2024

Patterns of Genetic Structure in Spiders Beetles

Naiya Sims

PATTERNS OF GENETIC STRUCTURE IN SPIDER BEETLES

A Capstone Experience/Thesis Project Presented in Partial Fulfillment

of the Requirements for the Degree Bachelor of Science

with Mahurin Honors College Graduate Distinction

at Western Kentucky University

By

Naiya B. Sims

April 2024

*****

CE/T Committee:

Jarrett R. Johnson, Ph.D. (Chair)

T. Keith Philips, Ph.D.

Naomi Rowland, M.S.

ABSTRACT

Spider beetles are a poorly known group of North American beetles in the family Ptinidae. Protection of biodiversity requires knowledge of species- and population-level distinctiveness, so data regarding genetic structure of spider beetles can shed light on conservation priorities. Spider beetles in the genus, *Coleotestudus*, are distributed broadly in the southwestern US. Given their small size and lack of flight, it is presumed that the dispersal capability of *Coleotestudus* is limited, leading to strong population structure on a small spatial scale. We used a three-enzyme restriction-site-associated DNA sequencing (3RAD) strategy to identify single nucleotide polymorphisms (SNPs) throughout the genome of three species: *C. abditus, C. giuliani,* and *C. ventriculus*. We used SNP genotype frequencies to describe genetic structure among populations. Our objective was to use a multilocus SNP dataset to evaluate taxonomic questions in the genus *Coleotestudus* raised by previous work that used three gene fragments – CO1, 16S, and 28S. We evaluated the following questions stemming from an existing three-gene phylogeny: 1) Are the three sampled species best described as three distinct genetic clusters? 2) Can each species be further subdivided into distinct genetic clusters? 3) Do the clusters revealed by our analysis make sense in a geographic context? Our results suggest the three sampled *Coleotestudus* species are best described as five genetic clusters (K=5) rather than three (K=3) using both STRUCTURE and DAPC analyses. Our analysis further suggests *C. abditus* and *C. giuliani* are each best described by K=1, and *C. ventriculus* is best described by K=3. Lastly, our analyses reveal that the clusters of *C. ventriculus* appear to correspond to major biogeographic regions of the

southwestern US. Future sampling at a finer geographic scale will help to further explore the patterns of gene flow in spider beetles, particularly among populations of *C. ventriculus*.

I dedicate this thesis to my parents, Nicole and Jason Sims, who have always believed in me. I dedicate this thesis to my friends and all the people who have inspired me. Most of all, I dedicate this thesis to myself, to the countless hours in lab, and resilience when faced with challenges.

ACKNOWLEDGMENTS

I would like to thank Dr. Jarrett Johnson for his continual support and help throughout these last three years. He has always been understanding and patient, he has believed in me and known I would reach my goals. These years in his lab have been educational and have taught me various skills. This involvement allowed me to cultivate lab techniques and work on a research team. More importantly, it has taught me that patience and persistence are vital to research. I would like to thank Dr. Keith Philips for trusting me with his collected spider beetle samples and for agreeing to be my second reader. I am appreciative to Robert Chambliss for his research and 75 spider beetle samples. I would like to thank Chris DiMeo for all their guidance and support. I thank the Mahurin Honors College for believing in me and awarding me the Honors Development Grant. I thank the FUSE grant committee for the financial assistance. I would like to thank Jerica Eaton and Elizabeth Strasko for being in the lab with me and always ready to help me and answer any questions. I am indebted to Andrew Jackson for his work and efforts in library preparation. He is incredibly hardworking, and I know he will excel in future endeavors. Lastly, I am grateful for my family and friends, specifically Hemali Patel, who spent hours with me in research and told me to never give up in lab and life.

VITA

EDUCATION

Western Kentucky University, Bowling Green, KY — May 2024
- Mahurin Honors College Graduate
  - B.S. in Biology
  - B.S. in Chemistry
  - Honors CE/T: *Pattern of Genetic Structure in Spider Beetles*

Elizabethtown Community and Technical College, Elizabethtown, KY — May 2020
- Associate in Science

John Hardin High School, Elizabethtown, KY — May 2020

PROFESSIONAL EXPERIENCE

Emergency Department, HCA TriStar Regional Hospital — Jan. 2022-
- Advanced Care Technician — Present

Physics and Astronomy Department, WKU — Jan. 2022-
- Teaching Assistant — Present

Housing and Residence Life, WKU — Feb. 2021-
- Desk Assistant — Present

AWARDS & HONORS

Summa Cum Laude, WKU, May 2024
Department of Biology Outstanding Biotechnology Student Award, WKU, April 2024
Department of Biology Lancaster Scholarship, WKU, March 2024
Mahurin Honors College Honors Development Grant, WKU, October 2023
National Gilman Scholarship Study Abroad, Summer 2023
Faculty Undergraduate Student Engagement Grant, WKU May 2022
Department of Biology Outstanding Biodiversity Student Award, WKU, April 2022

PROFESSIONAL MEMBERSHIPS

Association of Southeastern Biologists (ASB)
Alpha Epsilon Delta National Pre-Health Professional Honors Society (AED)
Beta Beta Beta Biological Honors Society (Tri-Beta)

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

Biological diversity can be measured at various scales, from ecosystems and communities of interacting species to the distribution of genetic lineages across landscapes or the combination of alleles within an individual's genome. Understanding the variation comprising plants and animals and the diversity of life is the basis for many fields of scientific inquiry focusing on ecological relationships and evolutionary history. Descriptions and analysis of biological diversity can have far reaching implications such as the prioritization of conservation efforts and the discovery of novel systems that inform medical research. Our research focuses on understanding the distribution of genetic diversity using genomic sequencing data to better understand the potential need for conservation of unique spider beetle lineages. This information can be used to inform management decisions regarding the potential effects of climate change and human impacts on gene flow in spider beetles and other taxa in the southwest US.

Spider beetles are small, flightless beetles whose round bodies and long legs resemble spiders (Fig. 1). Spider beetles of the genus *Coleotestudus* are distributed throughout the arid and semi-arid regions in the western United States (Fig. 2). Dispersal capability of spider beetles is limited due to their small size and lack of flight. Their limited vagility might be exacerbated by the harsh arid environment in which they live, leading to a prediction of strong genetic structure.

Spider beetles are a diverse group of insects with varying characteristics. From the 1980s to the 2000s, the number of described species of spider beetles has increased

from 300 to approximately 600 species (Akotsen-Mensah, 2009), and the taxonomic work on spider beetles continues to expand our understanding of their diversity. Sexual dimorphism is found in some spider beetle species, for example in the genus *Ptinus*, the males tend to be slender while the females have a rounder body. Other species do not have sexual dimorphism and both male and females have either an elongated body or a stout/rounded body (Spilman, 1987). Some spider beetles are pests and can be found contaminating various dry foods, while some species are scavengers and can survive on animal and plant material. Spider beetles have been found located around bee and bird nests, caves and rock shelters containing vertebrate residents, as well as in businesses that handle and process food (Spilman, 1987). Spider beetles seek out humid environments and are often found near avian and mammal excrement, due to moisture and availability of nutrients.

Over recent decades, there has been several investigations of spider beetle phylogenetics and debate regarding spider beetle taxonomy (e.g. Philips 2000, Bell and Philips 2012). Morphological traits are frequently used to describe species and group species into genera, but molecular investigations have been limited to small number of gene sequences. Additional molecular genetic research is needed to better understand genetic relatedness and discontinuities that could affect our understanding of evolutionary relationships among spider beetle species. A recent inquiry has hypothesized potential "phylospecies" in some spider beetle lineages, which may represent cryptic species diversity, evolutionarily significant units, or distinct population segments (Chambliss 2022).

The previous work by Chambliss (2022) entailed the collection of spider beetle samples throughout the dry and arid environments of the southwest US that contain the vegetation and vertebrate excrement required to support spider beetle populations. Samples were collected from Texas, New Mexico, Arizona, California, Nevada, Colorado, Utah, and Wyoming (Fig. 2). Chambliss (2022) sequenced two mitochondrial genes (CO1 and 16S) and one nuclear gene (28S) and built a concatenated gene tree. The gene tree indicated the potential for "phylospecies" within two of the three widely distributed species used in the present study: *Coleotestudus abditus* and *C. ventriculus* (Fig. 3). In contrast, *C. giuliani* demonstrated a lack of distinct lineages (Fig. 3).

The objective of this study is to test the taxonomic hypotheses raised by the phylogeny of Chambliss (2022) using a multilocus dataset comprising single nucleotide polymorphisms (SNPs) distributed throughout the genome of *C. abditus, C. giuliani,* and *C. ventriculus*. We hypothesize that the multilocus SNP dataset will support the monophyly and distinctiveness of the three sampled *Coleotestudus* species by revealing three distinct population clusters and a lack of admixture between species. We further predict that clustering of SNP genotypes within each species will reveal distinct population segments and reveal genetic structure that may support the phylospecies proposed by Chambliss (2022). Lastly, we predict that the genetic clusters revealed at the sub-species level will correspond to geographic regions in the southwest US and contribute to our understanding of gene flow in this group.

Our data will contribute to the protection of biological diversity of spider beetles and similar species. It will also reinforce existing views on species-level diversity, answer questions regarding existing phylogenetic trees, and provide molecular data regarding

contemporary patterns of admixture among distinct populations or species. Furthermore, this data is important in identification of distinct genetic clusters, specifically to see if there are isolated populations, and geographical barriers to gene flow that would limit genetic exchange among populations that could hinder adaptation in changing environments.

METHODS

Sample Collection

Spider beetles were collected as part of a previous project (see Chambliss 2022 for details). Briefly, they located terrain that showed signs of vertebrate habitation and placed traps around plants, grass, and sand dunes, near caves or rocks. Traps were baited with feces, and once captured all specimens were preserved in 100% ethanol. Chambliss (2022) collected specimens from four species in the genus *Coleotestudus* and a few specimens in the genus *Ptinus,* as outgroups, from western US states. There was sufficient material from *C. abditus, C. ventriculus,* and *C. giuliani* for inclusion in this project.

DNA Extraction

DNA extractions were performed as part of the Chambliss (2022) project to build the three-gene phylogeny. They used the E.Z.N.A. Insect DNA Kit protocol from Omega Bio-Tek (E.Z.N.A. Insect DNA Kit, 2024) using manufacturer's instructions. In short, specimens are ground using a mortar and pestle and transferred to microcentrifuge tubes. Buffers and enzymes are used to degrade proteins and RNA and samples are bound to a filter column via centrifugation to isolate the DNA (E.Z.N.A. Insect DNA Kit, 2024). Isolated DNA was eluted with a small quantity of buffer and the concentration was quantified using a NanoDrop 2000 spectrophotometer.

DNA library preparation

A DNA library was generated for each extracted DNA sample. A DNA library is simply a collection of genomic DNA fragments that are used to generate DNA sequence data for each individual. A DNA barcode or adapter index is added to each DNA fragment within an individual library, uniquely marking each individual. This allows pooling of all individual DNA libraries into a single sequencing library prior to sequencing.

The libraries for this study are quadruple indexed to allow pooling early in the preparation of the DNA sequencing library. Libraries were generated for each sample using a 3-enzyme restriction-site associated digestion (3RAD) protocol (Bayona-Vásquez, 2019). The sample concentrations were first normalized to 10ng/uL of DNA using a Qubit fluorometer, then digested using restriction enzymes (BamHI, ClaI, and MspI) to generate the genomic fragments for each sample. Individual libraries were pooled subsequent to the addition of internal adapter indices (Fig 4.) and prior to addition of outer iTrue5 and iTrue7 barcode indices (High-Throughput 3RAD Protocol, Bayona-Vásquez, 2019).

The quadruple-indexed pooled library was sent to the North Carolina State University Genomic Sequencing Laboratory for 150bp paired-end next-generation sequencing (NGS) on the Illumina NovaSeq platform. Illumina uses the indices to code the DNA and allows for genomic analysis (Fig 5.).

Statistical Analysis

Upon receiving sequence data from the genomics facility, a bioinformatics pipeline comprised of filtering and de-multiplexing in Stacks3 (Catchen, 2013) and locus identification in iPyrad4 (Eaton, 2020) was implemented. In Stacks, the command "process_radtags" was used with the following parameters: paired reads, 3RAD protocol, inline-inline barcodes, ClaI, BamHI, MspI enzymes, one adapter mismatch, and disable rad-check. Relevant iPyrad parameters were as follows: pair3rad data type, ATCGG, CGAT enzyme overhang, 1000 max cluster depth within samples, and 0.88 clustering threshold. All parameters not mentioned were left at default condition.

Clustering of SNPs was assessed using STRUCTURE5 (Pritchard et al. 2000) implemented by StructureThreader (Pina-Martins et al. 2017) and Discriminant Analysis of Principal Components (DAPC) implemented in RStudio6 (RStudio Core team, 2021). The ΔK method (Evanno et al., 2005) and the 'find.clusters()' procedure was implement in StructureSelector (Li and Liu 2018) and Adegenet2 (Jombart, 2015) R package, respectively. Relevant STRUCTURE parameters were as follows: 10,000 burnin, 100,000 iterations, 5 replicates for each K. Other parameters were left in default condition.

STRUCTURE is a software program that is used in population genetics to visualize the prevalence of patterns with the genetic samples. It assesses the clustering of samples based on allele frequencies and underlying population genetic theory. Basically, STRUCTURE attempts to group the individuals of a data set into genetic clusters that do not violate assumptions of the Hardy-Weinberg Equilibrium Principle. The Hardy-Weinberg Equilibrium assumes that the cluster (i.e., genetic population) is not under the

strong influence of mutation, random mating, gene flow, selection, and genetic drift. For some values of K, samples cluster well, and for other values, they cluster poorly. STRUCTURE uses an iterative process to probabilistically assign individuals to clusters and variation among replicate runs of this process can assess whether a particular value of K fits the patterns present in the data well or not.

Discriminant Analysis of Principal Components (DAPC) is a multivariate analysis that can identify genetic clusters based on overall similarity across all SNP loci. It identifies genetically similar individuals and groups them into clusters, K, and evaluates whether any values of K are more likely than others. It provides graphical illustration of clusters which can be portrayed as biplots with clusters represented by ellipses, or bar plots with clusters represented by colors. We followed the Jombart (2015) tutorial to implement the DAPC analysis.

RESULTS

We received 12.9 GB of read 1 data and 14.3 GB of read 2 data from North Carolina State University Genomic Sequencing Laboratory comprising sequences for 60 individual samples. Eight samples were omitted due to missing barcode combinations, and three samples were omitted due to low sequence coverage. Six sample were also omitted due to excessive file size and analysis time constraints. These latter samples will be added to the analysis at a later date. Demultiplexing of sequence data in Stacks yielded 84,527,001 raw sequence reads remaining in the 43 individual working data set for locus alignment and SNP identification. iPyrad filtering of reads, ported from Stacks, resulted in a 99% retention rate and 83,558,723 sequences for further analysis.

Following processing in iPyrad, the full data set, comprising 43 individuals across all three species, yielded 283 loci shared by at least 15 (35%) individuals. Each species was further investigated singly in iPyrad, and unique sets of loci were generated for each, using the same parameters explained above. The *C. abditus* analysis included 11 samples and yielded 254 loci shared by at least 9 (82%) individuals, the *C. giuliani* analysis included 9 individuals and yielded 232 loci shared by all 9 (100%) individuals, and the C. *ventriculus* analysis included 18 samples and yielded 259 loci shared by 12 (67%) individuals. One *C. ventriculus* sample was removed due to excessive file size, and one *C. abditus* sample was removed for low sequence coverage.

DAPC Analysis

DAPC analysis provides a line plot with "Number of Clusters" tested on the x axis and "Bayesian Information Criterion (BIC)" on the y axis. The BIC value represents the likelihood of a model given the data, and a lower score represents models that are more likely to be "true". The genetic samples are tested against models that allow for different number of clusters to infer the most likely number of population clusters contained in the data. The model with a K value resulting in the lowest BIC is most likely given the data. When multiple K values reveal low BIC values, preference is given to the most parsimonious, or smallest estimate of K. DAPC analysis of all samples using BIC and number of tested clusters results in the most likely number of clusters is K = 5 (Fig. 6A), and for *C. ventriculus* the lowest BIC occurs at K = 3 (Fig. 6B). *C. giuliani* (Fig. 6C) and *C. abditus* (Fig. 6D) do not have their lowest BIC estimate until number of tested clusters approximates the number of samples in the data set, so we default to K=1 for each.

The DAPC analysis can be presented as a bar plot with all samples color-coded based on the cluster to which an individual belongs. For the full data set analysis there are five distinct clusters (Fig. 7). *C. abditus* and *C. giuliani* are represented by a single color on the plot due to a lack of clustering at the sub-specific level, while *C. ventriculus* samples are split between three colors, each illustrating a distinct population segment. Each species had a 1.0 assignment probability (Fig. 7). When the K value for the DAPC full dataset bar plot is three, *C. giuliani* and *C. ventriculus* cluster 3 are denoted by their own colors and *C. abditus* and *C. ventriculus* clusters 1 and 2 are all represented as the same color (Fig. 8). Figure 9 shows the same K=5 bar plot from Figure 7, re-organized by

sample name for cross-referencing with Figures 2 and 3. The DAPC data are also often depicted as a biplot as is typical for principal components analysis of multivariate data. The K=5 biplot for the full data set shows grouping of sampled individuals into a single cluster for *C. abditus* and *C. giuliani* and three clusters for *C. ventriculus* (Fig. 10). The K=3 biplot for the full data set shows *C. giuliani* and *C. ventriculus* cluster 3 as their own clusters and *C. abditus* and *C. ventriculus* clusters 1 and 2 are all clustered together (Fig. 11). The K = 3 biplot of *C. ventriculus* excludes the other two species, and allows clearer visualization of *C. ventriculus* clusters 1, 2 and 3 (Fig. 12).

STRUCTURE Analysis

Replicate STRUCTURE runs across numerous values of K can be visualized as line plots showing the average log-likelihood values (LnP(K)) for each K value (Fig 13). Additionally, a ΔK value is generated by a formula proposed by Evanno et al. (2005) and represents an evaluation of variance associated with moving from a particular value of K–1 to K. During a preliminary analysis for all samples (Fig. 13A), the maximum Mean LnP(K) is approximately -2000 at K = 5 and ΔK = 3, which may indicate support for three-clusters in the data despite the highest likelihood value being at K=5. The large margin of error for K = 4, ranging from -4000 to -1000, indicates a lack of support for four genetic clusters in the data. In a subsequent analysis including values of K ranging up to K = 10, the average log-likelihood values (LnP(K)) begin to plateau and the highest likelihood value at K = 8 (Fig. 13E). In this extended analysis, for all samples, ΔK = 5 has the highest support, which suggests that distinct population segments exist in *C. ventriculus*, while *C. giuliani* and *C. abditus* are each best represented as a single genetic

cluster. For *C. ventriculus* (Fig. 13B) the Mean LnP(K) is slightly larger than -1500, with a K = 4 and a ΔK = 3. There are multiple small error bars for K greater than 4, the largest error range is approximately 500. For *C. giuliani* (Fig. 13C) the Mean LnP(K) is roughly -750 and the K value with the highest likelihood is K=9, which is the same number of samples in *C. giuliani*. Furthermore, the ΔK=2. There is an error bar for both K = 8 and K=9. For *C. abditus* (Fig. 13D) the peak Mean LnP(K) is -750 with a K=10, and a ΔK=3.

The STRUCTURE bar plot depicts cluster assignment probabilities for each individual. For the whole data set at K = 5 most of the samples have a 1.0 assignment probability into one of the five clusters, while some of the subjects show genomic admixture and assignment to multiple clusters (Fig. 14). For the whole data set at K = 3, *C. ventriculus* 2 and 3 are genetically clustered and most of the *C. giuliani* samples have a 1.0 assignment probability, whereas *C. abditus* has admixing with *C. ventriculus* 1 (Fig. 15). Using the data from the STRUCTURE assignment probabilities, the function 'floating.pie()' in the R package 'plotrix' (Lemon 2006) was used to plot the colored bar plot data onto a map of the collection localities (Fig. 16). Each sample has a location associated with the subject. Samples align into clusters that make sense geographically (Fig. 16A), specifically, *C. giuliani* is found in California, Nevada, Utah, and Wyoming, and *C. abditus* samples are concentrated in Utah, Colorado, Arizona, and New Mexico. Some of the samples show admixing with *C. ventriculus*, especially where clusters are overlapping. The *C. ventriculus* 1 cluster are concentrated in upper Texas with few species along the border to New Mexico. The *C. ventriculus* 3 cluster is dispersed in the midwestern states, localized in Utah, Arizona, New Mexico, and Texas (Fig. 16B). Some

C. ventriculus 3 samples have similar genetic components to the *C. ventriculus* 2 cluster found in California.

DISCUSSION

Our goal in next-generation sequencing of spider beetles is to reveal genetic structure of *C. ventriculus*, *C. abditus*, and *C. giuliani* collected from the western US. Additionally, the molecular data can be compared to earlier work, such as Chambliss (2022). The sequences show supporting and conflicting data to the suggested phylospecies' of Chambliss' (2022). Particularly, our results demonstrate a lack of admixture between the three species under current taxonomy, but also reveal important clustering of genetic diversity within *C. ventriculus*, supported by a logical geographic context.

There is sufficient evidence for species-level genetic structure in *Coleotestudus*. The ΔK value from STRUCTURE analysis of full data set suggests ΔK = 5, corresponding to the three named species in this study, and two additional clusters within *C. ventriculus*. The DAPC clustering analysis suggest K = 5 is the best K and retains the distinctiveness of the three named species. But for DAPC, K = 3 does not recover the three named species, and instead lumps *C. abditus* with *C. ventriculus* clusters 1 and 2. K=3 for STRUCTURE also lumps *C. abditus* with *C. ventriculus* cluster 1. This is indicative of major divergence within *C. ventriculus* equivalent to level of divergence between the three currently named species.

There is some evidence for distinct population segments or "phylospecies" in *Coleotestudus*. Genetic structure is apparent within *C. ventriculus* from both STRUCTURE and DAPC analyses, specifically into three distinct population segments.

ΔK values from STRUCTURE suggest the possibility of clusters within *C. giuliani* and *C. abditus* but the relation needs further investigation. Within both species, there is a possibility that genetic structure exists at a fine scale (i.e., each individual is from a different cluster), but our sampling density is insufficient to investigate this fully. Overall, our data match the conclusions of Chambliss (2022) for *C. giuliani* and *C. ventriculus* but not for *C. abditus*. Essentially our interpretation foregoes the delineation of phylospecies for *C. abditus* as described in Chambliss (2022), but our data does not contradict the overall pattern of their three-gene phylogeny.

The genetic variation in *Coleotestudus* in a geographic context only pertains to *C. ventriculus* based on earlier conclusions. STRUCTURE pie chart mapping illustrates the three clusters are largely non overlapping and may correlate with major biogeographical features of southwest US. For instance, there have been descriptions of limited gene flow in stoneflies (Peterson, 2017) due to the Sierra Nevada Mountain Ranges. This terrain overlaps with the species *C. ventriculus* cluster 1 and *C. ventriculus* cluster 3, suggesting similar separation.

Our research can be used in further genomic sequencing and investigation of spider beetles. Specifically analyzing their isolated populations and evolutionary biology. These results can be used in future research, to build upon our work with 3RAD and spider beetles. Within spider beetles this can shed light on the gene flow and further phylospecies within microhabitats. From a conservation perspective, isolated populations are subject to greater risk of local extirpation, with rapidly changing global climate further increasing this risk in already harsh arid environments. Further research based on our results can help scientists understand how human and environmental influence affect

spider beetle populations and their ability to maintain connectivity and genetic diversity via gene flow. It could provide a basis for mutations or specific segments of DNA that a scientist was investigating. There are several directions a scientist could explore with spider beetles; this research is one method of investigation.

CONCLUSION

In conclusion, the three-enzyme restriction-site-associated DNA sequencing (3RAD) strategy was successful in identifying single nucleotide polymorphisms (SNPs) throughout the genome of three species within the genus *Coleotestudus*. Both analysis strategies (STRUCTURE and DAPC) yielded highly concordant results. For the three species studied, *C. abditus, C. giuliani, and C. ventriculus,* K=5 was the best supported value for species-level variation. Our multilocus dataset supports *Coleotestudus abditus*, *C. giuliani*, and *C. ventriculus* as non-admixing distinct species. Our analysis further suggests *C. abditus* and *C. giuliani* are each best described either by K=1 or by a K value approximating the number of sampled sites. For *C. abditus* and *C. giuliani* additional sampling is required to evaluate the presence of cryptic species-level variation and we did not recover adequate support for the "phylospecies" groups of the earlier three-gene phylogeny for *C. abditus*. For *C. ventriculus* we concluded significant differentiation exists among three, geographically separated subpopulations, even with our limited sampling, sufficient genetic structure exists to warrant further investigation regarding cryptic species-level variation in *C. ventriculus*. The genetic structure present in our *C. ventriculus* dataset supports the "phylospecies" presented in the prior phylogeny. Continued research and sampling will help to further explain genetic structure and gene flow in spider beetles. Our research is only the beginning of genomic investigation that can be completed for spider beetles.

REFERENCES

Akotsen-Mensah, C., & Philips, T. K. (2009). Description of a new genus of spider beetle (Coleoptera: Ptinidae) from South Africa. Zootaxa, 2160(1), 51-67.

Bayona-Vásquez, N. J., Glenn, T. C., Kieran, T. J., Pierson, T. W., Hoffberg, S. L., Scott, P. A., ... & Faircloth, B. C. (2019). Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). PeerJ, 7, e7724.

Bell, K. L., & Philips, T. K. (2012). Molecular systematics and evolution of the Ptinidae (Coleoptera: Bostrichoidea) and related families. Zoological Journal of the Linnean Society, 165(1), 88–108.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. Molecular Ecology, 22(11), 3124-3140.

Chambliss, R. (2022). Description, diversity, and phylogeography of the North American spider beetle (Coleoptera: Ptinidae) genus Coleotestudo. [Master's Thesis, Western Kentucky University]. TopScholar Research and Creative Activity Database of WKU. https://digitalcommons.wku.edu/theses/3601/

Eaton, D. A., & Overcast, I. (2020). ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics, 36(8), 2592-2594.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

E.Z.N.A.® Tissue DNA kit - Solutions for DNA and RNA Purification. Omega Bio-tek. (2024, February 13). https://omegabiotek.com/product/genomic-dna-kit-e-z-n-a-tissue-dna-kit/?cn-reloaded=1

Gearner, O. M. (2019). A phylogenetic analysis of Bostrichoidea (Coleoptera) and revisions of the Southern African spider beetle genera Meziomorphum and Eutaphroptinus (Ptinidae: Coleoptera).

High-Throughput 3RAD Protocol. The University of Georgia. (n.d.). https://www.biorxiv.org/content/biorxiv/suppl/2017/10/19/205799.DC1/205799-4.pdf

Lemon, Jim. (2006). Plotrix: A package in the red light district of R. R-News. 6. 8-12.

Li, Y. L., & Liu, J. X. (2018). StructureSelector: A web-based software to select and visualize the optimal number of clusters using multiple methods. Molecular ecology resources, 18(1), 176–177.

Jombart, T., Kamvar, Z. N., Collins, C., Lustrik, R., Beugin, M.P., Knaus, B.J., & Jombart, M. T. (2018). Package 'adegenet'. Github repository:< https://github.com/thibautjombart/adegenet.

Next-generation sequencing (NGS). Explore the technology. (n.d.). https://www.illumina.com/science/technology/next-generation-sequencing.html

Peterson, M. G., O'Grady, P. M., & Resh, V. H. (2017). Phylogeographic comparison of five large-bodied aquatic insect species across the western USA. Freshwater Science, 36(4), 823–837.

Philips, T. K. (2000). Phylogenetic analysis of the New World Ptininae (Coleoptera: Bostrichoidea). Systematic Entomology - SYST ENTOMOL. 25. 235-262.

Pina-Martins, F., Silva, D. N., Fino, J., & Paulo, O. S. (2017). Structure_threader: An improved method for automation and parallelization of programs structure, fastStructure and MavericK on multicore CPU systems. Molecular ecology resources, 17(6), e268–e274.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Spilman, T. J. (1987). Spider beetles (Ptinidae, Coleoptera). Insect and Mite Pests in Food: An Illustrated Key, (655), 137-148.

FIGURES

Figure 1. Two representative spider beetles, *Coleotestudus ventriculus* (top) and *C. abditus* (bottom). Photos by K. Philips.

Figure 2. Map of sampling sites for *Coleotestudus abditus*, *C. giuliani*, and *C. ventriculus* in the Western US. The red colors correlate to the species *C. giuliani*, blue is *C. abditus*, and orange represents *C. ventriculus*. The cross marks represent locations where there was low sample collection, less than 5 samples collected. The dots indicate a moderate number of samples collected, more than 5, less than 10. The check marks denote locations where there were more than 10 samples collected. Throughout all species and locations, some locations contained only one collected sample to 80 collected samples.

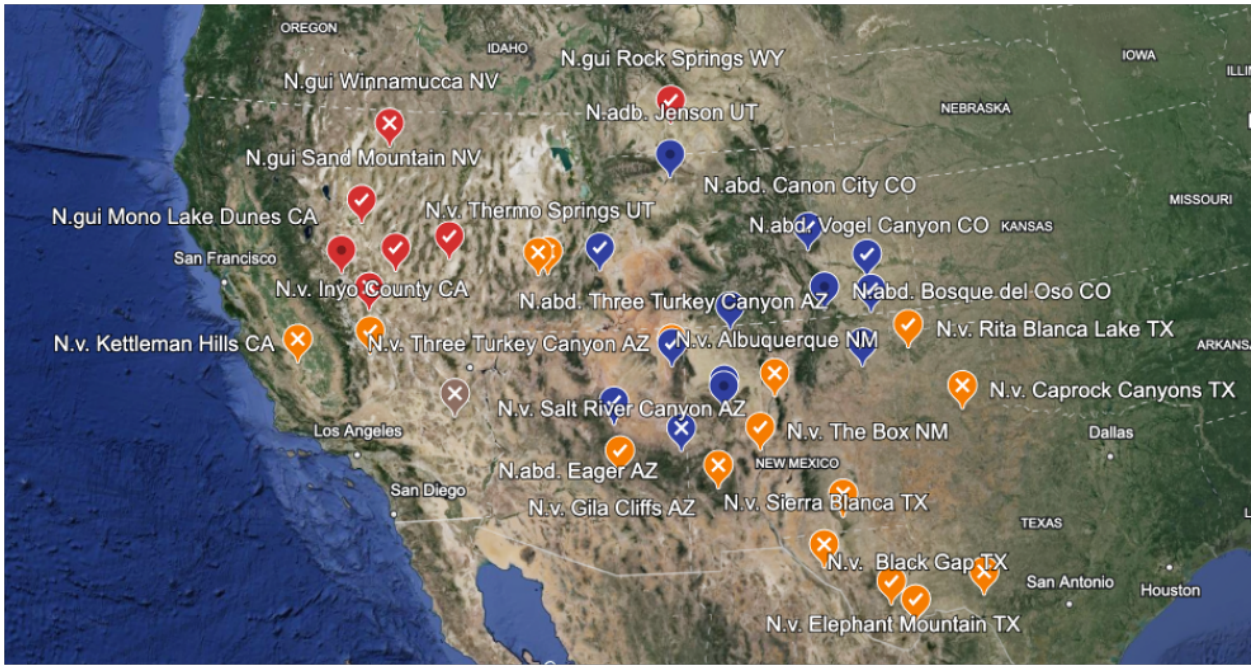

Figure 3. Bayesian phylogenetic tree of the genus *Colestestudus* showing three main species-level clades and hypothesizes population structure depicted as "phylospecies" by Chambliss (2022).

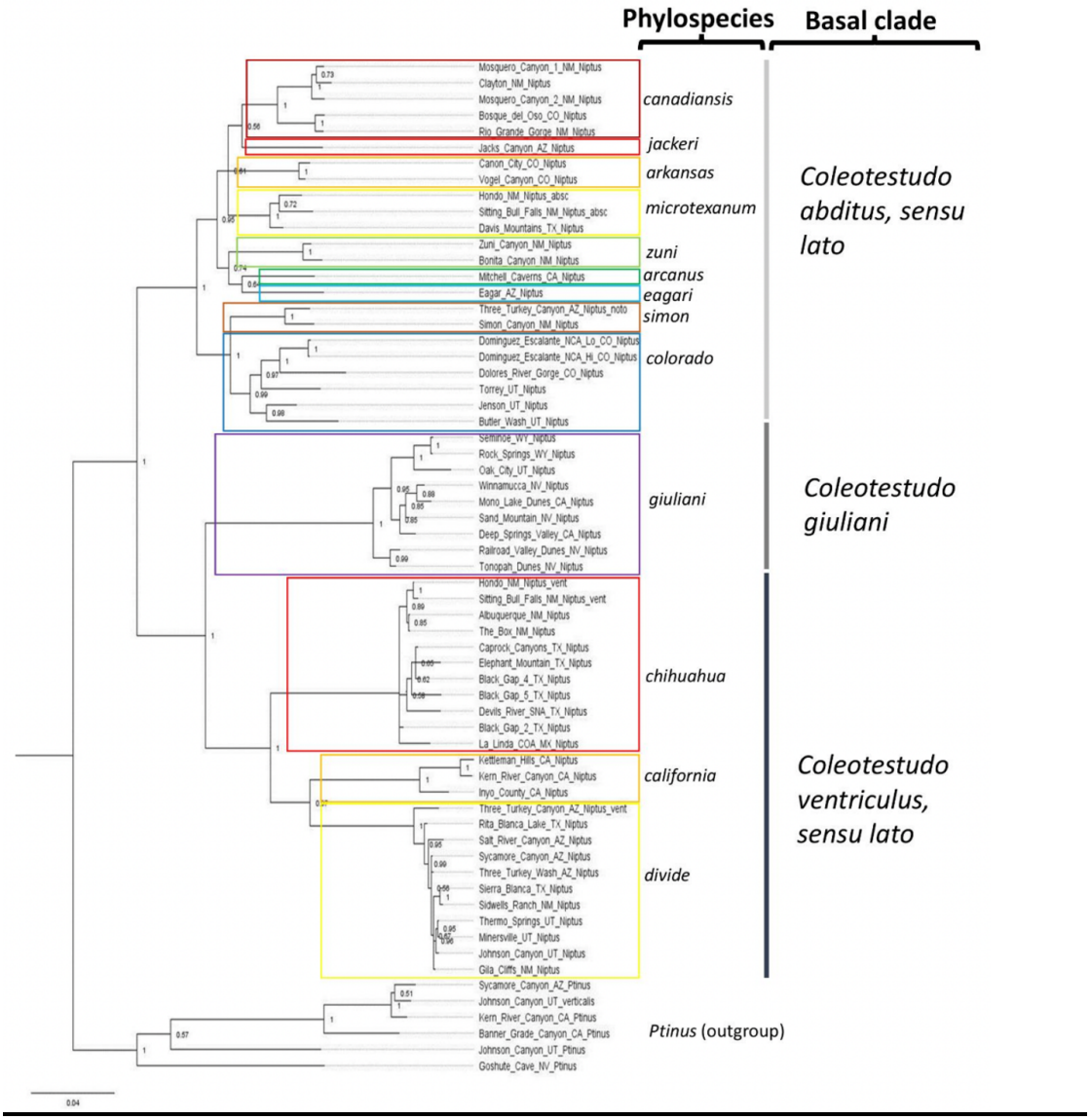

Figure 4. Visual representation of 3RAD Library Protocol. Demonstration of digestion, ligation of adaptors and indices to create DNA library. Figure from Bayona-Vásquez, 2019.

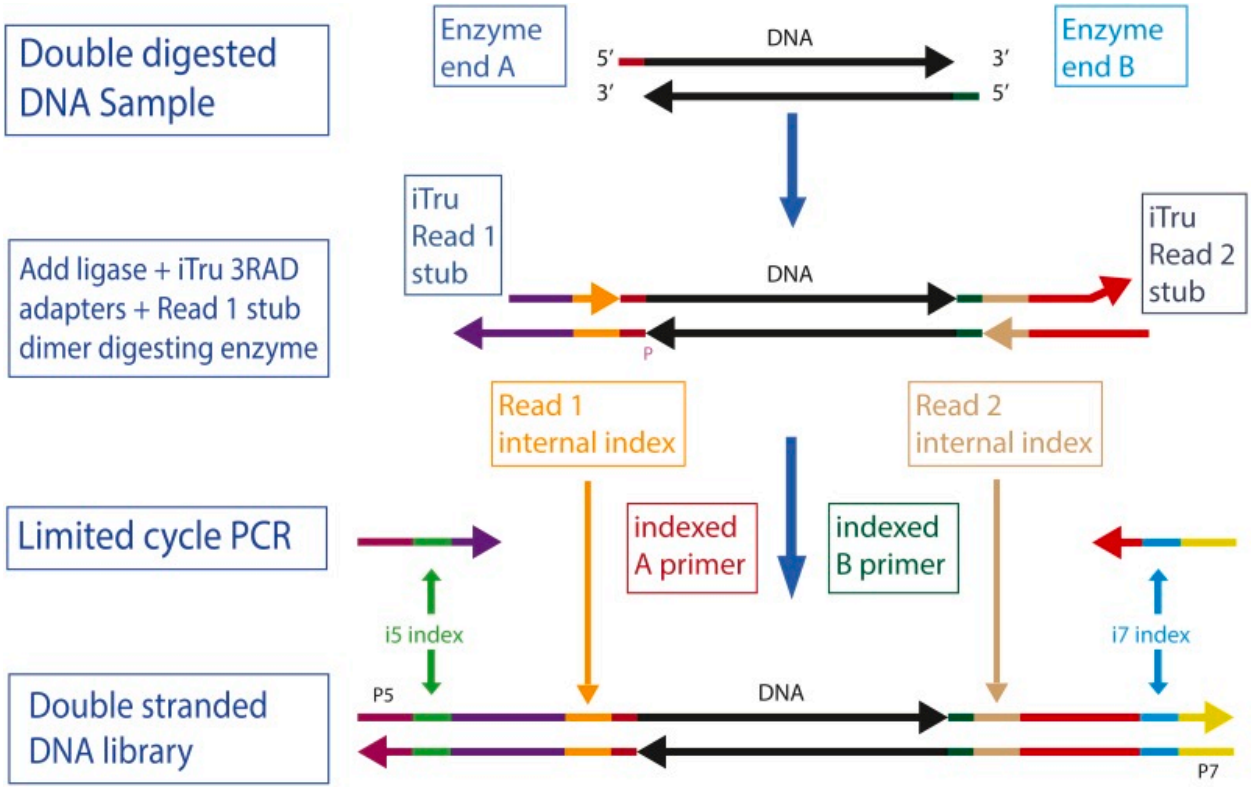

Figure 5. Illumina Next Generation Sequencing Steps. A. Library Preparation, B. Cluster Amplification, C. Sequencing, D. Alignment and Data Analysis. Figure from, *An introduction to Next-Generation Sequencing Technology*, Illumina, 2024.

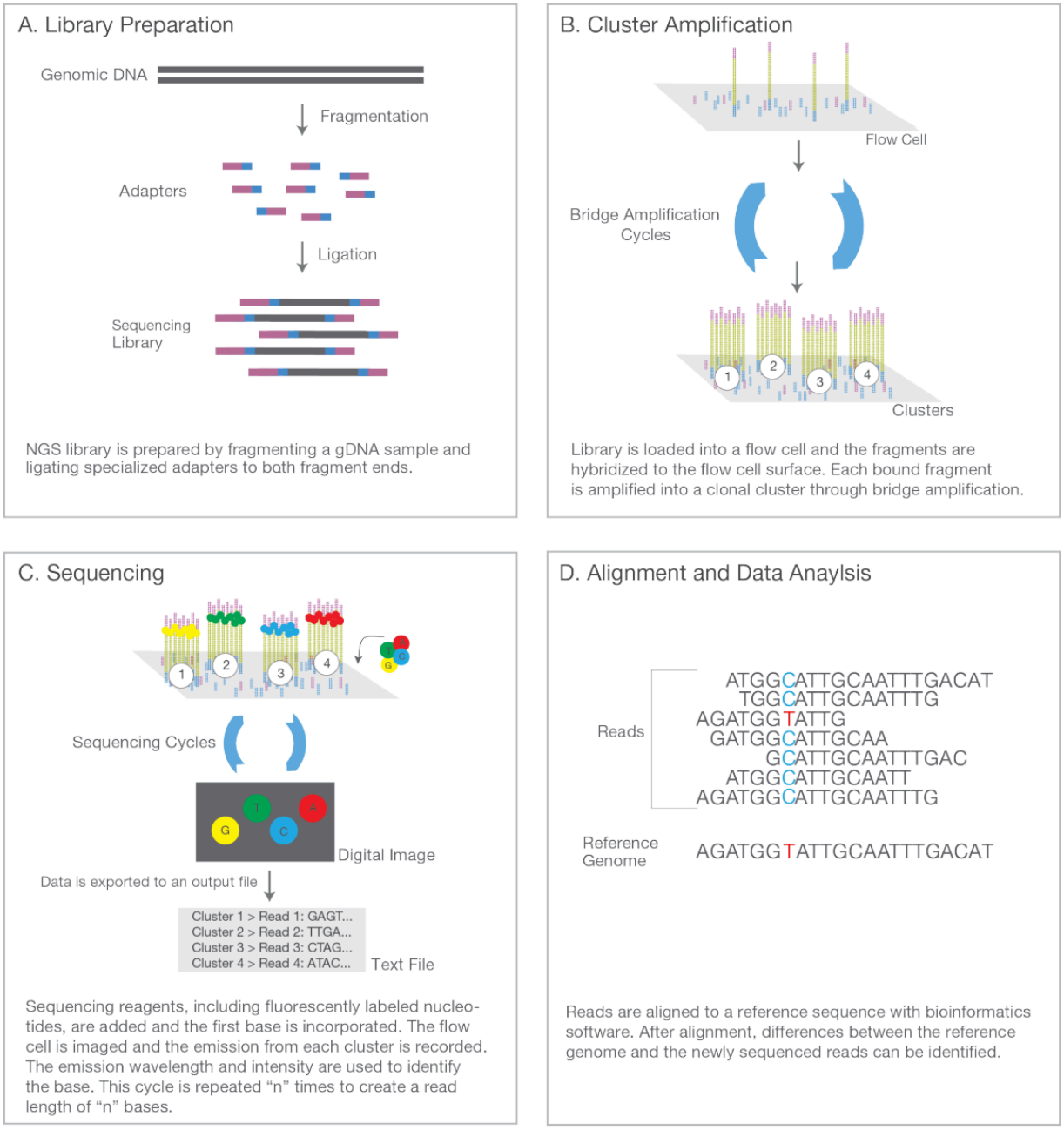

Figure 6. DAPC Line Plots. The line plot of Bayesian Information Criterion (BIC) for all tested number of clusters (K). The K value with the lowest BIC estimate is denoted by the red arrow.

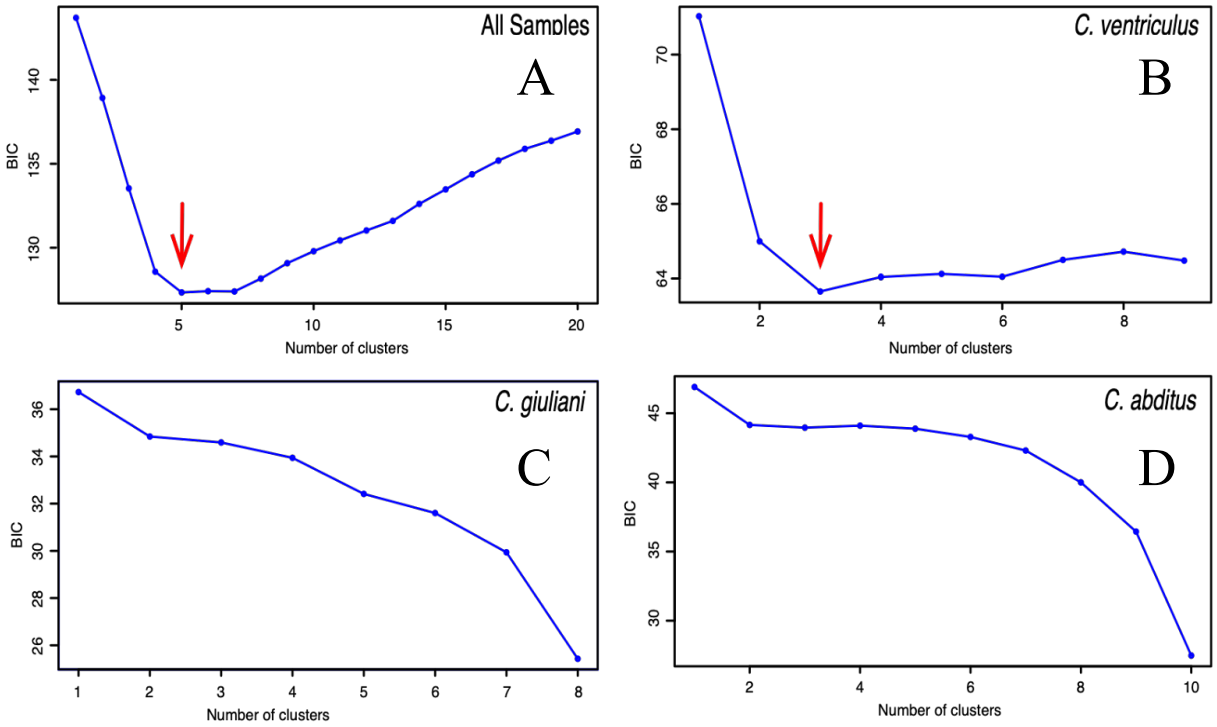

Figure 7. Full dataset DAPC bar plot. Bar plot showing individual assignment probabilities to each cluster (denoted by color) for K=5. Individuals are sorted.

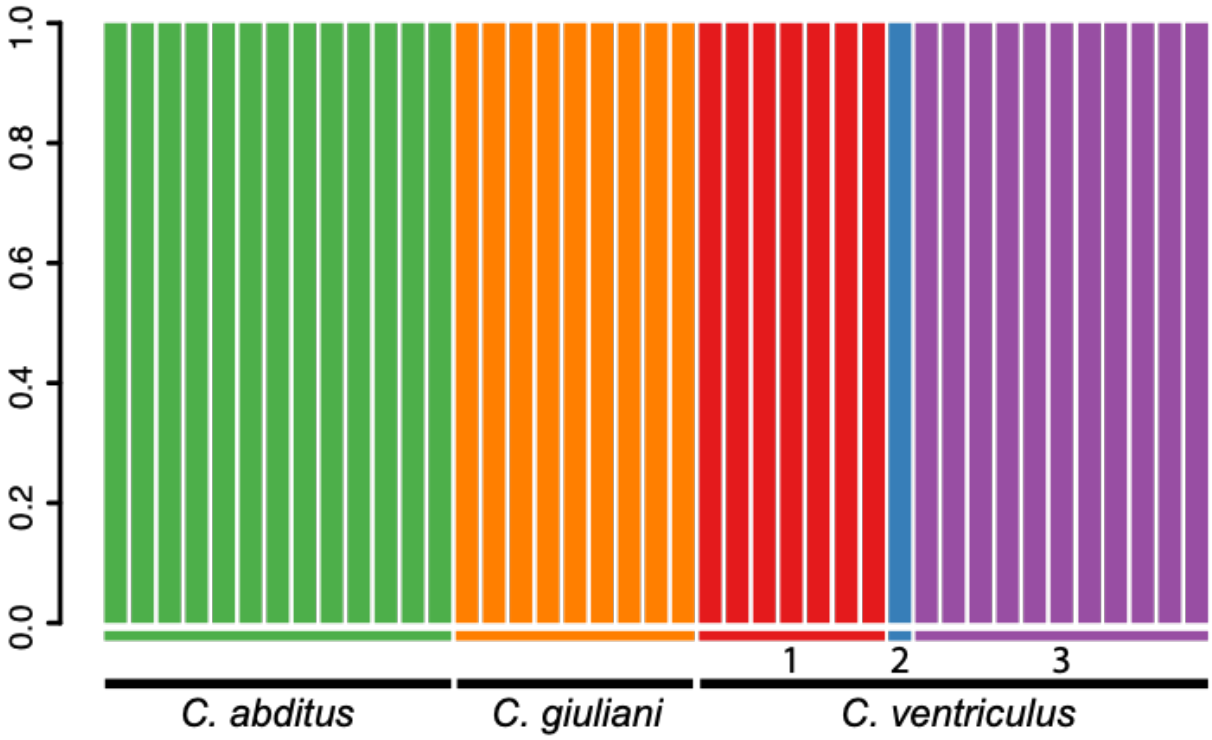

Figure 8. K=3 DAPC Bar Plot. The bar plot includes the full dataset, and the K value was set to three. The bar plot shows the individual assignment probability to the three genetic clusters (denoted by color).

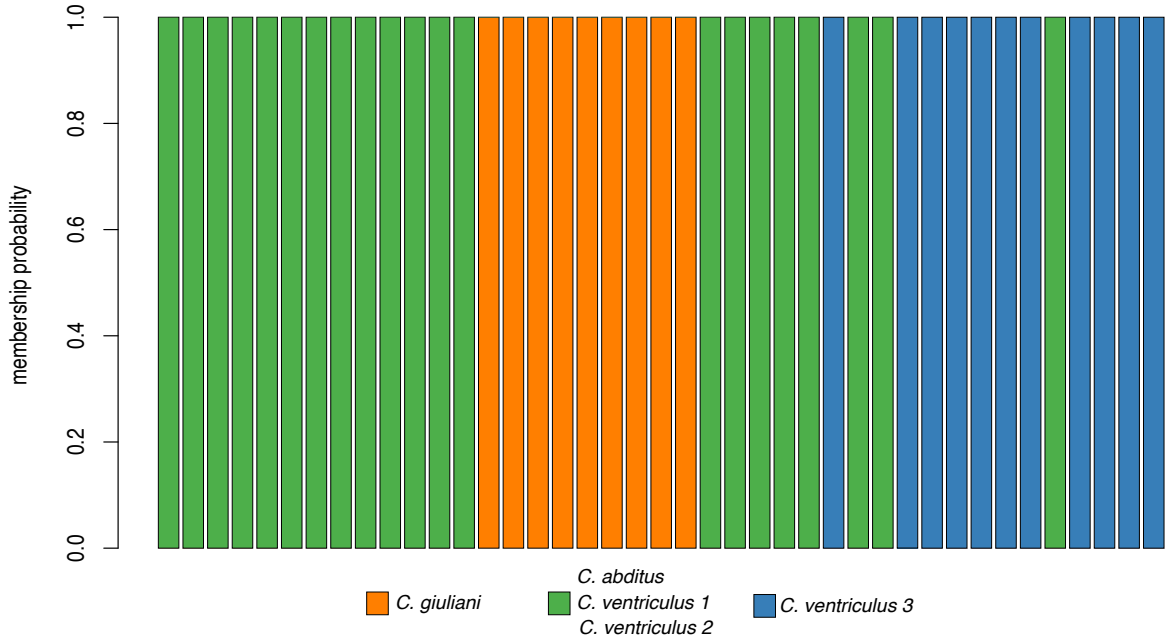

Figure 9. Full dataset DAPC bar plot with sample names. Bar plot showing individual assignment probabilities to each cluster (denoted by color) for K=5. Each sample is labeled with sample name, they correspond to location sampled from. *Coleotestudus abditus* is in green, *C. giuliani* is denoted by orange, *C. ventriculus* 1 is red, *C. ventriculus* 2 is blue, and *C. ventriculus* 3 is purple. Individuals are unsorted.

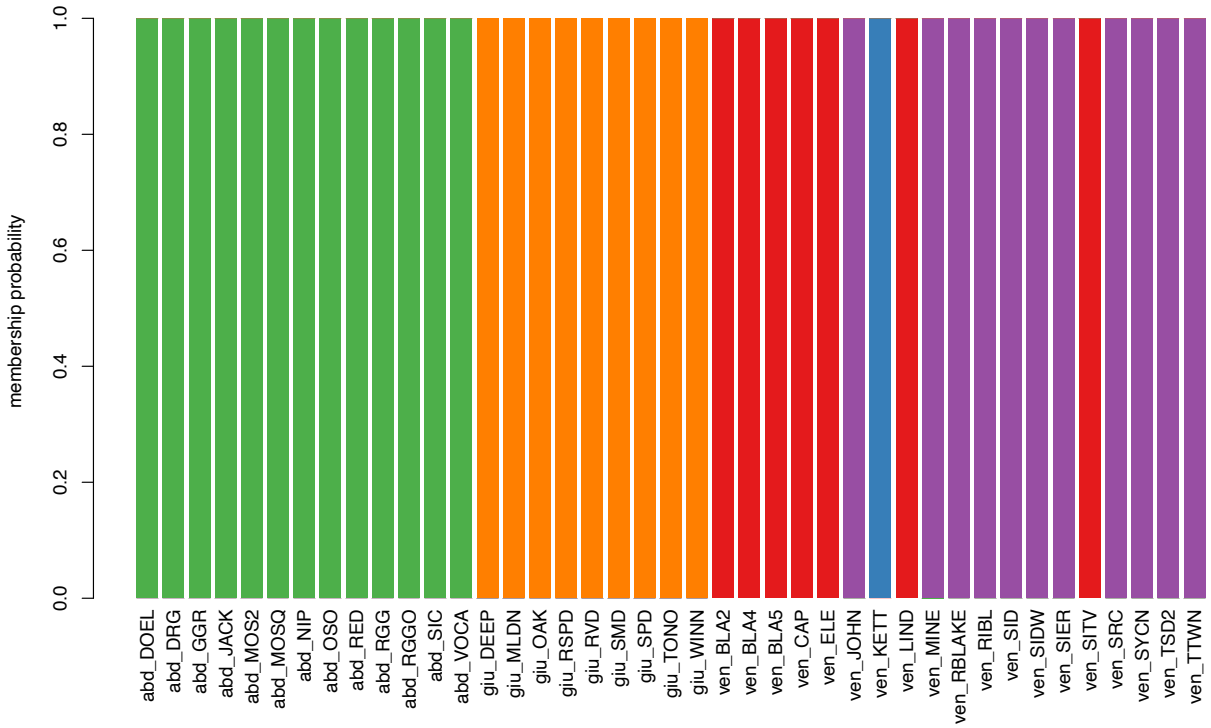

Figure 10. Full dataset DAPC Biplot K=5. K=5 genetic clustering from analysis of full data set

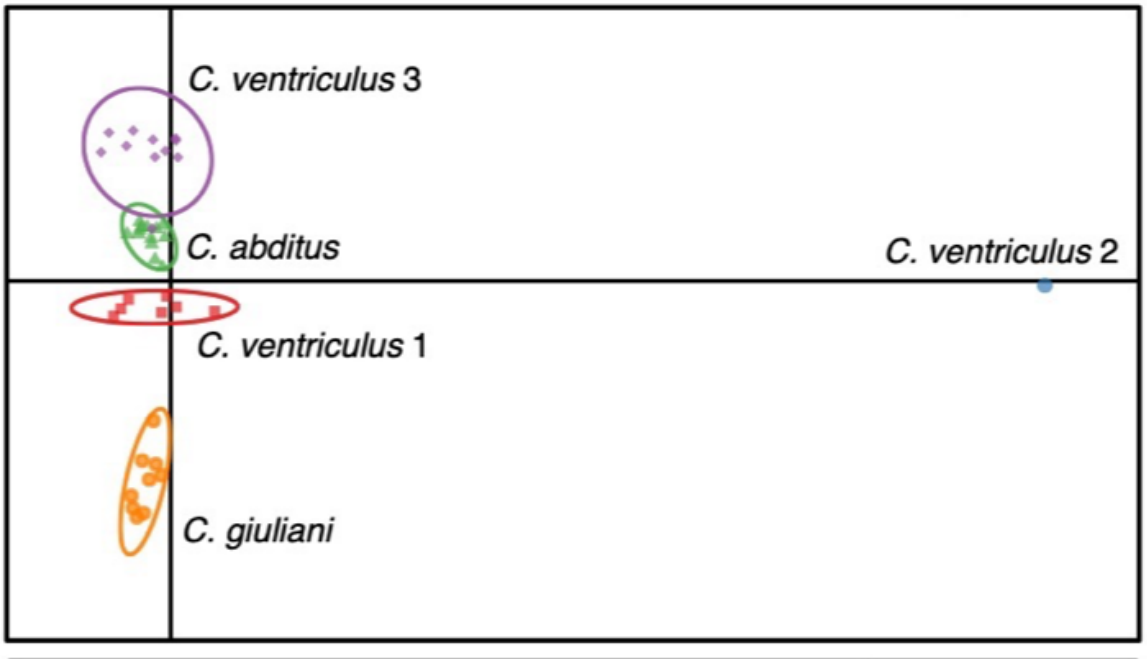

Figure 11. Full dataset DAPC Biplot K=3. K=3 genetic clustering from analysis of full data set.

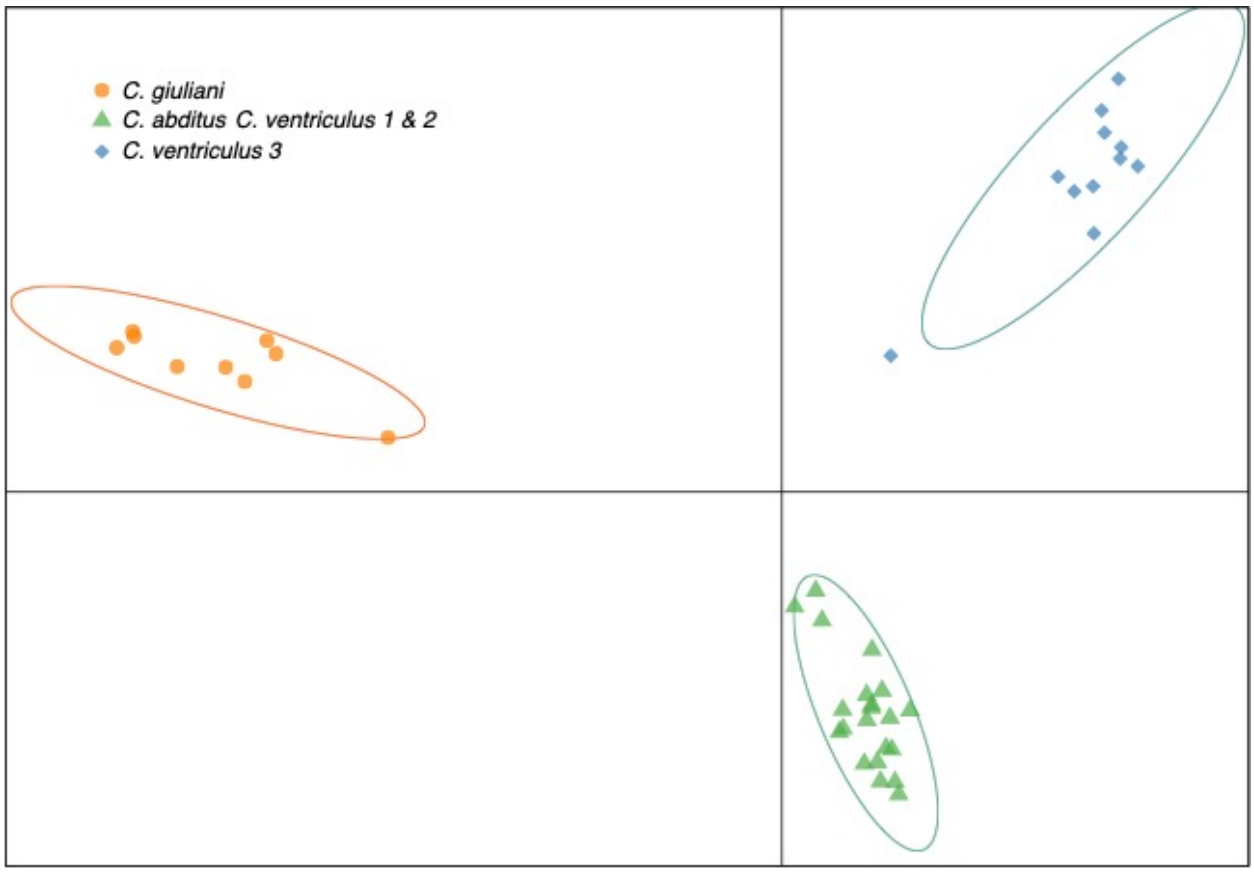

Figure 12. *C. ventriculus* DAPC biplot K=3. K=3 genetic clustering analysis of *C. ventriculus* data.

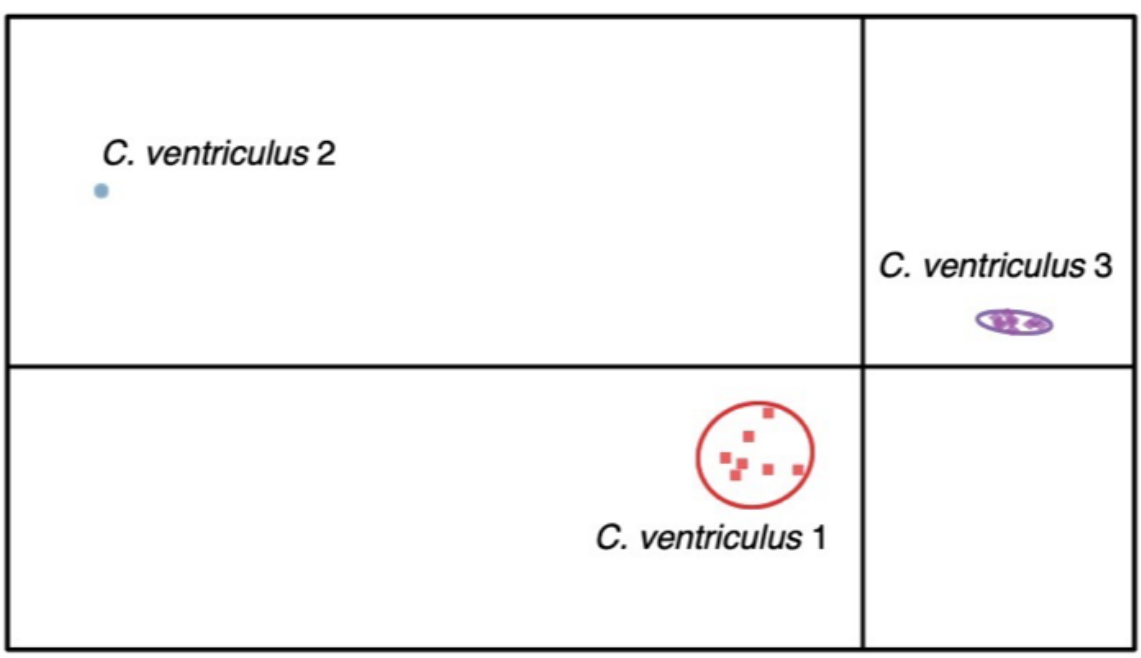

Figure 13. STRUCTURE Line Plot. Line plots of log-likelihood values (Ln P(K)) for all tested number of clusters (K). Error bars represent variation across five replicates. The K value with the highest likelihood estimate is bracketed in red. These plots are part of the STRUCTURE analysis. A contains all samples, B consists of *C. ventriculus*, C consists of *C. giuliani*, D contains *C. abditus.* E contains all samples when K is higher. F is the ΔK calculation for all samples with population models.

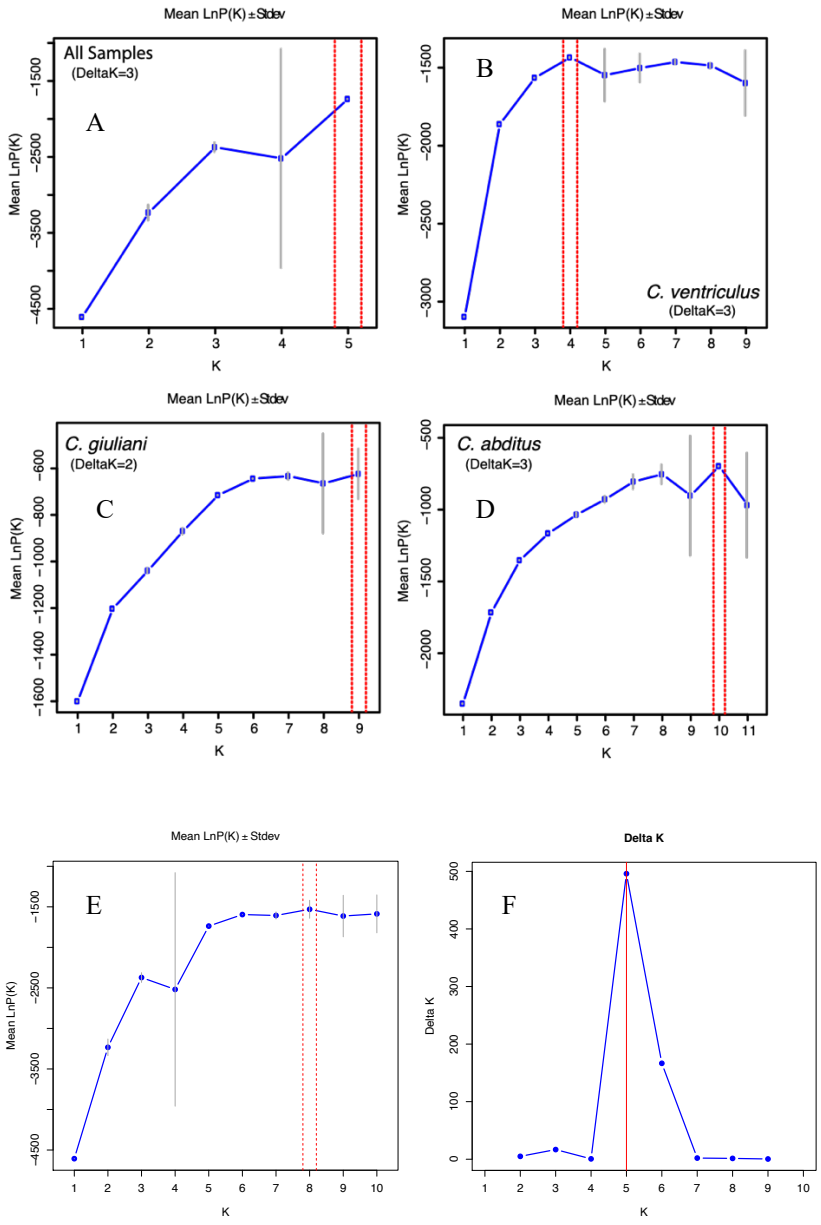

Figure 14. K=5 STRUCTURE bar plot. The bar plot includes the full data set and shows individual assignment probabilities to each cluster (denoted by color) for K=5. Individuals are unsorted.

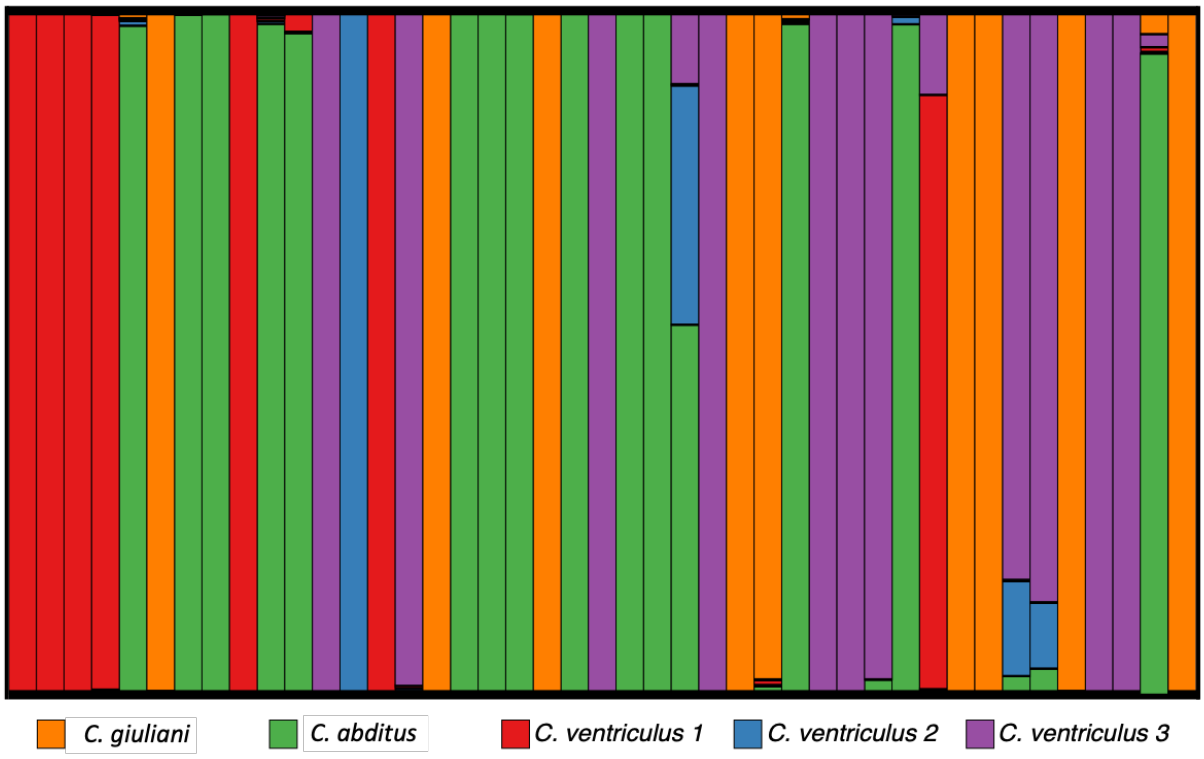

Figure 15. K=3 STRUCTURE bar plot. The bar plot includes the full dataset and shows individual assignment probabilities to each cluster (denoted by color) for K=3.

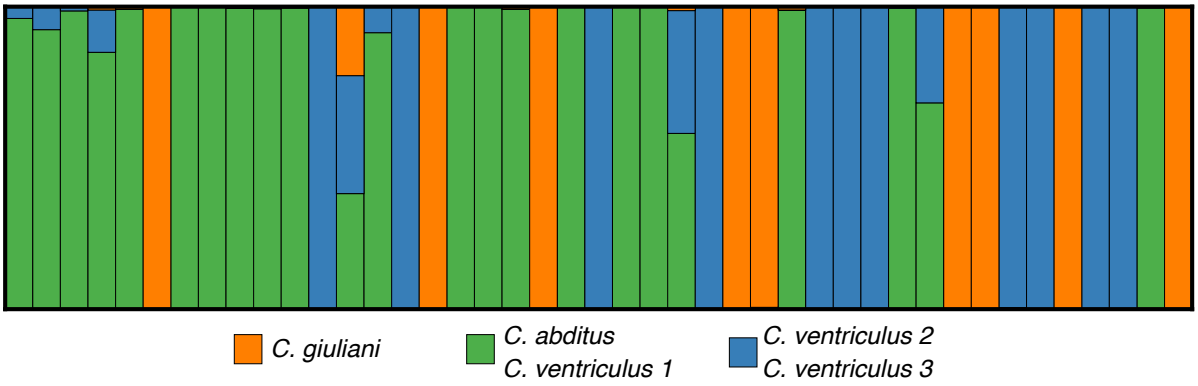

Figure 16. STRUCTURE Pie Chart Map of A) K=5 genetic clustering from STRUCTURE analysis of full data set, and B) K=3 genetic clustering from STRUCTURE analysis of *C. ventriculus* data.