Gene to Genus: Systematics and Population Dynamics in Lamiini Beetles (Coleoptera: Cerambycidae) With Focus on Monochamus Dejean

GENE TO GENUS: SYSTEMATICS AND POPULATION DYNAMICS IN LAMIINI BEETLES (COLEOPTERA: CERAMBYCIDAE) WITH FOCUS ON *MONOCHAMUS* DEJEAN

A dissertation presented

by

Patrick Scott Gorring

to

The Department of Organismic and Evolutionary Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biology

Harvard University

Cambridge, Massachusetts

April 2019

GENE TO GENUS: SYSTEMATICS AND POPULATION DYNAMICS IN LAMIINI BEETLES (COLEOPTERA: CERAMBYCIDAE) WITH FOCUS ON *MONOCHAMUS* DEJEAN

Abstract

Herbivorous insects make up more than a third of the multicellular species on Earth. The discovery, description, and organization of this diversity is necessary to enable study of the mechanisms involved in the dynamic speciation process of phytophages. In this dissertation, each of these levels is addressed using molecular systematics of the tribe Lamiini and genus *Monochamus*, and population genomics with a focus on differentiation in the sky island inhabiting pine sawyer, *Monochamus clamator*. With this data, we have evaluated several variables that could promote herbivore speciation including geography, host-plant diversity, and climate.

The genus *Monochamus* Dejean has long been considered to be a combination of species that do not belong in the same genus, but morphological characters have failed to delineate natural groups. We are the first to use multi-gene molecular data and coalescent modeling to estimate the phylogeny of this economically-important genus. *Monochamus rev. nov.* is a monophyletic group of Holarctic conifer-feeding species based on this data. The angiosperm feeding species currently placed in *Monochamus* are revealed to belong to other genera. We find that *Monochamus* is a derived conifer feeding genus that likely originated in southern Asia in the late Miocene and dispersed over the second Bering Bridge to North America and subsequently

diversified. Modern multispecies coalescent species delimitation techniques exposed varied evolutionary histories for current species, that some subspecies are unsupported and others should be elevated to species status, and helped to discover a new species of *Monochamus*. The approach of phasing nuclear sequences to alleles for phylogeny and delimitation revealed that phylogeny node heights are impacted at multiple taxonomic levels by ambiguity codes in sequences. This can lead to incorrect divergence times and delimit incorrect numbers of species.

The tribe that *Monochamus* belongs to, the Lamiini, is diverse and shares morphological characteristics with other tribes in the Lamiinae subfamily. We use molecular data to build a "backbone" phylogeny of the Lamiini and related tribes to test for evolutionary independence and determine if any morphological characters support the result. The data supports the tribes Batocerini, Gnomini, Monochamini, and Acridocephalini being synonyms of the Lamiini *sensu novo*. There are no synapomorphic morphological characters found to support the new tribe, but a group of characters will place most specimens and therefore allow more efficient identifications in this economically impactful and widespread group.

At the population level, we address adaptive processes in *Monochamus clamator* in the climate-change produced Great Basin sky islands of western North America. With the use of RNA-Seq to build a genomic and gene expression dataset for multiple habitat islands with differing host plant composition, the relationship between genetic differentiation and habitat factors was explored in a geographically explicit framework. Genomic data revealed low genetic differentiation at the island level and a high importance of immigration between islands across the Basin. Multi-matrix regression on factors of geography, host-plant diversity, and environment showed that only the environment has a significant relationship with genetic

distance in the beetles. Gene expression measures reveal hundreds of differentially expressed genes between island beetle groups that may be the first sign of adaptation to their habitat.

As a whole, this work contributes novel findings concerning the description, organization, and origins of diversity in herbivorous insects, a hyperdiverse group of organisms.

Table of Contents

Acknowledgements

Support has been critical to bringing me to this point in my life and to the completion of my dissertation project. This began with the freedom allowed by my parents to explore what excited me, which as a young boy was nature and studying science. My brother Adam and I started in our suburban Michigan backyard, picking cicada exuviae off of trees, observing leaf impressions in cement, and digging beneath the sand in our sandbox to discover the bones of a squirrel. With my family, I continued to strengthen my connection with nature on trips to natural history museums, in the rivers and forests of Michigan, and on beaches during family vacation.

My science instructors were influential in my decision to study the life sciences, even encouraging me in elementary school to explore summer courses at a local community college. Middle and high school teachers continued to foster my scientific interests in class and in extracurricular activities like the Science Olympiad competition. I was fortunate to attend a resident math and science camp at the University of Wisconsin supported by the government TRIO program which allowed me to get a taste of what university research entails. These experiences in youth primed me for Cornell University where I pursued as much science as possible. While at Cornell I deepened my love of nature in the entomology program and again found myself in a research lab. I was working on social spiders with Dr. Linda Rayor, but in addition to the research I learned what it is like to collaborate with the goal of discovering something new. She demonstrated how exciting research can be and was influential in my decision to pursue such a career over one in medicine. The Cornell experience is one I am thankful to have experienced and I am always excited to reminisce with Cornell friends from clubs, football or academics. Fortunately, I encounter many at national science conferences!

A systematic project of this magnitude would not have been possible without help through funding, provision of specimens, land access, and collaborators. My lab research was funded by the OEB department at Harvard, a Goelet research award through the Museum of Comparative Zoology, and in large part by a Milton award to the Farrell Lab from the Harvard Medical School. Field expeditions were funded by Putnam and Barbour grants from the MCZ, the Huron Mountain Wildlife Foundation, and Gorongosa National Park via the Carr Foundation. The David Rockefeller Center for Latin American Studies also provided travel funds as part of the enriching experience of being a graduate student associate of the center. Many sites provided access to enable the collections for this project: The Nature Conservancy's Southmost Preserve and Davis Mountains sites, the Huron Mountain Club, Death Valley National Park, Craters of the Moon National Monument, Great Basin National Park, Gorongosa National Park, Mojave National Preserve, Coronado National Forest, Humboldt-Toiyabe National Forest, Devil's Punchbowl Natural Area, and Merriam Powell Research Station. Important specimens for my systematics chapters were provided by multiple individuals and institutions. For these specimens I would like to thank Mike Ivie (Montana State University), Amy Berkov & Lin Li (City College of New York), Hannah Nadel (USDA OTIS lab), Sindhu Krishnankutty (USDA, Xavier University), Gino Nearns (Purdue University, USDA), Ian Swift (Los Angeles, CA), Petr Svacha (Czech Academy of Sciences), Lech Karpinski (Polish Academy of Sciences), Adam Slipinski (CSIRO), Masaru Nonaka (Japan), Jay Shetterly (Cambridge, MA), John Leavengood (USDA Entomologist), Patrick Sullivan (Hereford, AZ), Sang Il Kim (Harvard University), and Bruno de Medeiros (Harvard University). An MoU with the USDA Otis lab was very helpful for European species. I would like to thank Larry Hanks (University of Illinois) for providing chemical lures

that helped in field studies. Rob Mitchell (University of Wisconsin) has been a great help in thinking about and analyzing genes that may be experiencing selection in beetles.

The Farrell Lab was a wonderful place to call home during my studies. My advisor, Brian Farrell, allowed me to independently develop my project and remained excited about its intricacies as it progressed. Every time I met with Brian I came away feeling invigorated and ready to tackle another of the challenges these beetles presented. He also dedicated time to bring in money for lab research and to update equipment for genomic labwork. My fellow lab members made the lab a collaborative and fruitful environment. Bruno de Medeiros was a companion while collecting in the field and was a great help in the origin of my project. His assistance with statistics in R was crucial for my work. Sang Kim was a great help in the field in Southern Asia, was always up for local collecting, and is always ready with a camera to commemorate a moment. He was also a constant friend during late-night lab work. Alyssa Hernandez was a friend in and out of the lab. Gary Alpert was always ready to help with photography advice, databasing tips or new methods to collect data in the field. Piotr Naskrecki presented the opportunity to be a part of a survey effort and teaching of local students at Gorongosa National Park in Mozambique. Ricardo Perez-de la Fuente was always open to help stack photos and listen to Frank Zappa music four floors underground. Amie Jones was a great help with budgeting funds and teaching me to use graphics software. Whit Farnum provided me with essential training at the bench and always has an interesting historical story.

Students in Naomi Pierce's entomology lab at Harvard were a constant source of inspiration. Hearing their ideas at weekly lab meetings was a constant event in my graduate years and I appreciate the invitation to attend these and annual retreats with the group. Conversations with Jack Boyle, Leonora Bittleston, Rachel Hawkins, James Crall, Matthew Lim, Chris Baker,

Kadeem Gilbert, Shayla Salzmann, and Sarah Kocher helped me to refine my ideas and explore new avenues of research. Richard Childers was especially helpful with differential gene expression analyses for my data.

The other graduate students in the OEB program and people in the community made my free time enjoyable, and often productive. The OEB department provided unmatched resources and care for me as a student and person. The students that I entered Harvard with, Russ Corbett-Detig, Kolea Zimmerman, Sebastian Akle, Leonora Bittleston, James Crall, Talia Moore, Glenna Clifton, Danny Park, and Shelbi Russell, and their partners improved my health through conversation and stimulating activity. I also enjoyed the company of other students at departmental events, in the intramural leagues, and in the graduate student orchestra. Jim Wheeler was my guide upon entering graduate school and was a great friend in the wood shop on campus. I was a member of the Cambridge Entomological Club and it provided me with a number of friends that enjoy insects. I hunted insects locally with Jay Shetterley, Scott Smyers, Andrea Golden, and Jess Walden-Gray while other members provided good conversation before meetings.

Professionals in the Harvard community were a great resource. Faculty outside of my committee were always happy to speak on research or life in general and I gained a lot from their courses and advice. Entomology staff members Phil Perkins, Stefan Cover, Rachel Hawkins, and Crystal Maier assisted me in the collections and were great to speak with. The Ernst Mayr library team, especially Mary Sears and Ronnie Broadfoot were my go to for literature and helped me to track down every obscure reference. The FAS informatics team at Harvard helped me tremendously as I developed my skills in bioinformatics. Bob Freeman was great at introducing me to bioinformatics. Adam Freedman was patient with me as I worked with my transcriptome

data. Tim Sackton, Allison Shultz and Brian Arnold helped as I worked through population genomics problems.

My dissertation committee was integral to my growth as a graduate student. Brian Farrell was a great help as I developed ideas and grew as a scientific writer. He always had an encouraging word or interesting anecdote. Scott Edwards challenged me with new ways of thinking about phylogeny and population genetic processes. Our meetings were always useful and grounded in current research. Naomi Pierce was very encouraging and was a helpful mentor. I had a great time teaching in the lab of her insect biology course, speaking in lab meetings, and attending gatherings at her home. She is very generous with her time and it shows that students are one of her first priorities. Chuck Davis was also a big help with discussion on sequencing and biogeography as a member of my qualifying exam committee.

My family has been my biggest support while a graduate student. I married my wife, Kelly while in the program. She has been extremely tolerant with the changes graduate school brought: living at a distance for a time, leaving for months to collect with little contact, many late nights on campus, and the sometimes challenging living situation around Boston. On October 29th, 2016, we brought our son Sawyer into the world. He has provided extra inspiration to finish a project we can all be proud of. During my stressful final writing phase Kelly was patient and understanding of my sleep deprivation and picked up my slack at home. I am so thankful to have them, and for her enduring support I dedicate this dissertation to my wife Kelly.

Disclaimer

All taxonomic actions in this work are hereby disclaimed for nomenclatural purposes, as recommended in Article 8 of the International Code of Zoological Nomenclature.

CHAPTER 1

Convergence and major host plant shifts revealed in multilocus phylogeny of economically-important *Monochamus* longhorned beetles

Note: supplemental figures can be found in appendix A

Introduction

With plant-feeding insects exceeding 400,000 described species and estimates of double that number on Earth (Mitter et al. 1991) this is the most species-rich class of interactions known. Among beetles alone, there are over 135,000 described herbivores (Farrell 1998). Understanding the plant-insect interface is therefore important if we are to understand the role of ecological interactions in generating diversity. Determining how host relationships emerge across the evolutionary tree of a clade can provide insight into the macroevolutionary processes that have shaped the diversity. Here we present the first geographically and taxonomically broad phylogeny of sawyer beetles and discuss what factors may have impacted their diversification.

The genus *Monochamus* Dejean 1821, as currently defined, is a worldwide group of wood-feeding beetles, often called sawyers, inhabiting temperate and tropical forests. A total of 124 (www.biolib.cz) or 142 (http://titan.gbif.fr) species are placed in the genus, divided into several subgenera. As a long-standing genus, taxonomic history is somewhat disorganized with short descriptions leading to species inclusions that seem out of place. The subgenus *Monochamus* includes 18 conifer feeding species (the only conifer feeders in *Monochamus*) as well as many angiosperm feeders. Some of the most economically important species feed on conifers in the northern hemisphere. The larvae of sawyers feed extensively in the heartwood of recently dead trees, fulfilling an ecological role as decomposers, but also effectively decreasing the wood grade (quality) to pulp. Adults of the genus feed on the bark of young shoots in the conifer canopy, an apparent requirement for maturation of their cuticular hydrocarbons and gonads (Cherepanov 1990, Brodie et al. 2012). This life history trait facilitates the transmission

of the nematode *Bursaphelenchus xylophilus* Nickle, the causal agent of Pine Wilt Disease in Eurasian pines (Vicente et al. 2012), to living conifers. This adds to annual losses of millions of dollars in timber value attributed to *Monochamus* beetles (Allison et al. 2001). Elucidating the phylogenetic relationships of the conifer feeding *Monochamus* species is critical for accurate identification of invaders and exploration of the evolution of plant-insect-parasite relationships.

The systematics of several Eurasian *Monochamus* species has been recently explored using mitochondrial markers. For example, Cesari et al. (2005) used a portion of 12S and complete COI sequences to build a phylogenetic hypothesis for most European species. Toki and Kubota (2010) produced a 16S and COI phylogeny for species of the tribe Lamiini (=Monochamini) in Japan, including the conifer feeding and angiosperm feeding members of *Monochamus sensu lato*. The North American species have not been investigated with molecular techniques. In summary, researchers have used three mitochondrial genes, with little overlap in taxa or DNA sequences among the datasets. Because mitochondrial DNA evolves in a linked manner across all genes of the mitochondrion (Avise 1991, Moore 1995), there are phylogenetic data available for only a single locus for *Monochamus* beetles. We now know that single marker analyses can be misleading due to incomplete lineage sorting (Degnan and Rosenberg 2009) and introgression (Linnen and Farrell 2007). Analyses can be prone to terminate at local optima with the most general history of a genome only revealed through a diversity of gene trees (Edwards 2009). For these reasons, we include multiple nuclear and mitochondrial gene sequences to test existing phylogenetic hypotheses of the Eurasian species. Moreover, our study of all conifer feeding *Monochamus* species worldwide, using both mitochondrial and nuclear markers, will be the first analysis of North American taxa and the first attempt at a unified treatment of this important genus of timber-attacking beetles.

The sometimes cryptic nature of speciation, when species retain similar morphology or converge on a similar form due to similar selection pressures, can impede the taxonomic task of delimiting species. With recent or rapid diversification, the lack of evolutionary time for new morphological changes to become fixed can exacerbate this problem. *Monochamus* delimitation (Dillon and Dillon 1941, Hellrigl 1971, Linsley and Chemsak 1984, Wallin et al. 2013) has focused on diagnostic morphological traits such as the shape of elytra and patterns of elytral pubescence and microsculpture in geographically-delimited subsets of the genus. However, *Monochamus* classification has long been conflicting, and the characters that best delimit species-level taxa remain uncertain. By using many genetic characters, which give the opportunity for observing fixed differences not present in morphology, this study will test delimitations for current and proposed species and subspecies of conifer-feeding *Monochamus*.

We explore the phylogenetic relationships of a subset of *Monochamus* species including all known conifer feeding species and inhabited biogeographic regions. A 6882bp, seven marker, molecular dataset comprising mitochondrial DNA, ribosomal DNA, and nuclear DNA provides a robust estimate. Bayesian and maximum likelihood phylogenetic estimates (both concatenated and multispecies coalescent) are interrogated using species delimitation and topology tests. The impact of phasing nuclear markers, and their phasing groupings, in the context of a species level, multigene, phylogeny is explored. Results are used to discuss current classification and how geography and host plants may have shaped the evolution of the sawyers.

Materials and Methods

Species Sampling

Our goal is to produce the most inclusive phylogeny of *Monochamus* Dejean species and close relatives to date. *Monochamus* currently contains 22-28 subgenera according to the Titan (http://titan.gbif.fr) and BioLib (www.biolib.cz) databases, respectively. The subgenera *Monochamus* Dejean and *Tibetobia* Frivaldsky are present in the Palearctic (Lobl and Smetana 2010, Danilevsky 2018), and all other subgenera are endemic to Africa. Only the subgenus *Monochamus* is present in the Nearctic region. Most of these subgenera were described and categorized as genera by Lawrence and Elizabeth Dillon (Dillon and Dillon 1959a, 1959b, 1959c, 1959d, 1961). Subsequently, various published and unpublished lists of taxa have adjusted ranks, sometimes converted genera to subgenera, and have made unpublished taxonomy commonplace. This makes determining the current status of these taxa difficult as even the most up to date databases conflict. In addition, *Monochamus* is a well-known name, and its broad description allows it to continue as a destination for species that cannot be assigned elsewhere. This paper cannot address many of these issues, but it is important to recognize that the names presented here may change.

Fresh tissues were obtained for all of the conifer feeding species of *Monochamus* (hereafter *Monochamus* ***sensu novo***) and preserved in 95+% ethanol, at cryogenic temperatures or dried (Table 1.1). These specimens came from multiple geographic regions, when available. Several angiosperm feeding species from Africa and Asia were collected from *Monochamus sensu lato*. The taxon set represents 20 species in the subgenus *Monochamus* as well as five other *Monochamus* subgenera endemic to Africa. The African genera *Oxylamia* and *Pseudhammus*

were included due to morphological similarity and availability. In a broader survey of the tribe Lamiini (Gorring, unpub.), some genera outside *Monochamus* were found to be closely allied; therefore we included the genera *Goes* LeConte*, Microgoes* Casey*, Hebestola* Chevrolat and *Neoptychodes* Dillon & Dillon from North America and *Pharsalia (Cycos) subgemmata* (Thomson) from Thailand in the analysis. *Tetraopes linsleyi* Chemsak and *Tetraopes tetrophthalmus* (Forster) are outgroups from tribe Tetraopini (see Table 1.1). As there is no comprehensive resource to identify *Monochamus* and allies, we used an array of literature including Linsley and Chemsak (Linsley and Chemsak 1984) for North America, Cherepanov (Cherepanov 1990) or Wallin et al. (Wallin et al. 2013) for Eurasia and Duffy or Craighead's publications for larval identification (Craighead 1923, Duffy 1968). The nomenclatural status of species followed the most modern publications or updated lists available (Bezark 2017; Titan database). *M. mutator* LeConte was recently placed as a synonym of *M. maculosus* Haldeman (Bousquet et al. 2017) so most analyses were run with the former name. The sampling was not exhaustive, but included *Monochamus sensu lato* from all inhabited continents and assumed groupings, including all of the species from the conifer-feeding focal group.

DNA Code	Short code	Species	Host clade	Origin	COI-1718-3014	COI-barcode	WNG	EF1	AK	CAD	28S	TOPO
PSG000100	>PSG100_carolinensis_FL	*Monochamus carolinensis* (Olivier)	Conifer	FL, USA	x	x	x	x	x	x	x	x
PSG000557	>PSG557_carolinensis_Canada	*Monochamus carolinensis* (Olivier)	Conifer	New Brunswick, Canada	x		x		x	x		
PSG000782	>PSG782_carolinensis_MI	*Monochamus carolinensis* (Olivier)	Conifer	MI, USA	x	x	x		x	x		x
PSG000107	>PSG107_carolinensis_LA	*Monochamus carolinensis* (Olivier)	Conifer	LA, USA	x	x	x	x		x	x	x
PSG000365	>PSG365_cnevadensis_Craters	*Monochamus clamator latus*	Conifer	ID, USA	x	partial	x	x	x		x	x
PSG000449	>PSG449_clamator_Pinal	*Monochamus c. clamator*	Conifer	AZ, USA	x		x	x	x	x		x
PSG000183	>PSG183_clamator_Davis	*Monochamus c. clamator*	Conifer	TX, USA	x	x	x	x	x	x	x	x
PSG000472	>PSG472_cclamator_Colcord	*Monochamus c. clamator*	Conifer	AZ, USA	x		x		x	x	x	x
PSG000612	>PSG612_c_latus_MT	*Monochamus clamator latus*	Conifer	MT, USA	x		x		x	x		x
PSG000386	>PSG386_clatus_OR	*Monochamus clamator latus*	Conifer	OR, USA	x	x	x	x	x		x	x
PSG000402	>PSG402_clamator_Mendo	*Monochamus clamator latus*	Conifer	CA, USA	x	x	x	x	x	x	x	x
PSG000722	>PSG722_clatus_WA	*Monochamus clamator latus*	Conifer	WA, USA	x	x	x		x	x	x	x
PSG000224	>PSG224_clinsleyi_Panamint	*Monochamus c. clamator*	Conifer	CA, USA	x	x	x	x	x	x	x	x
PSG000253	>PSG253_cnevadensis_Toiyabe	*Monochamus c. clamator*	Conifer	NV, USA	x		x	x	x	x	x	x
PSG000261	>PSG261_cnevadensis_Desatoya	*Monochamus c. clamator*	Conifer	NV, USA	x		x	x	x	x	x	x
PSG000277	>PSG277_cnevadensis_Monitor	*Monochamus c. clamator*	Conifer	NV, USA	x		x	x	x	x	x	x
PSG000288	>PSG288_cnevadensis_Wpine	*Monochamus c. clamator*	Conifer	NV, USA	x		x	x	x	x	x	
PSG000323	>PSG323_cnevadensis_Schell	*Monochamus c. clamator*	Conifer	NV, USA	x		x	x	x	x		x
PSG000327	>PSG327_cnevadensis_Snakebristle	*Monochamus c. clamator*	Conifer	NV, USA	x	x	x	x	x	x	x	x
PSG000211	>PSG211_cnevadensis_Spring	*Monochamus c. clamator*	Conifer	NV, USA	x		x	x	x	x	x	x
PSG000089	>PSG89_crubigineus_AZ	*Monochamus c. clamator*	Conifer	AZ, USA	x	x	x	x	x	x	x	x
PSG000487	>PSG487_crubigineus_Mex	*Monochamus c. clamator*	Conifer	Oaxaca, Mexico	x	x	x		x	x	x	x
PSG000696	>PSG696_clam_MX	*Monochamus c. clamator*	Conifer	Mexico	x	x	x			x	x	x
PSG000701	>PSG701_clam_MX	*Monochamus c. clamator*	Conifer	Mexico	x	x	x		x	x		
PSG000541	>PSG541_marmorator_MI	*Monochamus marmorator* Kirby	Conifer	MI, USA	x	x	x	x	x	x	x	x
PSG000552	>PSG552_marmorator_Canada	*Monochamus marmorator* Kirby	Conifer	New Brunswick, Canada	x	x	x	x		x		x
PSG000073	>PSG73_mutator_MI	*Monochamus maculosus* LeConte	Conifer	MI, USA	x	x	x	x	x	x	x	x
PSG000563	>PSG563_mutator_Canada	*Monochamus maculosus* LeConte	Conifer	New Brunswick, Canada	x	x	x	x		x		
PSG000085	>PSG85_notatus_MA	*Monochamus notatus* (Drury)	Conifer	MA, USA	x	x	x	x	x	x	x	x
PSG000575	>PSG575_notatus_Saskatch	*Monochamus notatus* (Drury)	Conifer	Saskatchawan, Canada	x	x	x			x		x
PSG000611	>PSG611_notatus_NewBruns	*Monochamus notatus* (Drury)	Conifer	New Brunswick, Canada	x		x					x
PSG000110	>PSG110_notatus_MI	*Monochamus notatus* (Drury)	Conifer	MI, USA	x	x	x	x	x	x	x	x
PSG000396	>PSG396_oobtusus_OR	*Monochamus obtusus* Casey	Conifer	OR, USA	x	x	x	x	x	x	x	x
PSG000405	>PSG405_ofulvomac	*Monochamus obtusus* Casey	Conifer	CA, USA	x	x	x		x		x	x
PSG000723	>PSG723_oobtusus_WA	*Monochamus obtusus* Casey	Conifer	WA, USA	x		x			x	x	x
PSG000724	>PSG724_oobtusus_MT	*Monochamus obtusus* Casey	Conifer	MT, USA	x		x		x	x		x
PSG000079	>PSG79_sscutellatus_MI	*Monochamus scutellatus* (Say)	Conifer	MI, USA	x	x	x	x	x	x	x	x
PSG000313	>PSG313_scutellatus_Independenc	*Monochamus scutellatus* (Say)	Conifer	NV, USA	x	x	x	x	x	x	x	x
PSG000376	>PSG376_scutellatus_Willamette	*Monochamus scutellatus* (Say)	Conifer	OR, USA	x	x	x	x	x	x	x	x
PSG000426	>PSG426_Scutellatus_Graham	*Monochamus scutellatus* (Say)	Conifer	AZ, USA	x	x	x	x	x	x	x	x
PSG000569	>PSG569_scutellatus_NovaScotia	*Monochamus scutellatus* (Say)	Conifer	Nova Scotia, Canada	x	x	x		x	x		x
PSG000783	>PSG783_scutellatus_MT	*Monochamus scutellatus* (Say)	Conifer	MT, USA	x		x		x	x		x
PSG000104	>PSG104_titillator_FL	*Monochamus titillator* (Fabricius)	Conifer	FL, USA	x	x	x	x	x	x	x	x
PSGL0003	>PSGL3_titillator_KY	*Monochamus titillator* (Fabricius)	Conifer	KY, USA	x	x	x	x	x	x	x	x
PSG000106	>PSG106_titillator_LA	*Monochamus titillator* (Fabricius)	Conifer	LA, USA	x	x	x			x	x	x
PSG000784	>PSG784_titillator_MI	*Monochamus titillator* (Fabricius)	Conifer	MI, USA	x		x		x	x		x
PSG000039	>PSG39_Hebestola	*Hebestola pullata* (Haldeman)	Angiosperm	MA, USA	x	x	x	x	x	x	x	x
PSG000084	>PSG84_Microgoes	*Microgoes oculatus* (LeConte)	Angiosperm	MI, USA	x	x	x		x	x	x	x
PSG000131	>PSG131_Neoptychodes	*Neoptychodes trilineatus* (Linnaeus)	Angiosperm	AZ, USA	x	x	x	x	x	x	x	x
PSG000620	>PSG620_Goes_pulv	*Goes pulverulenta* (Haldeman)	Angiosperm	KY, USA	x	x	x	x	x	x		x
PSG000714	>PSG714_Goes_tig	*Goes tigrina* (DeGeer)	Angiosperm	AR, USA	x	x	x		x	x		x
PSG000753	>PSG753_Goes_pulcher	*Goes pulchra* (Haldeman)	Angiosperm	MI, USA			x					
PSG000756	>PSG756_Goes_debilis	*Goes debilis* LeConte	Angiosperm	MI, USA		x	x					
B065	>B065_adamitus_afr	*M. (Quasiochamus) adamitus* Thomson	Angiosperm	Sofala Prov., Mozambique	x	x	x	x	x	x	x	x
PSG000543	>PSG543_Anthores_Africa	*M. (Anthores) leuconotus* Pascoe	Angiosperm	Zimbabwe	x	x	x	x	x	x	x	x
PSG000606	>PSG606_homoeus_afr	*M. (Meliochamus) homoeus* Jordan	Angiosperm	Centre Reg., Cameroon	x	x	x	x	x	x		x
PSG000607	>PSG607_xfulvum_afr	*M. (Laertochamus) x-fulvum* Bates	Angiosperm	Centre Reg., Cameroon	x	x	x	x	x	x		x
PSG000608	>PSG608_Pseudhammus_afr	*Pseudhammus (Pseudhammus) myrmidonum* Kolbe	Angiosperm	Centre Reg., Cameroon	x	x	x	x	x	x		x
PSG000609	>PSG609_Oxylamia_afr	*Oxylamia (Oxylamia) fulvaster* (Jordan)	Angiosperm	Centre Reg., Cameroon	x	x	x	x	x			x
PSG000718	>PSG718_M_M_olivaceus	*Monochamus (M.) olivaceus* Breuning	Angiosperm	D.R. Congo			x			x		
PSG000523	>PSG523_spectabilis_Afr	*M. (Opepharus) spectabilis* (Perroud)	Angiosperm	Sofala Prov., Mozambique	x	x	x	x	x	x	x	x
PSG000591	>PSG591_alternatus_SK	*Monochamus alternatus endai* Makihara	Conifer	South Korea	x	x	x	x	x	x	x	x
PSG000630	>PSG630_alternatus	*Monochamus alternatus endai* Makihara	Conifer	Japan	from PSG821	x	x	x	x	x		x
PSG000821		*Monochamus alternatus endai* Makihara	Conifer	Japan	x							
PSG000682	>PSG682_alternatus_CH	*Monochamus a. alternatus* Hope	Conifer	China	x	x	x	x	x	x	x	x
PSG000649	>PSG649_gallo_FR	*Monochamus galloprovincialis* (Olivier)	Conifer	France	x	x	x		x	x	x	x
PSG000666	>PSG666_gallo_RU	*Monochamus galloprovincialis* (Olivier)	Conifer	Russia	x	x	x		x	x	x	x
PSG000671	>PSG671_gallo_IT	*Monochamus galloprovincialis* (Olivier)	Conifer	Italy	x	x	x		x	x		x
PSG000694	>PSG694_gallo_TK	*Monochamus galloprovincialis* (Olivier)	Conifer	Turkey	x	x	x	x	x	x	x	x
PSG000700	>PSG700_gallo_UK	*Monochamus galloprovincialis* (Olivier)	Conifer	Ukraine	x	x	x	x	x	x	x	x
PSG000622	>PSG622_grandis	*Monochamus grandis* Waterhouse	Conifer	Japan	x	x	x	x	x	x	x	x
PSG000588	>PSG588_guttulatus_SK	*Monochamus guttulatus* Gressitt	Angiosperm	South Korea	x	x		x	x	x	x	x
PSG000624	>PSG624_nitens	*Monochamus nitens* (Bates)	Conifer	Japan	x	x	x	x	x	x	x	x
PSG000626	>PSG626_nitens	*Monochamus nitens* (Bates)	Conifer	Japan	x		x			x	x	x
PSG000496	>PSG496_saltuarius_Czech	*Monochamus saltuarius* (Gebler)	Conifer	Czech	x	x	x	x	x	x	x	x
PSG000586	>PSG586_saltuarius_SK	*Monochamus saltuarius* (Gebler)	Conifer	South Korea	x	x	x		x	x	x	x
PSG000631	>PSG631_sp	*Monochamus sp.*	Conifer	Japan	x	x	x	x	x	x	x	
PSG000583	>PSG583_urussovi_Poland	*M. sartor urussovii* (Fisher von Waldheim)	Conifer	Poland		x	x		x	x		x
PSG000651	>PSG651_sartor_RU	*M. sartor urussovii* (Fisher von Waldheim)	Conifer	Russia	x	x	x	x	x	x	x	x
PSG000684	>PSG684_sartor_Hung	*Monochamus s. sartor* (Fabricius)	Conifer	Hungary	x	x	x	x	x	x	x	x
PSG000687	>PSG687_sartor_RU	*M. sartor urussovii* (Fisher von Waldheim)	Conifer	Russia	x	x	x	x	x	x	x	x
PSG000690	>PSG690_sartor_TK	*Monochamus s. sartor* (Fabricius)	Conifer	Turkey	x	x	x	x	x	x	x	
PSG000691	>PSG691_sartor_IT	*Monochamus s. sartor* (Fabricius)	Conifer	Italy	x	x	x	x	x	x	x	x
PSG000592	>PSG592_subfaciatus_SK	*Monochamus subfasciatus* (Bates)	Angiosperm	South Korea	x	x	x	x	x	x	x	x
PSG000494	>PSG494_sutor_Czech	*Monochamus sutor* (Linnaeus)	Conifer	Czech	x	x	x	x	x	x	x	x
PSG000656	>PSG656_sutor_IT	*Monochamus sutor* (Linnaeus)	Conifer	Italy		x	x	x	x	x	x	x
PSG000790	>PSG790_impluviatus_MON	*Monochamus i. impluviatus* (Motschulsky)	Conifer	Mongolia	x	x	x					
PSG000599	>PSG599_Pharsalia_TH	*Pharsalia (Cycos) subgemmata* (Thomson)	Angiosperm	Phrae Prov., Thailand	x	x	x	x	x	x	x	x
PSG000504	>PSG504_Tlinsleyi_AZ	*Tetraopes linsleyi* Chemsak	Angiosperm	AZ, USA	x		x	x		x	x	x
DDM0397	>Tetraopes_tetrophthalmus	*Tetraopes tetrophthalmus (Forster)*	Angiosperm	MD, USA	AF267478	AF267478	KP813603	KP677866	KP812442	KP813042	KP419667	
Gene completeness					0.97	0.79	0.99	0.67	0.86	0.90	0.69	0.87

Table 1.1. Sampled taxa and feeding habits. Short codes correspond to tree tips, when shaded they were removed in the StarBEAST2 analysis. In gene matrix: an 'x' means a successful sequencing, GenBank numbers indicate sequences that were downloaded. The bottom row indicates the completeness of the individual gene matrix.

DNA extraction, amplification, and sequencing

We used the DNeasy column extraction kit (Qiagen) to extract DNA from ethanol preserved, dried and frozen samples. Whole leg or thorax muscle was taken from adult beetles , vacuum centrifuged to remove ethanol, and ground with a pestle before an overnight lysis incubation in a shaker at 56C. DNA was eluted into 200 or 300ul of Qiagen buffer AE. The columns of dried samples were eluted once with 200ul and then again with 150ul pulled from the first elution to optimize yield and DNA concentration. DNA extracts acquired through an agreement with the United States Animal and Plant Health Inspection Service (APHIS) were extracted using the methods in Wu et al. (Wu et al. 2017)

High agreement at all nodes of the phylogeny and consensus among gene trees were the main goals of the project. Seven partial genes of varied evolutionary rate were sequenced for this analysis: ~1468bp cytochrome oxidase subunit I mitochondrial DNA (COI), ~1327bp 28S ribosomal DNA (28S), ~438bp wingless (*wg*), ~1153bp Elongation Factor 1 alpha (EF1α), ~742bp arginine kinase (ArgK), ~742bp topoisomerase I (Topo) and ~943bp rudimentary (CAD), totaling ~6882bp of aligned sequence data. Gene matrix completion percentages can be seen in Table 1.1. Due to differences in mutation rates, these genes contribute to phylogenetic support at nodes from species to genus level and above (Wild and Maddison 2008). PCR amplification of these genes was carried out using the primers in Table 1.2 according to established beetle protocols and optimized if needed (Maddison 2012, Mckenna et al. 2015). PCR success was confirmed using gel electrophoresis. PCR products were cleaned using EXO-SAP cleaning (COI, wg, CAD) or gel extraction (28S, EF1α, ArgK, Topo) using the Qiagen QIAquick column extraction kit. The sequencing reaction was performed using Applied Biosystems BigDye terminator v. 3.1 and followed by ethanol precipitation. One specimen's

1718-3014 COI sequence (PSG630) was lengthened using the COI barcode from a different specimen from the same collecting event (PSG821). Some COI barcodes were sequenced at the USDA OTIS lab (Buzzard's Bay, MA) as a part of their woodborer intercept project and were combined with new sequencing of the 1718-3014 segment (see Table 1.1).

Gene	Primer	Direction	Sequence	Reference
28S	ZX1	F	ACCCGCTGAATTTAAGCATAT	Van der Auwera et al. 1994
	rd5b	R	CCACAGCGCCAGTTCTGCTTAC	Whiting 2002
COI	LCO1490	F	GGTCAACAAATCATAAAGATATTGG	Hebert et al. 2003
	C1-J-1718	F	GGAGGATTTGGAAATTGATTAGTTCC	Simon et al. 1994
	HCO2198	R	TAAACTTCAGGGTGACCAAAAATCA	Hebert et al. 2003
	C1-J-2183	F	CAACATTTATTTTGATTTTTTGG	Simon et al. 1994
	TL2-N-3014	R	TCCAATGCACTAATCTGCCATATTA	Simon et al. 1994
CAD	CD338F	F	ATGAARTAYGGYAATCGTGGHCAYAA	Moulton & Wiegmann 2004
	CD668R	R	ACGACTTCATAYTCNACYTCYTTCCA	Wild & Maddison 2008
	CD688R	R	TGTATACCTAGAGGATCDACRTTYTCCATRTTRCA	Wild & Maddison 2008
WG	wg550F	F	ATGCGTCAGGARTGYAARTGYCAYGGYATGTC	Wild & Maddison 2008
	wg578F	F	TGCACNGTGAARACYTGCTGGATG	Ward & Downie 2005
	wgAbrZ	R	CACTTNACYTCRCARCACCARTG	Wild & Maddison 2008
	wgAbr	R	ACYTCGCAGCACCARTGGAA	Ward & Downie 2005
AK	ForB2	F	GAYTCCGGWATYGGWATCTAYGCTCC	Danforth, Lin, Fang 2005
	RevB1	R	TCNGTRAGRCCCATWCGTCTC	Danforth, Lin, Fang 2005
	ForB4	F	GAYCCCATCATCGARGACTACC	Jordal 2007
TOPO	TP643F	F	GACGATTGGAARTCNAARGARATG	Wild & Maddison 2008
	TP932R	R	GGWCCDGCATCDATDGCCCA	Wild & Maddison 2008
	TP675F	F	GAGGACCAAGCNGAYACNGTDGGTTGTTG	Wild & Maddison 2008
EF1a	For1deg	F	GYATCGACAARCGTACSATYG	Danforth & Ji 1998
	Cho10mod1	R	ACRGCVACKGTYTGHCKCATGTC	Danforth & Ji 1998
	Cho10r1	R	AGCATCDCCAGAYTTGATRGC	Mckenna & Farrell 2009

Table 1.2. PCR and Sanger sequencing primers for this study

Sequencing was carried out in the Harvard University Bauer Core on the Applied Biosystems 3730xl Genetic Analyzer machine. Resulting chromatograms were loaded into *Sequencher* software v. 5.1 (Gene Codes Corp.) to manually edit, pair and export the resulting consensus gene sequences. For the *MrBayes* dataset, heterozygous sites in nuclear genes are

coded with IUPAC ambiguity codes; in coalescent datasets, heterozygotes are phased in several ways (see following).

Sequence phasing

The allele is the smallest unit being acted upon by evolution, and the multispecies coalescent model is built on single allele analysis (Degnan and Rosenberg 2009, Andermann et al. 2018). Therefore, when exploring the coalescence of alleles in phylogeny, it should be important to separate the alleles of multiallelic species. Two phased datasets were constructed to test the effects of different levels of phasing and to compare with an unphased dataset. The most conservative way to do this is to phase within interbreeding species, not across species that do not share alleles. It is also optimal to use a representative population sample of the species concerned. Therefore, for dataset *SpeciesPhased*, we phased any nuclear gene, after *Gblocks* if used, for species with three or more individuals. To do this, we separated each species into multi-individual .fasta files for each gene. This file was input in the *SeqPHASE* webtool (Flot 2010) for conversion to a PHASE input file and a constant sites file. The input file was run at the UNIX command line in PHASE v2.1.1. (Stephens et al. 2001). The .out output file from *PHASE* is used in conjunction with the constant sites file, created earlier, in step two of the web tool to create a final phased allele fasta file. This locus file is now ready for *BEAST2* input in Starbeast2 or STACEY for species tree estimation under the multispecies coalescent. For dataset *GenusPhased*, an approach of phasing all *Monochamus sensu novo* individuals in the MP taxa set was done for comparison. The data preparation followed the same methodology as *SpeciesPhased*, except all species were input and phased simultaneously. The third dataset, *UnPhased*, was unphased and heterozygous alleles were collapsed using IUPAC ambiguity

codes. Each analysis was run on only the five single-copy nuclear genes due to the focus on differences arising from phasing strategy.

To test the impact of these three phasing strategies each dataset was input to the *BEAST2* v.2.4.8 (Bouckaert et al. 2014) package STACEY v. 1.2.4 (Jones 2017). STACEY is ideal for this test because each allele can be assigned as its own taxon to allow freedom of tip movement. This avoids the violation of the multispecies coalescent model that often comes from samples being assigned to taxa not representing their true coalescent group in *BEAST (Andermann et al. 2018). STACEY was run as outlined in Andermann et al. 2018 except running 500 million generations, sampling trees and logging every 20,000, and setting collapse height to 1.0E-5. The resulting species trees were compared in topology, node height, and mean number of delimited species. Runs that involved single nuclear gene input (*BEAST2, BPP*) used the *SpeciesPhased* phasing strategy based on published phasing simulation results (Andermann et al. 2018) and this paper's phase testing results.

2.4 Data alignment and partitioning

There were several processing and quality checks performed on the gene data. Within *Sequencher*, chromatograms were assembled into contigs, and the primer regions were trimmed. Bases of low quality or that conflicted between forward and reverse reads were manually edited. After export from *Sequencher*, each gene was aligned using *MAFFT* v. 7 (Katoh and Standley 2013). Gene matrices were then constructed by importing the aligned sequence .fasta files into the *Mesquite* software package v. 3.2 (Maddison and Maddison 2017). Within the Mesquite editor, sequences were realigned by eye if needed, trimmed, and translated to amino acids to check for problematic stop codons. Any stop codons were checked against the raw

chromatogram to confirm the nucleotides. *Gblocks* (Castresana 2000, Talavera et al. 2007) with less stringent options was used for genes 28S, AK and EF to increase the signal to noise ratio of the data and allow for easier viewing of alignments. The outgroups in the alignments introduced most of the problematic portions. The resulting gene matrices were combined into one multi-matrix .nexus file in Mesquite that gives various export options for downstream analysis. For single marker input, each gene was analyzed with *Jmodeltest* v. 2.1.4 (Posada 2008) to determine the best fitting model of sequence evolution.

For *MrBayes* and *RAxML*, the concatenated dataset of 6882 sites was analyzed in *Partitionfinder* v2.1.1 (Lanfear et al. 2016) using unlinked branch lengths, the greedy search algorithm, and AICc as the selection metric. With a potential statistical problem stemming from gamma + I models (Yang 2014), the potential model set in *Partitionfinder* was reduced to those not combining these two parameters (Kim et al. 2018). Genes with an intron were separated into two coding pieces and the intron for input. The preferred scheme partitioned the data into four subsets: 1) COI_2ndpos, 28S, EF1a_2ndpos, AKb_1stpos, EF1a_1stpos, CAD_2ndpos, CAD_1stpos, TOPO_1stpos, wg_1stpos, AKb_2ndpos, TOPO_2ndpos, wg_2ndpos 2) TOPO_3rdpos, wg_3rdpos, EF1a_3rdpos, AKa_3rdpos, AKb_3rdpos 3) COI_3rdpos 4) CAD_3rdpos, COI_1stpos, AK_intron, EF_intron, AKa_2ndpos, AKa_1stpos. The best models were: 1) TVM+I 2) K81UF+G 3) GTR+G 4) GTR+I. In *MrBayes*, since not all models are supported, the first two models are best represented by the parameterization for GTR+I and GTR+G, respectively. A 4087bp nuclear gene only dataset was also partitioned. This resulted in two partitions: 1) AK_coding_2ndpos, AK_coding_1stpos, EF1a_2ndpos, TOPO_3rdpos, AK_intron, CAD_2ndpos, EF1a_1stpos, AK_coding_3rdpos, TOPO_2ndpos, EF1a_b_3rdpos, TOPO_1stpos, CAD_1stpos, EF1a_b_2ndpos, wg_2ndpos, wg_3rdpos, wg_1stpos 2)

CAD_3rdpos, EF_intron, EF1a_3rdpos, EF1a_b_1stpos. The best-fitting models were TIM+G and TRN+I, respectively. These are coded as GTR+G and GTR+I in *MrBayes*.

Phylogenetic analyses

This phylogeny will include recent and more ancient divergence, and tree building methods may respond differently to this problem. In order to compare phylogeny estimates, and to be thorough, trees were built using Bayesian methods in both a coalescent and forward time phylogenetic analyses as well as Maximum Likelihood. Data were analyzed using a concatenated supermatrix approach and a Bayesian multispecies coalescent (MSC) multigene species tree approach.

The concatenated MP dataset was analyzed using *MrBayes* v3.2.6 (Ronquist et al. 2012) using three independent instances of two runs with four MCMC chains each (one cold). The dataset uses the same preparation as the *UnPhased* data. The sequence was partitioned according to the best model determined by Partitionfinder. Each was run for 20 million generations, for a total of 120 million, with sampling every 10000 generations. Stationarity and convergence were evaluated by deviation of split frequencies < .01, potential scale reduction factor values ~ 1.00, and effective sample size (ESS) >200 as measured in *Tracer* v. 1.6 (http://beast.community/tracer). The tree files were combined using mcmcp in *MrBayes* with a burnin of 25% to produce a consensus tree. All trees were viewed and manipulated in *FigTree* v. 1.4 (https://github.com/rambaut/figtree). The same method was used to estimate a concatenated five marker nuclear protein-coding tree with a run of 100 million total generations sampling every 1000.

Maximum likelihood analyses ran in *RAxML* v. 8.2.11 (Stamatakis 2014) on the Harvard Odyssey computing cluster. The supermatrix dataset ran using the optimal partitioning scheme from *PartitionFinder* and the GTRGAMMA nucleotide model. The threaded version of *RAxML* was used to run rapid bootstraps and 1000 independent starting trees.

Independent matrices of each of the seven genes were loaded into the *BEAST2* v2.4.8 StarBEAST2 v. 0.14.0 (Ogilvie et al. 2017) template in *BEAUti* for species tree estimation under the MSC. See Table 1.1 for an outline of the reduced taxon set. Each gene was treated as an independent partition and site, and clock models were assigned to each. A total of four independent chains of 200 million samples were run, recording every 5000. After verifying convergence in *Tracer*, *LogCombiner* v. 2.4.8 was used to combine the runs, and a maximum clade credibility tree was produced using *Treeannotator* v. 2.4.8 with the posterior probability limit=.5 and median node heights.

To determine the contribution of individual markers to the phylogenetic estimate seven additional six gene StarBEAST2 runs were done, each with one of the genes removed. Preservation of topology across these trees indicates that support for nodes is coming from multiple markers. Changes in topology indicate that excessive signal may be coming from a single marker.

Topology testing

When support values are low on a phylogenetic tree, placement of clades can be misleading and should be confirmed statistically before drawing conclusions. To this end we used the Swofford–Olsen–Waddell–Hillis (sowh) test implemented in *SOWHAT* v. 0.36 (Church et al. 2015). This test compares the log likelihood difference of two topologies to a null

distribution of likelihood differences produced through simulation of data under the same evolutionary parameters. To prepare competing topologies we manually created constraint trees where one node is constrained to test clade placement. The taxa included for these tests are the conifer feeding *Monochamus* individuals, *Goes*, *Hebestola*, and *Pharsalia* as an outgroup. Each *SOWHAT* analysis was run 100 repetitions, with the GTRGAMMA model and optimal partitioning scheme in *RAxML*, on 12-20 Intel cores on Harvard University's Odyssey cluster.

For this study, topology tests were performed with a constraint on the grouping for *Monochamus clamator latus* in a nuclear gene *RAxML* tree and a constraint of the *Goes* + *Hebestola* clade as sister to North American *Monochamus* using all seven genes. The *M. c. latus* test was restricted to nuclear data to eliminate the impact of COI since the subspecies is monophyletic in the 7-gene *RAxML* tree (Fig. A1). The unconstrained nuclear *RAxML* phylogeny in *SOWHAT* shows *M. c. latus* specimens are mixing with other subspecies of *M. clamator*. The input constraint tree for *SOWHAT* constrains *M. c. latus* to be monophyletic. The second run constrained *Goes* and *Hebestola* in a clade with the North American *Monochamus* species. This topology showed up in several missing gene tests (Table 1.3) and was highly supported in an analysis of the tribe Lamiini (Gorring, unpub.). A constraint of *Monochamus* conifer feeders as monophyletic was also planned, but the *RAxML* analysis found this clade monophyletic with no constraint.

Species delimitation

When delimiting species, a necessary aside is the investigator's definition of what constitutes a species. The unified species concept agrees in targeting independent divergence of metapopulation lineages (De Queiroz 2007), but differences remain in what criterion of

divergence is critical for the separation of species. A species is a lineage that has a strong pre- or post-zygotic barrier to reproduction with another lineage. While the biological species concept (Mayr 1942) forms an ideal endpoint for sexual organisms, it may be too conservative during the period when semi-independent populations are undergoing lineage-sorting and fixing characters. Mallet (1995) proposed that the biological species concept does not allow for the gradual restriction in gene flow necessary to separate a well-mixed population. This is a valid argument and points out a limitation in its practical functionality if taken to the extreme of zero flow allowed in sympatry (Futuyma and Mayer 1980). We use influence from the BSC, and its criticism (Coyne and Orr 2004), pragmatically for delimitation by not *requiring* that any single characteristic be shared among all individuals of a set of populations in order to diagnose them as constituting a species but rather that an integrative case be made for genetic independence. Evidence can come from sequences, morphology, ecology or other genetically influenced traits. This means showing over a number of independent markers that gene flow is low between the proposed species. Higher levels of flow, or low divergence, may warrant the designation of subspecific status with corroborating evidence. The subspecies taxonomic unit should always predict the distribution of characters other than those used to delimit it (Cohn 1965), and not just describe some found variation. That is where the unit's usefulness will be derived. Special arguments can be made for traits of particular importance such as cuticular hydrocarbon incompatibilities or polyploidy that can even impose isolation on sympatric populations. It is important to remember that proposed species are informed hypotheses that are, like the speciation process, not static.

Species boundaries can be difficult to delimit in rapidly evolving, geographically widespread complexes of morphologically very similar taxa, as is demonstrated in long synonym

lists for many herbivore species. Through the taxonomic history of the genus *Monochamus*, many species have been proposed and many synonymies have been made. Several populations are also described as putative subspecies. With a sampling of all conifer feeding species and many populations, a goal is to delimit these species in particular as a basis for subsequent revisionary and ecological work. Coalescent models are ideal for tree building and delimitation because they incorporate the phylogenetic uncertainty of gene trees, address incomplete lineage sorting, and cope with gene tree-species tree discordance (Degnan and Rosenberg 2009). The multi-gene coalescent species delimitation approaches *Bayesian Phylogenetics and Phylogeography* (*BPP*)(Yang 2015) and *BEAST2* package STACEY were employed to evaluate the conifer feeding *Monochamus* species.

A codified approach to delimit taxa is to compare the probability of trees with closely related taxa treated as split versus lumped. *BPP* makes this method efficient by using an iterative rjMCMC process that collapses nodes on the species tree and evaluates the differences in posterior probability between a split or lumped group of taxa (Yang and Rannala 2010). The A01 method uses a user set species tree to run this process on and A11 both determines a best MSC species tree and delimits species according to the taxa given. Phased alleles are assigned by the user to an individual taxon but remain attached to the individual and are accounted for in the model. *BPP* was run for the Eurasian species and North American species to reduce run time. The taxa chosen were those from the data matrices for StarBEAST2, including phased nuclear genes. Separate runs were done for the combined five nuclear genes and COI+28S. This is necessary because phasing is a variable that is assigned to the entire dataset of a *BPP* run; it also allows a comparison of these datasets.

STACEY is a second delimitation method that is a part of *BEAST2*. It allows the run of a StarBEAST2 analysis on alleles and improves efficiency by integrating out population size parameters. The STACEY run setup is described in the phasing methods section. After a run was complete, custom python scripts (T. Andermann unpub.) were used to rescale the STACEY species trees to the average clock rate for each MCMC step. The rescaled distribution of node heights at individual nodes of various taxonomic levels were output for comparison. The SpeciesDelimitationAnalyser package in *BEAST2* (Jones et al. 2015) was run to summarize posterior frequencies of clusterings and *R* scripts provided by the same paper were adapted to graphically display the delimitation results of the rescaled tree file in a pairwise similarity matrix heatmap (simmatrix) indicating probability of belonging to the same cluster.

Results

Phylogeny of conifer-feeding Monochamus and relatives

The Bayesian phylogenies produced using gene concatenation (*MrBayes*, Fig. 1.1) and the multispecies coalescent species tree method (StarBeast2, Fig. 1.2) agree in higher level nodes but have some species relationship discrepancies. Both trees show the genus *Pharsalia* (subgenus *Cycos*) Pascoe as the sister genus to a combination of all conifer feeding *Monochamus* (*Monochamus sensu novo*) and the clade including the angiosperm feeding genera *Goes* LeConte and *Hebestola* Haldeman (1.0 BPP). The Goes clade, including North American genera *Goes+Hebstola*, was found sister to *Monochamus sensu novo* in the concatenated tree (PP=1.0) and StarBeast2 (PP >.9) analyses but is placed in a clade with North American *Monochamus spp.* in all STACEY analyses (PP=1). Outside *Pharsalia (Cycos)* sits a clade including Asian angiosperm feeding *Monochamus sensu lato* and the North American genus *Microgoes* Casey.

Monochamus and related genera from Africa are even further from the conifer feeding *Monochamus* clade. These groupings are consistent in both phylogenies and conflict with current classification.

Within *Monochamus sensu novo*, some species relationships differed between the concatenated and coalescent trees. There are two well-supported clades, one including all North American species and one including all Eurasian conifer feeding species. *M. carolinensis* and *M. titillator* are sister species in the Bayesian concatenated tree (Fig. 1.1) while *M. carolinensis* and *M. maculosus* are sisters in the coalescent (Fig. 1.2). In the North American species, *M. notatus*, *M. marmorator*, *M. scutellatus,* and *M. obtusus* are closely related but differ in topology between the two trees. Eurasian species *M. galloprovincialis*, *M. sutor*, *M. grandis*, *M. nitens*, and *M. sartor* show different sister relationships and higher level branching patterns between concatenated and coalescent analyses.

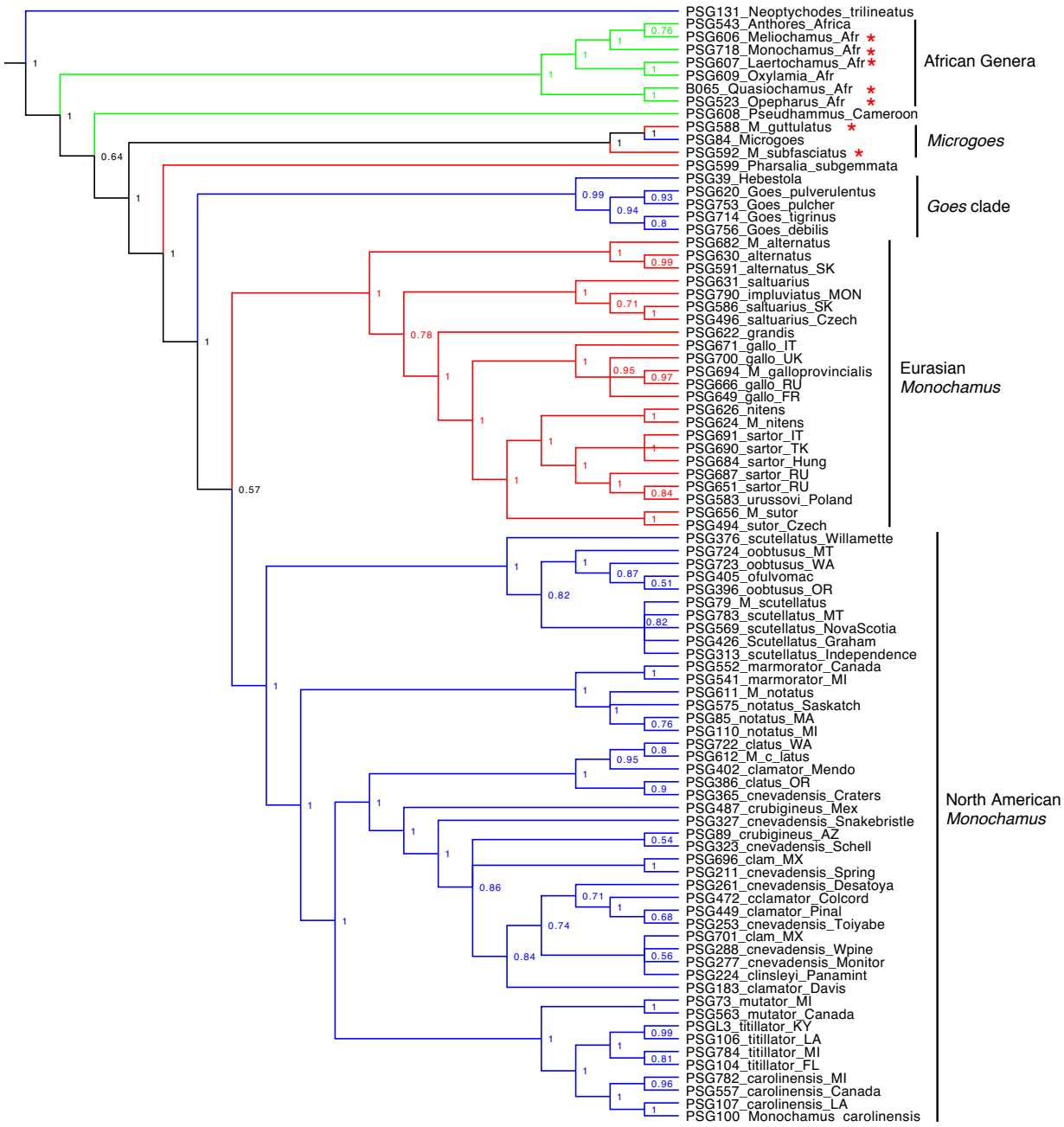

Figure 1.1 (not to scale) Bayesian supermatrix cladogram estimated using MrBayes. Asterix's indicate angiosperm feeding species currently in *Monochamus*. Blue branches- Nearctic, red branches-Eurasian, green-African. Branch labels are posterior probabilities

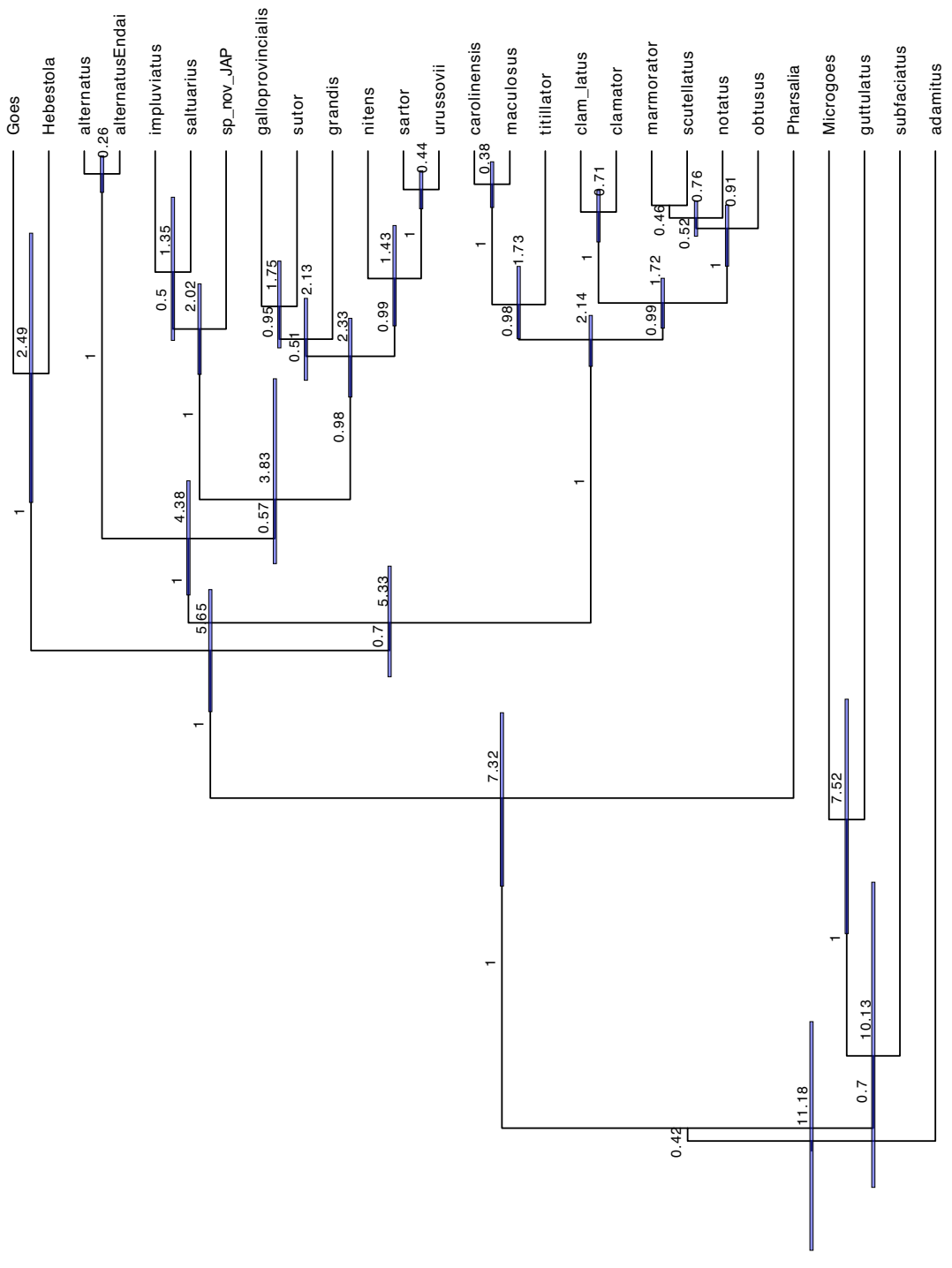

Figure 1.2. Dated StarBEAST2 SpeciesPhased species tree based on all seven markers. Branch labels are PP; node labels are node heights with bars indicating 95% confidence intervals on height. Scale in million years before present.

The StarBEAST2 runs of phylogenies with missing genes gave largely congruent topology (Figs. A2-A8). Support values for some groupings of most general interest are included in Table 1.3. The phylogenies are divided into those that support the *Goes* Clade as sister to all *Monochamus sensu nov.* or as sister to North American *Monochamus* species. There was a large effect of removing COI, with many relationship changes (Fig. A8). For other missing gene analyses, within *Monochamus sensu nov.* they tend to agree with the StarBEAST2 full data analysis (Fig. 1.2). Without *wg*, there is a shift to having *M. saltuarius* + *sp. nov.* sister species. Without 28S *M. alternatus* moves from being sister to all Eurasian species.

Clade	Concat. MrBayes	StarBEAST2	RAxML	COI	no 28S	no AK	no EF	no COI	no CAD	no wg	no TOPO
Mono NA+EUR	0.57	0.7	50	x	x	x	x	x	0.75	0.95	0.82
Goes + Mono	1	1	100	1	1	1	1	1	1	1	1
Goes + NAMono	x	x	x	0.94	0.47	0.45	0.53	0.67	x	x	x
Microgoes + Angio Mono	1	0.7	100	1	0.73	0.74	0.73	1	0.55	0.7	x

Table 1.3. Branch support values (in posterior probability except for bootstrap values in *RAxML* analysis) for higher level groupings over multiple analyses, 'x' indicates the branch was not present in a given analysis.

The coalescent species tree was dated using a strict clock enlisting the mitochondrial rate of evolution in *Tetraopes* cerambycid beetles (1.5%/my). The split between *Monochamus* conifer feeders in North America and Eurasia was 5.34 million years ago. The divergence of the *Monochamus sensu novo* clade from the *Goes*/*Hebestola* clade is estimated at 5.65 mya.

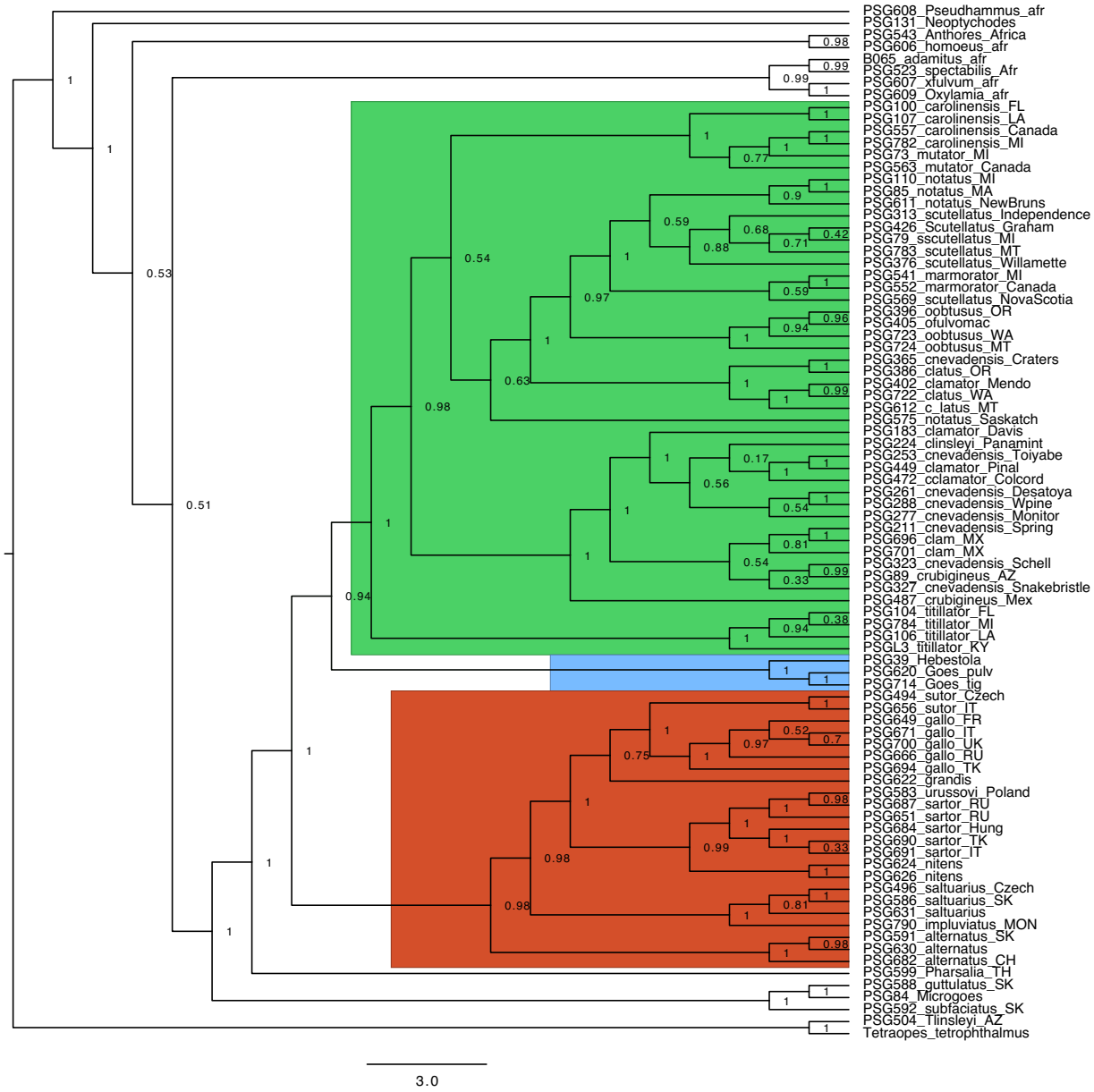

Figure 1.3. (not to scale) StarBEAST2 COI gene tree cladogram, green highlight=North American *Monochamus*, blue= *Goes* clade, red=Eurasian *Monochamus*

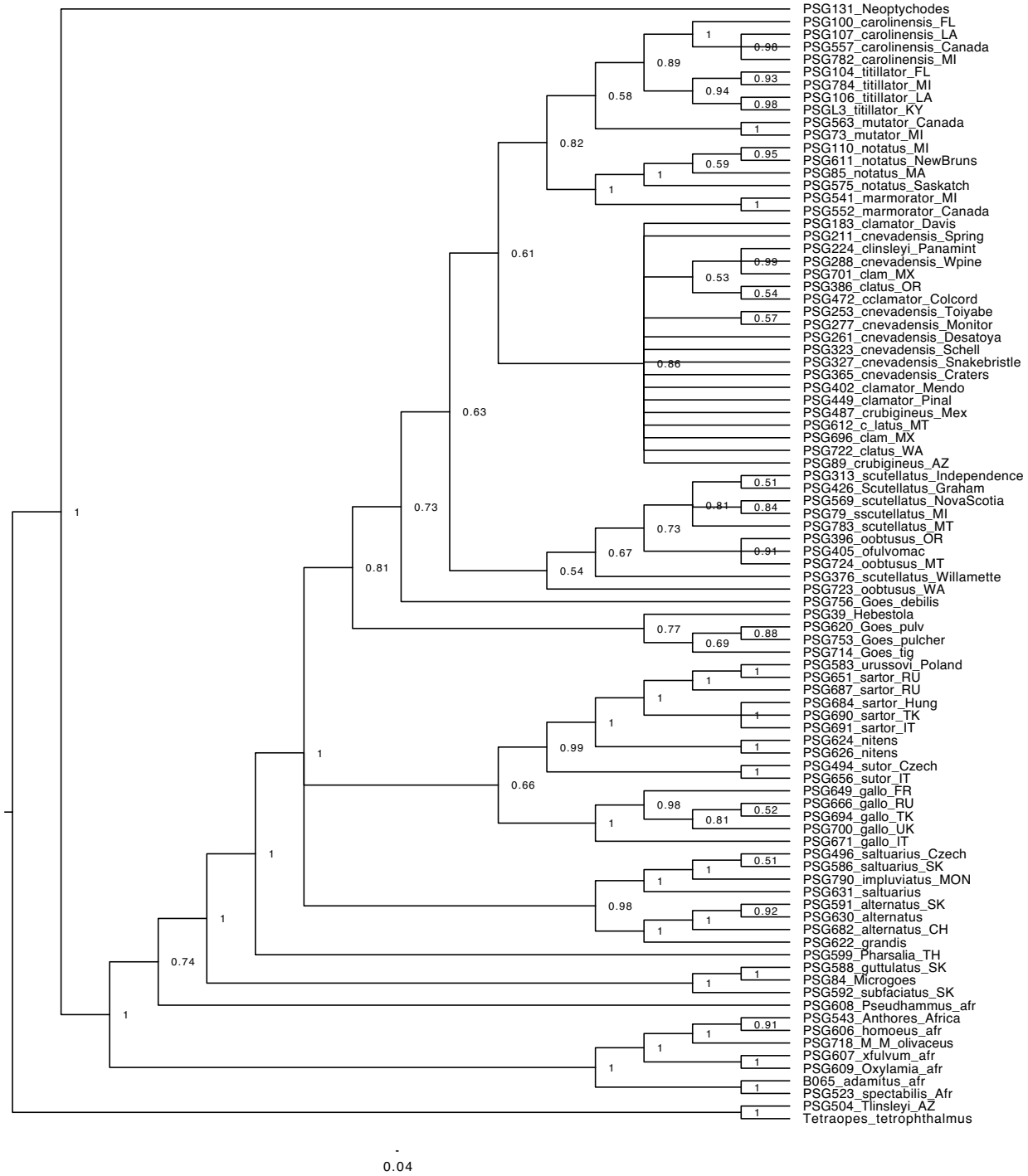

Figure 1.4. (not to scale) MrBayes nuclear only concatenated data cladogram, support=posterior probability

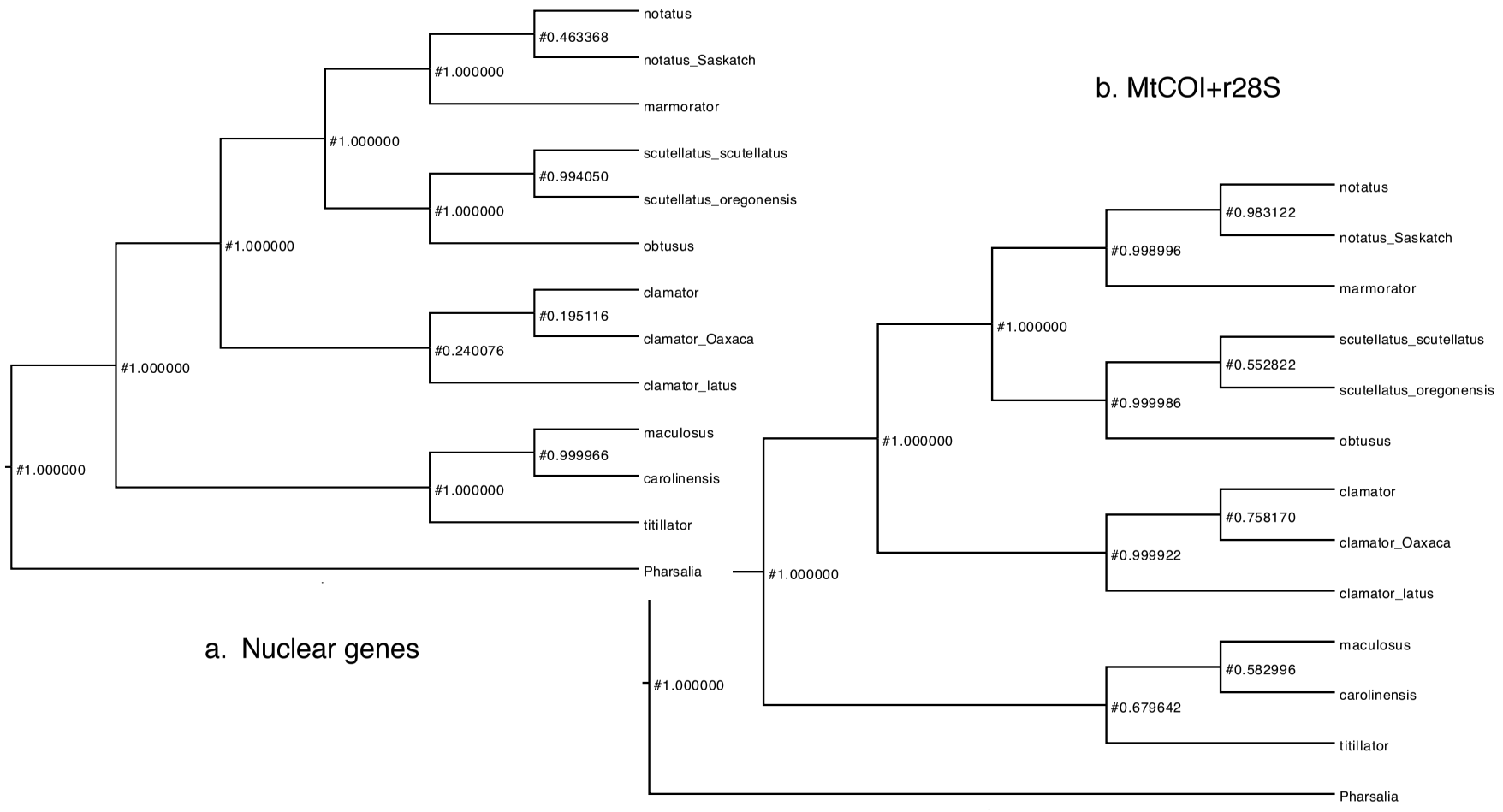

Figure 1.5. BPP, North American *Monochamus* A01 analysis, using guide tree, node values are support for a species split.
a) phased 5 nuclear genes, b) Mitochondrial + ribosomal genes

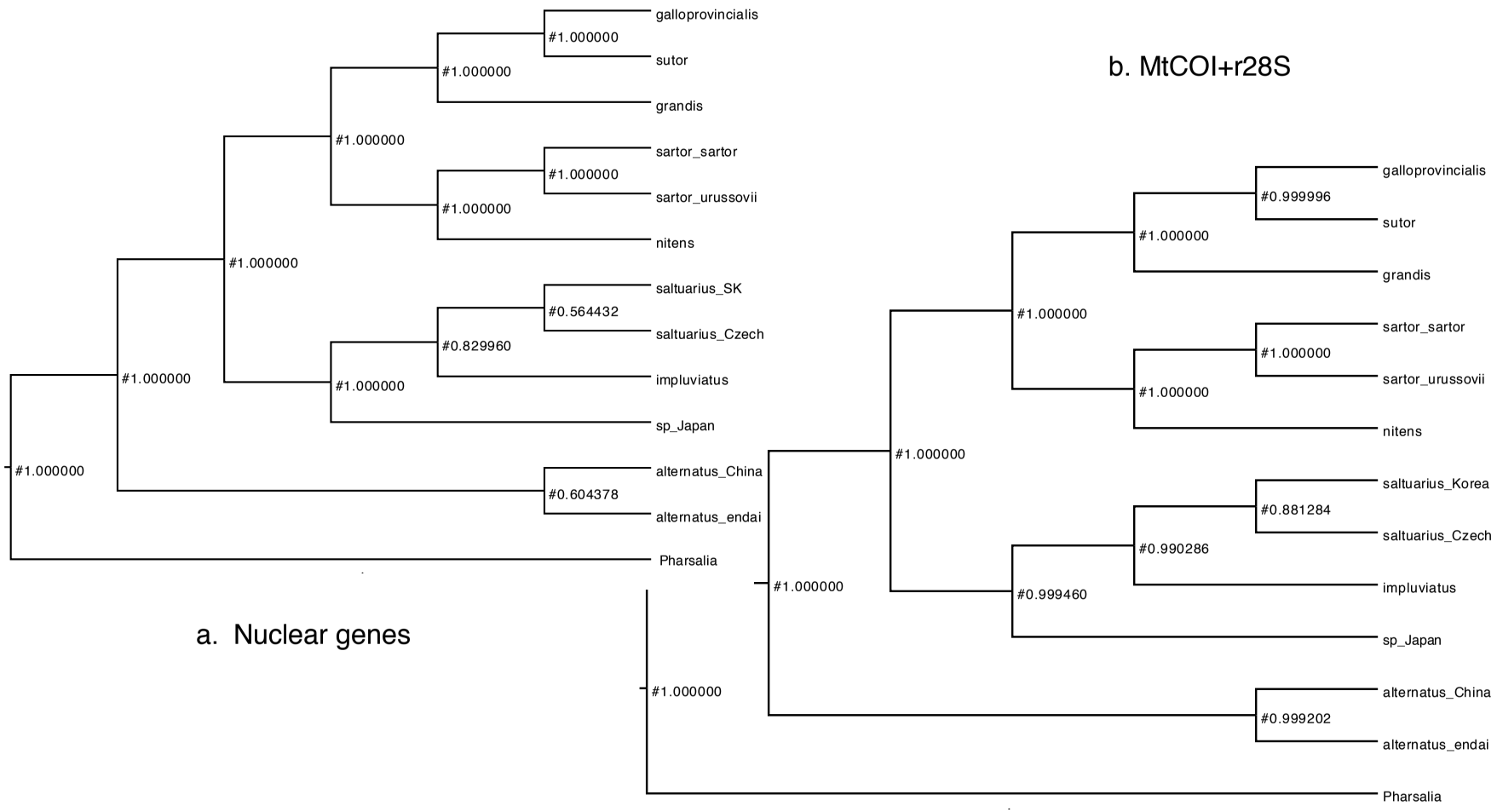

Figure 1.6. BPP, Eurasian *Monochamus* A01 analysis, using guide tree, node values are support for a species split.
a) phased 5 nuclear genes, b) Mitochondrial + ribosomal genes

Species delimitation

The phylogenies and the delimitation methods support the existence of the modern species of *Monochamus sensu nov.* but highlight issues with current subspecies assignments and species circumscriptions (Linsley and Chemsak 1984, Wallin et al. 2013, Plewa et al. 2018). North American species *M. carolinensis* and *M. titillator* were found to be well separated using mitochondrial data (Fig. 1.3) and either sister taxa in the concatenated analysis (PP=1) (Fig. 1.1) or separated by *M. maculosus* in coalescent analyses (Fig. 1.2). In the multigene trees, each of these three species is monophyletic and well supported (PP=1). Possible paraphyly of *M. carolinensis* and *M. maculosus* is evident only in the mitochondrial tree (Fig. 1.3). *M. scutellatus* and *M. obtusus* are together monophyletic but mixed in the Bayesian nuclear analysis (Fig. 1.4, pp=.54) and the Bayesian supermatrix tree (Fig. 1.1, pp=1). In Eurasia, *M. saltuarius* and *M. impluviatus* are some of the most recently diverged at ~1.37 million years (Fig. 1.2). The Bayesian concatenated tree (PP=.71) and Bayesian nuclear tree (Fig. 1.4, pp=1), as well as the species tree (Fig. 1.2, pp=.5) favor *M. impluviatus* as the sister species to *M. saltuarius*. STACEY delimitation shows *M. impluviatus* at a more derived position (Fig. 1.8) rendering *M. saltuarius* paraphyletic. A new species from Japan sits sister to *M. saltuarius* + *M. impluviatus* in the concatenated (pp=1), StarBEAST2 (pp=1), *BPP* (node integrity 1.0 for nuclear & MitoRibo) and STACEY analyses (Figs. 1.1,1.2,1.6,1.7-1.9).

In North America, there are currently subspecies for *M. clamator*, *M. scutellatus*, and *M. obtusus*. We sampled all of these subspecies. *M. clamator* shows mixing of the subspecies *M. c. clamator*, *M. c. rubigineus*, *M. c. nevadensis* and *M. c. linsleyi*. *M. c. rubigineus* from Oaxaca, Mexico does show evidence of possible separation, while other samples from this subspecies from further north are interspersed with the other subspecies. The only subspecific entity found

reciprocally monophyletic is *M. c. latus* (Concat PP=1, COI PP=1, *BPP* MitoRibo node integrity=1, *BPP* nuclear=.24). The dated tree (Fig. 1.2) places this split at ~690,000 years ago. *M. obtusus obtusus* and *M. o. fulvomaculatus* are found paraphyletic in all specimen level combined analyses (Figs 1.1, A1) and the mitochondrial tree (Fig. 1.3). The sister of *M. o. fulvomaculatus* in all trees is the geographically closest specimen of *M. o. obtusus* from Oregon. Putative *M. scutellatus oregonensis* from the west coast separated from *M. scutellatus* samples to the east (and *M. obtusus*) in the concatenated tree (Fig. 1.1; PP=1). Support for *M. s. oregonensis* as sister to other *M. scutellatus* from the COI gene tree is pp=.88 (Fig. 1.3), but the single specimen from Nova Scotia is sister to *M. marmorator*. *Monochamus notatus* was split into eastern and western subspecies in the past (Hopping 1945). Saskatchewan, Canada and eastern North American samples have separation support with pp=1 in the nuclear tree (Fig. 1.4) and MLB=100 (Fig. A1). The Bayesian supermatrix tree has a partial polytomy in the *M. notatus* grouping involving the Saskatchewan specimen (Fig. 1.1), and the COI gene tree has the specimen falling outside a clade including *M. notatus* and most other NA species (Fig. 1.3). The *BPP* MitoRibo analysis gave node integrity of .98 for the subspecific split (nuclear=.46)(Fig. 1.5). The STACEY *SpeciesPhased* topology of the *M. notatus* clade showed the alleles of the Saskatchewan specimen confidently (pp=1) mixed with those of Massachusetts and sister to alleles from Michigan (Fig. 1.8).

Eurasia contains many described *Monochamus* subspecies (Danilevsky 2018). The sampling of the current study can evaluate those of *M. sartor*, *M. saltuarius,* and *M. alternatus*. Samples of *M. saltuarius* from the Czech Republic and South Korea form a monophyletic group sister to *M. impluviatus* supported by all analyses (concat. PP=.71, *BPP* nuclear node integrity=.83, BPP MitoRibo=.99). Mitochondrial and nuclear genes support the separation of *M.*

s. urussovii from Poland/Russia and *M. s. sartor* from Italy/Turkey/Hungary (Figs. 1.3, 1.4). The concatenated tree (Fig. 1.1) separates them with PP=1, all phasing strategies in the STACEY nuclear analysis find the subspecies alleles to be monophyletic at PP=1, both *BPP* trees have a node integrity score of 1.0, and the dated species tree indicates their diverging ~440,000 years ago. *M. alternatus* samples diverge along the boundaries of the current subspecies, with *M. a. alternatus* from China sister to the group of *M. a. endai* from Japan and South Korea (concat. pp=1, ML=100, COI pp=1)(Figs 1.1,1.3,1.4). The species phased STACEY result (Fig. 1.8) shows both *M. a. alternatus* alleles monophyletic while in the genus phased analysis (Fig. 1.9) an allele from Japan groups with those from China.

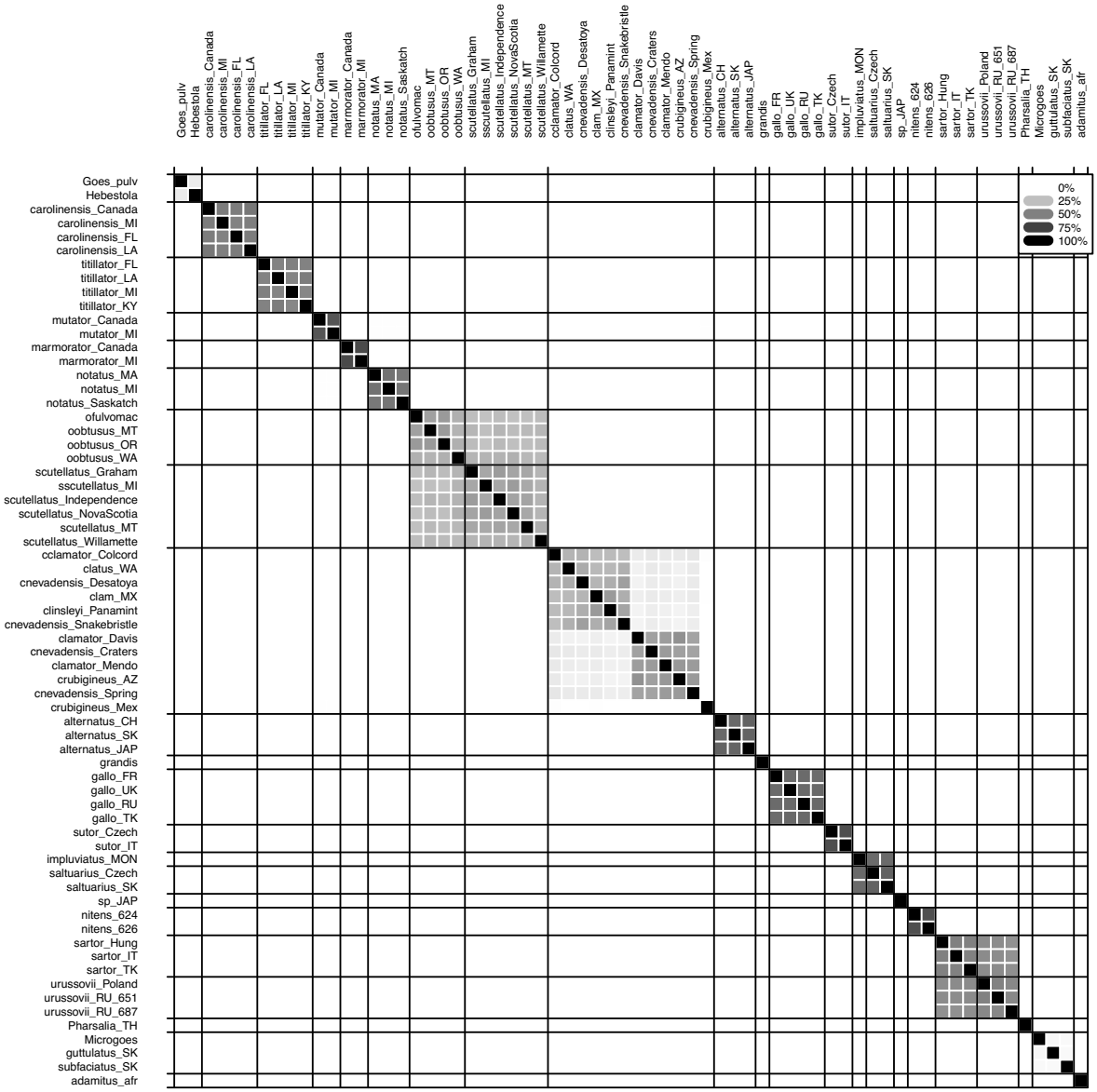

Figure 1.7. Unphased 'heterozygote' STACEY resulting simmatrix. Thin lines are proposed species boundaries, darker=higher probability they are same species

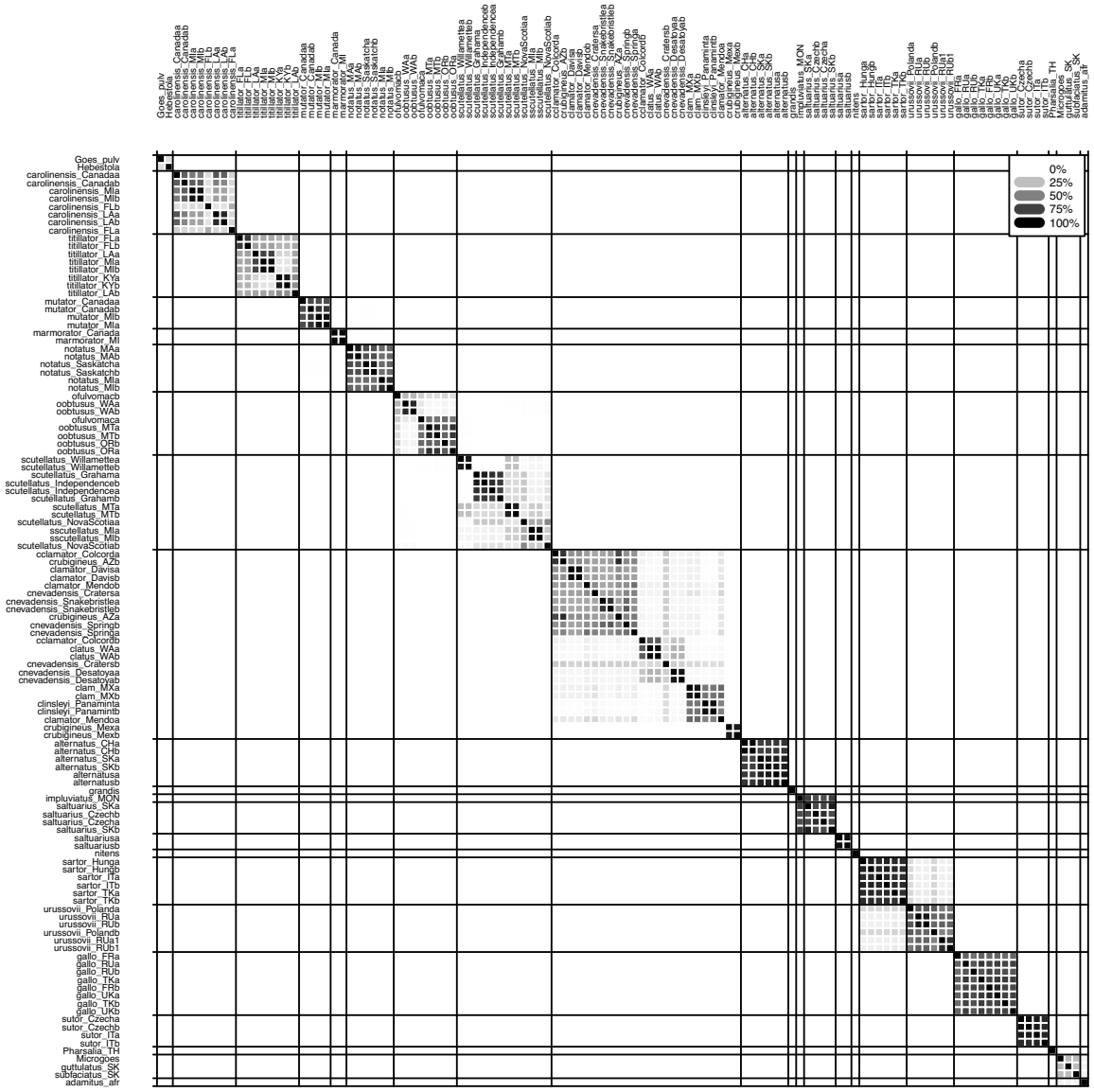

Figure 1.8. SpeciesPhased STACEY resulting simmatrix. Thin lines are proposed species boundaries, darker=higher probability they are same species

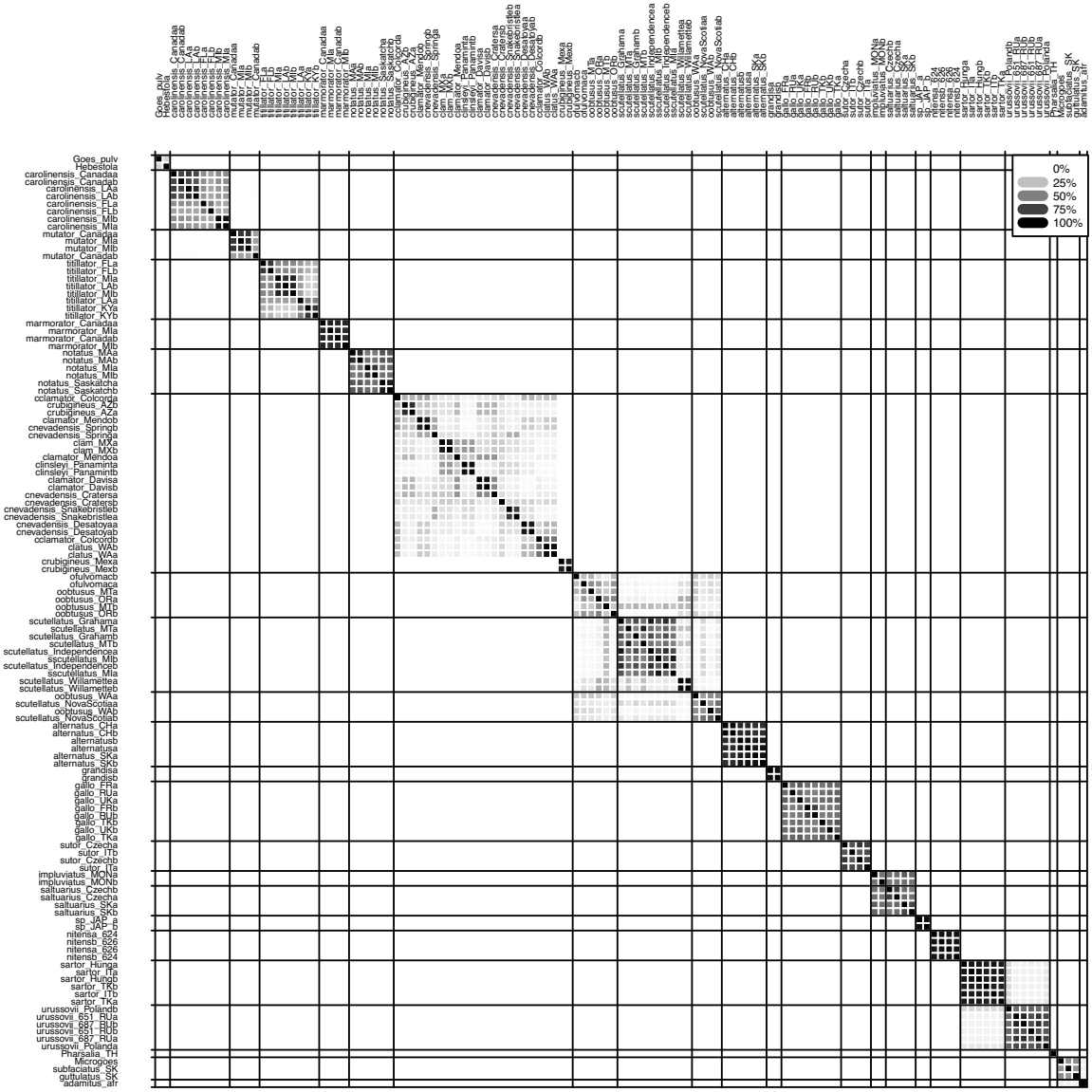

Figure 1.9. GenusPhased STACEY resulting simmatrix. Thin lines are proposed species boundaries, darker=higher probability they are same species

Nuclear allele phasing

Diploid heterozygote phasing strategies, as tested using STACEY, can produce conflicting results in topology, number of delimited species, and node height. We observed several topological discrepancies, but the species and higher level patterns were relatively stable. *M. carolinensis* and *M. titillator* are sister species in UnPhased and SpeciesPhased but *M. maculosus* renders them paraphyletic in GenusPhased results (Figs. 1.7-1.9). *M. scutellatus* and *M. obtusus* mix in all analyses with support values <.5 for most intraclade branches. The number of species predicted was the measure that shifted most drastically between analyses (Table 1.4). Each analysis had the same number of individuals, when divided into alleles species number prediction went from 38.3 species to 44.2 (GenusPhased) or 44.6 (SpeciesPhased).

The node heights in the species tree, critical information for accurate dating and cluster collapse in height-based delimitations, are shown to vary depending on phasing method and classification level (Fig. 1.10). The most drastic difference is seen at the species height, with unphased sequences the height is very close to zero while phased sequences average around .0008, above the user set cluster height. Relative measures at the sister species level show a similar trend of unphased sequences having a lower mean node height. At higher taxonomic levels (genus or multi-genus clade), the pattern switches with unphased mean node height higher than in phased sequences (Fig. 1.10c,d).

Delimitation analysis	**"real" species**	**predicted species**
BPP Nuclear A11, NA	9	27
BPP Nuclear A11, EUR	11	11
STACEY unPhased	25	38.25
STACEY SpeciesPhased	25	44.58
STACEY GenusPhased	25	44.2

Table 1.4. Species delimitation results, based on nuclear data, numbers include non-Monochamus species. STACEY results are mean number of clusters from the posterior distribution

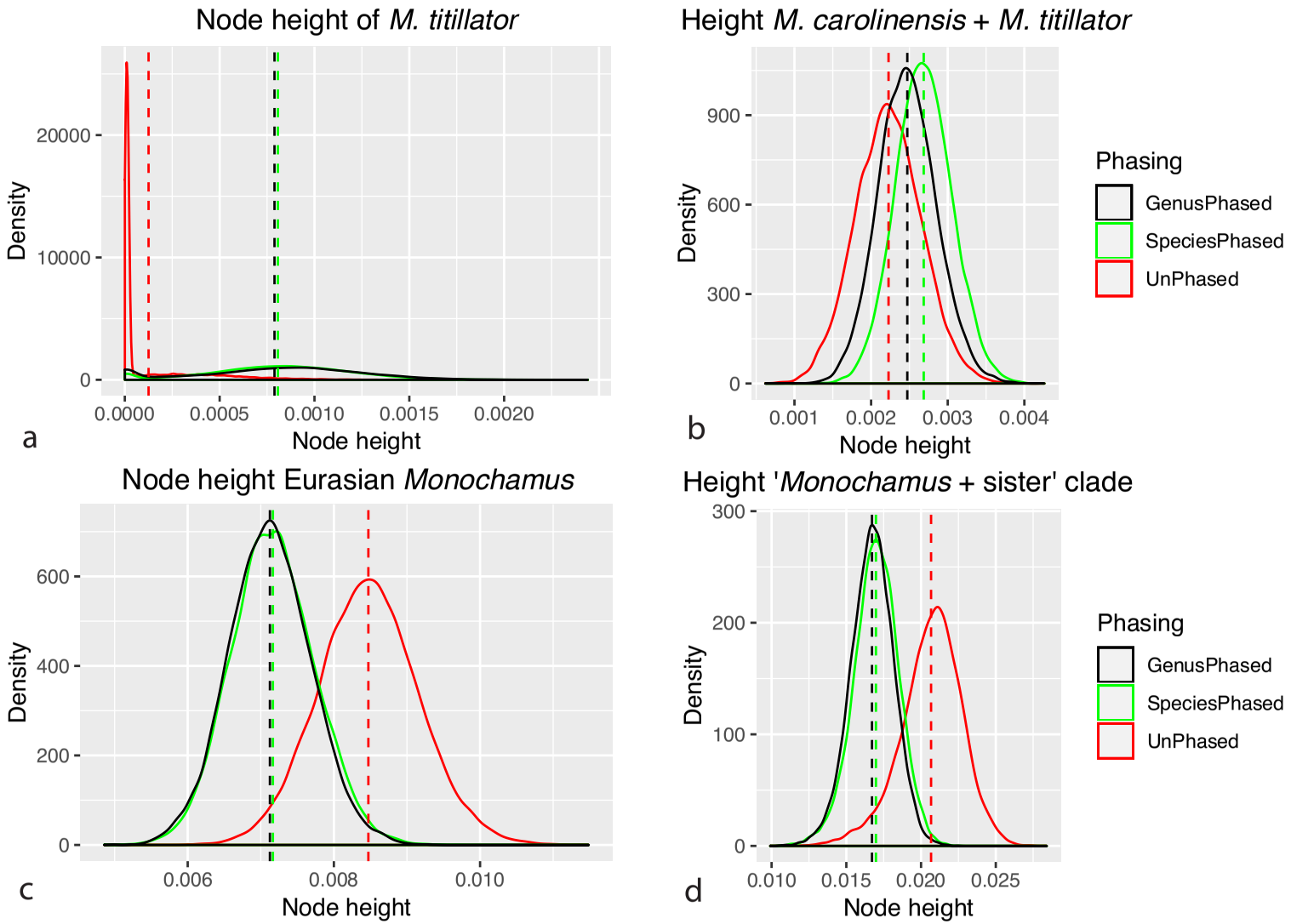

Figure 1.10. STACEY produced divergence time distributions. Each panel represents one node in the species tree. a) species *M. titillator*, b) sister species *M. carolinensis* + *M. titillator*, c) all Eurasian species of *Monochamus*, d) clade including *Monochamus*, *Goes*, and *Pharsalia*. Dashed lines are the mean node height, colored as density

Topology testing

Two constrained topologies different from the unconstrained RAxML tree were tested using the SOWH test in SOWHAT. Using the five marker nuclear data, the native RAxML tree showed *Monochamus clamator latus* as polyphyletic among *M. clamator* subspecies. With *M. c. latus* constrained monophyletic, the likelihood difference favoring the unconstrained tree is not significant: p=.07 95% CI: [.029,.139], likelihood difference 8.54. Constraining the *Goes* Clade as sister to the North American *Monochamus* using the full data reveals no topological signal. The unconstrained RAxML tree shows the *Goes* Clade as sister to Eurasian *Monochamus* + NA *Monochamus*. Testing the constrained tree shows an insignificant likelihood difference between the two topologies: p=.92 CI [.85,.96], likelihood difference -2.21.

Discussion

This study presents the first comprehensive phylogeny of the conifer feeding *Monochamus* species, including all North American and Eurasian species. It is also the first to explicitly include a worldwide sampling of angiosperm feeding *Monochamus sensu lato*. The only other phylogenetic study to include angiosperm feeding *Monochamus* has three non-conifer feeding species which are also restricted to Japan (Toki and Kubota 2010). This is the first large phylogeny to sample protein-coding loci outside of the mitochondrion, though a study of *M. galloprovincialis* and *M. sutor* used 28S (Koutroumpa et al. 2013). Support was lower in coalescent methods, including delimitation, as is usually seen when gene trees are recognized individually in smaller scale analyses (Liu et al. 2015, Edwards 2016). Given weak support at some nodes, and uncertainty of some clade placements, an increased number of genes and characters is warranted to arrive at a completely robust tree. Improved taxon sampling is required

to explore missing subgenera of *Monochamus* and rare taxa that have little or no representation in the phylogeny, like those inhabiting southwestern China and Africa.

Increased gene sampling was found sufficient for resolving most nodes, especially at the genus level. Figure 1.2 shows the best topology for the Palearctic species. This phylogeny agrees with the previously published mitochondrial tree (Cesari et al. 2005) and the treatment of *M. galloprovincialis* and *M. sutor* as sister species (Koutroumpa et al. 2013). There is some conflict, however, with the mitochondrial subtree of the *Monochamus* species of Japan(Toki and Kubota 2010). This tree placed *M. rosenmuelleri* (=*M. urussovii*) sister to *M. sutor*, and *M. nitens* was near the base of the clade. The topological arrangement of closely related species can sometimes get confused in a mitochondrial gene tree, as is seen with NA subspecies M. c. latus moving far from other M. clamator (Fig. 1.3). Discounting any mitochondrial introgression, this seems to have happened in the Japanese phylogeny since *M. nitens* is morphologically very similar to *M. urussovii*. Our tree is the first to integrate *M. grandis* and M. *nitens* with the rest of the Eurasian species. *M. nitens* is sister to *M. urussovii* + *M. sartor*. *M. grandis* is sister to *M. galloprovincialis* + *M. sutor*, and its morphology agrees with this placement. This was also the first time *M. impluviatus* was included in a phylogeny, confirming a sister relationship with *M. saltuarius*. The southern Asian *M. alternatus* is at the base of all trees to date and may indicate an origin of the genus near the generic diversity of the tribe. *Pharsalia (Cycos)* as sister to *Monochamus* promotes this origin with some features similar to *M. alternatus* and a modern overlapping range where broadleaved and coniferous forest would have been present in the Miocene (Henrot et al. 2016).

The topology of the North American species is represented in the subtree of Figure 1.1, for sister species at least. The rapid divergences that seem to have occurred ~1mya (Fig. 1.2)

have made it difficult to find an agreed 'species group' level phylogeny. The widespread *M. scutellatus* and west coast *M. obtusus* appear to be sister species but may experience gene flow in sympatry (see below). *M. notatus* and *M. marmorator* are clearly sister species and have overlapping ranges but preferences for different host plants. *M. carolinensis* and *M. titillato*r are similar looking species confirmed to be sisters. Their collective sister species, *M. maculosus,* is morphologically similar and may have hybridized with M. carolinensis in the recent past. This group being near the base of the tree agrees with the hypothesis that a mottled orange-brown vestiture may have been the ancestral condition. *M. titillator* and *M. alternatus* are still very similar in appearance after millions of years.

Higher Classification recommendations

Monochamus sensu novo includes only conifer-feeding species. All of the known species in the *Monochamus* conifer-feeding group fall geographically into North American and Eurasian subclades. These are either split or are sister groups depending on the uncertain placement of the *Goes* + *Hebestola* clade (*Goes* Clade). The placement of the *Goes* Clade is divided when single genes are removed (Table 1.3, Figs. A2-8), indicating that the topological signal varies among genes. Large population sizes and fast evolution can produce a situation where the most common gene tree does not agree with the true species tree topology due to incomplete lineage sorting. This is called an anomaly zone, which may occur near the basal splits of *Goes* and *Monochamus* clades (Xu and Yang 2016). Since mitochondrial genes can coalesce more quickly (Moore 1995, Hudson and Turelli 2003), the COI tree (Fig. 1.3) placing the *Goes* clade as sister to North American *Monochamus* is more likely in this short internal branch situation. Since the placement of the *Goes* clade can have impacts on biogeographic and diversification hypotheses, an

expanded gene study focused on its placement is warranted. The *Goes* and two *Monochamus* clades combined form a very stable clade (pp=1, all analyses) as sister to *Pharsalia (Cycos) subgemmata*. Previous treatments have focused on either Nearctic or Palearctic species, but this is the first evidence to show that they are separate monophyletic groups. Toki and Kubota (Toki and Kubota 2010) found the conifer-feeding *Monochamus* of Japan as a monophyletic group sister to angiosperm feeding species described under *Monochamus*. The genera (*Goes*, *Pharsalia*) that bound the conifer feeders in the present study are not present in Japan, where the Toki study restricted their taxon sampling. *Pharsalia (Cycos)* may be the true sister to *Monochamus sensu novo* + the Goes clade, but a thorough molecular sampling of the tribe is required since no morphological synapomorphy is evident. While there are many genera in the tribe Lamiini, the short internal branches separating the *Goes* and *Monochamus* clades indicate that they are either sister groups or the *Goes* clade renders *Monochamus sensu novo* paraphyletic. As the type species for *Monochamus* is the conifer feeding *M. sutor*, the conifer feeding species should retain the genus name.

Microgoes Casey is where small angiosperm feeding *Monochamus* belong. The genus *Microgoes* is currently monotypic and occurs in Eastern North America. This small-bodied species was found to be part of a clade of similar looking Asian *Monochamus* species that also feed in broadleaf trees. Some of these Asian species are in the clade of Japanese species found sister to Monochamus *sensu novo* (Toki and Kubota 2010). This clade of species is confidently placed outside of the conifer feeding *Monochamus* and is separated by the *Goes* group and *Pharsalia (Cycos)* (Fig. 1.1). Given its phylogenetic monophyly and the distinctive morphology consisting of longer filiform antennae in both sexes, smaller body size, procoxal cavities closed behind, and small lateral pronotal tubercles, this clade should be recognized as the genus

Microgoes rev. nov. The species belonging to this genus are listed in Table 1.5 and make it a Holarctic genus. They also all feed on broadleaf trees. The genus *Xenohammus* Schwarzer 1931 is very similar in morphology to *Microgoes* and with further research could prove to be synonymous. More thorough sampling of this clade through Asia is needed to show its evolutionary cohesiveness and define its morphological variability.

The African subgenera of Monochamus are distinct genera. *Monochamus sensu lato* contains multiple subgenera in Africa, a vestige of the description of many new genera by Dillon and Dillon in the late 1950s (Dillon and Dillon 1959a, 1959b, 1959c, 1959d, 1961). These subgenera are in no way cohesive with *Monochamus sensu nov*. and group as an African clade with deeper divergences separating them (Figs. 1.1,A9). Recognized genera like *Oxylamia* are intermixed with *Monochamus sensu lato* subgenera with high support. Since most of these are diagnosable, and are definitely outside *Monochamus sensu nov*., they should be re-elevated to (or remain at) the genus level. A number of these taxa contain very few species, as is common in the rest of the tribe which includes many monotypic genera. While several genera may have issues of monophyly, they should be addressed on a case by case basis which would be hindered by including them under a genus with unrelated taxa. A well-sampled revision of the Lamiini and related tribes is warranted and will be a large undertaking. An attempt at a revision of tribal classification using molecular data is underway (Gorring, unpub.).

Table 1.5. Classification updates of *Monochamus sensu lato*

Monochamus Dejean Eurasia includes:	*M. alternatus alternatus, M. a. endai, M. saltuarius, M. nitens, M. grandis, M. sartor, M. urussovii, M. sutor, M. impluviatus, M. galloprovincialis*
Monochamus Dejean North America includes:	*M. carolinensis, M. clamator clamator, M. c. latus, M. maculosus, M. marmorator, M. notatus, M. scutellatus, M. titillator*
Monochamus Dejean incerte sedis	*M. nigromaculatus* Gressitt, *M. talianus* Pic
Microgoes Casey	*Microgoes oculatus, Monochamus subfasciatus stat. nov., M. guttulatus stat. nov., M. masaoi stat. nov., M. maruokai stat. nov., M. rectus stat. nov., M. abruptus stat. nov., M. foraminosus stat. nov., M. sparsutus stat. nov.*
As genera	All current *Monochamus* subgenera
Lamiini Latreille incerte sedis	all unplaced species from *Monochamus* subgenus

Species boundaries in Monochamus sensu novo

One of the primary goals of this study is to delimit the species in the genus *Monochamus*. This means analyzing the species down to the allele level. Taxon sampling was attempted across species' ranges to give a representation of intraspecific variation. Delimitation followed a thorough investigation of the data. Potential incipient species get recognized at the subspecific level, with the definition that they are diagnosable entities differentiated from other populations but likely still reproductively compatible. The delimitation has revealed nuances in the data including mitochondrial introgression, incomplete lineage sorting, and mitonuclear discordance. The tools used to delimit species were *BPP* (Yang and Rannala 2010) and the STACEY package (Jones 2017), each of which employs the multispecies coalescent model. Using the MSC for delimitation has raised concerns of it recognizing populations and not species, especially with hundreds of loci (Sukumaran and Knowles 2017, Leaché et al. 2019). Isolation by distance can

be pronounced in widespread species, like many in *Monochamus,* and presents a challenge for the connectivity of taxon sampling. Using our nuclear delimitations as an example, when sampling from a broader part of a species geographic range in North America, more species were predicted in *BPP* and STACEY. In the *BPP* A11 analysis for Eurasia, data for what we felt were ten valid species resulted in the highest posterior probability (pp) for ten species (though the groupings were not all ideal). With broader sampling for a putative eight species in NA, there was highest pp for 26 species. Based on our valid species, *BPP* predicted 195% (37spp./19) and STACEY SpeciesPhased predicted 178% (44.43spp./25). In the end, despite inflated delimitations, integrated datasets can be used to test these reasonable species hypotheses and model parameters can be fine-tuned to trusted data. Intraspecies diversity could be a factor impacting species inflation in this data. Eurasian species tend to be more cohesive, as shown by dark blocks (strong pairwise affinity) in STACEY simmatrices (Figs. 1.7-1.9, A10) and more nodes collapsed in BPP.

Eight species are recognized in North America. This diversity seems to have evolved quickly with a crown age of about two million years (Fig. 1.2). The North America species have received much taxonomic attention and the classification is fairly settled (Hopping 1921, Dillon and Dillon 1941, Linsley and Chemsak 1984). The questioned validity of the morphologically similar, and often sympatric, species *M. titillator* and *M. carolinensis* has been one of the most prominent arguments. Due to the intraspecific variability in some of the diagnostic morphological characters, like the armature of the elytral apices, and geographical similarity they have been placed in synonymy in the past (LeConte 1852). Recent studies have shown differing genitalic (Pershing and Linit 1985) and behavioral (Walsh and Linit 1985) characters. We find that these two species are distinct using molecular data (Figs. 1.1,1.3,A1) and remain

monophyletic at the allelic level in STACEY delimitation analyses (Fig. 1.8). Though they have retained similar morphology, they diverged ~1.7 mya (Fig. 1.2). They are often found sister to one another but are sometimes split by *M. maculosus* which shows evidence of mitochondrial introgression with *M. carolinensis* where they meet near the US Canadian border. There is evidence of this in the COI tree where samples of *M. carolinensis* and *M. maculosus* from Canada and Michigan form a clade exclusive of Southern US samples (Fig. 1.3). Dating is also more likely to represent when these species hybridized (~380kya) and not a true divergence time. Other trees that include the COI evidence (Fig 1.2,A1) show a pulling of *M. maculosus* toward *M. carolinensis* while nuclear gene-based trees show *M. titillator* and *M. carolinensis* as sisters with pp > .9 (Fig. 1.4, 1.8). Introgression in conifer feeding insects that may share hosts is common in *Neodiprion* sawflies (Linnen and Farrell 2007, 2008) and may not be rare. The presence of mitochondrial and not nuclear exchange may align with a situation where one species is in low abundance, like at the edge of its range, and females are more likely to encounter aggressive males of the dominant species (Chan and Levin 2005). With multiple *Monochamus* species feeding on the same species of ephemeral host resources, situations like this could be present whenever one species is in low abundance. The argument could be made that introgression of both nuclear and mitochondrial genes could also be happening in these sympatric environments. This hybridization would result in the incorrect topology and deserves further study at the population level.

Two species that are seemingly very morphologically distant nevertheless show a close relationship on the molecular level, *M. scutellatus* and *M. obtusus*. These two species co-occur over much of the western coastal conifer forest of North America. While BPP delimited them confidently in nuclear and MitoRibo trees (Fig. 1.5) nuclear analyses showed the mixing of these

species (Fig. 1.4, 1.8). One potential reason that nuclear genes did not sort, but mitochondrial did, could stem from the short branches in this area of the phylogeny not giving a proportion of genes enough time to coalesce (Degnan and Rosenberg 2009). One of the best ways to confirm topologies that may experience lineage sorting is to look at many independent gene genealogies

In North America, *M. clamator*, *M. scutellatus,* and *M. obtusus* each currently has recognized subspecies (Linsley and Chemsak 1984). *M. scutellatus* has two subspecies: *M. s. oregonensis* from the west coast to a proposed hybrid zone on the border area of British Columbia and Alberta, Canada and *M. s. scutellatus* to the east. These entities were inferred to have been discrete species at one point, but an analysis of the proposed hybrid and parental species found only ambiguous quantitative differences and not discrete characters for the species (Raske 1973). Furthermore, experimental crosses produced fertile offspring, and therefore these eastern and western populations were relegated to subspecies (Raske 1973). Our sampling included a west coast individual, some from the intermountain west, and samples from the middle and east USA. The subspecies did tend to separate in analyses, sometimes with *M. obtusus* causing paraphyly (Fig. 1.3, 1.4). In the BPP analysis, more support was found in nuclear (5 diagnostic sites in ArgK) than the MitoRibo dataset (Fig. 1.5), perhaps indicating that while an ancestral separation was not long enough to build reproductive barriers some of the genes and morphological traits were subject to drift. Mitochondrial DNA and morphology could now be re-homogenizing since contact between east and west populations was re-established. Because there seems to be free reproduction in contact and loss of geographic differentiation, we recommend removing subspecies designations from *M. scutellatus*.

Monochamus obtusus is separated into the nominate and *M. o. fulvomaculatus* subspecies, diagnosed by differences in color and abundance of pubescence. They are not

monophyletic in our analyses, but *M. o. fulvomaculatus* does fall sister to an *M. o. obtusus* sample from Oregon at a derived position in multiple analyses aligning with a peripatric mechanism. If this has occurred, gene trees representing the topology of *M. o. fulvomaculatus* as sister should be present and mtDNA should diverge early. No evidence exists indicating divergence of *M. o. fulvomaculatus* and the color and amount of pubescence may be plastic and environmentally determined so subspecies status should be removed. *Monochamus clamator* from western NA has the largest number of named subspecies in North America: the nominate subspecies *clamator*, plus *rubigineus*, *latus*, *nevadensis* and *linsleyi*. After the inclusion of multiple examples from each proposed population, the only one found to group monophyletically is *M. c. latu*s, and only in analyses that include mitochondrial information. When SH topology tests are used on the nuclear data, there is a lower likelihood for a tree with *M. c. latus* constrained, but with p=.07, indicating the nuclear data is lacking signal. Since *M. c. latus* shows strong mitochondrial isolation, but nuclear alleles do not, we support retaining subspecies status. The remaining entities, which show differences that appear clinal and occur syntopically, should have their subspecies status removed. The one sample of *M. clamator* from Oaxaca, Mexico does have delimitation support in STACEY, but this may be due to its geographic distance from other samples and requires further sampling through Mexico and Central America. *M. notatus* has also been split in the past using limited sampling in British Columbia for the subspecies *M. n. morgani* (Hopping 1945). Delimitation results are similar to *M. clamator*, with support from mitochondrial but not nuclear analyses. Its range overlaps in large part with *M. scutellatus,* but *M. notatus* seems to have a different evolutionary history. More sampling from the western extent of its range is necessary before making any taxonomic decisions.

Eight *Monochamus* species are presently known in Europe and Asia; we have determined one more previously undescribed species. We find support for the current species delimitations, and our analyses are unambiguous for most of them. The clade began diversifying about two million years before the North American species group (4.3MYA), about one species split per branch per million years (Mayr 1942). The *saltuarius* species group contains three similar looking species, two are continental and one occurs on Japan (sp. nov.). We find support for these in both delimitation methods with BPP having strong node integrity for the *M. saltuarius/M. impluviatus* split in the MitoRibo tree but weaker integrity in the nuclear tree. The *speciesPhased* STACEY analysis shows some mixing with weak node support, stemming from low nuclear signal on account of partially degraded *M. impluviatus* tissue that did not amplify for multiple nuclear markers. Despite this, male genitalia have been able to diverge with a novel sclerite present in *M. impluviatus* (Wallin et al. 2013). The new species in Japan has high support from all analyses, and all alleles have sorted with eight fixed differences in AK. All individual StarBEAST2 gene trees except COI agree with the new species as sister, a result unlikely if it were the same as *M. saltuarius* and there was a peripatric origin of *M. impluviatus*. A COI gene tree incorporating GenBank samples confirms the (M. sp. nov.,(*M. saltuarius*, *M. impluviatus*)) topology with high support (Gorring, unpub.). *M. impluviatus* is a *Larix* specialist, a very unusual host association among mostly Pinaceae generalist species. This may have contributed to its divergence from the more generalist-feeding *M. saltuarius*.

In Eurasia, subspecies have been proposed for six of eight *Monochamus* species (Danilevsky 2018) most of which inhabit both subcontinents. The subspecies of *M. galloprovincialis* (see Koutroumpa et al. 2013), *M. impluviatus* and *M. sutor* are not treated here and deserve future evaluation. *M. saltuarius* was recently separated into the subspecies *M. s.*

saltuarius in eastern Eurasia and *M. s. occidentalis* in western Eurasia (Sláma 2017). We gathered samples from the Czech Republic and South Korea which form a monophyletic group sister to *M. impluviatus,* which is expected from similar morphology. The alleles of the two proposed subspecies were not monophyletic and even mixed with those of M. impluviatus (Fig. 1.8). *M. impluviatus* can be confidently separated from *M. saltuarius* by male genitalia (Wallin et al. 2013). While complete sorting of alleles is not a prerequisite for subspecies status, more isolation would be expected between Czech and Korean samples. In a Bayesian COI gene tree (Gorring, unpub.), there is no split between samples from Europe and the Russian Far East. It may have been premature to name subspecies without a complete sampling of the geographic range as the characters used for the subspecies description may be purely clinal. Recent work by European researchers has focused on showing that *M. sartor sartor* and *M. s. urussovii* are valid subspecies (Plewa et al. 2018). They used low COI divergence around 1%, a single EF1a haplotype, and nesting of *M. s. sartor* inside of *M. s. urussovii* in their COI tree as arguments for subspecies status. The authors also note that there is limited gene flow, distinct morphology of wing veins (Rossa et al. 2016), and each subspecies has distinct strains of *Wolbachia* bacteria. Another researcher has found gene flow between populations of *M. s. sartor* and *M. s. urussovii* (J. Goczal, pers. Comm.). We find mitochondrial and nuclear genetic support for the separation of *M. s. urussovii* from Poland and Russia and *M. s. sartor* from Italy, Turkey, and Hungary (Figs 1.3, 1.4). EF1a, *wg*, TOPO, and CAD results are similar to the Plewa et al. study, finding the same haplotype for all samples. Arginine Kinase had three fixed differences in coding regions, 28S showed two, and COI had nine. Though COI divergence is lower than the ~3% found in the comparison of *M. galloprovincialis* and *M. sutor*, nuclear divergence was close to the study's one polymorphism and two deletions in 28S (Koutroumpa et al. 2013).

A depression of COI divergence could be caused by a low level of mitochondrial introgression as seen in *M. carolinensis*. With all trees supporting the separation, a distinct morphology, and little evidence of gene flow after contact for 10,000+ generations since the last glacial maximum, *M. s. sartor* and *M. s. urussovii* should be different species. According to the dated StarBEAST analysis (Fig. 1.2), species divergence times less than one million years (~1.5% COI) are not uncommon in this genus and could have been encouraged in part by glacial refugia. Continued isolation of species does not necessitate a certain level of neutral divergence and other characteristics to recognize two populations as species are stochastic in the chronology of appearance (De Queiroz 2007). The *M. s. urussovii* situation is different from that in *M. scutellatus* because the entities are remaining distinct in mtDNA and morphology despite secondary contact. One future direction for this comparison would be to test their cuticular hydrocarbons, which *Monochamus* beetles use to recognize conspecifics and maturity in females (Brodie et al. 2012).

The final subspecies pair analyzed were *M. alternatus alternatus* and *M. a. endai*. With only one sample from the nominate population, this was not a complete analysis, but results show a genetic pattern consistent with their proposed ranges (Makihara 2004). Evidence to the contrary was found using the COII gene, with samples from China mixing with those from Japan (Kawai et al. 2006). There was no evaluation of the proposed morphological differences in that study. A vicariant event is consistent with Pleistocene connections of Japan and Korea and the past forest makeup of eastern China (Qiu et al. 2011). These subspecies may be valid, thorough sampling of alleles and morphology across their ranges would allow for an in-depth evaluation of population structure in this important southern Asian species.

The results of phasing nuclear data

The phasing of nuclear genes into their alleles and not just using ambiguity codes has been promoted in systematics (Kubatko et al. 2011, Andermann et al. 2018) and is very important when analyzing gene variants. The true impact of phasing in phylogenetics has not really been investigated and is rarely a step in phylogenetics pipelines. It is logical that phasing alleles captures more of the variability in the data and can provide more statistical power, but whether that is important for species level phylogenetics, or if subspecies delimitations can benefit, requires more data. While one of the most obvious effects is on the number of species predicted in a delimitation analysis (Table 1.4), the root of species estimation differences is the node height that the model uses for collapsing and clustering similar terminals in STACEY. Many delimitation methods seek a similar break in species/population patterns. Simulations have shown deviation from true node heights, with IUPAC consensus sequences consistently overestimating heights and phased sequences closest to the true height values (Andermann et al. 2018). High nodes can lead to unrealistic deeper divergence times and potentially less species predicted overall. In the *Monochamus* empirical dataset, we found what seems to be a similar pattern in higher taxonomic nodes (Fig. 1.10c,d), but there is no true value to compare to with empirical data. For nodes at the species and sister species level the relative pattern presented by phasing strategy switches, with unphased sequences at a lower mean node height than phased sequences. The result from this change will be less species delimited with unphased data (see Table 1.4) if the height falls under the set collapse height value. In the example of *Monochamus titillator* (Fig. 1.10a), the unphased node distribution would completely fall under the assigned collapse height, giving high posterior probability that all samples belong to the same species. The phased data for *M. titillator* has a higher mean node height, meaning the clustering probability

for a single species will be lower. If a dataset contains some species the researcher is confident in and some that are being questioned, this method can be run to determine the node height of phased 'good' species and that value can be used as collapse height for a complete STACEY run. Node height is also important when estimating species divergence times on a phylogeny and could potentially move nodes out of narrow date ranges that could be important for hypothesis testing (eg. climate based refugial hypotheses).

There was not a drastic impact on topology in our phasing comparisons, and there was generally low support for allele grouping below the species level. There is potential for alleles to show a level of sorting within structured species that could support subspecies designations before there is a species level divergence. In phased results for *Monochamus* (Fig. 1.8) there are blocks within species that are more cohesive than others (higher pp). This can be investigated further with more genes or allele frequency analysis. Based on the simulation results on node height (Andermann et al. 2018) and the sorting of potential species in our analysis, we promote phasing data by species if possible (*SpeciesPhase*). This will provide more accurate divergence time estimates, improved statistical power, and the ability to test hypotheses of species using the fundamental unit that most species tree models are built on, the allele. There just needs to be an awareness of the potential to split species into multiple populations by the MSC model (Leaché et al. 2019), this can be more severe when using allele sequences.

Missing gene analyses

There were very few topological differences between phylogenies missing one gene (Table 1.3, A2-8), and they were a good match to the full data StarBEAST2 run (Fig. 1.2). The topology of all analyses except no-COI differed from the preferred topology in having *M.*

carolinensis not sister to *M. titillator*, *M. marmorator* not sister to *M. notatus*, and *M. scutellatus* not sister to *M. obtusus*. Removing COI had the most drastic effect: the topology moved toward the preferred topology. Interestingly, the no-COI analysis gave the minority topology of *Goes* Clade sister to NA *Monochamus*, matching the COI gene tree (Fig. 1.3). The missing gene results show that COI may have a disproportionate impact on the topology, and strict analyses of nuclear genes should be taken into account when interpreting the relationships of species.

Dating and biogeography

Several biogeographic trends are worth discussion, though a model-based analysis was not a part of this study. Dating of the tree using a strict clock following COI evolutionary rates determined for *Tetraopes* longhorned beetles (Farrell 2001) places the crown group age of *Monochamus sensu nov*. around 5.3 million years. The COI evolutionary rate used was similar to that measured for other insects which span 1.5-2.3%/MY (Brower 1994, Quek et al. 2004, Sota and Hayashi 2007). With species that are closely related, and that are lacking calibration information, this method can give reasonable results. However, shallow divergence can also show inflated rates of sequence divergence before 1.5my (Ho et al. 2005, Sota and Hayashi 2007). Therefore, with deep, potentially saturated divergences or many shallow nodes, other dating methods should be explored. Clade age represents a Miocene/Pliocene dispersal to become Holarctic over the second Beringian Bridge, which was covered in coniferous taiga from 14-3.5 million years before present (Sanmartín et al. 2001). The geographic origins of the clade are a bit more tenuous and depend on the uncertain placement of the *Goes* clade. This is problematic as the taiga only contains coniferous trees and all *Goes* and *Hebestola* species feed in broadleaf trees. A Nearctic origin could be imagined where the ancestor of *Goes* and

Monochamus clades splits in a hospitable portion of North America, and a conifer feeder migrates across the land bridge with subsequent diversification on each continent. If the *Goes* clade falls sister to North American (NA) *Monochamus,* this promotes a Palearctic origin of the ancestor of Eurasian *Monochamus* and NA Goes and *Monochamus* clades, with a subsequent split in NA mixed forest. Most of the generic diversity basal to these clades resides in southern Asia. In agreement with our higher level discussion, we prefer the placement of the *Goes* clade sister to NA *Monochamus* species which in turn would support an Asian origin of *Monochamus*. Other beetle taxa, including *Agonum* ground beetles (Liebherr and Schmidt 2007) and *Plateumaris* leaf beetles (Sota et al. 2008) dispersed, in part, across the second Beringian Bridge with events in both directions.

Past land bridge connections between eastern Asia and Japan align with patterns of ancient vicariance and recent dispersal in multiple *Monochamus* clades. A land bridge connecting the Korean Peninsula to southern Japan was present for a significant total time during the late Miocene, Pliocene, and even during Pleistocene glacial maxima (Kitamura et al. 2001, Comes et al. 2014). Conifer forest connections to northeastern Asia through Sakhalin were present as well in the late Miocene and Pleistocene maxima (Pietsch et al. 2012). *M. grandis* is endemic to Japan and geographically proximate islands of the Russian Far East. Historical land connections allow for histories of species shared between Japan and the mainland as well as species closely related to mainland species. *M. nitens* diverged from mainland *M. urussovii* at 1.4mya with another more recent introduction as *M. urussovii* is also present in Japan. *M. nitens* was likely insular until a Pleistocene connection was present with Asia. Now a small population is present in South Korea. A similar situation likely occurred with *M. alternatus* which has a more recent connection with the Korean population (subspecies *M. a. endai* for both regions) and

is more genetically distant from China (*M. a. alternatus*). *M. grandis* diverged from mainland *M. galloprovincialis* + *M. sutor* at 2.1mya with *M. sutor* re-dispersing recently. *Plateumaris* leaf beetles that share similar modern ranges show similar trends: older (Pliocene) vicariant events and recent (Pleistocene) colonization events from Sakhalin through Hokkaido (Sota and Hayashi 2007). The ancestor of flightless cerambycid genus *Mesechthistatus* Breuning also shows evidence of late Pliocene colonization of Japan (Nakamine and Takeda 2008). Unfortunately, there is no sampling of the northern Asia *Monochamus* species (*M. urussovii* & *M. sutor*) shared between Japan and the mainland in this study. Exploration of the intricacies of species movement through the Pleistocene using a geographically sampled tree of Asian taxa is a future direction.

Host plant relationships

Conifer feeding insects introduce something of a paradox, they feed on an abundant, widespread resource but are depauperate in species diversity relative to angiosperm feeding allies (Farrell 1998). Understanding their origins and diversification can potentially illuminate any differences between conifers and angiosperms as substrates for the evolution of herbivorous insects. The present study reveals conifer feeding *Monochamus* as a derived clade representing a single switch to conifer feeding. Ancestral diversity in the tribe Lamiini is large (> 1500 spp.) and is purely angiosperm feeding. Scolytine bark beetles are a relatively ancient conifer-feeding group, they are ancestrally conifer feeding but originated from angiosperm-feeding ancestors within the weevils. Scolytines had one shift to a successful angiosperm clade, and few shifts of derived genera in these two host groups to the conifer or angiosperm feeding habit (Sequeira et al. 2000). Almost no insect species feed on both angiosperms and on conifers. Genera that are found on both host groups are usually feeding on decayed wood or fungus. Most often,

associations with these different divisions of plants are very conservative in herbivorous insects, with tribes or subfamilies typically associated with either conifers or flowering plants, but not both. Because ranks such as genera are arbitrary, it would be helpful for comparing different lineages of herbivores to standardize the rate of host shifts per speciation event or per unit of time (Farrell and Mitter 1993).

This general observation opens the question as to whether there is a connection between such conservatism and diversification. If there is a difference between conifers and angiosperms that is relevant to insect population biology and speciation, perhaps through their different kinds of defensive traits or population structures, then we would expect to see consistent differences in the evolutionary trajectories of their associated insect groups (Farrell 1998). The hypothesis of coevolution (Ehrlich and Raven 1964) and the host-plant population architecture hypothesis (Barton and Charlesworth 1984) are two major ideas of how the diversity of angiosperm feeding insect species has arisen. Ehrlich and Raven (1964) proposed that herbivore insects specialized for feeding on chemically defended plants show increased diversity stemming from plant-insect arms races and further research has provided evidence of this (Mitter et al. 1991, Futuyma and Agrawal 2009, Agrawal et al. 2009). Alternatively, the greater population structure (or 'patchiness') of many flowering plants, enabled by insect pollination, may confer similar structure in herbivores (but see Peterson and Denno 1998). Tree patchiness may reach an extreme in tropical forests, where one species of tree can be as far as possible from its nearest conspecifics to avoid specialist herbivores (Janzen 1970, 1973). In contrast, temperate forests tend to be more homogeneous with fewer tree species, and a relatively high proportion of wind-pollinated species (such as oaks, maples and birches) with contiguous distributions.

In contrast to the processes hypothesized to promote diversification in angiosperm feeders, conifer feeding species have abundant plant biomass and a relatively narrow range of resin-based defenses (Farrell et al. 1991). Natural selection for specialization and ecological speciation is generally thought more intense and pervasive in tropical forests than in temperate forests (Schluter 2001), where more often climate-related vicariance may result in range fragmentation leading to speciation. Host-specialist *Neodiprion* sawflies show that geographic separation likely initiated speciation, followed by a host shift (Linnen and Farrell 2010). The angiosperm-feeding *Goes* clade and conifer-feeding *Monochamus* clade in North America are approximately the same age and have equivalent diversity, indicating that their use of these different groups of temperate trees does not have an obvious influence on diversification rates. There are still few studies indicating a role of host-plant population structure in ecological or geographic speciation (Farrell and Mitter 1993, Denno et al. 1995, Schluter 2000). Overall, the patterns in these *Monochamus* beetles support hypotheses of geographic vicariance between and within continents. Extended periods of separation seem necessary, as they are surprisingly vagile and recent glacial cycling with host isolation in North American sky islands has not produced appreciable genetic structure (see chapter 3).

Conclusions

The first expansive phylogeny of the genus *Monochamus* shows that conifer feeding species are the true *Monochamus*, small Asian angiosperm feeders belong to the genus *Microgoes*, and African species are genetically distant. We recognize eighteen conifer feeding species in the revised definition of *Monochamus* rev. nov. Many challenges were present in this dataset including incomplete lineage sorting, mitonuclear discordance, and hybridization.

Species tree and MSC species delimitation methods were able to delimit most species with some expected inflation, and gave a good indication of the evolutionary processes at play when methods broke down. Evaluations of the best sequence phasing strategy reveal that using ambiguity codes can result in misleading divergence times and care must be taken to understand how the samples are structured when using cut-off type delimitation methods. The *Monochamus* species present in the new world were shown to have dispersed from Asia over the second Bering Bridge during the late Miocene/early Pliocene and subsequently split into conifer feeding and angiosperm feeding clades. Climate induced landscape changes have left their mark on *Monochamus* species, especially those of Japan where there have been multiple waves of immigration. Though geography, and thus allopatry, has been implicated in the speciation history of many *Monochamus* species, host plant influences are still a viable isolating factor. Some of the most recently diverged sister species seem tend to differ in host preference which may have resulted from sympatric differentiation or recent isolation and adaptation to limited hosts instigated by glacial cycling.

Literature cited

Agrawal, A. A., M. Fishbein, R. Halitschke, A. P. Hastings, D. L. Rabosky, and S. Rasmann. 2009. Evidence for adaptive radiation from a phylogenetic study of plant defenses. Proceedings of the National Academy of Sciences of the United States of America 106:18067–72.

Andermann, T., A. M. Fernandes, U. Olsson, M. Töpel, B. Pfeil, B. Oxelman, A. Aleixo, B. C. Faircloth, and A. Antonelli. 2018. Allele Phasing Greatly Improves the Phylogenetic Utility of Ultraconserved Elements. Systematic Biology 0:1–15.

Avise, J. C. 1991. Ten Unorthodox Perspectives On Evolution Prompted By Comparative Population Genetic Findings On Mitochondrial Dna. Annual Review of Genetics 25:45–69.

Barton, N. H., and B. Charlesworth. 1984. Genetic revolutions, founder effects, and speciation. Annual Review of Ecology and Systematics 15:133–164.

Bezark, L. G. 2017. Checklist of the Oxypeltidae, Vesperidae, Disteniidae and Cerambycidae, (Coleoptera) of the Western Hemisphere.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C. H. Wu, D. Xie, M. A. Suchard, A. Rambaut, and A. J. Drummond. 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology 10:1–6.

Bousquet, Y., S. Laplante, H. E. Hammond, and D. W. Langor. 2017. Cerambycidae (Coleoptera) of Canada and Alaska: identification guide with nomenclatural, taxonomic, distributional, host-pant, and ecological data. Nakladatelstvi Jan Farkac, Prague.

Brodie, B. S., J. D. Wickham, and S. a. Teale. 2012. The effect of sex and maturation on cuticular semiochemicals in Monochamus scutellatus (Coleoptera: Cerambycidae). The Canadian Entomologist 144:801–808.

Brower, A. V. 1994. Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proceedings of the National Academy of Sciences 91:6491–6495.

Castresana, J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Molecular Biology and Evolution 17:540–552.

Cesari, M., O. Marescalchi, V. Francardi, and B. Mantovani. 2005. Taxonomy and phylogeny of European Monochamus species: first molecular and karyological data. Journal of Zoological Systematics and Evolutionary Research 43:1–7.

Chan, K. M. a, and S. a Levin. 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution; international journal of organic evolution 59:720–9.

Cherepanov, A. I. 1990. Cerambycidae of Northem Asia. Vol. 3. Lamiinae Part I. Page (A.

Dhote and V. Kothekar, Eds.). English Tr. Oxonian Press Pvt. Ltd., New Delhi.

Church, S. H., J. F. Ryan, and C. W. Dunn. 2015. Automation and evaluation of the SOWH test with SOWHAT. Systematic Biology 64:1048–1058.

Cohn, T. J. 1965. The Arid-Land Katydids of the North American Genus Neobarrettia (Orthoptera: Tettigoniidae): Their Systematics and a Reconstruction of Their History. Miscellaneous Publications of the Museum of Zoology, University of Michigan:179.

Comes, H., Y.-X. Qiu, S. Sakaguchi, N. Yuan, and X.-S. Qi. 2014. A strong 'filter' effect of the East China Sea land bridge for East Asia's temperate plant species: inferences from molecular phylogeography and ecological niche modelling of Platycrater arguta (Hydrangeaceae). BMC Evolutionary Biology 14:41.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Craighead, F. 1923. North American cerambycid larvae: a classification and the biology of North American cerambycid larvae. Dominion of Canada, Department of Agriculture Bulletin 27.

Danilevsky, M. L. 2018. CATALOGUE OF PALAEARCTIC CERAMBYCOIDEA.

Degnan, J. H., and N. a Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in ecology & evolution 24:332–40.

Denno, R., M. McClure, and J. Ott. 1995. INTERSPECIFIC INTERACTIONS IN PHYTOPHAGOUS INSECTS: Competition Reexamined and Resurrected. Annual review of entomology 40:297–331.

Dillon, E. S., and L. S. Dillon. 1959a. The Monochamini (Cerambycidae) of the Ethiopian Faunistic Region I. Subtribe Monochamidi. Pseudhammus and Allies.

Dillon, E. S., and L. S. Dillon. 1959b. The Monochamini (Cerambycidae) of the Ethiopian Faunistic Region VI. The subtribe Docohammidi. The Coleopterists' Bulletin 13:7–12.

Dillon, L. S., and E. S. Dillon. 1941. The tribe Monochamini in the Western Hemisphere.

Dillon, L. S., and E. S. Dillon. 1959c. The Monochamini (Cerambycidae) of the Ethiopian Faunistic Region III. Subtribe Monochamidi. Oxylamia and related genera. Entomologische Arbeiten aus dem Museum G. Frey Tutzing Bei München 10:415–463.

Dillon, L. S., and E. S. Dillon. 1959d. The Monochamini (Cerambycidae) of the Ethiopian Faunistic Region IV. Subtribe Monochamidi. Melanopolia and Allies. Annals of the Entomological Society of America 52:552–566.

Dillon, L. S., and E. S. Dillon. 1961. The Monochamini (Cerambycidae) of the Ethiopian Faunistic Region II. Subtribe Monochamidi. Genera related to Monochamus. Bulletin of the British Museum (Natural History). Entomology. 11:61–96.

Duffy, E. A. . 1968. A Monograph of the Immature Stages of Oriental Timber Beetles (Cerambycidae). Thanet Press at Margate, London.

Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? Evolution 63:1–19.

Edwards, S. V. 2016. Phylogenomic subsampling: a brief review. Zoologica Scripta 45:63–74.

Ehrlich, P. R., and P. H. Raven. 1964. Butterflies and Plants: A Study in Coevolution. Evolution 18:586–608.

Farrell, B. D. 1998. "Inordinate Fondness" explained: why are there So many beetles? Science (New York, N.Y.) 281:555–9.

Farrell, B. D. 2001. Evolutionary assembly of the milkweed fauna: cytochrome oxidase I and the age of Tetraopes beetles. Molecular phylogenetics and evolution 18:467–78.

Farrell, B. D., D. E. Dussourd, and C. Mitter. 1991. Escalation of Plant Defense: Do Latex and Resin Canals Spur Plant Diversification ? The American Naturalist 138:881–900.

Farrell, B., and C. Mitter. 1993. Phylogenetic Determinants of Insect/Plant Community Diversity. Pages 253–66 *in* R. E. Ricklefs and D. Schluter, editors. Species diversity in ecological communities : historical and geographical perspectives. University of Chicago Press, Chicago, IL.

Flot, J. F. 2010. Seqphase: A web tool for interconverting phase input/output files and fasta sequence alignments. Molecular Ecology Resources 10:162–166.

Futuyma, D. J., and A. A. Agrawal. 2009. Macroevolution and the biological diversity of plants and herbivores. Proceedings of the National Academy of Sciences 106:18054–18061.

Futuyma, D. J., and G. Mayer. 1980. Non-allopatric speciation in animals. Systematic Biology 29:254–271.

Hellrigl, K. G. 1971. The bionomics of the European species of Monochamus (Coleopt. Cerambycid.) and their importance in forest and timber management (Die Bionomie der europaischen Monochamus-Arten (Coleopt., Cerambycid.) und ihre Bedeutung fur die Forst- und Holzwirtschaft). Redia 52:367–509.

Henrot, A.-J., T. Utescher, B. Erdei, M. Dury, N. Hamon, G. Ramstein, M. Krapp, N. Herold, A. Goldner, E. Favre, G. Munhoven, and L. François. 2016. Middle Miocene climate and vegetation models and their validation with proxy data. Palaeogeography, Palaeoclimatology, Palaeoecology 467:95–119.

Ho, S. Y. W., M. J. Phillips, A. Cooper, and A. J. Drummond. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Molecular Biology and Evolution 22:1561–1568.

Hopping, G. R. 1945. A new subspecies of Monochamus notatus (Coleoptera: Cerambycidae). Proceedings of the Entomological Society of British Columbia 42:17–18.

Hopping, R. 1921. A review of the genus Monochamus Serv. (Cerambycidae , Coleoptera). The Canadian Entomologist 53:252–259; plates XI, XII.

Hudson, R., and M. Turelli. 2003. Stochasticity Overrules the "Three-Times Rule": Genetic Drift, Genetic Draft, and Coalescence Times for Nuclear Loci versus Mitochondrial DNA. Evolution 57:182–190.

Janzen, D. H. 1970. Herbivores and the Number of Tree Species in Tropical Forests. The American Naturalist 104:501–528.

Janzen, D. H. 1973. Host Plants as Islands. II. Competition in Evolutionary and Contemporary Time. The American Naturalist 107:786–790.

Jones, G. 2017. Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. Journal of Mathematical Biology 74:447–467.

Jones, G., Z. Aydin, and B. Oxelman. 2015. DISSECT: An assignment-free Bayesian discovery method for species delimitation under the multispecies coalescent. Bioinformatics 31:991–998.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution 30:772–780.

Kawai, M., E. Shoda-Kagaya, T. Maehara, Z. Zhou, C. Lian, R. Iwata, A. Yamane, and T. Hogetsu. 2006. Genetic Structure of Pine Sawyer Monochamus alternatus (Coleoptera: Cerambycidae) Populations in Northeast Asia: Consequences of the Spread of Pine Wilt Disease. Environmental Entomology 35:569–579.

Kim, S., B. A. S. de Medeiros, B. K. Byun, S. Lee, J. H. Kang, B. Lee, and B. D. Farrell. 2018. West meets East: How do rainforest beetles become circum-Pacific? Evolutionary origin of Callipogon relictus and allied species (Cerambycidae: Prioninae) in the New and Old Worlds. Molecular Phylogenetics and Evolution 125:163–176.

Kitamura, A., O. Takano, H. Takata, and H. Omote. 2001. Late pliocene-early pleistocene paleoceanographic evolution of the Sea of Japan. Palaeogeography, Palaeoclimatology, Palaeoecology 172:81–98.

Koutroumpa, F. A., D. Rougon, C. Bertheau, F. Lieutier, and G. Roux-Morabito. 2013. Evolutionary relationships within European Monochamus (Coleoptera: Cerambycidae) highlight the role of altitude in species delineation. Biological Journal of the Linnean Society 109:354–376.

Kubatko, L. S., H. L. Gibbs, and E. W. Bloomquist. 2011. Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in sistrurus rattlesnakes. Systematic Biology 60:393–409.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34:772–773.

Leaché, A. D., T. Zhu, B. Rannala, and Z. Yang. 2019. The Spectre of Too Many Species. Systematic biology 68:168–181.

LeConte, J. 1852. An attempt to classify the longicorn Coleoptera of the part of America north of Mexico. Journal of the Academy of Natural Sciences Philadelphia s.2 (2):139–178.

Liebherr, J. K., and J. Schmidt. 2007. Phylogeny and biogeography of the Laurasian genus Agonum Bonelli (Coleoptera, Carabidae, Platynini). Deutsche Entomologische Zeitschrift (neue Folge) 51:151–206.

Linnen, C. R., and B. D. Farrell. 2007. Mitonuclear discordance is caused by rampant mitochondrial introgression in Neodiprion (Hymenoptera: Diprionidae) sawflies. Evolution; international journal of organic evolution 61:1417–38.

Linnen, C. R., and B. D. Farrell. 2008. Phylogenetic analysis of nuclear and mitochondrial genes reveals evolutionary relationships and mitochondrial introgression in the sertifer species group of the genus Neodiprion (Hymenoptera: Diprionidae). Molecular phylogenetics and evolution 48:240–57.

Linnen, C. R., and B. D. Farrell. 2010. A test of the sympatric host race formation hypothesis in Neodiprion (Hymenoptera: Diprionidae). Proceedings. Biological sciences / The Royal Society 277:3131–8.

Linsley, E. G., and J. A. Chemsak. 1984. The Cerambycidae of North America, Part VII, No. 1: Taxonomy and classification of the subfamily Lamiinae, Tribes Parmenini through Acanthoderini. University of California Publications in Entomology 102:i–xi, 1–258.

Liu, L., S. Wu, and L. Yu. 2015. Coalescent methods for estimating species trees from phylogenomic data. Journal of Systematics and Evolution 53:380–390.

Lobl, I., and A. Smetana, editors. 2010. Catalogue of Palaearctic Coleoptera Vol. 6. Chrysomeloidea. Apollo Books, Stenstrup, Denmark.

Maddison, D. R. 2012. Phylogeny of Bembidion and related ground beetles (Coleoptera: Carabidae: Trechinae: Bembidiini: Bembidiina). Molecular phylogenetics and evolution 63:533–76.

Maddison, W. P., and D. R. Maddison. 2017. Mesquite: a modular system for evolutionary analysis.

Makihara, H. 2004. Two new species and a new subspecies of Japanese Cerambycidae (Coleoptera). Bulletin of the Forestry and Forest Products Research Institute 3:15–24.

Mallet, J. 1995. A species definition for the modern synthesis. Trends in ecology & evolution

10:294–9.

Mayr, E. 1942. Systematics and the Origin of Species. Columbia University Press, New York.

Mckenna, D. D., A. L. Wild, K. Kanda, C. L. Bellamy, R. G. Beutel, M. S. Caterino, C. W. Farnum, D. C. Hawks, M. A. Ivie, M. L. Jameson, R. A. B. Leschen, A. E. Marvaldi, J. V. Mchugh, A. F. Newton, J. A. Robertson, M. K. Thayer, M. F. Whiting, J. F. Lawrence, A. Ślipiński, D. R. Maddison, and B. D. Farrell. 2015. The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. Systematic Entomology 40:835–880.

Mitter, C., B. D. Farrell, and D. J. Futuyma. 1991. Phylogenetic Studies of Insect-Plant interactions : Insights into the genesis of diversity. Trends in Ecology & Evolution 6:290–93.

Moore, W. S. 1995. Inferring Phylogenies From mtDNA Variation: Mitochondrial-Gene Trees Versus Nuclear-Gene Trees. Evolution 49:718–726.

Nakamine, H., and M. Takeda. 2008. Molecular phylogenetic relationships of flightless beetles belonging to the genus Mesechthistatus Breuning, (Coleoptera: Cerambycidae) inferred from mitochondrial COI gene sequences. Journal of Insect Science 8:1–11.

Ogilvie, H. A., R. R. Bouckaert, and A. J. Drummond. 2017. StarBEAST2 brings faster species tree inference and accurate estimates of substitution rates. Molecular Biology and Evolution 34:2101–2114.

Pershing, J. C., and M. J. Linit. 1985. A Structural Difference in the Male Genitalia of Monochamus carolinensis (Olivier) and M. titillator (Fabricius)(Coleoptera: Cerambycidae). Journal of the Kansas Entomological Society 58:543–546.

Peterson, M., and R. Denno. 1998. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. The American naturalist 152:428–46.

Pietsch, T. W., V. V. Bogatov, S. Y. Storozhenko, A. S. Lelej, V. Y. Barkalov, H. Takahashi, S. L. Joneson, S. K. Kholin, K. A. Glew, J. Harpel, P. Krestov, E. Makarchenko, N. Minakawa, M. Ôhara, D. J. Bennett, T. R. Anderson, R. Crawford, L. Prozorova, Y. Kuwahara, M. Shedko, M. Yabe, P. Woods, and D. E. Stevenson. 2012. Biodiversity and Biogeography of Sakhalin Island. Pages 11–78 *in* В. В. Богатов, В. Ю. Баркалов, А. С. Лелей, Е. А. Макарченко, and С. Ю. Стороженко, editors. Flora and fauna of North-West Pacific islands (Materials of International Kuril Island and International Sakhalin Island Projects). Dalnauka, Vladivostok.

Plewa, R., K. Sikora, J. M. Gutowski, T. Jaworski, G. Tarwacki, M. Tkaczyk, R. Rossa, J. Hilszczanski, G. Magoga, and Ł. Kajtoch. 2018. Morphology, genetics and Wolbachia endosymbionts support distinctiveness of Monochamus sartor sartor and M. s. urussovii (Coleoptera: Cerambycidae). Arthropod Systematics and Phylogeny 76:123–135.

Posada, D. 2008. jModelTest: Phylogenetic model averaging. Molecular Biology and Evolution

25:1253–1256.

Qiu, Y. X., C. X. Fu, and H. P. Comes. 2011. Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Molecular Phylogenetics and Evolution 59:225–244.

De Queiroz, K. 2007. Species concepts and species delimitation. Systematic biology 56:879–86.

Quek, S.-P., S. J. Davies, T. Itino, and N. E. Pierce. 2004. Codiversification in an Ant-Plant Mutualism: Stem Texture and the Evolution of Host Use in Crematogaster (Formicidae: Myrmicinae) inhabitants of Macaranga (Euphorbiaceae). Evolution 58:554–570.

Raske, A. 1973. Taxonomic relationship between Monochamus scutellatus and M. oregonensis (Coleoptera: Cerambycidae). The Canadian Entomologist 105:795–806.

Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–542.

Rossa, R., J. Goczał, and A. Tofilski. 2016. Within- and between-species variation of wing venation in genus Monochamus (Coleoptera: Cerambycidae). Journal of Insect Science 16:1–7.

Sanmartín, I., H. Enghoff, and F. Ronquist. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biological Journal of the Linnean Society 73:345–390.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

Schluter, D. 2001. Ecology and the origin of species. Trends in ecology & evolution 16:372–380.

Sequeira, A. S., B. B. Normark, and B. D. Farrell. 2000. Evolutionary assembly of the conifer fauna: distinguishing ancient from recent associations in bark beetles. Proceedings of the Royal Society B: Biological Sciences 267:2359–2366.

Sequencher® DNA sequence analysis software. (n.d.). . Gene Codes Corporation, Ann Arbor, MI.

Sláma, M. 2017. A contribution to the recognition of two Longicorn species Cerambyx cerdo Linnaeus, 1758 and Monochamus saltuarius (Gebler, 1830) (Coleoptera, Cerambycidae). Humanity space. International Almanac 6:933–938.

Sota, T., L. Bocak, and M. Hayashi. 2008. Molecular phylogeny and historical biogeography of the Holarctic wetland leaf beetle of the genus Plateumaris. Molecular Phylogenetics and Evolution 46:183–192.

Sota, T., and M. Hayashi. 2007. Comparative historical biogeography of Plateumaris leaf beetles

(Coleoptera: Chrysomelidae) in Japan: Interplay between fossil and molecular data. Journal of Biogeography 34:977–993.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. American Journal of Human Genetics 68:978–989.

Sukumaran, J., and L. L. Knowles. 2017. Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences 114:1607–1612.

Talavera, G., J. Castresana, K. Kjer, R. Page, and J. Sullivan. 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Systematic Biology 56:564–577.

Toki, W., and K. Kubota. 2010. Molecular Phylogeny Based on Mitochondrial Genes and Evolution of Host Plant use in the Long-Horned Beetle Tribe Lamiini (Coleoptera: Cerambycidae) in Japan. Environmental Entomology 39:1336–1343.

Wallin, H., M. Schroeder, and T. Kvamme. 2013. A review of the European species of Monochamus Dejean, 1821 (Coleoptera, Cerambycidae) – with a description of the genitalia characters. Norwegian Journal of Entomology 60:11–38.

Walsh, K. D., and M. J. Linit. 1985. Oviposition Biology of the Pine Sawyer, Monochamus carolinensis (Coleoptera: Cerambycidae). Annals of the Entomological Society of America 78:81–85.

Wild, A. L., and D. R. Maddison. 2008. Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. Molecular Phylogenetics and Evolution 48:877–891.

Wu, Y., N. F. Trepanowski, J. J. Molongoski, P. F. Reagel, S. W. Lingafelter, H. Nadel, S. W. Myers, and A. M. Ray. 2017. Identification of wood-boring beetles (Cerambycidae and Buprestidae) intercepted in tradeassociated solid wood packaging material using DNA barcoding and morphology. Scientific Reports 7:1–12.

Xu, B., and Z. Yang. 2016. Challenges in species tree estimation under the multispecies coalescent model. Genetics 204:1353–1368.

Yang, Z. 2014. Molecular evolution: a statistical approach. Oxford University Press, Oxford.

Yang, Z. 2015. A tutorial of BPP for species tree estimation and species delimitation. Current Zoology 61:854–865.

Yang, Z., and B. Rannala. 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences of the United States of America 107:9264–9.

CHAPTER 2

Multigene phylogeny of the Lamiini and related tribes (Coleoptera: Cerambycidae) reveals polyphyly prompting a revised tribal classification

Note: supplemental material can be found in appendix B

Introduction

The subfamily Lamiinae is the most diverse in the beetle family Cerambycidae with close to 20,000 described species (database Titan, titan.gbif.fr) distributed worldwide. The synonymous tribes Lamiini and Monochamini account for almost 10% of these species and are together one of the most species-rich tribes in the Lamiinae containing the second most generic diversity. These beetles are large, often greater than 2cm in length, and feed in the heartwood of living, dying, and recently dead trees. This feeding habit, combined with the requirement by adults for feeding on living bark (termed maturation feeding), makes these beetles economically important vectors of parasites. Moreover, clarifying the evolutionary cohesiveness of tribe-level taxa is essential for study of the possible factors leading to their high diversity. Finally, tribes that can be unambiguously diagnosed are essential for their utility in identification and classification.

The tribes of focus in the present study are the Lamiini, Monochamini, Gnomini, and Batocerini. All were established in the 1800s based on only a small proportion of the genera that are currently placed in these tribes. Today, the members of these tribes are collectively distributed worldwide and are especially diverse in tropical Asia. The individual cohesiveness of these tribes has been a topic of taxonomic research for over a century.

The Lamiini was first described at the tribe level by Mulsant (Mulsant 1839) based on genera *Lamia*, *Morimus*, and *Monochamus*. Gistel erected the Monochamini in 1848 (Gistel 1848). These initial divisions relied on very limited numbers of genera, and as available material grew so did opinions on their divisions. Through multiple revisions in the 1860s the number of

tribes under the Lamiinae varied from 6 to 35, the genera included under tribes shifted, and subdivisions changed based on divergent opinions concerning the interpretations of morphological characters that vary among and within tribes (Thomson 1860, 1864, Bates 1861). Pascoe (Pascoe 1864) gives a thoughtful discussion of the problem of characterizing higher taxa in the Lamiinae- that none of the tribes he delineates have any uniquely distinguishing features (i.e. synapomorphies) or even diagnostic series of characters. He notes that due to the variability present, genera may seem like members of a particular tribe by having some characters and lacking others and anomalous members are present in each tribe. In the end, he felt that Monochamini, Gnomini, and Batocerini should be synonymized under Lamiini while flightless genera such as *Phrissoma* should be placed in the Dorcadionini based on one of his primary characters-the lack of humeral angles (i.e., being without wings or apterous)(Pascoe 1866). At the end of the decade, Lacordaire (Lacordaire 1869) retained the separation of flightless genera into their own tribe and upheld many of the divisions (but not synonymies) made through the decade but, like Pascoe, included characters of more convincing nature such as the cicatrix of the antennal scape. The turnover in classification and associated uncertainty are understandable as limited genera were available to some authors and more informative characters were only discovered after increased study. In the end, divisions were still based on opinion as no consistent characters could be found. The Coleopterum Catalogus (Aurivillius 1922) recognized all of the tribes erected in the 1860s and introduced the Potemnemini. Acridocephalini was the only current tribe not yet separated by this date. Geographically restricted works added to the mass of genera before a worldwide revision of the Lamiinae was attempted (Breuning 1943). In this revision, Breuning combined the Agniini (including Acridocephalini), Lamiini, and Monochamini. The Phrissomini, Batocerini, Gnomini, and Ancylonotini remained as tribes. Even

after this point, possibly due to Breuning's poor reasoning and outright mistakes, synonymized tribes like Monochamini were still in use (see Dillon and Dillon 1959). A call was later issued for a worldwide treatment to address the well-known problems of Lamiinae tribal classification (Linsley and Chemsak 1984). This call has not been answered, and New World publications still use Monochamini for resident taxa. In the Old World, some work has been attempted to decrease the inflation of tribes with varied success. Sama (Sama 2008) synonymized the Dorcadionini and Phrissomini under Lamiini, trying to correct convergent character based classification. In recent works, the tribes Lamiini, Monochamini, Gnomini, Batocerini, and Dorcadionini are considered valid (Bouchard et al. 2011). Considering that taxonomic history has produced multiple tribes in the Cerambycidae that are likely not reciprocally monophyletic groups, we attempt to address some of these ambiguities by reviewing morphology in the light of new genetic data. Without any evolutionary or morphological support, there is really no purpose for erecting a tribal classification.

A tumultuous classification history and obvious discontent in the taxonomic community has instigated the need for an independent dataset to evaluate the reality of these tribes. A single phylogeny of the Monochamini has only sampled taxa from Japan with mitochondrial data (Toki and Kubota 2010). This study did sample the generic diversity present in Japan but focused on host plant use rather than classification, per se. With a new genetic character set consisting of genes that provide phylogenetic signal at genus and tribe levels, we can evaluate the monophyly of tribes and their relationships using phylogenetic estimates and statistical topology tests. This backbone will then be used to characterize diagnosable entities and make a classification representative of evolutionary history. The taxon and genetic sampling of the current study

provides a start to the necessary reorganization of the subfamily Lamiinae by tackling a large, recognizable, and economically important subset of tribes.

Methods

Taxon Sampling

The main goal of this project is to create a backbone phylogeny for the large tribes Monochamini and Lamiini as well as tribes that have a similar morphology: the Batocerini, Gnomini, Acridocephalini, Ancylonotini, and Mesosini. With the Monochamini alone containing over 250 genera (database Titan, http://titan.gbif.fr/index.html), many of which are monotypic, there was neither the opportunity nor resources available to sample exhaustively all genera. Availability of fresh tissues determined the sampling of genera across the tribes involved. An attempt was made to sample multiple genera from multigeneric tribes. Other potentially related tribes unsampled in this genetic study are the Dorcadionini, Oculariini, Xenoleini, and the former Phrissomini. A total of 53 genera and subgenera (Table 2.1) from the target tribes give a practical representation of each tribe across their geographic breadth. Outgroups came from the morphologically distant tribe Tetraopini. Sampling occurred across all continents except Antarctica. Fresh samples were collected into 95+% ethanol or frozen at cryogenic temperatures. Recently collected dried specimens were acquired from individuals across the globe. Sampling included the type genera of tribes Lamiini, Monochamini, Ancylonotini, Gnomini, Acridocephalini, and Batocerini. Scattered identification resources are needed to identify beetles from these tribes, including geographic treatments and original descriptions. Some especially helpful tools are the photographic catalog of the Cerambycidae of the world

(apps2.cdfa.ca.gov/publicApps/plant/bycidDB), the worldwide Cerambycoidea site (Cerambycoidea.com), The Lamiinae of Laos faunistic treatment (Rondon and Breuning 1970), and the Monochamini in the Western Hemisphere monograph (Dillon and Dillon 1941).

Short code	Species	Tribe (Titan Database)	locality	COI-long	WG	CAD
B065_adamitus_afr	*Monochamus (Quasiochamus) adamitus* Thomson	Monochamini	Mozambique			
DDM1110_Lamia	*Lamia textor* Linne	Lamiini	Slovakia	no barcode		
DDM694_Taeniotes_ama	*Taeniotes amazonum* Thomson	Monochamini	Costa Rica	no		
PSG100_carolinensis_FL	*Monochamus carolinensis* (Olivier)	Monochamini	USA			
PSG131_Neoptychodes	*Neoptychodes trilineatus* Linne	Monochamini	USA			
PSG39_Hebestola	*Hebestola nebulosa* Haldeman	Monochamini	USA			
PSG503_Plectrodera_NM	*Plectrodera scalator* (Fabricius)	Monochamini	USA	no barcode		
PSG504_Tlinsleyi_AZ	*Tetraopes linsleyi* Chemsak	Tetraopini	USA	no barcode		
PSG523_spectabilis_Afr	*Monochamus (Opepharus) spectabilis* (Perroud)	Monochamini	Mozambique			
PSG543_Anthores_Africa	*Monochamus (Anthores) leuconotus* (Pascoe)	Monochamini	Zimbabwe			
PSG592_subfaciatus_SK	*Monochamus subfasciatus* (Bates)	Monochamini	South Korea			
PSG594_A_malasiaca_SK	*Anoplophora malasiaca* (Thomson)	Monochamini	South Korea	no barcode		
PSG595_Acalolepta1_SK	*Acalolepta (Acalolepta) cf. fraudatrix*	Monochamini	South Korea	no barcode		
PSG596_Acalolepta2_SK	*Acalolepta (Acalolepta) cf. fraudatrix*	Monochamini	South Korea	no barcode		
PSG597_Sarothrocera_TH	*Sarothrocera lowii* White	Monochamini	Thailand	no barcode		
PSG598_Cerosterna_TH	*Cerosterna pollinosa* Buquet	Monochamini	Thailand	no barcode		
PSG599_Pharsalia_TH	*Pharsalia (Cycos) subgemmata* (Thomson)	Monochamini	Thailand			
PSG600_Epepeotes_TH	*Epepeotes luscus* (Fabricius)	Monochamini	Thailand	no barcode		
PSG601_Cremnosterna_TH	*Cremnosterna carissima* (Pascoe)	Monochamini	Thailand	no barcode		
PSG602_Aristobia_TH	*Aristobia approximator* (Thomson)	Monochamini	Thailand	no barcode		
PSG603_Acalolepta_TH	*Acalolepta sp.*	Monochamini	Thailand	no barcode		
PSG604_Anamera_TH	*Anamera cf. densemaculata* Breuning	Monochamini	Thailand	no barcode		
PSG605_Batocera_TH	*Batocera rufomaculata* (Degeer)	Batocerini	Thailand	no barcode		
PSG606_homoeus_afr	*Monochamus (Meliochamus) homoeus* Jordan	Monochamini	Cameroon			
PSG607_xfulvum_afr	*Monochamus (Laertochamus) x-fulvum* (Bates)	Monochamini	Cameroon			
PSG608_Pseudhammus_afr	*Pseudhammus myrmidonum* Kolbe	Monochamini	Cameroon			
PSG609_Oxylamia_afr	*Oxylamia fulvaster* (Jordan)	Monochamini	Cameroon			no
PSG611_notatus_NewBruns	*Monochamus notatus* (Drury)	Monochamini	Canada	no barcode		no
PSG612_c_latus_MT	*Monochamus latus* Casey	Monochamini	USA	no barcode		
PSG615_Golsinda_TH	*Golsinda basicornis* Gahan	Mesosini	Thailand	no barcode		
PSG617_Taeniotes_scal	*Taeniotes scalatus* (Gmelin)	Monochamini	Costa Rica	no		no
PSG618_Ptychodes_mix	*Ptychodes mixtus* Bates	Monochamini	Costa Rica	no barcode		
PSG620_Goes_pulv	*Goes pulverulentus* (Haldeman)	Monochamini	USA			
PSG624_nitens	*Monochamus nitens* (Bates)	Monochamini	Japan			
PSG633_Palimna	*Palimna annulata* (Olivier)	Ancylonotini	Thailand	no barcode		no
PSG634_Imantocera	*Imantocera penicillata* (Hope)	Gnomini	Thailand	no barcode		
PSG635_Apriona_swain	*Apriona swainsoni* (Hope)	Batocerini	Thailand	no barcode		
PSG636_Acalolepta_AU1	*Acalolepta sp.*	Monochamini	Australia	no barcode		no
PSG637_Acalolepta_AU2	*Acalolepta sp.*	Monochamini	Australia	no barcode		
PSG638_Paraleprodera_TH	*Paraleprodera insidiosa* (Gahan)	Monochamini	Thailand			
PSG641_cf_Epicedia	*cf. Trachystolodes*	Lamiini	Thailand	barcode only		

Table 2.1. Sampled specimens for DNA, 'no' indicates that the marker was not sequenced and completeness is # successfully sequenced/#samples. Codes indicate GenBank sequences and gray shading indicates USDA barcode sequences in Genbank were obtained from the OTIS lab in Buzzard's Bay, MA

PSG642_Acalolepta_TH	*Acalolepta (Acalolepta) cf. cervina* (Hope)	Monochamini	Thailand	no barcode		no
PSG656_sutor_IT	*Monochamus sutor* (Linnaeus)	Monochamini	Italy	CA13_18.02		
PSG682_alternatus_CH	*Monochamus alternatus* Hope	Monochamini	China	CA12_3.01		
PSG689_Batocera_CH	*Batocera lineolata* Chevrolat	Batocerini	China	TX13_10.01		
PSG694_gallo_TK	*Monochamus galloprovincialis* (Olivier)	Monochamini	Turkey	CA14_32.01		
PSG714_Goes_tig	*Goes tigrinius* (Degeer)	Monochamini	USA			
PSG716_Pseudhammus_myr	*Pseudhammus (P.) myrmidonum* Kolbe	Monochamini	Democratic Republic of Congo	no		
PSG718_M_M_olivaceus	*Monochamus (M.) olivaceus* Breuning	Monochamini	Democratic Republic of Congo	barcode only		
PSG725_Psacothea	*Psacothea hilaris* (Pascoe)	Monochamini	North Korea	no barcode		no
PSG756_Goes_debilis	*Goes debilis* LeConte	Monochamini	USA	barcode only		no
PSG758_Agnia_Phil	*Agnia casta* Newman	Monochamini	Philippines			no
PSG759_A_asuanga_Phil	*Anoplophora asuanga* Schultze	Monochamini	Philippines	no barcode		
PSG614_Blepephaeus_TH	*Blepephaeus succinctor* (Chevrolat)	Monochamini	Thailand	no barcode		
PSG760_Blepephaeus_Phil	*Blepephaeus mindanaoensis* (Schultze)	Monochamini	Philippines	no barcode		
PSG763_Epepeotes_Phil	*Epepeotes plorator* (Newman)	Monochamini	Philippines	no barcode		no
PSG764_Gnoma_Phil	*Gnoma luzonicum* Erichson	Gnomini	Philippines			no
PSG765_Pelargoderus_Phil	*Pelargoderus luzonicus* (Breuning)	Monochamini	Philippines			no
PSG766_Paranhammus_Phil	*Paranhammus marcipor* (Newman)	Monochamini	Philippines			
PSG767_Nemophas_Phil	*Nemophas subterrubens* Heller	Monochamini	Philippines			
PSG769_Cyriotasiastes_Phil	*Cyriotasiastes rhetenor* (Newman)	Monochamini	Philippines	no barcode		
PSG771_Acalolepta_lux	*Acalolepta (Acalolepta) luxuriosa* (Bates)	Monochamini	China	bar only (WA13_20.1)		no
PSG774_Plagiohammus_thor	*Hammatoderus thoracicus* (White)	Monochamini	Costa Rica	no barcode		
PSG775_Taeniotes_prae	*Taeniotes cf. praeclarus* Bates	Monochamini	Costa Rica	no barcode		
PSG776_Neoptychodes_cret	*Neoptychodes cretatus* (Bates)	Monochamini	Costa Rica	no barcode		
PSG777_Plagio_eman	*Hammatoderus emanon* (Dillon & Dillon)	Monochamini	Costa Rica	no barcode		
PSG778_Arctolamia	*Arctolamia fruhstoferi* Aurivillius	Lamiini	Vietnam	no barcode		no
PSG786_Doesburgia	*Doesburgia celebiana* Tippmann	Batocerini	Indonesia	no barcode		
PSG787_Idactus	*Idactus cristulatus* (Fairmaire)	Ancylonotini	Mozambique	no barcode		
PSG788_Laziopezus	*Lasiopezus longimanus* (Thomson)	Ancylonotini	Mozambique	no barcode		no
PSG789_Ancylonotus	*Ancylonotus tribulus* (Fabricius)	Ancylonotini	Mozambique	no barcode		no
PSG79_sscutellatus_MI	*Monochamus scutellatus* (Say)	Monochamini	USA			
PSG798_Pharsalia_ant	*Pharsalia (Antennopharsalia) antennata* Gahan	Monochamini	Vietnam			no
PSG807_Stegenagapanthia	*Stegenagapanthia albovittata* Pic	Monochamini	Vietnam	barcode only		no
PSG817_Acridocephala	*Acridocephala seriata* Jordan	Acridocephalini	Cameroon	no		
PSG819_Ethiopiochamus_scabiosus	*Monochamus (Ethiopiochamus) scabiosus* (Quedenfeldt)	Monochamini	Cameroon	no barcode		
PSG84_Microgoes	*Microgoes oculatus* (LeConte)	Monochamini	USA			
PSG588_guttulatus_SK	*Monochamus guttulatus* Gressitt	Monochamini	South Korea		no	
PSG757_Acthophora_Phil	*Acthophora tristis* Newman	Monochamini	Philippines	barcode only	no	no
PSG779_Blepephaeopsis	*Blepephaeopsis vietnamensis* Hayashi	Monochamini	Vietnam	no barcode	no	no
T. tetrophthalmus	*Tetraopes tetrophthalmus* (Forster)	Tetraopini	USA	AF267478	KP813603	KP813042
Completeness				0.94	0.96	0.75

Table 2.1 (Continued)

Sequence data

This study applied Sanger-based sequencing techniques and followed the PCR, sequencing and processing methodology of Chapter 1. The markers explored were mitochondrial COI (~1468bp), Nuclear CAD (~943bp) and Nuclear *wg* (~441bp). A total aligned matrix of ~2852 DNA characters was used for concatenated analyses. The *wg* fragment has two amino acid indels. The genus *Blepephaeus* has an exclusive single AA insertion, and there is an AA deletion present in all taxa except the Tetraopini, Mesosini, and Ancylonotini. CAD and COI show no indels.

Phylogenetic analyses

Concatenated phylogenetic analyses were conducted using both Bayesian and Maximum Likelihood approaches. To find optimal data partitions the concatenated dataset was analyzed in *PartitionFinder* v2.1.1 (Lanfear et al. 2016) using unlinked branch lengths, the greedy search algorithm, and AICc as the selection metric. The ideal partitioning scheme has three partitions: 1) GTR+G+I for all first and second positions 2) TRN+I+G for CAD and *wg* third positions 3) GTR+G for COI third positions. When restricting to the models available in *MrBayes* for single gene runs, first and second partitions get model GTR+G+I and partition three GTR+G. For the concatenated run, each partition was run under GTR+G, since a gamma model can account for invariant sites (Yang 2014). *RAxML* was run using GTRGAMMA for the same three partitions given the software's restriction to one model.

The Bayesian analysis was run using *MrBayes* v. 3.2 (Ronquist et al. 2012). Single gene runs were encapsulated as two runs of four chains (one cold) for 20-50 million generations with sampling every 1000 generations. Stationarity and convergence were confirmed by deviation of

split frequencies < .01, potential scale reduction factor values ~ 1.00, and effective sample size (ESS) >200 as measured in *Tracer* v. 1.6 (http://beast.community/tracer). The consensus tree was summarized using sumt command in *MrBayes* with a burnin of 25%. The concatenated run was the same except spread over eight cores for a total of 400 million generations over eight 50 million generation runs on the Harvard 'Odyssey' computing cluster (www.rc.fas.harvard.edu/odyssey). These runs were combined using the mcmcp command in *MrBayes*. Posterior probabilities are used as support values.

A maximum likelihood tree was estimated using *RAxML* v. 8.2.11 (Stamatakis 2014) on the Harvard computing cluster. The threaded version was used with eight cores and -N 1000 (-f a, -x) to run a rapid bootstrap and search for the best-scoring ML tree of 1000 independent starts. *RAxML* was run with the optimal *PartitionFinder* partitions under the GTRGAMMA nucleotide model, and the bootstrap score is used to indicate confidence.

Topology testing

When support values are low on a phylogenetic tree, placement of clades can be misleading and should be confirmed statistically before drawing any conclusions. To this end, we used the Swofford–Olsen–Waddell–Hillis test implemented in *SOWHAT* (Church et al. 2015). This test compares the log-likelihood (lnL) difference of two topologies to a null distribution of log-likelihood differences produced through simulation of data under the same model. To prepare competing topologies we manually created constraint trees where one node is constrained to test clade placement. Each *SOWHAT* analysis was run 100 repetitions, with the GTRGAMMA model and the optimal *PartitionFinder* partitioning scheme in RAxML, on 12-20 Intel cores on Harvard University's Odyssey cluster. The output of each analysis is a p-value

based on 100 simulation repetitions calculated by dividing the number of simulated lnL differences greater than or equal to the empirical lnL difference between the unconstrained and constrained trees. A confidence interval falling within the significance level ($p<.05$) indicates that the sample size is sufficient.

Six separate *SOWHAT* analyses were completed in order to explicitly investigate the placement of tribes within and near the Lamiini. The native *RAxML* analysis places Monochamini+Lamiini+Gnomini+Batocerini+Acridocephalini within the same clade. Non-Lamiini taxa that grouped confidently with those assigned to Lamiini were made members of that tribe for this analysis to reduce conflict, irrespective of online database assignments. The constraints are as follows (Fig. 2.1): constrained samples of Monochamini sensu the Titan database (except the samples moved to Lamiini); Gnomini constrained outside of a clade including Monochamini, Lamiini, Batocerini, and Acridocephalini; Batocerini constrained outside Monochamini+Lamiini+Gnomini+Acridocephalini; constrained Monochamini+Acridocephalini; constrained Monochamini+Lamiini+Acridocephalini; constrained Ancylotonini+Monochamini+Lamiini+Batocerini+Acridocephalini. There is no justification given in the literature for elevating Acridocephalidi from subtribe to tribe level, and it should have been synonymized with other subtribes under Monochamini. We also constrained the conifer feeding *Monochamus* species to explore the placement of the clade including *Goes* + *Hebestola*.

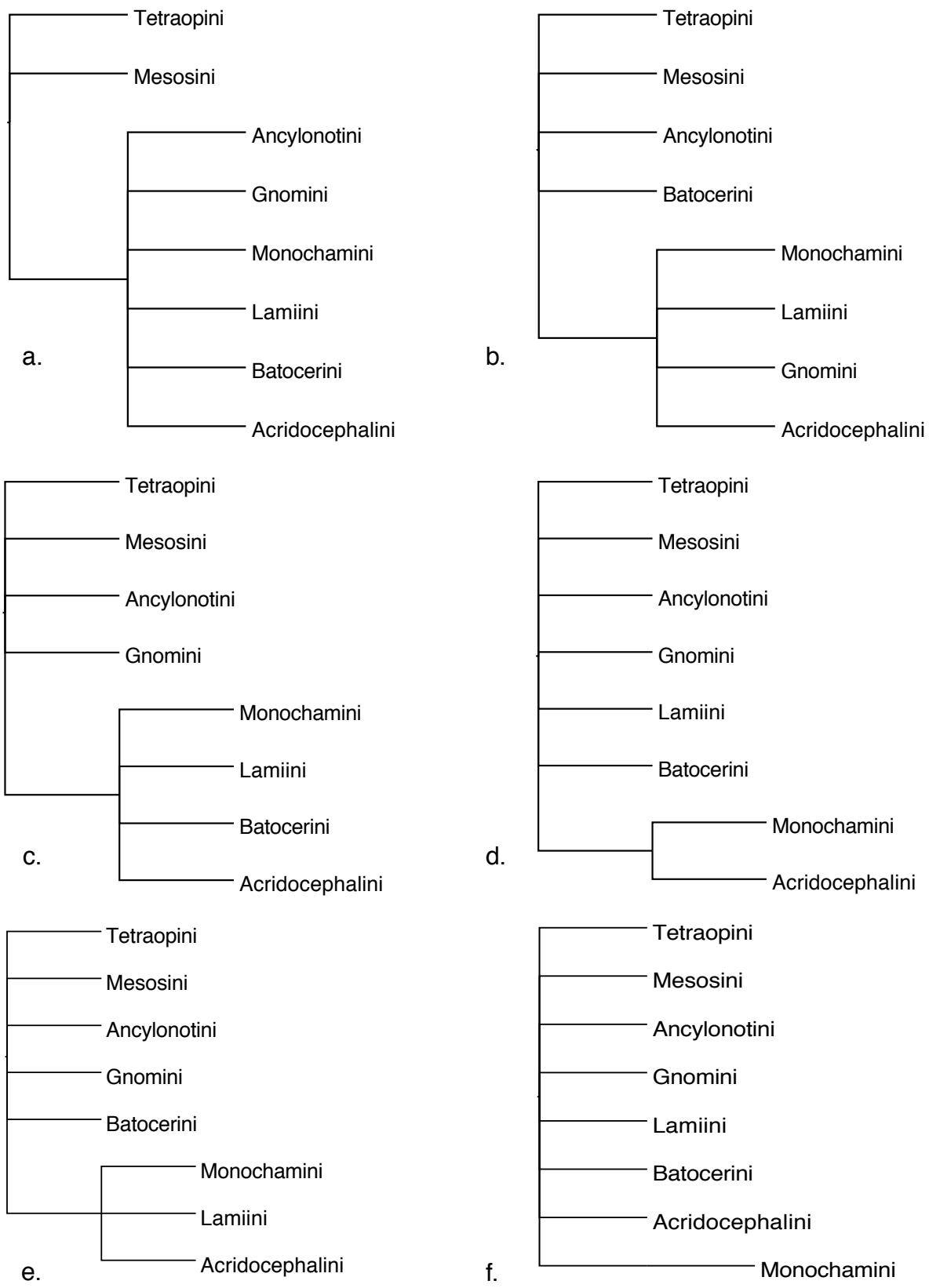

Figure 2.1. SOWHAT topology constraints (a)Ancylonotini in (b) Batocerini out (c) Gnomini out (d) Monochamini+Acridocephalini in (e) Monochamini+Acridocephalini+Lamiini in (f) Monochamini monophyletic

Morphology

The main goal of a classification is to "serve as an efficient information storage and retrieval system" (Mayr 1969). Classifications of organisms store information about relatedness, as estimated from the distributions of the characteristics of samples of species. A goal of modern classification is to recognize monophyletic groups which then lend themselves to study of the historical mechanisms that have led to the diversity of species. The information reflected in classifications is useful also for identifying the groups to which organisms belong, from species up through the taxonomic ranks. Even with the spread of molecular tools the most efficient way to identify organisms is still through morphological study. Comparative morphology is also the only way to connect fossil samples to living taxa and into phylogenetic studies of relatedness. Several morphological characters have been used in the past to define the tribes of this study. We will evaluate some of the most promising traits to determine if they have any utility in correspondence with the molecular phylogeny. Table 2.2 lists these characters and their states. We mapped these morphological characters on the Bayesian phylogeny to visualize character utility in diagnosing monophyletic groups. To further investigate the variability of these characters, the character states in 28 genera in the Monochamini and six genera of the Lamiini were coded for summary purposes in addition to the genera represented by molecular data.

Character	**States**
tarsomere 4+5	fused or unfused
prosternum length	transverse/subquadrate or longer than wide
scape cicatrix	ridged open; ridged closed; simple granulate; absent
mesotibial furrow	present or not present (only dense setae)
mesocoxal cavities	open to touch epimeron or closed from epimeron
lateral pronotal spines	present or not present

Table 2.2. Characters reviewed for tribal signal and their states in studied tribes

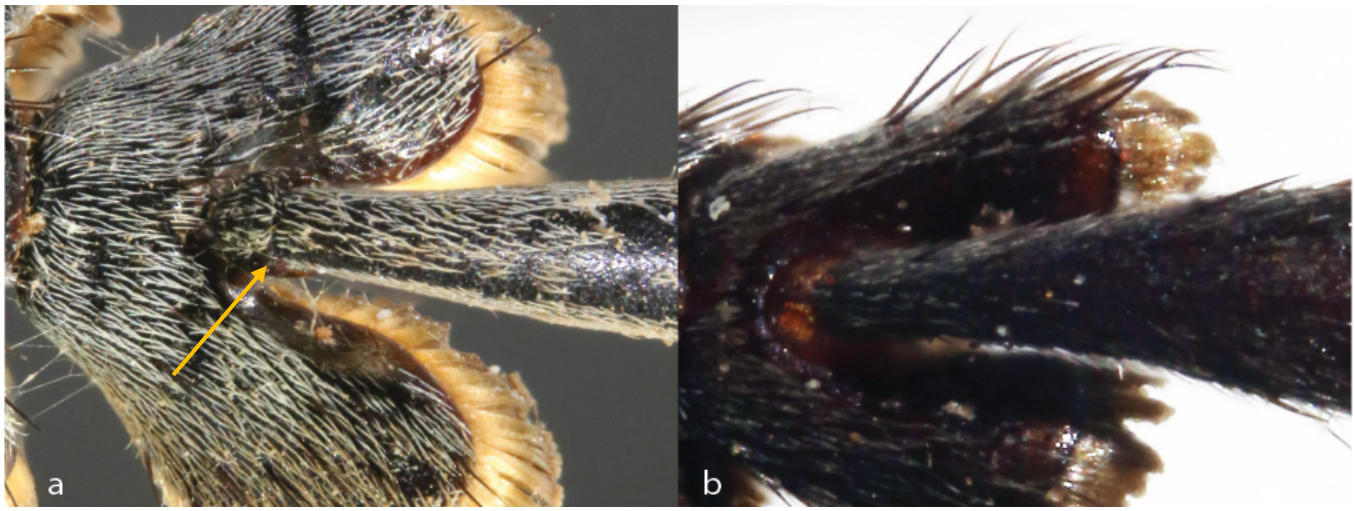

Figure 2.2. Unfused and fused tarsomere 4+5. a) unfused (*Batocera*); b) fused (*Monochamus*)

Figure 2.3. Variation in the cicatrix of the scape. a) *Ancylonotus*; b) *Mesosa*; c) *Batocera*; d) *Abatocera*; e) *Paraepepeotes* (Monochamini)

Figure 2.4. Tribal morphology: mesotibial furrow (*Monochamus*); b) elongate prosternum (*Macrochenus*); c) epipleuron open to mesocoxal cavity, the arrow indicates coxal cavity opening (*Batocera*); d) lateral pronotal spine (*Batocera*); e) lateral pronotal armature absent (*Mesosa*)

Results

Phylogenetic estimation

The Bayesian supermatrix tree was well resolved overall with the majority of branches supported at pp>.9 (Fig. 2.5). The individual gene trees (Figs. 2.7,B1,B2) show contribution of signal at different tree levels indicating varied evolutionary rates and congruence without much much conflict towards the backbone of the tree. The tribal backbone tree agreed across Bayesian and ML analyses (Figs. 2.5,2.6). As traditionally defined, the tribes Batocerini, Gnomini, and Ancylonotini were found to be monophyletic (pp=1, MLB=99-100). The tribes Lamiini and Monochamini, as defined in the Titan database, were not found to be monophyletic. Members of the Gnomini and Monochamini split the samples of the Lamiini, revealing the Lamiini to be polyphyletic. Monophyly of the Mesosini could not be determined with the limited sampling, and Acridocephalini only contains one genus. The *Acalolepta* clade formed the most basal group in the Bayesian phylogeny (Fig. 2.5). It is worth noting that all of the South and Central American genera included in the tree formed a monophyletic grouping (pp=1, MLB=100).

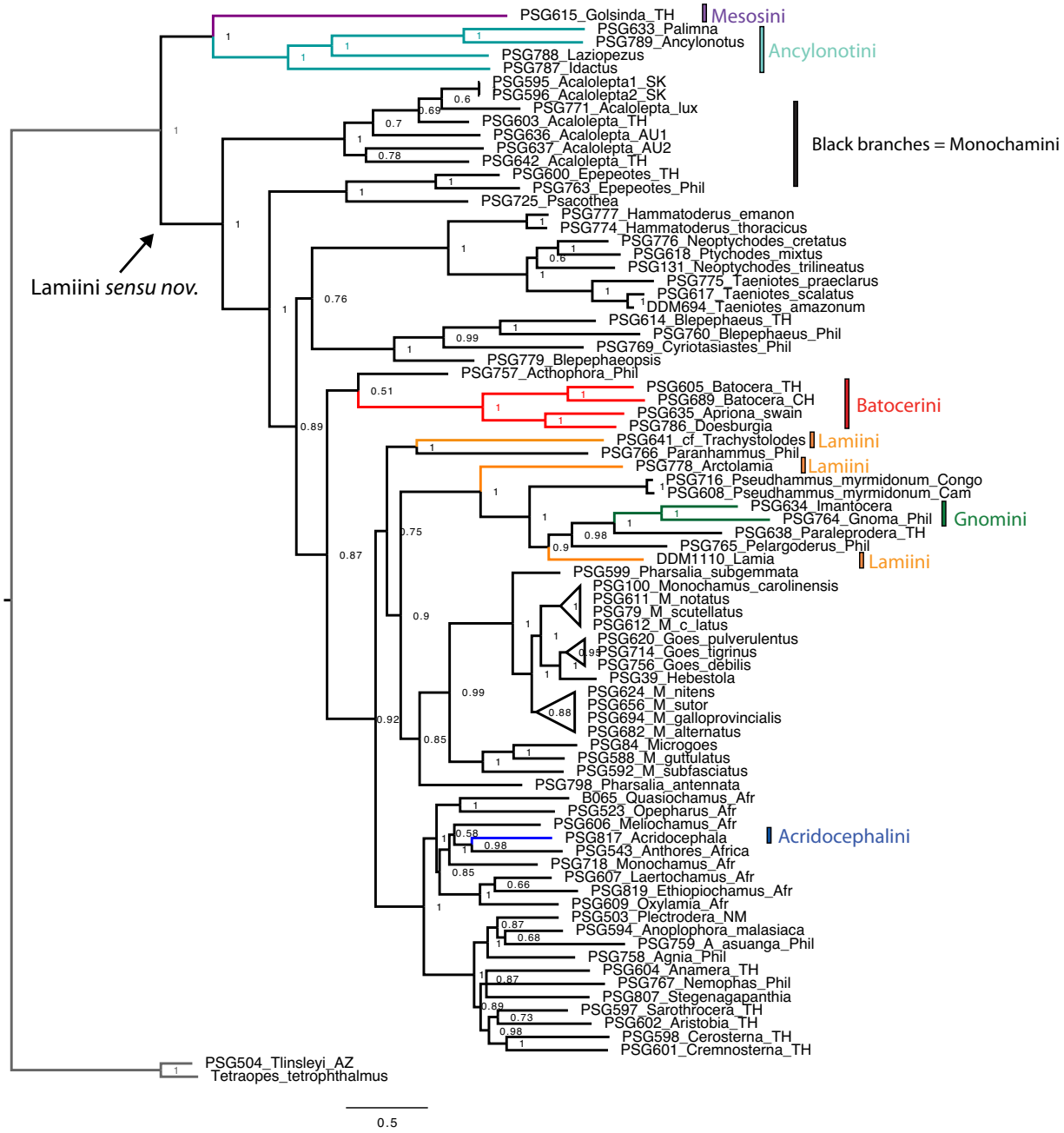

Figure 2.5. *MrBayes* supermatrix tree, colored by tribal affiliation, support values posterior probability

Figure 2.6. *RAxML* supermatrix tree, colored by tribe assignment. Bootstrap values on branches

The sampling within genera was not dense enough for most taxa to draw firm conclusions, but some preliminary results indicate some possibly problematic relationships. The six species sampled from the genus *Acalolepta* from the continents of Asia and Australia combined as a monophyletic group without separation by continent. Two subgenera of *Pharsalia*, *Cycos* and *Antennopharsalia* were found to be separated by the genus *Microgoes*. The new world genus *Ptychodes* rendered *Neoptychodes* paraphyletic. Within the genus *Pseudhammus*, the *wg* gene tree shows the subgenus *Pseudhammus* related to *Lamia* while the subgenus *Litigiosus* was closer genetically to other African genera (Fig. 2.7). Finally, the genus *Monochamus* is rendered paraphyletic by a clade including the genera *Goes* and *Hebestola* from the Nearctic region.

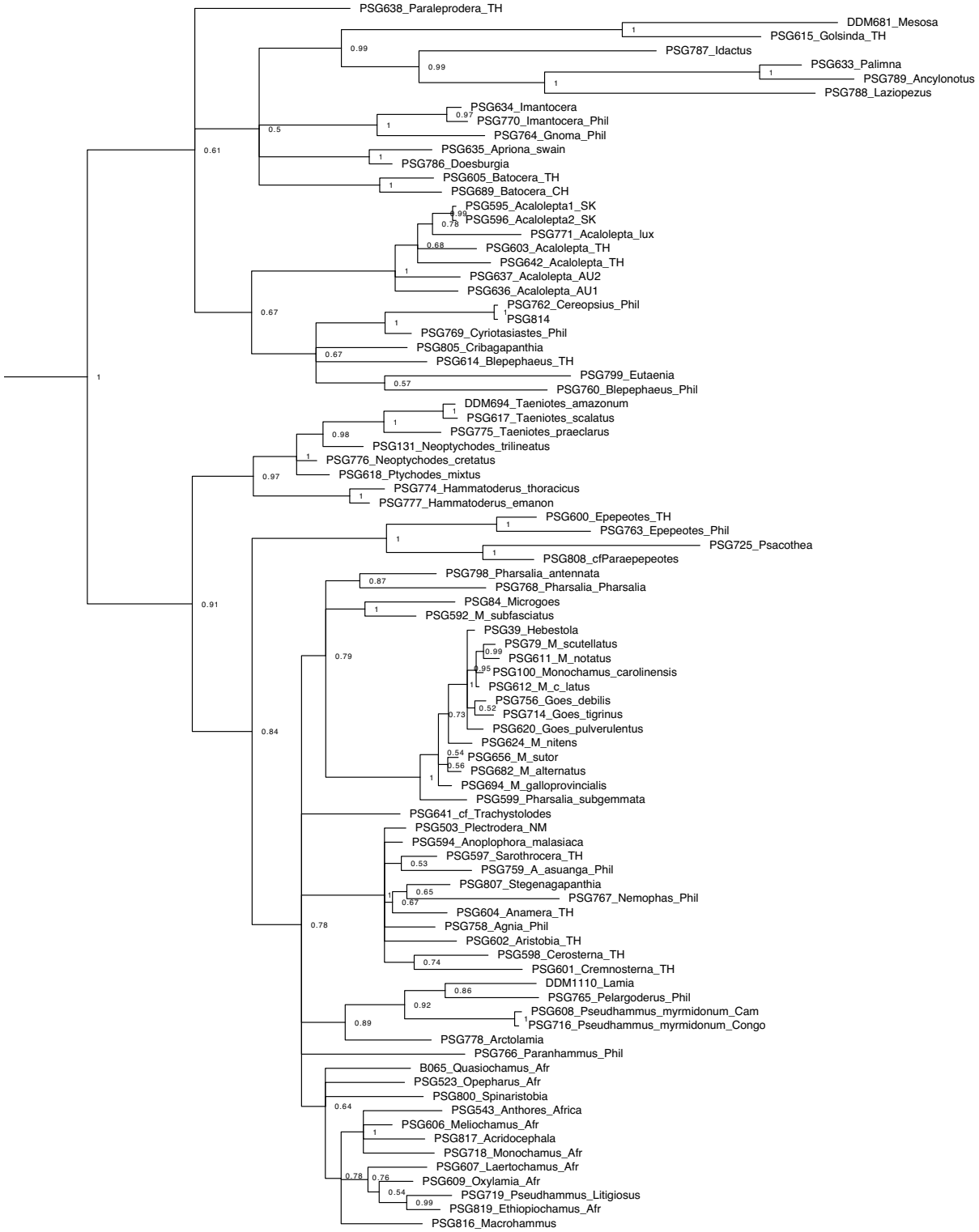

Figure 2.7. *MrBayes wg* gene tree with posterior probability values

Topology testing

The unconstrained maximum likelihood tree determined using the *SOWHAT* software matches the *RAxML* tree topology presented in Figure 2.6. The unconstrained tree was supported with significance in each constraint analysis (Table 2.3). One constraint performed within a clade was constraining the conifer feeding *Monochamus* monophyletic, excluding *Goes* + *Hebestola*. This analysis also resulted in significant support for the unconstrained topology: p<.01, 95% CI [0,.036], likelihood difference 102.61.

Analysis	constraint figure	pvalue [95% CI]	test stat (empirical lnL)
Constrain Ancylonotini inside	a	<.01 [0,.036]	96.61
Constrain Batocerini outside	b	<.01 [0,.036]	129.48
Constrain Gnomini outside	c	<.01 [0,.036]	226.06
Monochamini monophyletic (with Acridocephalini)	d	<.01 [0,.036]	162.71
Monochamini w/only Lamiini inside	e	<.01 [0,.036]	280.74
Monochamini monophyletic	f	<.01 [0,.036]	274.97

Table 2.3. SOWHAT testing results, the test stat is the difference in lnL between the unconstrained and constrained ML trees

Morphology

None of the characters explored can be recognized as synapomorphies for Lamiini *sensu novo* or any of the previously recognized tribes that fall within Lamiini. Characters such as fused tarsomeres (Fig. 2.8), a mesotibial furrow, and an elongated pronotum are present in multiple lineages and are variable within tribes (Table 2.4). Many of these characters are also plesiomorphic and present in the outgroup tribes Mesosini and Ancylonotini. The rimmed cicatrix on a cylindrical scape is present in most, but not all, of the taxa in Lamiini sensu nov., and can be greatly reduced or absent in species i.e. *Trachystola* (Lamiini) and *Abatocera* (Batocerini). Lateral pronotal spines arising from the middle of the pronotum are also present in almost all taxa in the Lamiini *sensu nov*., while some genera including *Hebestola, Ptychodes*, and *Macrochenus* do not have them.

Tribe	fused tarsomere 4+5	elongate prosternum	cicatrix	mesotibial furrow	mesocoxal cavities	lateral pronotal spines
Lamiini	yes	no	present, variable in form	yes	open	yes
Monochamini	variable	variable	present, variable in form	variable	open	variable
Batocerini	no	no	variable	yes	open	yes
Gnomini	yes	variable	present, ridged	yes	open	variable
Acridocephalini	yes	no	present, ridged	yes	open	yes
Ancylonotini	no	no	present, ridged and granulate	variable	open	variable
Mesosini	no	no	present, prominent expanded lateral ridge	variable	open	no

Table 2.4. Presence of discussed morphology in former and current tribes. Summary based on genera from the molecular analysis and additional museum specimens.

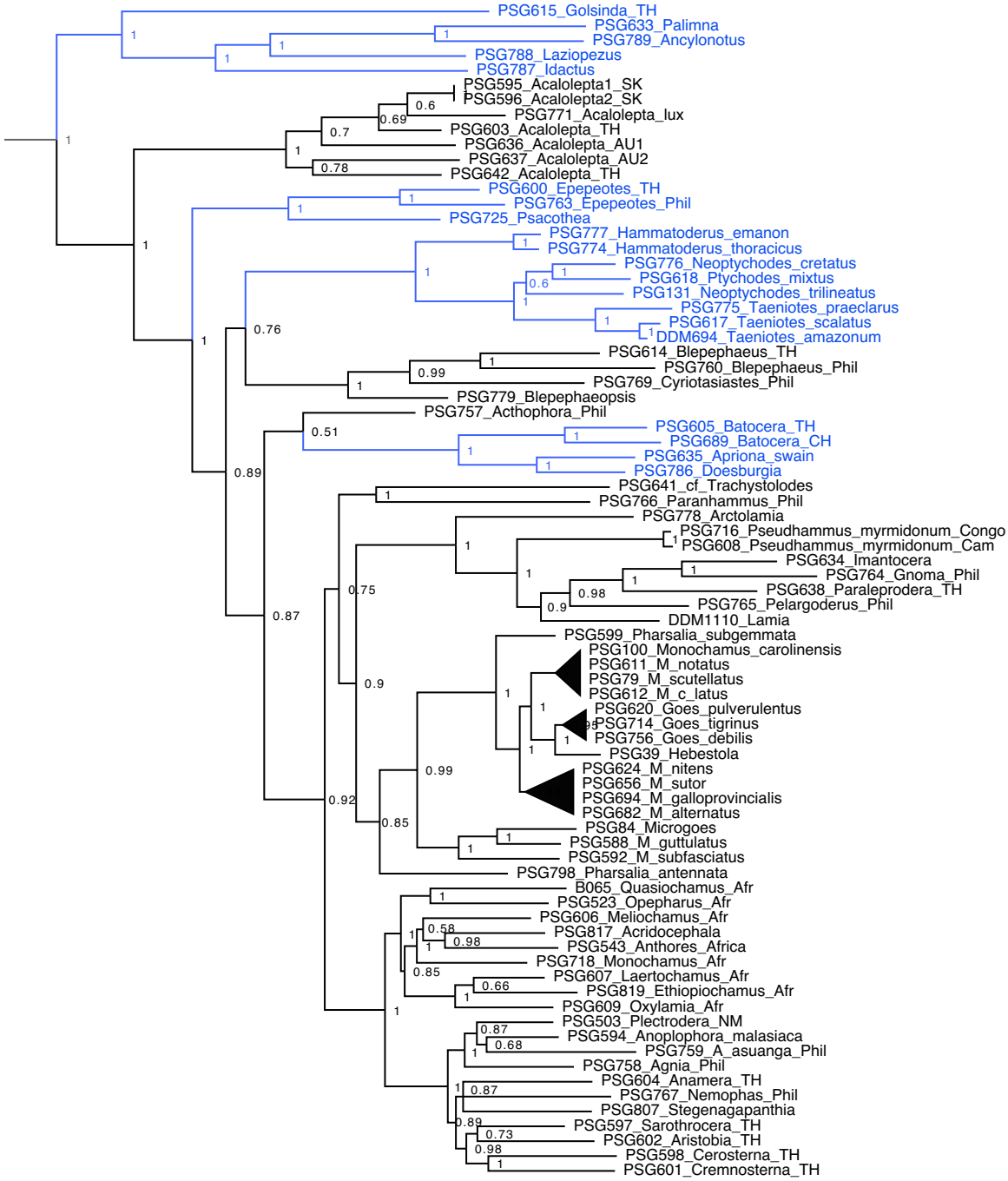

Figure 2.8. *MrBayes* supermatrix phylogeny, branches colored blue indicate unfused tarsomere 4+5

Discussion

We present the first study to sample across the full geographic range of the Lamiini and related tribes. The only other phylogenetic study of this group focused on the fauna of Japan, with few representatives outside of the Monochamini and was rooted using the Batocerini as an outgroup (Toki and Kubota 2010). In that tree, the Monochamini was not monophyletic, and no other phylogenetic studies have explored the higher level relationships of these groups. With limited polytomy formation, topological congruence, and high support along the Bayesian backbone the markers we have used seem to combine to give an informative signal for this higher level analysis. The topology does agree in the monophyly of certain clades previously considered as tribes but in positions that would cause a drastic increase in the number of tribes if they were to remain at this taxonomic rank. Nevertheless, despite strong support for the backbone topology in this analysis, there is a limited sampling of genes and taxa. Increasing the number of independent markers can provide more characters and also allow for advanced averaging over gene trees in a multispecies coalescent framework (Xu and Yang 2016). The inclusion of more taxa would allow a view of how the genera in these tribes group and allow for the determination of possible higher level structure of this very large clade. Increased taxon sampling has the potential to break long branches and lend an overall improved phylogenetic inference (Hedtke et al. 2006) but can also confound relationships at the species tree level because the addition of more taxa often means shorter internal branches (Degnan and Rosenberg 2009). Finally, though morphological traits traditionally thought to separate these tribes were analyzed, none of these characters were found to be synapomorphies for any of the tribes or of

Lamiini *sensu novo*. A full morphological investigation should be performed, including internal and larval anatomy to determine if diagnostic characters can be discovered.

Tribal monophyly and generic evaluation

This is the first study to reveal the suspected polyphyly in the largest of these tribes, the Monochamini. We find that the tribes Batocerini, Gnomini, Acridocephalini are monophyletic within a broader tree of mixed Monochamini and Lamiini taxa (Figs 2.5,2.6). Since the tribe Lamiini takes nomenclatural precedence, the other tribes should be synonymized under Lamiini *sensu nov.* An earlier revision (Breuning 1943) accomplished some of these same changes, synonymizing tribes Lamiini and Monochamini under Agniini, but was not widely followed because Breuning failed to respect the nomenclatural priorities of the names affected. The Monochamini and Lamiini have been treated as separate tribes after this publication (Dillon and Dillon 1959, Linsley and Chemsak 1984). There are flightless lamiine tribes that share characteristics with Lamiini *sensu nov.* but no tissues were available for this study. These are the Phrissomini Thomson, 1860 and Dorcadionini Swainson, 1840. These tribes have been synonymized under Lamiini (Sama 2008), yet Dorcadionini is still recognized as a taxon (Bouchard et al. 2011). Due to the similarity in morphology with genera in the Lamiini and characters specific to a flightless lifestyle being the reasoning for exclusion, the tribe Phrissomini should remain synonymized under Lamiini *sensu nov*. pending further evidence. The Dorcadionini was not included in the present study with either molecular or morphological approaches, but it does show similar characters to the Lamiini (fused tarsi, pronotal spines, some have a cicatrix) so it may belong within or close to the Lamiini. The Xenicotelini, Oculariini, and Xenoleini are additional tribes that could fit in the Lamiini with their own synapomorphies but

this could not be tested due to a lack of available tissues. Proposed tribal synonomies are presented in the classification section of appendix B.

Topology testing can bolster the hypothesis that the tribes synonymized under Lamiini sensu novo are properly placed and not a product of weaknesses of the data. Constraint trees were created to test existing hypotheses of tribal placement (Fig. 2.1). The tribes Lamiini and Monochamini have been proposed as synonymous in past publications (Breuning 1943, 1961) and online lists. Simply constraining the members of these tribes as monophyletic results in a significant likelihood difference supporting the unconstrained tree with other tribes included (Table 2.3). Each of the other analyses, either constraining tribes outside or inside of Lamiini *sensu novo*, gave similar results in support of the unconstrained *RAxML* topology (Fig. 2.6). All sowh tests of alternative topologies were rejected with a p-value of <0.01. Low p-values indicate that the ML tree estimate is extremely robust. The monophyly of the conifer feeding *Monochamus* species was found uncertain in a study focused on the genus (Gorring, unpub.). We tested the placement of the clade including *Goes* and *Hebestola* which that study found as sister to conifer feeding *Monochamus* or rendering it paraphyletic. The topology testing showed significant support for the *Goes* clade as sister to North American *Monochamus* species. This supports a biogeographic origin of the ancestor of North American *Monochamus* and *Goes* clades in Asia and subsequent diversification on both conifers and angiosperms, respectively, of these genera in North America.

The discussion of several genera follows from their placement in the phylogeny. The genus *Acalolepta* is one of the largest in the Lamiini with over 250 species. With samples from Australia, Thailand, and Korea we find the genus monophyletic and not separated by geography. There are three *Acalolepta* subgenera, but we could not evaluate their validity by using only

samples from the subgenus *Acalolepta*. There is potential for *Acalolepta* to be a paraphyletic genus, with the genus *Mimorsidis* rendering it paraphyletic in the Lamiini tree of Japan (Toki and Kubota 2010). *Mimorsidis yayeyamensis*, the species used in that study, was also described as *Acalolepta ishigakiana* by S. Breuning, the author that erected *Mimorsidis* (Breuning and Villiers 1973). It has a similar morphology so may belong to *Acalolepta*. This data also shows that the characters used by lamiine taxonomists in the past to identify subgenera better represent the genus level. Two subgenera of *Pharsalia* were included: *Cycos* and *Antennopharsalia*. These do not form a monophyletic group and are separated by the genus *Microgoes*. Sampling including the other subgenera could reveal if some form monophyletic groups as is shown between the subgenera *Pharsalia* and *Antennopharsalia* in the *wg* gene tree. Either way, *Cycos* should be elevated to genus. The subgenera sampled from the African genus *Pseudhammus*, *Pseudhammus* and *Litigiosus*, show a similar paraphyletic pattern in the *wg* tree with *Pseudhammus* grouping with the European genus *Lamia* and *Litigiosus* grouping with other African genera in the tree. The most extreme example is in the subgenera of *Monochamus*. These are mostly present in Africa, and the phylogenetic analyses have them polyphyletic in a clade with other African genera. The subgenus *Monochamus*, which is geographically spread through Africa, Asia, and North America, is divided into four groups in the phylogeny: Nearctic *Monochamus*, Palaearctic conifer feeding *Monochamus*, part of the *Microgoes* clade, and an African branch (represented by *Monochamus (M.) olivaceus* (PSG718). In the end, only the conifer feeding species represent true *Monochamus,* and they seem to be paraphyletic. The issue would benefit from more gene sampling since concatenated trees with more genes find conifer feeding *Monochamus* monophyletic while this study's concatenated three-gene analysis and topology tests find their group to be paraphyletic. A species tree analysis over many markers should help to determine the

universal signal of the genome and if incomplete lineage sorting is possibly associated with rapid diversification following entry into North America (Liu et al. 2015). Another relationship to note is that of the new world genus *Ptychodes* which renders *Neoptychodes* paraphyletic. These two genera are similar morphologically and deserve further sampling to clarify relationships. When thinking about expanded taxon sampling for future work, it would be advisable to sample all subgenera and multiple samples from any morphologically diverse genus. A distinct clade could also be explored in depth to re-evaluate the morphological characters that seem informative at the genus level for guidance in taxon choice for an expanded study.

Morphology

Comparative morphology is the foundation of the science of taxonomy and classification. While many of the characters used to define the tribes that we now consider synonyms of the Lamiini were insufficient to delineate those taxa, some can be useful to describe many of the genera within the tribe's new boundaries. One character rarely used in the past is the fusion of tarsomeres four and five. These beetles are essentially tetramerous. This character is found in most of the tribes proposed to be within Lamiini *sensu nov*. and is also found in the Dorcadionini and former Phrissomini. This fusion is not present in the observed genera of the former Batocerini. It is found to be variable in the Monochamini, not present in some Monochamini genera in the molecular phylogeny estimate (Fig. 2.8), and also not observed in *Macrochenus*, *Paraepepeotes*, and *Gerania* of the additional Monochamini genera coded. Interestingly, all South American Lamiini genera share the plesiomorphic unfused state with the Mesosini and Ancylonotini. These patterns hold for the genera observed but we did not have available for this study all of the genera or species within genera, and so fusion absence in certain tribes and

apparent constancy in others could change. Regardless of the reversals, this is a very strong character since beetle tarsal formula changes are often restricted to the family level. Two characters that are also significant in the Lamiini *sensu nov*. are lateral pronotal tubercles and a cicatrix on the scape. With several exceptions, the genera included have lateral pronotal projections that are often acute. The Mesosini often have small anterolateral projections, and the Ancylonotini are variable in their pronotal armature. The cicatrix, or scar, at the apex of the scape is present in almost all species observed in the revised Lamiini. The level of development does differ from a lightly granulate apex as in *Rosenbergia* to a heavily rimmed and closed cicatrix. This is not the only tribe with a cicatrix; some tribes are variable while others like Mesosini have a consistent shape. Overall, no morphological diagnostic characters are constant for the Lamiini, as has been discussed as far back as the 1860s (Bates 1861, Pascoe 1864). With the combination of these tribes, much of the ambiguity with specimen assignment to tribe will be relieved with a combination of characters. These are fused tarsomeres 4+5, lateral mid-pronotal projections, and a cicatrix on the scape. Fortunately, while there are anomalous genera that are exceptions for each of these characters, most tend towards a distinctive elongate body shape and this combination of characters. Currently outside of the Lamiini *sensu nov.,* the tribe Dorcadionini also has fused tarsomeres and pending future study, the Xenoleini, Oculariini, and Xenicotelini may as well. This seems to be one of the most useful characters for placement of genera within Lamiini and further morphological and molecular investigation can corroborate or falsify this hypothesis. A full morphological phylogenetic analysis will also potentially reveal other characters with strong phylogenetic signal.

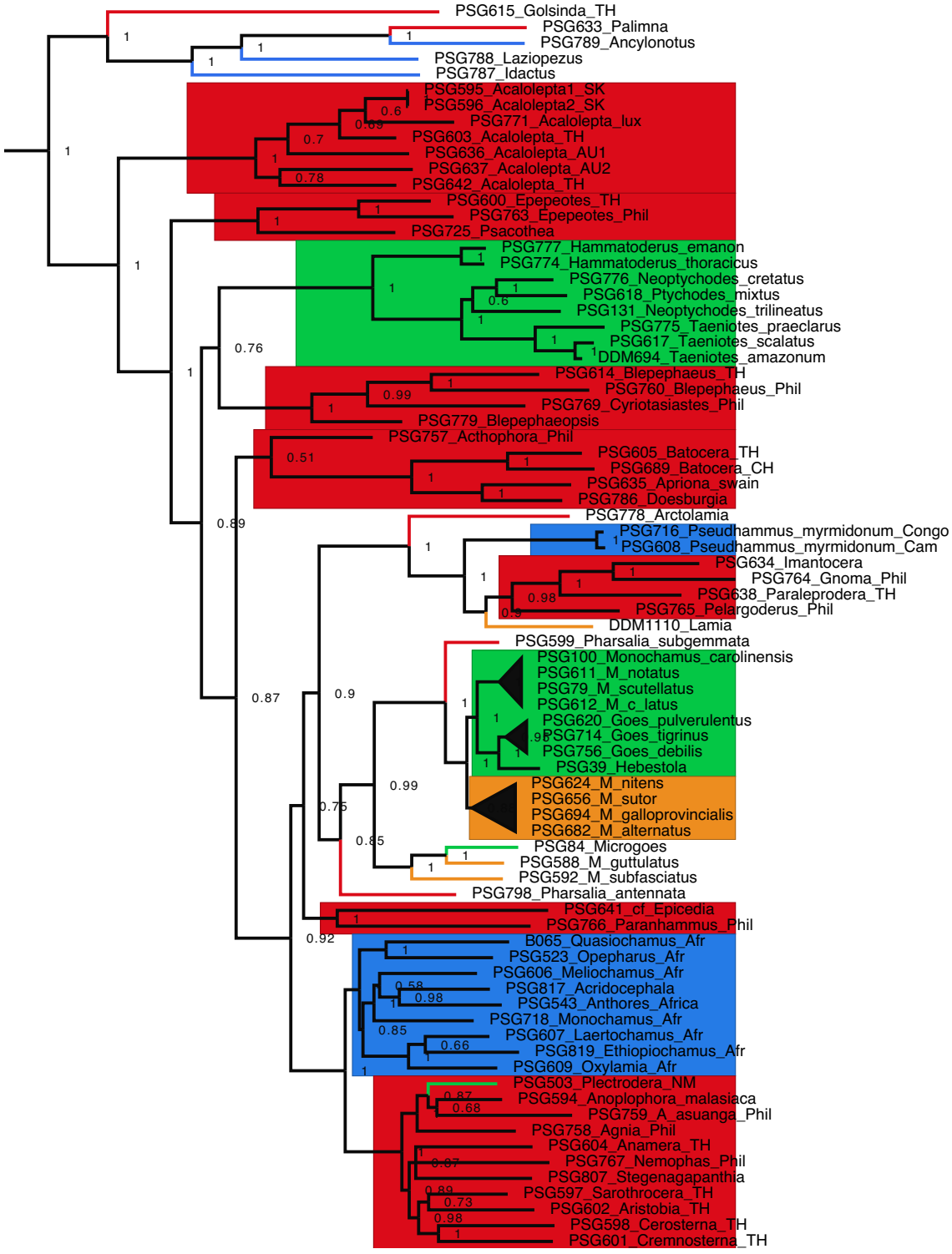

Figure 2.9. Phylogeny colored by geographic region. red=Oriental, green=Western Hemisphere, blue=Africa, orange=Palearctic

Clade origins

The topology based on these genera indicates some historical biogeographic patterns although dating and model-based biogeography analyses were not performed (Fig. 2.9). The clade that comprises genera from South and Central America is found near the base of the tree and separate from North American genera. With its closest relatives in south Asia, this group could have a climate based vicariant origin similar to the cerambycid genus *Callipogon* which ostensibly dispersed over the first Bering Bridge and diverged from an Asian relative ~34 mya (Kim et al. 2018). Up until that point in time, a boreotropical forest covered the Holarctic with a connection through Beringia (Wolfe 1978). Also in support of presence in North America during the Paleogene, a fossil assigned to the tribe Lamiini was described from Florissant beds (Scudder 1878). We found no direct Oriental or Palearctic relative to the *Taeniotes* clade. This clade does carry the plesiomorphic character of an unfused fourth tarsomere which is rare in the tribe so a putative Asian relative may also have this. The North American clades of *Monochamus* and *Goes* + *Hebestola* have a late Miocene-early Pliocene origin according to COI evolutionary rates (Gorring, unpub.). According to the current study, a widespread Holarctic species experienced a vicariant event giving rise to the ancestor of both North American clades. Finally, most of the genera sampled from Africa had a single origin. Its sister group includes mostly southeast Asian genera. The divergence of these clades is not ancient, and they likely separated within the past 25 million years. The only African genus to fall outside of the African clade is the *Pseudhammus* subgenus *Pseudhammus* which is closely related to *Lamia* and allies.

Conclusions

We have found through a molecular analysis and morphological review that the tribes now included in Lamiini *sensu novo* are not reciprocally monophyletic. Some of these former tribes do show a higher level of variable morphology (such as unfused tarsomeres in the former Batocerini clade) that can help to identify them as derived clades within the Lamiini. While the tribes Acridocephalini, Batocerini, Gnomini, and Monochamini are shown to be synonyms of the Lamiini with no support for their separation, the tribes Dorcadionini, Oculariini, Xenoleini, and Xenicotelini which also show similar characters need further analysis to determine if they belong within the tribe. The Phrissomini has been synonymized under the Lamiini (Sama 2008), but the true placement of some of its genera will require further investigation.

Future studies can build upon this work with increased taxon, morphological, and genetic sampling. The taxa sampled for this study represent a substantial enough proportion of total diversity in the clade to get an idea of tribal placements, but determining groupings of genera that may ease identification in this large group will require more genera to be sampled. Ideally, a molecular method that can take advantage of museum specimens of rare and monotypic genera would allow thorough generic and subgeneric sampling to solve abundant taxonomic issues. A morphology-based, or combined molecular and morphological phylogeny of the Lamiini would be another step towards characterizing similar tribes of the Lamiinae. This would enable an evaluation of the morphological traits, their consistency with genetic character evidence, and whether a tribal classification is supported or useful for this variable and diverse subfamily.

Literature cited

Aurivillius, C. 1922. Coleopterum Catalogus. Pars 73 [vol. 23]. Cerambycidae: Lamiinae I. Page (S. Schenkling, Ed.). W. Junk, Berlin.

Bates, H. 1861. Contributions to an Insect Fauna of the Amazon Valley. Coleoptera: Longicornes. The Annals and Magazine of Natural History, London 3:40–52.

Bouchard, P., Y. Bousquet, A. E. Davies, M. a Alonso-Zarazaga, J. F. Lawrence, C. H. C. Lyal, A. F. Newton, C. a M. Reid, M. Schmitt, S. A. Slipiński, and A. B. T. Smith. 2011. Family-group names in Coleoptera (Insecta). ZooKeys 972:1–972.

Breuning, S. 1943. Études sur les Lamiaires (Coleoptera: Cerambycidae). Douzième tribu: Agniini Thomson. Novitates Entomologicae. Revue Mondiale d'Entomologique Systématique - Biologie Troisième:137–144.

Breuning, S. 1961. Catalogue des Lamiares du Monde (Col. Céramb.) Part 5. Museums G. Frey Tutzing, München.

Breuning, S., and A. Villiers. 1973. Trois nouveaux Lamiaires des îles Ryu-Kyu (Col. Cerambycidae). Bulletin de la Société Entomologique de France 78:48–50.

Church, S. H., J. F. Ryan, and C. W. Dunn. 2015. Automation and evaluation of the SOWH test with SOWHAT. Systematic Biology 64:1048–1058.

Degnan, J. H., and N. a Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in ecology & evolution 24:332–40.

Dillon, E. S., and L. S. Dillon. 1959. The Monochamini (Cerambycidae) of the Ethiopian Faunistic Region I. Subtribe Monochamidi. Pseudhammus and Allies.

Dillon, L. S., and E. S. Dillon. 1941. The tribe Monochamini in the Western Hemisphere.

Gistel, J. 1848. Naturgeschichte des Thierreichs für höhere Schulen. Hoffmann, Stuttgart.

Hedtke, S. M., T. M. Townsend, and D. M. Hillis. 2006. Resolution of Phylogenetic Conflict in Large Data Sets by Increased Taxon Sampling. Systematic Biology 55:522–529.

Kim, S., B. A. S. de Medeiros, B. K. Byun, S. Lee, J. H. Kang, B. Lee, and B. D. Farrell. 2018. West meets East: How do rainforest beetles become circum-Pacific? Evolutionary origin of Callipogon relictus and allied species (Cerambycidae: Prioninae) in the New and Old Worlds. Molecular Phylogenetics and Evolution 125:163–176.

Lacordaire, T. 1869. Histoire naturelle des insectes. Genera des coléoptères ou exposé méthodique et critique de tous les genres proposés jusqu'ici dans cet ordre d'insectes. Tome neuvième. Famille des longicornes (suite). Nouvelles suites à Buffon 6:1–409.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34:772–773.

Linsley, E. G., and J. A. Chemsak. 1984. The Cerambycidae of North America, Part VII, No. 1: Taxonomy and classification of the subfamily Lamiinae, Tribes Parmenini through Acanthoderini. University of California Publications in Entomology 102:i–xi, 1–258.

Liu, L., Z. Xi, S. Wu, C. C. Davis, and S. V. Edwards. 2015. Estimating phylogenetic trees from genome-scale data. Annals of the New York Academy of Sciences 1360:36–53.

Mayr, E. 1969. Principles of systematic zoology. McGraw-Hill, New York.

Mulsant, E. 1839. 1re livraison. - Longicornes. Histoire naturelle des Coleopteres de France. Maison, Paris.

Pascoe, F. P. 1864. Longicornia Malayana; or, a descriptive catalogue of the species of the three longicorn families Lamiidæ, Cerambycidæ and Prionidæ collected by Mr. A. R. Wallace in the Malay Archipelago. (Part I). Transactions of the Entomological Society of London 13:1–96, 5pls.

Pascoe, F. P. 1866. Longicornia Malayana; or, a descriptive catalogue of the species of the three longicorn families Lamiidæ, Cerambycidæ and Prionidæ collected by Mr. A. R. Wallace in the Malay Archipelago. (Part III). Transactions of the Entomological Society of London 13:225–336, 5pls.

Rondon, J. A., and S. Breuning. 1970. Lamiines du Laos. Pacific Insects Monograph 24:315–571.

Ronquist, F., M. Teslenko, P. Van Der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–542.

Sama, G. 2008. PRELIMINARY NOTE ON THE CERAMBYCID FAUNA OF NORTH AFRICA WITH THE DESCRIPTION OF NEW TAXA (Insecta Coleoptera Cerambycidae). Quaderno di Studi e Notizie di Storia Naturale della Romagna 27:217–245.

Scudder, S. H. 1878. An Account of Some Insects of Unusual Interest from the Tertiary Rocks of Colorado and Wyoming. Bulletin of the United States Geological and Geographical Survey of the Territories 4:519–542.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Thomson, J. 1860. Essai d'une classification de la famille des cérambycides et matériaux pour servir à une monographie de cette famille. Paris.

Thomson, J. 1864. Systema cerambycidarum ou exposé de tous les genres compris dans la famille des cérambycides et familles limitrophes. H. Dessain, Liége.

Toki, W., and K. Kubota. 2010. Molecular Phylogeny Based on Mitochondrial Genes and Evolution of Host Plant use in the Long-Horned Beetle Tribe Lamiini (Coleoptera: Cerambycidae) in Japan. Environmental Entomology 39:1336–1343.

Wolfe, J. a. 1978. A paleobotanical interpretation of Tertiary climates in the Northern Hemisphere: Data from fossil plants make it possible to reconstruct Tertiary climatic changes, which may be correlated with changes in the inclination of the earth's rotational axis. American Scientist 66:694–703.

Xu, B., and Z. Yang. 2016. Challenges in species tree estimation under the multispecies coalescent model. Genetics 204:1353–1368.

Yang, Z. 2014. Molecular evolution: a statistical approach. Oxford University Press, Oxford.

CHAPTER 3

Not geography but climate influences population structure of sky island pine beetles (Cerambycidae: *Monochamus*) in the Great Basin of North America

Introduction

Millions of species have arisen through evolutionary time, and diverse factors have been implicated in initiation of the speciation process. Herbivorous insects are the most species-rich group of multicellular organisms on the planet with over 400,000 described species (Mitter et al. 1991), so understanding their modes of speciation will help to directly explain a large portion of the Earth's diversity and may shed light on how other groups develop evolutionary isolation. The ecological impact of this high diversity cannot be overstated and their multitude of connections to other species and the environment hold the web of life together. Studying the impact of potential isolating factors in a natural environment is the best way to to understand how multiple factors may contribute to the diversification process.

Geographic isolation is the prevailing hypothesis for the separation of populations (Coyne and Orr 2004) but hypotheses that invoke sympatric processes have gained traction (Nosil et al. 2002, Nosil 2012). Increased speciation rates in insects are also thought to correlate strongly with changes in host-plant geographic range (Howden 1969). These changes directly influence the insects feeding on the plant species adapting to new climate regimes. Rapid isolation relating to host phenological adaptation is evident in *Rhagoletis pomonella* host race formation in sympatry in less than a century (Bush 1969, Feder et al. 1990). Recent research has indicated that the genetic diversity necessary for this host shift originated from introgression from another *R. pomonella* population (Feder et al. 2003). Nevertheless, at minimum genetic sorting and reinforcement took place in sympatry. As with sympatric sorting in host fruit phenology, a vicariant event, which isolates subpopulations with different communities of host plants for thousands of generations, could induce rapid speciation (Futuyma and Mayer 1980).

Over a contracted timescale, these divergent selective forces could be manifest as increased divergence between islands with different host composition when compared to those with the same host(s). The geographic separation of sky islands may also contribute a divergent signal either as isolation by distance (with near islands closer genetically) or as a completely isolating mechanism where islands will show distinctive genetic admixtures.

A natural experiment has been constructed by climate change in mountainous regions of western North America. Climate induced range changes through the Quaternary Period have effected both plants and animals, with distinct phylogeographic units found in many groups of organisms. Conifers are often separated into different species groups on different mountains (Wells 1983). Birds have also diversified through the Pleistocene with divergence times attributable to refugia (Avise and Walker 1998). *Melanoplus* grasshoppers in northern sky islands reveal evidence of glacial refugia, multiple speciation cycles and potential sexual selection (Knowles 2000, 2001). *Habronattus* spiders in the AZ sky islands show evidence of geographic structure and sexual selection in males (Masta 2000, Masta and Maddison 2002). *Moneilema* cactus longhorns show a pattern of recent range expansions in more xeric desert conditions which expanded in recent time (Smith and Farrell 2005). Climate change has moved these groups of varied vagility up and down mountains and in and out of contact with one another. The net effect of these oscillations is isolation that can produce groups that differ in genetics, morphology and behavior. Many of these groups are hypothesized to have experienced a level of sexual selection leading to their rapid divergence. Other selective forces may have had the same effect on separated populations of pine sawyer beetles.

Conifer feeding beetles on sky islands are an ideal system for the study of multiple factors influencing herbivore divergence since potential factors of geographic, ecological, and

environmental isolation can be explored. *Monochamus clamator* (LeConte) pine sawyer beetles (Coleoptera: Cerambycidae) are longhorned beetles that feed on dying and live tissues of various genera of the Pinaceae. This restricts them to conifer forests and presents an opportunity for adaptation to the defenses of conifer communities on habitat islands. They can also fly, giving them moderate dispersal ability and potential for a pattern of isolation by distance (Peterson and Denno 1998). The unique geology of the Great Basin of western North America presents a landscape of some 207 (Charlet 1996) 'sky islands', ~150 of which have species of the family Pinaceae, that are separated by great swathes of desert or steppe habitat. Packrat midden evidence shows a long history of subalpine conifers in the region and more recent (~at 6kya) dispersal of pinyon pine woodland from southern Nevada and California refugia (Wells 1983, Thompson 1990, Cole et al. 2013). This island-like separation of habitats and variation in host-plant presence allows the study of multiple factors that may promote population differentiation.

Population genomic data was used to explore the structure of *M. clamator* across seven Great Basin sky islands. SNPs extracted from aligned RNASeq reads were used to answer a number of questions about the species. The first is whether or not island beetle populations show expected differentiation in standard population statistics such as Jost's *D* (Jost 2008) and allele frequency measures. A second question is whether or not there is substantial migration between the geographically distant islands. Limited gene flow knowledge exists for insects in the Great Basin (Britten et al. 1995, Simpkin et al. 2000) and most measures were of low elevation species with limited numbers of loci. We will measure migration parameters using a modern Bayesian approach. Answering these questions will give an overview of how this moderately vagile organism is moving and interacting with conspecifics at a geographic scale.

In addition to determining how the sky island sawyers are genetically structured, we will focus on biotic and abiotic factors resulting in that structure. These regression measures have been colloquially termed isolation by adaptation, isolation by distance (or resistance in the case of barriers) and isolation by environment. Since divergences between these islands were likely to be taking place over the Pleistocene and into modern time, both next generation genome wide sequencing using transcriptomics and gene expression interrogation were undertaken to capture information on the shorter timescale. Using matrix regression methods, we test geographic isolation and resistance to movement between habitat islands, potential host plant adaptation testing for selected genes and differentially expressed genes, and environmental variables that may be isolating populations. Measuring this selection of population statistics and correlations produces a clear picture of how sawyer beetles are structured in the sky islands of the Great Basin and the factors that may be responsible for this structure.

Methods

RNA-Seq beetle population sampling

In order to assess population structure and gene flow across the habitat islands of the Great Basin, samples of *Monochamus clamator* LeConte pine sawyer beetles were collected from multiple high elevation localities across Nevada (Table 3.1, Fig. 3.1). A minimum of five individuals of each sex were collected from each location using panel traps baited with genus specific aggregation pheromones and host-plant volatiles, UV light setups, and hand collecting. The majority of specimens were collected live into these traps and maintained with fresh hostplant material before being processed. Samples for transcriptome sequencing were photographed and flash-frozen in liquid nitrogen in the field then stored frozen at -80C upon

return to the laboratory. Six individuals were chosen from seven habitat islands to represent its population.

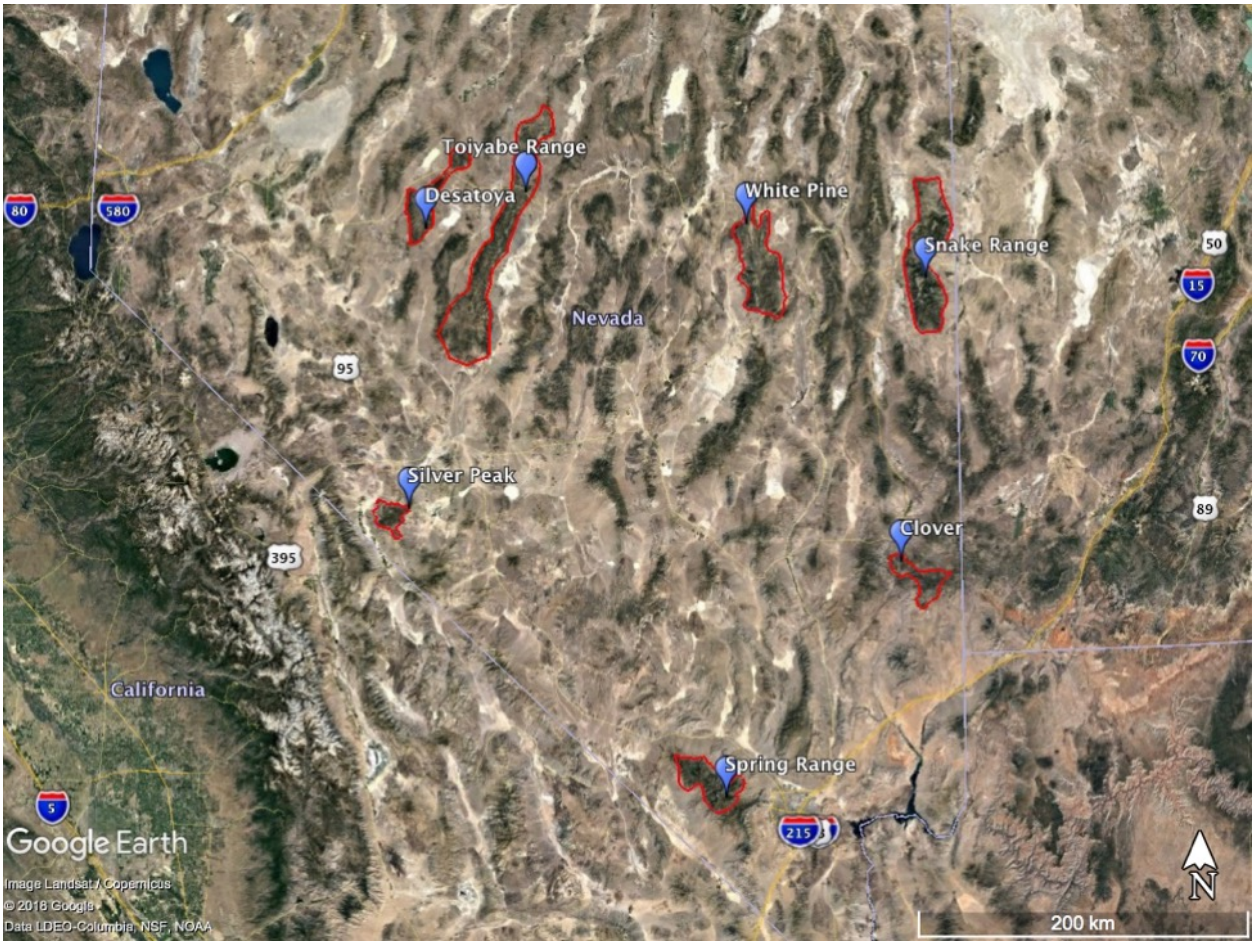

Figure 3.1. Collecting localities in the Great Basin, USA. Pertinent mountain ranges are outline in red. Map image from Google Earth.

Sample	Sex	Specimen	Habitat Island; county, Nevada	Hospitable host species present
PSG000211	M	M.clamator_spring2	Spring Mountains; Clark/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, Abies concolor*
PSG000213	F	M.clamator_spring3	Spring Mountains; Clark/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, Abies concolor*
PSG000214	F	M.clamator_spring4	Spring Mountains; Clark/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, Abies concolor*
PSG000219	M	M.clamator_spring1	Spring Mountains; Clark/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, Abies concolor*
PSG000221	F	M.clamator_spring6	Spring Mountains; Clark/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, Abies concolor*
PSG000225	M	M.clamator_silver1	Silver Peak Range; Esmeralda	*Pinus monophylla*
PSG000226	F	M.clamator_silver3	Silver Peak Range; Esmeralda	*Pinus monophylla*
PSG000227	F	M.clamator_silver4	Silver Peak Range; Esmeralda	*Pinus monophylla*
PSG000231	M	M.clamator_silver2	Silver Peak Range; Esmeralda	*Pinus monophylla*
PSG000232	M	M.clamator_silver5	Silver Peak Range; Esmeralda	*Pinus monophylla*
PSG000233	F	M.clamator_silver6	Silver Peak Range; Esmeralda	*Pinus monophylla*
PSG000236	M	M.clamator_clover1	Clover Mountains; Lincoln	*Pinus monophylla, P. ponderosa*
PSG000237	M	M.clamator_clover2	Clover Mountains; Lincoln	*Pinus monophylla, P. ponderosa*
PSG000238	M	M.clamator_clover3	Clover Mountains; Lincoln	*Pinus monophylla, P. ponderosa*
PSG000239	F	M.clamator_clover4	Clover Mountains; Lincoln	*Pinus monophylla, P. ponderosa*
PSG000240	F	M.clamator_clover5	Clover Mountains; Lincoln	*Pinus monophylla, P. ponderosa*
PSG000242	F	M.clamator_clover6	Clover Mountains; Lincoln	*Pinus monophylla, P. ponderosa*
PSG000250	M	M.clamator_toiyabe1	Toiyabe Range; Nye/Lander	*Pinus monophylla, P. flexilis*
PSG000251	M	M.clamator_toiyabe2	Toiyabe Range; Nye/Lander	*Pinus monophylla, P. flexilis*
PSG000252	M	M.clamator_toiyabe3	Toiyabe Range; Nye/Lander	*Pinus monophylla, P. flexilis*
PSG000256	F	M.clamator_toiyabe4	Toiyabe Range; Nye/Lander	*Pinus monophylla, P. flexilis*
PSG000257	F	M.clamator_toiyabe5	Toiyabe Range; Nye/Lander	*Pinus monophylla, P. flexilis*
PSG000258	F	M.clamator_toiyabe6	Toiyabe Range; Nye/Lander	*Pinus monophylla, P. flexilis*
PSG000261	M	M.clamator_desatoya1	Desatoya Mountains; Churchill/Lander	*Pinus monophylla*
PSG000262	F	M.clamator_desatoya3	Desatoya Mountains; Churchill/Lander	*Pinus monophylla*
PSG000263	M	M.clamator_desatoya2	Desatoya Mountains; Churchill/Lander	*Pinus monophylla*
PSG000268	F	M.clamator_desatoya4	Desatoya Mountains; Churchill/Lander	*Pinus monophylla*
PSG000269	F	M.clamator_desatoya5	Desatoya Mountains; Churchill/Lander	*Pinus monophylla*
PSG000272	M	M.clamator_desatoya6	Desatoya Mountains; Churchill/Lander	*Pinus monophylla*
PSG000286	M	M.clamator_whitepine1	White Pine Range; White Pine/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, Abies concolor*
PSG000287	M	M.clamator_whitepine2	White Pine Range; White Pine/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, Abies concolor*
PSG000288	M	M.clamator_whitepine3	White Pine Range; White Pine/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, Abies concolor*
PSG000289	F	M.clamator_whitepine4	White Pine Range; White Pine/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, Abies concolor*
PSG000292	F	M.clamator_whitepine5	White Pine Range; White Pine/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, Abies concolor*
PSG000294	F	M.clamator_whitepine6	White Pine Range; White Pine/Nye	*Pinus monophylla, P. flexilis, P. ponderosa, Abies concolor*
PSG000326	M	M.clamator_snakebristle1	Snake Range; White Pine	*Picea engelmannii, Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, P. albicaulis, Abies concolor*
PSG000335	M	M.clamator_snakecamp1	Snake Range; White Pine	*Picea engelmannii, Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, P. albicaulis, Abies concolor*
PSG000336	M	M.clamator_snakecamp2	Snake Range; White Pine	*Picea engelmannii, Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, P. albicaulis, Abies concolor*
PSG000338	F	M.clamator_snakecamp3	Snake Range; White Pine	*Picea engelmannii, Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, P. albicaulis, Abies concolor*
PSG000340	F	M.clamator_snakecamp4	Snake Range; White Pine	*Picea engelmannii, Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, P. albicaulis, Abies concolor*
PSG000354	M	M.clamator_snake9300	Snake Range; White Pine	*Picea engelmannii, Pinus monophylla, P. flexilis, P. ponderosa, P. longaeva, P. albicaulis, Abies concolor*

Table 3.1. Samples collected from the habitat islands of the Great Basin for RNA-Seq.

RNA-Seq transcriptome sequences

The mRNA was extracted from *Monochamus clamator* antennae to build individual tissue specific transcriptomes. Antennae were removed for mRNA extraction from frozen specimens one individual at a time. Both antennae were ground in liquid nitrogen and mRNA was directly extracted using the Dynabeads mRNA direct kit (Thermo Fisher kit 61011) or total RNA was extracted using the Nucleospin RNA plus extraction kit (Clontech 740984.50). With total RNA, the Dynabeads were subsequently used to purify mRNA from the extract (Thermo Fisher kit 61006). After extraction, the Qubit machine was used to estimate RNA quantity. RNA was then visualized for ribosomal RNA contamination and successful extraction using the Agilent Bioanalyzer machine with the RNApico chip.

A processing robot was used to reduce RNA-Seq library preparation variability and work with small amounts of mRNA. Libraries were prepared using the Apollo 324^{TM} NGS Library Prep robot in Harvard's Bauer Core facility. This machine allows for automated cDNA prep using kits produced by Wafergen, reducing bench time and potential variation over samples prepared. The Wafergen PrepX^{TM} RNA-Seq library kit was used to create stranded cDNA. All extracted mRNA (~60ng) was used for the reaction mixture. Adapter ligation PCR was performed for 13 cycles using the small RNA primer provided in the PrepX kit and different illumina index primers for each sample to be pooled together. A 1X Ampure XP bead cleanup was then done on the Apollo following PCR. PCR and cleanup followed the PrepX RNA-Seq protocol. After prep, a DNA qubit measure was done for each library and Tapestation (High Sensitivity D1000 tape) or Bioanalyzer (high sensitivity DNA chip) were run to estimate average library size. Finally, either qpcr or the formula (library concentration ng/ul x 1000000ul/1L x 1nmol/660ng x 1/avg. fragment size in bp) was used to measure library molarity.

Indexed libraries were combined into pools and diluted to 2nm concentration for loading. Paired end 150bp (rapid run) and 125bp (v4) sequencing was done at the Bauer Core facility on an illumina HiSeq 2500 machine, targeting similar read counts for each library. All following bioinformatics procedures were run on Harvard's 'Odyssey' Linux computing cluster.

Read processing and reference transcriptome build

In order to create a clean read set for each individual, several quality control steps were performed after the reads were returned. Approximately 1.4 billion reads were sequenced for this study's samples. Reads were corrected with Rcorrector (Song and Florea 2015) using default settings to correct reads that could lead to erroneous kmers. Unfixable reads were discarded. TrimGalore v. 0.4.2 built on cutAdapt (Martin 2014) was used to trim reads by sample for quality and adapter contamination with a quality cut of q-score 25, length minimum of 36bp, and default values for other options. Since the library prep adapters were non-standard their sequences were manually added for trimming. This resulted in 1,174,385,104 reads to use for downstream analyses (~179 million per habitat island). A blacklisting procedure was also done to remove ribosomal RNA in the dataset. This seemed to be more of an issue for samples using the direct mRNA protocol and rRNA was less abundant in samples that had total RNA extracted first. To do this, Bowtie2 v. 2.2.2 (Langmead and Salzberg 2012) was used to create an index from available small and large subunit rRNA curated and available on the Silva website (www.arb-silva.de). Bowtie2 was set to very-sensitive-local, and any sequences that aligned to the rRNA index were removed from the read set. Anywhere from 100,000 to over a million reads were removed from samples indicating that rRNA can be captured along with mRNA on the

dynabeads. Overall, these filtering steps may be stringent, but with the high number of reads sequenced more can be removed in order to ultimately build better contigs.

A reference antennal transcriptome was built *de novo* from a subset of individuals in the dataset to align reads for gene expression counts and SNP calling. The individuals used are indicated in table 1 and total nine individuals representing every habitat island, both sexes, and ~288 million reads. To build this transcriptome Trinity v. 2.2.0 (Haas et al. 2013) was used in a strand specific mode, with in silico normalization, min_kmer_cov 2, max_reads_per_loop 5000000, SS_lib_type set to FR according to the Wafergen kit strand protocol, and group_pairs_distance 800.

Reference annotation

To further filter this transcriptome and increase its utility for beetle expression measures the transcripts were annotated using multiple reference databases. Using *Blast* v. 2.2.29+ (Camacho et al. 2009), blastx and blastp searches were performed on the uniref90 database and a custom database with available (at GenBank, www.ncbi.nlm.nih.gov) beetle transcripts (65,325) and proteins (75,413) from *Tribolium castaneum, Anoplophora glabripennis, Monochamus alternatus, Agrilus planipennis,* and *Dendroctonus ponderosae*. A blastn search was also done on the beetle nucleotides. Results were output in the .outfmt6 format and combined into an annotation matrix using the PandAnnotate.py script from *PandAnnotate* (https://github.com/harvardinformatics/PandAnnotate). This matrix was then pruned, removing transcripts not annotated to invertebrates and those with no annotation. *RSEM* v. 1.2.29 (Li and Dewey 2011) was used to create an expected counts matrix for the transcripts and those with

combined expression across samples less than five were removed for an ultimate transcript list of 107,533 transcripts.

With this number of transcripts, it is likely that some fragments belong to the same isoform so transcript clustering would benefit downstream expression quantification. The reduced transcript list was extracted from the *RSEM* counts matrix and those sequences were pulled from the original trinity fasta output using the UCSC genome browser's utility faSomeRecords (http://genome.ucsc.edu). The RNA-Seq quasi-mapper *Salmon* v. 0.8.0 (Patro et al. 2017) was used to set up the clustering of transcripts by building an index based on the reduced transcript list then using the quant command with option -dumpEq to map reads and provide the clustering tool with equivalence classes from the mapping. *RAPclust* (https://github.com/COMBINE-lab/RapClust) was used to cluster the transcripts. This tool uses mapping information of the samples to create a mapping ambiguity graph that is then clustered in their treatment groups, if provided. We chose to make each habitat island a 'treatment' group which, if anything, will inflate the final number of clusters. The end result was 69,831 clusters and a list of transcripts with no assigned cluster, data that can be used to build a map linking transcripts to their respective clusters. The consensus gene symbol from multiple blast searches was attached to the cluster and transcript.

SNP calling

For assessment of population structure, differentiation measures, and to identify potentially selected genes, a set of SNPs was called from the clustered transcript set reduced to those over 500bp. This fasta formatted sequence list was made into a BWA index using *BWA* v. 0.7 (Li and Durbin 2009). Each of the 41 individuals in the study were aligned to this indexed

transcriptome using the BWA-mem command with -M option to make results *Picard Tools* compatible. The .bam formatted output files were then sorted using *SAMtools* v. 1.4 (Li et al. 2009). Since this is a SNP calling procedure and read depth is not essential, duplicate reads from the alignment were removed using *Picard Tools* v. 2.9.0 (http://broadinstitute.github.io/picard) to increase computing efficiency. All sample .bam files were then aligned using the *SAMtools* mpileup command with -B option to disable read realignments and a minimum mapping quality (q) of 1. This result was piped to *VarScan*.v. 2.3.7 (Koboldt et al. 2009) for variant calling using mpileup2cns with default options except a minimum coverage of 5. To further reduce this set to usable downstream SNPs, *VCFtools* v. 0.1.14 (Danecek et al. 2011) was used to constrain to biallelic sites, remove indels, and recode data. Two datasets were produced: one with no missing data and one that required presence in at least 2 individuals per habitat island. For the population thinned dataset a minor allele count (--mac) of 2 was given and a minimum genotype depth (minDP) of 3, but not for the no missing set. All SNP work in *R* was done under *R* version 3.4.2 (R Core Team 2017). For linkage disequilibrium pruning of the SNP set the package SNPRelate v. 1.4.2 (Zheng et al. 2012) was used with an LD threshold of .2 to remove SNPs in LD within a sliding window. This outputs a list of SNPs in linkage equilibrium that can be used with *VCFtools* to produce an LD free file.

Gene expression

Differentially expressed genes, expression distance, and expression clusterings were measured across islands. For comparisons an expected counts matrix was produced using *RSEM* and the rsem-calculate-expression function which produces isoform and gene level counts. This is done by first building a reference index in *RSEM* based on a gene to transcript map produced

from the reduced transcript list. The summarized count matrix across all samples is in expected counts which are non-integer due to resulting from posterior probabilities of alignment which are not always 1.

Differential expression was measured between habitat islands according to host plant presence. A design array was formed with specimens assigned to low (1 sp.), medium (2 spp.), and high (>2 spp.) host-plant diversity according to the habitat island they originated from. These were assigned in contrasts of host diversity, setting pairwise comparisons of the habitat types. Differential expression testing was very similar to the tutorial found at https://ucdavis-bioinformatics-training.github.io/2018-June-RNA-Seq-Workshop/thursday/DE.html. The *RSEM* expected counts matrix had normalization factors calculated and was transformed to counts per million and rows without at least 5 samples with expression greater than 1 were removed. To compare for differentially expressed genes *R* v. 3.4.1 packages edgeR v. 3.12.1 (Robinson et al. 2010) and limma v. 3.26.9 (Ritchie et al. 2015, Phipson et al. 2016) were used. The voom function in limma was used to find regression coefficients according to the contrast matrix and hypotheses were tested as contrasts of the fitted linear models. Differentially expressed genes were decided at a false discovery p-value <.05 and minimum log fold change of 1.5. A second analysis of elevation and tree species separated specimens in the Snake Range was also done in the same manner.

To measure and visualize gene expression distance between samples a combination of base R v. 3.5.1 and packages DESeq2 v. 1.22.2 (Love et al. 2014), pheatmap v. 1.0.12 (https://CRAN.R-project.org/package=pheatmap), ggplot2 v. 3.0.0 (Wickham 2016) and the genefilter v. 1.64.0 bioconductor package were the main packages used (see Appendix 2 for complete code). The RSEM expected counts matrix was rounded to integer values to allow use in

DESeq2, then this matrix was transformed using DESeq2 function vst() to normalize the sometimes extreme counts and allow more genes to contribute signal to distance measures. Between sample distance was measured with Euclidean distance using the dist() function. Clustering of samples and genes was done with the hclust() function and complete linkage. To create a pairwise distance matrix between island clusters the meandist () function in the vegan v. 2.5-2 (https://CRAN.R-project.org/package=vegan) package was used on the vst standardized distance object to give average Euclidean distances between manually assigned island clusters.

Population measures

Because the beetles in this study were collected on conifer trees restricted to mountain tops with desert or steppe habitat in between, the genetic structure of the putative populations was of interest. For these measures either the LD pruned SNP set with no missing data (16k set) or the population thinned dataset (44k set) were used. These datasets were converted from VCF to genepop format using *PGDspider* v. 2.1.1.0 (Lischer and Excoffier 2012). This .gen file was imported to *R* and made into a genind object using adegenet v. 2.1.1 (Jombart and Ahmed 2011). The hierfstat package v. 0.04-22 (https://CRAN.R-project.org/package=hierfstat) was used to translate the genind to a hierfstat object (genind2hierfstat()) and by population diversity statistics *Hs* and *Ho* were calculated using the basic.stats() function. The MMOD v. 1.3.3 (Winter 2012) package was used to calculate pairwise Jost's *D*, Nei's F_{ST} and Hedrick's G_{ST} for the islands using the genid object. This was chosen as the differentiation statistic since traditional stats like F_{ST} have been shown to level and decrease as allelic differentiation of subpopulations grows (Jost 2008). This could be an issue with organisms isolated from one another for many generations, as may be the case in island inhabiting species. The other statistics are included for

comparison of utility. MMOD does not support pairwise stats by locus, so locus specific pairwise differentiation statistics were calculated using the diffCalc () function in the diveRsity v. 1.9.90 (Keenan et al. 2013) package.

To genetically cluster individuals we used adegenet and the 44k SNP set. The k-means driven find.clusters () function was called with a maximum number of clusters set to ten and retaining 20 pca axes. The ideal cluster number was chosen using the Bayesian information criterion. The hierfstat function indpca () was used to create a PCA based on allele frequencies of the individuals.

To measure gene flow between islands the software *Migrate* v. 3.6.11 (Beerli and Felsenstein 2001, Beerli 2006) was used in Bayesian mode with 5,000 of the SNPs from the 16k no missing data set. The *Migrate* formatted document was created from the VCF file in *PGDspider*. Migration directionality was free to vary, constant mutation rate across loci and prior distributions for θ and *M* were uniform. Two runs of 400,000 reps with a burn-in of 2,000 and recording every 20 steps were performed. Chain stationarity was gauged with effective sample sizes and similarity of the results over the two runs.

To determine if any SNPs may be representative of selected markers across the islands we used the software *BayeScan* v. 2.1 (Foll and Gaggiotti 2008). This Bayesian method uses allele frequency differences of populations and can account for population sizes and immigration rates. Default settings, which make the neutral model 10x more likely than the selective model, were used with the 16k and 44k SNP sets and seven populations were assigned, one for each habitat island. A run with two populations was also done, one population contained the islands with one host (Silver and Desatoya) and the other population included all other islands.

Matrix regression

In order to compare multiple variables potentially relating to genetic diversity in these beetles matrices were constructed and matrix regression analysis performed. The genetic distance matrix was constructed based on pairwise Jost's *D* differentiation values from the 44k SNP set as described above. The host-plant diversity dissimilarity matrix was calculated in *R* using the vegan package and vegdist () function with jacaard distance based on the number of hospitable tree species present on the island (Table 3.1). To calculate environmental distances all 19 bioclim variables were downloaded from Worldclim (http://worldclim.org/version2) at 2.5 minute resolution and cropped to the area of the study by making a stack and cropping it using the raster v. 2.6-7 (https://CRAN.R-project.org/package=raster) package. The sp v. 1.3-1 (Bivand et al. 2013) package was used to make a spatial points dataframe from the locality information of the island samples. Raster::extract() was used to pull the bioclim variables per specimen which were averaged by population. A PCA was done on the averaged bioclim variables using the prcomp() function to account for correlation of bioclim variables with one another. PC1 accounted for 99.1% of the variance in the data so was used to create a Euclidean distance matrix with the dist() function. Geographic distance was measured using spherical distance functions implemented in the Geographic Distance Matrix Generator v. 1.2.3 software (http://biodiversityinformatics.amnh.org/open_source/gdmg).

A multiple matrix regression with randomization was performed in *R* using the MMRR function (Wang 2013). Each of the matrices for genetic distance, expression distance, host distance, geographic distance, and environmental distance were brought into the same environment. Distance matrices were standardized by subtracting the mean and then multiplying

the result by the square root of result/result^2. Column and row names were also assigned to be the same for all matrices. The MMRR script was run for 1000 permutations.

Results

Sequencing and reference transcriptome

Library preparation and sequencing was successful and produced more than 25 million reads for all individuals (Table 3.2). A reference *Monochamus clamator* antennal transcriptome was built with multiple males and females from across habitat islands. This transcriptome has 258,123 Trinity genes and 354,446 transcripts. There is a mean contig length of 444.68bp and N50 of 478 based on all contigs with a total of 157,613,300 assembled bases. Using BUSCO v. 2.0.1 benchmarking (Simão et al. 2015) with the Endopterygota gene set, 80.6% complete and 14.3% fragmented BUSCO orthologs (2318/2442) were recovered.

Table 3.2. Sequencing and processed read results, gray fill indicates the sample was used as part of the reference transcriptome assembly. Island totals correspond to all samples of the island of the corresponding sample

Sample	Specimen	Raw reads	Clean reads (post trimming)	Total island reads (125-150bp)
PSG000211	M.clamator_spring2	32736168	27,165,858	
PSG000213	M.clamator_spring3	28258446	24,275,568	
PSG000214	M.clamator_spring4	36570618	30,183,826	
PSG000219	M.clamator_spring1	52805394	38,423,712	
PSG000221	M.clamator_spring6	32003558	28,975,030	149,023,994
PSG000225	M.clamator_silver1	45893698	33,216,734	
PSG000226	M.clamator_silver3	32041334	27,916,316	
PSG000227	M.clamator_silver4	27171960	24,486,914	
PSG000231	M.clamator_silver2	50108754	36,845,198	
PSG000232	M.clamator_silver5	36376204	30,470,610	
PSG000233	M.clamator_silver6	38525390	29,373,594	182,309,366
PSG000236	M.clamator_clover1	34616240	25,221,922	
PSG000237	M.clamator_clover2	33946676	28,395,938	
PSG000238	M.clamator_clover3	31965542	27,511,606	
PSG000239	M.clamator_clover4	32659444	29,373,588	
PSG000240	M.clamator_clover5	27420190	23,975,698	
PSG000242	M.clamator_clover6	30698458	26,907,172	161,385,924
PSG000250	M.clamator_toiyabe1	36221286	27,558,180	
PSG000251	M.clamator_toiyabe2	33642068	28,423,968	
PSG000252	M.clamator_toiyabe3	39643094	33,309,418	
PSG000256	M.clamator_toiyabe4	33145070	27,540,664	
PSG000257	M.clamator_toiyabe5	32424676	27,113,482	
PSG000258	M.clamator_toiyabe6	29037196	25,725,842	169,671,554
PSG000261	M.clamator_desatoya1	37041456	27,708,156	
PSG000262	M.clamator_desatoya3	32522204	28,378,882	
PSG000263	M.clamator_desatoya2	37306246	31,446,876	
PSG000268	M.clamator_desatoya4	30280732	24,728,700	
PSG000269	M.clamator_desatoya5	32801218	29,648,996	
PSG000272	M.clamator_desatoya6	39915282	33,604,174	175,515,784
PSG000286	M.clamator_whitepine1	28829334	24,643,764	
PSG000287	M.clamator_whitepine2	32030568	26,109,156	
PSG000288	M.clamator_whitepine3	26668456	21,779,650	
PSG000289	M.clamator_whitepine4	31503176	29,498,572	
PSG000292	M.clamator_whitepine5	27195832	24,195,764	
PSG000294	M.clamator_whitepine6	28704412	26,086,880	152,313,786
PSG000326	M.clamator_snakebristle1	43687666	30,692,492	
PSG000335	M.clamator_snakecamp1	34408810	26,652,526	
PSG000336	M.clamator_snakecamp2	33080656	27,526,976	
PSG000338	M.clamator_snakecamp3	39418272	33,944,556	
PSG000340	M.clamator_snakecamp4	36166202	32,233,762	
PSG000354	M.clamator_snake9300	38548984	33,114,384	184,164,696

SNP calling

The SNP calling workflow resulted in high numbers of SNPs for use in population genomic calculations and selection testing. After the initial VCF filtering done on each dataset (16k & 44k) the no missing data set had 65,922 SNPs and the population thinned dataset that required at least two representatives of that SNP per island had 143,306 sites. Following the LD thinning process the set with no missing data had 16,036 SNPs while the population thinned set had 44,038.

Population measures

Genetic clustering analysis gives BIC support for two distinct genetic clusters in the 44k set of SNP data (Fig. 3.2). One of these two allelic clusters contains all the Silver Peak individuals and the other all other island individuals. This pattern can also be seen in an allele frequency based PCA plot of all individuals (Fig. 3.3). The results of *Migrate* runs show low population sizes and a high importance of immigration relative to mutation (Table 3.3, Fig. 3.4). The *M* value is the immigration rate per generation divided by the mutation rate per generation so larger values show that immigration is more important for new variants in a population.

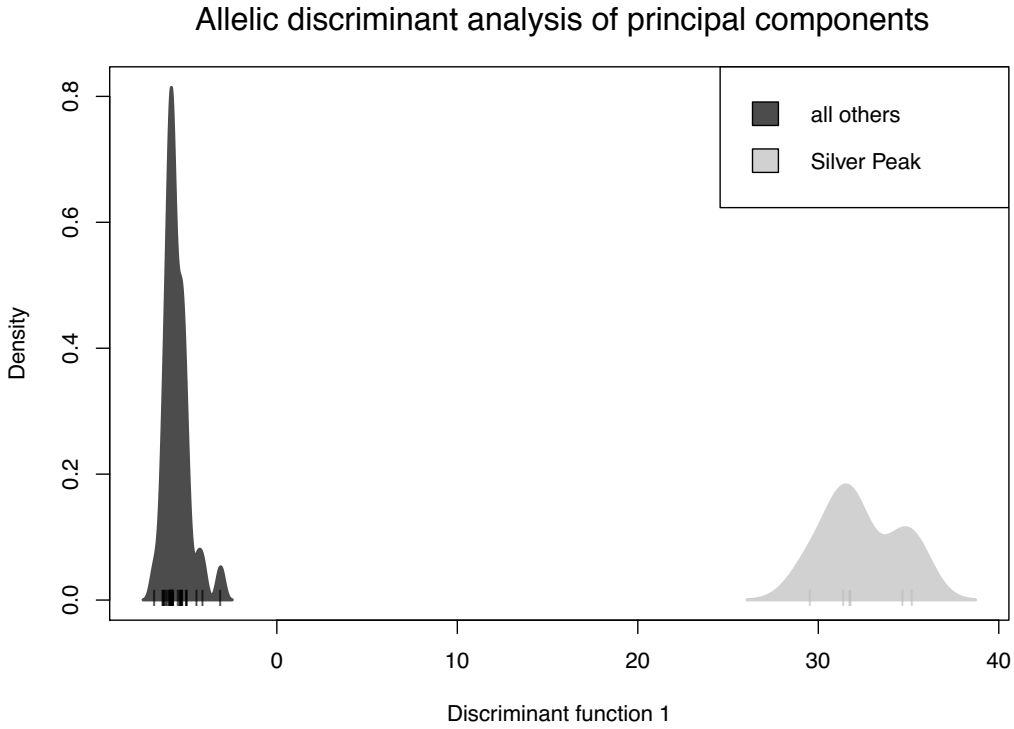

Figure 3.2. Discriminant analysis density plot based on the allele level genomic data

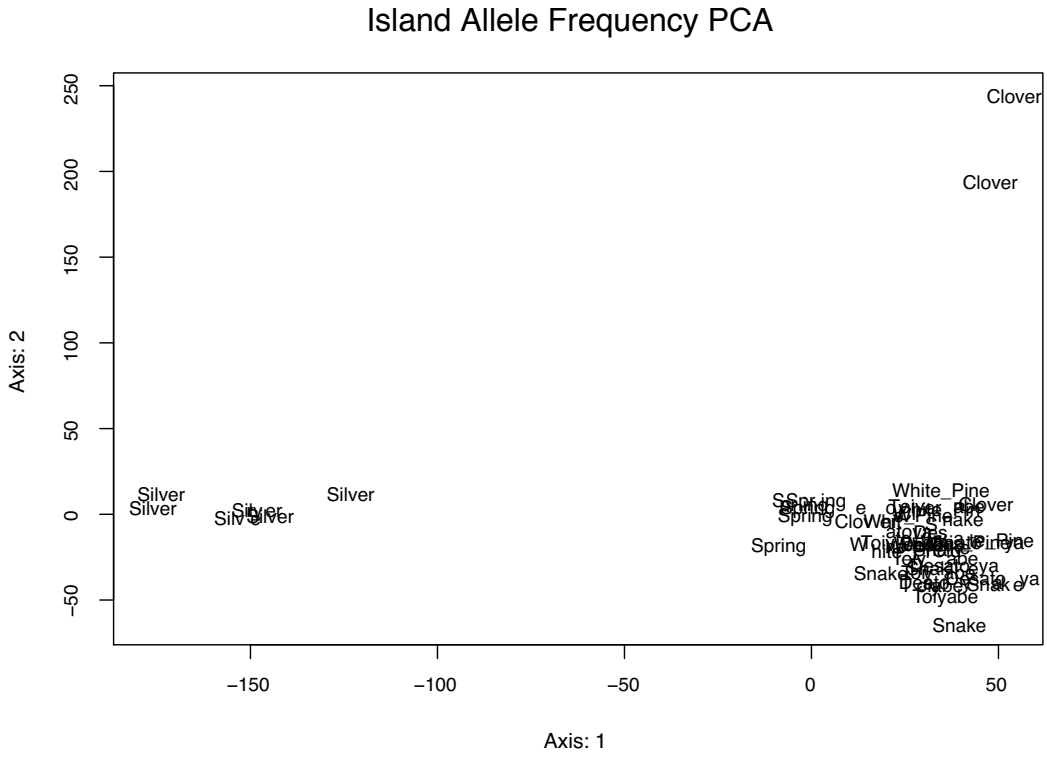

Figure 3.3. Allele frequency PCA plot with individuals labeled by island of residence, based on 44k SNP set

Locus	Measure	Mean	Locus	Measure	Mean
All	M_Silver->Spring	109.57	All	M_Desatoya->Toiyabe	74.57
All	M_Clover->Spring	110.99	All	M_White_Pine->Toiyabe	87.21
All	M_Toiyabe->Spring	100.82	All	M_Snake->Toiyabe	74.07
All	M_Desatoya->Spring	99.32	All	M_Spring->Desatoya	85.3
All	M_White_Pine->Spring	105.65	All	M_Silver->Desatoya	69.61
All	M_Snake->Spring	109.95	All	M_Clover->Desatoya	86.85
All	M_Spring->Silver	63.9	All	M_Toiyabe->Desatoya	83.17
All	M_Clover->Silver	112.87	All	M_White_Pine->Desatoya	86.89
All	M_Toiyabe->Silver	72.3	All	M_Snake->Desatoya	85.55
All	M_Desatoya->Silver	81.06	All	M_Spring->White_Pine	63.58
All	M_White_Pine->Silver	85.56	All	M_Silver->White_Pine	68.14
All	M_Snake->Silver	79.01	All	M_Clover->White_Pine	85.83
All	M_Spring->Clover	72.9	All	M_Toiyabe->White_Pine	62.97
All	M_Silver->Clover	77.43	All	M_Desatoya->White_Pine	73.27
All	M_Toiyabe->Clover	80.06	All	M_Snake->White_Pine	96.41
All	M_Desatoya->Clover	99.3	All	M_Spring->Snake	92.76
All	M_White_Pine->Clover	86.19	All	M_Silver->Snake	71.31
All	M_Snake->Clover	71.65	All	M_Clover->Snake	98.16
All	M_Spring->Toiyabe	71.93	All	M_Toiyabe->Snake	75.9
All	M_Silver->Toiyabe	71.69	All	M_Desatoya->Snake	92.87
All	M_Clover->Toiyabe	90.22	All	M_White_Pine->Snake	76.35

Table 3.3. *Migrate* results of the mean effective immigration rate *M* for a no missing data set of 5000 SNPs, -> indicates migration direction

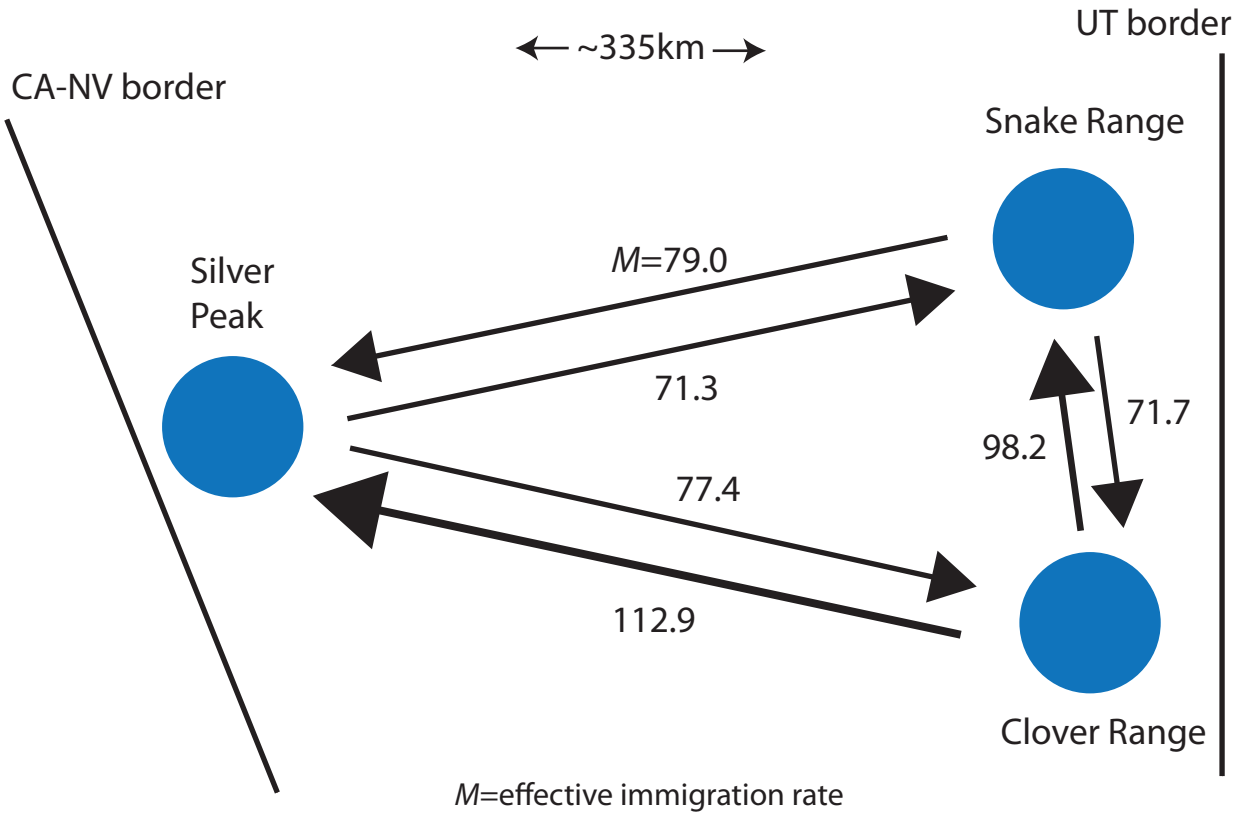

Figure 3.4. Simplified three island diagram of migration between Great Basin sky islands. *M* value is the effective immigration rate from the *Migrate* program and indicates relative importance of migration compared to mutation for new diversity in populations.

	Mean Hs	Mean Ho	Mean theta
Spring	0.1177	0.0879	0.0047
Silver	0.1141	0.0907	0.0053
Clover	0.1197	0.0807	0.0054
Toiyabe	0.1255	0.0939	0.0062
Desatoya	0.1243	0.0931	0.0054
White_Pine	0.1201	0.0799	0.0054
Snake	0.1282	0.0963	0.0061
overall	0.12	0.09	

Table 3.4. Population summary statistics. *Ho* and *Hs* calculated using the *R* hierfstat package, θ using *Migrate*

	Spring	Silver	Clover	Toiyabe	Desatoya	White_Pine
Silver	0.01768855					
Clover	0.01027234	0.02114295				
Toiyabe	0.00876456	0.01846116	0.00624622			
Desatoya	0.00894792	0.01856683	0.00600976	0.00364496		
White_Pine	0.00920414	0.01950948	0.00624592	0.00447083	0.00447471	
Snake	0.0090062	0.01893147	0.00602618	0.00413527	0.00419804	0.00482585

Table 3.5. Pairwise Jost's *D* on 44k SNP set, mmod package

	Spring	Silver	Clover	Toiyabe	Desatoya	White_Pine
Silver	0.10665622					
Clover	0.04583103	0.12478424				
Toiyabe	0.03458992	0.10937692	0.01917173			
Desatoya	0.04010717	0.10946643	0.02159211	0.00526574		
White_Pine	0.04257716	0.11411697	0.01772174	0.00983973	0.00988585	
Snake	0.04063318	0.11129417	0.02003133	0.00806051	0.00925196	0.01241613

Table 3.6. Pairwise Hedrick G_{ST} calculation on 44k SNP set using mmod package

	Spring	Silver	Clover	Toiyabe	Desatoya	White_Pine
Silver	0.05285941					
Clover	0.02199671	0.06239723				
Toiyabe	0.01646521	0.05411567	0.00905891			
Desatoya	0.01915447	0.05419913	0.01022092	0.00246657		
White_Pine	0.02038828	0.05671804	0.00838564	0.0046257	0.0046505	
Snake	0.01937052	0.05503758	0.00945592	0.00377301	0.00433595	0.00583595

Table 3.7. Pairwise Nei G_{ST} (F_{ST}) calculation on 44k SNP set using mmod package

Genetic diversity measures were similar across the island populations (Table 3.4). θ values are low for all populations indicating a very low effective population size. Since these samples are all within the same species it is probably safe to say mutation rates for each SNP is similar so comparisons can be done but all values are close indicating small effective population sizes irrespective of island size. The observed heterozygosity (*Ho*) and gene diversity (*Hs*) on each island are similar to the overall mean. Jost's *D* differentiation values were all low in pairwise comparisons of the islands (Table 3.5). With *D* around .02 across pairwise comparisons, the differentiation between Silver Peak and other islands is about twice that of other island comparisons (Fig. 3.5). Some differentiation measures can falter when individual populations gain more unique alleles (Jost 2008), our reason for using Jost's *D* for matrix regression. While the alleles of our populations are not showing much insularity, a comparison of Nei's F_{ST} and Hedrick's G_{ST} (Tables 3.6,3.7) shows a similar trend to Jost's *D* with pairwise comparisons including Silver Peak double the value of other comparisons.

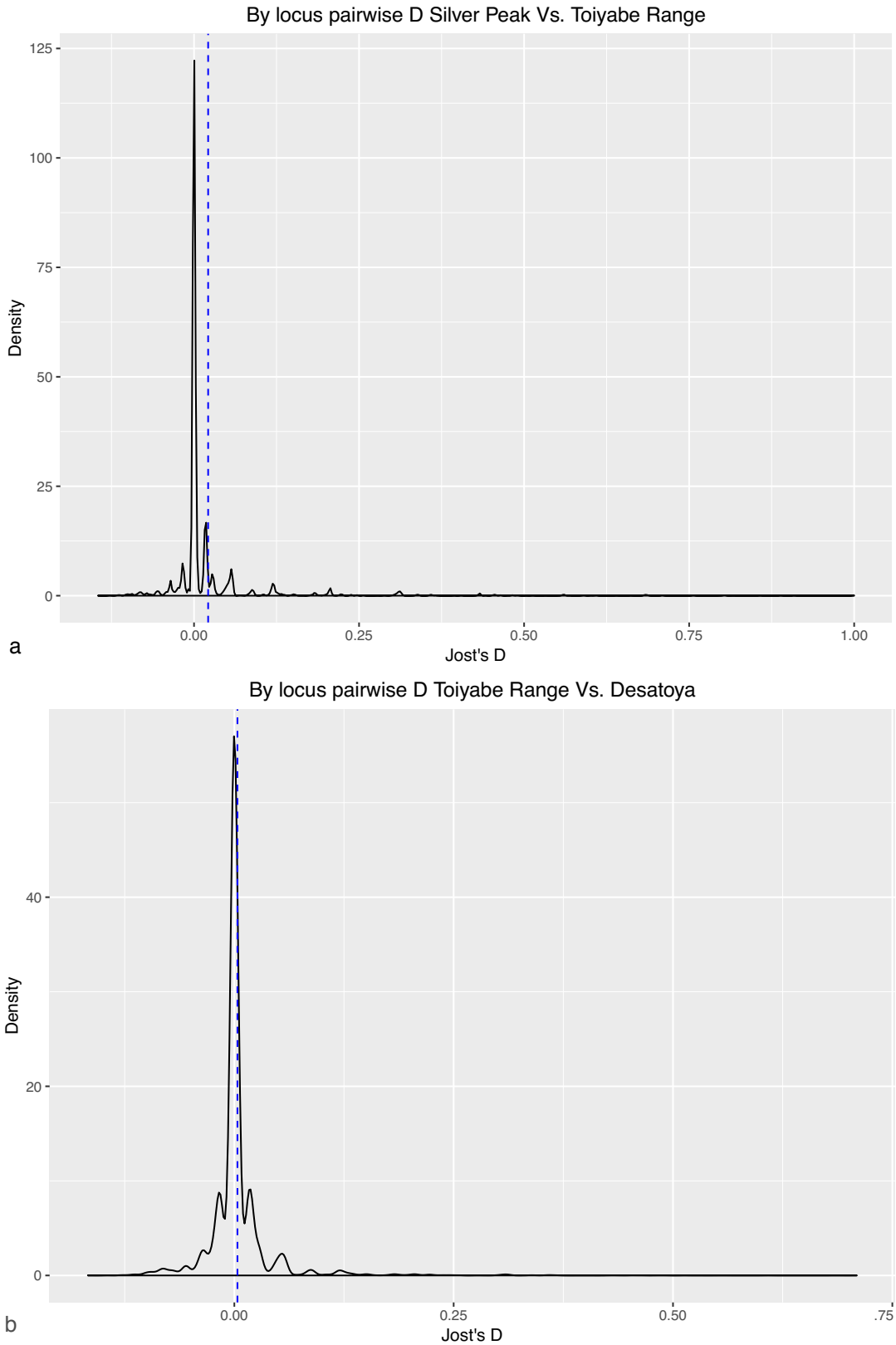

Figure 3.5. Pairwise Jost's *D* values density over all loci from 44k SNP set. Mean value of *D* shown by dashed line. Indicates higher differentiation in Silver Peak comparisons (a).

Matrix regression

Multi-variable matrix regressions of both genetic distance and gene expression distance on the variables of host, geography, and environment were done. The distance matrices for each of these variables are in tables 3.8,3.9, and 3.10. Dendrograms to visualize the information found in these matrices are in Figure 3.6. The multi-matrix regression of these variables with genetic distance resulted in r^2=.24 and *F*-statistic=1.81 (p=.06). The coefficient and t-test results are in Table 3.11 and plots of individual variables are in Figure 3.7. The only significant coefficient in the analysis was environment. In a similar regression using gene expression distances between the islands there were no significant results.

	Clover	Desatoya	Silver	Snake	Spring	Toiyabe
Desatoya	0.5					
Silver	0.5	0				
Snake	0.7142857	0.8571429	0.8571429			
Spring	0.6	0.8	0.8	0.2857143		
Toiyabe	0.6666667	0.5	0.5	0.7142857	0.6	
White_Pine	0.5	0.75	0.75	0.4285714	0.2	0.5

Table 3.8. Host distance matrix (Jaccard distance)

	Clover	Desatoya	Silver	Snake	Spring	Toiyabe
Desatoya	342972					
Silver	290057.2	165363.2				
Snake	168462.3	299638.8	334556.2			
Spring	173195.8	382044.4	254052.8	330097.4		
Toiyabe	307262.8	66304.7	198265.7	240112.9	375755.7	
White_Pine	215860.6	193136.4	260139.5	109172.6	337953.4	131405.3

Table 3.9. Geographic distance matrix, in meters

	Clover	Desatoya	Silver	Snake	Spring	Toiyabe
Desatoya	0.0275					
Silver	0.0379	0.0104				
Snake	0.0032	0.0243	0.0348			
Spring	0.0275	0.0001	0.0105	0.0243		
Toiyabe	0.0088	0.0188	0.0292	0.0056	0.0187	
White_Pine	0.0034	0.0309	0.0413	0.0065	0.0308	0.0121

Table 3.10. Environmental distance matrix, Euclidean distance on principal component 1 accounting for 99% of variance from the 19 Bioclim variables

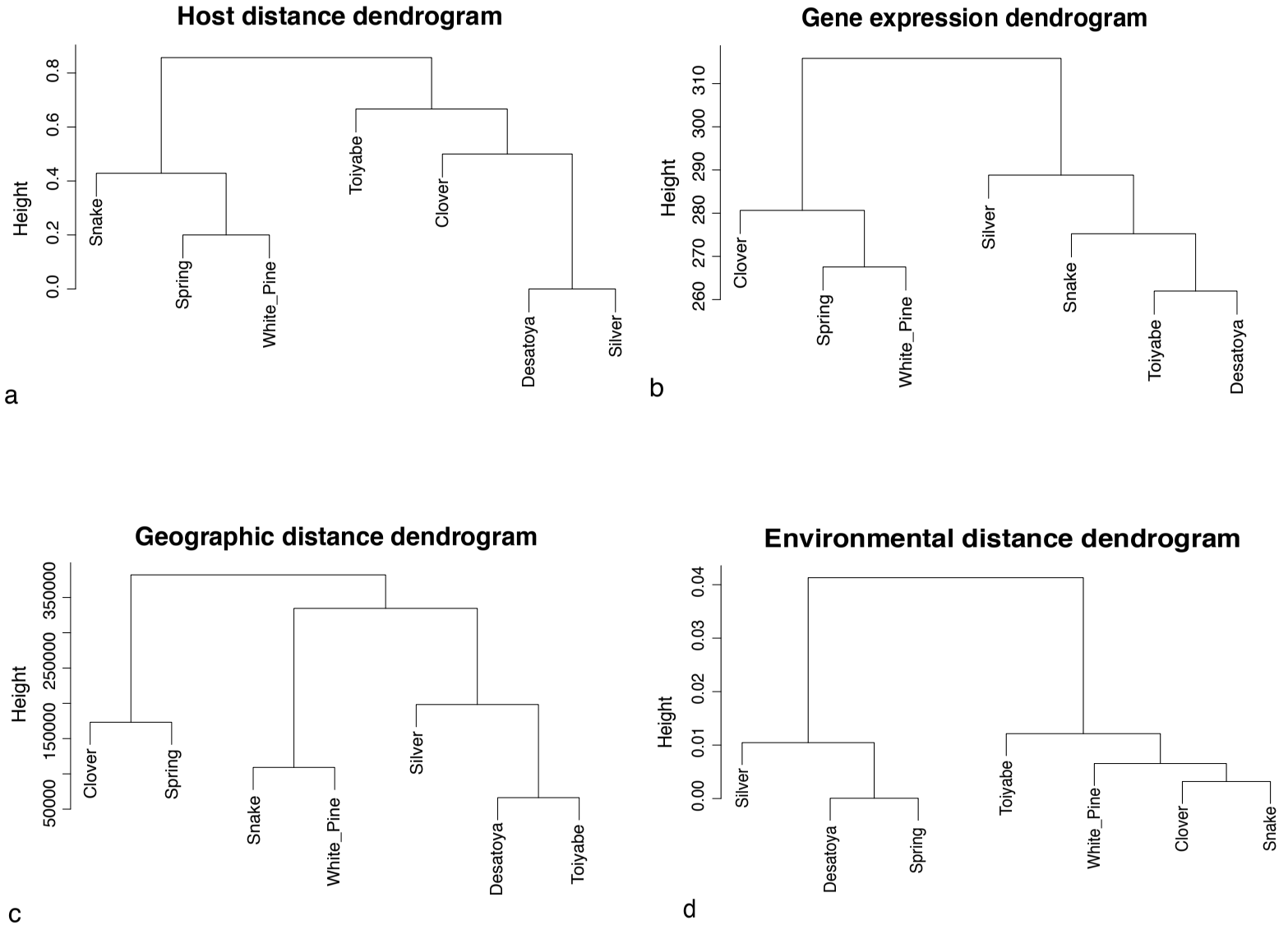

Figure 3.6. Cluster dendrograms based on the distance matrices produced for matrix regression, complete linkage distance

	Intercept	host	distance	environment
coefficients	-1.22E-16	-1.13E-01	1.22E-01	4.44E-01
t-statistic	-5.78E-16	-5.24E-01	5.48E-01	2.05E+00
t p-value	0.16	0.355	0.439	0.035

Table 3.11. MMRR matrix regression results for genetic distance (Jost's *D*)

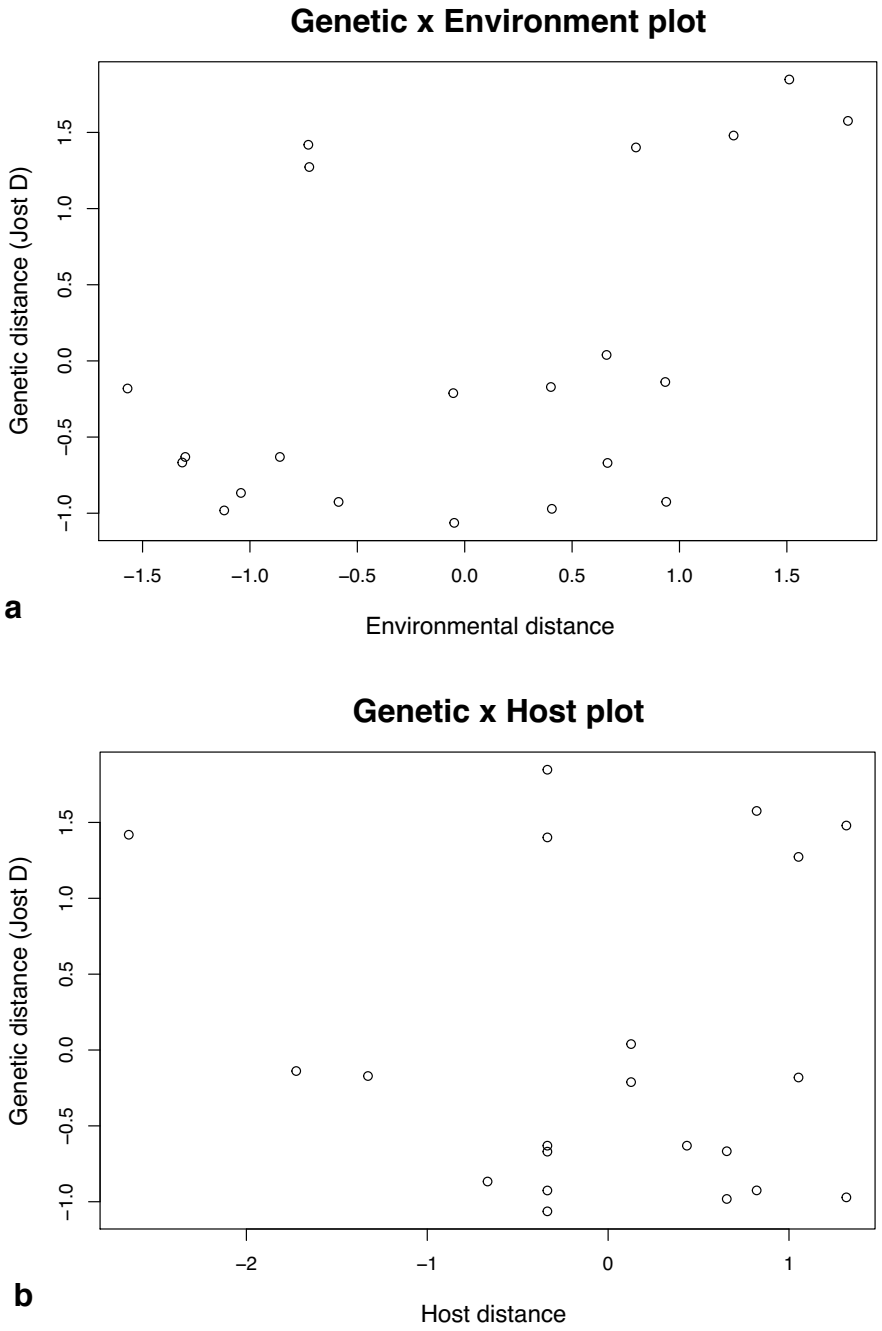

Figure 3.7. Pairwise comparison plots for matrix variables environment (a), host (b), and geography (c) against genetic distance (Jost's *D*)

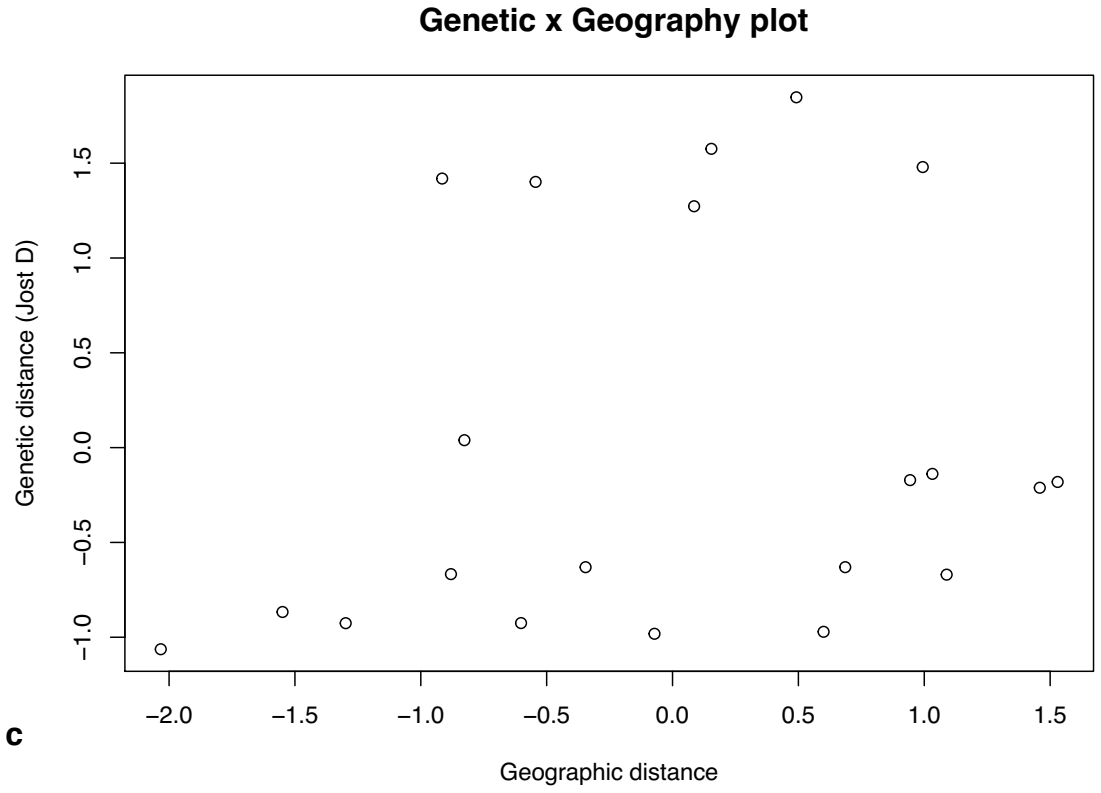

Figure 3.7. (Continued)

Selected markers

Tests for selected alleles were performed on the 16k, no missing dataset and the 44k population thinned set. When assigning the seven island populations, with the 16k set *BayeScan* nine F_{ST} outlier SNPs were found. With the 44k set 35 outliers were found. The results from the 16k set were traced to the annotated gene clusters where they originated and the gene function was searched. Most of the genes matched to the closely related Asian Longhorned Beetle genome but functions like 'equilibrative nucleoside transporter 4' cannot be clearly tied to pathways relating to variables studied. Two population runs of *BayeScan* returned no SNPs under selection.

Gene expression

Gene expression results were mixed for the populations concerned. No signal was found for island specific expression patterns but differentially expressed genes were revealed in some comparisons. When the distances of the gene expression matrices for each individual are clustered in a dendrogram (Fig. 3.8) a common expression pattern is not found for any island population. When choosing the top 50 genes most variable in expression there is a pattern revealed for the clustering of the genes and of the beetle samples (Fig. 3.9), but these beetle clusters seem random and do not correspond to any variables measured in this study.

Testing for differential expression showed high numbers of significant differences in expression of individual genes, with greater than 1.5 log fold change. The scenarios to explain the largest results were comparing low host diversity (1 host species) islands to medium diversity (2 host species) islands (321 DE genes) and comparing low host to high host diversity (> 2 host species, multiple genera) islands (424 DE genes). Both of these comparisons include Silver Peak as one of the low host islands. A contrast of specimens from high and low elevations in the Snake Range produced no differentially expressed genes at $p<.05$ and $logFC>1.5$.

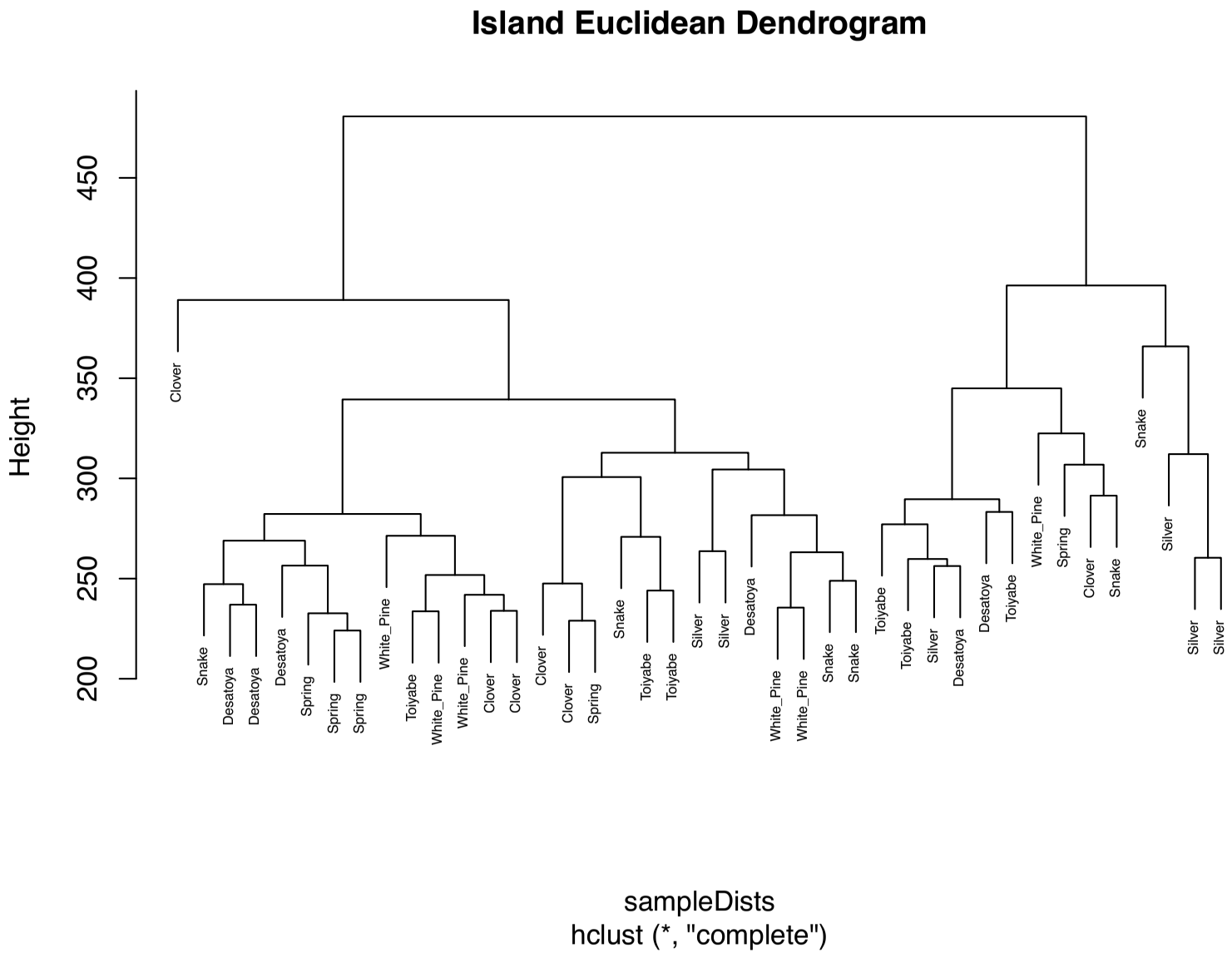

Figure 3.8. Grouped gene expression Euclidean distance dendrogram, complete linkage distance

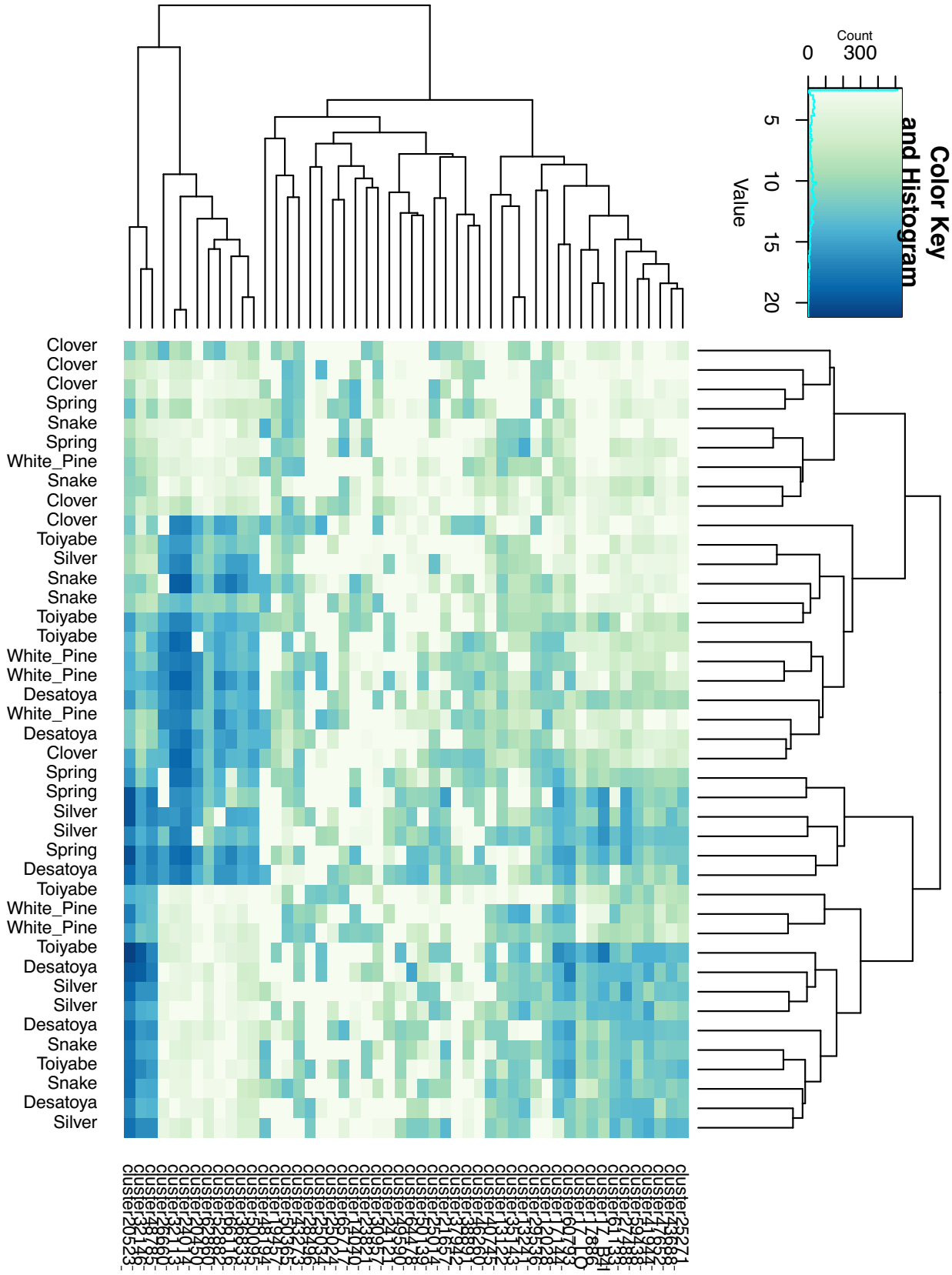

Figure 3.9. Top 50 variable expression genes heatmap, based on vst transformed counts. Shows no expression pattern similarity by island population.

Discussion

The investigation of factors leading to the assumed isolation of Great Basin *Monochamus clamator* sawyer beetle populations revealed that they are freely exchanging alleles and consequently are not well separated. The only island to show some minor genetic separation was Silver Peak Range, a range in the more arid west of the Great Basin with only *Pinus monophylla* present as a host, but this island is not unique among those sampled in having these characteristics. Silver Peak is the closest to the Sierra Nevada range, the rain shadow of which may impose environmental factors not experienced by other populations. Other island populations are relatively homogeneous according to genomic data. There is no evidence of direct selection on the genome, but gene expression data, which may be expected to reveal the initial adaptations to natural environments, show that the structuring between Silver Peak and the other islands may be due to interactions with biotic or abiotic factors.

Population differentiation

The sawyer beetles living in the conifer forests of the habitat islands of the Great Basin are much more vagile than would be expected by witnessing their clumsy flight. Population differentiation statistics close to zero (pairwise Jost's *D* range =.004-.02) show that there is high overall similarity in the genetic structure across islands spanning the state of Nevada (Table 3.5, Fig. 3.3). Low Ho and Hs values indicating low population heterozygosity may indicate a recent bottleneck. *Dendroctonus* bark beetles show genetic structuring (F_{ST}~.3), indicating isolation by distance, through contiguous pine forest in the western USA (Mock et al. 2007). This study showed high divergence for populations east and west of the Great Basin meaning that while

Monochamus beetles seem to be traveling well across the basin other pine associates may not. No *Dendroctonus* sampling was done in the Great Basin area, so direct comparison of *Monochamus* and *Dendroctonus* is not yet possible.

One of the most intriguing results of this research is the importance of migration between habitat islands, even hundreds of miles apart. *Migrate* software results show a high importance of immigration over mutation (Table 3.3) even from near the Utah border in the Snake Range to Silver Peak near the Nevada-California border (Fig. 3.4). Values for other islands are similar or even higher. Since *Nm* is the product of θ and *M* divided by 4 for diploids, the small theta values calculated by *Migrate* give *Nm* values <1 for the beetles. A low number is expected considering the distance and habitat between these sky islands, but there is not a geographic trend with more migration between neighbor islands. Using SNPs to calculate theta may also impact the calculation so using longer loci is a future direction. Dragonflies, which are extremely vagile organisms, in the Great Basin show rates that peak around *Nm*=35 for distances <200km across lowland habitat (Simpkin et al. 2000). The comparatively less vagile boreal *Euphydryas* butterflies show *Nm* as high as ~11 in the Great Basin (Britten et al. 1995) even with sampling from ends of the same habitat island. The movement of *Monochamus* beetles is surprising given their slow, erratic flight pattern, but they have been shown to disperse more than 20km in the European species *M. galloprovincialis* (David et al. 2014). It seems very unlikely that larvae within tree trunks could be moved by humans from one mountaintop to another. It is conceivable that flying beetles that rise high enough could be moved long distances by wind, but movement was similar in both directions between all of the sampled islands, while wind would be more unidirectional. *Tetraopes* longhorned beetles show some migration between milkweed patches of less than 60m and only carry enough energy to survive one to two days without feeding

(McCauley 1983, Davis 1984) but anecdotally these beetles have been said to show up in potted milkweeds in pine forest (B. Farrell, pers. Comm.). Moreover, wind pollinated pines themselves show strong population structure and their pollen should be easier to transport by wind (Jørgensen et al. 2002).

Isolation by distance, gene flow

Most organisms studied in sky island systems show some level of isolation by distance. While some of these species are poor movers due to flightlessness, inhospitable terrain should be a formidable barrier for flighted species, and therefore greater genetic similarity with similar geographically close habitat is the expected pattern. Herbivorous insects of varied vagility were reviewed for isolation by distance (Peterson and Denno 1998). The authors describe the expected patterns of no correlation of genetic and geographic distance in very good dispersers due to apparent panmixis and very locally structured populations in weak dispersers, also resulting in no IBD. Insects with moderate dispersal ability are expected to show the pattern of increasing genetic distance with geographic distance, at least in contiguous habitat. Sky island organisms would be expected to follow this trend, possibly with some added level of differentiation due to habitat barriers. However, the near absence of studies with detailed descriptions of both population structure and geographically-explicit sampling was limiting in 1998.

The surge over the last 20 years in development and application of molecular techniques suitable for population studies has resulted in an increase in relevant observations, however, including those of the present study. Sawyer beetles of the Great Basin sky islands do not show any significant pattern of isolation by distance (Fig. 3.7c, MMRR p=.44). Lower elevation montane species such as flightless *Moneilema* cactus longhorns (Smith and Farrell 2005) and

Baetis mayflies (Polato et al. 2017) show patterns of IBD migrating preferentially to nearby areas and are either flightless (*Moneilema*) or adults for a matter of days (*Baetis*). *Baetis* from higher elevations show the local structuring and weak IBD expected of low vagility herbivores (Polato et al. 2017). Boreal *Euphydryas* butterflies, expected to be decent dispersers, showed no significant pattern of IBD in the Great Basin sky islands, attributed to the insular nature of habitats on mountaintops (Britten et al. 1995). The pattern of no IBD due to panmixis has not been recorded in sky island insect species and the dispersal barrier of the intervening lowlands has been thought to eliminate this possibility. The evidence of non-significant correlation between geography and genetic distance, low differentiation in alleles (Fig. 3.3), and low differentiation values show that *M. clamator* is close to panmictic across the Great Basin.

Environment and Host-plant influence

There is no evidence that host plant differences among sky islands are influencing the genomes of the sky island sawyers, but there may be an impact on the expression of genes. In the matrix regression, host diversity actually had a negative coefficient and no significance in the result. One thing to consider with this result is how the plant community was reduced to the dissimilarity of tree species present. A method that weighted presence of additional tree genera or better captured the phylogenetic difference in species could lead to a different result. The overall gene expression distance regression also did not show host as a significant factor. While the hypothesis that isolation with different host compositions can lead to adaptation has not been truly tested (since these beetles are not really isolated), the case of the slightly differentiated island Silver Peak is interesting. When contrasting this island (in combination with the other low host diversity island Desatoya) with islands that have more host plant availability, hundreds of

genes are differentially expressed. This could be related to the host composition of the islands, and we are seeing the start of adaptation that will continue if gene flow lessens. The number of differentially expressed (DE) genes increases as host diversity distance increases, suggesting a connection. The possibility of geographic grouping in just the differentially expressed genes merits further investigation. Since the multiple matrix regression attempts to disentangle the variables of host diversity and environment which may be correlated and shows no significance for host, this DE pattern may also be attributable (in part or exclusively) to environmental factors.

Selected genes could provide guidance on whether host or environment is leading to high numbers of DE genes. Surprisingly, there are no markers showing a signal of selection in the *BayeScan* two population analysis, when islands were grouped by single host and multiple host. Since selection is only found when more islands are their own populations, selection may be occurring in response to variables we have not explored. Further study could place the selected markers into known biological pathways to gain some insight into the selective force. Additionally, the contrast testing for differential gene expression by elevation in the Snake Range did not result in any DE genes, when a change of both climate and host plant community should have been experienced by the beetles. *Baetis* mayflies show selected genes associated with elevation differences associated with temperature and habitat changes (Polato et al. 2017). A lack of difference in sawyers may indicate that the variables associated with pine woodland and subalpine conifer habitats are not different enough to instigate gene expression adaptation. There was also low statistical power for this test due to limited samples so patterns may be revealed with more individuals.

The impact of the environment, termed isolation by environment (IBE)(Wang and Bradburd 2014), is the only factor to show a significant correlation with genetic distance in matrix regression and the western islands do cluster together in environmental distance (Fig. 3.6). The increasing aridity that results from approaching the Sierra Nevada's rain shadow and the Mojave Desert could be influencing these populations. The Bioclim variable for temperature seasonality showed the most loading in the PCA analysis used to create the environmental distance matrix. The western Great Basin is subject to extremely high summer temperatures and may not cool as much as the central Great Basin 'cold desert' areas. With the regression data associating environmental with genetic differentiation and host diversity differences lying on the same East-West axis, the evidence seems to point to environment influencing gene expression. Testing new island groupings for DE, informed by the evidence, can help to clarify this result.

Gene expression in natural environments

In field studies, the expression of genes has been underexplored relative to population genetic investigation, but there is potential for expression differences to be the first response in changing adaptive landscapes. The natural environment is hard to control, but it is also extremely difficult to replicate (Eberhardt and Thomas 1991, Lemoine et al. 2016). Gene expression patterns can be one of the most plastic aspects of the biochemical makeup of an insect which is good for adaptation but problematic for research. Common gardens are one approach where external variables can be controlled for some organisms, but finding the correct environment often is difficult logistically and species may have unforeseen responses (Cheviron et al. 2008). Expression patterns not only change due to extrinsic but also intrinsic factors like age of the specimens and tissue sampled. Combining mRNA sampling in a natural environment and

gauging plasticity in a simultaneous common garden in specimens of known age and development may be an ideal strategy. Natural laboratories like sky islands can permit accurate measurements of gene expression and realistic assessment of the process of adaptation.

The collections for this study were made in the field and were as consistent as possible. Beetles were kept in a similar way with a local host plant and samples from an island were killed together in the same manner directly into liquid nitrogen. This strategy permitted the influence of the natural environment at these sites while minimizing change in variables that may distort gene expression patterns.

Gene expression has been implicated as a first indicator of the divergence process in animals (Wolf et al. 2010). What we may be seeing in the differentially expressed gene evidence presented here is the early stages of adaptation to environment or host diversity, which is not able to progress due to immigration and mixing of populations. This may be a common pattern in nature, where shifting population ranges, due to change in climate or other factors, create instances of temporary isolation sufficient for the local accumulation of novel mutations or allele frequencies which are subsequently lost when populations merge again (Futuyma 1987). Speciation itself may be a kind of ratchet that preserves such differentiation from loss.

The fact that overall expression patterns do not group by population could indicate a plastic response to micro-habitat variables or could be a product of genetic variation in individuals impacting mRNA production. Plasticity could potentially be controlled for by field based common gardens (Cheviron et al. 2008) considering the challenges discussed above. The results of gene expression in this study are not clear but add to considerations for future field-based gene expression sampling. Future work on this dataset can be to reduce the gene set

analyzed to decrease noise or to focus on differentially expressed genes and their patterns across the landscape.

Conclusions

Altogether, we find the sky island conifer associated sawyer beetle community to be amazingly homogeneous across the Great Basin. With a genomic scale SNP dataset and a geographically clear sampling scheme over habitat islands that vary in host composition the factors that have led to genetic differentiation among islands can be appropriately interrogated. This test, in a robust multiple matrix regression form, revealed that geography and host do not impact the genetic structure of these beetles but the environment does. While especially surprising that geography does not have a structuring influence, the environmental factors acting on these populations should be explored further as they may become increasingly important for many creatures as climate on these mountain areas continues to warm. This study, considering it is spatially explicit and care was taken to gather information on potentially adaptive factors, moves forward the understanding of the spatial scale of adaptation. This type of information is needed for more species and environments (Peterson and Denno 1998) to achieve a consensus understanding of how ecological factors may structure populations and the spatial scale at which adaptation can evolve to promote speciation.

Literature cited

Avise, J. C., and D. Walker. 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings. Biological sciences / The Royal Society 265:457–63.

Beerli, P. 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22:341–45.

Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proceedings of the National Academy of Sciences 98:4563–68.

Bivand, R. S., E. Pebesma, and V. Gómez-Rubio. 2013. Applied Spatial Data Analysis with R: Second Edition. Springer, New York.

Britten, H. B., P. F. Brussard, D. D. Murphy, and P. R. Ehrlich. 1995. A test for isolation-by-distance in central rocky mountain and great basin populations of edith's checkerspot butterfly (Euphydryas editha). Journal of Heredity 86:204–210.

Bush, G. L. 1969. Sympatric Host Race Formation and Speciation in Frugivorous Flies of the Genus Rhagoletis (Diptera, Tephritidae). Evolution 23:237–251.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC bioinformatics 10:421.

Charlet, D. 1996. Atlas of Nevada Conifers A Phytogeographic Reference. University of Nevada Press, Reno, NV.

Cheviron, Z. A., A. Whitehead, and R. T. Brumfield. 2008. Transcriptomic variation and plasticity in rufous-collared sparrows (Zonotrichia capensis) along an altitudinal gradient. Molecular Ecology 17:4556–4569.

Cole, K. L., J. F. Fisher, K. Ironside, J. I. Mead, and P. Koehler. 2013. The biogeographic histories of Pinus edulis and Pinus monophylla over the last 50,000 years. Quaternary International 310:96–110.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

David, G., B. Giffard, D. Piou, and H. Jactel. 2014. Dispersal capacity of Monochamus galloprovincialis, the European vector of the pine wood nematode, on flight mills. Journal of Applied Entomology 138:566–576.

Davis, M. A. 1984. The flight and migration ecology of the red milkweed beetle (Tetraopes tetraophthalmus). Ecology 65:230–234.

Eberhardt, L. L., and J. M. Thomas. 1991. Designing environmental field studies. Ecological Monographs 61:53–73.

Feder, J. L., S. H. Berlocher, J. B. Roethele, H. Dambroski, J. J. Smith, W. L. Perry, V. Gavrilovic, K. E. Filchak, J. Rull, and M. Aluja. 2003. Allopatric genetic origins for sympatric host-plant shifts and race formation in Rhagoletis. Proceedings of the National Academy of Sciences 100:10314–10319.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1990. The Geographic Pattern of Genetic Differentiation between Host Associated Populations of Rhagoletis pomonella (Diptera: Tephritidae) in the Eastern United States and Canada. Evolution 44:570–594.

Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977–93.

Futuyma, D. J. 1987. On the Role of Species in Anagenesis. The American Naturalist 130:465–473.

Futuyma, D. J., and G. Mayer. 1980. Non-allopatric speciation in animals. Systematic Biology 29:254–271.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. Macmanes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. Leduc, N. Friedman, and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 8:1494–512.

Howden, H. F. 1969. Effects of the pleistocene on north american insects. Annual Review of Entomology 14:39–56.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–71.

Jørgensen, S., J. L. Hamrick, and P. V. Wells. 2002. Regional patterns of genetic diversity in Pinus flexilis (Pinaceae) reveal complex species history. American Journal of Botany 89:792–800.

Jost, L. 2008. G(ST) and its relatives do not measure differentiation. Molecular ecology 17:4015–26.

Keenan, K., P. Mcginnity, T. F. Cross, W. W. Crozier, and P. A. Prodöhl. 2013. DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution 4:782–788.

Knowles, L. L. 2000. Tests of Pleistocene Speciation in Montane Grasshoppers (Genus Melanoplus) from the Sky Islands of Western North America. Evolution 54:1337–1348.

Knowles, L. L. 2001. Did the pleistocene glaciations promote divergence? Tests of explicit refugial models in montane grasshopprers. Molecular ecology 10:691–701.

Koboldt, D. C., K. Chen, T. Wylie, D. E. Larson, M. D. McLellan, E. R. Mardis, G. M. Weinstock, R. K. Wilson, and L. Ding. 2009. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25:2283–2285.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357–9.

Lemoine, N. P., A. Hoffman, A. J. Felton, L. Baur, F. Chaves, J. Gray, Q. Yu, and M. D. Smith. 2016. Underappreciated problems of low replication in ecological feld studies. Ecology 97:2554–2561.

Li, B., and C. N. Dewey. 2011. RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. BMC Bioinformatics 12:323.

Li, H., and R. Durbin. 2009. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–60.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Lischer, H. E. L., and L. Excoffier. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299.

Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology 15:550.

Martin, M. 2014. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17:10–12.

Masta, S. E. 2000. Phylogeography of the jumping spider Habronattus pugillis (araneae: salticidae): recent vicariance of sky island populations? Evolution; international journal of organic evolution 54:1699–711.

Masta, S. E., and W. P. Maddison. 2002. Sexual selection driving diversification in jumping spiders. Proceedings of the National Academy of Sciences of the United States of America 99:4442–7.

McCauley, D. E. 1983. Gene Flow Distances in Natural Populations of Tetraopes tetraophthalmus. Evolution 37:1239–46.

Mitter, C., B. D. Farrell, and D. J. Futuyma. 1991. Phylogenetic Studies of Insect-Plant interactions : Insights into the genesis of diversity. Trends in Ecology & Evolution 6:290–93.

Mock, K. E., B. J. Bentz, E. M. O'Neill, J. P. Chong, J. Orwin, and M. E. Pfrender. 2007. Landscape-scale genetic variation in a forest outbreak species, the mountain pine beetle (Dendroctonus ponderosae). Molecular Ecology 16:553–568.

Nosil, P. 2012. Ecological Speciation. Oxford University Press, New York.

Nosil, P., B. J. Crespi, and C. P. Sandoval. 2002. Host-plant adaptation drives the parallel evolution of reproductive isolation. Nature 417:440–3.

Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford. 2017. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. Nature Methods 14:417–419.

Peterson, M., and R. Denno. 1998. The influence of dispersal and diet breadth on patterns of genetic isolation by distance in phytophagous insects. The American naturalist 152:428–46.

Phipson, B., S. Lee, I. J. Majewski, W. S. Alexander, and G. K. Smyth. 2016. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Annals of Applied Statistics 10:946–63.

Polato, N. R., M. M. Gray, B. A. Gill, C. G. Becker, K. L. Casner, A. S. Flecker, B. C. Kondratieff, A. C. Encalada, N. L. Poff, W. C. Funk, and K. R. Zamudio. 2017. Genetic diversity and gene flow decline with elevation in montane mayflies. Heredity 119:107–116.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids research 43:e47.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–40.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Simpkin, J. L., H. B. Britten, and P. F. Brussard. 2000. Effects of habitat fragmentation and differing mobility on the population structures of a Great Basin dragonfly (Sympetrum corruptum) and damselfly (Enallagma carunculatum). Western North American Naturalist 60:320–332.

Smith, C. I., and B. D. Farrell. 2005. Phylogeography of the longhorn cactus beetle Moneilema appressum LeConte (Coleoptera: Cerambycidae): was the differentiation of the Madrean sky islands driven by Pleistocene climate changes? Molecular ecology 14:3049–65.

Song, L., and L. Florea. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience 4:48.

Thompson, R. 1990. Late Quaternary vegetation and climate in the Great Basin. Pages 200–239 *in* J. Betancourt, T. Van Devender, and P. Martin, editors. Packrat Middens The Last 40,000 years of biotic change. University of Arizona Press, Tucson, AZ.

Wang, I. J. 2013. Examining the full effects of landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. Evolution 67:3403–3411.

Wang, I. J., and G. S. Bradburd. 2014. Isolation by environment. Molecular Ecology 23:5649–5662.

Wells, P. V. 1983. Paleobiogeography of Montane Islands in the Great Basin since the Last Glaciopluvial. Ecological Monographs 53:341–382.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

Winter, D. J. 2012. MMOD: An R library for the calculation of population differentiation statistics. Molecular Ecology Resources 12:1158–1160.

Wolf, J. B. W., T. Bayer, B. Haubold, M. Schilhabel, P. Rosenstiel, and D. Tautz. 2010. Nucleotide divergence vs. gene expression differentiation: comparative transcriptome sequencing in natural isolates from the carrion crow and its hybrid zone with the hooded crow. Molecular ecology 19 Suppl 1:162–75.

Zheng, X., D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328.

Appendix A

Supplemental Figures

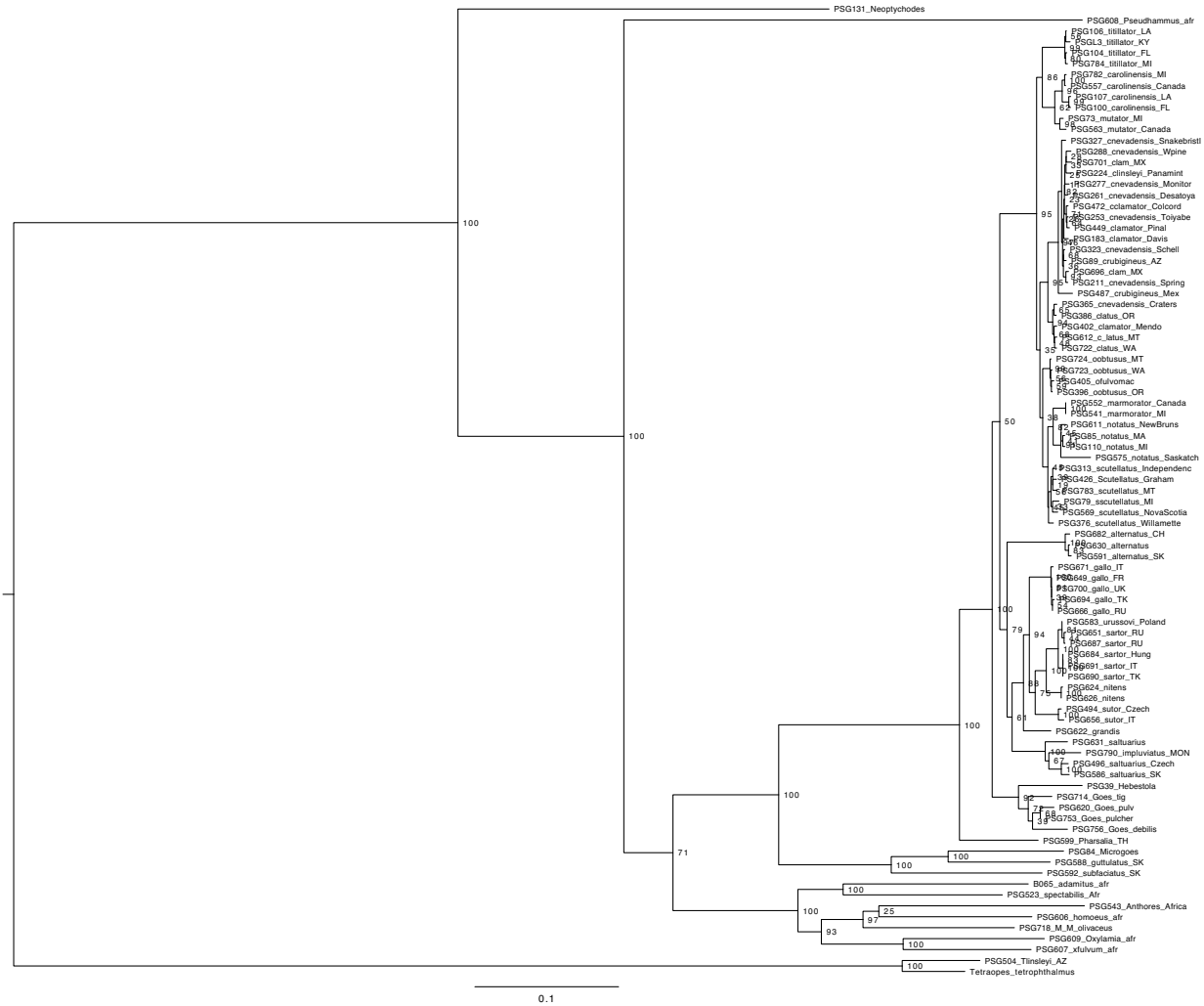

Figure A1. *Monochamus* concatenated all genes RAxML tree partitioned with *PartitionFinder* (to scale)

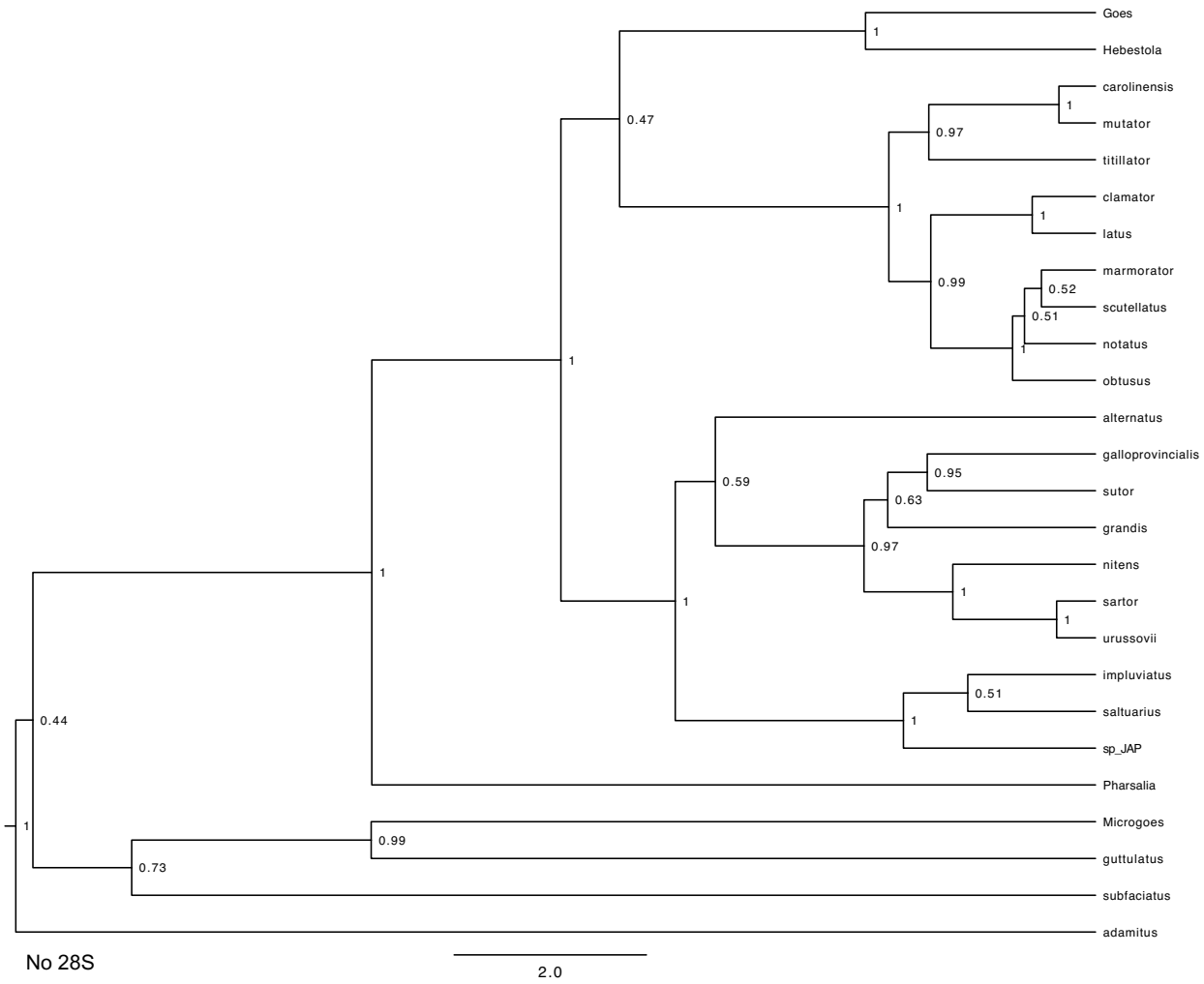

Figure A2. Concatenated StarBEAST2 species tree missing 28S

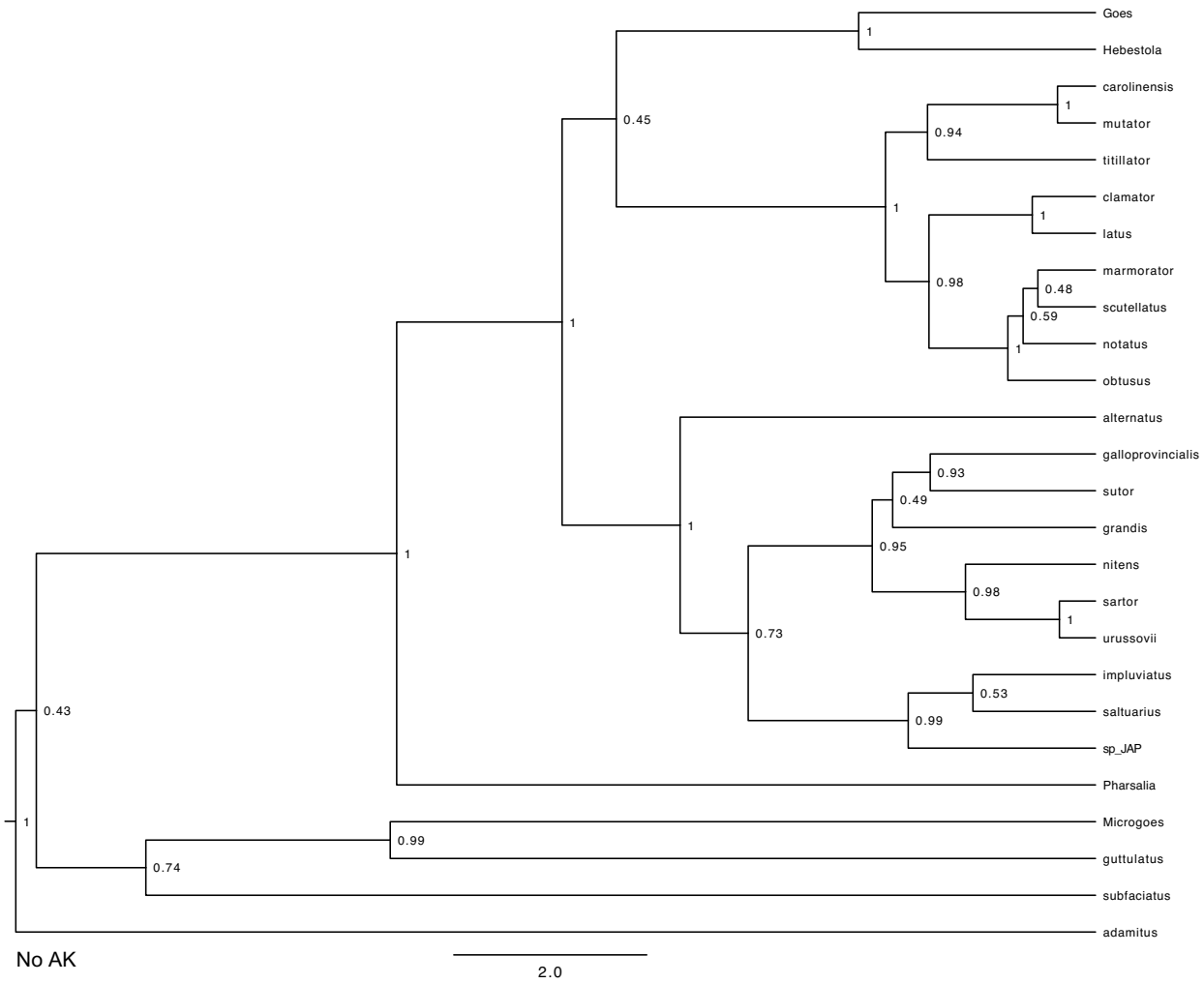

Figure A3. Concatenated StarBEAST2 species tree missing AK

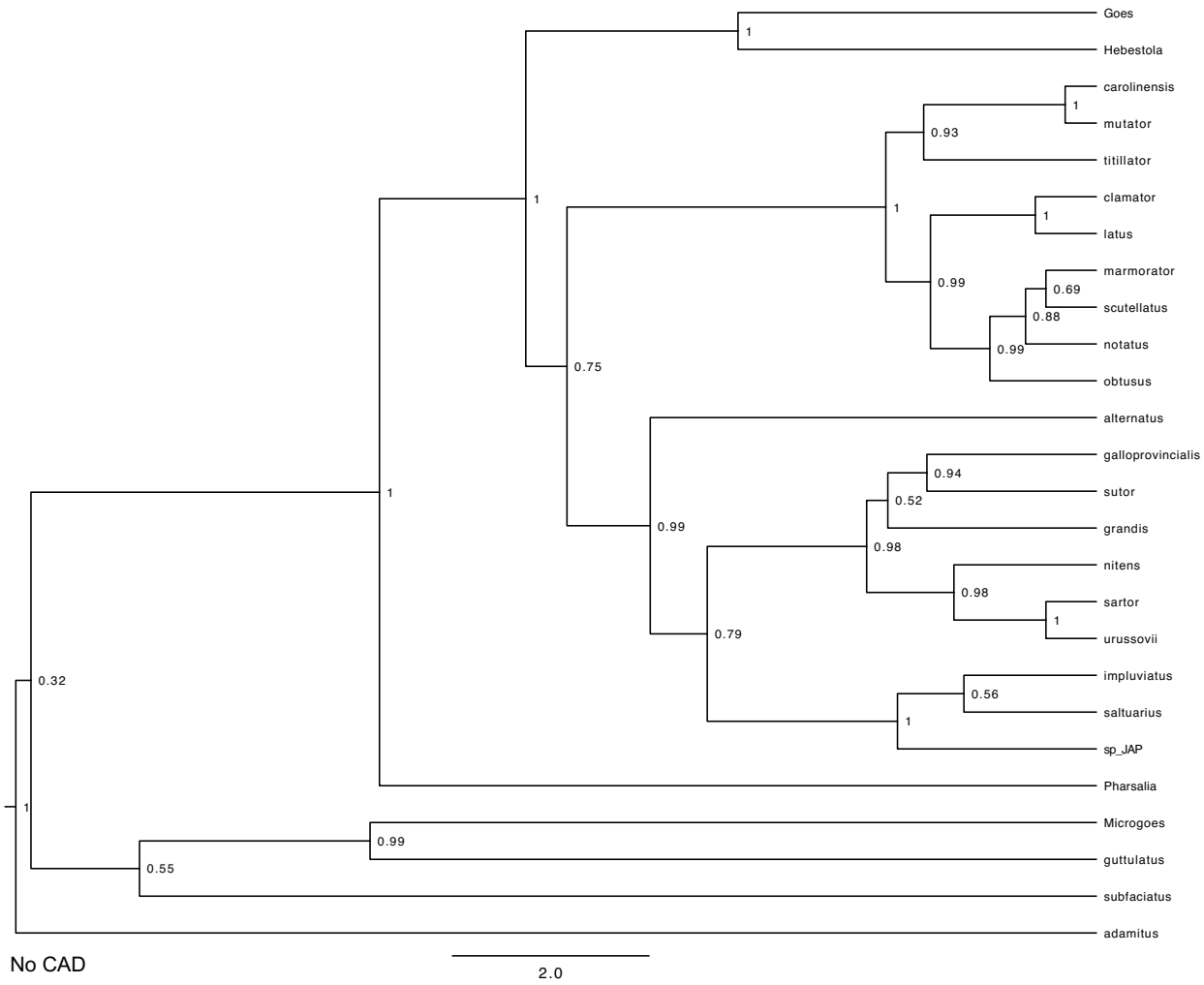

Figure A4. Concatenated StarBEAST2 species tree missing CAD

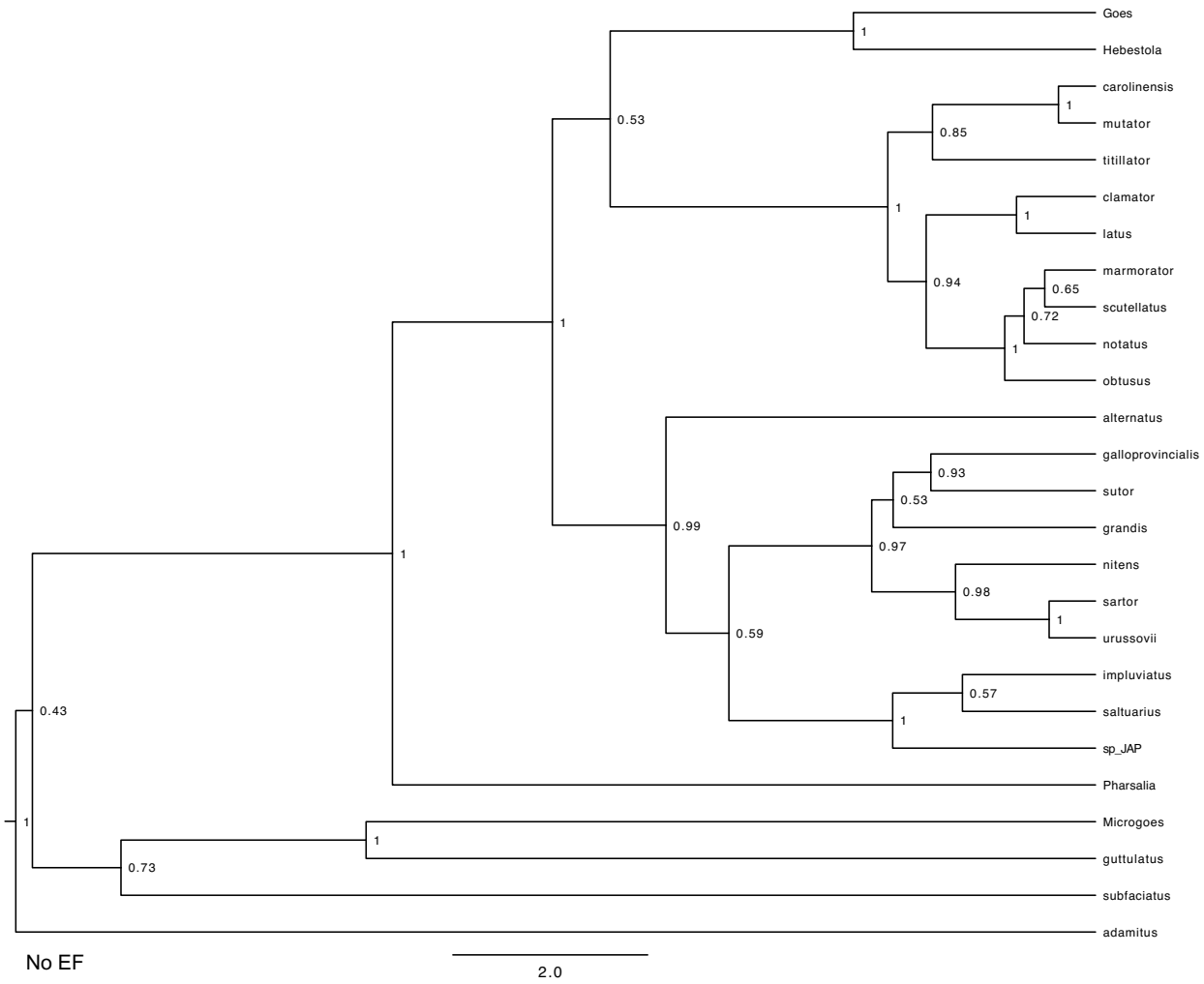

Figure A5. Concatenated StarBEAST2 species tree missing EF

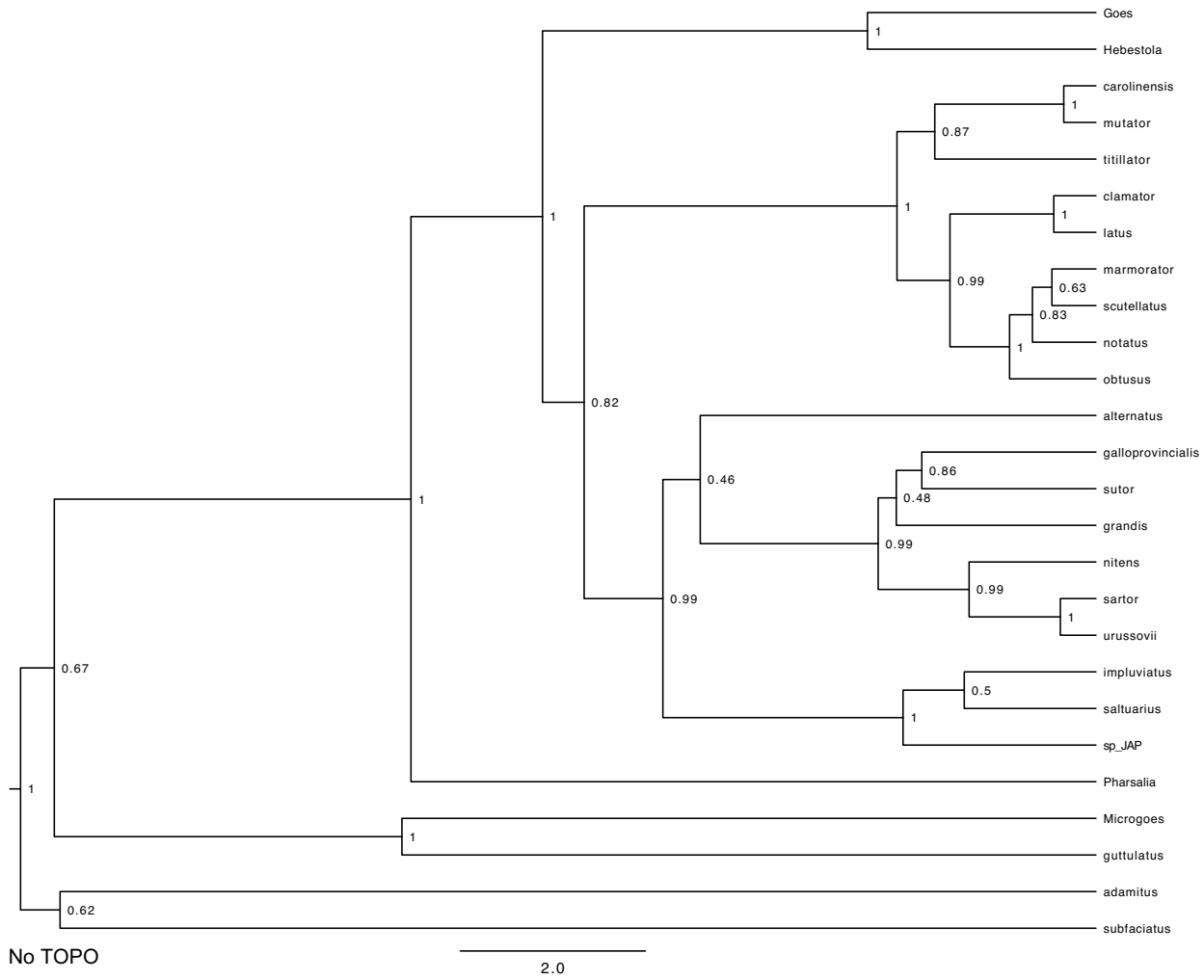

Figure A6. Concatenated StarBEAST2 species tree missing TOPO

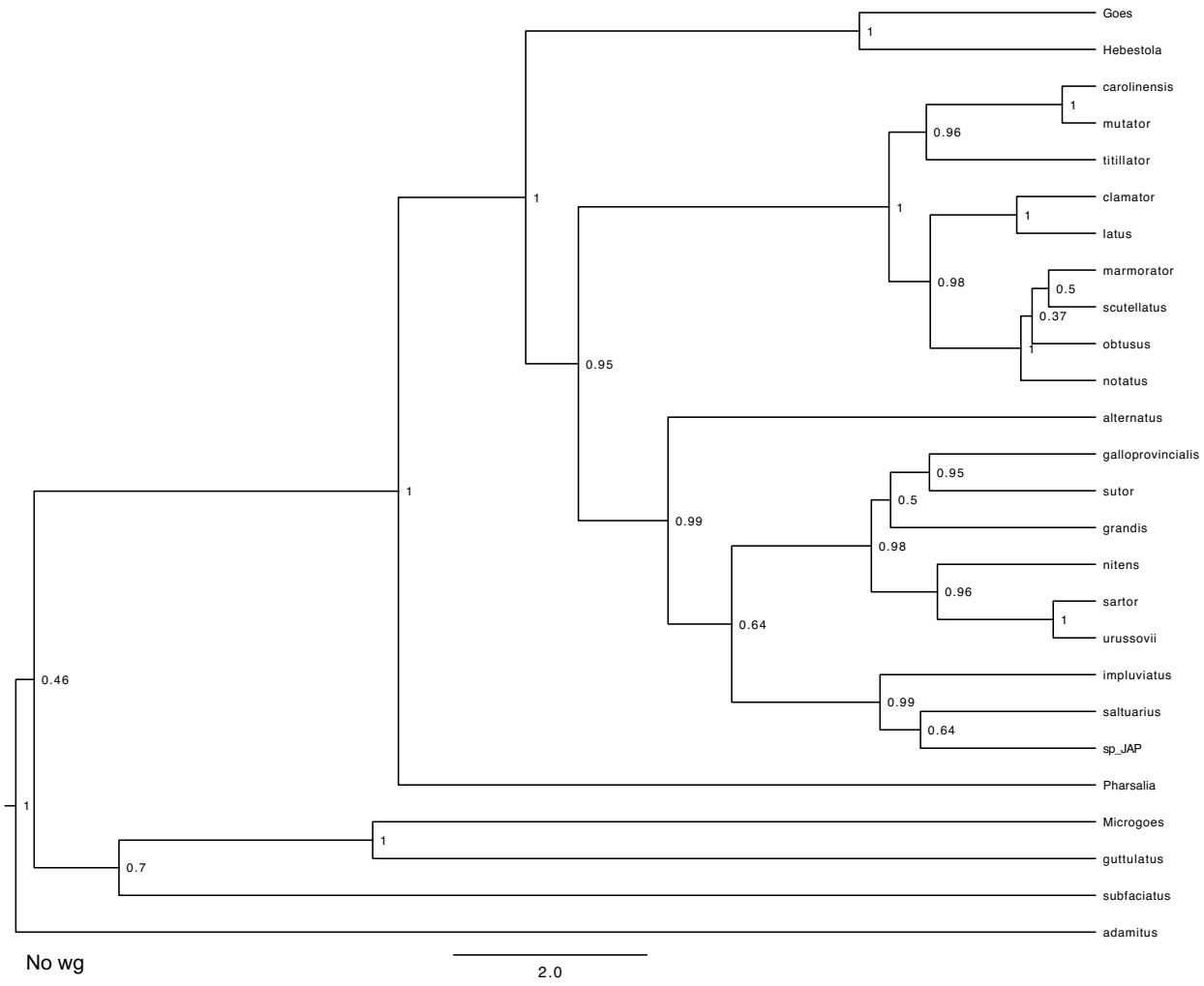

Figure A7. Concatenated StarBEAST2 species tree missing *wg*

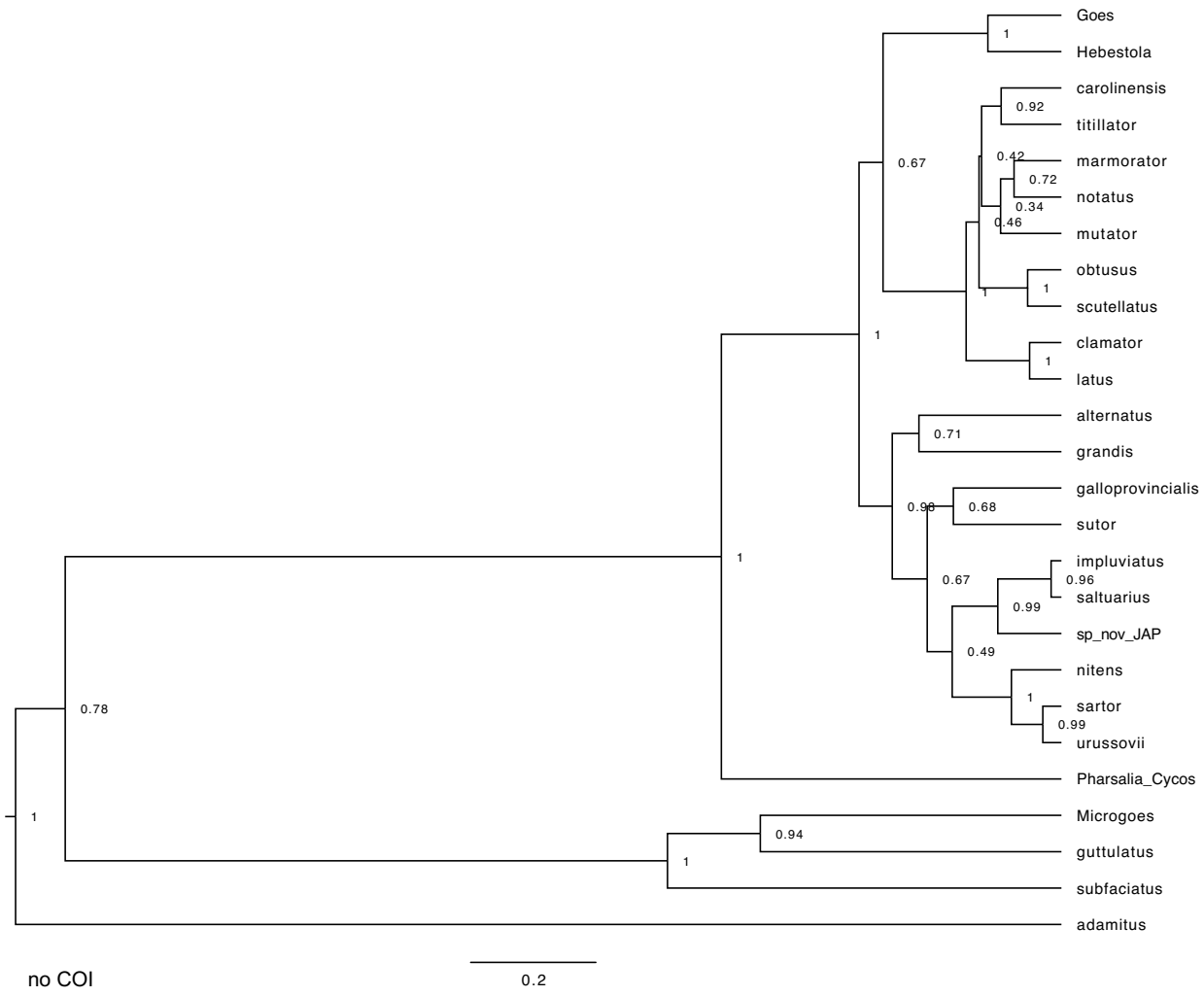

Figure A8. Concatenated StarBEAST2 species tree missing COI

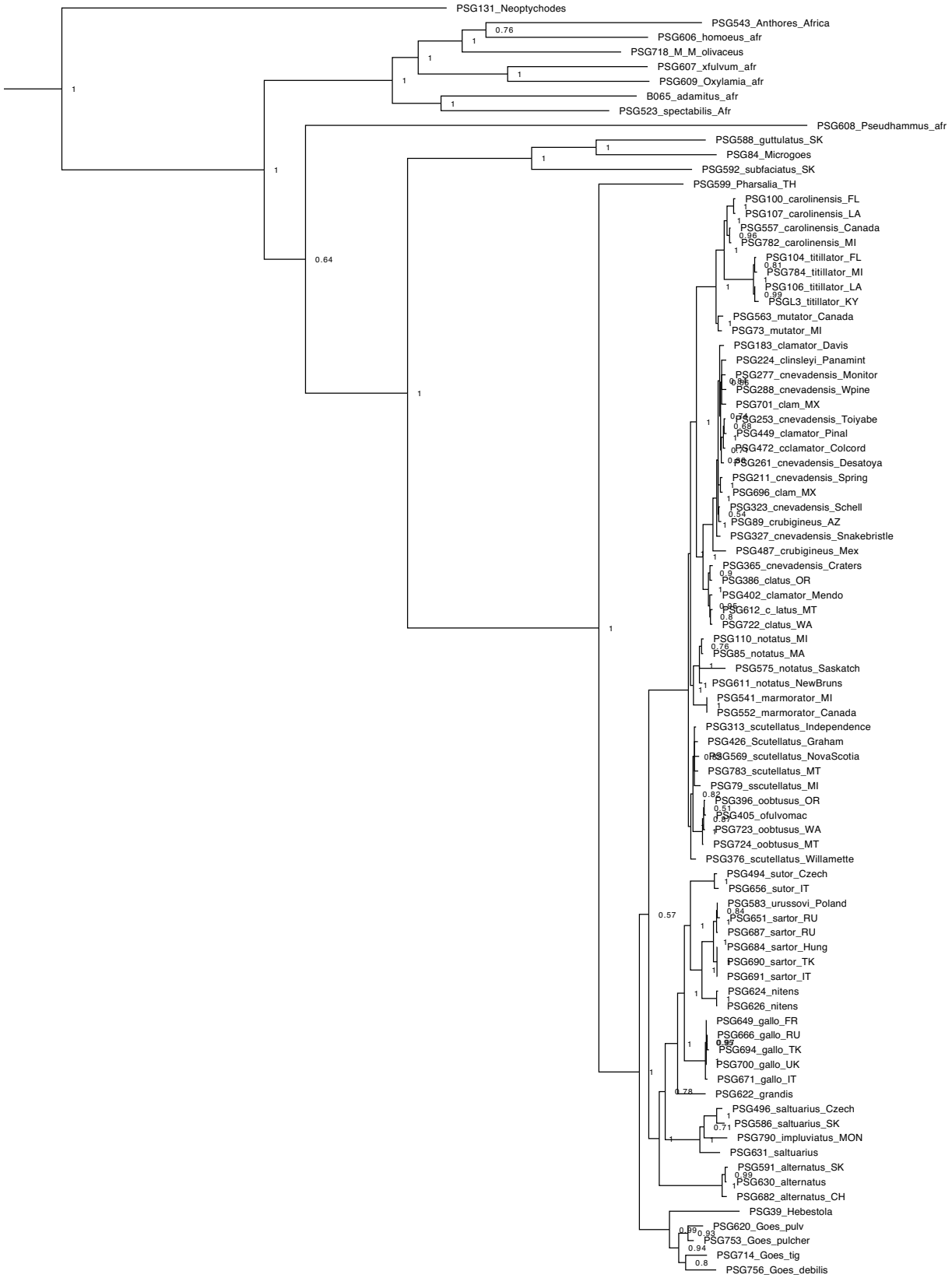

Figure A9. *Monochamus* MrBayes phylogram, using concatenated mitochondrial and nuclear dataset

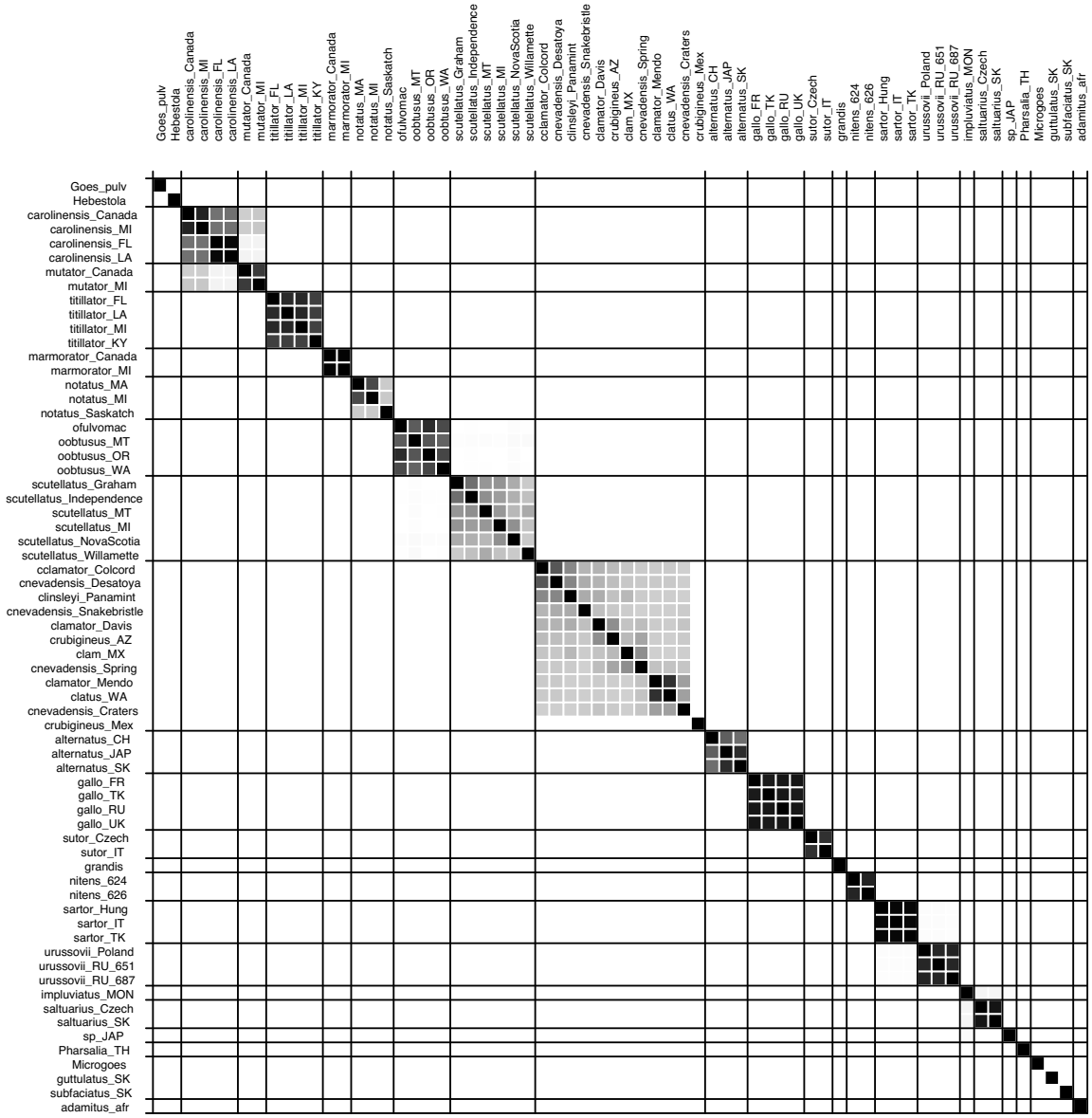

Figure A10. *Monochamus* all genes (mitochondrial + nuclear) unphased STACEY Simmatrix at 1e-5 node height cutoff

Appendix B

Supplemental Material

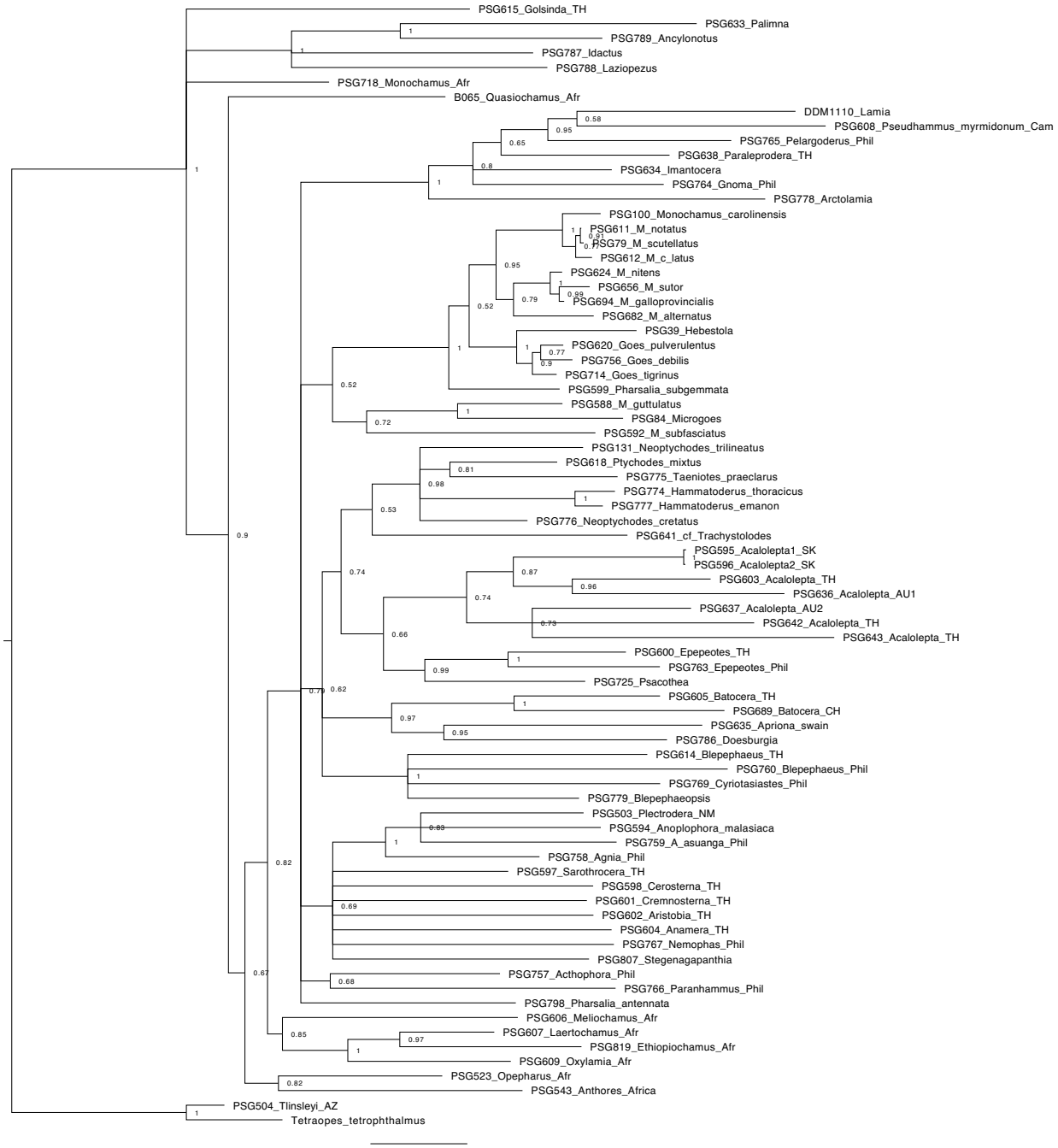

Figure B1. Lamiini tribe MrBayes COI gene tree

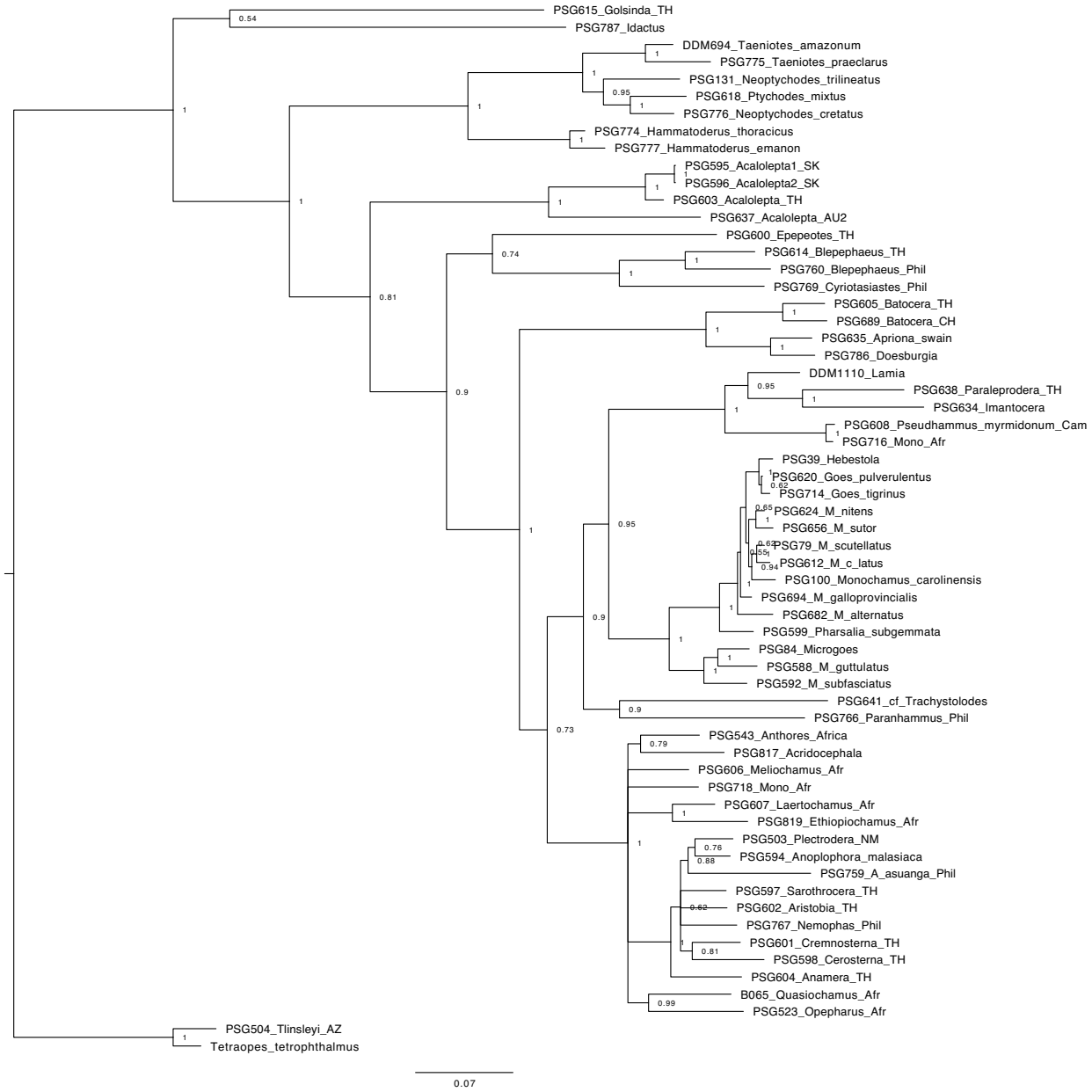

Figure B2. Lamiini tribe MrBayes CAD gene tree

Classification

Tribe LAMIINI Latreille, 1825

Lamiini Latreille (1825): 401. Type genus Lamia Fabricius, 1775 by original designation.

Synonyms: Lamiides Blanchard, 1845; Monohammidae Gistel, 1848; Pachystolaeidae Gistel, 1848; Gnomitae Thomson, 1860; Agnitae Thomson, 1864; Batoceritae Thomson, 1864; Geranitae Thomson, 1864; Hebestolitae Thomson, 1864; Morimitae Thomson, 1864; Phrissomitae Thomson, 1864; Taeniotitae Thomson, 1864; Potemnemini Aurivillius, 1922; Docohammidi Dillon & Dillon, 1959; Acridocephalidi Dillon & Dillon, 1959