Ancestral State Reconstruction and Evolutionary Modeling of Eusociality in Hymenoptera

BINF 6203 Final Project

Samantha A. Taylor

14 December 2021

Introduction

There are many levels of organization of sociality in animals. The highest level is termed eusociality, and this type of sociality has three defining characteristics: overlapping generations within adults in a colony, reproductive division of labor where adults are divided into reproductive (queens) and non-reproductive (sterile workers) groups, and cooperative broad care where the offspring is cared for by non-parental individuals. While eusociality is primarily found and studied in the Hymenoptera order, it can also be found in Isoptera, and a few crustaceans and mammals. In all species of the Hymenoptera and Thysanoptera orders, haplodiploidy occurs, with females being heterozygous, and males being homozygous or hemizygous.

Suzanne Batra coined the term "eusocial" in 1966 when studying Halictine bees, with Charles D. Michener adding to the definition in 1969, by defining the three main characteristics known today. Eusociality was initially thought of as an evolutionary paradox, since within a gene-centered view of evolution, natural selection should not produce sterile individuals. This led to several theories as to the origins of eusociality: group selection (multi-level selection), and kin selection (inclusive fitness).

Within the kin selection hypothesis, the altruistic care is due to haplodiploidy. This theory suggests that animals are more likely to care for another that is related to them then for one that is not related, and considers relatedness, fitness benefits of the recipient, and fitness cost of the altruist. Relatedness can be calculated using variable r, with sisters are more closely related to each other (r = 3/4) than daughters are to their mothers (r = 1/2). The group hypothesis focuses on phenotype, with selection acting simultaneously on multiple levels of organization. Although being unselfish may be costly for the selfless individual, groups that contain all altruistic individuals may be more fit than groups containing all selfish individuals due to the selfless group being more productive with less confrontation. In essence, with kin selection, you genetically benefit due to your sister passing on some of your shared genes to her kids, whereas with multi-level selection, your family benefits because your sister (with superior genes) passes on family traits to her kids. Both theories hypothesize that once a species goes eusocial, they hit the "point of no return". [1, 2]

However, there are multiple levels to eusociality. While a species can be solitary (fit none of the three eusociality characteristics), they can be primitively eusocial (only sometimes having overlapping generations), or highly eusocial (fit all three characteristics).

The objective is to see if primitive sociality is a precursor to more complex social behaviors in the Hymenoptera order, through the use of ancestral state reconstruction, treating eusociality-level as both a discrete trait and a continuous trait. This order comprises of sawflies, wasps, bees, and ants. In bee and wasp species, all three levels occur, whereas in ants and sawflies, only highly eusocial and solitary social behaviors occur, respectively.

Methods

In past studies, complications occur during the gene selection step of answering this question. Therefore, new genes have been chosen that have a possible high correlation to the evolution of eusociality because of the genes' expressions. The five genes compared are BRI3 (Brain Protein I3), FASN (Fatty Acid Synthase), GRM1 (Glutamate Metabotropic Receptor 1), MRJP1 (Major royal jelly protein 1), and Oamb (octopamine receptor). A fresh selection of species were also chosen, with the outgroup being the thrip, Frankliniella occidentalis, due to it being haplodiploidy, and including protein coding genes such as those for royal jelly.

```
##                    SPECIES CLASSIFICATION     SOCIALITY.TYPE
## 1  Frankliniella occidentalis     Thrip              Outgroup
## 2             Cephus cinctus     Sawfly              Solitary
## 3              Athalia rosae     Sawfly              Solitary
## 4          Trachymyrmex zeteki        Ant     Highly Eusocial
## 5       Camponotus floridanus    Ant\xa0     Highly Eusocial
## 6         Monomorium pharaonis        Ant     Highly Eusocial
## 7              Apis mellifera        Bee     Highly Eusocial
## 8              Apis laboriosa        Bee     Highly Eusocial
## 9           Bombus terrestris        Bee Primitively eusocial
## 10           Bombus impatiens        Bee Primitively eusocial
## 11         Megachile rotundata        Bee              Solitary
## 12          Eufriesea mexicana        Bee              Solitary
## 13       Vespula pensylvanica       Wasp     Highly Eusocial
## 14          Venturia canescens       Wasp     Highly Eusocial
## 15            Polistes fuscatus       Wasp Primitively Eusocial
## 16          Polistes canadensis       Wasp Primitively Eusocial
## 17         Nasonia vitripennis       Wasp              Solitary
```

NOTE: All data was created by hand, using NCBI reference sequences in the linked table

Creating Concatenated Alignments & Creating Partition Files in R   To get link the alignments, the data from multiple gene sequences were concatenated and partitioned. The data set contained 5 FASTA files, with each file corresponding to a different gene sequence from 16 individuals in the Hymenoptera order, and one individual from the Thysanoptera order (all in the Insecta class). The files were read into R, and the alignment was completed separately for each gene (in alphabetical order), the alignments were concatenated, and then output files were appropriately created for RAxML usage (phylip format). To do this, the msa and phangorn libraries were used. The alignments were performed using ClustalOmega and the lengths of each alignment were stored in a vector.

Maximum Likelihood with RAxML   Next, partition files were created in R that defined where the start and end of each partition (in this case each gene) was in the concatenated alignment. Once the created text file was uploaded to the student hpc cluster, a job script was created for a RAxML run. The -f a option was used to specify the rapid bootstrap algorithm, instead of the check algorithm and the -m GTRGAMMA (or General Time Reversible model (+ Gamma)) option was used because it assumes unequal base frequencies, a different substitution rate for every possible base substitution, and that the mutation rates across the sequence are Gamma distributed. The specification -N 1000 was used to perform 1000 bootstrap replicates. The -o Frankliniella_occidentalis flag was used to specify the outgroup.

Phylogenetic Tree and Assign Character States   Once the RAxML tree was loaded into R, an ordered, named, numeric vector of character states was created, with three possible states: 25 for solitary, 50 for primitively eusocial, and 75 for highly eusocial. A discrete vector was made from assigned species with a score less than 50 ("0" for solitary), and species with a score greater than 50 ("1") for having eusocial

characteristics. Then, a numeric, ordered, vector of continuous scores was formed, with ' "1", "2", and "3" levels, in order to show all three social states.

Fitting Maximum Likelihood Models   To perform a maximum likelihood ancestral state reconstruction, I used the ace() function from the phytools package. This function requires a vector of character states, a phylogenetic tree, an option indicating whether the data is discrete or continuous, and a transition probability model. To fit a discrete model, the "ARD" model was specified, meaning "All Rates Different." To fit a Stochastic model (Markov), which allows sight of sample character state histories, a transition matrix was created. To fit a continuous model, a different data type and a Brownian Motion (BM) model were specified. In order to make a sort of pie-chart on each node, the maximum possible score from the current state values and the predicted ancestral ones needed to be found, along with shaping the scaling factors to be less than one. The cex option was used to set the scale for each pie chart.
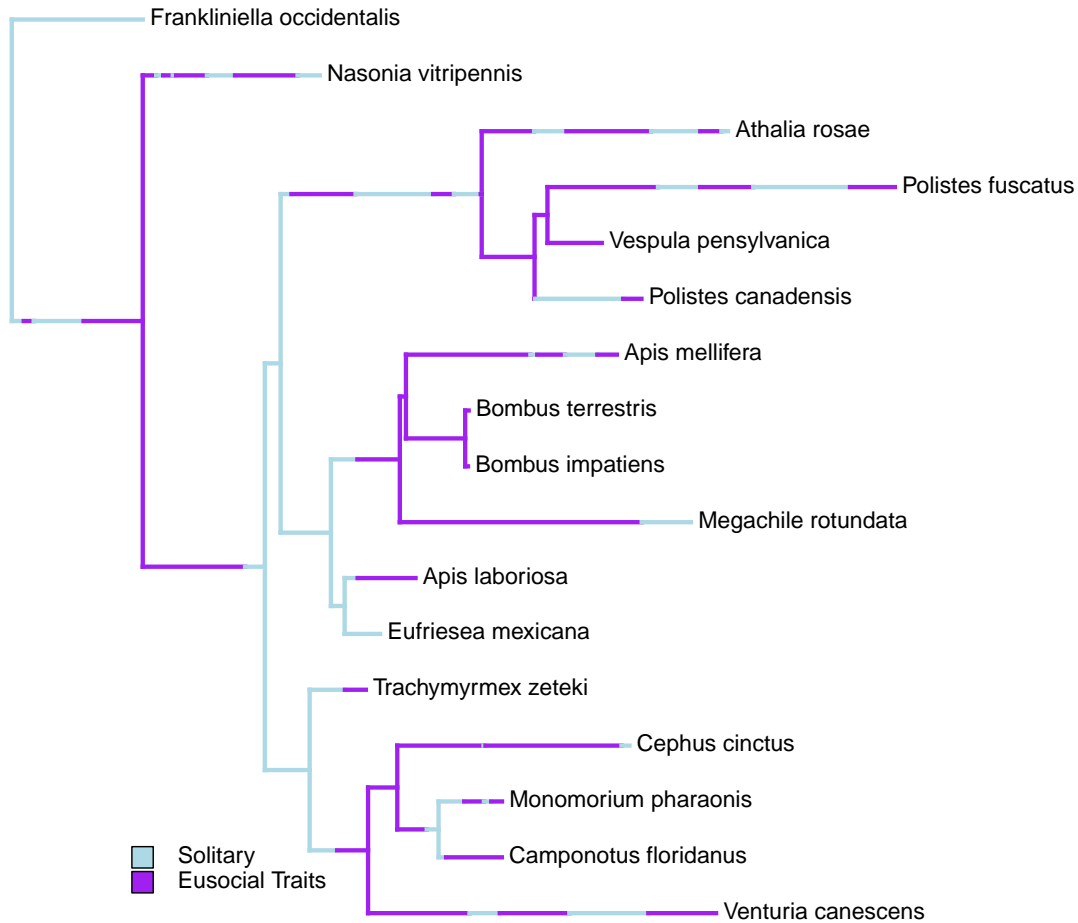

Results & Discussion

I performed an ancestral reconstruction with reconstructed sociality phenotype probabilities at each node. Based on the Figure I discrete model plot, there appear to have been 4 independent social events in the wasps and bees. The first occurred in the Apis laboriosa/Eufrisea (bee) lineage; the second occurred in the Polistes/Vespula pensylvanica lineage; the third occurred in the Apis mellifera/Bombus lineage; the fourth occurred in the Cephus cinctus/Camponotus floridanus/Monomorium pharaonis lineage. Based on the discrete maximum likelihood model (with all different rates), it seems to primarily support the idea that eusociality evolves from solitariness. However, it seems that both Apis mellifera and Apis laboriosa evolved from a solitary species, as there are no primitive species in their lineages. Cephus cinctus breaks the mold altogether, sharing a common ancestor with the highly eusocial Venturia canescens.
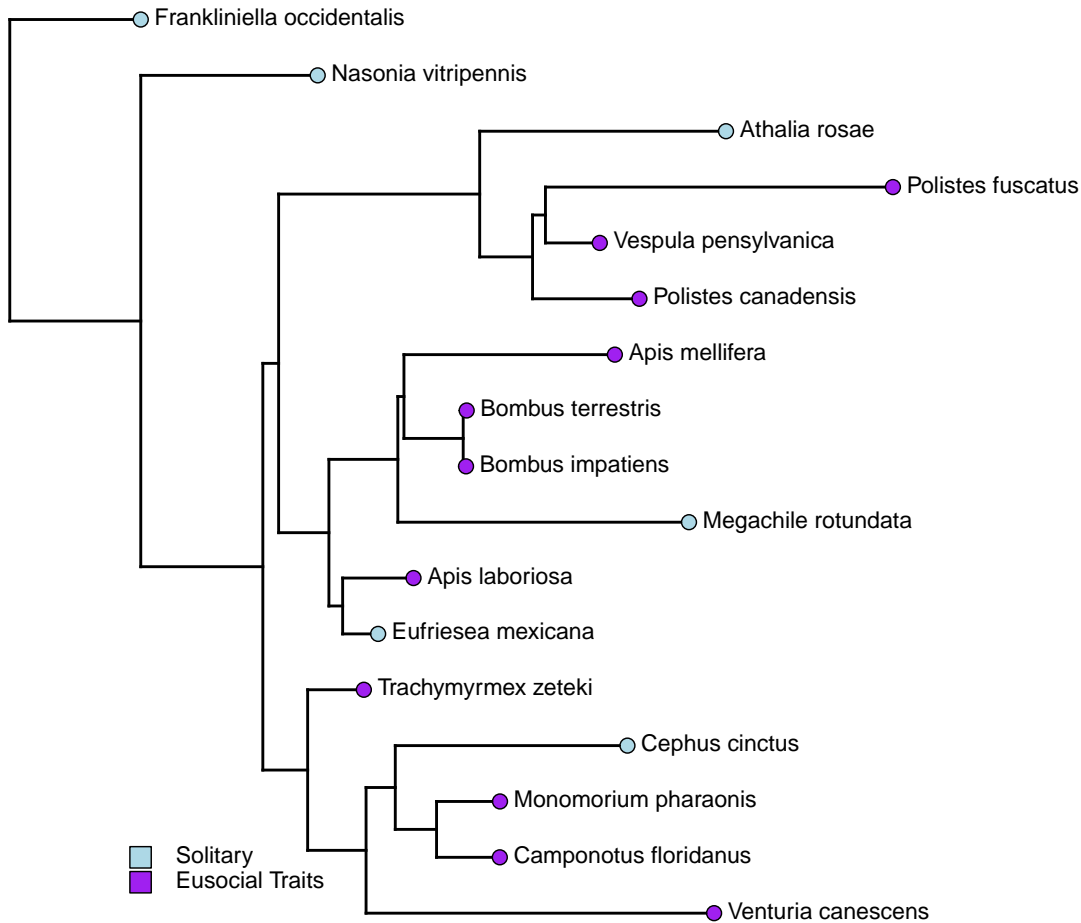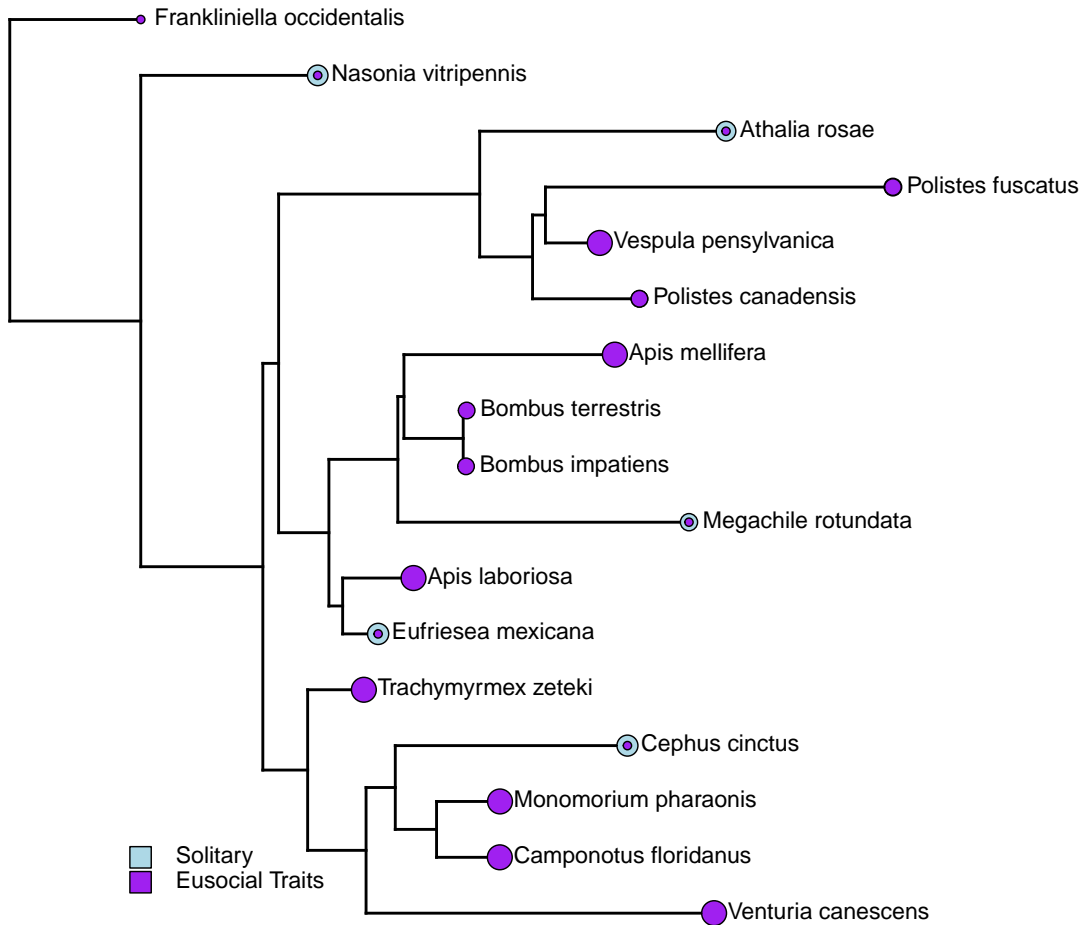
Figure I. Discrete Maximum Likelihood Model

I performed an ancestral reconstruction with reconstructed sociality phenotype probabilities at each node and sample character state histories. Based on the Figure II stochastic model plot, multiple transitions between states along different branches in the tree can be seen. The Apis laboriosa/Eufrisea lineage appears to remain the same, however, Megachile rotundata, Athalia rosae, and Cephus cinctus seems to have somewhat of a transition between social states. The ancestor of Venturia canescens and Cephus cinctus shows as possibly being eusocial, therefore, still presenting the idea that it is possible to return to solitariness.

Figure II. Stochastic Modelling of Possible Character Evolutionary Histories

I performed an ancestral reconstruction with reconstructed sociality phenotype probabilities at each node. Based on the Figure III continuous model plot, circle size can be used to track changes in sociality along different lineages. The most obvious observation is that 5 of the solitary species (that are not the outgroup) show having higher solitariness, but also low amounts of eusociality. Because none of the eusocial species show any number of solitary traits (even the primitively social species) the theory that species must evolve to eusocial tendenciess could be supported.

Figure III. Continuous Maximum Likelihood Model

To plot continuous character changes for all three states separately, a heat map was plotted. The heatmap colors at the internal nodes are indicative of the estimated most likely character states at those points, so, either solitary, primitive, or highly eusocial. Looking at Figure IV, Cephus cinctus is still an enigma; its lineage clearly shows that it evolved from not just a possible primitive eusocial line, but specifically a highly eusocial line.
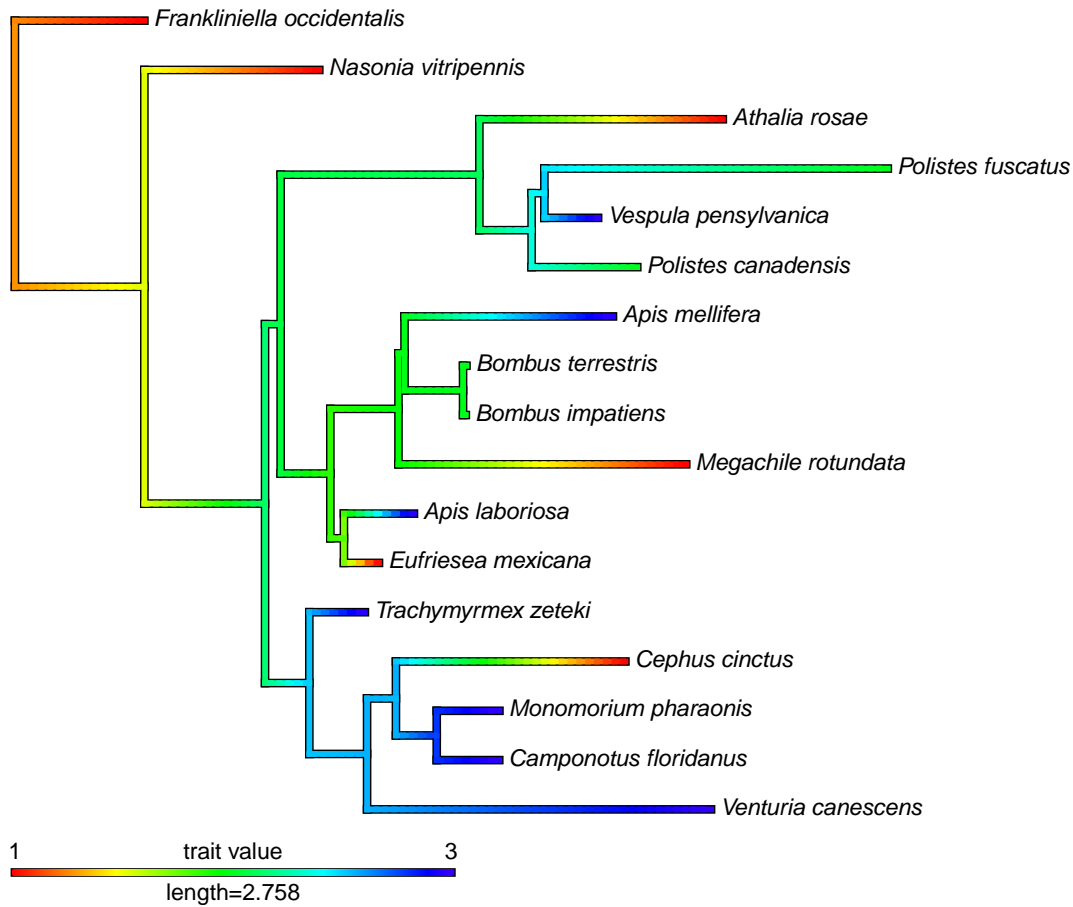
Figure IV. Continuous Character Change Heat Map

Conclusions

The objective is to see if primitive sociality is a precursor to more complex social behaviors in the Hymenoptera order, through the use of ancestral state reconstruction, treating eusociality-level as both a discrete trait and a continuous trait. Based on the results of this study, it seems that typically, solitariness is a precursor to more complex social behaviors, but that high levels of eusociality can be evolved from solitary species directly, without the need for primitive eusociality first. The only exception to this "rule" was Cephus cinctus, that evolved being solitary from high levels of eusociality.

In conclusion, more studies need to be conducted, as studies like this one have also shown "mixed" results. Further gene annotations need to be conducted on all 17 species, and other species should be added as well. If I were to continue to work on this topic, I would have field scientists study the behaviors of additional ant, wasp, and bee species that we do not yet have a full understanding of their social behaviors, in order to better link phenotypes to genotypes, along with do full genome annotations of multiple male and female members of each species.

Sources

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

[2] Plowes, N. (2010) An Introduction to Eusociality. Nature Education Knowledge 3(10):7

[3] NCBI reference sequences in the linked table