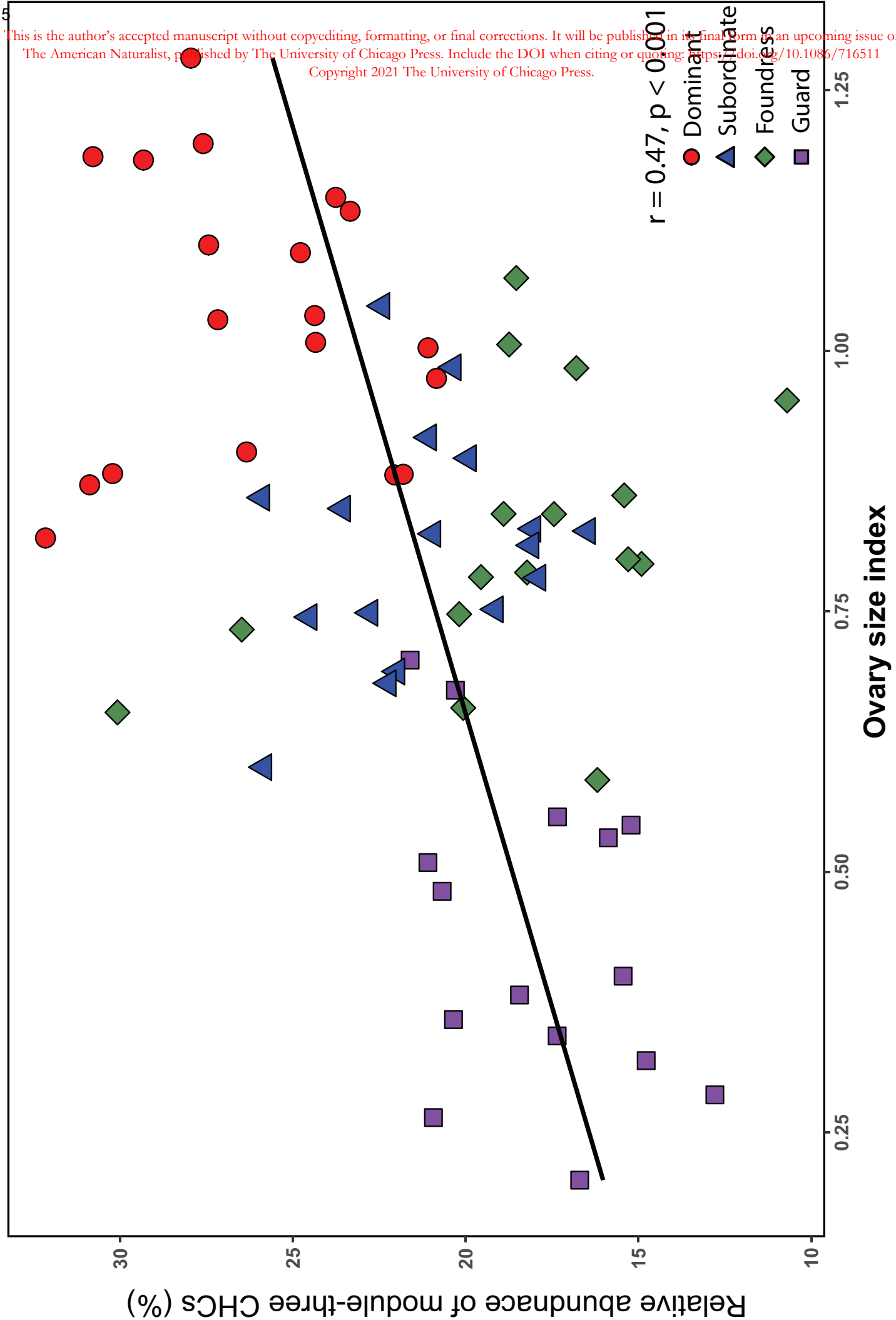
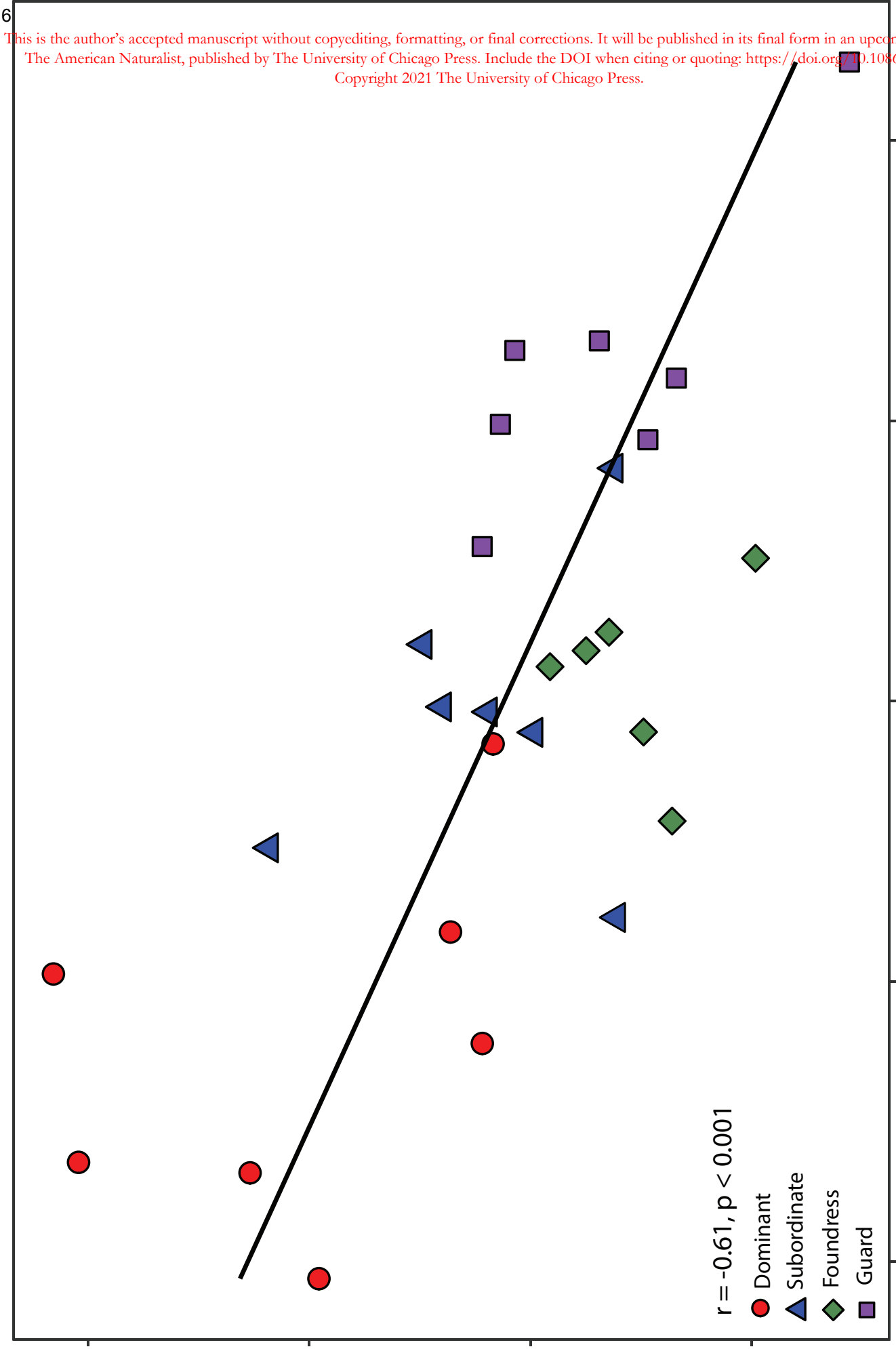


Figure 6





**The American Naturalist**

# Social behavior, ovary size, and population of origin influence cuticular hydrocarbons in the orchid bee, *Euglossa dilemma*

Nicholas W. Saleh[1], Kirsten Hodgson[1], Tamara Pokorny[2], Aaron Mullins[3], Thomas Chouvenc[3], Thomas Eltz[4], Santiago R. Ramírez[1]

[1] Center for Population Biology, University of California, Davis, CA, USA

[2] Institute of Zoology, University of Regensburg, Regensburg, Germany

[3] Entomology and Nematology Department, Ft. Lauderdale Research and Education Center, Institute of Food and Agricultural Sciences, University of Florida, FL, USA

[4] Department of Animal Ecology, Evolution and Biodiversity, Ruhr University Bochum, Bochum, Germany

Corresponding author: Nicholas Saleh, nsaleh@ucdavis.edu

**Supplemental File 1.**

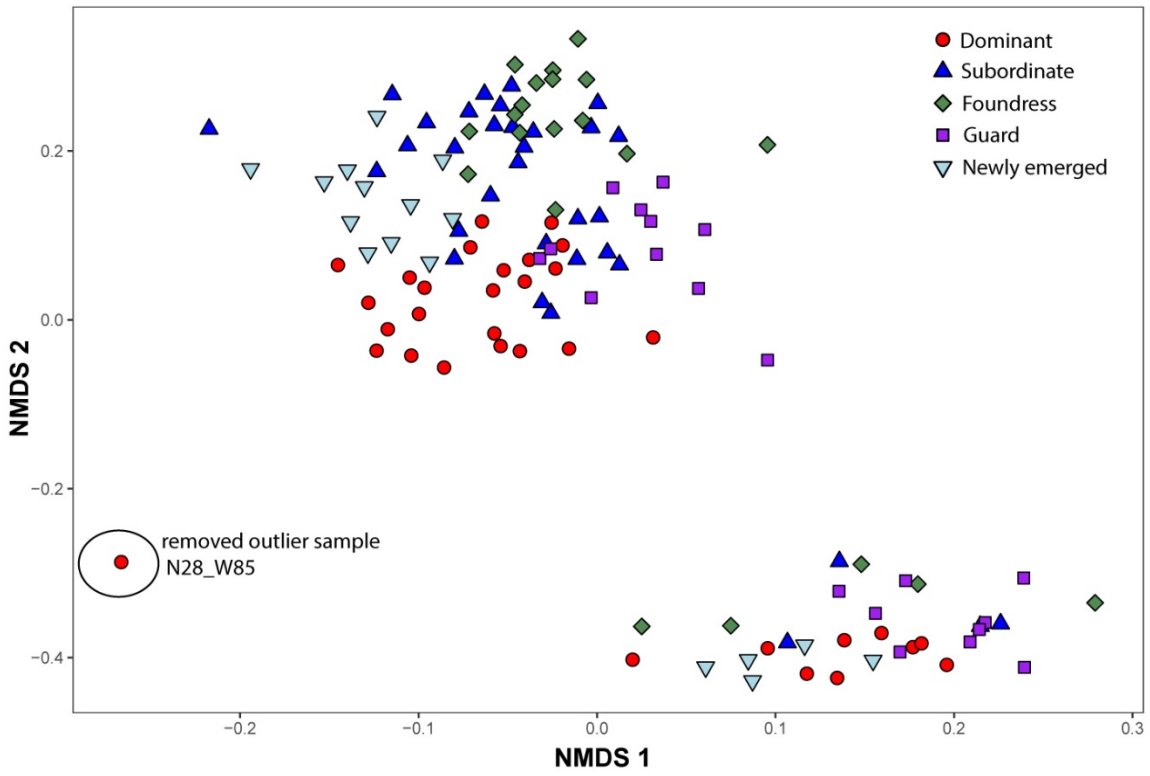



**Figure S1.** NMDS plot of CHC variation across *E. dilemma* female behavioral groups in the Florida population with outlier individual included. Unique color and symbol combinations represent the different behavioral groups. Stress value for NMDS configuration = 0.042.

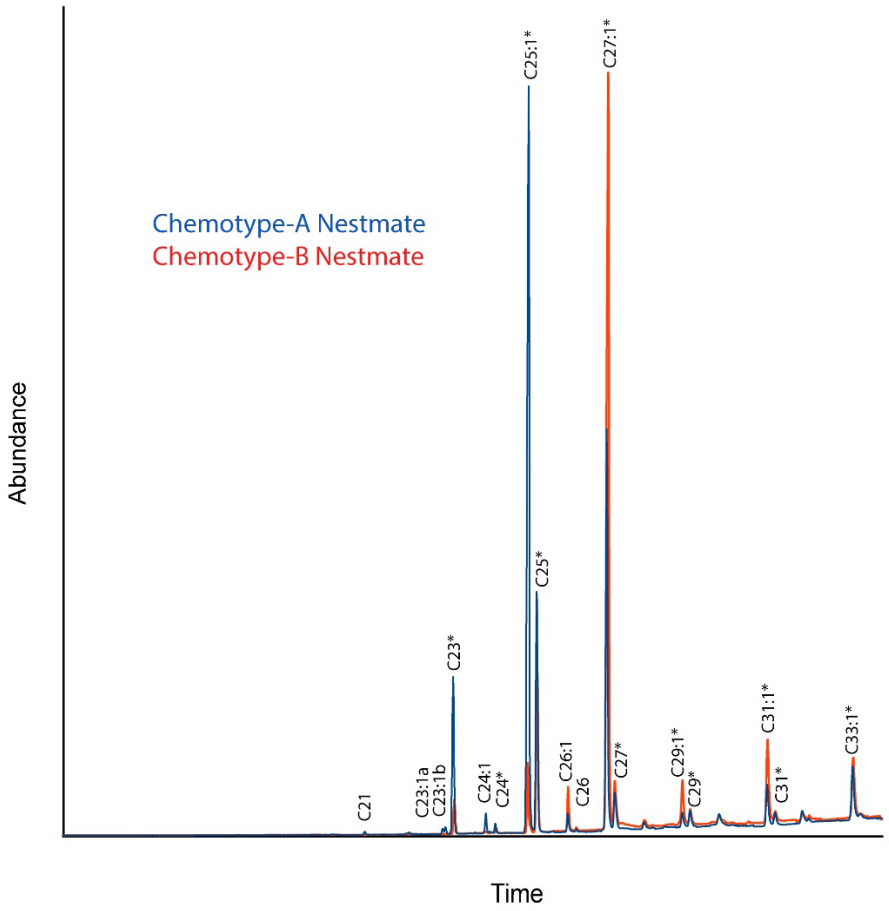



**Figure S2.** Overlaid chromatograms from two *E. dilemma* adult females collected from the same nest in Florida. CHC peaks are labeled with their identity. Unlabeled peaks were not identified consistently as CHCs and not included in analysis. Asterisks indicate that this CHC was one of 11 CHCs used in the analysis across sexes, populations, and species. All 17 detected CHCs were used to analyze FL female samples. Blue and Red lines indicate the chromatograms from separate individuals, with the blue individual expressing chemotype-A and the red individual expressing chemotype-B.

**Appendix 1.** Comparison of wing, body, and abdomen CHC extractions.

Here we evaluate the degree to which wing CHC extractions (the primary source of data in this study) are representative of the cuticular surface by comparing CHCs extracted from the wings, the dorsal abdominal surface, and the whole-body from the same individuals.

Methods

For this comparison, we used 16 *E. dilemma* females that had been removed from nests collected from Ft. Lauderdale, FL, and then maintained together in a single mesh cage in a growth chamber at 27°C and 60% humidity on a 12hr light/dark cycle. Honey-water was provided ad-libitum in the cage and individuals were sampled 4-8 weeks after removal from nests. Individuals were frozen in a -20°C freezer until CHC extraction.

Three types of extractions were conducted on the set of 16 individuals. First, wings were removed from all individuals and extractions from the wings were conducted according to the methods detailed in the main manuscript text. Next, to sample the dorsal abdominal surface, we adapted methods from Turillazzi et al., 1998, using sterilized cotton applicators to extract CHCs. For each individual extraction, a sealed cotton applicator was removed from its packaging, submerged for five seconds in hexane, and partially air-dried for 10 seconds in a fume hood. Following this, the tip of the cotton applicator was rubbed over the dorsal surface of the abdomen, from the stinger to the point of attachment near the thorax for 60 seconds. After 60 seconds, the tip of the cotton applicator was cut off and placed in a glass scintillation vial. 100µl of hexane was added to the vial and it was swirled occasionally, soaking for 10 minutes. Next, the 100µl of hexane was transferred to a GCMS vial and allowed to completely evaporate overnight in a closed fume hood. The next day, 20µl of hexane was added to concentrate the extract. To sample the entire body, wingless individuals were placed into glass scintillation vials with 500µl of hexane and swirled occasionally for 10 minutes. This extract was then transferred to a GCMS vial.

The three types of extracts were run together using the same GCMS program described in the manuscript text. However, given that the amount of CHCs differed among the three extraction types, we adjusted the injection volume and split/splitless approach for each extraction

method so that the ion counts among sample types was more comparable. Wings and abdomen extractions were both run splitless with injection volumes of 2µl. The whole-body extraction, which yielded a higher concentration of CHCs than the other two extraction types, was run with a 5 to 1 split on a 1µl injection volume.

GCMS chromatograms were processed and analyzed according to the approaches in the manuscript text. We used NMDS plots to visualize the overall relationship among samples and pairwise PERMANOVA with FDR p-value adjustment to assess statistical differences among the methods, considering the entire set of CHCs. We also assessed individual CHC differences by performing Kruskal-Wallis tests with Steel-Dwass multiple comparisons (see main text for citations) on the relative abundance of individual peaks.

Results

Data was successfully generated using all three extraction methods, though the abdomen extraction failed for one individual (abdomen n = 15, wings n = 16, whole-body n = 16). All three extraction types yielded quantitatively comparable results. No compounds unique to extraction type were found in notable abundance (>1% of the largest peak area) and no unique compounds were found in greater than 50% of the samples from a given extraction method. Consequently, we proceeded with analysis of the 17 CHCs discussed in the main manuscript text, which were also identified in the extractions here.

NMDS analysis (Fig. S3) showed that chemotypes A and B were detected regardless of extraction type. In addition, an individual's chemotype assignment was consistent among all three extraction methods. PERMANOVA analysis of the dissimilarity matrix underlying the NMDS plot showed that Wing and Abdomen extractions are statistically indistinguishable (p =

0.42) while Wing and Whole-body ($p = 0.04$) and Abdomen and Whole-body ($p = 0.041$) extractions are significantly different.

Analysis among specific peaks showed that differences are found in all three methods, with these differences most pronounced between the whole-body extraction and the other two methods (Fig. S4). Considering the Abdomen versus the Wings extraction, 3 of 17 peaks show statistically significant differences. Considering the Abdomen versus the Whole-body extraction, 8 of 17 peaks show statistically significant differences. Finally, considering the Wings versus the Whole-body, 6 of 17 peaks show statistically significant differences.

## Discussion and Conclusions

Given that all three methods are quantitatively comparable, and that wing and abdomen extractions produce largely indistinguishable chromatograms, the wings appear to accurately represent the cuticular surface of the bees. The whole-body samples group separately from the other two methods and show a greater number of individual peak differences. Whole-body extractions may also include glandular secretions or internal lipids which can affect the quantitative or qualitative characteristics of the sample (Choe et al., 2012; Lacey et al., 2008; Monnin et al., 1998). Consequently, it is possible that the larger qualitative differences seen in whole-body extractions could result from skews introduced by glandular secretions not representative of the direct cuticular surface. Given that the cuticular surface is expected to be most important in CHC based communication among individuals, using the wing extracts appears to be a useful approach to assess the sources of variation structuring *E. dilemma* CHCs.

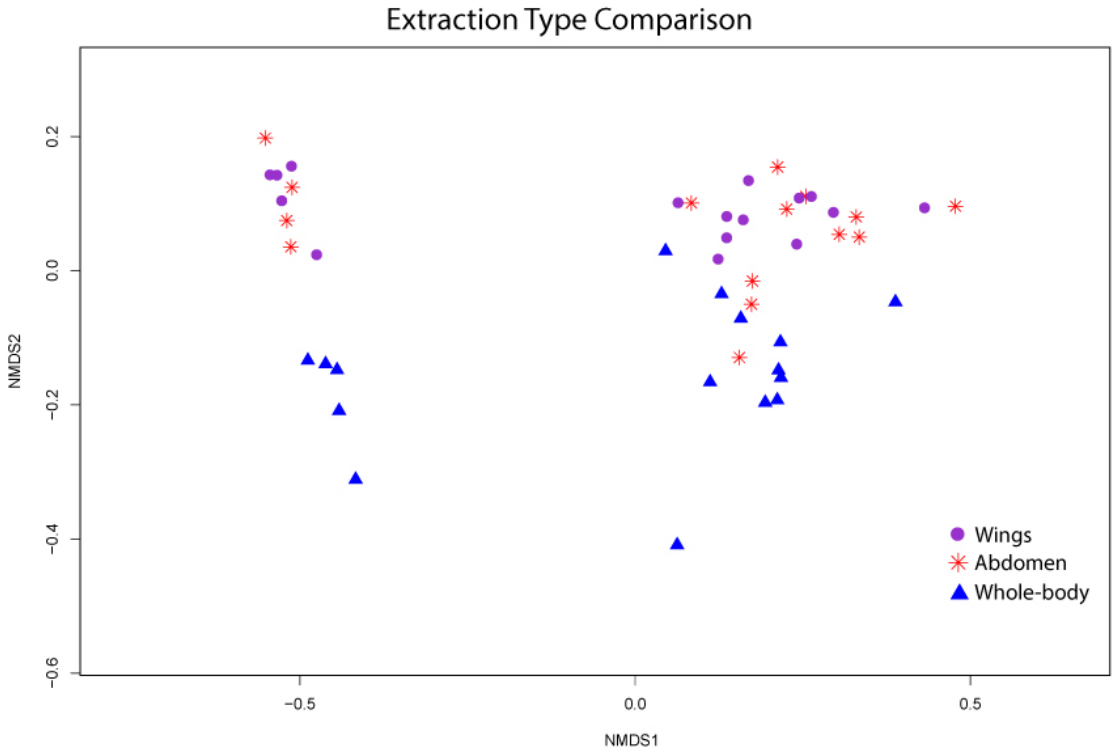



**Figure S3.** NMDS plot showing three extraction types from the same individuals. Stress = 0.041.

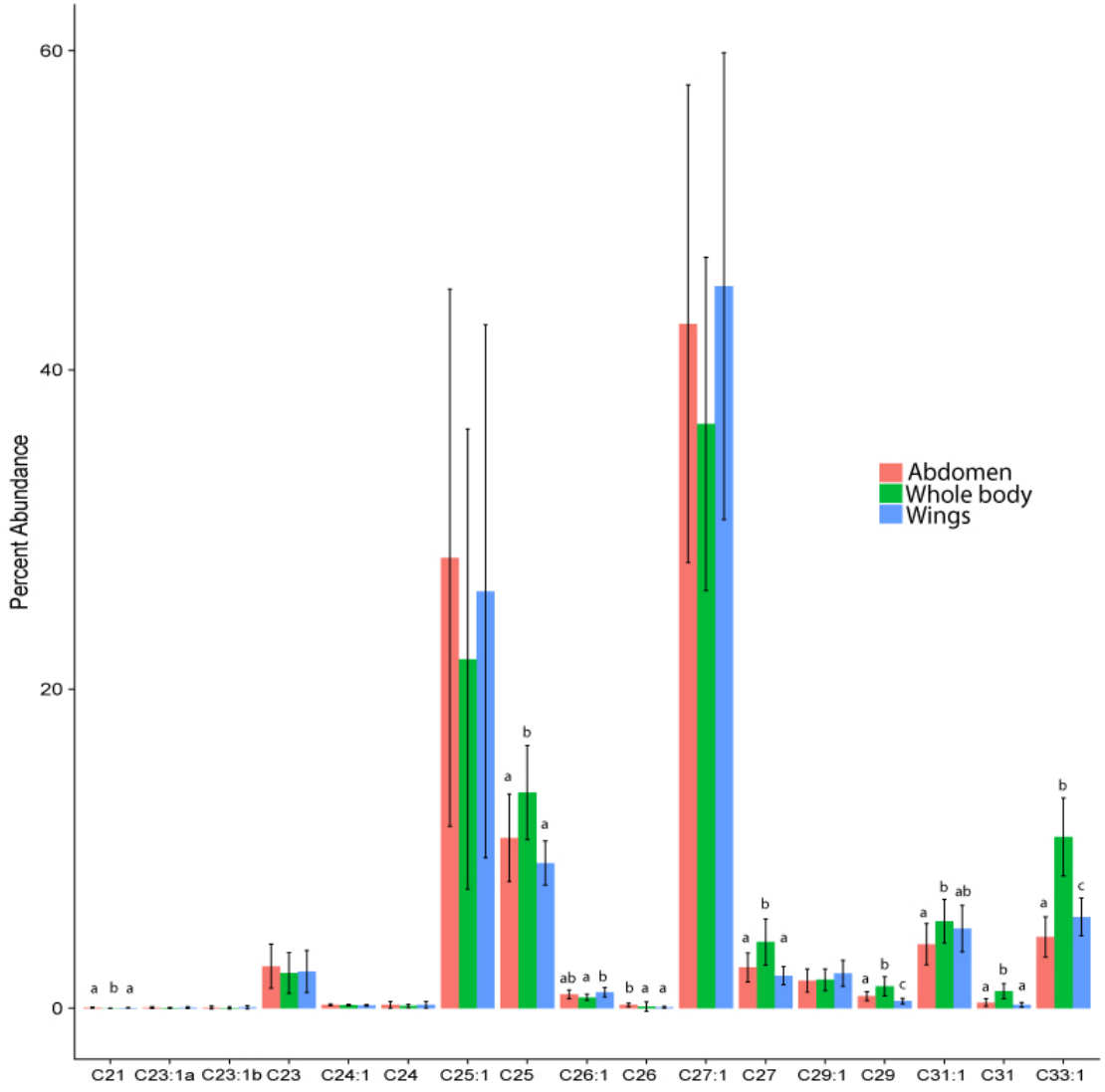



**Figure S4.** Relative abundance comparison of CHC peaks among the three extraction methods. Peaks found to be statistically different in Kruskal-Wallis tests were assigned groupings with Steel-Dwass tests. Groups with the same letter label are in the same statistical group. Peaks are only labeled with statistical groupings if at least one extraction method produced statistically significant differences at a given peak. Error bars represent the standard deviation. The first bar at each CHC peak in the figure is the abdomen, the second is the whole-body, and the third is the wings.

References for Appendix 1:

Choe, D. H., Ramírez, S. R., & Tsutsui, N. D. 2012. A Silica Gel Based Method for Extracting Insect Surface Hydrocarbons. *Journal of Chemical Ecology*, *38*(2), 176–187. https://doi.org/10.1007/s10886-012-0074-1

Lacey, E. S., Ginzel, M. D., Millar, J. G., & Hanks, L. M. 2008. 7-methylheptacosane is a major component of the contact sex pheromone of the cerambycid beetle Neoclytus acuminatus acuminatus. *Physiological Entomology*, *33*(3), 209–216. https://doi.org/10.1111/j.1365-3032.2008.00624.x

Monnin, T., Malusse, C., & Peeters, C. 1998. Solid-phase microextraction and cuticular hydrocarbon differences related to reproductive activity in queenless ant Dinoponera quadriceps. *Journal of Chemical Ecology*, *24*(3), 473–490. https://doi.org/10.1023/A:1022360718870

Turillazzi, S., Sledge, M. F., & Moneti, G. 1998. Use of a simple method for sampling cuticular hydrocarbons from live social wasps. *Ethology Ecology & Evolution*, *10*(3), 293–297. https://doi.org/10.1080/08927014.1998.9522859

**Appendix 2.** Genotyping and relatedness methods and additional details.

Assessing relatedness among individuals in our samples represents a challenge due to the high levels of inbreeding present in our dataset. Because of this, we provide additional information here about methodology so that areas of uncertainty in the analysis can be made more explicit.

Our samples should show a high level of inbreeding for two reasons. First, our study population in Florida is the result of an accidental introduction approximately 15-20 years ago, probably from a small number of individuals (Skov and Wiley, 2005). In addition, all these nests are from one city in FL and many of these nests were collected on the same building. Given that individuals often show very short-range dispersal (personal observation, N. Saleh), nests from a single building are likely founded by the offspring of nearby nests. Thus, we likely have multiple extended families present in our dataset. Most relatedness estimators assume an outbred population when calculating allele frequencies among samples (Wang, 2014). Although several relatedness estimators are robust to violations of this assumption, this violation can result in a bias in relatedness values. In addition, relatedness estimates from datasets with a relatively low number of samples and a high marker density (hundreds or thousands of SNPs), will be

underestimated (Wang, 2017), which we see in our analysis. Given these biases, we performed several iterations of parameters in STACKS, to see if any particular combination gave us the ability to discern known relationships among our control nests (Fig. S3, below). We also independently extracted DNA from the thorax and leg of one individual, to serve as a basic control.

Our four control nests consisted of three guard individuals that were observed to have provisioned their own offspring and one non-kin control where we observed an outside bee (N28_W85) usurp the dominant position in a social nest. We collected the guarding individuals and 1-4 of their subsequently emerging offspring, which were their daughters. After identifying and then sampling the guarding mother, nests were moved into a lab growth chamber (12hr light/dark cycle, 80°F, 70% humidity) to facilitate collecting emerging offspring. One of the three guarding mothers was maintained in the lab for an extended period with the developing offspring and was thus not included in the broader chemical/statistical analysis (individual N51B_R66_1). Individuals that emerged from control nests in the growth chamber were not included in the broader chemical/statistical analysis, due to the different context of their emergence and sampling. In addition, several genotyped samples were collected on or emerging from nests sampled in the field and lacked behavioral observation. Consequently, these samples were used in the genotyping dataset as possible nestmates but not used in the broader chemical/statistical analysis of behavioral CHC data. Individuals only involved in the genotype dataset and not the chemical/behavioral analysis, for any of the reasons outlined above, are labeled as "Genotype_only" under the behavior column in the supplemental CHC data set.

Considering our various controls, we first confirmed that the duplicated control sample from the same individual showed the highest relatedness value in the dataset. This was indeed

the case, regardless of varied parameters, so we dropped the duplicated sample and proceeded to assess the relationship of other known controls. Specifically, we were looking to optimize a parameter combination to show that known sisters have high relatedness values with each other and relative to the entire dataset, with mothers/daughters somewhat below that. In addition, we expected our non-kin control to fall around the average relatedness value of the dataset, which we expect to be representative of non-kin individuals.

Initially, we started with less strict parameters that generated thousands of SNPs. For instance, we did not require that all individuals have each SNP (instead setting values between 60-80%) and we kept the minor allele frequency low (0.01-0.05). In addition, we did not set a maximum heterozygosity. Although these combinations gave us thousands of SNPs, we found that the relatedness values among all samples was extremely high, swamping the signal we were hoping to pull out from our controls. Given this, we then varied parameters to stricter settings (see methods in main text), which resulted in a much smaller, but more informative set of SNPs. With these strict parameters, known sisters show the highest relatedness to each other and they show high relatedness relative to the rest of the dataset, which we expect, as sisters should be more related than any other possible relationship among females. Furthermore, all mothers and daughters in a nest showed lower relatedness than sisters, but still a much higher relatedness than the dataset average. Finally, the pairwise relatedness of the non-kin control nest fell close to the average pairwise relatedness value, suggesting that the average value in the distribution was representative of non-kin. Given this, we felt that we could tentatively assign most nests to broad categories of kin (mothers, daughters, sisters) and non-kin (other relationships).

Based on this, we considered relatedness values of less than 0.21 as "non-kin," as 0.21 was both the relatedness value for our non-kin control and about the average relatedness value

for a random pair of individuals. We considered values above 0.29 to be kin, as none of our known kin controls had values less than this. Finally, values between 0.21 and 0.29 were defined as "ambiguous," as they fall in between our known controls.

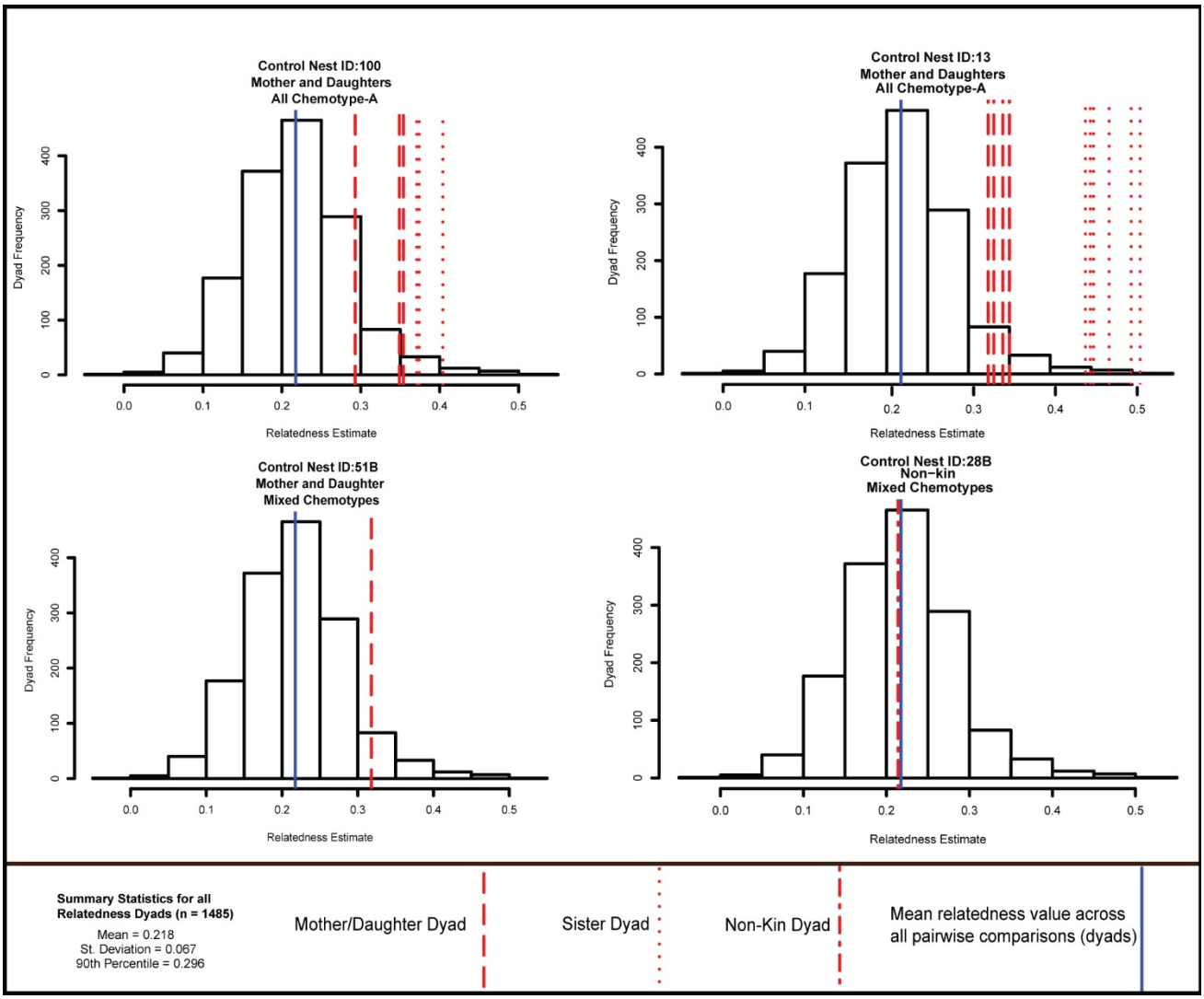



**Figure S5.** Histograms showing the pairwise relatedness estimates (dyads) between nestmates of the four control nests genotyped. The total dataset is comprised of 1485 dyads where a relatedness value between two individuals is estimated.

**Table S1**. Relatedness calculations from individuals of the same nest. We assigned putative kin/non-kin status based on values from our control individuals. Any individuals that fall below the lowest Mother-daughter value (0.29) and above the non-kin control value (0.21) we label as ambiguous. Mixed chemotype nestmate comparisons are bolded.

| Nest | Individual 1 | Chemotype | Individual 2 | Chemotype | Relatedness | Putative Kin (Y/N)? |
|---|---|---|---|---|---|---|
| 100 | N100_24_27 | A | N100_24_32 | A | 0.371 | Y |
| 100 | N100_24_27 | A | N100_48_1 | A | 0.404 | Y |
| 100 | N100_24_27 | A | N100_w82_94 | A | 0.3493 | Y |
| 100 | N100_24_32 | A | N100_48_1 | A | 0.3736 | Y |
| 100 | N100_24_32 | A | N100_w82_94 | A | 0.3541 | Y |
| 100 | N100_48_1 | A | N100_w82_94 | A | 0.2934 | Y |
| 13 | N13_24_12 | A | N13_24_26 | A | 0.51 | Y |
| 13 | N13_24_12 | A | N13_48_2 | A | 0.4986 | Y |
| 13 | N13_24_12 | A | N13_g9_2 | A | 0.4727 | Y |
| 13 | N13_24_12 | A | N13_y2_77 | A | 0.3305 | Y |
| 13 | N13_24_26 | A | N13_48_2 | A | 0.4426 | Y |
| 13 | N13_24_26 | A | N13_g9_2 | A | 0.4494 | Y |
| 13 | N13_24_26 | A | N13_y2_77 | A | 0.3501 | Y |
| 13 | N13_48_2 | A | N13_g9_2 | A | 0.4526 | Y |
| 13 | N13_48_2 | A | N13_y2_77 | A | 0.3242 | Y |
| 13 | N13_g9_2 | A | N13_y2_77 | A | 0.3422 | Y |
| 16_A | N16A_n17_165 | A | N16A_n57_166 | A | 0.3345 | Y |
| **16_B** | **N16B_w97_75** | **B** | **N16B_w98_76** | **A** | **0.2842** | **Ambiguous** |
| 20 | N20_R19_82 | A | N20_w94_80 | A | 0.3809 | Y |
| 20 | N20_R19_82 | A | N20_w95_81 | A | 0.3108 | Y |
| 20 | N20_w94_80 | A | N20_w95_81 | A | 0.3484 | Y |
| 26 | N26_n34_154 | B | N26_n61_153 | B | 0.2355 | Ambiguous |
| **2** | **N2_n10_98** | **B** | **N2_n61_99** | **A** | **0.3573** | **Y** |
| 27 | N27_24_16 | A | N27_w90_78 | A | 0.4086 | Y |
| 27 | N27_24_16 | A | N27_w91_79 | A | 0.4643 | Y |
| 27 | N27_w90_78 | A | N27_w91_79 | A | 0.3759 | Y |
| 28A | N28A_n67_133 | A | N28A_n80_134 | A | 0.3602 | Y |
| 28A | N28A_n67_133 | A | N28A_n66_132 | A | 0.4296 | Y |
| 28A | N28A_n67_133 | A | N28A_n81_135 | A | 0.4611 | Y |
| 28A | N28A_n80_134 | A | N28A_n66_132 | A | 0.3625 | Y |
| 28A | N28A_n80_134 | A | N28A_n81_135 | A | 0.3788 | Y |
| 28A | N28A_n66_132 | A | N28A_n81_135 | A | 0.4537 | Y |
| **28B** | **N28B_w47_83** | **A** | **N28B_w85_84** | **B** | **0.2143** | **N** |
| 29 | N29_w76_158 | A | N29_G23_159 | A | 0.4486 | Y |
| 29 | N29_w76_158 | A | N29_Y88_157 | A | 0.3915 | Y |
| 29 | N29_G23_159 | A | N29_Y88_157 | A | 0.3851 | Y |
| **33** | **N33_w80_173** | **A** | **N33_G27_171** | **B** | **0.2828** | **Ambiguous** |
| 33 | N33_w80_173 | A | N33_w98_174 | A | 0.3853 | Y |
| **33** | **N33_G27_171** | **B** | **N33_w98_174** | **A** | **0.3371** | **Y** |
| 44 | N44_n47_111 | A | N44_n85_118 | A | 0.2968 | Y |

| | | | | | | |
|---|---|---|---|---|---|---|
| 47 | N47_w74_111 | B | N47_Y87_112 | B | 0.3762 | Y |
| **5** | **N5_w23_67** | **B** | **N5_w24_68** | **A** | **0.1745** | **N** |
| 51A | N51A_24_4 | A | N51A_24_5 | A | 0.3508 | Y |
| 51A | N51A_24_4 | A | N51A_w78_89 | A | 0.3551 | Y |
| 51A | N51A_24_4 | A | N51A_y6_88 | A | 0.4697 | Y |
| 51A | N51A_w78_89 | A | N51A_y6_88 | A | 0.3516 | Y |
| 51A | N51A_24_5 | A | N51A_y6_88 | A | 0.3345 | Y |
| **51B** | **N51B_24_47** | **A** | **N51B_R66_1** | **B** | **0.3184** | **Y** |
| 59A | N59A_n75_97 | A | N59A_n53_96 | A | 0.3563 | Y |
| 59B | N59B_y15_91 | A | N59B_y14_90 | A | 0.3487 | Y |
| **75** | **N75_w20_64** | **B** | **N75_w21_65** | **A** | **0.2598** | **Ambiguous** |
| 82 | N82_y10_86 | A | N81_y35_87 | A | 0.4087 | Y |

References for Appendix 2:

Skov, C., and Wiley, J. 2005. Establishment of the Neotropical Orchid Bee Euglossa Viridissima (Hymenoptera: Apidae) in Florida. *Florida Entomologist*, *88*(2), 225–227. https://doi.org/10.1653/0015-4040(2005)088[0225:EOTNOB]2.0.CO;2

Wang, J. 2014. Marker-based estimates of relatedness and inbreeding coefficients : an assessment of current methods. *Journal of Evolutionary Biology, 27*, 518–530. https://doi.org/10.1111/jeb.12315

Wang, J. 2017. Estimating pairwise relatedness in a small sample of individuals. *Heredity*, *119*(5), 302–313. https://doi.org/10.1038/hdy.2017.52

**Appendix 3:** Additional details on the CHCs of newly emerged females.

Newly emerged females, sampled within 24hrs of emergence, appear to have several unique features of their CHC profiles that led us to determine that including them with other female behavioral groups in statistical analysis was problematic. We briefly discuss these features, as they suggest interesting aspects of early life CHC development in *E. dilemma* that may be useful for future follow up. As seen in other social insects, newly emerged females often have lower quantities of certain CHCs or show different CHC relative abundances (Monnin et al, 1998; Culliver-Hot et al., 2001; Kelstrup et al., 2018). First, when running the samples on the

GC-MS, we noticed that newly emerged females appear to have a much lower total abundance of CHCs across the whole chromatogram, a feature that is not apparent when only comparing relative abundances, as is done in our other analyses. To illustrate this, we compare the total ion abundances from chromatograms across all female behaviors (Fig. S4). This has been used elsewhere to serve as a rough estimate of the total abundance of material in a chemical sample (Arriaga-Osnaya et al, 2017). Given that all behaviors were extracted and analyzed with the same approach, this suggests that newly emerged females are likely still in the process of secreting CHCs and have not reached the amount typical of most females performing normal adult nest behaviors. In addition, nearly all newly emerged females are missing the three shortest chain length CHCs in the chromatogram (C21, C23:1a, C23:1b), which may further skew relative abundance estimates when comparing to other behaviors. Finally, when hierarchical clustering is run, as in Figure 3, a very different correlational structure is revealed when only newly emerged females are run, with many peaks showing patterns of correlation different from the otherwise consistent patterns of the females collected from nests. Additional work should assess the timeline of CHC deposition and change across the early life of an individual. For now, we restrict analysis of newly emerged females to the NMDS plots because of these qualitative and quantitative differences from the rest of the female dataset.

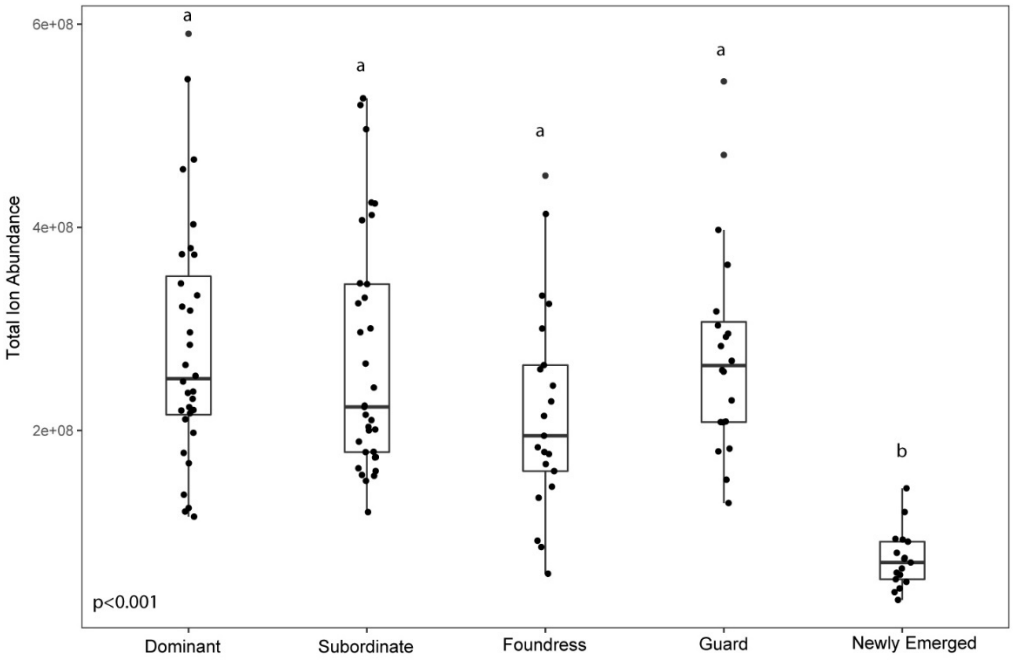



**Figure S6.** Total ion abundance across behaviors. Letters denote statistical groupings determined by a Steel-Dwass test and the p-value is calculated using a Kruskal Wallis test, due to non-normal data.

References for Appendix 3:

Arriaga-Osnaya, B. J., Contreras-Garduño, J., Espinosa-García, F. J., García-Rodríguez, Y. M., Moreno-García, M., Lanz-Mendoza, H., Cueva del Castillo, R. 2017. Are body size and volatile blends honest signals in orchid bees? *Ecology and Evolution*, *7*(9), 3037–3045. https://doi.org/10.1002/ece3.2903

Cuvillier-Hot, V., Cobb, M., Malosse, C., & Peeters, C. 2001. Sex, age and ovarian activity affect cuticular hydrocarbons in Diacamma ceylonense, a queenless ant. *Journal of Insect Physiology*, *47*(4–5), 485–493. https://doi.org/10.1016/S0022-1910(00)00137-2

Kelstrup, H. C., Hartfelder, K., Lopes, T. F., & Wossler, T. C. 2018. The Behavior and Reproductive Physiology of a Solitary Progressive Provisioning Vespid Wasp : Evidence for a Solitary-Cycle Origin of Reproductive Castes, *191*(2). https://doi.org/10.1086/695336

Monnin, T., Malusse, C., & Peeters, C. 1998. Solid-phase microextraction and cuticular hydrocarbon differences related to reproductive activity in queenless ant Dinoponera quadriceps. *Journal of Chemical Ecology*, *24*(3), 473–490. https://doi.org/10.1023/A:1022360718870

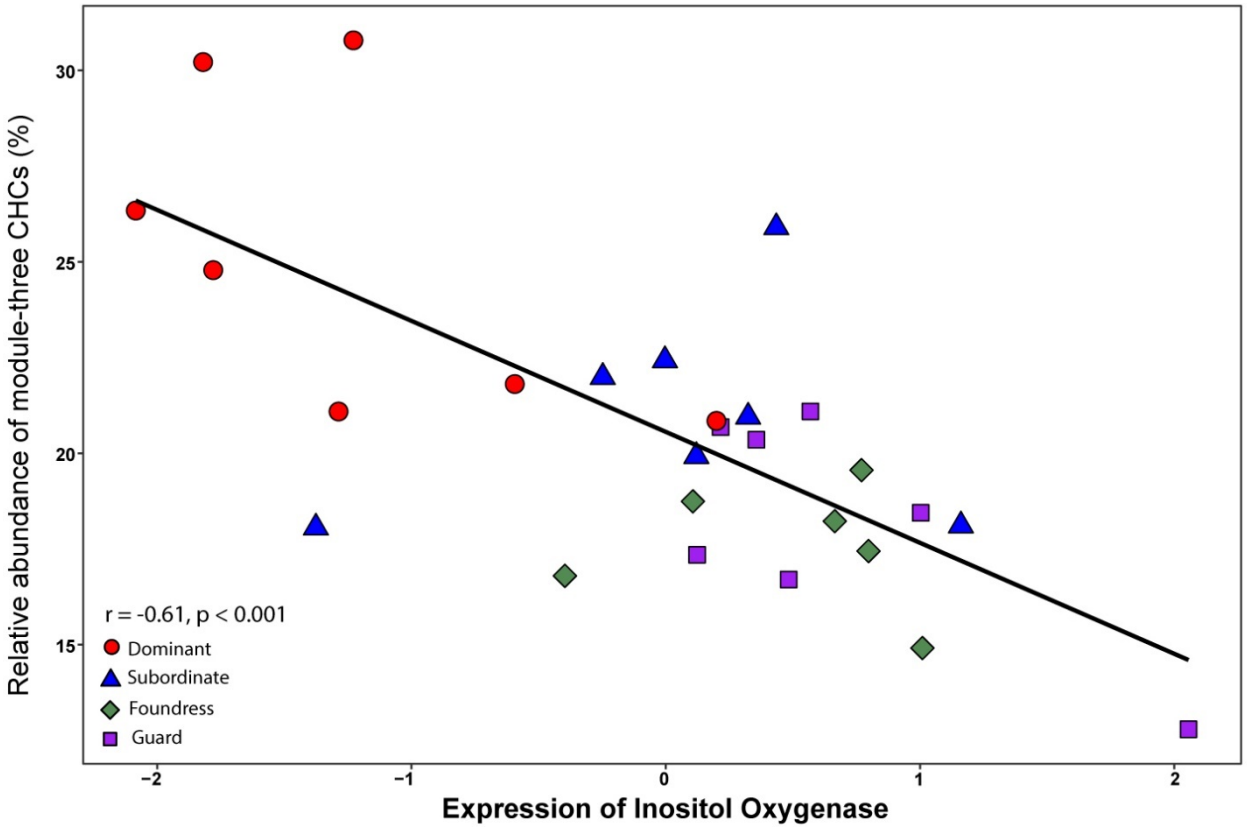



**Figure S7.** Correlation between ovary expression of inositol Oxygenase and the relative abundances of the module-three (C25, C26, C27, C29, C31) CHCs across behavioral groups. The x-axis shows log2 expression levels for each individual relative to the mean value for that gene across individuals. Unique shape and color combinations correspond to the four sampled behavioral groups. The correlation coefficient and FDR adjusted p-value were calculated with Spearman's rho.

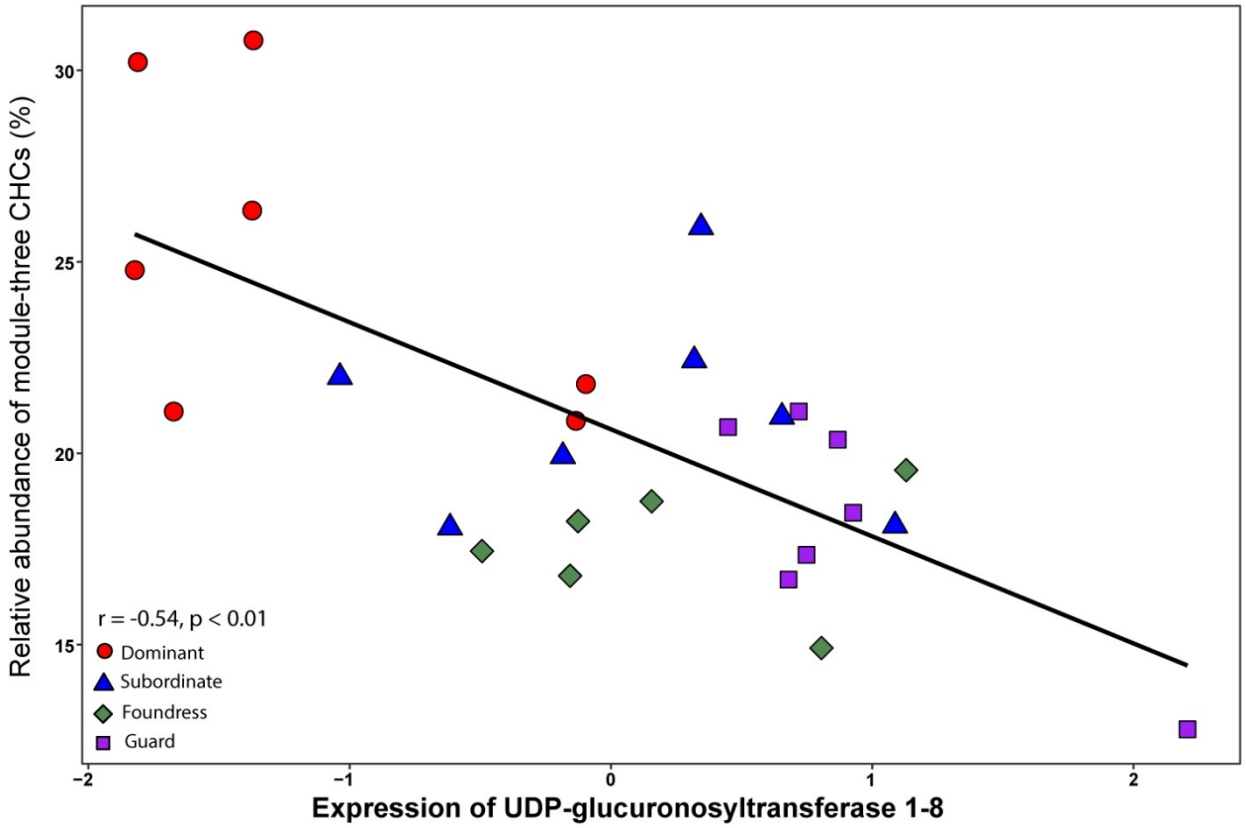



**Figure S8.** Correlation between ovary expression of inositol Oxygenase and the relative abundances of the module-three (C25, C26, C27, C29, C31) CHCs across behavioral groups. The x-axis shows log2 expression levels for each individual relative to the mean value for that gene across individuals. Unique shape and color combinations correspond to the four sampled behavioral groups. The correlation coefficient and FDR adjusted p-value were calculated with Spearman's rho.

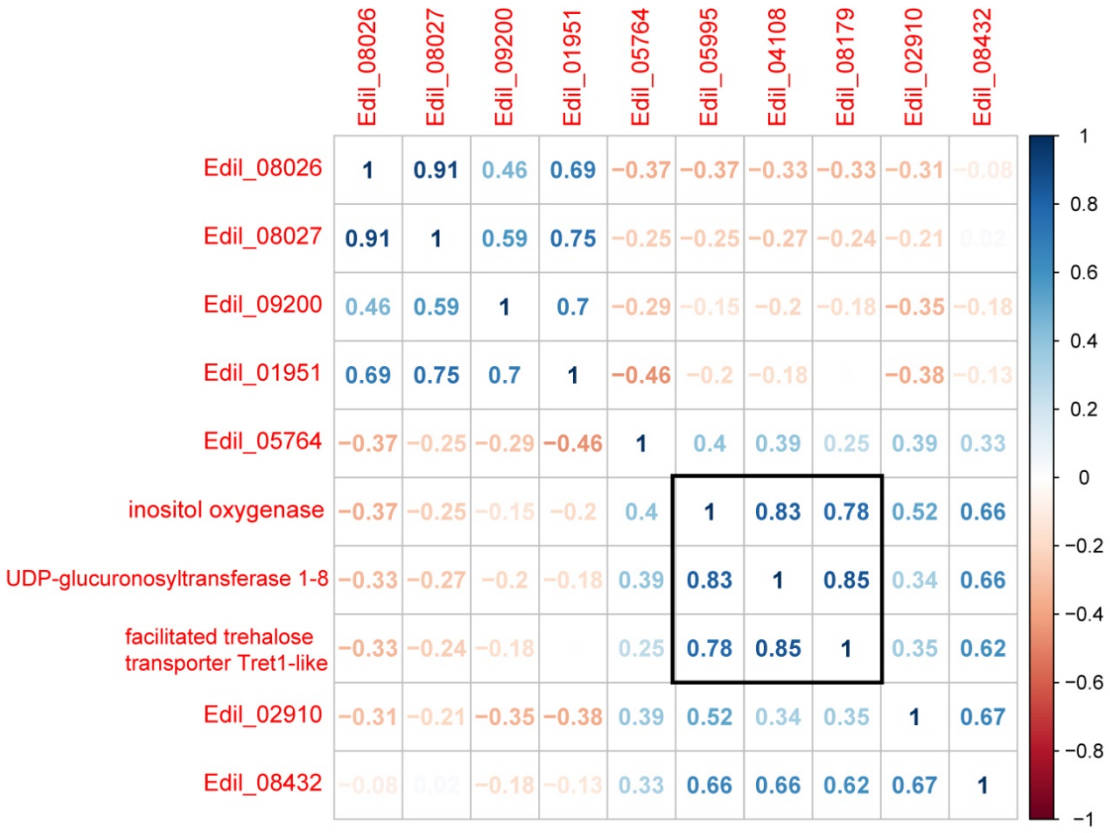



**Figure S9.** Correlation matrix showing each pairwise combination of the 10 genes differentially expressed between dominants and subordinates in *E. dilemma* ovaries. The number in each light gray square represents the Spearman's rho correlation coefficient between the expression values (log2 normalized counts per million transcripts) of two genes for all individuals in the gene expression data set. A black box has been drawn to highlight the correlations between the three genes identified to have correlated expression with ovary size and CHC module-three expression. The annotations for these three genes have been added on the left vertical axis, while their *E. dilemma* gene ID is maintained on the horizontal axis.

**Table S2**. Correlation between individual CHCs in module-three, ovary size index, and three DEGs between dominants and subordinates in the ovaries. Pearson's r is shown for ovary size index and Spearman's rho is shown for gene expression correlations. Unadjusted p-values are shown.

| CHC | Ovary Size Index | facilitated transporter Tret 1-like | inositol oxygenase | UDP-glucornosyltransferase 1-8 |
|---|---|---|---|---|
| C25 | R = 0.47, p < 0.001 | R = -0.51, p = 0.0061 | R = -0.47, p = 0.013 | R = -0.35, p = 0.07 |
| C26 | R = 0.16, p = 0.19 | R = -0.58, p = 0.0016 | R = -0.57, p = 0.002 | R = -0.5, p = 0.0077 |
| C27 | R = 0.34, p = 0.005 | R = -0.45, p = 0.017 | R = -0.62, p = 0.00061 | R = -0.46, p = 0.015 |
| C29 | R = 0.35, p = 0.004 | R = -0.29, p = 0.14 | R = -0.32, p = 0.1 | R = -0.44, p = 0.021 |
| C31 | R = 0.44, p < 0.001 | R = -0.44, p = 0.022 | R = -0.31, p = 0.12 | R = -0.52, p = 0.0053 |

**Appendix 4**. Assessing ovary size and dominance information in modules 1 and 2 (chemotype related peaks).

Due to the bimodal nature of the chemotype data, we looked at fertility and dominance patterns in chemotype-A individuals only, as chemotype-B individuals were less abundant and had insufficient statistical power on their own, with several behaviors having sample sizes of only 4 or 5 individuals  However, looking at chemotype-A individuals still allows for examination of patterns in the chemotype-B peaks, as these are all still expressed in all individuals, just at a lower relative abundance. We first confirmed that patterns of significance were the same for CHC module 3 when run with the reduced set of chemotype-A individuals; this was indeed the case. However, although we found significant effects of behavior on CHC modules 1 and 2 (supplemental figure S7A-B), dominant individuals did not uniquely express a higher relative abundance of these CHCs. Furthermore, these modules were not significantly correlated with ovary size (supplemental figure S8-S9). Consequently, these results suggest that the other CHC modules that we identify in our analysis (1 and 2) do not reflect dominance status or ovary size.

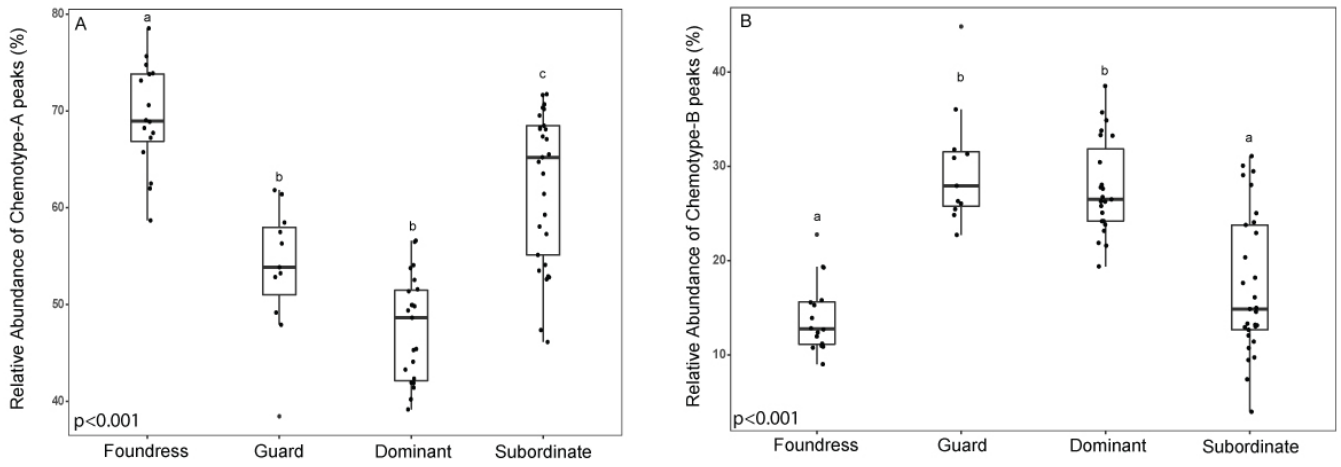



**Figure S10. A)** Relative abundance of CHCs from the chemotype-A module across behaviors. Letters denote statistical groupings determined by a Steel-Dwass test and the p-value is calculated using a Kruskal Wallis test, due to data failing ANOVA assumptions. **B)** Relative abundance of CHCs from the chemotype-B module across behaviors. Letters denote statistical groupings determined by a Steel-Dwass test and the p-value is calculated using a Kruskal Wallis test, due to data failing ANOVA assumptions. The box plots show the mean value in each group.

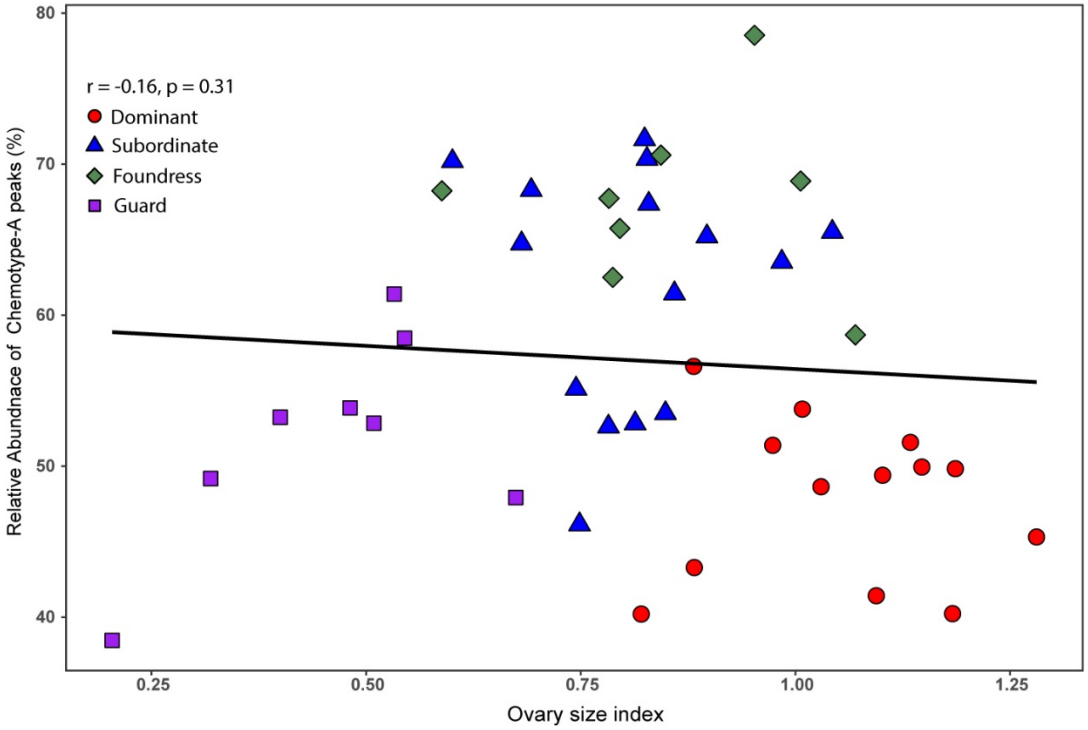


**Figure S11.** Correlation between ovary size index and the chemotype-A peak abundances across behaviors. Unique shape and color combinations correspond to the four sampled behaviors. The correlation coefficient and p-value were calculated with Pearson's r.

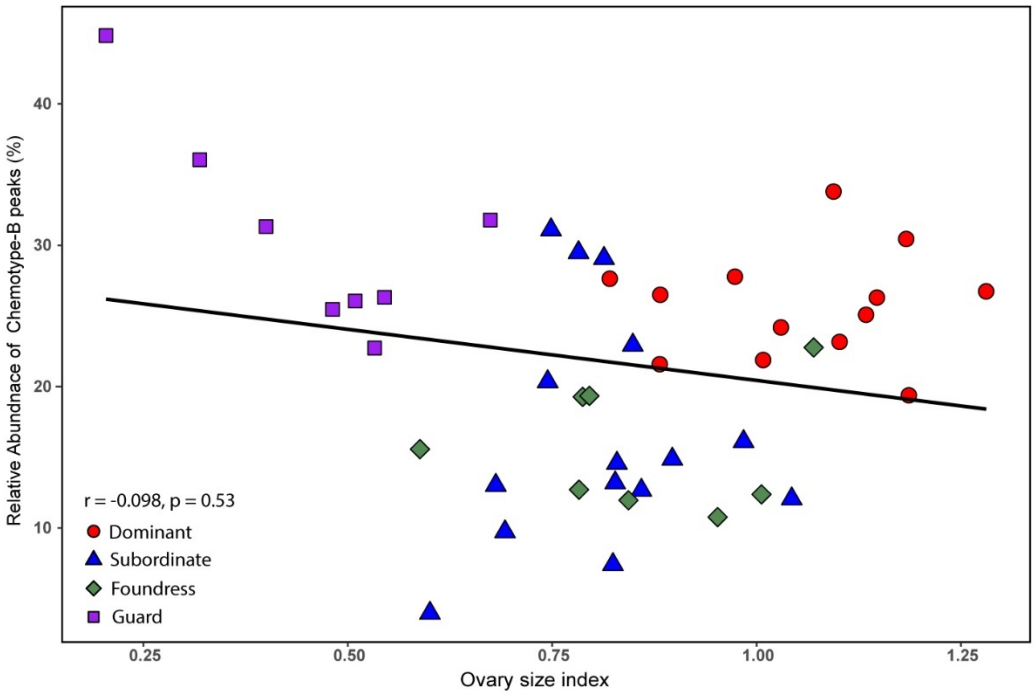



**Figure S12.** Correlation between ovary size index and the chemotype-B peak abundances across behaviors. Unique shape and color combinations correspond to the four sampled behaviors. The correlation coefficient and p-value were calculated with Spearman's rho.