UNIVERSITY OF CALIFORNIA
RIVERSIDE

AND

SAN DIEGO STATE UNIVERSITY

Species Delimitation and Biogeography of the Thorn Harvestmen (*Acuclavella*) and Their Placement Within the Ischyropsalidoidea (Arachnida, Opiliones, Dyspnoi)

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Evolutionary Biology

by

Casey H. Richart

December 2018

Dissertation Committee:
Dr. Marshal Hedin, Co-Chairperson
Dr. Cheryl Y. Hayashi, Co-Chairperson
Dr. Tod W. Reeder
Dr. Mark S. Springer

The Dissertation of Casey H. Richart is approved:

_____

_____

_____

Committee Co-Chairperson

_____

Committee Co-Chairperson

University of California, Riverside
San Diego State University

ACKNOWLEDGEMENTS

We will now discuss in a little more detail the Struggle for Existence

- Charles Darwin, 1859

I did not foresee how tumultuous my doctoral studies would be. This period of scientific learning was commensurately accompanied with intense personal learning. So many of you kindly broke my fall and stepped up beyond reasonable expectations. In the process you have made a life-long friend. I will be forever grateful. Thank you so much. I wish I was a poet instead of a biologist. Because of you, intense personal learning became personal growth. I love you all.

Thank you to Robin Keith Hedin, Marshal Hedin, Molly Hedin, and Ole Hedin, to Kevin Burns and Kevin O'Neal, and to Shahan Derkarabetian for becoming a surrogate family when I needed it most. I don't know how to thank you; I'm crying as I reflect on your love and support. Thank you for the roof, thank you for the food, thank you for the work, thank you for your love. As my San Diego community support strengthened, so too did my family back home. To my mother and stepfather, Sheryl and Mike Fairhart I could not have finished without your support. I know I don't need to explain myself, I needed some time and you provided it. To my father and stepmother Dale Richart and Jina Bailey, thank you for providing a shelter from the storm, for making me feel comfortable and at home. To my sister and brother-in-law Poppy and Ian Louthan, thank you for flying down here just to help me move, thank you for checking in on me more

than anyone, and for being patient and sympathetic. To Tyler, thank you for your loyal support and the Ph.inished mug! To Nick, your decisions to turn your life around helped me make positive decisions too; I'm proud of you, let's keep at it. To my brother and sister Jason and Allegra Lampley, thank you for being unwavering in your love and support. Teri and Barry Armstrong, Janice Chaufty, and Bill and Vicki Leonard - I clearly could have gone to any of you for as much support as I needed, thank you. Thanks to my nieces and nephews Hunter, Lily, Azalea, Bayani, Archer, and Vera - *Sit tibi semper habetis mirantibus in cordibus vestris, et multa adventures cochlea*.

Thanks to B.J. Stacey for offering me a roof and for the time we spent giving thanks to the cathedral of Mother Nature's beauty. Vanesa Margarita Ramírez Bello, Cariño, gracias por tu apoyo y por la confianza que me has dado. Dr. Salu Elizabeth John, MD was integral to getting my life back on track, thank you so much. Recently, many dear friends and family reached out with personal support, and I love them for it: Courtney Achenbach, James Barklage, Elizabeth Beckett, Mike Burger, Mercedes Burns, Jay and Heidi Carlisle, Julie Charernsuk, Kimberly Cole, Caleigh Cornell, Aghop and Armine Der-Karabetian, Niccolo Eck, Jeff Fairhart, Sam Fellows, Rebecca Godwin, Marc Hayes, Steve Herman, Miguel Hernandez-Patricio, Dovi Kacev, Paul Maier, Seabird McKeon, Tyler Kai LoRusso, Tara Luckau, Makenzie Mabry, Andre Nguyen, Cassie O'Hara, Niki O'Neal, Ketan Patel, Amy Porter, Rosalyn Price-Waldman, Ty Roach, Jason Ryberg, Jordan Satler, Emily Schisler, Rachel Shoop, Lee Simpson, Eric Stiner, Alexandra Sumarli, Jill Sundlof, Nick Vinciguerra, Michael Wall, Marlene Wagner, and Yue Zhang. Thank you to All Classical Portland and its contributors for the

solace, alleviated stress, and increased productivity, and to Igor Stravinsky for making things Rite. I would like to thank Anderson.paak for reminding me to let it go and Celebrate while I still can.

My advisers Marshal Hedin and Cheryl Hayashi are extraordinary persons, and I am deeply honored to be of their academic legacy. Marshal, you are better at life than anyone I know. Your research improves each year, your ability to stay current with theory and technology astounds me. All the while, you regularly get in the field and are among the best naturalists I have ever met. You perceive beauty in so many corners of life, and are incredibly adept at sharing this beauty with all around you - in the classroom, in the lab, even on the internet available to everyone. You are a caring and devoted husband and father. Because of your example I am better at life. I will always be grateful. Cheryl, thank you for showing me how it's done. Your words are powerful and they are always in my head. Your insight is sharp - always saying the right thing at the right time. I am almost always glad to hear it :-). Also, thanks for always referring to my grammatical and spelling errors as "typos".

I was fortunate to be surrounded by many excellent educators. I am particularly appreciative of my year at UC Riverside, and the academic atmosphere of evolution and systematics under the tutelage of David Reznick, Mark Springer, and John Gatesy. Many students from that time too worked hard at their academics, and I fondly recall debating Michael Forthman and John Hash on the topic of the day. A number of undergraduate students and advanced high school interns helped collect and manage data: thank you very much to Stas Vidyakin for taking images and sorting specimens, Alexa Feist for

characterizing transcriptome alignments and field work, Lisa Kairy for curatorial duties, Rebecca Friedrich and Teta Charernsuk for help collecting morphometric data, Yixuan Xia for georeferencing and DNA extractions, and Amy Downey for morphometric data collection and data entry. Wet lab work was facilitated with the support and tutelage of Mercedes Burns, Shahan Derkarabetian, Andrew Gottscho, James Starrett, and Megan Morris. Also, thank you to the Dinsdale Lab, Rohwer Lab, and Reeder Lab for providing access to laboratory equipment. Thank you Adrienne Fusek for tuition waivers. Thank you and in solidarity with current and former lab mates Tierney Bougie, Brendan Boyer, Allan Cabrero, Dave Carlson, Erik Ciaccio, Angela DiDomenico, Kristen Emata, Erika Garcia, Jordan Satler. I am appreciative to philosophical waxing during conversations with Nate Robinett. Jordan Satler, you always helped when I needed it, thanks bud. Thank you Sean Harrington for helping to trouble-shoot phylogenetic analyses and being receptive when I needed to bounce ideas off of someone. Shahan Derkarabetian, Timothy Shaw, and Paul Maier helped troubleshoot initiating various analyses, thank you. A big thanks to Joe Warfel who helped with the beautification of multiple presentations. Discussions with William Shear improved my appreciation for the importance of a name for the retrieval of biological information, also thanks for helping identify *Taracus* specimens for chapter 1. I thank the blogger valen for their very useful entry (http://populationgeneticsnotes.blogspot.com/2013/07/how-to-run-structure-structure.html) on using STRUCTURE and downstream applications. Axel Schönhofer and Jochen Martens provided *Ischyropsalis* genomic material. Shahan Derkarabetian, Jim Starrett, and Amy Hubert aided in RNA extractions, and S. Derkarabetian and Dave

Carlson aided in transcriptome assembly. This research was improved via discussions with Kevin Burns, Dean Leavitt, John Gatesy, Rob Meredith, Tod Reeder, Bill Shear, Mark Springer, and J. Starrett. I thank Jean Valjean for asking tough questions. Specimens were collected with aid from D. Carlson, S. Derkarabetian, Damian Elias, Marc P. Hayes, S. Huber, Robert Keith, Paul Marek, Steve Merkley, T. Novak, Adrienne Richart, Dan Richart, Jordan Satler, A. Schönhofer, L. Slana, J. Starrett, B. Shear, and Jeff Underwood. Chapter 1 was improved through comments from Mercedes Burns, Allan Cabrero, D. Carlson, S. Derkarabetian, K. Emata, J. Starrett, and two anonymous reviewers. Chapter 1 was funded in part by the University of California Riverside Newell Travel Award, and the American Arachnological Society Vincent Roth Fund for Research in Systematics. Some data collection was supported by a California State University Program for Education and Research in Biotechnology (CSUPERB) grant to M. Hedin.

DEDICATION

I dedicate this dissertation to my teachers, for they have defined my perception of beauty, and to my students, where these definitions descend.

ABSTRACT OF THE DISSERTATION

Species Delimitation and Biogeography in the Thorn Harvestmen (*Acuclavella*) and Their Placement Within the Ischyropsalidoidea (Arachnida, Opiliones, Dyspnoi)

by

Casey H. Richart

Doctor of Philosophy, Graduate Program in Evolutionary Biology
University of California, Riverside, joint with San Diego State University, San Diego
December 2018
Dr. Marshal Hedin, Co-chairperson
Dr. Cheryl Hayashi, Co-chairperson

This dissertation lays the foundation for a research program that highlights the important contribution that poorly-dispersing short-range endemic (SRE) organisms can have on advancing empirical and theoretical evolutionary biology. Here, SREs are represented by members of the harvestmen (Opiliones) superfamily Ischyropsalidoidea. In Chapter 1, phylogenetic analyses of an ancient rapid diversification at the base of Ischyropsalidoidea uncovers considerable incongruence across gene trees. Incongruent gene trees are not recovered in frequencies expected from a simple multispecies coalescent model, and incomplete lineage sorting is rejected as the sole contributor to gene tree conflict. In Chapter 2, an integrative taxonomy tests existing, and puts forward novel species-level taxonomic hypotheses within the *Acuclavella* species endemic to the inland temperate rainforest of the U.S. northern Rocky Mountains. Incongruence among

taxonomic disciplines suggests rapid morphological evolution across multiple recent diversifications. Results suggest at least two undescribed species exist within inland populations of *Acuclavella*. Alternative combinations of incongruence across taxonomic disciplines possibly represent different stages of the speciation continuum, and suggest that different evolutionary mechanisms underlie speciation in thorn harvestmen, thus highlighting the importance of integrative species delimitation for evolutionary research. The disjunct inland temperate rainforest has become a model for recent biogeography, with a complex and recent evolutionary history suggesting that this inland forest was a refuge from Pleistocene glaciation that was compartmentalized into variously isolated pockets during the last glacial maximum (LGM). In Chapter 3, genomic-scale data are used to evaluate the evolutionary consequences of the LGM. This research finds that rivers were a major cause of refugia compartmentalization, with populations and species structured by rivers well before the LGM. Additionally, a negative correlation between elevation and heterozygosity suggests that populations were confined to lower valleys during the LGM with post-glacial colonization of higher elevations. The combined effects of rivers and glaciers paints a picture of a complex Pleistocene glacial refugia system during the LGM. This relatively high-resolution inference of past evolutionary events within an established biogeographic model system points towards the undervalued importance of using poorly-dispersing short-range endemic taxa for biogeography and phylogeography.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

An appreciation for biodiversity benefits society in myriad ways. Obvious examples include food, pharmaceuticals, pesticides, and companions. The greatest inventor of all time - natural selection - has provided endless adaptations and diversity upon which we may broaden our understanding, make discoveries, and apply them to the improvement of our lives as well as Earth. We cannot know what our future explorers may need to improve ourselves. It could be a protein, or a metabolic pathway, or a mechanical adaptation that inspires an engineer. We have already lost so much. Many books have been burned that haven't even been read - species that have succumbed to human-caused extinct that were never and will never be documented. Those adaptations - that diversity - will not be available for our advancement. The flood tide of the current mass extinction is not likely to ebb anytime soon. We need to read. We need to read broadly, thoroughly, and quickly. The contemporary documentation and conservation of biodiversity is one of the best investments we can make to our descendents.

Currently, the majority of conservation efforts focus on the species level of biodiversity, with habitats and intraspecific diversity also important considerations. Thus, this measure of biodiversity, the number and welfare of species, is integral to the quality of our legacy. Concurrently, societal prioritization and funding of conservation efforts lags behind its value. Conservation biologists must make decisions regarding resource allocation. Species are going to fall through the cracks, but we need to do what we can in the short-term to minimize the number that do. These decisions are usually based on a

species. However, what taxonomists designate as species is a working hypothesis to be tested as new data become available. Sometimes a single species is named twice, sometimes a named species is comprised of multiple lineages. The taxonomist is tasked with the maintenance and evaluation of species hypotheses. There has never been a more important time to fund and invest in taxonomy, for taxonomists are the book readers. What they learn is organized and made available to as many facets of society as possible. There has also never been a time when taxonomists need to be so broadly versed. Still incumbent on the taxonomist is the curation of a research collection, as well as writing and disseminating information. Today, taxonomists are expected to be proficient in multivariate statistics, phylogenetics, and population genetics.

Both theoretically and empirically, speciation is the protracted process of one lineage becoming two. This gradual, generation-by-generation divergence makes the delimitation of incipient species a very difficult problem. In order to put forward robust species hypotheses they must be scrutinized using many lines of evidence based robustly on theory. Species bridge the gap between phylogenetics and phylogeography, thus it is important for the alpha taxonomist to be well-versed in both phylogenetic and population genetic techniques. The motivation of this research is to robustly delimit species, organize them into the tree of life, and to document intraspecific biodiversity.

Another motivation is to highlight our complete lack of knowledge concerning the diversity of leaf-litter dwelling organisms. Here, the focus is on our fundamental lack of understanding of the species- and population-level diversity of these organisms. Further, we have no idea what is going on regarding the community ecology of these diverse

communities. For example, leaf-litter denizens include myriad decomposers, organisms that chemically defend themselves, crawlies that can pass through wood and compact soils, any or all of which could benefit society even presently. In order to understand these processes we need to be able to communicate about who the players are. Thus, the first step is naming species. Once we understand who the species are, we can ask downstream questions. A frequent inquiry regarding specific and intraspecific biodiversity is the biogeographical and phylogeographical processes that led to this diversity. This dissertation also highlights the fine-scale resolution that poorly-dispersing litter-dwelling organisms have for these evolutionary historical inquiries.

This dissertation lays the foundation for a research program that seeks to prioritize conservation efforts through the identification, delimitation, organization, and dissemination of biodiversity knowledge. Chapter 1 focuses on phylogenetic analyses and the organization of species into high-level clades. Chapter 2 focuses on robust and objective species delimitation, and the theoretical and empirical approaches that help assure that biodiversity is discovered. Lastly, Chapter 3 highlights the insight that can be attained from documenting allopatric boundaries and intraspecific biodiversity. This foundation sets up future research that explores the power of genetic diversity to predict biodiversity, and the extent to which poorly-dispersing litter-dwelling organisms undergo a composite mode of speciation, co-distributed concordance, and the identification of unique endemics and areas of endemism.

CHAPTER 1

Phylogenomic analyses resolve an ancient trichotomy at the base of Ischyropsalidoidea (Arachnida, Opiliones) despite high levels of gene tree conflict and unequal minority resolution frequencies

Phylogenetic resolution of ancient rapid radiations has remained problematic despite major advances in statistical approaches and DNA sequencing technologies. This chapter reports on a combined phylogenetic approach utilizing transcriptome data in conjunction with Sanger sequence data to investigate a tandem of ancient divergences in the harvestmen superfamily Ischyropsalidoidea (Arachnida, Opiliones, Dyspnoi). Analyses on Sanger sequences resolve nodes within and between closely related genera, and RNA-seq data from a subset of taxa resolves a short and ancient internal branch. Several analytical approaches are used to explore this succession of ancient diversification events, including concatenated and coalescent-based analyses and maximum likelihood gene trees for each locus. The robustness of phylogenetic inferences is evaluated using a randomized locus sub-sampling approach, and find congruence across these methods despite considerable incongruence across gene trees. Incongruent gene trees are not recovered in frequencies expected from a simple multispecies coalescent model, and incomplete lineage sorting is rejected as the sole contributor to gene tree conflict. These approaches attain robust support for higher-level phylogenetic relationships within Ischyropsalidoidea.

Harvestmen (Opiliones) are among the most species-rich arachnid orders (Harvey 2002), and have an ancient diversification history (Hedin et al. 2012, Sharma and Giribet 2014). Within Opiliones there are 46 recognized families, approximately 1500 genera, and more than 6500 described species (Machado et al. 2007, Kury et al. 2014). Phylogenomics of higher-level relationships within Opiliones (Hedin et al. 2012) has found strong support for four primary clades (suborders Cyphophthalmi, Laniatores, Dyspnoi, and Eupnoi), with recent amendments based on novel fossil discoveries (Giribet and Sharma 2015). This research builds upon Hedin et al. (2012) to analyze relationships within the Dyspnoi superfamily Ischyropsalidoidea. The Ischyropsalidoidea are confined to the northern hemisphere, with the age of the root estimated to be as recent as 137 MYA (Schönhofer et al. 2013) or as old as 240–360 MYA (Sharma and Giribet 2014). Currently, 85 species are classified into seven genera: *Ischyropsalis*, *Sabacon*, *Taracus*, *Ceratolasma*, *Acuclavella*, *Hesperonemastoma*, and *Crosbycus* (Kury, 2013). The superfamily is defined on the basis of genitalic characters (Martens 1976), palpal morphology (Martens et al. 1981), and by having metapeltidial sensory cones (Shear 1986, though see Shultz 1998). Each genus is morphologically distinct (Figure 1.1), and there has been little controversy regarding their respective monophyly. In fact, monogeneric families have been proposed for a number of genera (e.g., Dresco 1970, Martens 1976, Shear 1986, Schönhofer 2013), but this taxonomic solution was criticized by Gruber (1978). Conversely, family level hypotheses within Ischyropsalidoidea have been problematic, with two of the last three non-monogeneric familial hypotheses

(Sabaconidae of Giribet et al. 2010, Taracidae of Schönhofer 2013) failing to identify diagnostic morphological synapomorphies.

I. INTRODUCTION

Harvestmen (Opiliones) are among the most species-rich arachnid orders (Harvey 2002), and have an ancient diversification history (Hedin et al. 2012, Sharma and Giribet 2014). Within Opiliones there are 46 recognized families, approximately 1500 genera, and more than 6500 described species (Machado et al. 2007, Kury et al. 2014). Phylogenomics of higher-level relationships within Opiliones (Hedin et al. 2012) has found strong support for four primary clades (suborders Cyphophthalmi, Laniatores, Dyspnoi, and Eupnoi), with recent amendments based on novel fossil discoveries (Giribet and Sharma, 2015). This research builds upon Hedin et al. (2012) to analyze relationships within the Dyspnoi superfamily Ischyropsalidoidea. The Ischyropsalidoidea are confined to the northern hemisphere, with the age of the root estimated to be as recent as 137 MYA (Schönhofer et al. 2013) or as old as 240–360 MYA (Sharma and Giribet 2014). Currently, 85 species are classified into seven genera: *Ischyropsalis*, *Sabacon*, *Taracus*, *Ceratolasma*, *Acuclavella*, *Hesperonemastoma*, and *Crosbycus* (Kury 2013). The superfamily is defined on the basis of genitalic characters (Martens 1976), palpal morphology (Martens et al. 1981), and by having metapeltidial sensory cones (Shear 1986, though see Shultz 1998). Each genus is morphologically distinct (Figure 1.1), and there has been little controversy regarding their respective monophyly. In fact,

monogeneric families have been proposed for a number of genera (e.g., Dresco 1970, Martens 1976, Shear 1986, Schönhofer 2013), but this taxonomic solution was criticized by Gruber (1978). Conversely, family level hypotheses within Ischyropsalidoidea have been problematic, with two of the last three non-monogeneric familial hypotheses (Sabaconidae of Giribet et al. 2010, Taracidae of Schönhofer 2013) failing to identify diagnostic morphological synapomorphies.

Systematics has moved into an era where phylogenetic hypotheses are being resolved at an unprecedented rate. For example, well-studied systems, such as mammalian interordinal relationships, now contain few controversial nodes (e.g., Murphy et al. 2001, Meredith et al. 2011). This development is due in part to the arrival of next-generation sequencing (NGS) technologies and continued advance in statistical phylogenetics. With the ability to generate matrices containing hundreds of loci (e.g. Hedin et al. 2012, Faircloth et al. 2012), NGS data have both supported preexisting hypotheses, and recovered novel taxonomic hypotheses that are robustly supported. Furthermore, NGS technologies have allowed for the identification of rapid, ancient radiations (e.g., McCormack et al. 2013, Teeling and Hedges 2013). These radiations are notoriously hard to resolve (e.g. Faircloth et al. 2012, Springer and Gatesy 2014), with short internal branches that are in part characterized by high levels of gene tree conflict. Gene trees can conflict with a species tree for numerous reasons (Maddison 1997, Maddison and Knowles 2006, Degnan and Rosenberg 2009) including undetected paralogy, recombination, hybridization, saturation, and long-branch attraction (LBA). A major source of gene tree conflict inherent to rapid diversifications is coalescent

stochasticity – the random sorting of ancestral polymorphisms across successive speciation events (Kingman 1982, Degnan and Rosenberg 2009). This phenomenon is known as incomplete lineage sorting (ILS), and its occurrence is expected to increase as a function of shorter internal branch lengths and larger ancestral population sizes (Maddison 1997). ILS can occur to such an extent that the most likely gene tree is incongruent with the species tree, a situation that has been defined as an "anomaly zone" (Degnan and Rosenberg 2006).

A problem inherent to molecular phylogenetic reconstruction of ancient and rapid successive diversification events is that short internal branches do not provide enough time for slowly evolving loci to accumulate informative substitutions, whereas more rapidly evolving loci accumulate homoplastic substitutions along descending long branches (Regier et al. 2008). Such internal branches can have so little phylogenetic signal that even small amounts of non-phylogenetic signal can yield support for an incorrect phylogeny (Huelsenbeck and Hillis 1993, Swofford et al. 2001, Philippe et al. 2011), and this can occur to such an extent that it is positively misleading (Huelsenbeck and Hillis 1993, Bull et al. 1993). Therefore, although the selective use of slowly evolving coding regions (nucleotides or amino acids) has been a successful approach for reconstructing the backbone of numerous higher level phylogenies (e.g., Iwabe et al. 1989, Hedin et al. 2012, Zhang et al. 2012, Lang et al. 2013, Raymann et al. 2014), these loci are not expected to contain enough informative characters to resolve the branching order of ancient internal nodes across short branches, and simply adding more data does

not guarantee that analyses will resolve the correct topology (Swofford et al. 2001, Philippe et al. 2011).

Longer loci have been shown to improve phylogenetic inference in situations where long terminal branches relative to short internal branches cause "zones" of inconsistent estimation (Swofford et al. 2001). Accordingly, independent loci have traditionally been concatenated into a supermatrix, with the assumption that this will allow for the emergence of hidden support, or the increased support for a clade relative to the sum of support for the clade when data partitions are analyzed separately (Gatesy et al. 1999). The theoretical argument against the supermatrix approach is that recombination and coalescent stochasticity result in genes having different evolutionary histories, and that concatenating these loci into a supermatrix (which in effect treats all data as a single locus) violates the assumption of recombination (Kubatko and Degnan 2007). Simulation studies under these conditions (short internal branches with high levels of gene tree conflict) have supported this contention, showing that concatenation can result in support for incorrect topologies (Seo 2008), with support increasing as more loci are added (Kubatko and Degnan 2007). To address these concerns, methods of phylogenetic inference have been developed under multispecies coalescent models that co-estimate gene trees, divergence times, population sizes, and a species tree from multiple unlinked loci (e.g., BEST, Liu and Pearl 2007, *BEAST, Heled and Drummond 2010). Simulation studies have shown that fully-parametric multispecies coalescent methods can be highly accurate even with high levels of gene tree incongruence (Liu and Edwards 2009), and outperform supermatrix methods (Heled and Drummond 2010).

Currently, the major shortcoming of fully-parametric coalescent analyses is that the very large parameter space is too computationally demanding to be applied to analyses with a large number of loci and/or taxa (Edwards et al. 2007, Liu et al. 2009, Bayzid and Warnow 2013, O'Neill et al. 2013). As such, many phylogenomic analyses use partially-parametric coalescent analyses such as STAR (Liu et al. 2009) or MP-EST (Liu et al. 2010). These partially-parametric coalescent analyses use gene trees constructed in isolation as input for phylogenetic inference, and assume that gene trees are correctly inferred and all gene tree discordance is due to ILS. However, gene trees can conflict at ancient and rapid divergences for numerous reasons (see above). Further, careful re-analyses of partially-parametric coalescent results (e.g., the data from Song et al. 2012) have shown that such methods can provide high support for likely erroneous topologies when the assumption of correctly inferred gene trees is violated (Springer and Gatesy 2016). Simulations have long shown that phylogenetic inference of deep divergences associated with short internal branches produces a large number of incorrect gene trees simply do to sampling error (e.g., Huelsenbeck and Hillis 1993, Swofford et al. 2001); empirically this problem is more extensive with shorter loci (Gatesy and Springer 2014). As such, there is continued debate as to which phylogenomic methods are preferable for resolution of ancient rapid diversifications (Liu et al. 2010, Leaché and Rannala 2011, Song et al. 2012, Lemmon and Lemmon 2013, Patel et al. 2013, Gatesy and Springer 2014, Lanier and Knowles 2015, Springer and Gatesy 2016, Edwards et al. 2016).

In this paper a combined phylogenetic approach was employed utilizing transcriptome data in conjunction with Sanger sequences (e.g., Leaché et al. 2014b) to

analyze two nearly independent phylogenetic matrices, including an "expanded panel" that contains 14 loci for 12 ingroup terminals (with some missing data), and a "transcriptome panel" that contains 672 loci for 3 ingroup terminals (no missing data). Preliminary phylogenetic analysis of the expanded panel identified a weakly supported topology deep in the ischyropsalidoid species tree. Therefore, transcriptome data was generated from ischyropsalidoid exemplars descending from these ancient and rapid diversifications to specifically target these problematic nodes. With this combined strategy a robustly supported phylogeny for every node sampled within Ischyropsalidoidea to the rank of genus is attained. Analyses identify a near trichotomy at the base of the superfamily that has resulted in high levels of gene tree incongruence, and show that the minority resolution frequencies of alternative topologies are unequal. For this ancient and short internal branch multiple phylogenetic methods are compared that are congruent in their support for a topology not previously recovered for Ischyropsalidoidea. Further, analyses of this dataset suggest that the supermatrix approach recovers the agreed upon phylogeny with fewer loci and higher support than do partially-parametric coalescent analyses.

II. MATERIALS AND METHODS

(1) Primer design, PCR, and Sanger sequencing

Protein-coding genes annotated as single-copy single-exon in Ixodes scapularis, a well-annotated arachnid genome, were downloaded and filtered from VectorBase

(http://iscapularis.vectorbase.org/). BLAST was used to query these loci against three published transcriptome assemblies (*Hesperonemastoma*, *Ortholasma*, *Trogulus*; Hedin et al. 2012) to generate alignments for PCR primer design. PCR primers were manually designed based on these alignments in Geneious Pro 5.5 (Kearse et al. 2012) and characterized using Primer3 (Rozen and Skaletsky 2000). Primers were tested against *Hesperonemastoma*, *Ortholasma* and *Trogulus* (*HOT*) genomic DNA extractions, and primer combinations successful on any member of the *HOT* panel were then tested on an expanded panel of ischyropsalidoid genera (*Sabacon*, *Taracus*, *Acuclavella*, *Ceratolasma*, *Ischyropsalis*, and an additional *Hesperonemastoma*). In addition to newly designed loci, the expanded panel and outgroups (*Ortholasma* and *Trogulus*) were amplified for gene regions previously used at deeper levels in Opiliones. These included EF-1a (Hedin et al. 2010), 18S and 28S (Giribet et al. 1999, Shultz and Regier 2001), COI (many authors, e.g., Richart and Hedin 2013, Derkarabetian and Hedin 2014), polII (Shultz and Regier 2001), and wingless (Wnt2, Richart and Hedin 2013).

PCR primers were designed by comparing alignments of *Hesperonemastoma*, *Ortholasma*, and *Trogulus* (*HOT*) transcripts derived from published transcriptome assemblies (Hedin et al. 2012). VectorBase (http://iscapularis.vectorbase.org/) was used to identify Ixodes protein-coding sequences that are both single-copy and single-exon (SCSE), and over ≥ 375 aa, resulting in a protein set of 360 Ixodes loci. This set of proteins was used to perform a custom BLAST search against the *HOT* transcripts using tblastn in Geneious Pro 5.5 (Kearse et al. 2012). Liberal parameters were used (max hits = 20, max e value 1e-1) in order to evaluate as many alignments as possible. Candidate

loci were further filtered by only including those: 1) with BLAST hits from all taxa (56 loci excluded, 13 loci with no BLAST hit from any transcript); 2) with no evidence of paralogs (i.e., only single non-overlapping BLAST hits along an alignment per taxon, 147 loci excluded), and 3) with BLAST hits > 400 aa from *Hesperonemastoma* and at least one of the other two transcripts (72 loci excluded). Occasionally, multiple transcripts for a single taxon were adjacent or minimally overlapping (<31 bp) in alignments and did not contain ambiguities within overlapping regions; these transcripts were manually merged into a super-transcript. These filtering criteria were not mutually exclusive, excluding a total of 237 loci and retaining 123 candidate loci.

PCR primers for Sanger sequencing were developed manually in Geneious on a subset of 52 loci selected for length (> 1000 bp). Annotated primers were generally between 21 and 27 bp, with up to four degenerate sites per primer. The *Hesperonemastoma* transcript anchored primer design, with degenerate nucleotides encompassing one or both of *Ortholasma* or *Trogulus* transcripts. Forward and reverse annotated primers flanking 400-1200 bp of protein-coding sequence were characterized in Geneious Pro using the Primer3 application (Rozen and Skaletsky 2000). Primers were filtered, excluding those with: GC content < 40%; hairpin > 5.0; primer-dimer > 5.0; pair hairpin > 5.0; pair primer-dimer > 5.0; and pair melting temperature difference (Tm) > 5°. Primers compatible with an optimal number of complimentary primers were preferred. Primers designed on 34 alignments passed these filters, and a total of 196 oligonucleotides (99 forward, 97 reverse) were synthesized by Bioneer (us.bioneer.com).

Initial PCR primer testing targeted the *HOT* panel, using genomic DNA from specimens collected from localities near where the individuals used for the transcriptome sequencing were obtained. These localities did not always coincide, and in the case of *Hesperonemastoma* and *Ortholasma* specimens were utilized from more than one locality (Appendix A). 50 μM PCR reactions included: 2 μl of each primer at 10 μM, 1 μl of genomic DNA, 1 μl of dNTPs, 5 μl of 10× buffer, 0.1 μl of Taq, and 39 μl dH2O. PCR cycle conditions were: 94°C for 5 min; 40× (94°C for 1 min, 54°C for 1 min, 72°C for 1 min); 72°C for 5 min. Unsuccessful amplifications were repeated with the same conditions except for annealing temperatures of 50° and 58°C. Primer combinations that successfully amplified expanded panel exemplars are available in Table 1.1. Product purification used the Bioneer AccuPrep® PCR Purification Kit with a slight modification to their standard short protocol (step 7, increased time for drying by centrifugation to 3 minutes; step 8, increase wait time for elution to 5 minutes). Sequencing reactions of purified products used PCR primers, and were Sanger sequenced in both directions at the Genomics Core Facility, University of California, Riverside.

(2) Expanded panel phylogenetics

To evaluate ischyropsalidoid intergeneric relationships an expanded panel was targeted comprised of six ischyropsalidoid genera (see above), including two species from each genus, plus outgroups. The intrageneric sampling scheme targeted species spanning the root node of each genus with the intention of subdividing long branches. This sampling was informed by previous research in *Acuclavella* (Richart and Hedin

2013), *Sabacon* (Schönhofer et al. 2013), *Ischyropsalis* (Schönhofer et al. 2015), and *Hesperonemastoma* (unpublished: Richart, Hayashi, and Hedin). Exemplars of *Taracus* and *Ceratolasma* were chosen from distant localities within their respective geographic distributions. Original sequence data were augmented with GenBank sequences. Also, the 14 OTUs in the expanded matrix were occasionally represented by multiple intraspecific individuals or relatively closely related species (Table 1.2). Expanded panel specimens were field-collected and stored at 80° C in 100% EtOH (Vink et al. 2005) with the exception of *Ischyropsalis* which was preserved in a urea buffer (Asahida et al. 1996). All extractions were conducted using the Qiagen DNeasy Blood & Tissue Kit, per manufacturer's protocol; most extractions were performed using half of a bilaterally divided individual, with the other half saved as a voucher.

Expanded panel alignments were generated from newly developed markers (eight loci) and six previously-used loci (see above). Some alignments were further populated using transcriptome-derived sequence data, and trimmed to the start at the nearest first base pair of an open reading frame. GenBank accession numbers, data matrix coverage, and alignment lengths are provided in Table 1.3. All alignments were conducted in Geneious using MAFFT 6 (Katoh et al. 2002), and regions of alignment uncertainty were removed with GBlocks 0.91b (Castresana 2000). Partitions and models of evolution were jointly estimated using PartitionFinder 1.1.1 (Lanfear et al. 2012) for protein coding loci using linked branch lengths, BIC criterion, and a greedy search algorithm, with analyses run separately to inform *BEAST, RAxML, and MrBayes analyses. Substitution models for translated AA sequences for the eleven nuclear protein-coding loci were estimated

using MEGA 6.06, using ML model selection (Tamura et al. 2013). Evolutionary models for the ribosomal regions 28S and 18S utilized jModelTest 2.1.6 (Guindon and Gascuel 2003, Darriba et al. 2012), considering 24 models evaluated using AIC criterion to choose optimal models under a ML search. Further methods for model selection and resulting models are available in Table 1.4.

ML gene trees and concatenated phylogenetic analyses were run using RAxML-HPC2 8.0.24 (Stamatakis et al. 2008) on the CIPRES Science Gateway 3.3 (Miller et al. 2010). A rapid bootstrap analysis and search for the best-scoring ML tree (-f a) was conducted using the GTRGAMMA model. The RAxML concatenated phylogeny was repeated three times. Bayesian phylogenetic reconstruction used both concatenation (via MrBayes 3.2.1, Ronquist et al. 2012) and the coalescent-based *BEAST (Heled and Drummond 2010). MrBayes 3.2.1 was run for 10 million generations at which point the average standard deviation of split frequencies was < 0.0001. Parameters were logged every thousand generations, and the first 25% of samples were discarded as burn-in. Lognormal relaxed clock models were specified for all 14 loci in preliminary *BEAST analyses in order to determine whether a strict clock could be statistically rejected for any of the loci. Estimates of the coefficient of variation for each locus indicated that a strict clock model could not be rejected for 9 of the 14 loci; a strict clock model was specified for each of these 9 loci in subsequent analyses. Six independent *BEAST runs were conducted for 100,000,000 generations, sampling every 10,000 generations. ESS values and stationarity of likelihood values were evaluated in Tracer 1.6, then combined using LogCombiner 1.8.0 with a 60% burn-in. This output was used to reconstruct a maximum

clade credibility tree in TreeAnnotator 1.8.0. The concatenated and unpartitioned AA alignment was analyzed in PhyML 3.0 under the JTT+G model and using the best of NNI and SPR tree search algorithm (Guindon et al. 2010).

When absent (primarily *Sabacon* and *Acuclavella*), transcriptomic data were added to the preliminary expanded panel alignments by first identifying homologs to specific loci in the de novo assemblies (blastn, e-value: 1e-10), then performing alignments with a cost matrix of 65% similarity, and a gap open and extension penalty of 20 and 3 respectively. Intronic regions from Sanger sequences were manually removed. Subsequently, all alignments were realigned in Geneious using MAFFT 6 (Katoh et al. 2002) with the FFT-NS-2 algorithm and the 200 PAM long-odds matrix (Jones et al. 1992) calculated with k = 2 (Kimura 1980), gap open penalty = 2, and offset value = 0.123. Nucleotide alignments with indels had regions of alignment uncertainty removed with GBlocks 0.91b (Castresana 2000) with minimum length of a block = 3, allowed gap position = half, minimum number of sequences used to identify a conserved position and a flanking position = half the number of alignment sequences, and maximum number of contiguous non-conserved positions = 8; putative protein coding sequences had their open reading frame conserved. The same number of partitions (n=35) was recovered for the MrBayes and *BEAST query, though sometimes the same partitions recovered different models as more likely; the RAxML query resulted in a scheme of 31 partitions (Table 1.4). For the Pol II alignment, the addition of transcriptome data resulted in 20 bp and 76 bp gaps in the RT-PCR generated sequences on GenBank (Shultz and Regier

2001). Thus, transcriptome-derived sequences were analyzed instead of these GenBank sequences and resulting gaps were coded as missing data.

(3) Transcriptome generation and phylogenomics

Preliminary expanded panel phylogenetic analyses showed a near trichotomy with long-branches separating three ingroup lineages including (1) *Sabacon*, (2) *Hesperonemastoma* and *Taracus*, and (3) *Ischyropsalis*, *Acuclavella*, and *Ceratolasma*. Thus exemplars were selected representing each of these lineages (*Sabacon*, *Hesperonemastoma*, and *Acuclavella*) for transcriptome-based phylogenomics. The transcriptome panel also included two outgroup taxa (*Trogulus* and *Ortholasma*). Three of the panel transcriptomes were previously published (Hedin et al. 2012; SRX450964, SRX451776, and SRX450937 for *Trogulus*, *Ortholasma*, and *Hesperonemastoma* respectively). RNA extractions for *Sabacon* and *Acuclavella* were conducted using TRIzol, and purified with the Qiagen RNeasy MinElute Cleanup Kit. Extractions used whole body tissues from single individuals with the midgut removed to reduce contamination. RNAs were sent to the Genomic Services Lab at HudsonAlpha Institute for Biotechnology (www.hudsonalpha.org) where non-normalized libraries were prepared using the Illumina TruSeq RNASeq kit and then sequenced using Illumina HiSeq technology with paired-end, 50-bp reads. For all transcriptomes adaptors were trimmed using TrimGalore! 0.2.7 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and sequences with >1% ambiguity or less than thirty base pairs were removed using PRINSEQ Lite 0.20.2 (Schmieder and Edwards 2011). Sequence

reads passing these filters were assembled de novo using the Trinity platform (Grabherr et al. 2011, Haas et al. 2013).

Sets of orthologous sequences were filtered from the five transcriptomes by first identifying putative homologs to 5470 Ixodes loci annotated as single-exon, and also the 367 harvestmen loci used previously by Hedin et al. (2012). Identification of putative homologs was conducted with a liberal BLAST e-value (1e - 1; Altschul et al. 1990). Queries resulting in missing data (n = 3931) or with multiple highly-overlapping hits from a single transcriptome (n = 985) were not considered further. Paralogy was further assessed in two ways. First, a representative sequence from each alignment was again subjected to BLAST against the Dyspnoi transcripts, with a conservative e-value (1e - 50). If this BLAST returned multiple sequences per exemplar, then alignments were discarded (n = 2). Second, gene trees not recovering (n = 20) or not supporting (with a BSV < 70, n = 30) a monophyletic Ischyropsalidoidea were discarded under the assumption that incongruent nodes for this otherwise well-supported clade are the result of paralogy or LBA. These criteria were not mutually exclusive, and a total of 672 loci passed query and paralogy filters.

Gene trees for 672 loci were estimated using ML in PhyML 3.0 (Guindon et al. 2010) using default parameters including the HKY85+G substitution model and the NNI tree search algorithm. Nodal support was assessed via 100 bootstrap replicates, which were rooted using the re-root tool on the STRAW web server (Shaw et al. 2013). STRAW was also used to conduct partially-parametric coalescent-based analyses, using PhyML gene trees as input. For these analyses both MP-EST (Liu et al. 2010), which

uses the frequency of triplets of taxa to estimate the topology and branch lengths, and STAR (Liu et al. 2009), which computes the pairwise topological distance among pairs of taxa were used to determine the average placement of nodes across a collection of gene trees. The coalescent arises as a large-population approximation of the Wright-Fisher model (Nordborg 2001), thus coalescent analyses make the same simplifying assumptions including constant populations sizes and no selection, and attribute all gene tree incongruence to ILS (Kubatko and Degnan 2007, Springer and Gatesy 2016). These partially-parametric coalescent methods were chosen because the size of the transcriptome panel was too computationally demanding to implement fully-parametric coalescent analyses (e.g., *BEAST). Additionally, the transcriptome panel is analyzed via concatenation using RAxML-HPC2 on CIPRES. This supermatrix was partitioned by gene using the default "new rapid hill-climbing" tree search algorithm, with a GTRGAMMA model applied to each partition.

(4) Comparison of concatenation versus coalescent phylogenomic analyses

The performance and consistent estimation of concatenation versus partially-parametric methods was evaluated by randomly sub-sampling transcriptome-derived loci. Ten replicates each of 25, 50, 100, 200, 300, 400, 500, and 600 loci were selected, resulting in a total of 80 replicates. Phylogenetic analyses for each of these replicates was performed using MP-EST, STAR, and RAxML, using parameters as outlined above.

(5) Evaluation of unequal minority resolution frequencies

Under the basic multispecies coalescent model the frequency of minority resolution gene trees should be equal (Pamilo and Nei 1988). The 672 PhyML gene trees were used to test the equality of minority resolution frequencies using a two-sided binomial test. The symmetry of gene tree frequency across different attributes of the data were evaluated to determine if UMRFs were the result of methodological bias. Additionally, a subset of the analyzed loci that were retrieved from transcriptomes using a different methodological pipeline (loci from Hedin et al. 2012). The assumption is that if minority asymmetry persists across different attributes or treatments of the data then UMRFs are not a methodological artifact, but are caused by biological aspects such as structured ancestral populations or paraphyletic gene flow (see Discussion). Two locus attributes that have been suggested to improve phylogenomic pipelines were chosen, including high AT3 (Romiguier et al. 2013), and high phylogenetic support values (Salichos and Rokas 2013, though see Betancur-R et al. 2014). Further evaluation into the frequency of alternative topologies in the data was accomplished by analyzing CFs, or the proportion of the sampled genome that agree with a given bipartition, within the 672-locus dataset (Baum 2007). This was done using a Bayesian Concordance Analysis (BCA, Ané et al. 2007) in the program BUCKy (Larget et al. 2010). BUCKy uses independent Bayesian analysis of each gene as input. These analyses were conducted in MrBayes 3.2.1 (Ronquist et al. 2012), based on 100,000 generation runs, sampling every 100 trees, and discarding the first 250 trees as burn-in. Each locus included two partitions, one combining the 1st and 2nd bp position, and another for the 3rd, with nst =

6 and rates = gamma. BUCKy was used to map the posterior sample of trees to alternative topologies using an a priori expectation of gene tree discordance. For this analysis, the prior level of discordance (a) was chosen to give equal likelihood to each of the three possible rooted triplets. The probability that two loci share the same tree is about 1/(1 + a), thus a = 2 was set. This analysis can be used to reject the hypothesis that all gene tree discordance is due to incomplete lineage sorting (Ané 2010).

III. RESULTS

(1) Expanded panel phylogenetics

Primer design resulted in the development of eight molecular markers with phylogenetic utility in Ischyropsalidoidea (Table 1.3). Data augmented from *Sabacon* and *Acuclavella* transcriptomes improved the average percentage of loci sampled per OTU in the expanded panel (Table 1.5). The final expanded panel included 14 loci for 12 ingroup taxa, with a concatenated alignment length of 9407 bp (26.5% missing). All phylogenetic analyses were rooted with the troguloid genera *Ortholasma* and *Trogulus* except for the Wnt2 matrix that contained data for *Trogulus* only. Expanded panel RAxML gene trees are deposited in the Dryad Digital Repository (http://dx.doi.org/10.5061/dryad.3mr26) and available in Appendix A. Ischyropsalidoidea and all genera are recovered with high support in the majority of gene trees. Twelve of the 14 loci recovered Ischyropsalidoidea with BSVs ≥ 97. Occasionally genera were not recovered as monophyletic, though

paraphyly was always with respect to closely related genera. Furthermore, gene trees tended to recover (*Ischyropsalis*, (*Acuclavella*, *Ceratolasma*)) with strong support.

It has been shown that paralogous sequences of EF-1α exist in mite harvestmen (Clouse et al. 2013). Therefore, we queried an ischyropsalidoid sequence from our EF-1α alignment against all five Dyspnoi transcriptomes using blastn in Geneious Pro 5.5 (Kearse et al. 2012). Using an e-value of 1e-50, troguloid transcripts are not queried, and ischyropsalidoid transcripts were returned without paralogs. At more liberal E-values (1e-20, 1e-10), alternate sequences were detected for EF-1α, however these putative paralogous sequences were not recovered from all transcripts, and when they did occur consisted of few short sequences on the 3' end of the fragment. Further, the EF-1a gene tree was very similar to the species tree by recovering all genera (five of six genera with strong support), Ischyropsalidoidea with strong support, *Acuclavella* sister to *Ceratolasma* (BSV = 99), and *Ischyropsalis* sister to *Acuclavella* + *Ceratolasma* (BSV = 99, see s4 below). This suggests that paralogous EF-1α sequences were not captured by our approach, and that the rampant gene duplication of EF-1α known from a single species of Cyphophthalmi is not problematic in Dyspnoi (see also Schönhofer et al. 2015, Appendix A).

Despite recovery of superfamily and "tip" relationships, nearly all backbone nodes within Ischyropsalidoidea lack support in gene tree analyses – i.e., it is unclear how most genera are related by examining individual gene trees. In contrast, combined phylogenetic analyses of the expanded panel recovered the same topology across most methods of inference (Figure 1.2). The one exception is the PhyML analysis of translated

AAs which recovered *Sabacon* sister to remaining ischyropslidoids, though this relationship is not well supported (BSV = 52; Appendix A). This analysis also fails to recover the genus *Ischyropsalis* as monophyletic, with these two taxa by far having the most missing data (Table 1.3, 1.5). All other combined analysis nodes were strongly supported by MrBayes, RAxML and *BEAST, with the exception of a node associated with a short branch deep in the ischyropsalidoid phylogeny. This node was most strongly recovered in the MrBayes concatenated analysis, which had only one tree in the 99% credible set – the only other sampled tree recovered (*Sabacon*, (*Hesperonemastoma*, *Taracus*)). This node was less well-supported in RAxML (BSV = 68) and *BEAST (PP = 0.82) analyses.

(2) Transcriptome panel phylogenomics

Transcriptome assembly statistics, and comparison to previously published (Hedin et al. 2012) transcriptome assemblies, are reported in Table 1.6. The final transcriptome panel included 672 loci, 3 ingroup and 2 outgroup taxa, with a concatenated alignment length of 536,124 bp. These data are nearly distinct from the expanded panel, with 5 loci with various levels of overlap totaling 1335 bp. Phylogenetic analyses of the transcriptome panel further resolved phylogenetic relationships at the base of Ischyropsalidoidea despite high levels of gene tree conflict. Evaluation of PhyML gene trees indicate high levels of gene tree incongruence (Figure 1.3), with (*Hesperonemastoma*, (*Acuclavella*, *Sabacon*)) recovered in 37.5%; (*A*, (*H*, *S*)) in 35.4%; and (*S*, (*H*, *A*)) in 27.1% of gene trees. Concatenated analysis of all data (RAxML), and

coalescent-based analyses of PhyML gene trees (MP-EST and STAR) were congruent in their recovery of *Hesperonemastoma* as sister to the other remaining ischyropsalidoid lineages, but differed in their support and inferred branch lengths (Figure 1.4). Partially-parametric coalescent analyses tended to recover a very short branch just inside of Ischyropsalidoidea with only moderate support values. The concatenated analysis recovered a longer internal ischyropsalidoid branch with higher support for this topology.

(3) Comparison of phylogenomic analyses from sub-sampled loci

Concatenated and partially-parametric coalescent analyses tended to recover the same topology in any particular replicate, though concatenated analyses more consistently recovered (*H*, (*A*, *S*)) with higher support than in coalescent analyses, which did not settle on this topology until after 300 or more loci were analyzed. Perhaps most conspicuous is a 600-loci replicate that was recovered as (*A*, (*H*, *S*)) by both STAR and MP-EST. Examination of the results from sub-sampled loci shows that recovering *Hesperonemastoma* as sister to *Sabacon* + *Acuclavella* could not have reasonably been recovered without using over 400 loci (Figure 1.5).

(4) Unequal minority resolution frequencies

The minority resolution frequencies in the 672-locus dataset were unequal (two-sided binomial test, $p = 0.0072$). This trend persisted across treatments (Table 1.7), though this was not significant for the subset of loci that were generated by Hedin et al. 2012 ($p = 0.1203$). BCA analyses reject the hypothesis that all gene tree discordance is

due to ILS with 99% confidence (Figure 1.6). The BUCKy concordance tree (Figure 1.4) recovers the same topology as concatenation and coalescent-based analyses, with a CF of 0.475. The 99% highest posterior density interval of trees in the posterior sample (0.394–0.475) does not overlap with either of the minority resolution topologies. Both of the alternative topologies (*A*, (*H*, *S*)) and (*S*, (*H*, *A*)) were frequently recovered with non-overlapping CFs in the 99% posterior tree sample, with CFs of 0.351 (0.313–0.390) and 0.214 (0.179–0.250) respectively (Figure 1.6).

(6) Data availability

A spreadsheet characterizing the 672 loci alignments, as well as all alignments, matrices, trees, and partition files are deposited in the Dryad Digital Repository <http://dx.doi.org/10.5061/dryad. 3mr26>. Illumina raw reads for *Sabacon* (SRR2924723) and *Acuclavella* (SRR2924718) have been submitted to NCBI Short Read Archive. All Sanger sequence data generated in this study have been deposited to GenBank (Table 1.3).

IV. DISCUSSION

(1) Resolution of an ancient trichotomy is aided by increased taxon sampling

Randomly sampling loci from the transcriptome panel shows that hundreds of loci were necessary to reliably infer the topology at the base of Ischyropsalidoidea. The 25-loci sub-sample analyses recovered (*H*, (*A*, *S*)) in only 50% of replicates (Figure 1.5). As

such, with only 14 loci in the expanded panel, the topological congruence between the expanded panel and transcriptome panel may simply be due to chance. An alternative explanation is that the increased taxon sampling in the expanded panel aids phylogenetic inference by shortening the branches leading from the base of Ischyropsalidoidea (Pollock et al. 2002). To explore this possibility, the expanded panel was trimmed to only include OTUs represented in the transcriptome panel (this matrix includes only 1.4% missing data). Phylogenetic analysis of this reduced matrix with RAxML, MrBayes, and *BEAST using the same parameters as on the full expanded panel returns mixed results (Appendix A.2). RAxML does not recover *Hesperonemastoma* as sister to *Acuclavella* and *Sabacon* (BSV = 52 for *Sabacon* as sister), MrBayes does (PP = 49), though both of these concatenated analyses are weakly supported. On the other hand, *BEAST recovers *Hesperonemastoma* as sister to the other ischyropsalidoids with PP = 97.8. For ancient radiations, *BEAST may be more robust to reduced taxon sampling compared to other methods of phylogenetic inference used here, though just a single example is provided and this should be further explored. Conversely, *BEAST is typically used to infer shallow evolutionary events, and sampling more than one individual per species is explicitly recommended. In *BEAST, sampling multiple individuals per species allows for more accurate population size estimation, and this in turn may allow for better estimates of divergence times and topology (Heled and Drummond 2010).

These findings suggest that increased taxon sampling along descending branches from an ancient near-trichotomy helps with the phylogenetic inference of these diversifications. It has been assumed this would not be the case, because the number of

lineages to evaluate sorting along the short critical branch is not increased (Degnan and Rosenberg 2006, Kubatko and Degnan 2007). Likely this is due to the additional taxa diminishing phylogenetic artifacts by breaking up long external branches, thus resulting in less LBA (Hillis 1998). For deep phylogenetic questions, variant sites can become saturated, resulting in abundant homoplasy due to convergence, which is thought to be positively correlated with branch length (Felsenstein 2004). It is likely that increased taxon sampling diminishes the amount of saturation, which in effect unmasks synapomorphic information along the short internal branch. To this end, the data support the findings of Heled and Drummond (2010) that increased taxon sampling contributes to accurate species tree estimations of rapid radiations. However, Heled and Drummond (2010) couched this argument for shallow phylogenetic inferences, and suggested that increased locus sampling is more important for accurate estimation of deep phylogenetic questions. Additionally, the results suggest using caution when attempting to resolve ancient diversifications using few terminals.

(2) Emergence of support with supermatrix analyses

High levels of gene tree conflict (Figure 1.3) characterize the root of Ischyropsalidoidea. Despite the reported success of partially-parametric coalescent analyses compared to concatenation (e.g., Liu et al. 2009, 2010; Kubatko et al. 2009), and claims that widespread ILS can result in poor performance of concatenated methods (Kubatko and Degnan 2007), however results seem to indicate that hidden support emerges when analyzing data via concatenated analyses (Gatesy and Springer 2014), and

concatenation appears more robust than partially-parametric coalescent methods (e.g., Leaché and Rannala 2011, Patel et al. 2013). The randomly sub-sampled loci shows that the supermatrix approach to phylogenetic inference, even in the face of high levels of gene tree conflict, results in robust support for nodes otherwise weakly supported by partially-parametric coalescent analyses. This is notable, since others (Edwards et al. 2016) have claimed that sub-sampling phylogenomic data as a test for phylogenetic consistency produces a condition where "the deficiencies of concatenation become even more glaring". They make this claim despite only a single supporting example (from Song et al. 2012); the example here does not support this claim. In analyses of mammal relationships, Gatesy and Springer (2014) were able to uncover extreme hidden support for 10 clades by concatenating loci (from McCormack et al. 2012) that strictly conflict with the species tree. To evaluate if hidden support is present in the data, the 420 loci from the transcriptome panel were combined with gene trees strictly conflicting with the species tree, and analyzed this partitioned supermatrix in RAxML using the parameters outlined above. This analysis did not recover extreme hidden support (Appendix A.1). However, this may be caused by the unequal minority frequency of the gene trees generated from analysis of these loci. The analyses of Gatesy and Springer included 25 ingroup taxa, and the 5 OTUs representing members of Glires were scattered through the gene trees. Thus, it is difficult to calculate the symmetric frequencies of the many alternate gene tree topologies in their data.

(3) Interpreting gene tree incongruence and unequal minority resolution frequencies

The divergence at the base of Ischyropsalidoidea is inferred to have happened rapidly enough to allow for ILS to occur. However, gene tree incongruence can arise from factors other than ILS. Springer and Gatesy (2016) were able to show that the high levels of incongruent gene trees, attributed by Song et al. (2012) to ILS, were instead caused by other factors. Among their criticisms of the phylogenetic analyses of gene trees by Song et al. (2012) are factors that potentially apply to this research, including inadequate search algorithms and LBA. First, the PhyML gene tree analyses used default parameter settings including the HKY85 model of nucleotide substitution and the NNI tree search algorithm. More rigorous model selection would likely return gene trees with better log-likelihood values (e.g., Springer and Gatesy 2016), and NNI has been shown to be prone to being stuck in local optima (Money and Whelan 2011). However, each gene tree was also run using more rigorous methods using MrBayes for the BUCKy analysis. The data for both sets of gene tree analyses agree that there is a large amount of gene-tree discordance in the data set, and they agree in the frequency that these alternative topologies are recovered. Second, higher frequencies of a gene tree than its true frequency can be caused by LBA. In the data however, the longest ingroup branch in both the PhyML and RAxML analyses is associated with *Hesperonemastoma*. The average branch lengths in PhyML analyses is 0.400891 for *Hesperonemastoma*, 0.299665 for *Acuclavella*, and 0.299444 for *Sabacon*; in the RAxML phylogeny it is 0.443753, 0.333928, and 0.350993, respectively. Under LBA, longer *Hesperonemastoma* branches would be pulled toward outgroup sequences, causing the gene tree stoichiometry to be

skewed toward an (*H*, (*A*, *S*)) topology. If this is the case, there is more gene tree incongruence in this system than reported, for (*H*, (*A*, *S*)) is the most frequently recovered topology. As such, ILS is a likely source of some gene tree incongruence in the data.

That being said, the lack of symmetry in the frequency of minority gene trees in the 672 loci does not match theoretical expectations. Since ILS is a random process, expectations are that one topology would be favored, with alternative topologies occurring at equal frequencies (Pamilo and Nei 1988, Knowles and Kubatko 2010). For example, in great ape relationships, a study using 11,945 loci found that 76.6% of gene trees support a (*Gorilla*, (*Homo*, *Pan*)) topology, with 11.5% and 11.4% supporting (*H*, (*G*, *P*)) and (*P*, (*H*, *G*)) respectively (Ebersberger et al. 2007). In Ischyropsalidoidea such symmetry is not observed (Figure 1.3): 37.5% of gene trees recover (*Hesperonemastoma*, (*Acuclavella*, *Sabacon*)), 35.4% recover (*A*, (*H*, *S*)), and 27.1% recover (*S*, (*H*, *A*)). This is a classic rooted triplet case, where the probabilities of the three possible gene trees should all equal 33.3% when s = 0 – a hard polytomy. UMRFs in rooted triplets have been shown to arise from violation of the assumptions of the multispecies coalescent model (Zwickl et al. 2014). Potentially, any of the artifactual or biological causes of gene tree/species tree discordance (Maddison 1997, Degnan and Rosenberg 2006), other than ILS, could lead to UMRFs. A recent study by Zwickl et al. (2014) uncovered UMRFs due to both analytical artifacts and biological processes. In some cases, they were able to restore equal minority frequencies of gene trees by using alternative alignment strategies, suggesting that unequal frequencies were caused by artifacts in their methodological

pipeline. In other cases, UMRFs persisted across alternative alignment strategies, suggesting that they are the result of biological processes (e.g., introgression).

Zwickl et al. (2014) were able to reduce or eliminate UMRFs that were the product of methodological biases by including intron sequences in their alignments. This strategy is not employed here, for the transcriptome-generated data do not include intronic sequences. Therefore, an attempt was made to remove UMRFs from the dataset by considering loci that were generated using a different methodological pipeline (Hedin et al. 2012), or by considering locus/gene tree traits thought to improve phylogenomic analyses, including phylogenetic support values (Salichos and Rokas 2013) and high AT3 content (Romiguier et al. 2013). UMRFs were consistent across these treatments (Table 1.7), and thus conclude that UMRFs are likely the result of biological processes. However, it is possible that methodological biases persist across these subsets of the data. For example, convergence in the base composition between two taxa could skew the stoichiometry of a topology combining these taxa to be more common that its true frequency (Springer and Gatesy 2016).

The cause of UMRFs in this system could result from any violation of the multispecies coalescent model. Although undetected paralogy is not considered as the likely cause of this discordance, due to low frequency of paralogs detected by the filtering criteria, this is one possibility. Other biological processes that are more likely to apply to Ischyropsalidoidea include ancestral population structure and paraphyletic gene flow. Population structure has been shown to cause UMRFs when subdivision is present in the ancestor of three lineages and persists through both speciation events (Slatkin and

Pollack 2008). This may initially seem unlikely, but many harvestmen lineages are known to show extreme population structure, as are many non-vagile terrestrial arthropods (e.g., Derkarabetian et al. 2011, Keith and Hedin 2012). Likewise, it can be inferred from Leaché et al. (2014a) that paraphyletic gene flow, or gene flow between species that are not sister taxa, can increase the frequency of gene trees grouping these taxa together. Also, the total branch lengths of the transcriptome panel RAxML analysis from the base of Ischyropsalidoidea to the tip of *Hesperonemastoma*, *Acuclavella*, and *Sabacon* are 35.0%, 33.2%, and 31.8% of the total of these sums respectively, which more closely matches the asymmetry of gene trees above than do theoretical expectations, suggesting that selection or evolutionary rates may be playing a role.

To us, inferring the population genetics of lineages that underwent successive diversifications around 200 MYA (Schönhofer et al. 2013, Sharma and Giribet 2014) seems a near-futile effort. Thus far, species tree analyses for the most part have been robust in their inference of nodes with UMRFs (Zwickl et al. 2014), though estimating the correct species tree can become difficult when the biological processes underlying this discordance are severe (e.g. Leaché et al. 2014a). Since species tree resolution within Ischyropsalidoidea is the primary goal, the source the UMRFs recovered here is not further searched for, under the assumption that it is not severely affecting the phylogenetic inference. Future research on difficult phylogenetic nodes could employ a strategy where the likelihood of the data are analyzed with respect to a priori models that vary in ancestral population structure, timing of divergences, etc., with the most likely model selected using a criterion score (e.g., Carstens et al. 2013). Clearly the causes of

UMRFs and the consequences to phylogenetic inference should be the focus of future research.

V. CONCLUSIONS

A single short branch deep in the phylogeny of Ischyropsalidoidea is inferred. Despite high levels of gene tree conflict, *Hesperonemastoma* + *Taracus* sister to remaining ischyropsalidoids is consistently recovered across different analytical methods with different strategies of taxon and locus sampling. Though the short internal branch deep in the ischyropsalidoid phylogeny is consistently recovered, it is associated with high levels of gene tree conflict and relatively poor support values. These characteristics are precisely those associated with topological conflict between coalescent- and supermatrix-based methods of phylogenetic inference (Lambert et al. 2015). That being said, the causes of gene tree conflict associated with ancient short internal branches should continue to be explored. Particularly, simulation analyses should explore the effects of ILS and UMRFs at the time of divergence on phylogenetic analyses, with these simulations extended to deep time to assess if this signal degrades though time. If gene tree conflict initially associated with ILS degrades to gene tree conflict associated with sampling error in ancient diversifications, than coalescent methods that attribute all such conflict to ILS may be inappropriate. As phylogenomics comes of age, transposon insertions have been used to independently assess incongruence at ancient nodes, suggesting that incongruence in gene trees is largely due to sampling error (Gatesy and

Springer 2014), and assessing this signal could inform the appropriateness of phylogenomic analyses. Further, these analyses should manipulate the underlying assumptions of the simple coalescent model, such as selection and variation in population size, to assess the impacts of these parameters on the frequency of gene tree incongruence (Nordborg 2001, Scally et al. 2012, Springer and Gatesy 2016).

Arguably the strongest evidence for phylogenetic hypotheses occurs when clades are recovered from independent lines of evidence (Rota-Stabelli et al. 2011). The expanded panel results suggest that increased taxon sampling may be as important for estimating ancient radiations as increased locus sampling, for these additional taxa may decrease saturation and increase phylogenetic signal along problematic branches. The impact of taxon sampling on the phylogenetic reconstruction of ancient radiations should also be an area of future research.

VI. REFERENCES

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

Ané, C. (2010). Reconstructing concordance trees and testing the coalescent model from genome-wide data sets. In: Knowles, L.L., Kubatko, L.S. (Eds.), Estimating Species Trees: Practical and Theoretical Aspects. Wiley-Blackwell, Hoboken, New Jersey (Chapter 3).

Ané, C., B. Larget, D. A. Baum, S. D. Smith, A. Rokas (2007). Bayesian estimation of concordance among gene trees. Molecular Biology and Evolution 24 (2): 412–426. http://dx.doi. org/10.1093/molbev/msl170.

Asahida, T., T. Kobayashi, K. Saitoh, I. Nakayama (1996). Tissue preservation and total DNA extraction from fish stored at ambient temperature using buffers containing high concentration of urea. Fisheries Science 62 (5): 727–730.

Baum, D. A. (2007). Concordance trees, concordance factors, and the exploration of reticulate genealogy. Taxon 56 (2): 417–426.

Bayzid, M. S., T. Warnow (2013). Naive binning improves phylogenomic analyses. Bioinformatics 29 (18): 2277–2284. http://dx.doi.org/10.1093/bioinformatics/ btt394.

Betancur-R, R., G. J. P. Naylor, G. Ortí (2014). Conserved genes, sampling error, and phylogenomic inference. Systematic Biology 63 (2): 257–262. http://dx.doi.org/10.1093/ sysbio/syt073.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, P. J. Waddell, (1993). Partitioning and combining data in phylogenetic analysis. Systematic Biology 42 (3): 384–397. http://dx.doi.org/10.2307/2992473.

Carstens, B. C., R. S. Brennan, V. Chua, C. V. Duffie, M. G. Harvey, R. A. Koch, C. D. McMahan, B. J. Nelson, C. E. Newman, J. D. Satler, G. Seeholzer, K. Posbic, D. C. Tank, J. Sullivan (2013). Model selection as a tool for phylogeographic inference: an example from the willow *Salix melanopsis*. Molecular Ecology 22 (15): 4014–4028. http://dx.doi.org/10.1111/mec.12347.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis Molecular Biology and Evolution 17 (4): 540–552.

Clouse, R. M., P. P. Sharma, G. Giribet, W. C. Wheeler (2013). Elongation factor-1a, a putative single-copy nuclear gene, has divergent sets of paralogs in an arachnid. Molecular Phylogenetics and Evolution 68 (3): 471–481. http://dx.doi.org/10.1016/j. ympev.2013.04.025.

Conesa A., S. Götz, J. M. García-Gómez, M. Terol, M. Robles (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21 (18): 3674–3676. doi:10.1093/bioinformatics/bti610

Darriba, D., G. L. Taboada, R. Doallo, D. Posada (2012). JModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9 (8): 772.

Degnan, J. H., N. A. Rosenberg (2006). Discordance of species trees with their most likely gene trees. PLoS Genet 2 (5): e68. http://dx.doi.org/10.1371/journal. pgen.0020068.

Degnan, J. H., N. A. Rosenberg (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in Ecology & Evolution 24 (6): 332–340. http://dx. doi.org/10.1016/j.tree.2009.01.009.

Derkarabetian, S., M. Hedin (2014). Integrative taxonomy and species delimitation in harvestmen: a revision of the western North American genus *Sclerobunus* (Opiliones: Laniatores: Travunioidea). PLoS One 9 (8): e104982. http://dx.doi. org/10.1371/journal.pone.0104982.

Derkarabetian, S., J. Ledford, M. Hedin (2011). Genetic diversification without obvious genitalic morphological divergence in harvestmen (Opiliones, Laniatores, *Sclerobunus robustus*) from montane sky islands of western North America. Molecular Phylogenetics and Evolution 61: 844–853.

Dresco, E. (1970). Recherches sur la variabilité et la phylogénie chez les Opiliones du genre *Ischyropsalis* CL Koch (Fam. Ischyropsalidae), avec la creation de la famille nouvelle des Sabaconidae. Bulletin du Muséum National d'Histoire Naturelle, 2e Serie 41: 1200–1213.

Ebersberger, I., P. Galgoczy, S. Taudien, S. Taenzer, M. Platzer, A. von Haeseler (2007). Mapping human genetic ancestry. Molecular Biology and Evolution 24 (10): 2266–2276. http://dx.doi.org/10.1093/molbev/msm156.

Edwards, S. V., L. Liu, D. K. Pearl (2007). High-resolution species trees without concatenation. Proclamations of the National Academy of Sciences 104 (14): 5936–5941.

Edwards, S. V., Z. Xi, A. Janke, B. C. Faircloth, J. E. McCormack, T. C. Glenn, B. Zhong, S. Wu, E. M. Lemmon, A. R. Lemmon, A. D. Leaché, L. Liu, C. C. Davis

(2016). Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Molecular Phylogenetics and Evolution Part A 94: 447–462. http:// dx.doi.org/10.1016/j.ympev.2015.10.027.

Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, T. C. Glenn (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Systematic Biology 61 (5): 717–726. http:// dx.doi.org/10.1093/sysbio/sys004.

Felsenstein, J. (2004). Inferring Phylogenies. Sinauer Associates.

Gatesy, J., M. S. Springer (2014). Phylogenetic analysis at deep timescales: unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Molecular Phylogenetics and Evolution 80: 231–266. http://dx.doi.org/10.1016/j. ympev.2014.08.013.

Gatesy, J., P. O'Grady, R. H. Baker (1999). Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level Artiodactyl Taxa. Cladistics 15 (3): 271–313. http://dx.doi.org/ 10.1111/j.1096-0031.1999.tb00268.x.

Giribet, G., A. B. Kury (2007). Phylogeny and biogeography. In: Pinto-da-Rocha, R., Machado, G., Giribet, G. (Eds.), Harvestmen: The Biology of Opiliones. Harvard University Press, Cambridge, MA, pp. 62–87 (Chapter 3).

Giribet, G., P. P. Sharma (2015). Evolutionary biology of harvestmen (Arachnida, Opiliones). Annual Review of Entomology 60 (1): 157–175. http://dx.doi.org/10.1146/ annurev-ento-010814-021028.

Giribet, G., M. Rambla, S. Carranza, J. Baguñà, M. Riutort, C. Ribera (1999). Phylogeny of the arachnid order Opiliones (Arthropoda) inferred from a combined approach of complete 18S and partial 28S ribosomal DNA sequences and morphology. Molecular Phylogenetics and Evolution 11 (2): 296–307. http://dx. doi.org/10.1006/mpev.1998.0583.

Giribet, G., L. Vogt, A. Pérez González, P. Sharma, A. B. Kury (2010). A multilocus approach to harvestman (Arachnida: Opiliones) phylogeny with emphasis on biogeography and the systematics of Laniatores. Cladistics 26 (4): 408–437. http://dx.doi.org/10.1111/j.1096-0031.2009.00296.x.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev (2011). Trinity: reconstructing a full-length transcriptome

without a genome from RNA-Seq data. Nature Biotechnology 29 (7): 644–652. http://dx.doi.org/ 10.1038/nbt.1883.

Gruber, J. (1978). Redescription of *Ceratolasma tricantha* Goodnight and Goodnight, with notes on the family Ischyropsalidae (Opiliones, Palpatores). Journal of Arachnology 6: 105–124.

Guindon, S., O. Gascuel (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52 (5): 696–704. http://dx. doi.org/10.1080/10635150390235520.

Guindon, S., J. F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, O. Gascuel (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59 (3): 307–321. http://dx. doi.org/10.1093/sysbio/syq010.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. Couger, D. Eccles, B. Li, M. Lieber, M. 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, A. Regev (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8 (8): 1494–1512. http://dx.doi.org/ 10.1038/nprot.2013.084.

Harvey, M. S. (2002). The neglected cousins: What do we know about the smaller arachnid orders? Journal of Arachnology 30 (2): 357–372. http://dx.doi.org/10.1636/0161- 8202(2002)030[0357:TNCWDW]2.0.CO;2.

Hedin, M., S. Derkarabetian, M. McCormack, C. Richart, J. W. Shultz (2010). The phylogenetic utility of the nuclear protein-coding gene EF-1a for resolving recent divergences in Opiliones, emphasizing intron evolution. Journal of Arachnology 38 (1): 9–20. http://dx.doi.org/10.1636/HA09-49.1.

Hedin, M., J. Starrett, S. Akhter, A. L. Schönhofer, J. W. Shultz (2012). Phylogenomic resolution of Paleozoic divergences in Harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome data. PLoS One 7 (8): e42888. http:// dx.doi.org/10.1371/journal.pone.0042888.

Heled, J., A. J. Drummond (2010). Bayesian inference of species trees from multilocus data. Molecular Biology and Evolution 27 (3): 570–580. http://dx.doi.org/10.1093/molbev/ msp274.

Hillis, D. M. (1998). Taxonomic sampling, phylogenetic accuracy, and investigator bias. Systematic Biology 47 (1): 3–8. http://dx.doi.org/10.1080/106351598260987.

Huelsenbeck, J. P., D. M. Hillis (1993). Success of phylogenetic methods in the four-taxon case. Systematic Biology 42 (3): 247–264. http://dx.doi.org/10.1093/sysbio/42.3.247.

Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa, T. Miyata (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proceedings of the National Academy of Sciences 86 (23): 9355– 9359.

Katoh, K., K. Misawa, K. Kuma, T. Miyata (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30 (14): 3059–3066. http://dx.doi.org/10.1093/nar/gkf436.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28 (12): 1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.

Keith, R., M. Hedin (2012). Extreme mitochondrial population subdivision in southern Appalachian paleoendemic spiders (Araneae: Hypochilidae: *Hypochilus*), with implications for species delimitation. Journal of Arachnology 40 (2): 167–181. http://dx.doi.org/10.1636/A11-49.1.

Kimura M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution, 16 (2): 111–120. doi:10.1007/BF01731581.

Kingman, J. F. (1982). The coalescent. Stochastic Processes and their Applications 13 (3): 235–248. http://dx. doi.org/10.1016/0304-4149(82)90011-4.

Knowles, L. L., L. S. Kubatko (2010). Estimating species trees: an introduction to concepts and models. In: Knowles, L.L., L. S. Kubatko (Eds.), Estimating Species Trees: Practical and Theoretical Aspects. Wiley-Blackwell, Hoboken, New Jersey (Chapter 1).

Kubatko, L. S., J. H. Degnan (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. Systematic Biology 56 (1): 17–24. http://dx.doi. org/10.1080/10635150601146041.

Kubatko, L. S., B. C. Carstens, L. L. Knowles (2009). STEM: species tree estimation using maximum likelihood for gene trees under coalescence. Bioinformatics 25 (7): 971–973. http://dx.doi.org/10.1093/bioinformatics/btp079.

Kury, A. B. (2013). Order Opiliones Sundevall, 1833. In: Zhang, Z.-Q. (Ed.), Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa 3703 (1): 27–33. http://dx.doi.org/ 10.11646/zootaxa.3703.1.7.

Kury, A., A. Mendes, D. Souza (2014). World checklist of Opiliones species (Arachnida). Part 1: Laniatores–Travunioidea and Triaenonychoidea. Biodiversity Data Journal 2: e4094. http://dx.doi.org/10.3897/BDJ.2.e4094.

Lambert, S. M., T. W. Reeder, J. J. Wiens (2015). When do species-tree and concatenated estimates disagree? An empirical analysis with higher-level scincid lizard phylogeny. Molecular Phylogenetics and Evolution Part A 82: 146–155. http://dx. doi.org/10.1016/j.ympev.2014.10.004.

Lanfear, R., B. Calcott, S. Y. Ho, S. Guindon (2012). PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29 (6): 1695–1701. http://dx.doi.org/10.1093/molbev/ mss020.

Lang, J. M., A. E. Darling, J. A. Eisen (2013). Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. PLoS One 8 (4): e62510. http://dx.doi.org/10.1371/journal.pone.0062510.

Lanier, H. C., L. L. Knowles (2015). Applying species-tree analyses to deep phylogenetic histories: challenges and potential suggested from a survey of empirical phylogenetic studies. Molecular Phylogenetics and Evolution 83: 191–199. http://dx. doi.org/10.1016/j.ympev.2014.10.022.

Larget, B. R., S. K. Kotha, C. N. Dewey, C. Ané (2010). BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26 (22): 2910–2911. http://dx.doi.org/10.1093/bioinformatics/btq539.

Leaché, A. D., B. Rannala (2011). The accuracy of species tree estimation under simulation: a comparison of methods. Systematic Biology 60 (2): 126–137. http://dx.doi. org/10.1093/sysbio/syq073.

Leaché, A. D., R. B. Harris, B. Rannala, Z. Yang (2014a). The influence of gene flow on species tree estimation: a simulation study. Systematic Biology 63 (1): 17–30. http://dx. doi.org/10.1093/sysbio/syt049.

Leaché, A. D., P. Wagner, C. W. Linkem, W. Böhme, T. J. Papenfuss, R. A. Chong, B. R. Lavin, A. M. Bauer, S. V. Nielsen, E. Greenbaum, M.-O. Rödel, A. Schmitz, M. LeBreton, I. Ineich, L. Chirio, C. Ofori-Boateng, E. A. Eniang, S. Baha El Din, A. R. Lemmon, F. T. Burbrink (2014b). A hybrid phylogenetic–phylogenomic approach for species tree estimation in African Agama lizards with applications to biogeography, character evolution, and diversification. Molecular

Phylogenetics and Evolution 79: 215–230. http://dx.doi.org/10.1016/j.ympev.2014.06.013.

Lemmon, E. M., A. R. Lemmon (2013). High-throughput genomic data in systematics and phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44 (1): 99–121. http://dx.doi.org/ 10.1146/annurev-ecolsys-110512-135822.

Liu, L., S. V. Edwards (2009). Phylogenetic analysis in the anomaly zone. Systematic Biology 58 (4): 452–460. http://dx.doi.org/10.1093/sysbio/syp034.

Liu, L., D. K. Pearl (2007). Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Systematic Biology 56 (3): 504–514. http://dx.doi.org/10.1080/ 10635150701429982.

Liu, L., L. Yu, D K. Pearl, S. V. Edwards (2009). Estimating species phylogenies using coalescence times among sequences. Systematic Biology 58 (5): 468–477. http://dx.doi. org/10.1093/sysbio/syp031.

Liu, L., L. Yu, S. Edwards (2010). A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evolutionary Biology 10 (1): 302. http://dx.doi.org/10.1186/1471-2148-10-302.

Machado, G., R. Pinto-da-Rocha, G. Giribet (2007). What are harvestmen? In: Pinto-da-Rocha, R., Machado, G., Giribet, G. (Eds.), Harvestmen: The Biology of Opiliones. Harvard University Press, Cambridge, MA, pp. 1–13 (Chapter 1).

Maddison, W. P. (1997). Gene trees in species trees. Systematic Biology 46 (3): 523–536. http:// dx.doi.org/10.1093/sysbio/46.3.523.

Maddison, W. P., L. L. Knowles (2006). Inferring phylogeny despite incomplete lineage sorting. Systematic Biology 55 (1): 21–30. http://dx.doi.org/10.1080/ 10635150500354928.

Martens, J. (1976). Genitalmorphologie, system und phylogenie der Weberknechte (Arachnida: Opiliones). Entomologica Germanica 3 (1–2): 51–68.

Martens, J. (2015). *Sabacon* Simon, 1879 in the Palaearctic: a survey of new and known species from France, Nepal, India, China, Russia and Japan (Arachnida: Opiliones: Sabaconidae). In: Biodiversität und Naturausstattung im Himalaya V. Hartmann and Weipert, Erfurt, pp. 167–210.

Martens, J., U. Hoheisel, M. Götze (1981). Vergleichende anatomie der legeröhren der Opiliones als beitrag zur phylogenie der ordnung (Arachnida). Zoologische

Jahrbücher. Abteilung für Anatomie und Ontogenie der Tiere Abteilung für Anatomie und Ontogenie der Tiere 105: 13–76.

McCormack, J. E., B. C. Faircloth, N. G. Crawford, P. A. Gowaty, R. T. Brumfield, T. C. Glenn (2012). Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Research 22 (4): 746–754. http://dx.doi.org/10.1101/ gr.125864.111.

McCormack, J. E., M. G. Harvey, B. C. Faircloth, N. G. Crawford, T. C. Glenn, R. T. Brumfield (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8 (1): e54848.

Meredith, R. W., J. E. Janečka, J. Gatesy, O. A. Ryder, C. A. Fisher, E. C. Teeling, A. Goodbla, E. Eizirik, T. L. L. Simão, T. Stadler, D. L. Rabosky, R. L. Honeycutt, J. J. Flynn, C. M. Ingram, C. Steiner, T. L. Williams, T. J. Robinson, A. Burk-Herrick, M. Westerman, N. A. Ayoub, M. Springer, W. J. Murphy (2011). Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334 (6055): 521–524. http://dx.doi.org/10.1126/science.1211028.

Miller, M. A., W. Pfeiffer, T. Schwartz (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop (GCE), 14 November 2010, New Orleans, LA, pp. 1–8. http://dx.doi.org/10.1109/GCE.2010.5676129.

Money, D., S. Whelan (2011). Characterizing the phylogenetic tree-search problem. Systematic Biology: syr097. http://dx.doi.org/10.1093/sysbio/syr097.

Murphy, W. J., E. Eizirik, S. J. O'Brien, O. Madsen, M. Scally, C. J. Douady, E. Teeling, O. A. Ryder, M. J. Stanhope, W. W. de Jong, M. S. Springer (2001). Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294 (5550): 2348–2351. http://dx.doi.org/10.1126/science.1067179.

Nordborg, M. (2001). Coalescent theory. In: Balding, D., M. Bishop, C. Cannings, (Eds.), Handbook of Statistical Genetics. Wiley, Chichester, pp. 179–212.

O'Neill, E. M., R. Schwartz, C. T. Bullock, J. S. Williams, H. B. Shaffer, X. Aguilar-Miguel, G. Parra-Olea, D. W. Weisrock (2013). Parallel tagged amplicon sequencing reveals major lineages and phylogenetic structure in the North American tiger salamander (*Ambystoma tigrinum*) species complex. Molecular Ecology 22: 111–129. http://dx.doi.org/10.1111/mec.12049.

Pamilo, P., M. Nei (1988). Relationships between gene trees and species trees. Molecular Biology and Evolution 5 (5): 568–583.

Patel, S., R. T. Kimball, E. L. Braun (2013). Error in phylogenetic estimation for bushes in the tree of life. Journal of Phylogenetics & Evolutionary Biology 1 (2): 110. http://dx.doi.org/10.4172/ 2329-9002.1000110.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide, D. Baurain (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biology 9 (3): e1000602. http://dx.doi.org/ 10.1371/journal.pbio.1000602.

Pollock, D. D., D. J. Zwickl, J. A. McGuire, D. M. Hillis (2002). Increased taxon sampling is advantageous for phylogenetic inference. Systematic Biology 51 (4): 664–671. http:// dx.doi.org/10.1080/10635150290102357.

Raymann, K., P. Forterre, C. Brochier-Armanet, S. Gribaldo (2014). Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in Archaea. Genome Biology and Evolution 6 (1): 192–212. http://dx.doi.org/ 10.1093/gbe/evu004.

Regier, J. C., J. W. Shultz, A. R. D. Ganley, A. Hussey, D. Shi, B. Ball, A. Zwick, J. E. Stajich, M. P. Cummings, J. W. Martin, C. W. Cunningham (2008). Resolving arthropod phylogeny: exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence. Systematic Biology 57 (6): 920–938. http://dx.doi.org/10.1080/ 10635150802570791.

Richart, C., M. Hedin (2013). Three new species in the harvestmen genus *Acuclavella* (Opiliones, Dyspnoi, Ischyropsalidoidea), including description of male *Acuclavella quattuor* Shear, 1986. ZooKeys 311: 19–68. http://dx.doi.org/ 10.3897/zookeys.311.2920.

Romiguier, J., V. Ranwez, F. Delsuc, N. Galtier, E. J. Douzery (2013). Less is more in mammalian phylogenomics: AT-rich genes minimize tree conflicts and unravel the root of placental mammals. Molecular Biology and Evolution 30 (9): 2134–2144. http://dx.doi. org/10.1093/molbev/mst116.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61 (3): 539–542. http://dx.doi.org/10.1093/sysbio/sys029.

Rota-Stabelli, O., L. Campbell, H. Brinkmann, G. D. Edgecombe, S. J. Longhorn, K. J. Peterson, D. Pisani, H. Philippe, M. J. Telford (2011). A congruent solution to arthropod phylogeny: phylogenomics, microRNAs and morphology support monophyletic Mandibulata. Proceedings of the Royal Society B: Biological Sciences 278 (1703): 298–306. http://dx.doi.org/10.1098/rspb.2010.0590.

Rozen, S., H. Skaletsky (2000). Primer3 on the WWW for general users and for biologist programmers. In: Misener, S., S. A. Krawetz (Eds.), Bioinformatics Methods and Protocols. Humana Press, pp. 365–386.

Salichos, L., A. Rokas (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497 (7449): 327–331. http://dx.doi.org/ 10.1038/nature12130.

Scally, A., J. Y. Dutheil, L. W. Hillier, G. E. Jordan, I. Goodhead, J. Herrero, A. Hobolth, T. Lappalainen, et al. (2012). Insights into hominid evolution from the gorilla genome sequence. Nature 483 (7388): 169–175. http://dx.doi.org/ 10.1038/nature10842.

Schmieder, R., R. Edwards (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics 27 (6): 863–864. http://dx.doi.org/10.1093/ bioinformatics/btr026.

Schönhofer, A. L. (2013). A taxonomic catalogue of the Dyspnoi Hansen and Sørensen, 1904 (Arachnida: Opiliones). Zootaxa 3679 (1): 1–68.

Schönhofer, A. L., M. McCormack, N. Tsurusaki, J. Martens, M. Hedin (2013). Molecular phylogeny of the harvestmen genus *Sabacon* (Arachnida: Opiliones: Dyspnoi) reveals multiple Eocene–Oligocene intercontinental dispersal events in the Holarctic. Molecular Phylogenetics and Evolution 66 (1): 303–315. http://dx.doi.org/ 10.1016/j.ympev.2012.10.001.

Schönhofer, A. L., C. Vernesi, J. Martens, M. Hedin (2015). Molecular phylogeny, biogeographic history, and evolution of cave-dwelling taxa in the European harvestmen genus *Ischyropsalis* (Opiliones: Dyspnoi). Journal of Arachnology 43 (1): 40–53. http://dx.doi.org/10.1636/H14-39.1.

Seo, T. K. (2008). Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Molecular Biology and Evolution 25 (5): 960–971. http://dx.doi.org/ 10.1093/molbev/msn043.

Sharma, P., G. Giribet (2014). A revised dated phylogeny of the arachnid order Opiliones. Evolutionary and Population Genetics 5: 255. http://dx.doi.org/10.3389/ fgene.2014.00255.

Shaw, T. I., Z. Ruan, T. C. Glenn, L. Liu (2013). STRAW: species TRee Analysis Web server. Nucleic Acids Research 41 (W1): W238–W241. http://dx.doi.org/10.1093/nar/ gkt377.

Shear, W. A. (1986). A cladistic analysis of the opilionid superfamily Ischyropsalidoidea, with descriptions of the new family Ceratolasmatidae, the new genus *Acuclavella*,

and four new species. American Museum Novitates, 2844: 1–29. URI: http://hdl.handle.net/2246/3579.

Shultz, J. W. (1998). Phylogeny of Opiliones (Arachnida): an assessment of the "Cyphopalpatores" Concept. Journal of Arachnology 26 (3): 257–272.

Shultz, J. W., J. C. Regier (2001). Phylogenetic analysis of Phalangida (Arachnida, Opiliones) using two nuclear protein-encoding genes supports monophyly of Palpatores. Journal of Arachnology 29 (2): 189–200.

Slatkin, M., J. L. Pollack (2008). Subdivision in an ancestral species creates asymmetry in gene trees. Molecular Biology and Evolution 25 (10): 2241–2246. http://dx.doi.org/ 10.1093/molbev/msn172.

Song, S., L. Liu, S. V. Edwards, S. Wu (2012). Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proceedings of the National Academy of Sciences 109 (37): 14942–14947. http://dx.doi.org/10.1073/ pnas.1211733109.

Springer, M. S., J. Gatesy (2014). Land plant origins and coalescence confusion. Trends in Plant Science 19 (5): 267–269. http://dx.doi.org/10.1016/j.tplants.2014.02.012.

Springer, M. S., J. Gatesy (2016). The gene tree delusion. Molecular Phylogenetics and Evolution 94: 1– 33. http://dx.doi.org/10.1016/j.ympev.2015.07.018.

Stamatakis, A., P. Hoover, J. Rougemont (2008). A rapid bootstrap algorithm for the RAxML web servers. Systematic Biology 57 (5): 758–771. http://dx.doi.org/10.1080/ 10635150802429642.

Swofford, D. L., P. J. Waddell, J. P. Huelsenbeck, P. G. Foster, P. O. Lewis, J. S. Rogers (2001). Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Systematic Biology 50 (4): 525–539. http://www. jstor.org/stable/3070852.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, S. Kumar (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30 (12): 2725–2729. http://dx.doi.org/10.1093/molbev/mst197.

Teeling, E. C., S. B. Hedges (2013). Making the impossible possible: rooting the tree of placental mammals. Molecular Biology and Evolution 30 (9): 1999–2000. http://dx.doi.org/ 10.1093/molbev/mst118.

Vink, C. J., S. M. Thomas, P. Paquin, C. Y. Hayashi, M. Hedin (2005). The effects of preservatives and temperatures on arachnid DNA. Invertebrate Systematics 19 (2): 99–104. http://dx.doi.org/10.1071/IS04039.

Zhang, N., L. Zeng, H. Shan, H. Ma (2012). Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytologist 195 (4): 923–937. http://dx.doi.org/10.1111/j.1469-8137.2012.04212.x.

Zwickl, D. J., J. C. Stein, R. A. Wing, D. Ware, M. J. Sanderson (2014). Disentangling methodological and biological sources of gene tree discordance on *Oryza* (Poaceae) Chromosome 3. Systematic Biology 63 (5): 645–659. http://dx.doi.org/10.1093/ sysbio/syu027.

Table 1.1. Primer combinations used to isolate and amplify targeted loci and annotation from BLAST to Ixodes genome.

Primer Name	Primer Combination, 5' to 3'	Annotation
008_F_184_HO 008_R_968_HO	GGGGAAGGKTGCGTGGATAAACCWC CCAGAYTCRTCYGAAACGCTGTCCTC	Similar to *Ixodes* ISCW000585-PA; fam43, putative
017_F_242a_HOT 017_F_242b_HOT 017_R_760_HOT	GCCCCAGGCGAGTATCTTACGCAGG GCCCCAGGCGAGTATCTTACGCAGGTTC CGCTRCCCCAGAGMARTTCGGTTTTAAC	Similar to *Ixodes* ISCW001223-PA; Mid1-interacting protein, putative
049_F_194_HOT 049_R_845_HT	GCAACAAYGARCATTGTCAGGAAGG CCYTCGTCTTCCATYTTGATGTGGTASG	Similar to *Ixodes* ISCW004228-PA; hypothetical protein
069_F_163_HOT 069_R_545_HOT	GCYCTTCGWCCCTGGTCATTCAGYG CCTCCAGTGTARAGAAAMGAACTGG	Similar to *Ixodes* ISCW005714-PA; 1,4-dihydroxy-2-naphthoate octaprenyltransferase, putative
156_F_790_HOT 156_R_1283_HO	CGYCARTATCTTTTCTGCATGACACC GCMGAMACAAARACTTGMGCAATATCATC	Similar to *Ixodes* ISCW012356-PA; conserved hypothetical protein
281_F_1353_HT 281_F_1469_HO 281_R_2019_HOT	CCGATTATTCMAARAAYCARGATG GCCACCAARTTGATMCTSTTYAACG GGWGTCATGAACTGCTGKATCATRTTC	Similar to *Ixodes* ISCW020128-PA; mediator of RNA polymerase II transcription subunit, putative
300_F_730_HT 300_R_1565_HOT	CCACCTCCACCMGTYATGCATTC GGATCTTCTTCYTCTTGYTTYTC	Similar to *Ixodes* ISCW021146-PA; ski interacting protein, putative
334_F_200_HO 334_R_760_HO	GCAGAGAASCATTCMGWAGAGC CRCAWGTYTGTTTRAAGAARC	Similar to *Ixodes* ISCW023542-PA; conserved hypothetical protein

Table 1.2. Operational taxonomic units (OTUs) and data type used to populate expanded panel phylogenetic analyses.

OTU	Species	Data Type	Voucher#	Locality	Latitude°	Longitude°	Date
Ortholasma	*O. coronadense* Cockerell, 1916	Transcriptome	SRX451776	CA: San Diego Co., San Diego			
		PCR+Sanger	OP2879	CA: Santa Barbara Co., Figueroa Mtn Rd	34.7409	-120.0621	25-Mar-11
Trogulus	*T. martensi* Chemini, 1983	Transcriptome	SRX450964	Germany, Baden-Württemberg, Unteruhldingen			
		PCR+Sanger	OP2857	Germany, Bodendsee	47.7480	9.2487	28-Aug-10
Hesperonemastoma A	*H.* cf. *modestum* (Banks, 1894)	Transcriptome	SRX450937	CA: San Diego Co., Palomar Mt			
		PCR+Sanger	OP2848	CA: Kern Co., Water Canyon Rd.	35.0839	-118.4947	8-Oct-10
		PCR+Sanger	CHR3794	CA: San Diego Co., Palomar Mt	33.3272	-116.8892	5-Jul-12
		PCR+Sanger	AXLS1180	CA: Tulare Co., Three Rivers	36.4538	-118.8252	15-Mar-11
Hesperonemastoma B	*Hesperonemastoma kepharti* (Crosby & Bishop, 1924)	PCR+Sanger	AXLS1219	NC: Buncombe Co., Candler, Mt. Pisgah Hwy	35.4497	-82.7226	16-Jun-12
	H. cf. *kepharti*	PCR+Sanger	OP2611	AR: Montgomery Co., E. Albert Pike Rec. Area	34.3818	-93.8785	25-Jun-09
Taracus A	*Taracus packardi* Simon, 1879	PCR+Sanger	OP1254	CO: Chaffee Co., North Monarch Pass	38.4979	-106.3323	29-Jun-07
Taracus B	*Taracus* cf. *pallipes* Banks, 1894	PCR+Sanger	OP1787	OR: Clatsop Co., Saddle Mountain State Park	45.9626	-123.6900	17-Jun-07
Sabacon A	*Sabacon cavicolens* (Packard, 1884)	Transcriptome		NC: Swain Co., Nantahala River Gorge			23-Oct-12
		PCR+Sanger	OP2576	IL: Johnson Co., Fern Clyffe State Park	37.5385	-88.9801	20-Jun-09
Sabacon B	*Sabacon occidentalis* (Banks, 1894)	PCR+Sanger	OP2700	WA: Clallam Co., Hoko Falls	48.1940	-124.4499	5-Aug-13
Acuclavella A	*Acuclavella makah* Richart & Hedin, 2013	PCR+Sanger	OP2717	WA: Clallam Co., tributary of Sitka River	47.9571	-124.2215	6-Aug-13
Acuclavella B	*Acuclavella merickeli* Shear, 1986	Transcriptome		ID: Idaho Co., Falls Point Rd	46.0422	-115.2958	23-Jul-11
		PCR+Sanger	OP2238	ID: Idaho Co., FS Rd 443	46.0385	-115.2943	7-Jul-12
Ceratolasma A	*Ceratolasma tricantha* Goodnight & Goodnight, 1942	PCR+Sanger	OP1810	OR: Coos Co., Charleston, OR Inst. Marine Bio.	43.3443	-124.3287	26-May-07
Ceratolasma B	*Ceratolasma tricantha*	PCR+Sanger	OP2229	WA: Lewis Co., Rainbow Falls State Park	46.6301	-123.2330	9-Aug-12
Ischyropsalis A	*Ischyropsalis kollari* Koch, 1839	PCR+Sanger	AXLS914	Austria, Carinthia, Treffling, Tschiernock Mtn	46.8528	13.5879	19-Jun-10
		PCR+Sanger	AXLS296	Austria, Steiermark, Koralpe, Wolfsberg	46.7987	14.9509	31-Aug-09
	Ischyropsalis nodifera Simon, 1879	PCR+Sanger	CJM5570	Spain: Álava, País Vasco	43.0733	-2.6619	27-Apr-07
Ischyropsalis B	*Ischyropsalis h. hellwigi* (Panzer, 1794)	PCR+Sanger	AXLS240	Slovenia: Sadni Tracniki, Olševa Mountains	46.4571	14.6861	27-Aug-09

Table 1.3. GenBank numbers, expanded panel matrix, and alignment lengths. Italicized GenBank accession numbers represent sequences downloaded from GenBank. Newly developed loci are in the upper table, previously used loci are continued on the following page.

OTU	**8**	**17**	**49**	**69**	**156**	**281**	**300**	**334**
Ortholasma	KU168429	KU168438	KU168456	KU168457	KU168473	KU168483	KU168491	KU168498
Trogulus	KU168430	KU168439	KU168446	KU168458	KU168474	KU168476	KU168492	KU168499
Hesperonemastoma A	KU168431	KU168440	KU168447	KU168459	KU168466	KU168477	KU168493	KU168500
Hesperonemastoma B	KU168432	KU168441	KU168448		KU168467	KU168478		KU168495
Taracus A			KU168449	KU168460	KU168468	KU168479	KU168485	KU168496
Taracus B			KU168450		KU168469	KU168480	KU168486	KU168497
Sabacon A	KU168437	KU168442	KU168451	KU168464	KU168475	KU168481	KU168494	KU168501
Sabacon B		KU168443	KU168452	KU168461		KU168482		
Acuclavella A	KU168433		KU168453	KU168462	KU168470		KU168487	
Acuclavella B	KU168434	KU168445	KU168454	KU168465	KU168471	KU168484	KU168488	KU168502
Ceratolasma A	KU168435		KU168455				KU168489	
Ceratolasma B	KU168436			KU168463	KU168472		KU168490	
Ischyropsalis A		KU168444						
Ischyropsalis B								
n=	9	8	11	9	10	9	10	8
Align. Length (BPs) =	729	456	588	324	438	612	789	517

Table 1.3 (continued). GenBank numbers, expanded panel matrix, and alignment lengths for previously used loci. Italicized GenBank accession numbers represent sequences downloaded from GenBank. Newly developed loci are reported on the previous page.

OTU	**COI**	**EF1a**	**polII**	**r18S**	**r28S**	**Wnt2**	***n*=**
Ortholasma	*GQ912870*	*KU168506*	KU168516	KU168520	KU168533		13
Trogulus	*GQ912872*	*AF240880*	KU168517	KU168521	KU168534	KU168542	14
Hesperonemastoma A	*EF108588*	*AF240869*	KU168518	KU168522	KU168535	KU168543	14
Hesperonemastoma B	KU168503	KU168507		KU168523	KU168536	KU168544	11
Taracus A		KU168508		KU168524	*JX573592*	KU168545	10
Taracus B	*GQ912867*	KU168509	*AH010475*	KU168525		KU168546	10
Sabacon A	KU168505	KU168510	*AH010471*	KU168526	KU168537	KU168547	14
Sabacon B	*JX573670*	KU168511		KU168527	KU168538	KU168548	9
Acuclavella A	*GQ870647*	KU168512		KU168528	KU168539	KU168549	10
Acuclavella B	*GQ870645*	KU168513	KU168519	KU168529	KU168540	KU168550	14
Ceratolasma A	*GQ912865*	KU168514	*AH010458*	KU168530	KU168541		8
Ceratolasma B		KU168515		KU168531		KU168551	7
Ischyropsalis A	KU168504	*JX573604*	*AH010464*	KU168532	*JX573546*	KU168552	7
Ischyropsalis B	*JX573639*	*JX573603*			*JX573545*		3
n=	12	14	8	13	12	11	
Align. Length (BPs) =	1098	672	1137	547	1110	390	**9407**

Table 1.4. Partitions and models used for expanded panel phylogenetic analyses.

ID	Partition	Beast Order	Best Model	Beast Model	RAxML order	RAxML model	MrBayes Order	MrBayes Model
8	1-729\3	1	SYM+I	GTR+I+G	1	GTR+G	1	SYM+I
	2-729\3	2	HKY				2	HKY
	3-729\3	3	HKY+G		2	GTR+G	3	HKY+G
17	1-456\3	4	K80+G	HKY+G	3	GTR+G	4	JC+G
	2-456\3							
	3-456\3	5	HKY+G		4	GTR+G	5	HKY+G
49	1-588\3	6	TrN+G	TN93+G	5	GTR+G	6	HKY+G
	2-588\3							
	3-588\3	7	HKY+G		6	GTR+G	7	HKY+G
69	1-324\3	8	TrNef+G	GTR+G	7	GTR+G	8	SYM+G
	2-324\3	9	GTR		8	GTR+G	9	JC
	3-324\3	10	K80+G		9	GTR+G	10	K80+G
156	1-438\3	11	SYM+I	GTR+I+G	10	GTR+G	11	SYM+I
	2-438\3	12	K80				12	K80
	3-438\3	13	HKY+G		11	GTR+G	13	HKY+G
281	1-612\3	14	SYM+I	GTR+I+G	12	GTR+G	14	SYM+I
	2-612\3	15	TrNef+I		13	GTR+I+G	15	JC+I
	3-612\3	16	TrNef+G		14	GTR+G	16	K80+G
300	1-789\3	17	GTR+I	GTR+I+G	15	GTR+G	17	GTR+I
	2-789\3	18	HKY+I		16	GTR+G	18	F81+I
	3-789\3	19	TrN+G		17	GTR+G	19	HKY+G
334	1-517\3	20	K80+G	GTR+G	18	GTR+G	20	K80+G
	2-517\3	21	GTR+G				21	GTR+G
	3-517\3	22	TrN+G		19	GTR+G	22	HKY+G
COI	1-1098\3	23	SYM+I	GTR+I+G	20	GTR+G	23	SYM+G
	2-1098\3	24	HKY+I		21	GTR+I+G	24	HKY+I
	3-1098\3	25	GTR+G		22	GTR+I+G	25	GTR+I+G
EF1a	1-672\3	26	TrNef+I	TN93+I+G	23	GTR+I+G	26	K80+I
	2-672\3							
	3-672\3	27	TrNef+G		24	GTR+G	27	HKY+G
r18S	1-547	28	K80+I	HKY+I	25	GTR+G	28	K80+I
r28S	1-1110	29	GTR+G	GTR+G	26	GTR+G	29	GTR+G
polII	1-1137\3	30	GTR+I	GTR+I+G	27	GTR+G	30	GTR+I
	2-1137\3	31	HKY+I		28	GTR+G	31	F81+I
	3-1137\3	32	TrN+G		29	GTR+G	32	HKY+G
Wnt2	1-390\3	33	SYM+I	GTR+I+G	30	GTR+G	33	SYM+I
	2-390\3	34	K80				34	K80
	3-390\3	35	HKY+G		31	GTR+G	35	HKY+G

Table 1.5. Voucher numbers for sequences generated via Sanger sequencing or Transcript for sequences mined from transcriptome genomics. Newly developed loci are in the upper table, previously used loci are on the following page.

OTU	**8**	**17**	**49**	**69**	**156**	**281**	**300**	**334**
Ortholasma	OP2879	OP2879	Transcript.	OP2879	Transcript.	Transcript.	Transcript.	Transcript.
Trogulus	OP2857	OP2857	OP2857	OP2857	Transcript.	OP2857	Transcript.	Transcript.
Hesperonemastoma A	OP2848	OP2848	OP2848	OP2848	CHR3794	OP2848	Transcript.	Transcript.
Hesperonemastoma B	OP2611	OP2611	OP2611		OP2611	OP2611		OP2611
Taracus A			OP1254	OP1254	OP1254	OP1254	OP1254	OP1254
Taracus B			OP1787		OP1787	OP1787	OP1787	OP1787
Sabacon A	Transcript.	OP2576	OP2576	Transcript.	Transcript.	OP2576	Transcript.	Transcript.
Sabacon B		OP2700	OP2700	OP2700		OP2700		
Acuclavella A	OP2717		OP2717	OP2717	OP2717		OP2717	
Acuclavella B	OP2238	Transcript.	OP2238	Transcript.	OP2238	Transcript.	OP2238	Transcript.
Ceratolasma A	OP1810		OP1810				OP1810	
Ceratolasma B	OP2229			OP2229	OP2229		OP2229	
Ischyropsalis A		AXLS296						
Ischyropsalis B								
n=	9	8	11	9	10	9	10	8
Align. Length (BPs) =	729	456	588	324	438	612	789	517

Table 1.5 (continued). Voucher numbers for sequences generated via Sanger sequencing or Transcript for sequences mined from transcriptome genomics. Newly developed loci are in the upper table, previously used loci are on the following page.

OTU	COI	EF1a	polII	r18S	r28S	Wnt2
Ortholasma	GQ912870	OP2879	Transcript.	OP2879	OP2879	
Trogulus	GQ912872	AF240880	Transcript.	OP2857	OP2857	OP2857
Hesperonemastoma A	EF108588	AF240869	Transcript.	OP2848	AXLS1180	CHR3794
Hesperonemastoma B	AXLS1219	OP2611		OP2611	OP2611	OP2611
Taracus A		OP1254		OP1254	JX573592	OP1254
Taracus B	GQ912867	OP1787	AH010475	OP1787		OP1787
Sabacon A	Transcript.	OP2576	AH010471	OP2576	OP2576	OP2576
Sabacon B	JX573670	OP2700		OP2700	OP2700	OP2700
Acuclavella A	GQ870647	OP2717		OP2717	OP2717	OP2717
Acuclavella B	GQ870645	OP2238	Transcript.	OP2238	OP2238	OP2238
Ceratolasma A	GQ912865	OP1810	AH010458	OP1810	OP1810	
Ceratolasma B		OP2229		OP2229		OP2229
Ischyropsalis A	AXLS914	JX573604	AH010464	AXLS296	JX573546	CJM5570
Ischyropsalis B	JX573639	JX573603			JX573545	
n=	12	14	8	13	12	11
Align. Length (BPs) =	1098	672	1137	547	1110	390

Table 1.6. Transcriptome data and assembly information.

Taxon	# of paired-end reads	# Gb	# transcripts >200 bp	mean length (>200 bp)	Max length
Ortholasma	80.7 M (50-bp)	4.04	34,357	839.6	11,074
Trogulus	54.7 M (50-bp)	2.74	46,840	937.4	9,614
Hesperonemastoma	120.0 M (50-bp)	6.00	42,007	999.1	8,952
Acuclavella	60.9 M (100-bp)	6.09	20,926	1494.9	36,044
Sabacon	43.3 M (100-bp)	4.33	24,135	1121.7	12,424

Table 1.7. Occurrence of minority gene trees and probability of equal RFs using a two-sided binomial test.

Treatment	A,(H,S)	S,(H,A)	p =
All Loci	238	181	0.0062
BS Values	67	28	< 0.0001
% AT3 Content	63	42	0.0504
Hedin et al. 2012	69	51	0.1203

Figures

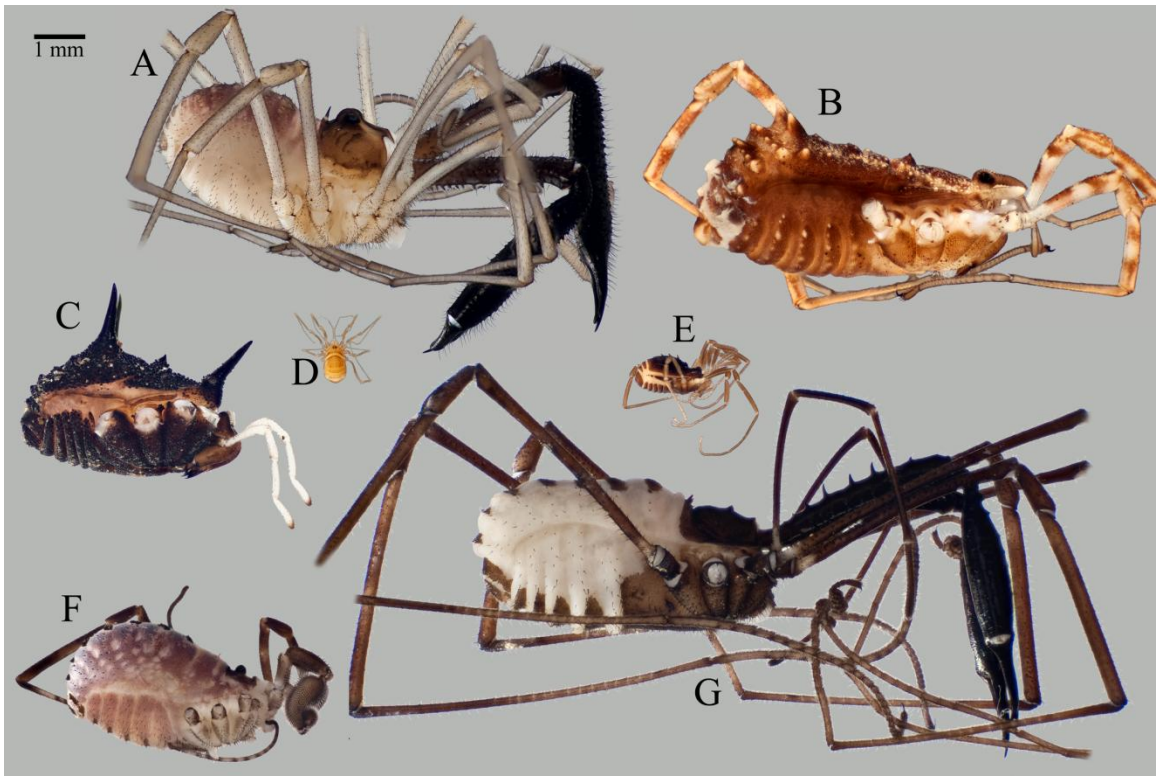


Figure 1.1. Ischyropsalidoidea. Generic representatives from the superfamily Ischyropsalidoidea. A. *Taracus gertschi* (851092), B. *Ceratolasma tricantha* (850889), C. *Acuclavella makah* (829726), D. *Crosbycus dasycnemus* (851086), E. *Hesperonemastoma* sp. (851085), F. *Sabacon* sp. (851091), and G. *Ischyropsalis h. hellwigi* (851090). Full sized high-resolution images can be seen at MorphBank using the specimen identification numbers listed behind each name. Specimens were imaged using a Visionary Digital BK Plus system (http://www.visionarydigital.com) with composite images combined using Zerene Stacker 1.04 (http://www.zerenesystems.com), and edited with Adobe Photoshop CS6.

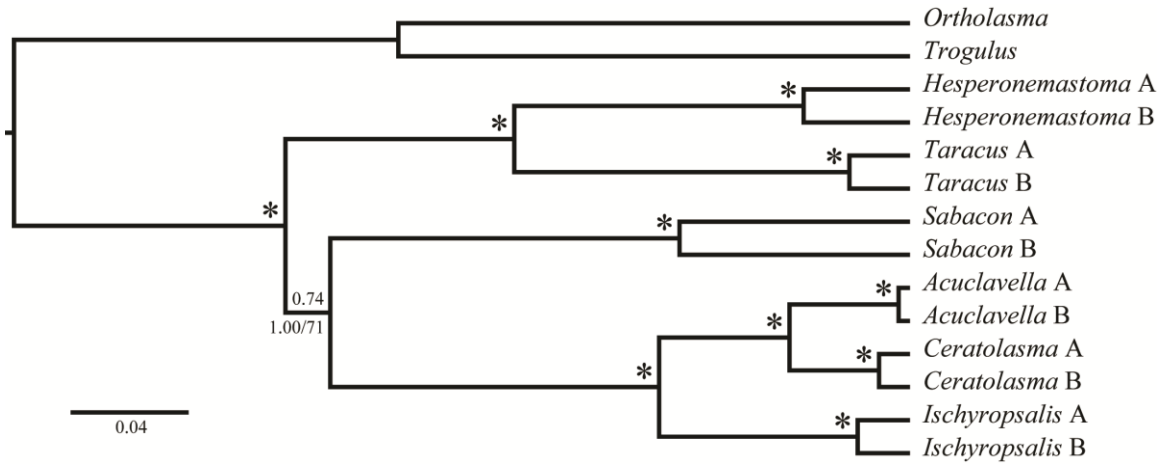


Figure 1.2. Expanded panel phylogeny. Phylogeny of Ischyropsalidoidea based on 14 genes analyzed via coalescent (*BEAST; topology pictured) and concatenated (MrBayes, RAxML) methods. The node with support values shows *BEAST above, and MrBayes/RAxML support values below the parent branch. All other nodes were recovered with *BEAST posteriors ≥ 0.96, MrBayes posteriors of 1.0, and RAxML bootstrap values ≥ 99. Identification of samples used to populate OTUs follows Appendix A and Table 1.

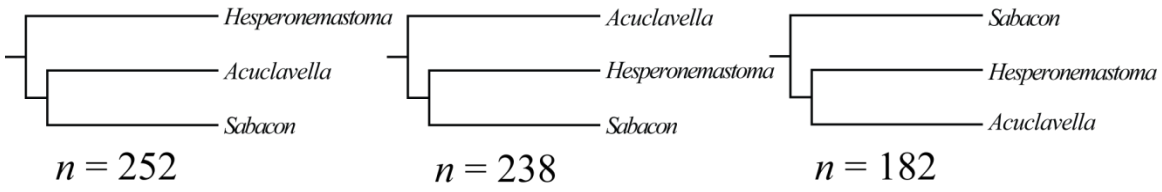


Figure 1.3. Gene tree synopsis. Results of PhyML gene tree analyses of 672 loci.

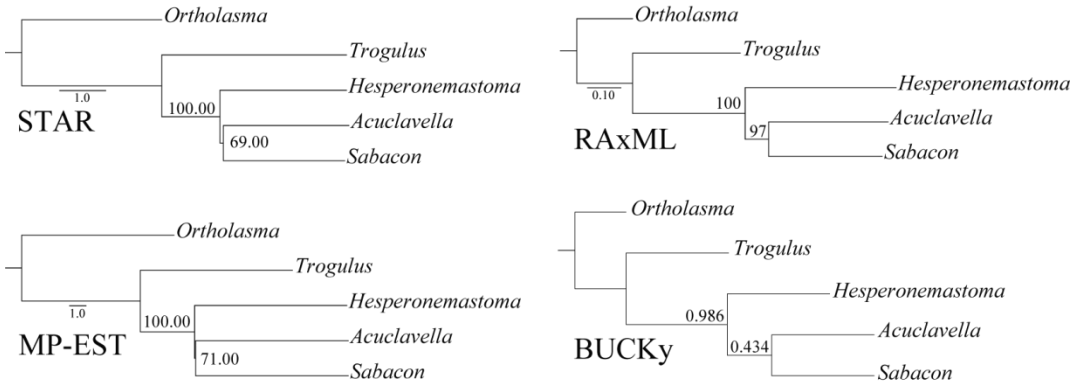
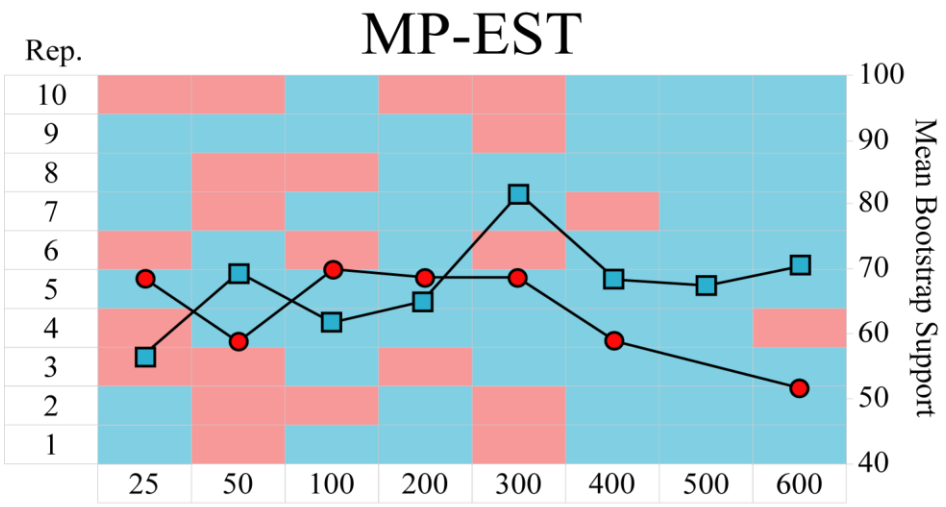
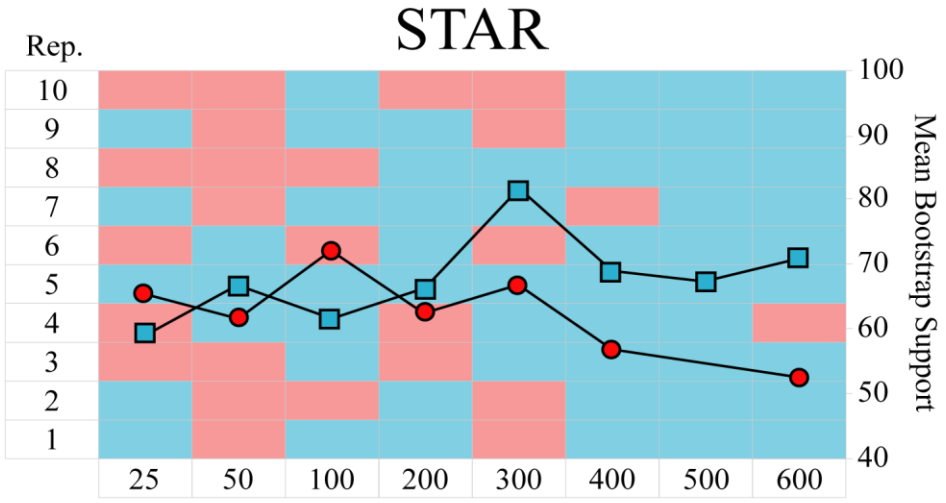
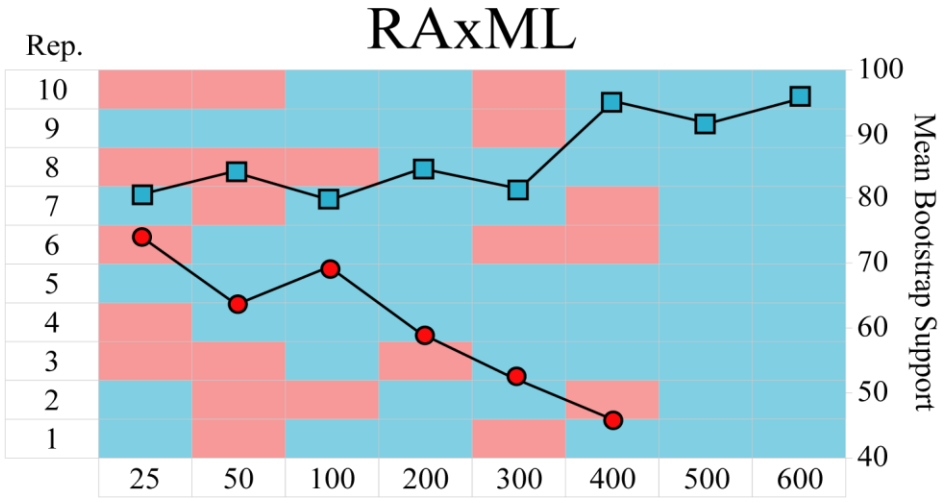


Figure 1.4. Transcriptome panel phylogenies and concordance tree. Phylogenies from analyses of 672 loci derived from transcriptomics. Bootstrap support values are shown for partially-parametric coalescent (STAR and MP-EST) and concatenation (RAxML) analyses. Also shown is the BUCKy primary concordance tree and associated concordance factors. Scale bars for STAR and MP-EST are in coalescent units; the RAxML scale depicts the number of substitutions per site.

Figure 1.5. (following page). Comparison across inference methods of randomly sampled loci. Comparison of partially-parametric coalescent (MP-EST and STAR) and concatenated supermatrix (RAxML) methods of phylogenetic inference across randomly sampled loci. Replicates are color-coded to represent recovered topologies. Blue: (*Hesperonemastoma*, (*Sabacon*, *Acuclavella*)). Red: (*A*, (*S*, *H*)). The x-axis is the number of loci per replicate. The y-axis is the mean bootstrap value, the average values of alternate topologies recovered from each replicate.


MP-EST
Rep.
10
9
8
7
6
5
4
3
2
1
25
50
100
200
300
400
500
600
100
90
80
70
60
50
40
Mean Bootstrap Support


STAR
Rep.
10
9
8
7
6
5
4
3
2
1
25
50
100
200
300
400
500
600
100
90
80
70
60
50
40
Mean Bootstrap Support


RAxML
Rep.
10
9
8
7
6
5
4
3
2
1
25
50
100
200
300
400
500
600
100
90
80
70
60
50
40
Mean Bootstrap Support

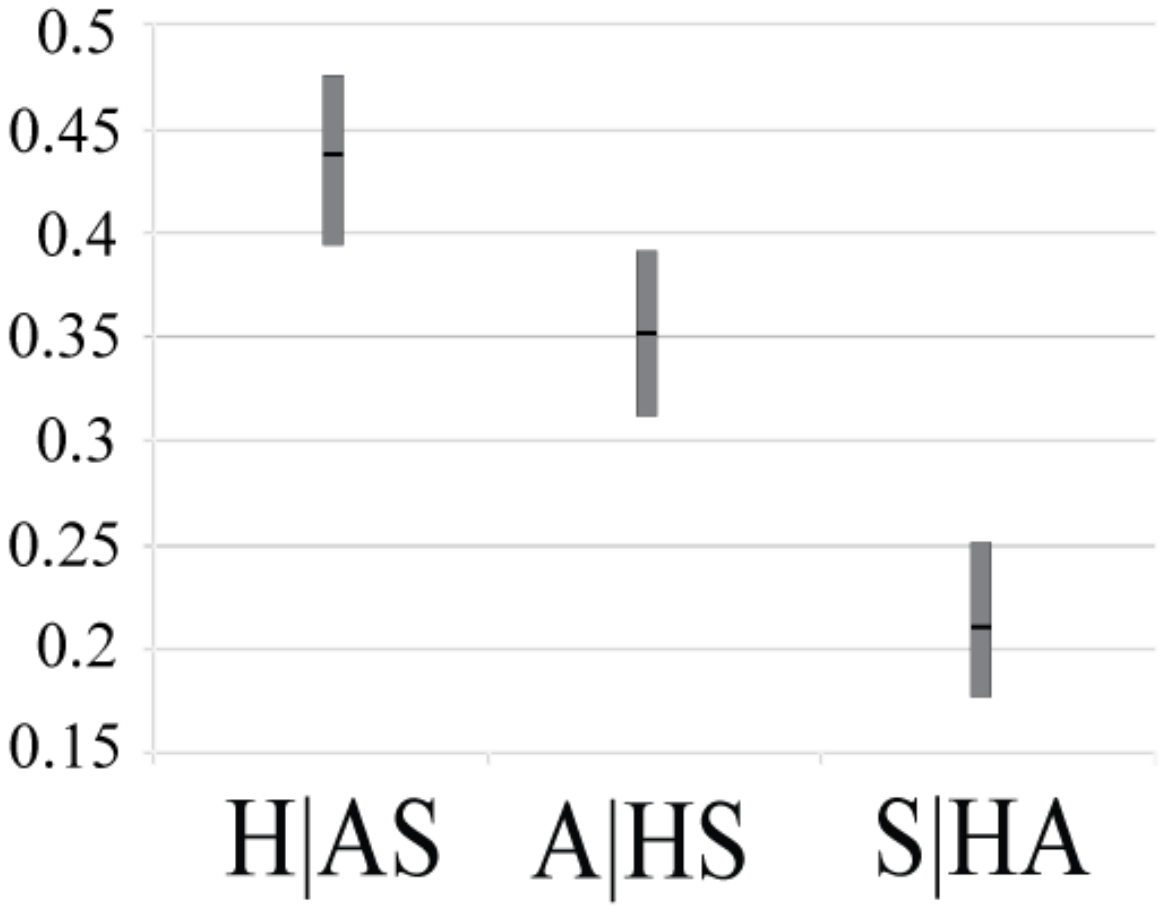


Figure 1.6. Concordance factors. Concordance factors of three alternate hypotheses. (H): *Hesperonemastoma*, (A): *Acuclavella*, and (S): *Sabacon*. Bars show the estimated concordance factor (black) and 99% confidence interval (gray).

CHAPTER 2

Integrating ddRAD-seq data and morphology to delimit species in the Thorn Harvestmen *Acuclavella* (Opiliones, Dyspnoi, Ischyropsalidoidea)

Accurate species delimitation is fundamental to biology. The delimitation of recently diverged species is the nontrivial process of identifying the temporal transition between micro- and macroevolutionary processes. In this research, species are delimited in the thorn harvestmen *Acuclavella* (Arachnida, Opiliones) using genomic-scale data and morphology. Simpson's species definition provides the theoretical underpinning for analyses operationally applied via the Generalized Lineage Construct in an integrative taxonomic framework. Morphological and molecular data are used to test existing taxonomic hypotheses, as well as to generate and reciprocally test novel species-level hypotheses. The inland species of *Acuclavella* from the Pacific Northwest of North America provide a challenging system for species delimitation, with incongruence among taxonomic disciplines suggesting rapid morphological evolution across multiple recent diversifications. Results suggest at least two undescribed species exist within inland populations of *Acuclavella*. Combinations of incongruence across taxonomic disciplines possibly represent different stages of the speciation continuum. Alternative combinations of incongruence suggest that different evolutionary mechanisms underlie speciation in thorn harvestmen, thus highlighting the importance of integrative species delimitation for evolutionary research.

I. INTRODUCTION

Species are fundamental units of biology. They are objectively real, in that they exist apart from our ability to detect them, but the species taxon is a hypothesis that may not reliably track the true number of species (Wheeler 2004, Hey 2009). Taxonomic reliability is important, for species hypotheses are the cornerstone of many research programs (Bik 2017). Numerous evolutionary processes are bound to the confines of a species. Population genetic analyses are more appropriately interpreted when species limits are accurately defined. Species inform sampling strategies in systematics, and allow for more correct inference of the rate and extent of diversification. Given the degree of importance these hypotheses have on evolutionary research, the species designation needs to be made clearly and carefully (Schlick-Steiner et al. 2010, Freudenstein et al. 2017). Similarly, formal species hypotheses are fundamental to other fields of biology (Sites and Marshall 2004). They are the primary unit of biodiversity and are essential to the successful execution of conservation biology (Agapow et al. 2004, Mace 2004). Many ecological studies would be impossible without the diagnostic resources provided by taxonomists. Accurate species delimitation is important in the interpretation of medicinal and biochemical research (Bortolus 2008). Taxonomic decisions have significant downstream consequences.

The field of species delimitation continues to undergo major theoretical and analytical advances (Camargo and Sites 2013, Conix 2018). A central goal of species delimitation is to develop inference methodologies that are rigorous and repeatable.

Reproducibility of research is a cornerstone of science (Packer et al. 2018). This is a non-trivial problem in species delimitation where the prolonged divergence of one lineage into two is a continuous process that can occur in myriad ways. The characteristic(s) that first reveals a diversification event might be identified via genomic or morphological analyses, by the observed response to a pheromone or a song, to an ecological shift, and/or to an inability to interbreed, etc. In order to detect diversification under such vast possibilities, species hypotheses are best generated and evaluated using multiple species criteria in an integrative taxonomy (Schlick-Steiner et al. 2010). Lineage divergence is often a time extended process (Darwin 1859, de Queiroz 1998, Avise 2000, Sukumaran and Knowles 2017; though see Leaché et al. 2018), and the extent of divergence can be assessed through the evaluation of many lines of evidence. Confidence in delimitation is highest when disparate criteria yield congruent results (Carstens et al. 2013). However, the thresholds of each criterion are arbitrary (Sites and Marshall 2004), and these thresholds are not achieved at the same time or in a particular order (de Queiroz 2007). Delimitation criteria must be explicit for reproducible species delimitation using an integrative approach (Dejaco et al. 2016). Given the enormity of the parameter space that encompasses all modes of speciation, any given method will fail in situations where the assumptions of that method are strongly violated (Carstens et al. 2013). Incongruence across delimitation can help identify complex evolutionary histories and reveal interesting evolutionary questions (Schlick-Steiner et al. 2014). Comprehending speciation necessitates investigation into the processes that initiate and maintain lineage diversifications. Under an integrative taxonomy, not only is our understanding of species

limits improved, but we can also better appreciate the evolutionary processes likely responsible for speciation.

The inland mesic forests of the Pacific Northwest (PNW) of North America hosts a unique and rich flora and fauna (Daubenmire 1975, Brunsfeld et al. 2001, Carstens et al. 2005). This mossy and fern-laden forest of the Northern Rocky Mountains is primarily located in northern Idaho and is home to many short-range endemic leaf litter-dwelling invertebrates such as terrestrial gastropods (Leonard et al. 2011), millipedes (Shelley et al. 2010), harvestmen (Briggs 1971), and beetles (Barr 2011), with many more endemics awaiting description and discovery. Some invertebrate species have diversified within this inland mesic forest island (e.g., Lucid et al. 2018), a phenomenon thus far best exemplified by the thorn harvestmen *Acuclavella* (Opiliones, Ischyropsalidoidea). Opiliones are an ancient and diverse group of arachnids (Pinto-da-Rocha et al. 2007, Kury 2013) often characterized by low vagility and habitat specialization. Harvestmen have been the focus of many applied examples of a modern integrative taxonomy (e.g. Arthofer et al. 2013, Richart and Hedin 2013, Derkarabetian and Hedin 2014, Wachter et al. 2015). Presently, *Acuclavella* contains seven nominal taxa all endemic to the PNW (Shear 1986, Richart and Hedin 2013). All thorn harvestmen are black, heavily sclerotized and robust-bodied, with rigid and acute spines arising from the scutum. They are crenophilic habitat specialists, occurring adjacent small, perennial water features such as headwater streams, springs, and seeps (Richart and Hedin 2013). Inland *Acuclavella* species are arranged in an allopatric series of short-range endemics (*sensu* Harvey et al. 2002) with distributions ranging roughly from 500 - 5,000 km^2. Four species described

by Shear (1986), are endemic to the mesic inland forest (Richart and Hedin 2013). These species are primarily diagnosed by their dorsal spines (Figure 2.1).

In the revision by Richart and Hedin (2013), a lack of phylogenetically informative characters did not allow explicit testing of the previous taxonomic hypotheses by Shear (1986). Additionally, Richart and Hedin discovered multiple new populations with novel morphologies that could not be sufficiently evaluated with molecular data and these populations were not ascribed to a species taxon. The taxonomic and species delimitation problems uncovered by Richart and Hedin (2013) form the basis of this research. Specifically, these problems (from south-to-north) include 1) inference of the molecular evolutionary relationship between *A. merickeli* and *A. quattuor*. Morphometric analyses strongly clustered samples by species, with morphospecies bound to disparate geographic areas. Phylogenetic analyses recovered these two morphospecies exclusive of other *Acuclavella* with strong support on a long branch, but the root node of this clade was very shallow and subtended paraphyletic morphospecies. Another finding 2) was a collection of populations morphometrically undistinguishable from *A. quattuor*, yet strongly inferred by few-gene molecular phylogenetics to be an independent lineage. This clade, dubbed there and here *A.* cf. *quattuor*, has a geographically cohesive distribution that is not adjacent to *A. quattuor*, with these two putative species bracketing the *A. merickeli* morphospecies. Perhaps most challenging 3) was a series of morphologies that included undescribed traits such as individuals with three pairs of spines on scute areas I-III and females with four pairs of spines on areas I-IV, with these morphologies apparently encompassing the diagnostic morphologies of *A. cosmetoides*

and *A. shoshone* in a continuous variation (Richart 2014). These individuals clustered into a discrete morphospace from other thorn harvestmen species, but without apparent morphological clustering within the group, and without sufficient molecular data to evaluate within-group relationships.

Speciation is particularly difficult to detect when a series of allopatric identities all contain high population structure, as is often seen in poorly dispersing short-range endemics (e.g., Hedin et al. 2015). Moreover, empirical research has found a positive association between population structure and speciation rates (Harvey et al. 2016). Species delimitation in light of these issues is non-trivial. The difficulty in distinguishing species from populations can be aided by incorporating next-generation sequencing technologies. Particularly useful for elucidating relatively recent evolutionary events in non-model organisms is restriction-site associated DNA sequencing (RADseq; Baird et al. 2008) which allows for the rapid generation of thousands of short-read loci scattered across the genome. Here, a two enzyme double-digest (ddRADseq) protocol (Peterson et al. 2012) is used to generate genome-wide data to aid in the delimitation of interior *Acuclavella* species. ddRADseq has allowed for the simultaneous sampling of population genomic and phylogenomic data, with the former dependent on the number of individuals sampled per lineage. The utility of phylogenomic data is dependent on potential allelic dropout for old divergences where a large amount of missing data can rapidly accumulate as a function of decreasing relatedness. This allows for the generation of sequence data suitable for genetic methods of species delimitation from both genealogical and

phylogenetic perspectives, as well analyses aimed at detecting the transition between micro and macroevolutionary processes (Shaffer and Thomson. 2007).

Here, the Evolutionary Species Concept (ESC) is the theoretical foundation used for species delimitation. The ESC includes species characterized as 'an entity composed of organisms that maintains its identity from other such entities through time and over space and that has its own independent evolutionary fate and historical tendencies" (Wiley 1978). Under the ESC, species are lineages maintained from other lineages by evolutionary mechanisms (Simpson 1951). The ESC has universal applicability and encompasses all secondary species concepts that contend species arise via evolutionary mechanisms (Wiley 1978). For this reason, the ESC is considered the primary species concept and is theoretically capable of recognizing all naturally occurring species (Frost and Kluge 1994, Mayden 1997, Wiley and Mayden 2000). The many secondary 'concepts' do not define species, rather they are operational criteria alternatively useful for delimiting species, contingent upon the evolutionary history and circumstances of a particular diversification event (de Queiroz 1998, Sites and Marshall 2004, Hey 2006, Schlick-Steiner et al. 2010). The ESC is operationally applied to species delimitation via the Generalized Lineage Construct (GLC) in an integrative taxonomy (Marshall et al. 2006, de Queiroz 2007, Padial et al. 2010).

The primary aim of this research is to more rigorously test species hypotheses within the inland thorn harvestmen in an integrative framework. This includes the evaluation of 1) the four species described by Shear (1986), 2) the species and unique populations and morphologies uncovered by Richart and Hedin (2013), as well as 3)

novel species-level hypotheses developed herein. This research adds to a growing body of work (e.g., Schlick-Steiner et al. 2010, Camargo et al. 2010) that contends that the best way to proceed with species delimitation is to attempt to identify the evolutionary mechanism that caused the diversification, and to rely more heavily on methods of inference that best capture the parameter space of the inferred mechanisms. *Acuclavella* are non-model taxa, and available lines of evidence are limited. Here morphological and genomic data are used in an integrative framework. Genomic-scale data in the form of ddRAD-sequences are collected from a larger sample of inland *Acuclavella* than has been previously analyzed. This data is used to validate existing species hypotheses, and is analyzed in many different ways. In the integrative protocol followed here, congruence within the various genomic analyses will precede evaluation of incongruence with morphological data (*sensu* Jacobs et al. 2018). This research points toward unanswered questions, the importance of which are crucial to the understanding of terrestrial biodiversity within the Pacific Northwest.

II. MATERIALS AND METHODS

(1) Taxon sampling

Details of specimen collecting and curation are outlined elsewhere (Richart and Hedin 2013). In this research, focal samples represent 233 individuals (115 males, 118 females) from 48 localities (Appendix B). All specimens are deposited in the San Diego State University Museum of Biodiversity Terrestrial Arthropod Collection (SDSU_TAC)

or genomics collection (SDSU_OP) unless otherwise noted. Molecular specimens were preserved in 100% EtOH and stored at -80°C; morphological specimens were preserved in 80% EtOH and stored at room temperature. A total of 82 individuals from 45 inland localities (northern Idaho and adjacent Montana) were sampled for ddRADseq analyses. Because of the potential for allelic dropout when using ddRADseq with distantly related individuals, as well as a bias towards loci with slower mutation rates and deeper coalescent times (Arnold et al. 2013), ddRADseq data was only collected from inland *Acuclavella* species. The localities represented with morphological data are also represented with ddRADseq data with three exceptions (Appendix B footnote). All individuals in ddRADseq analyses are included in morphological analyses with the exception of OP4063, OP2298, OP2340, OP2269, and OP2720 - all excluded due to some missing data (leg measurements). After removing individuals with missing data, morphometric analyses were conducted on a total of 230 individuals (115 males, 115 females) from 45 localities (Appendix B).

(2) Molecular data collection, ddRAD-sequencing, and data processing

Genomic DNA was extracted using a Qiagen DNeasy Blood and Tissue Kit (Qiagen Valencia, CA, USA) following manufacturers protocol, using lateral half bodies with intestines and eggs removed. Prior to enzymatic digestion, genomic extractions were quantified using a Qubit 2.0 fluorometer (Thermo Fisher Scientific), and samples with < 25 ng/μL (n = 2) were replaced. Libraries for ddRADseq were prepared with slight modifications to the standard protocol (Peterson et al. 2012), as specified elsewhere

(Burns et al. 2017). A pilot study was conducted to assess sequence output using three combination pairs of five REs, with *Msp1* (5'-CCGG-3') and *EcoR1* (5'-GAATTC-3') resulting in the largest number of shared loci and unlinked SNPs between samples (Burns et al. 2017). These REs were subsequently applied to a panel of 80 inland *Acuclavella*. Genomic DNA (500 ng at 40 µl volume) was digested in a 50 µl reaction including 100 units (1 µl) of each RE (New England Biolabs, Ipswich, MA), 5 µl of 10x CutSmart Buffer (New England Biolabs), and 3 µl of water at 37° C for 4 h. Samples were purified using 1.5x the reaction volume of Agencourt AMPure XP beads (Beckman Coulter, Inc., Brea, CA) and quantified using Qubit 2.0. The standard deviation of genomic concentration was less than 1 within each of the 8 columns, thus dilution was not necessary. Adapters were ligated to digested and purified genomics in a 40 µl reaction using 33 µl of genomics, 1 µl of Msp1 P2 adaptor, 1 µl P1 adaptor (8 different adaptors, 1 per row), 4 µl of 10x ligase reaction buffer, and 1 µl of T4 DNA-ligase (New England Biolabs). Adapter ligation reactions were incubated at room temperature for 40 min, heat killed at 65°C for 10 min, then cooled to room temperature by 2°C every 90 s for 22 cycles. Columns were pooled, purified using AMPure XP, and fragments 400-600 bp were isolated with a Pippin Prep (Sage Science). Standard Illumina primers were added to the pooled samples using PCR in 50 µl reactions consisted of 20 µl of post size-selected DNA, 3 µl water, 1 µl Primer P1, 1 µl Primer P2 (10 different primers, 1 per pool), and 25 µl Phusion Master Mix (New England Biolabs). PCR amplification cycle conditions were 98°C for 30 seconds, 98°C for 10 seconds and 72°C for 20 seconds (12 iterations), and 72°C for 10 minutes. PCR products were purified with an AMPure XP

bead cleanup, then molarity was quantified using an Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Resulting libraries were sequenced on the Illumina HiSeq 2500 platform under the 100-bp single-end protocol at the University of California, Riverside, IIGB Genomics Core facility.

Raw sequence reads were assembled using pyRAD 3.0.5 (Eaton and Ree 2013, Eaton 2014). This program assembles short-read sequence data by demultiplexing, quality filtering, clustering within samples, filtering by error rate and heterozygosity, filtering by base calling and paralog detection, clustering between samples, and aligning across samples. Similarity thresholds for within and between sample clustering (Wclust) were set to 0.95 which has been shown to recover the majority of unique alleles, yet not be so divergent as to result in many over-split loci which can reduce genetic distance when using more stringent between sample clustering thresholds (Harvey et al. 2015). All analyses also shared the common settings: 5 maximum number of sites with a quality scores < 20 (NQual), 6 minimum coverage for a cluster (MinDepth), and 3 maximum number of individuals with a shared heterozygous site (MaxSH). Minimum number of samples in a final locus (MinCov) varied depending on downstream analysis, specified below where the method of analysis is discussed and shown in Appendix C. Generally, MinCov was set to ~50% of the samples, rounding up. The lone exception was data curated from all *Acuclavella* samples, where MinCov was lowered to a minimum of 26.8% with a minimum 22 of 82 samples needed to include a locus. In order to minimize the potentially serious problem of nonrandom patterns of missing data (Pritchard and Wen 2003, O'Leary et al. 2018), for analyses considering a subset of samples (e.g.,

intraspecific analysis of population structure) data was curated anew from pyRAD only considering these subsets. Also, these sets of data were generated with a more conservative minimum coverage (Appendix C).

(3) Phylogenetic inference

Phylogenetic analyses are used to 1) evaluate existing and newly generated species hypotheses for monophyly, 2) generate input trees for tree-based methods of species delimitation, and 3) estimate the geographical distribution of clades. Phylogenetic analyses included inference under maximum likelihood (ML), Bayesian, and coalescent approaches. All of these analyses used matrices from the same pyRAD output (Appendix C), with minimum coverage (MinCov) set to 22 (of 82 samples). Both maximum likelihood (ML) and Bayesian analyses were inferred from an unpartitioned concatenated alignment of entire reads from 4525 loci and 430,177 nucleotides with 57.3% missing data. The ML analysis used RAxML 8.2.4 (Stamatakis 2014) on the CIPRES Science Gateway (Miller et al. 2010). Phylogenetic support was quantified by executing 1000 rapid bootstrap inferences and a thorough ML search was optimized under the GAMMA model and a GTR substitution matrix. The resulting phylogeny was midpoint rooted. The Bayesian phylogeny was estimated with BEAST 2.4.5 (Drummond and Rambaut 2007, Bouckaert et al. 2014). The best-fit model of evolution was selected by PartitionFinder 1.1.1 (Lanfear et al. 2012) using the BIC criterion with linked branch lengths and a greedy search algorithm. Preliminary analyses of two replicates of 50 million generations sampling every 1000 used a log normal relaxed clock, a uniform prior ucld Stdev

distribution, and an average rate of evolution prior that was uniformly distributed which was used to root the phylogeny. The resulting ucldMean distribution did not overlap with zero, thus rejecting a strict clock and the analyses were retained. Two additional runs of 25 million generations were run and combined with initial runs after visualizing .log files in Tracer 1.6.0 (Rambaut et al. 2014) and discarding the first 9-37% of the generations sampled prior to stationarity as burn-in. Runs were combined using LogCombiner 2.4.0 and TreeAnnotator 2.3.2 (Rambaut and Drummond 2014). Two phylogenies were estimated under the coalescent model using the fully parametric yet computationally economical Singular Value Decomposition Scores for Species Quartets (SVDQuartets; Chifman and Kubatko 2014, 2015) in PAUP* 4.0a (Swofford, 2003). These two phylogenies differ in their inclusion of a rogue sample (see Results). Both SVDQuartets analyses were conducted on an alignment of 4474 unlinked SNPs with ~56.8% missing data, with all individuals as lineages, and all possible unrooted quartets sampled with the most probable assembled using the QFM quartet assembly tree inference (Reaz et al. 2014).

(4) Generation and integrative evaluation of species hypotheses

Theoretical approach and analytical framework - Robust species hypotheses are attained by integrating multiple types of data and analyses (Edwards and Knowles 2014). Here, molecular and morphological lines of evidence are integrated to formulate species hypotheses. Consistent with an integrative taxonomy, an effort is made to advance reasonable evolutionary hypotheses when results between these lines of evidence are

incongruent. The model of diversification in *Acuclavella* appears to be one of allopatry, as is common in harvestmen and other poorly dispersing taxa, and a strong correlation between monophyly and geography is expected (e.g., Wiens and Penkrot 2002, Derkarabetian and Hedin 2014, Wachter et al. 2015). Hence, since molecular data is heavily used to formulate species hypotheses, geographic information is not. Geography is considered supporting evidence when congruent with morphologically delimited *a priori* and *a posteriori* hypotheses. Also, all *Acuclavella* are microhabitat specialists - associated with cool, perennial, forested mountain streams. There is no indication of habitat-associated ecological speciation within the genus. Therefore, ecological species criteria are not advanced for the discovery or evaluation of putative *Acuclavella* species.

An attempt was made to assign individuals to *a priori* species hypotheses by identifying specimens using taxonomic keys created by the species authors (Shear 1986, Richart and Hedin 2013). The majority of the specimens north of the Middle Fork Clearwater have traits not currently ascribed to any named species and thus are *incertae sedis* regarding species-level classification (Richart and Hedin 2013). Analyses on the molecular data include an array of methods along the spectrum of microevolution (population genetics) and macroevolution (phylogenetics). Evaluating speciation from the microevolutionary perspective of population genetics helps to assure that species-level diversity is contained within the output and thus considered by other delimitation analyses. Most of the methods used here are discovery methods, meaning that putative species are identified without *a priori* assignment of samples to species. These including morphometric (Ezard et al. 2010), population genetic (STRUCTURE; Pritchard et al.

2000), and tree-based analyses (GMYC; Pons et al. 2006, Fujisawa and Barraclough 2013). These hypothesis-free analyses are used to validate *a priori* species hypotheses and to generate *a posteriori* species hypotheses. Congruence across these lines of evidence are interpreted as validation of evolutionary divergence (*sensu* Page et al. 2005, Wachter et al. 2015). Statistical noise can be problematic in some species delimitation analyses when differentiation among lineages is highly variable (i.e., large variation between some samples may obscure subtle variation between others). Therefore, *a posteriori* results are often used to inform nested analyses on the same data type with the same delimitation criterion. This approach is theoretically justified, for evolution results in groups nested within groups (e.g., Seifert et al. 2013). Since it is impossible to reject hypotheses not tested (Schlick-Steiner et al. 2014), this approach also insures that over-split species hypotheses are evaluated for species delimitation. The evaluation of monophyly is dependent on generated hypotheses and is here used to validate *a priori* and *a posteriori* hypotheses. Details for analyses are expanded in their respective sections below.

No new species are described in this dissertation and no formal names are ascribed. Two of the binomials (*Acuclavella merickeli* and *A. sheari*) follow the morphological diagnostics outlined in previous taxonomic works (Shear 1986, Richart and Hedin 2013). The name *Acuclavella quattuor* is herein used to refer to individuals matching previous descriptions and found between the Salmon River and South Fork Clearwater River, whereas *A.* cf. *quattuor* refers to a disparate population of individuals north of the Selway River that match the morphological diagnostics of *A. quattuor*. Two

other binomials (*A. cosmetoides* and *A. shoshone*) refer to clades that contain the type localities for these names, but in both clades the existing morphological variation greatly exceeds current definitions. Other names of putatively unique lineages either use geographic features (Hoodoo Mountains, Lolo Creek, North of Lochsa River, South of Lochsa River) or refer to (and are perhaps members of) existing binomials (e.g., *A.* cf. *cosmetoides*, *A.* cf. *shoshone*). These names are used herein to facilitate communication. All such names can be viewed in Figure 2.2. They do not imply species-level entities and, excepting the evaluation of previous (*a priori*) taxonomic hypotheses, do not represent assumptions that would bias analytical results. Species-level hypotheses were examined within and across many of these names.

Previous taxonomic hypotheses **-** The validity of the five described species of inland *Acuclavella* (Shear 1986; Richart and Hedin 2013) are statistically evaluated using multiple lines of evidence. All specimens were identified using the diagnostic characters specified in the available taxonomic keys (Shear 1986, Richart and Hedin 2013). Specimens from all five type localities are represented in both molecular and morphological data sets with the exception of *A. sheari* in the genomic sample - *A. sheari* is represented by an individual only 1 km from the type locality, within the same basin, and on the same side of East Fork Fall Creek. Specimens keying out as *Acuclavella merickeli* and *A. sheari* form geographically cohesive groups encompassing their respective type localities. Specimens diagnosed as *A. quattuor* occur in two disparate regions; the one south of *A. merickeli* populations includes the type locality, and another north of *A. merickeli* populations. The diagnostic characters for *A. cosmetoides* and

A. shoshone are localized to their type localities. The surrounding populations to these types show extensive variation including intermediate as well as morphologies not yet specifically ascribed to a specific taxon. Thus, no *a priori* assumptions were made regarding the validity of these species, other than that *A. cosmetoides* is the type species for the genus and thus some lineage will bear this name.

STRUCTURE The model-based Bayesian clustering method STRUCTURE 2.3.4 (Pritchard et al. 2000) was used to develop species hypotheses by assigning individuals to populations using unlinked SNPs. STRUCTURE assumes that populations are in Hardy-Weinberg equilibrium (HWE), and individuals are probabilistically assigned to a population to achieve HWE (or jointly to two or more populations if the genotypic markers indicate admixture). The "all inland *Acuclavella*" analysis used the same liberal-coverage pyRAD data set that was used for phylogenetic analyses (Appendix C). For this analysis, the optimal number of populations (K) were evaluated over K = 1-7 using the ΔK method based on the rate of change in the log probability between successive K values which has been shown to accurately detect the uppermost hierarchical level of structure (Evanno et al. 2005). Subsequent hierarchical STRUCTURE analyses were conducted on subsets of the exemplars inferred from higher level analyses, as recommended by several authors (Evanno et al. 2005, Vähä et al. 2007, Bryson et al. 2016, Janes et al. 2017). In order to minimize the potential problem of populations being inferred from nonrandom patterns of missing data (Pritchard and Wen 2003), nested analyses were conducted on data generated anew from pyRAD only considering the exemplars being analyzed. For nested analyses, K was typically evaluated over fewer

possible populations, to as low as K = 1-4. Because ΔK uses the rate of change between log probabilities it is not able to evaluate K = 1. Here, K = 1 is assumed when the log-likelihood value for K = 1 is higher than the log-likelihood value for $K = \Delta K$. All analyses were evaluated with 6 replicates which were assessed for convergence prior to combination of runs in STRUCTUREHARVESTER 0.6.94 (Earl and vonHoldt 2012). For example, in one set of analyses four of the six runs had a Ln likelihood of around -12,000, whereas the other two runs were -150,000 and -4,831,000; only the four convergent runs were combined in Structure Harvester. CLUMPAK was used to visualize and compare clustering results and visualize bar plots for resultant values of K (Kopelman et al. 2015). Results are graphically displayed using DISTRUCT (Rosenberg 2004).

Discriminant analysis of principal components Population structure was also explored using the non-model-based discriminant analysis of principal components (DAPC) using the dudi.pca and find.clusters functions in the R package Adegenet 2.0.1 (Jombart 2008, Jombart and Ahmed 2011) using R 3.3.2 (R Core Team 2016). *K*-means clustering and the Bayesian information criterion (BIC) were used to determine the optimum number of clusters and the first 20 principal components were retained. As in STRUCTURE analyses, DAPC analyses were conducted on exemplar subsets when inferred clusters were congruent with phylogenetic clades using specific pyRAD datasets (Appendix C).

Tree-based species delimitation Tree-based species delimitation used two methods, the Generalized Mixed Yule Coalescent (GMYC; Pons et al. 2006, Fujisawa

and Barraclough 2013) and the Bayesian Poisson Tree Processes (bPTP; Zhang et al. 2013). Both of these methods attempt to identify the transition point between relationships best described by phylogenetic connections (macroevolutionary processes) and relationships best described by the reticulate genealogies found within a single lineage (microevolutionary processes). The GMYC model optimization finds the ML transition between diversification events using a pure birth process (Yule model, Nee et al. 1994) and intraspecific genealogical branching events based on the neutral coalescent model (Hudson 1990). Using an ultrametric tree as input, it maximizes the likelihood of alternative numbers of species lineages by detecting the transition point between diversification events and intraspecific branching attributed to the coalescent. Similarly, the bPTP uses the difference between within- and between-species branch lengths to delimit species. The bPTP model assumes that the number of substitutions between species is significantly higher than the number of substitutions within species (Zhang et al. 2013). The bPTP is perhaps easier to implement than the GMYC, for it does not require an ultrametric tree as input. The two methods are comparable in that they both identify branching rate transition points. They differ in that the GMYC uses time to find these transitions and the bPTP uses the number of substitutions (Zhang et al. 2013).

Only including SNP data in phylogenetic analyses can inflate branch lengths, thus input trees for GMYC and bPTP analyses used phylogenies built from full sequence data which tend to have a more accurate topology and appropriate branch lengths (Leaché et al. 2015). Both GMYC and bPTP analyses were conducted on trees inferred from the same pyRAD output (Appendix C). Tree-based delimitation analyses were run on the

bPTP (http://species.h-its.org/ptp/) and GMYC (species.h-its.org/gmyc/) web servers. The GMYC analyses use the BEAST species tree as input and were run using both the single- and multiple-threshold approach, the latter of which iteratively explores different models of the number of lineages by splitting and fusing existing species clusters (Monaghan et al. 2009). An advantage of the multiple-threshold version is the inference of a 95% confidence set of delimited species, within which the single-threshold result is typically included though with reduced confidence (Fujisawa and Barraclough 2013). The bPTP analysis was initially run on the RAxML tree, then after surprising results was also run on the same ultrametric BEAST tree used as the GMYC input. Analyses on both trees used 50,000 MCMC generations, 500 thinning, and 0.15 burn-in.

Morphometrics - Details of data collection and the continuous morphological characters analyzed are reported elsewhere (Richart and Hedin 2013). All individuals with missing data were removed from analyses. Male and female data sets were analyzed separately due to strong sexual dimorphism. A total of 230 individuals are analyzed, evenly split between males and females, with fairly well-balanced sampling between clades as well as within clades for both sexes (Table 2.1).

Morphometric analyses use the algorithm described by Ezard (et al. 2010) to discover *a priori* species hypotheses. This method obtains orthogonal components that are robust to non-normally distributed data and reduces them to only include components with significant explanatory power (Croux and Ruiz-Gazen 2005, Pearson and Ezard 2013). The data are compressed to components with higher explanatory power (Peres-Neta et al. 2005), and the Bayesian Information Criteria (BIC) is used to identify the

optimal model of the number of clusters while accounting for outliers (Fraley and Raftery 2002, Filzmoser et al. 2008, Pearson and Ezard 2013). Parameters followed default settings with the exception that most analyses were run using the Tousseeuw-Croux estimator (*qn*) robust measure of scale as well as with the default median absolute deviation (*mad*). Morphometric analyses used a custom script (Derkarabetian and Hedin 2014; https://github.com/ShahanDerkarabetian/Ezard_etal_2010_wrapper) executed on a terminal using R 3.0.2 (R Core Team 2013). The *K*-values were adjusted for each analysis to match the number of sample localities per analysis. The analysis would not run on the full 16 character data set of Richart and Hedin (2013), but initiated when the two least variable measurements (trochanter and patella lengths) were removed. Morphological analyses were also conducted on data sets only including specimens recovered from one of two major clades inferred from phylogenetic analyses. Analyses on complete and nested data sets were run with and without outliers, as identified by the analyses. Thus, morphometric analyses take both a discovery and validation approach to delimitation. The problem of analyses occasionally not initializing was recurrent on these nested analyses, but also resolved with the removal of relatively invariant measurements.

Evaluation of monophyly - The Monophyletic Species Criterion requires reciprocal monophyly (Mishler and Donoghue 1982), which must be evaluated in the context of pre-assigned individuals (Wiens and Penkrot 2002). Ideally, monophyletic groups are accompanied by strong support values and subtended by relatively long branches. Here, monophyly is evaluated for validation of the *a priori* taxonomic hypotheses of Shear (1986) and Richart and Hedin (2013), as well as for select *a*

posteriori hypotheses derived from morphometric and non-phylogenetic molecular discovery methods.

III. RESULTS

(1) Sequencing and ddRADseq data processing

Sequencing resulted in 108.21 million raw reads with an average of 1.32 million reads across the 82 samples (range: 0.08-3.69). Just over 100 million reads (92.9%) with an average of 1.23 million per sample (0.07-3.44) passed quality controls (Appendix D). Between sampling clustering statistics for each pyRAD curated data set can be seen in Appendix C. The between-sample clustering considering all 82 samples resulted in 4525 loci and 4474 unlinked SNPs, with only 13 loci shared across all samples (Appendix C). This data set contained an average of 4975 polymorphisms per sample (533-13116), with an average of 1954 loci (127-2881) and 1932 SNPs (123-2842) per sample (Appendix E). Individuals with low numbers of shared loci are the individuals with relatively few total raw reads. As expected, the exercise of curating specific data sets for specific downstream analyses resulted in larger data sets by minimizing allelic dropout; not anticipated was that the resulting matrices also contained a smaller proportion of missing data (Figure 2.3). One of the samples, OP2234, has about 94% more missing data than the average sample, and 86.5% more missing data than the closest sample for this metric. This sample was retained when preliminary phylogenetic (RAxML) and population genetic (STRUCTURE) analyses placed it with the other two samples from that locality.

(2) Phylogenetic inference and the geography of clades

The Bayesian BEAST analysis used the GTR + I + G substitution model as chosen by PartitionFinder. By and large phylogenetic analyses are congruent in their topology and support, especially the RAxML and BEAST analyses which only differ topologically at a single node within a single locality. Both of these trees have high support values, with RAxML bootstrap values (BSV) lowering at shallower nodes (Figure 2.4), whereas BEAST posterior probabilities (pp) diminish at the oldest nodes (Figure 2.5). Both the RAxML analysis using midpoint rooting and the BEAST analysis using the molecular clock to root inland *Acuclavella* were congruent in their root placement.

Initial SVDQuartets analysis did not recover a monophyletic *A.* cf. *quattuor* and support values were very low for most internal nodes. This result is driven by the *A.* cf. *quattuor* sample OP2234 which has very high levels of missing data, as described above. Thus, another SVDQuartets analyses was conducted with this individual removed from the otherwise identical alignment. The removal of this sample result in a monophyletic *A.* cf. *quattuor* and improved support values at all internal nodes suggests that this rogue sample was jumping throughout the phylogeny (Figure 2.6). Support values for the SVDQuartets phylogeny remain lower than those seen in the concatenated analyses, which is not unexpected, but with one exception (*A. shoshone* sister to *A.* cf. *shoshone*) recovered entities evaluated as independent species with strong support (>80). Researchers using short-read RADseq data for phylogenetic analyses are still understanding how missing data impacts phylogenetic inference (Huang and Knowles

2016a). The OP2234 sample was recovered as sister to an individual from the same locality in both RAxML (BSV=96) and BEAST (pp=1) with high support. These results suggest that within a sample SVDQuartets reaches a problematic level of missing data before the concatenated approaches of RAxML and BEAST. Previous studies have suggested that up to 80% missing data per locus does not negatively affect tree topology, in fact numerous researchers have found that adding more loci tends to increase resolution though these matrices have higher levels of missing data (e.g., Rubin et al. 2012, Wagner et al. 2013, Huang and Knowles 2016a). Here however, it is not missing data at a locus, but missing data within a terminal. This problem could have been avoided by trimming samples based on the amount of missing data, but perhaps having some samples with orders of magnitude more missing data than other samples will not be uncommon in the phylogenomic era, for example when some samples are decades old museum specimens (e.g., Hedin et al. 2018).

The inferred root from ML and Bayesian analyses separates a southern and northern clade centered on the Middle Fork Clearwater River (MFCR; Figure 2.2). Since the SVDQuartets phylogeny is unrooted it cannot address this initial split, but all other clades discussed below are recovered in all phylogenetic analyses. South of the MFCR is a clade comprised of *A. sheari* located south of the Salmon River. *Acuclavella sheari* is sister to a shallow clade confined to an area south of the Selway River and north of the Salmon River. Within this clade, *Acuclavella quattuor* and *A. merickeli* are paraphyletic. This clade is the shallowest clade in the ultrametric phylogeny given the number and geographic coverage of the samples. These three species are sister to a *A.* cf. *quattuor*,

which is found north of the Selway River and encompasses the Lochsa Basin. Within *A.* cf. *quattuor* a deep node separates populations on either side of the Lochsa River. With the exception of the *A. sheari* localities, a close relationship between these areas has been previously detected in the Coeur d'Alene Salamander (Carstens et al. 2004). In these salamanders, samples south of the Selway River formed a clade, as did samples north of the Lochsa, and these clades were recovered as sister. All of these relationships were recovered with low support, though this is probably owing to the use of a single locus in the salamander study. Just north of the MFCR is a clade that extends north to the South Fork Clearwater River (SFCR) and contains the *A. cosmetoides* type locality. This clade has deep structure, with the southern three populations forming a clade (*A.* cf. *cosmetoides*), with this clade sister to clade containing a long branch leading to population just south of Lolo Creek that is sister to remaining populations including the type locality. Another clade is found southwest of the St. Maries River and northeast of the Palouse Prairie in the Hoodoo Mountains. All samples north of the NFCR and east of the St. Marys River ranging to westernmost Montana form a clade that is sister to a single locality just south of the NFCR (*A.* cf. *shoshone*). Lastly, it should be mentioned that most collection localities were recovered as monophyletic.

(3) Generation of species hypotheses

Previous taxonomic hypotheses **-** Individuals identified using taxonomic keys included geographic cohesive groups for *A. merickeli* and *A. sheari*. Specimens from two disparate geographic areas were identified as *A. quattuor*. Specimens matching the

diagnostics for *A. cosmetoides* and *A. shoshone* only occurred at the type localities for these species.

STRUCTURE The majority of STRUCTURE analyses resulted in $\Delta K > 3$ (Table 2.2), and strong population structure is required for the Evanno method to detect such structure (Waples and Gaggiotti 2006). Occasionally STRUCTURE and ΔK would recover "ghost populations", where a seemingly superfluous population would be spread across all or most of the samples in the analysis with low likelihood values (Figure 2.7). Only analyses resulting in some samples having a higher likelihood of belonging to one population and other samples having a higher likelihood of belonging to a different population were subjected to nested analyses. Thus, these "ghost populations" were not further considered. Three populations were inferred at the uppermost level of population structure, two of which form geographically cohesive units (Table 2.2, Figure 2.7). The first of these represented all *A. merickeli* and *A. quattuor* individuals. The second grouped all samples in the *A. shoshone, A.* cf. *shoshone,* and Hoodoo Mountain clades. The third inferred population placed *A. sheari* (south of the Salmon River) with *A.* cf. *quattuor* and *A. cosmetoides* (north of the Selway and south of the NFCR). Subsequent hierarchical analyses on subsets of the data uncovered further populations with high probability, validating the recommendations (e.g., Evanno et al. 2005) to explore data in this manner. Tier II analyses recovered 1) *A. sheari, A.* cf. *quattuor*, and the greater *A. cosmetoides* clade as separate populations; 2) the Hoodoo Mountains clade from the greater *A. shoshone* clade; but 3) did not recover *a priori* identified *A. merickeli* from *A. quattuor*. By tier III lower levels of population structure were only significant with the

A. cf. *quattuor*, *A. cosmetoides*, and Hoodoo Mountains clades. At this level, 1) *A.* cf. *quattuor* populations north and south of the Lochsa River were recovered as distinct, 2) within *A. cosmetoides* two populations were inferred, one of which was a paraphyletic grouping, and 3) the four localities in the Hoodoo Mountain clade were inferred to be three different populations. At tier IV, population structure was detected within the *A.* cf. *quattuor* clade south of the Lochsa and within the *A. cosmetoides* clade, with the five of the eight total populations inferred at this tier coming from single localities.

Populations with inferred introgression appear to vary across hierarchical tiers of analyses. There were fewer inferences of introgression in Tier II than in many of the further nested analyses, with almost no introgression inferred across inferred populations and corresponding clades. Tier III shows much more introgression across inferred populations, though no introgression is recovered in the three populations inferred from the Hoodoo Mountains. A small amount of introgression is inferred between populations of *A.* cf. *quattuor* north and south of the Lochsa River. Also at Tier III, the *A.* cf. *shoshone* population is inferred to be widely introgressed with the remaining *A. shoshone* exemplars. All three analyses run at Tier IV showed higher levels of introgression between all inferred populations.

Uneven sampling of populations has been shown to be one of a few violations of model assumptions that result in a downward-biased estimation of K (Puechmaille 2016, Lawson et al. 2018). Samples of *Acuclavella* in this study are unevenly sampled, and this likely effected the results. For example, *A. sheari* was not differentiated in nested

STRUCTURE analyses until the second hierarchical iteration, despite being one of the earliest diverging lineages and occupying the longest phylogenetic branch. Likewise, the *A.* cf. *shoshone* population found south of the NFCR is subtended by a relatively long branch and is sister to all remaining *A. shoshone* populations. STRUCTURE analyses did recover $\Delta K = 2$ within *A. shoshone*, with the samples south of the NFCR having a large proportion of their genome inferred as belonging to a different population than the remaining *A. shoshone* samples, however with the highest likelihood values for the same population assignment as all *A. shoshone* samples (Figure 2.7, tier III). Puechmaille (2016) was able to alleviate this problem by subsampling their data to produce even sampling. This strategy is not employed here, for only one population south of the NFCR was sampled.

DAPC Results for the discriminant analysis of principal components can be seen in Appendix F. All DAPC analyses recovered at least some paraphyletic groups. For analysis of all inland *Acuclavella*, *K* was evaluated for *K*=3, *K*=6, and *K*=15 (Appendix F.1, F.2, and F.3 respectively). Of these, *K*=6 was the only analysis resulting in BIC scores vs. the number of *K* with the characteristic elbow. This analysis assigned all individuals to clusters unequivocally, and with one exception the recovered groups were monophyletic in the phylogenies reported above. The other two assessments had BIC scores vs. the number of *K* continuing to decrease as *K* reached the number of samples in the analyses. They each also recovered multiple groups that are paraphyletic. Five nested analyses, each only considering samples within one of the five major clades (Appendix F.4-F.8; Figure 2.4), contrasted greatly in the amount of subgroup splitting. Three of

these analyses, the Hoodoo Mountain clade, *A. cosmetoides*, and *A.* cf. *quattuor* were split into multiple groups, often split to the locality, and in the case of *A. cosmetoides* splitting localities and clustering samples that are paraphyletic. Two groups were inferred within the *A. merickeli* + *A. quattuor* clade, one of the groups only consisting of *A. quattuor* samples, and the other cluster containing all *A. merickeli* samples as well as four *A. quattuor* individuals. Nested DAPC analysis recovered *K*=3 within *A. shoshone*, one of which corresponds to the *A.* cf. *shoshone* clade.

Tree-based species delimitation Results from tree-based species delimitation varied widely. The multiple threshold GMYC analysis (Figure 2.8) inferred 14 ML entities (8-22). The results of this analysis however were not intuitive - *A. sheari* was not recovered as distinct despite being recovered in a shallow relationship that is subtended by a long branch and has a relatively ancient most recent common ancestor to its sister clade. In contrast, the greater *A. cosmetoides* clade was split into 6 entities, including an occurrence of splitting individuals from the same locality. The GMYC has been shown to over-split lineages when population structure is extensive within species, especially when sampling is not extensive enough to detect potential isolation by distance (Lohse 2009, Keith and Hedin 2012, Fujisawa and Barraclough 2013, Satler et al. 2013, Fernández and Giribet 2014, Hamilton et al. 2014, Wachter et al. 2015). A considerable amount of population structure is inferred within inland *Acuclavella* species, as seen in the STRUCTURE results reported above. The single threshold GMYC analysis delimited 15 ML entities with a confidence interval of 8-22 (Figure 2.8). In examination of the single threshold analyses, a likelihood value for an alternative and more conservative

delimitation is nearly equal to the most likely partition (Figure 2.8, panels A and B). Because of the known issue of over-splitting and the very similar likelihood value, this conservative single-threshold (cST-GMYC) hypotheses is also used to evaluate species hypotheses. This partition infers 10 entities (Figure 2.8). One of the criticisms of the GMYC is its dependence on an accurate and ultrametric input tree (Zhang et al. 2013); the ultrametric BEAST tree used here is highly supported.

Results from the bPTP analyses were surprisingly split (Figure 2.5). The most supported partition found via heuristic search using the RAxML phylogeny identified 59 entities (range: 24 - 64, mean: 47.31). This analysis delimited 42 singletons with 17 occurrences of individuals from a single locality delimited as a different species. Using the ultrametric BEAST tree as input reduced the amount of splitting very little, delimiting 40 total species (range: 25-65, mean: 47.23). This analysis also delimited numerous singletons and often delimited individuals from the same locality as separate species. Hyper-split results has been recovered by some users of the GMYC (e.g., Satler et al. 2013), but unlike here, previous studies comparing PTP to GMYC tended to find that the GMYC splits samples into more putative species than does the PTP (e.g., Schwarzfeld and Sperling 2015). The PTP initiates the search for the maximum likelihood delimitation at the root of the input phylogeny, and delimitation estimates are more accurate when the phylogeny is correctly rooted (Zhang et al. 2013). Though RAxML (midpoint) and BEAST (molecular clock) trees agreed in their rooting, outgroup rooting was not able to be employed here. It has been recommended to use caution when using the PTP and the numbers of samples per species is unbalanced, with the branching patterns under this

sampling scheme not optimally parameterized (Zhang et al. 2013). Perhaps that is occurring with the phylogenies used here. Perhaps these results are caused by the nature of the data. The bPTP model assumes that the number of substitutions between species is significantly higher than the number of substitutions within species (Zhang et al. 2013), but perhaps in a phylogeny reconstructed from a genomic array of SNPs the robustly supported intraspecific branches are longer than expected by the model. Further studies of the effect of phylogenetic tree inference on PTP are needed.

Though designed for single-locus analysis, the GMYC can be used to delimit species from any ultrametric tree, including those reconstructed from concatenated SNPs such as used here (Fujisawa and Barraclough 2013). Use of the GMYC on concatenated SNPs has been previously reported (Singhal et al. 2018), but to my knowledge never justified. Comparatively, this approach is likely preferable to GMYC analysis of independent gene trees reconstructed from alignments of short sequence reads from nuclear loci (e.g., Toussaint et al. 2015). Relatively common evolutionary phenomena have been shown to mislead GMYC analyses, many of which are more likely to be confounding when using a single locus mitochondrial marker. Mitochondrial introgression across species boundaries would fail to detect different lineages that might be recovered using multiple nuclear marker. Also, extensive nuclear admixture in the face of deep mitochondrial (mtDNA) divergences (e.g., Miralles and Vences 2013), and/or when dispersal is heavily male biased (e.g. Satler et al. 2013), will result in a mitochondrial gene tree reflecting more structure than actually occurs between species. Furthermore, a gene tree is more likely than a phylogeny to, by chance, have a high

number of intraspecific deep coalescent events that approximate the Yule process, and the GMYC would infer too recent of an inflection point toward coalescent processes (Esselstyn et al. 2012). Similarly though with an opposite result, rapid speciation of lineages with large effective population sizes would scatter coalescent processes deeper into a gene tree than a phylogeny, and the GMYC would infer too few lineages (Esselstyn et al. 2012). Gene trees can differ from species trees for a number of reasons (Maddison 1997), and many of these (e.g., undetected paralogy, reconstruction artifacts, strong selection, and horizontal transfer) would befuddle tree-based delimitation approaches (Reid and Carstens 2012, Fujisawa and Barraclough 2013). Inclusion of multiple nuclear markers alleviate these concerns. Lastly, because of possible nuclearization of the mitochondrial COI (Richart and Hedin 2013), there is currently no suitable locus to build such a gene tree for *Acuclavella*. Given the high utility of SNP data for species delimitation, a simulation study comparing gene trees and concatenated SNP trees to known species trees should be an avenue of future research.

Morphometrics - In general, analyses of male morphometric data inferred more clusters than for female data. Also, the *mad* measure of scale inferred more clusters than did the *qn* measure (Table 2.3, Appendix G); analyses resulting in the most clusters are summarized in Figure 2.5. In each of the all-male analyses there are one or more groups of morphometric clusters that do not necessarily correspond to an *a priori* defined entity, or to a clade that might be considered different species, but are conterminous with the root node. This pattern is very nearly recovered in the analysis of female data as well. Since the root node corresponds with the MFCR, all samples north of this boundary are

being inferred to occupy a different morphospace than all samples south. In both the *mad* and *qn* male analyses *A. sheari* samples are recovered in an exclusive group. With this exception, the southern samples include 1) *A. merickeli*, *A. quattuor*, and *A.* cf. *quattuor*, and the northern samples include 2) *A. cosmetoides*, *A. shoshone*, and the Hoodoo Mountain clade. These two groups of putative species are further evaluated without including the other group in the analyses, which may introduce noise.

The all males analysis using *mad* infers nine clusters, the most for all morphometric analyses (Figure 2.9). The only *a priori* identified species recovered by this analysis was *A. sheari*. Four morphometric clusters are inferred within and exclusive to the *A. merickeli*, *A. quattuor*, and *A.* cf. *quattuor* samples, and four cluster are inferred within and exclusive to the Hoodoo Mountains, *A. cosmetoides*, and *A. shoshone* samples. Within the first of these groups, *A. merickeli* samples are exclusively inferred as belonging to two morphometric clusters, and both clusters are monophyletic. The *A. quattuor* and *A.* cf. *quattuor* samples are also exclusively in two clusters, though samples belonging to these disparate clades are scattered across both clusters. Within the second of these groups, the four individuals from the *A. cosmetoides* type locality were inferred as a unique morphometric cluster. Each of the other three inferred morphometric clusters contained individual from the Hoodoo Mountain clade, the *A. shoshone* clade, and the remaining *A. cosmetoides* samples. The all male *qn* analysis recovers 5 clusters, none of which are *a priori* defined species or clades, but it did recover *A. merickeli*, *A. quattuor*, and *A.* cf. *quattuor* samples in exclusive clusters than the others (Appendix G.2). The all female *mad* analysis also recovered the *A. merickeli*, *A. quattuor*, and

A. cf. *quattuor* samples exclusive of the others with a few exceptions in the *A. shoshone* clade, and a cluster is nearly congruent with *a priori* identified *A. merickeli* (Appendix G.3). Only two clusters were inferred in the all female *qn* analysis. These two clusters largely corresponded to samples north and south of the MFCR, again with a few *A. shoshone* inferred to belong to the southern samples. Unlike the *mad* analysis, in this analysis *A. sheari* samples cluster with southern samples (Appendix G.4).

The analyses on smaller data sets were largely similar to analyses of the entire dataset, for example analyses on male data recovered the *A. cosmetoides* type locality as a distinct cluster in a very divergent morphospace (Appendix G.5). Improvements from analyses on the full data set include, *A. merickeli* recovered by male data as a single cluster in a divergent morphospace (Appendix G.8), and female data recovers a distinct *A. merickeli* with a single exception (Appendix G.9). With these exceptions, *A. shoshone*, *A. cosmetoides*, and Hoodoo Mountain samples were either scattered across multiple clusters (*mad* male Appendix G.5, *mad* female Appendix G.7) or were inferred to belong to a single cluster (*qn* male Appendix G.6). Likewise, *A. quattuor* and *A.* cf. *quattuor* were either inferred as a single cluster (*mad* male Appendix G.8) or were inferred to occupy two clusters, with specimens from both clades scattered across them (Appendix G.9).

The Ezard (2010) method was specifically chosen as a discovery approach. Numerous clusters were inferred within inland *Acuclavella*, but very few of these inferred clusters correlated to clades or groups of *a priori* identified specimens. *Acuclavella sheari* was recovered as a distinct morphological cluster, was identified *a priori*, and was

recovered as a clade. This is also true of *A. cosmetoides,* where the diagnostic morphotypes have only been found at the type locality. Though the three *A. cosmetoides* samples from the type locality are recovered as a clade, this is a very shallow and well-nested group. Though never recovered in analyses as single groups, the clades and *a priori* identified species found north and south of the MFCR were recovered in exclusive clusters, perhaps pointing toward some morphometric signal deeper in the phylogeny than the species level. That morphometric analyses recovered so few clades that probably represent different species is not surprising, for the degree of morphological difference within a species can be greater than between related species (Simpson 1951). This is likely true within inland *Acuclavella*. For example, in the analyses of all females the morphological space occupied within *A. cosmetoides*, *A. shoshone*, and the populations from the Hoodoo Mountains is greater than or equivalent to the combined morphological space occupied by *A. merickeli*, *A. quattuor, A.* cf. *quattuor*, and *A. sheari* (Appendix G.3).

(4) Integrative evaluation of species hypotheses and taxonomic suggestions

Integrative evaluation **-** Herein, species are delimited based on hypotheses generated using multiple approaches and lines of evidence. The amount of conflict between data types and between analytical methods is substantial (Figure 2.5). Some authors contend that, in an integrative taxonomy, only congruent delimitations across all analyses should be considered ample evidence to support a species hypothesis (e.g., Carstens et al. 2013). Further, only congruent delimitations within a data type should be

considered, thereby mitigating incongruence before comparisons across data types (Jacobs et al. 2018). Both of these approaches are unnecessarily conservative, as further discussed below. A significant amount of morphological variation within inland *Acuclavella* species appears to not always or strongly track genetic lineage divergence. The requirement of congruence between morphology and molecular data would impede our understanding of *Acuclavella* diversity. Thus, more weight is given to congruent analyses within a data type, without requiring congruence across data types.

Morphometrics - Few hypotheses put forward by morphometric analyses seem to reliably track speciation in *Acuclavella*. Often, inferred clusters are scattered across the phylogeny and across the landscape, and often specimens from a single locality are inferred to belong to multiple clusters. This is likely not the result of inadequate analytical methods, for the Ezard (2010) morphometric approach has been successful in previous integrative taxonomy (e.g., Derkarabetian and Hedin 2014). Rather, this is probably due to other factors such as homoplasy (convergence or plesiomorphy) in the case of *A. quattuor* and *A.* cf. *quattuor*, or due to a somatic morphology that has evolved and/or is currently evolving faster than lineages diversified (the Hoodoo Mountain, *A. cosmetoides*, and *A. shoshone* clades). Morphometric analyses put forward a total of five hypotheses that will be further considered: 1) *A. sheari*, 2) *A. merickeli*, 3) *A. quattuor* + *A.* cf. *quattuor*, 4) *A. cosmetoides* + *A. shoshone* + Hoodoo Mountain populations, 5) *A. cosmetoides* type locality.

Tree-based delimitation - Tree-based methods varied tremendously in the number of entities delimited (Figure 2.5). Every hypothesis inferred by GMYC methods was also

recovered or further split (usually the case) by PTP analyses. Thus, reconciliation only needs to happen within the three GMYC analyses. Using a simple two-thirds majority rule, tree-based delimitation analyses generated the following species hypotheses: 1) *A. sheari*; 2) *A. merickeli* + *A. quattuor*; 3) *A.* cf. *quattuor* north, and 4) south of the Lochsa; 5) the Hoodoo Mountain clade; 6) *A.* cf. *shoshone*, with the remaining *A. shoshone* localities are split into 7) two southeastern populations in eastern Clearwater County, and 8) the northern populations that include the type locality; lastly, the deep nodes found in the greater *A. cosmetoides* clade suggest four more hypotheses that include 9) southern populations near the MFCR, 10) a population just south of Lolo Creek, 11) a central population near Orogrande Creek, and 12) the remaining localities which include the type locality.

STRUCTURE and DAPC - STRUCTURE and DAPC are both used to determine the number of population clusters (*K*) within the data. Thus, these two approaches will be reconciled for the generation of species hypotheses. Both analyses occasionally clustered paraphyletic entities, this was especially problematic at shallow nodes in DAPC analyses (Figure 2.5). When low hierarchical structure is recovered as paraphyletic, generated species hypotheses default to an upper-level monophyletic cluster. These clustering methods generate the following species hypotheses: 1) *A. sheari*; 2) *A. merickeli* + *A. quattuor*; 3) *A.* cf. *quattuor* north, and 4) south of the Lochsa; 5) the Hoodoo Mountain clade; 6) *A.* cf. *shoshone*, 7) the remaining *A. shoshone* localities; and within the greater *A. cosmetoides* clade 8) southern populations near the MFCR + Lolo Creek, 9) Orogrande Creek, and 10) the remaining localities including the type locality.

Reconciliation - Reconciling the conflict seen in delimitation analyses within inland *Acuclavella* species is not trivial. The only *a priori* species hypothesis regularly recovered across evidence and analyses is *A. sheari*. The only other generality between morphological and molecular analyses is the grouping of populations north of the MFCR exclusive of populations between the MFCR and the Salmon River. When considering the analyses on molecular data, tree-based and clustering methods were fairly congruent, differing only in the number of entities inferred within the greater *A. cosmetoides* and *A. shoshone* clades. Taking the conservative approach by only considering agreements between these different types of analyses on the molecular data suggests that there are 11 inland *Acuclavella* species. These eleven species combine *A. quattuor* and *A. merickeli* into a single lineage. Separation of these two lineages brings the species tally to a dozen.

Specimens *a priori* identified as *Acuclavella quattuor* and *A. merickeli* are not recovered as reciprocally monophyletic in the input tree, and thus not possibly delimited by the GMYC which explicitly assumes species are reciprocally monophyletic (Fujisawa and Barraclough 2013). Though monophyly is not a required criterion for species delimitation, strong evidence should back arguments for paraphyletic species. The evidence herein that argue for a two-species hypothesis are that these two taxa were 1) identified *a priori*, 2) confirmed to be morphologically distinct by multivariate analyses, and 3) these disparate morphologies were confined to allopatric geographies. The evolutionary mechanism that underlies this phenomenon is likely genic speciation (Wu 2001, Via 2009). Genic speciation is a difficult hypothesis to falsify, as further discussed below.

Taxonomic conclusions - How should these 12 putative species be taxonomically treated? Currently, five species taxa exist in northern Idaho. Of these, only *A. sheari* and *A. merickeli* have adequate morphological diagnostics. Here, recommendations are made to describe two new species, which would bring the number of specific *Acuclavella* taxa in the Greater Clearwater Refugium to seven. Additionally, the definitions of two existing species taxa need to be expanded. For both the two new species and the two expanded definitions, specimens should undergo further scrutiny in search for potential somatic characters that would aid in the diagnoses and identifications. Integrative taxonomy when conducted well removes the need to widely revisit the system for revisions. However, other entities delimited by the integrative approach employed here should await further sampling and analyses before taxonomic action is taken. More specific recommendations are made in the immediately following paragraphs.

The diagnostic morphology for *A. cosmetoides* is currently only known from the type locality and this finding is supported by multivariate morphometric analysis. However, this locality is genetically closely related to adjacent populations with very different morphotypes. Retaining the current definition would result in these adjacent populations either remaining *incertae sedis*, or would necessitate the description of numerous new species without the support of substantiating evidence. The integrative taxonomic delimitation taken here suggests three to five species exist within the greater *A. cosmetoides* clade, many of them currently known from single localities. Deep phylogenetic structure and extensive morphological variation suggest that the *A. cosmetoides* clade has been in place for many thousands of years. Further, the

delimitation of these lineages is associated with a high likelihood peak in GMYC analyses. Conversely, tier III and IV STRUCTURE analysis revealed extensive admixture between these inferred populations. Thus, future research should thoroughly search this area to better determine the contact zones between subclades and morphotypes, and then determine if additional taxonomic acts are necessary. For now, the concept of *A. cosmetoides* should be expanded to include the monophyletic group of populations found north of the MFCR and south of the NFCR.

Two species are delimited within the greater *A. shoshone* clade. A single sampled population just south of the NFCR, *A.* cf. *shoshone*, is on a long branch that is sister to remaining *A. shoshone* samples. This population is only 6 kilometers away from a sampled population on the northern bank of the NFCR, yet this north-shore sample is more closely related to individuals 90 kilometers further north. *Acuclavella* cf. *shoshone* likely represents a new species suitable for taxonomic recognition. However, this action is not yet recommended, pending the sampling of other populations to get a better idea of the distribution and morphological variation of this species. Meanwhile, the definition of *A. shoshone* should be expanded to include the extensive morphological variation found within this clade, and for the time being, include *A.* cf. *shoshone* within this definition.

The *A. merickeli* + *A. quattuor* clade is the antithesis of a cryptic species, with straightforward morphological discrimination resulting in two geographically cohesive units, without any detected molecular support for diverging lineages (e.g., Buyck et al. 2016). Among the many definitions under the Phylogenetic Species Criteria is the following (Frost and Kluge 1994), "Phylogenetic Species represent the set of naturally

occurring supraorganismal units *discoverable* by one operation, organismal character analysis, using the evidentiary rule of fixed characters as a ranking rule that extends from an operational simplifying assumption." Under this criterion, *A. quattuor* and *A. merickeli* are valid species. Morphological divergence is clear evidence when predicting that populations evolve independently (Simpson 1951). A number of evolutionary processes could underlie this incongruence, such as very strong selection coupled with the recent divergence of lineages with large effective population sizes, possibly with ongoing hybridization and introgression. To highlight the extreme conundrum posed in this instance, the *A. merickeli + A. quattuor* clade is the shallowest clade in the ultrametric phylogeny given the number of localities and given the geographic coverage of the samples.

A possible solution is to subsume *A. quattuor* as a subspecies of *A. merickeli*. Subspecies are a subjective phenomenon compared to the objective reality of species. Nonetheless, having an official subspecies taxon does help convey important information, such as the close relationship of two populations and of allopatry (Mayr 1963). Though subspecies are not commonly used in harvestmen taxonomy they are not unprecedented (e.g., Briggs 1971), and this taxonomic earmark would help communicate important information to down-stream users. Further, this designation is neither arbitrary or inconsistent (Wilson and Brown 1953; Huang and Knowles 2016b), for morphometric analyses clearly discriminate these two geographically cohesive forms, and this combination of fixed morphological differences without complimentary molecular phylogenetic distinctiveness does not occur elsewhere in the genus. This situation

matches a classic definition of subspecies, as being a collection of populations that are diagnosable and occupy different geographic areas (Mayr and Ashlock 1991). Morphology suggests two lineages, genetic approaches suggest one. The most likely explanation is that a locus conferring adaptive significance is not passed across the species boundary, without such strict barriers for other loci. If *A. quattuor* and *A. merickeli* are diverging, the most likely explanation is the genic speciation model (Wu 2001). This model is not explicitly tested here, but suggestions are made in the Future *Acuclavella* Research section that concludes this chapter. It is not a trivial endeavor to falsify a two-species hypothesis once it has been formulated (Miralles and Vences 2013, Hedin 2015). To conclude, a two-species hypothesis should be retained until the genic speciation model can be adequately tested. This situation is an empirical reminder on the importance of morphology in an integrative taxonomy (Schlick-Steiner et al. 2010).

The populations sampled from the Hoodoo Mountains were consistently recovered across molecular analyses and should be described as a new species. There are deep phylogenetic nodes within this species. Further, the log-likelihood of there being three populations in the analyzed samples is 60% higher than the log-likelihood of their being a single population (Table 2.2), and these populations are inferred to have very little if any introgression (Figure 2.7). These mountains, and other forested areas on the western edge of the interior mesic forest should be further surveyed and sampled for *Acuclavella* populations.

Acuclavella quattuor and *A.* cf. *quattuor* are morphologically similar yet molecularly and geographically divergent. Molecular analyses strongly infer these

species are independent lineages. In fact, these analyses reported suggest that two lineages exist in the greater *A.* cf. *quattuor* clade, one on either side of the Lochsa River. Sampling for each of these lineages includes multiple localities and individuals. However, any taxonomic treatment separating these two metapopulations should await further sampling. A network of streams that comprise the headwaters of the Lochsa River is proximal to Lolo Pass, U.S. Highway 12, numerous Forest Service roads and trails, thus offering an ideal scenario to elucidate the relationship between the headwaters of major rivers and the isolation of lineages. For now, these two sister lineages should be united under a new species description. Integrative taxonomy has a rich history in harvestmen species delimitation; when conflict between data sets arises it typically involves a lack of morphological evidence to corroborate a two-species hypothesis inferred by molecular analyses (e.g., Boyer et al. 2007, Arthofer et al. 2013, Derkarabetian and Hedin 2014, Wachter et al. 2015). Thus, the conflict in analyses across data sets between *A. quattuor* and *A.* cf. *quattuor*, and within *A.* cf. *quattuor*, is only surprising given the abnormal morphological diversity found within the genus.

IV. DISCUSSION

How to fail less often at species delimitation - The Evolutionary Species Concept (ESC) was introduced by Simpson (1951) as an alternative to the scientific discussion of alternative species concepts. In this seminal work, referring to the species concepts in use at the time, he states, "...it is possible.. to combine some of their apparently but not really

conflicting views into one consistent statement...". The ESC contends that species derive from evolutionary processes, thus the ESC encompasses all secondary species concepts. These secondary 'concepts' are independent criteria whose usefulness in species delimitation is contingent on the evolutionary processes relevant to that particular group of organisms. Criticism of the ESC include that it provides little or no guidance as to which traits are more important for delimiting species (Templeton 1989). This independence of not relying on a particular criterion is a strength of the ESC, not a weakness. Further confusion surrounding the ESC is that the word 'concept' has been used to refer to the secondary criteria, giving the impression that they are comparable when they are subsidiary. Hereafter, the word 'concept' is reserved for the ESC, and the word 'criteria' is used to refer to delimitation criteria such as the Biological Species Criteria (BSC). Widespread adoption of this language may prevent this misunderstanding from continuing. Mayden, a strong proponent of the ESC, considered it to be a theoretical concept that was not operational (Mayden 1997). However, theoretical advancements over recent decades show that the ESC is operationally applied to species delimitation through the Generalized Lineage Construct (GLC; de Queiroz 2007, *contra* Frost and Kluge 1994). Under this theoretical framework and operational strategy, even a single criterion is sufficient to propose a taxonomic species hypothesis when the evolutionary mechanism underlying the differentiating criterion is signal from the initiation of divergence and likely to result in the maintenance of that divergence (de Queiroz 2007). This allows a taxonomist to give more weight to a particular trait when that trait or suite of traits has been shown to drive diversification within that system (Padial et al. 2010).

Frost and Kluge (1994) claimed that a Phylogenetic Species Criterion (PSC) was the 'operational manifestation' of the ESC, specifically the "smallest naturally occurring diagnosable samples of organisms." This definition would appear to be congruent with the view that a single appropriately interpreted criterion is sufficient to delimit species. Yet, Frost and Kluge argue against paraphyletic species, "with its distortion of history, character generality, and hypotheses of homology, is so detrimental to understanding and communication in evolutionary biology that any payoff would have a very high price." Perhaps incongruence across data sets was not yet sufficiently appreciated, but a reliance on the criterion of monophyly precludes the PSC from being the operational manifestation of the ESC.

Decisions that populations will evolve separately involves prediction (Simpson 1951), and the rationale for these decisions must be clear and supported by the data (Lambertz 2017). Two desired outcomes of species delimitation is the objectivity and reliability of lineage classifications (Conix 2018). Many researchers advocate for objective species delimitation (e.g. Fujita et al. 2012, Derkarabetian and Hedin 2014). Objectivity is repeatability. Given a clear definition of what constitutes a species, given the delimitation criteria, and given the data, independent researchers arrive at the same conclusions (Hey and Pinho 2012). Reliability is accuracy. When there is conflict between species criteria, it must be true that one is not tracing lineage divergence. A reliable classification is one where delimited entities most closely track the true number of lineages in the group. Delimiting species and its independent replication is straightforward when there is congruence across data sets (de Queiroz 2007). More often

than not however, the integrative taxonomist is faced with incongruence across data rather than agreement (e.g., Schlick-Steiner et al. 2010, Willis 2017). Many practitioners of integrative taxonomy are conservative with their conclusions, cautiously favoring to err towards a single-species hypothesis (e.g., Bond 2012). Avoiding being misled by false positives (type I error) is more important than obscuring our understanding with false negatives (type II error), but avoiding both should be maximized (Wiley and Mayden 2000). Thus, there is a trade-off. How less repeatable are species delimitation efforts that are attempting to be more reliable? I posit that an increase in reliability can be achieved without a commensurate decrease in objectivity via a sophisticated integrative taxonomy.

All methods to delimit species will fail in certain scenarios, and in an integrative taxonomy various methods applied to the problem can return a spectrum of concordance across data sets (de Queiroz 2007). There are two different camps among practitioners of an integrative taxonomy (Conix 2018). In a congruence-required integrative taxonomy, disparate disciplines must all support a delimitation (e.g. Carstens et al. 2013). In a sophisticated integrative taxonomy, a single criterion, if appropriately interpreted, is sufficient for delimitation (e.g., Wachter et al. 2015). Under a congruent integrative taxonomy, false positives are conservatively avoided without regard for the minimization of false negatives. For example, Carstens (et al. 2013) suggests that researchers apply a wide range of species delimitation analyses and conservatively delimit species when analyses are congruent. A problem with this requirement of congruence is that as more and more delimitation criteria are used to evaluate species hypotheses the reliability (accuracy) of the delimitations deteriorate (Figure 2.10). This is analogous to areas of

statistical inconsistency, where the addition of more data results in stronger support for an incorrect hypothesis. Here though, it is the conceptual approach rather than the analytical method that obscures support.

Conversely, in a sophisticated integrative taxonomy accuracy improves with the addition of lines of evidence (Figure 2.10). With this approach, species are reliably and objectively delimited even with incongruence when plausible evolutionary explanations underpinning the incongruence are supported by data and theory (Schlick-Steiner et al. 2010, Conix 2018). Under a sophisticated integrative taxonomy, false negatives are decreased by not relying on a single criterion, threshold, or congruence to delimit species, and an effort is made to evade false positives by invoking evolutionary explanations for conflicting signals. A sophisticated integrative taxonomy is a rigorous taxonomic approach, for practitioners are obligated to search for evolutionary explanations for disagreement across disciplines, and if none are readily apparent, additional disciplines are added to elaborate on the extent and potential causes of discordance (Schlick-Steiner et al. 2010).

In the quest for objectivity, many practitioners of species delimitation require the achievement of some preset measure of significance. Using statistical significance tests does not make scientific research objective (Thompson 1999). Since adjusting alpha-values results in different delimitations, selecting these cut-off values is inherently arbitrary (Hey 2009). In the genomic era, an alpha-value of fashion has been comparing Bayes Factors (Kass and Raftery 1995) across alternative hypotheses using analyses based on the coalescent (Grummer et al. 2014, Leaché et al. 2014). Basically, there must

be significantly more coalescent events associated with a particular hypothesis than with an alternative, or the probability of the data given the model must by higher than the probability of alternative models. Thus, coalescent-based species delimitation falls under the family of criterion-based species delimitation with an affinity to Genealogical Species Criteria (GSC), which itself has close affinity to PSC (Harrison 1998).

In practice this research often takes an initial 'discovery approach', where the ESC is used to generate species hypotheses, then these hypotheses are validated only using GSC (e.g., Everson et al. 2018, Kuchta et al. 2018). No doubt, the principles of the GSC have philosophically and operationally advanced the field of species delimitation (e.g., Avise and Ball 1990). The multispecies coalescent model is well-suited for comparison of the neutral genome to estimate population and species divergences (Leaché et al. 2018). However, there is no objective reason that the GSC should be the operational default method of validating species hypotheses. For example, the BSC has a current and long species delimitation history in ornithology (see McKitrick and Zink 1988). Under the BSC philosophy, some ornithologists will operationally validate the presence of a prezygotic reproductive barrier between allopatric species by playing the songs of the alternate populations to see how response may differ across them (e.g., Freeman and Montgomery 2017). Similarly, Esselstyn (et al. 2012) used bat echolocation calls to validate putative species identified by the GMYC. Objective validation should not by perceived as the equivalent of meeting GSC criteria. If the validation method is not appropriate for the identification of the mechanism(s) that initiate and maintain species boundaries, then the research is operating under the threshold of a secondary species

criterion. Phenetic species criteria applied to morphological traits can be more informative for species delimitation than genetic data when traits are under selection, when migration is high, or when gene trees are highly discordant and/or poorly resolved (e.g., Solís-Lemus et al. 2015). To use coalescent-based molecular methods to "objectively" validate morphology-derived species hypotheses under these scenarios is illogical. Favoring operational validation via GSC without theoretical (or empirical) explanation as to why GSC is more appropriate than other delimitation criteria is itself subjective. Indeed, this is true for any *a priori* favoring of any operational criterion, especially when broadly applied to the tree of life (e.g., Fujita et al. 2012, Herrera and Shank 2016). Another shortcoming of coalescent-based species delimitation is that it treats data equally among all living things. However, organisms differ markedly in their propensity to violate the assumptions of the coalescent model. Many poorly dispersing short-range endemics violate the assumption of panmixia within a species (e.g., Keith and Hedin 2012, Hedin et al. 2015). Species delimitations using genomic data and GSC to delimit such species have been shown to over-split, preferring a two-species model given population structure (Jackson et al. 2017, Sukumaran and Knowles 2017), and this preference becomes asymptotically stronger as the number of loci increases (Leaché et al. 2018).

This researcher bias is referred to as the problem of operationalization-dependency. The inter-subjective objectivity of operational species delimitation relying on congruence across data sets does not prioritize operational criteria relevant to the evolutionary mechanisms fundamental to the diversification of the organisms being

evaluated (Lipscomb et al. 2003, Conix 2018). As a hypothetical, a macroevolutionary analysis suggests that sexual selection is playing a role in the diversification of a clade of jumping spiders. Analysis within this clade finds two groups that have congruent diagnoses between morphometric analysis and prezygotic barriers via female choice during courtship, yet they are paraphyletic to one another. The taxonomists should apply the evolutionary theory applicable to their system to interpret incongruent operational criteria. In the case of the jumping spiders, the conflicting criteria should illuminate the recent divergence between the two species. Rather than casting doubt on a two species hypothesis, this conflict stimulates evolutionary research. As another example consider cryptic species, where conflict in data is inherent. Molecular data and delimitation analyses have vastly aided in the discovery of cryptic diversity (Bickford et al. 2007; e.g., Hamilton et al. 2011, Satler et al. 2013). We shouldn't always expect morphological changes in species that have persisted in nearly identical microhabitats. So why would we put equal weight on morphological or ecological methods of species delimitation when discovery can only be made using molecular methods? Many groups of organisms have a suite of characteristics that includes poor dispersal ability, and morphological and ecological/niche conservatism (e.g., Starrett and Hedin 2007, Leavitt et al. 2015). In these situations, phylogenetic and genealogical criteria provide perspective that would otherwise go unperceived.

Reliable classifications have important consequences for macro- and microevolutionary research. Diversification rates would become more comparable across different taxonomic groups. Population genetic studies would more accurately identify

intra-specific processes. The model examples for reliability in species delimitation are precisely the ones where we understand the evolutionary mechanisms causing diversification. It is possible that a species can arise in only a few generations, but this would be an extraordinary claim for a taxonomist to make, and as the Truzzi aphorism goes, extraordinary claims require extraordinary proof. Hence, here's a real world example: a Neotropical tanager - one of Darwin's Finches - is the result of hybrid-speciation, has a distinct morphology, a pre-zygotic mating barrier in the form of a novel song, and has behaved as an independent lineage for six generations and 31 years (Lamichhaney et al. 2017). This species is not going to pass delimitation tests of reciprocal monophyly or models under the multi-species coalescent. Keen observation, insight, and evolutionary theory allow for this delimitation. Another example comes from *Drosophila* where, with the ease of husbandry and fast generation times, hybrid breakdown can be quantified between alternate lineages (Coyne and Orr 1989). These are systems where the evolutionary mechanisms causing speciation have been the target of extensive research. Taxonomic reliability is a function of familiarity - when researchers have strong natural history skills and an intimate understanding of the system (e.g., Andrés et al. 2013, Lamichhaney et al. 2017).

To a certain extent, all methods and delimitations require researchers to make somewhat qualitative judgments as to species boundaries (Sites and Marshall 2004). Some researchers have maintained that criterion consistency is important when delimiting species within a given group of closely related species (e.g., Delrieu-Trottin et al. 2017). Consistency should not, however, trump the evaluation of multiple processes, for the

mechanisms causing diversification within a group are not necessarily consistent. For example, within the box-eucalypts, lineages formed 1) through isolation and drift, 2) as the result of strong natural selection, and 3) from hybrid speciation (Flores-Rentería et al. 2017). Instead, consideration of the mechanism(s) of speciation should be consistently applied within a given group and across the entire tree of life (Camargo et al. 2010). This would allow for a more meaningful comparison of diversity or diversification rates between clades.

The function of theory is to guide scientific investigation (McMurray 1955). The ESC and a sophisticated integrative taxonomy based on a broad range of species criteria, yet lack of reliance on any one criterion, better guides scientific investigation by more readily identifying complex evolutionary mechanisms and in generating scientific hypotheses (Lipscomb et al. 2003, Schlick-Steiner et al. 2014). Practitioners who adhere to a particular delimitation criterion will be affected by that criterion itself in the discovery of the origin and maintenance of disparate lineages (Wiley 1978). The strength of the ESC is that all current knowledge on speciation can be used to asses diversification. Determining the cause of lineage origins allows for the more accurate delimitation of lineages (Padial et al. 2010). For example, evaluating change in gene flow during speciation, or studying the evolutionary mechanisms underlying ecological speciation, necessitates scrutiny of lineages that have not yet reached complete isolation (Via 2009, Noutsos et al. 2014). The close study of incipient species will allow for a better comprehension of the parameter space potentially relevant to species delimitation and provide new insights (Nolte and Tautz 2010). Further, to understand speciation it is

important to consider populations that have not fully achieved speciation, for after speciation evolutionary mechanisms that had nothing to do with the initiation and maintenance of the divergence will further differentiate the lineages (Harrison 1998, Via 2009). If an evolutionary biologist wants to better understand how speciation is initiated and maintained, congruence-based species delimitation obscures by increasing the likelihood that lineage divergence happened well in the past. The discordance of data sets makes species delimitation more difficult, but presents an opportunity for interesting evolutionary questions. By spending less time defining what species are *a priori*, and more time understanding the etiology of speciation, the principle of reciprocal illumination will better identify the boundaries of speciation within a given system. I suggest that species delimitation is more reliable when researchers use a sophisticated integrative taxonomy to make assumptions based on evolutionary theory that reasonably apply to their system. This is how we can fail less often at species delimitation.

The biodiversity crisis and the operational appeal of alpha values - The need for a reliable alpha taxonomy is perhaps most important to the field of conservation biology, both by bringing clarity to management programs (e.g., Hedin et al. 2015, Rix et al. 2018) and in documenting biodiversity before it is lost to extinction (Wheeler 2004, Schlick-Steiner et al. 2010). A sophisticated integrative taxonomy is best suited to the first of these needs, and should be conducted on many species of known conservation concern (e.g., Hedin 2015). The second is perhaps better addressed using threshold-based delimitation, with the desire to as quickly and accurately describe as many species as possible against growing threats to biodiversity (Wiens 2007). Threshold-based species

delimitation with alpha-values set to make false positives unlikely can greatly facilitate and accelerate the documentation of the world's biodiversity. Data can be collected, tossed into the black box of *au courant* analyses, and voila, look at the species, on to the next. Biodiversity is an immense place, and there are not enough taxonomists to place empirical delimitations within a strong theoretical background. It doesn't take an organismal expert to avoid false positives when all or most of the data analyzed are congruent, any statistician can do it. In the face of an ongoing major extinction (May 2002; Ceballos et al. 2015, 2017), taxonomy-oriented statisticians must continue to conduct rapid assessments and earmark putative species identified through the singularity of DNA-based taxonomy, whether that be from using the Phylogenetic, Phenetic, or Genealogical Species Criteria. There remains the problem of statisticians inferring multiple species, then not formally describing those species, perhaps due to a lack in taxonomic training, perhaps due to misaligned priority by the scientific community as to the academic importance of taxonomy (Carstens et al. 2013, Satler et al. 2013). Taxonomy should continue to be the focus of increase funding, such as the National Science Foundation's Partnerships for Enhancing Expertise in Taxonomy (PEET) program, which is no longer receiving proposals. An increase in trained taxonomists would facilitate the description of entities delimited by statisticians through collaboration. Taxonomists should welcome all of the help we can get, root each other on, and accomplishing as much as possible.

The practice of a singular molecular approach to species delimitation is revealing putative species at a faster pace than the current taxonomic workforce can apply names

and morphological diagnostics (Padial et al. 2010). It is imperative that rapid DNA-based species delimitation continue. The Multi-species Coalescent model (MSC) has been shown to outperform other rapid species delimitation methods (e.g., DNA barcoding) that rely on satisfying a threshold (Yang and Rannala 2017). With the ever-increasing ease of collecting genomic data, the MSC has become a standard in DNA taxonomy (e.g., Wagner et al. 2014). Just as an exclusively morphological approach would fail to detect cryptic species, an exclusively DNA-based approach will not document all species nor the important biodiversity found within species. Therefore, taking statistical delimitation one step further, congruence-required integrative taxonomy too may have an important role to play in a rapid assessment of biodiversity. Like DNA taxonomy, it allows for the discovery and classification of lineages without knowledge of the natural history of the organisms. Molecular and morphological data can be collected from tissues and specimens, and analyses produce conclusions.

The taxonomic information science - Taxonomy is the hypothesis driven science of describing and organizing biodiversity, and taxonomists must make theses hypotheses identifiable and accessible (Wheeler 2004). A consequence of an integrative taxonomy is an ever-increasing importance on the work that taxonomists are tasked. Taxonomists help improve the bioliteracy of non-scientists by assigning names and descriptions to defined groups. Perhaps more than ever, taxonomists connect directly to the public via citizen science projects and online repositories for species observations. Taxonomists need to continue describing and organizing biodiversity, making these products web-accessible, while simultaneously collaborating with phylogeneticists, ecologists, genomic

researchers, and bioinformaticians (Bik 2017). Taxonomists are the hub of an information retrieval system for all of biology.

Most taxonomists that practice species delimitation also concern themselves with the categorization and organization of species into arbitrary groups such as genera and families. Thus, most taxonomists consider themselves systematists (Dayrat 2005). The naming of higher taxa is important to our understanding and communication of biodiversity, and is imperative to the navigation of the vastness of biodiversity. The species taxon lays at the interface of macro- and microevolutionary processes, so taxonomists that practice species delimitation should equally concern themselves with the identification of populations and the intrinsic and/or extrinsic processes that have led to speciation. Too few taxonomists consider themselves population biologists, phylogeographers, or molecular ecologists, yet biodiversity extends into these areas. Taxonomic recognition below the species level has implications for conservation biology and in preserving evolutionary potential (Patten and Remsen 2017). Biodiversity within the species can be delimited and classified more logically than can genera and families, and formally naming and providing identifying features of this biodiversity would better accomplish major tasks of the taxonomic discipline. For example, the integrative approach used here identified a phenomenon where allopatric species have fixed morphological differences yet belong to the shallowest clade in the tree given the number of samples. This clade is likely an ideal system for evolutionary biologists interested in the causal mechanisms that could explain this pattern (e.g., genic speciation, divergent selection). How should a taxonomist best organize information to increase the likelihood

that it is discovered by a researcher? Without a formal taxonomy for classifying putative and within-species diversity, how can the information known within this species best be advertised to the world? Names are perhaps the most important tool for effective communication, for a name provides the link to knowledge (Schlick-Steiner et al. 2010).

An expansion of taxonomy could increase the number of downstream users and perhaps better highlight the importance of taxonomy. A formal taxonomy should expand to include evolutionary aspects and delimited biodiversity that is potentially below the species level to become an information science that curates our accumulated knowledge of biodiversity (Godfray 2002). Formal species-level taxonomy that makes a distinction between different types of delimited species would better convey this information to downstream users. For example, species robustly delineated across multiple data types might be better suited for conservation biologists when resources to protect species are limited (Mace 2004, Hedin 2015). A nomenclature for a well-supported clade without apparent morphological or ecological divergences from closely related clades, or populations with discrete and fixed morphologies without molecular differentiation in the majority of the genome, would better attract the attention of evolutionary biologists interested in specific speciation processes. A taxonomy allowing for the application of names to entities satisfying various combinations of species criteria would more effectively categorize this biodiversity for addressing particular evolutionary questions (de Queiroz 2007). This would help accomplish a major goal of taxonomy, but also openly acknowledge some level of doubt about the independence of evolutionary lineages. This in turn would ensure that taxonomic products are not unintentionally

misinterpreted by conservation biologists (Dussex et al. 2018). The practice of delimiting different types of species may also help to stabilize classifications, which under current practice can change frequently depending on the operational criteria favored by a particular taxonomist (Conix 2018). Of course, a taxonomic rank already exists below the species, but application of the subspecies trinomial in some universal manner would not distinguish between, for example, 'morphological' or 'deep population structure' sub-specific entities. With this expanded taxonomy, practitioners also perhaps wouldn't be compelled to take an ultraconservative approach to species delimitation (e.g., Carstens et al. 2013).

Future research - A w**e**ll-executed integrative taxonomy reduces the need to revisit the taxonomic hypotheses of its recipients. This ideal was not achieved here. An integrative taxonomic evaluation of inland *Acuclavella* suggests that future work is needed beyond the taxonomic recommendations advanced above. The diversification of inland *Acuclavella* lineages is characterized by a number of closely related lineages that vary both in their stage along the speciation continuum and apparently in speciation mechanism. As such, the thorn harvestmen of the inland Western Hemlock Zone have the potential to become a model in the study of speciation. Outlined here are avenues of future research that will advance our understanding of speciation.

Phenotypes are the object of natural selection. The morphological diversity of dorsal spination within and between *Acuclavella* species in northern Idaho suggests that perhaps adaptation is playing a role their divergences. Natural selection is thought to play a prominent role in initiating and maintaining speciation (Shapiro et al. 2016). Usually

phenotypic plasticity is the underlying mechanism when two entities are conspicuously different morphologically but very closely related genetically (e.g., Mallien et al. 2018). Invoking plasticity to explain the difference between the paraphyletic *A. merickeli* + *A. quattuor* species requires making unlikely assumptions. The fixed differences between them is likely due to fixed molecular differences and not differential epigenetic expression of a locus on either side of the SFCR. What is causing this apparent explosion in morphologies? This answer may be elucidated via in-lab feeding trials using the stream-associated amphibians endemic to the inland mesic forest such as the Coeur d'Alene Salamander, Mountain Tailed-Frog, and Idaho Giant Salamander.

The detection of genic speciation is non-trivial. Both theoretical arguments and empirical evidence suggest that prolonged divergent selection with recurrent migration and gene flow will result in the genomic clustering of sites under selection, with fewer, larger, and more tightly linked divergent alleles compared to divergent selection without gene flow (Yeaman and Whitlock 2011). Thus, if that portion of the genome is not sampled then the speciation genes will go undetected. Some have had success in the search for candidate speciation genes using proteomics (e.g., Andrés et al. 2008, Andres et al. 2013). This research sequenced the transcriptome of the male accessory gland of different cricket species. All known species of *Acuclavella* males have a gland on top of their first cheliceral segment (Shear 1986). The function of this gland is unknown, but being sexually dimorphic suggests that it plays a role in mating. Comparative transcriptomic analyses of these glands may be fruitful in the search for the genes underlying speciation (e.g., Andrés et al. 2013). However, there is no reason to suspect

that sexual selection is playing a role in *Acuclavella* diversification. Divergent selection is characterized by early extrinsic reproductive isolation, where fitness in hybrids is reduced when the genotypes underlying phenotypes are selected against by environment interactions (Seehausen et al. 2014). An annotated *Acuclavella* reference genome would greatly facilitate the association of divergent regions of the genome to mechanisms of reproductive isolation, as well as represent an important tool for all workers on Opiliones. High quality reference genomes improve the detection of selection by 1) matching candidate loci to linked coding regions, 2) assisting investigation into gene function, and 3) helping to ensure that highly variable regions of the genome that include functional genes are investigated (Manel et al. 2016).

A peninsula of mountains extending west of the crest of the Bitterroots, framed by the Lochsa to the south and the NFCR to the north, is a pinwheel of intrigue. Along the southern edge high in the Lochsa basin, is the boundary between the north and south of the Lochsa *A.* cf. *quattuor* clades. Do they stay separated by that river? Down-canyon, somewhere *A.* cf. *quattuor* and *A. cosmetoides* must come into close proximity. Where is this boundary? There are no major drainages that separate them, but perhaps a steep north-facing slope. The western bulb of this peninsula is occupied by *A. cosmetoides.* Deep genetic structure, especially in the southern populations suggests that this area was the heart of the Greater Clearwater Refugium; this area was associated with genetic diversity in the willow *Salix melanopsis* as well (Brunsfeld et al. 2007). On the northern edge, the boundary between *A. cosmetoides* and the *A.* cf. *shoshone* population needs to be established; a good place to start would be on either sides of Orogrande Creek, Weitas

Creek, and Fourth of July Creek. Certainly the geographic extent and molecular characterization of *A.* cf. *shoshone* needs to be made. Finally, what is the boundary of the *A.* cf. *shoshone* clade and the *A.* cf. *quattuor* clade? This is probably the best example of a ridge, separating the Lochsa from the NFCR, rather than a river being the boundary of two *Acuclavella* species.

Inland *Acuclavella* species monophyly is well supported, though determining the root within this clade has consistently been problematic (Richart and Hedin 2013). Determining the root would allow better assessment of the timing and spatial development of divergences within the clade, and a better understanding of the evolution of spine morphology variation between and within the species. Outgroup rooting using the western Washington species would be ideal. This would be facilitated by using a method that results in less missing data and longer reads, for example a transcriptomic (e.g. Richart et al. 2016) or ultra-conserved elements approach (Starrett et al. 2017).

V. CONCLUSIONS

This research uses multiple data types and discovery approaches to objectively and reliably delimit species of inland thorn harvestmen. Specifically, this research identifies two undescribed species within the inland *Acuclavella* radiation and identifies two previously described species that need to have their diagnostic definitions expanded. Additional populations that possibly represent independent lineages are recovered, but these are conservatively not recognized as species pending further sampling. This

research is objective in that it is repeatable, given the data, the theoretical approach of the Evolutionary Species Concept, and the operational application via the General Lineage Construct. The conclusions are an attempt at reliability, meaning that delimited entities more likely match the real number of lineages in this system, rather than if using a conservative congruence-required approach. This is made apparent when considering that *A. sheari* would be the only species delimited if congruence across data and analyses was required. Just as using multiple disciplines for species delimitation avoids the failure rate inherent in using a single discipline, a sophisticated integrative taxonomy avoids the failure rate inherent to a congruence-required integrative taxonomy (Schlick-Steiner et al. 2010). The *a priori* favoring of any single species criterion applied broadly to species delimitation is an impediment to rigorous evolutionary biology. The adoption of a rigorous sophisticated integrative taxonomy for all life will lead to a more precise understanding of biodiversity and speciation mechanisms.

VI. REFERENCES

Agapow, P. M., O. R. Bininda-Emonds, K. A. Crandall, J. L. Gittleman, G. M. Mace, J. C. Marshall, and A. Purvis (2004). The impact of species concept on biodiversity studies. The quarterly review of biology 79:161–179.

Andrés, J. A., E. L. Larson, S. M. Bogdanowicz, and R. G. Harrison (2013). Patterns of Transcriptome Divergence in the Male Accessory Gland of Two Closely Related Species of Field Crickets. Genetics 193:501–513.

Andrés, J. A., L. S. Maroja, and R. G. Harrison (2008). Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proceedings of the Royal Society B: Biological Sciences 275:1975–1983.

Arnold, B., R. B. Corbett-Detig, D. Hartl, and K. Bomblies (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Molecular Ecology 22:3179–3190.

Arthofer, W., H. Rauch, B. Thaler-Knoflach, K. Moder, C. Muster, B. C. Schlick-Steiner, and F. M. Steiner (2013). How diverse is *Mitopus morio*? Integrative taxonomy detects cryptic species in a small-scale sample of a widespread harvestman. Molecular Ecology 22:3850–3863.

Avise, J. C. (2000). Cladists in wonderland. Evolution 54:1828–1832.

Avise, J. C., and R. Ball (1990). Principles of genealogical concordance in species concepts and biological taxonomy. Oxford surveys in evolutionary biology 7:45–67.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE 3:e3376.

Barr, C. B. (2011). *Bryelmis* Barr (Coleoptera: Elmidae: Elminae), a New Genus of Riffle Beetle with Three New Species from the Pacific Northwest, U.S.A. The Coleopterists Bulletin 65:197–212.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das (2007). Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution 22:148–155.

Bik, H. M. (2017). Let's rise up to unite taxonomy and technology. PLOS Biology 15:e2002231.

Bortolus, A. (2008). Error Cascades in the Biological Sciences: The Unwanted Consequences of Using Bad Taxonomy in Ecology. AMBIO: A Journal of the Human Environment 37:114–118.

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 Comput Biol 10:e1003537.

Boyer, S. L., J. M. Baker, and G. Giribet (2007). Deep genetic divergences in Aoraki denticulata (Arachnida, Opiliones, Cyphophthalmi): a widespread "mite harvestman" defies DNA taxonomy. Molecular Ecology 16:4999–5016.

Briggs, T. S. (1971). The harvestmen of family Triaenonychidae in North America (Opiliones). California Academy of Sciences, San Francisco.

Brunsfeld, S., J. Sullivan, D. Soltis, and P. Soltis (2001). Comparative phylogeography of northwestern North America: a synthesis. In Integrating Ecology and Evolution in a Spatial Context (J. Silvertown and J. Antonovics, Editors). Blackwell Publishing, Williston, Vermont, pp. 319–339.

Brunsfeld, S. J., T. R. Miller, and B. C. Carstens (2007). Insights into the Biogeography of the Pacific Northwest of North America: Evidence from the Phylogeography of Salix Melanopsis. Systematic Botany 32:129–139.

Bryson, R. W., W. E. Savary, A. J. Zellmer, R. B. Bury, and J. E. McCormack (2016). Genomic data reveal ancient microendemism in forest scorpions across the California Floristic Province. Molecular Ecology 25:3731–3751.

Burns, M., J. Starrett, S. Derkarabetian, C. H. Richart, A. Cabrero, and M. Hedin (2017). Comparative performance of double-digest RAD sequencing across divergent arachnid lineages. Molecular Ecology Resources 17:418–430.

Buyck, B., I. Olariaga, J. Justice, D. Lewis, W. Roody, and V. Hofstetter (2016). The Dilemma of Species Recognition in the Field When Sequence Data are not in Phase with Phenotypic Variability. Cryptogamie, Mycologie 37:367–389.

Camargo, A, and J. J. Sites (2013) Species delimitation: a decade after the Renaissance. In: The species problem: ongoing issues (ed Pavlinov IY), pp 225-247. InTech – Open Access publisher.

Camargo Arley, Sinervo Barry, and Sites Jack W. (2010). Lizards as model organisms for linking phylogeographic and speciation studies. Molecular Ecology 19:3250–3270.

Carstens, B. C., S. J. Brunsfeld, J. R. Demboski, J. M. Good, and J. Sullivan (2005). Investigating the Evolutionary History of the Pacific Northwest Mesic Forest

Ecosystem: Hypothesis Testing Within a Comparative Phylogeographic Framework. Evolution 59:1639–1652.

Carstens, B. C., T. A. Pelletier, N. M. Reid, and J. D. Satler (2013). How to fail at species delimitation. Molecular Ecology 22:4369–4383.

Carstens, B. C., A. L. Stevenson, J. D. Degenhardt, and J. Sullivan (2004). Testing Nested Phylogenetic and Phylogeographic Hypotheses in the *Plethodon vandykei* Species Group. Systematic Biology 53:781–792.

Ceballos, G., P. R. Ehrlich, A. D. Barnosky, A. García, R. M. Pringle, and T. M. Palmer (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances 1:e1400253.

Ceballos, G., P. R. Ehrlich, and R. Dirzo (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. Proceedings of the National Academy of Sciences 114:E6089–E6096.

Chifman, J., and L. Kubatko (2014). Quartet Inference from SNP Data Under the Coalescent Model. Bioinformatics:btu530.

Conix, S. (2018). Integrative taxonomy and the operationalization of evolutionary independence. European Journal for Philosophy of Science:1–17.

Coyne, J. A., and H. A. Orr (1989). Patterns of Speciation in *Drosophila*. Evolution 43:362–381.

Croux, C., and A. Ruiz-Gazen (2005). High breakdown estimators for principal components: the projection-pursuit approach revisited. Journal of Multivariate Analysis 95:206–226.

Darwin, C. (1859). On the origin of species by means of natural selection. First Edition.

Daubenmire, R. (1975). Floristic Plant Geography of Eastern Washington and Northern Idaho. Journal of Biogeography 2:1–18.

Dayrat, B. (2005). Towards integrative taxonomy. Biological Journal of the Linnean Society 85:407–415.

de Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation. In Endless Forms: Species and Speciation (D. J. Howard and S. H. Berlocher, Editors). Oxford University Press, Oxford, England, pp. 57-75 (Chapter 5).

De Queiroz, K. (2007). Species Concepts and Species Delimitation. Systematic Biology 56:879–886.

Dejaco, T., M. Gassner, W. Arthofer, B. C. Schlick-Steiner, and F. M. Steiner (2016). Taxonomist's nightmare … evolutionist's delight: an integrative approach resolves species limits in jumping bristletails despite widespread hybridization and parthenogenesis. Systematic Biology 65:947–974.

Delrieu-Trottin, E., K.-N. Shen, C.-W. Chang, and P. Borsa (2017). One species hypothesis to rule them all: consistency is essential to delimitate species. Journal of Heredity 1:3.

Derkarabetian, S., and M. Hedin (2014). Integrative Taxonomy and Species Delimitation in Harvestmen: A Revision of the Western North American Genus *Sclerobunus* (Opiliones: Laniatores: Travunioidea). PLoS ONE 9:e104982.

Drummond, A. J., and A. Rambaut (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Dussex, N., H. R. Taylor, M. Irestedt, and B. C. Robertson (2018). When genetic and phenotypic data do not agree: the conservation implications of ignoring inconvenient taxonomic evidence. New Zealand Journal of Ecology 42:7.

Earl, D. A., and B. M. vonHoldt (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359–361.

Eaton, D. A. R., and R. H. Ree (2013). Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Systematic Biology 62:689–706.

Edwards, D. L., and L. L. Knowles (2014). Species detection and individual assignment in species delimitation: can integrative data increase efficacy? Proc. R. Soc. B 281:20132765.

Esselstyn, J. A., B. J. Evans, J. L. Sedlock, F. A. A. Khan, and L. R. Heaney (2012). Single-locus species delimitation: a test of the mixed Yule–coalescent model, with an empirical application to Philippine round-leaf bats. Proceedings of the Royal Society of London B: Biological Sciences:rspb20120705.

Evanno, G., S. Regnaut, and J. Goudet (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology 14:2611–2620.

Everson, K. M., K. B. P. Hildebrandt, S. M. Goodman, and L. E. Olson (2018). Caught in the act: incipient speciation across a latitudinal gradient in a semifossorial mammal from Madagascar, the mole tenrec *Oryzorictes hova* (Tenrecidae). Molecular Phylogenetics and Evolution. https://doi.org/10.1016/j.ympev.2018.02.024

Ezard, T. H., P. Pearson, and A. Purvis (2010). Algorithmic approaches to aid species' delimitation in multidimensional morphospace. BMC Evolutionary Biology 10:11.

Fernández, R., and G. Giribet (2014). Phylogeography and species delimitation in the New Zealand endemic, genetically hypervariable harvestman species, *Aoraki denticulata* (Arachnida, Opiliones, Cyphophthalmi). Invertebrate Systematics 28:401–414.

Filzmoser, P., R. Maronna, and M. Werner (2008). Outlier identification in high dimensions. Computational Statistics & Data Analysis 52:1694–1711.

Flores-Rentería, L., P. D. Rymer, and M. Riegler (2017). Unpacking boxes: Integration of molecular, morphological and ecological approaches reveals extensive patterns of reticulate evolution in box eucalypts. Molecular Phylogenetics and Evolution 108:70–87.

Fraley, C., and A. E. Raftery (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. Journal of the American Statistical Association 97:611–631.

Freeman, B. G., and G. A. Montgomery (2017). Using song playback experiments to measure species recognition between geographically isolated populations: A comparison with acoustic trait analyses. The Auk 134:857–870.

Freudenstein, J. V., M. B. Broe, R. A. Folk, and B. T. Sinn (2017). Biodiversity and the Species Concept—Lineages are not Enough. Systematic Biology 66:644–656.

Frost, D. R., and A. G. Kluge (1994). A consideration of epistemology in systematic biology, with special reference to species. Cladistics 10:259–294.

Fujisawa, T., and T. G. Barraclough (2013). Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets. Systematic Biology 62:707–724.

Fujita, M. K., A. D. Leaché, F. T. Burbrink, J. A. McGuire, and C. Moritz (2012). Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology & Evolution 27:480–488.

Godfray, H. C. J. (2002). Challenges for taxonomy. Nature 417:17–19.

Grummer, J. A., R. W. Bryson, and T. W. Reeder (2014). Species Delimitation Using Bayes Factors: Simulations and Application to the *Sceloporus scalaris* Species Group (Squamata: Phrynosomatidae). Systematic Biology 63:119–133.

Hamilton, C. A., D. R. Formanowicz, and J. E. Bond (2011). Species Delimitation and Phylogeography of *Aphonopelma hentzi* (Araneae, Mygalomorphae, Theraphosidae): Cryptic Diversity in North American Tarantulas. PLOS ONE 6:e26207.

Hamilton, C. A., B. E. Hendrixson, M. S. Brewer, and J. E. Bond (2014). An evaluation of sampling effects on multiple DNA barcoding methods leads to an integrative approach for delimiting species: A case study of the North American tarantula genus *Aphonopelma* (Araneae, Mygalomorphae, Theraphosidae). Molecular Phylogenetics and Evolution 71:79–93.

Harrison, R. G. (1998). Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In Endless Forms: Species and Speciation (D. J. Howard and S. H. Berlocher, Editors). Oxford University Press, pp. 19–31.

Harvey, M., G. F. Seeholzer, B. T. Smith, D. L. Rabosky, A. M. Cuervo, J. T. Klicka, and R. T. Brumfield (2016). A positive association between population genetic differentiation and speciation rates in New World birds. bioRxiv:085134.

Hedin, M. (2015). High-stakes species delimitation in eyeless cave spiders (*Cicurina*, Dictynidae, Araneae) from central Texas. Molecular Ecology 24:346–361.

Hedin, M., D. Carlson, and F. Coyle (2015). Sky island diversification meets the multispecies coalescent – divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. Molecular Ecology 24:3467–3484.

Hedin, M., S. Derkarabetian, M. J. Ramírez, C. Vink, and J. E. Bond (2018). Phylogenomic reclassification of the world's most venomous spiders (Mygalomorphae, Atracinae), with implications for venom evolution. Scientific Reports 8:1636.

Herrera, S., and T. M. Shank (2016). RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Molecular Phylogenetics and Evolution 100:70–79.

Hey, J. (2006). On the failure of modern species concepts. Trends in Ecology & Evolution 21:447–450.

Hey, J. (2009). On the arbitrary identification of real species. In Speciation and Patterns of Diversity (R. K. Butlin, J. R. Bridle and D. Schluter, Editors). Cambridge University Press, pp. 15–28.

Hey, J., and C. Pinho (2012). Population Genetics and Objectivity in Species Diagnosis. Evolution 66:1413–1429.

Huang, H., and L. L. Knowles (2016a). Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. Systematic Biology 65:357–365.

Huang, J.-P., and L. L. Knowles (2016b). The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single Bayesian approach in Hercules beetles. Systematic Biology 65:685–699.

Hudson, R. R. (1990). Gene genealogies and the coalescent process. Oxford surveys in evolutionary biology 7:44.

Jackson, N. D., B. C. Carstens, A. E. Morales, and B. C. O'Meara (2017). Species Delimitation with Gene Flow. Systematic Biology 66:799–812.

Jacobs, S. J., C. Kristoferson, S. Uribe-Convers, M. Latvis, and D. C. Tank (2018). Incongruence in molecular species delimitation schemes: what to do when adding more data is difficult. Molecular Ecology 0.

Janes, J. K., J. M. Miller, J. R. Dupuis, R. M. Malenfant, J. C. Gorrell, C. I. Cullingham, and R. L. Andrew (2017). The K = 2 conundrum. Molecular Ecology 26:3594–3602.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Jombart, T., and I. Ahmed (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071.

Kass, R. E., and A. E. Raftery (1995). Bayes Factors. Journal of the American Statistical Association 90:773–795.

Keith, R., and M. Hedin (2012). Extreme mitochondrial population subdivision in southern Appalachian paleoendemic spiders (Araneae: Hypochilidae: Hypochilus), with implications for species delimitation. Journal of Arachnology 40:167–181.

Kopelman, N. M., J. Mayzel, M. Jakobsson, N. A. Rosenberg, and I. Mayrose (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15:1179–1191.

Kuchta, S. R., A. D. Brown, and R. Highton (2018). Disintegrating over space and time: Paraphyly and species delimitation in the Wehrle's Salamander complex. Zoologica Scripta 47:285–299.

Kury, A. B. (2013). Order Opiliones Sundevall, 1833. In : Zhang, Z.-Q. (Ed.) Animal Biodiversity: An Outline of Higher-level Classification and Survey of Taxonomic Richness (Addenda 2013). Zootaxa 3703:27–33.

Lambertz, M. (2017). Taxonomy: retain scientific autonomy. Nature 546:600–600.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. Molecular Biology and Evolution 29:1695–1701.

Lamichhaney, S., F. Han, M. T. Webster, L. Andersson, B. R. Grant, and P. R. Grant (2017). Rapid hybrid speciation in Darwin's finches. Science:eaao4593.

Lawson, D. J., L. van Dorp, and D. Falush (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature Communications 9:3258.

Leaché, A. D., A. S. Chavez, L. N. Jones, J. A. Grummer, A. D. Gottscho, and C. W. Linkem (2015). Phylogenomics of Phrynosomatid Lizards: Conflicting Signals from Sequence Capture versus Restriction Site Associated DNA Sequencing. Genome Biology and Evolution 7:706–719.

Leaché, A. D., M. K. Fujita, V. N. Minin, and R. R. Bouckaert (2014). Species Delimitation using Genome-Wide SNP Data. Systematic Biology 63:534–542.

Leaché, A., T. Zhu, B. Rannala, and Z. Yang (2018). The Spectre of Too Many Species. bioRxiv:263996.

Leavitt, D. H., J. Starrett, M. F. Westphal, and M. Hedin (2015). Multilocus sequence data reveal dozens of putative cryptic species in a radiation of endemic Californian mygalomorph spiders (Araneae, Mygalomorphae, Nemesiidae). Molecular Phylogenetics and Evolution 91:56–67.

Leonard, W. P., L. Chichester, C. H. Richart, and T. A. Young (2011). *Securicauda hermani* and *Carinacauda stormi*, two new genera and species of slug from the Pacific Northwest of the United States (Gastropoda: Stylommatophora: Arionidae), with notes on *Gliabates oregonius* Webb 1959. Zootaxa 2746:43–56.

Lipscomb, D., N. Platnick, and Q. Wheeler (2003). The intellectual content of taxonomy: a comment on DNA taxonomy. Trends in Ecology & Evolution 18:65–66.

Lohse, K. (2009). Can mtDNA Barcodes Be Used to Delimit Species? A Response to Pons et al. (2006). Systematic Biology 58:439–442.

Lucid, M. K., A. Rankin, A. Espíndola, L. Chichester, S. Ehlers, L. Robinson, and J. Sullivan (2018). Taxonomy and biogeography of Hemphillia (Gastropoda: Pulmonata: Arionidae) in North American rainforests, with description of a new species (Skade's jumping-slug, *Hemphillia skadei* sp. nov.). Canadian Journal of Zoology:305–316.

Mace, G. M. (2004). The role of taxonomy in species conservation. Philosophical Transactions of the Royal Society of London B: Biological Sciences 359:711–719.

Maddison, W. P. (1997). Gene Trees in Species Trees. Systematic Biology 46:523–536.

Mallien, C., B. Porro, T. Zamoum, C. Olivier, J. Wiedenmann, P. Furla, and D. Forcioli (2018). Conspicuous morphological differentiation without speciation in *Anemonia viridis* (Cnidaria, Actiniaria). Systematics and Biodiversity 16:271–286.

Manel, S., C. Perrier, M. Pratlong, L. Abi-Rached, J. Paganini, P. Pontarotti, and D. Aurelle (2016). Genomic resources and their influence on the detection of the signal of positive selection in genome scans. Molecular Ecology 25:170–184.

Marshall, J. C., E. Arévalo, E. Benavides, J. L. Sites, J. W. Sites, and J. Hey (2006). Delimiting species: comparing methods for Mendelian characters using lizards of the *Sceloporus grammicus* (Squamata: Phrynosomatidae) complex. Evolution 60:1050–1065.

May, R. M. (2002). The future of biological diversity in a crowded world. Current Science 82:1325–1331.

Mayden, R. L. (1997). A hierarchy of species concepts: the denouement in the saga of the species problem. In Species: The Units of Biodiversity (M. F. Claridge, H. A. Dawah and M. R. Wilson, Editors). Chapman & Hall, pp. 381–424.

Mayr, E. (1963). Animal species and evolution. Harvard University Press, Cambridge, MA.

Mayr, E., and P. D. Ashlock (1991). Principles of systematic zoology. McGraw-Hill New York.

McKitrick, M. C., and R. M. Zink (1988). Species Concepts in Ornithology. The Condor 90:1–14.

McMurray, F. (1955). Preface to an Autonomous Discipline of Education. Educational Theory 5:129–140.

Miller, M. A., W. Pfeiffer, and T. Schwartz (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. IEEE, pp. 1–8.

Miralles, A., and M. Vences (2013). New Metrics for Comparison of Taxonomies Reveal Striking Discrepancies among Species Delimitation Methods in *Madascincus* Lizards. PLOS ONE 8:e68242.

Mishler, B. D., and M. J. Donoghue (1982). Species Concepts: A Case for Pluralism. Systematic Zoology 31:491–503.

Monaghan, M. T., R. Wild, M. Elliot, T. Fujisawa, M. Balke, D. J. G. Inward, D. C. Lees, R. Ranaivosolo, P. Eggleton, T. G. Barraclough, and A. P. Vogler (2009). Accelerated Species Inventory on Madagascar Using Coalescent-Based Models of Species Delineation. Systematic Biology 58:298–311.

Nee, S., R. M. May, P. H. Harvey, S. Nee, R. M. May, and P. H. Harvey (1994). The reconstructed evolutionary process. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 344:305–311.

Nolte, A. W., and D. Tautz (2010). Understanding the onset of hybrid speciation. Trends in Genetics 26:54–58.

Noutsos, C., J. O. Borevitz, and S. A. Hodges (2014). Gene flow between nascent species: geographic, genotypic and phenotypic differentiation within and between *Aquilegia formosa* and *A. pubescens*. Molecular Ecology 23:5589–5598.

O'Leary, S. J., J. B. Puritz, S. C. Willis, C. M. Hollenbeck, and D. S. Portnoy (2018). These aren't the loci you're looking for: Principles of effective SNP filtering for molecular ecologists. Molecular Ecology 27:3193–3206.

Packer, L., S. K. Monckton, T. M. Onuferko, and R. R. Ferrari (2018). Validating taxonomic identifications in entomological research. Insect Conservation and Diversity 11:1–12.

Padial, J. M., A. Miralles, I. De la Riva, and M. Vences (2010). The integrative future of taxonomy. Frontiers in Zoology 7:16.

Page, T. J., S. C. Choy, and J. M. Hughes (2005). The taxonomic feedback loop: symbiosis of morphology and molecules. Biology Letters 1:139–142.

Patten, M. A., and J. V. Remsen (2017). Complementary Roles of Phenotype and Genotype in Subspecies Delimitation. Journal of Heredity 108:462–464.

Pearson, P. N., and T. H. Ezard (2013). Evolution and speciation in the Eocene planktonic foraminifer *Turborotalia*. Paleobiology 40:130–143.

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. Computational Statistics & Data Analysis 49:974–997.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7:e37135.

Pinto-da-Rocha, R., G. Machado, and G. Giribet (2007). Harvestmen: the biology of Opiliones. Harvard University Press.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler (2006). Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Systematic Biology 55:595–609.

Pritchard, J. K., M. Stephens, and P. Donnelly (2000). Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959.

Pritchard, J. K., and W. Wen (2002). Documentation for STRUCTURE software: version 2. University of Chicago, Chicago.

Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: subsampling and new estimators alleviate the problem. Molecular Ecology Resources 16:608–627.

R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Rambaut A, A. J. Drummond (2014) TREEANNOTATOR v2. 1.2. Edinburgh: University of Edinburgh, Institute of Evolutionary Biology.

Reaz, R., M. S. Bayzid, and M. S. Rahman (2014). Accurate Phylogenetic Tree Reconstruction from Quartets: A Heuristic Approach. PLOS ONE 9:e104008.

Reid, N., and B. Carstens (2012). Phylogenetic estimation error can decrease the accuracy of species delimitation: a Bayesian implementation of the general mixed Yule-coalescent model. BMC Evolutionary Biology 12:196.

Richart, C. H. (2014). Taxonomic Revision of the Opilionid Genus *Acuclavella* Shear, 1986 (Opiliones, Ischyropsalidoidea) based on Molecular Phylogenetics and Morphometrics with Description of Four New Species. Master's Thesis. San Diego State University, Montezuma Publishing, San Diego, California. Available at http://scholarworks.calstate.edu/handle/10211.3/115826.

Richart, C. H., C. Y. Hayashi, and M. Hedin (2016). Phylogenomic analyses resolve an ancient trichotomy at the base of Ischyropsalidoidea (Arachnida, Opiliones) despite

high levels of gene tree conflict and unequal minority resolution frequencies. Molecular Phylogenetics and Evolution 95:171–182.

Richart, C., and M. Hedin (2013). Three new species in the harvestmen genus *Acuclavella* (Opiliones, Dyspnoi, Ischyropsalidoidea), including description of male *Acuclavella quattuor* Shear, 1986. ZooKeys 311:19–68.

Rosenberg, N. A. (2004). distruct: a program for the graphical display of population structure. Molecular Ecology Notes 4:137–138.

Rubin, B. E. R., R. H. Ree, and C. S. Moreau (2012). Inferring Phylogenies from RAD Sequence Data. PLoS ONE 7:e33394.

Satler, J. D., B. C. Carstens, and M. Hedin (2013). Multilocus Species Delimitation in a Complex of Morphologically Conserved Trapdoor Spiders (Mygalomorphae, Antrodiaetidae, *Aliatypus*). Systematic Biology 62:805–823.

Schlick-Steiner, B. C., W. Arthofer, and F. M. Steiner (2014). Take up the challenge! Opportunities for evolution research from resolving conflict in integrative taxonomy. Molecular Ecology 23:4192–4194.

Schlick-Steiner, B. C., F. M. Steiner, B. Seifert, C. Stauffer, E. Christian, and R. H. Crozier (2010). Integrative taxonomy: a multisource approach to exploring biodiversity. Annual Review of Entomology 55:421–438.

Schwarzfeld, M. D., and F. A. H. Sperling (2015). Comparison of five methods for delimitating species in *Ophion* Fabricius, a diverse genus of parasitoid wasps (Hymenoptera, Ichneumonidae). Molecular Phylogenetics and Evolution 93:234–248.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, Å. Brännström, A. Brelsford, et al. (2014). Genomics and the origin of species. Nature Reviews Genetics 15:176–192.

Seifert, B., M. Ritz, and S. Csösz (2013). Application of Exploratory Data Analyses opens a new perspective in morphology-based alpha-taxonomy of eusocial organisms. Myrmecological News 19:1–15.

Shaffer, H. B., and R. C. Thomson (2007). Delimiting Species in Recent Radiations. Systematic Biology 56:896–906.

Shapiro, B. J., J.-B. Leducq, and J. Mallet (2016). What Is Speciation? PLOS Genetics 12:e1005860.

Shear, W. A. (1986). A cladistic analysis of the opilionid superfamily Ischyropsalidoidea, with descriptions of the new family Ceratolasmatidae, the new genus *Acuclavella*, and four new species. American Museum novitates; no. 2844.

Shelley, R. M., C. H. Richart, and A. E. Bogan (2010). *Octoglena claraqua*, n. sp.(Polyzoniida: Hirudisomatidae), a new milliped from Idaho, USA; first record of the order from the western interior of North America. Zootaxa 2446:55–64.

Simpson, G. G. (1951). The species concept. Evolution 5:285–298.

Singhal, S., H. Huang, M. R. Grundler, M. R. Marchán-Rivadeneira, I. Holmes, P. O. Title, S. C. Donnellan, and D. L. Rabosky (2018). Does Population Structure Predict the Rate of Speciation? A Comparative Test across Australia's Most Diverse Vertebrate Radiation. The American Naturalist 192:432–447.

Sites Jr, J. W., and J. C. Marshall (2004). Operational criteria for delimiting species. Annual Review of Ecology, Evolution, and Systematics:199–227.

Solís-Lemus, C., L. L. Knowles, and C. Ané (2015). Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution 69:492–507.

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

Starrett, J., and M. Hedin (2007). Multilocus genealogies reveal multiple cryptic species and biogeographical complexity in the California turret spider *Antrodiaetus riversi* (Mygalomorphae, Antrodiaetidae). Molecular Ecology 16:583–604.

Starrett, J., S. Derkarabetian, M. Hedin, R. W. Bryson, J. E. McCormack, and B. C. Faircloth (2017). High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. Molecular Ecology Resources 17:812–823.

Sukumaran, J., and L. L. Knowles (2017). Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences:201607921.

Swofford, D. (2003). PAUP*: phylogenetic analysis using parsimony, version 4.0b10.

Templeton, A. R. (1989). The meaning of species and speciation: a genetic perspective. The Units of Evolution (Cambridge: MIT, 1992):159–83.

Thompson, B. (1999). Statistical Significance Tests, Effect Size Reporting and the Vain Pursuit of Pseudo-Objectivity. Theory & Psychology 9:191–196.

Toussaint, E. F. A., J. Morinière, C. J. Müller, K. Kunte, B. Turlin, A. Hausmann, and M. Balke (2015). Comparative molecular species delimitation in the charismatic Nawab

butterflies (Nymphalidae, Charaxinae, *Polyura*). Molecular Phylogenetics and Evolution 91:194–209.

Vähä, J.-P., J. Erkinaro, E. Niemelä, and C. R. Primmer (2007). Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. Molecular Ecology 16:2638–2654.

Via, S. (2009). Natural selection in action during speciation. Proceedings of the National Academy of Sciences 106:9939–9946.

Wachter, G. A., C. Muster, W. Arthofer, G. Raspotnig, P. Föttinger, C. Komposch, F. M. Steiner, and B. C. Schlick-Steiner (2015). Taking the discovery approach in integrative taxonomy: decrypting a complex of narrow-endemic Alpine harvestmen (Opiliones: Phalangiidae: *Megabunus*). Molecular Ecology 24:863–889.

Wagner, C. E., I. Keller, S. Wittwer, O. M. Selz, S. Mwaiko, L. Greuter, A. Sivasundar, and O. Seehausen (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular ecology 22:787–798.

Wagner, P., A. D. Leaché, and M. K. Fujita (2014). Description of four new West African forest geckos of the *Hemidactylus fasciatus* Gray, 1842 complex, revealed by coalescent species delimitation. Bonn Zoological Bulletin 63:1–14.

Waples, R. S., and O. Gaggiotti (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. Molecular Ecology 15:1419–1439.

Wheeler, Q. D. (2004). Taxonomic triage and the poverty of phylogeny. Philosophical Transactions of the Royal Society of London B: Biological Sciences 359:571–583.

Wiens, J. J. (2007). Species Delimitation: New Approaches for Discovering Diversity. Systematic Biology 56:875–878.

Wiens, J. J., and T. A. Penkrot (2002). Delimiting Species Using DNA and Morphological Variation and Discordant Species Limits in Spiny Lizards (*Sceloporus*). Systematic Biology 51:69–91.

Wiley, E. O. (1978). The Evolutionary Species Concept Reconsidered. Systematic Biology 27:17–26.

Wiley, E. O., and R. L. Mayden (2000). The Evolutionary Species Concept. In Species Concepts and Phylogenetic Theory: A Debate (Q. D. Wheeler and R. Meier, Editors). Columbia University Press, pp. 70–90.

Willis, S. C. (2017). One species or four? Yes!...and, no. Or, arbitrary assignment of lineages to species obscures the diversification processes of Neotropical fishes. PLOS ONE 12:e0172349.

Wilson, E. O., and W. L. Brown (1953). The Subspecies Concept and Its Taxonomic Application. Systematic Zoology 2:97–111.

Wu, C.-I. (2001). The genic view of the process of speciation. Journal of Evolutionary Biology 14:851–865.

Yang, Z., and B. Rannala (2017). Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses. Molecular Ecology 26:3028–3036.

Yeaman, S., and M. C. Whitlock (2011). The Genetic Architecture of Adaptation Under Migration–Selection Balance. Evolution 65:1897–1911.

Zhang, J., P. Kapli, P. Pavlidis, and A. Stamatakis (2013). A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29:2869–2876.

Table 2.1. Morphometric samples sizes.

Clade Name	Males	Females	Totals
A. sheari	4	2	6
A. cosmetoides	19	25	44
A. shoshone	27	25	52
A. cf. *quattuor*	22	18	40
Hoodoo Mtns	10	8	18
A. merickeli	19	20	39
A. quattuor	14	17	31
Totals	115	115	230

Table 2.2. STRUCTURE results: log-likelihood values for K=1 and K= ΔK are emboldened when the log-likelihood of K=1 > K=ΔK. *indicate terminally derived populations. † indicates that one of the inferred populations is a "ghost" (see text).

Tier	Missing Data	Sample Description	K set	# Local.	Runs	ΔK	K = 1	K = ΔK
1	57.0%	All Inland Acuclavella	1-7	49	3	3	-46346.6	-34880.6
2	28.5%	Asheari, Acfquattuor, Acosmetoides	1-7	22	3	†4	-24039.5	-19434.4
2	25.5%	AHoodoo, Ashoshone	1-6	15	4	†3	-37446.5	-28790.2
3	25.8%	Acfquattuor	1-5	7	4	2	-33881.9	-26376.9
3	23.4%	Ashoshone	1-5	11	6	2	-40669.8	-37701.9
3	27.8%	Acosmetoides	1-5	9	5	†3	-38095.6	-34186.2
4	24.0%	AcosmetoidesA	1-5	4	5	3	-43348.7	-42814.1
4	25.5%	AcosmetoidesB	1-6	5	3	2	-40241.3	-37779.7
3	18.6%	AHoodoo	1-5	4	6	*3	-31246.0	-18199.2
3	0.0%	Asheari	1-4	2	6	3	**-4688.1**	-4904.6
2	21.0%	Americkeli, Aquattuor	1-5	12	6	*†3	-41485.7	-40200.5
4	26.7%	AcfquattuorA	1-5	4	6	*3	-31480.1	-29735.9
4	16.6%	AcfquattuorB	1-4	3	6	3	**-22053.0**	-22620.5
5	0.0%	AcosmetoidesA2	1-4	2	6	2	**-11912.2**	-12637.6
5	24.1%	AcosmetoidesB1	1-4	3	6	2	**-36115.1**	-36425.9
5	23.9%	AcosmetoidesB2	1-4	2	6	2	**-30214.7**	-33430.9
5	24.6%	AcfquattuorA2	1-4	2	5	3	**-32988.5**	-36684.5

Table 2.3. Morphometric results.

Samples	**sex**	**Measure of Scale**	**# charact.**	**K =**
All	m	mad	14	9
	m	qn	14	5
	f	mad	12	4
	f	qn	12	2
A. cosmetoides	m	mad	12	5
A. shoshone	m	qn	14	1
Hoodoo Mtns	f	mad	12	3
A. merickeli	m	mad	10	2
A. quattuor *A.* cf. *quattuor*	f	mad	12	3

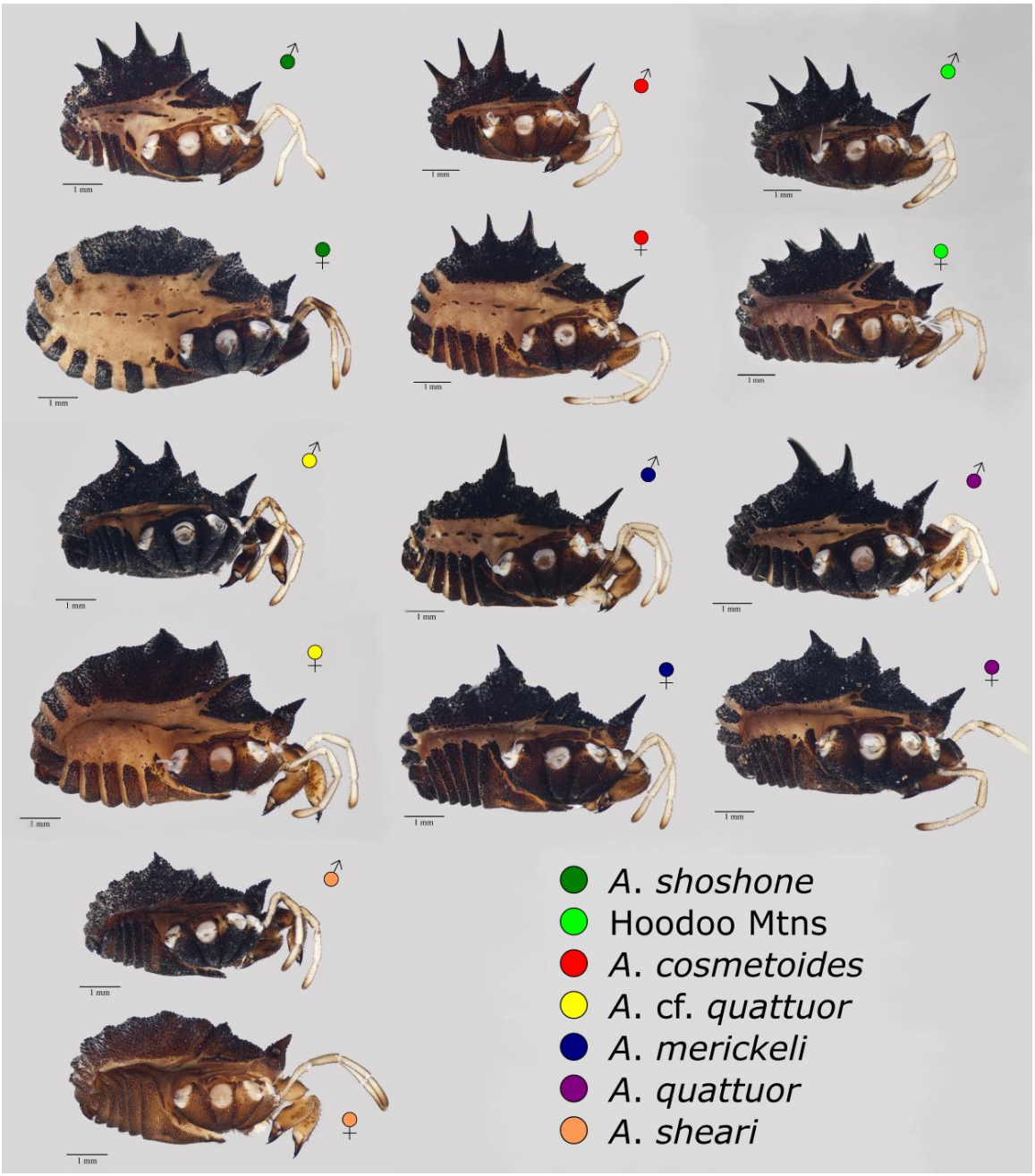


Figure 2.1. Thorn harvestmen images.

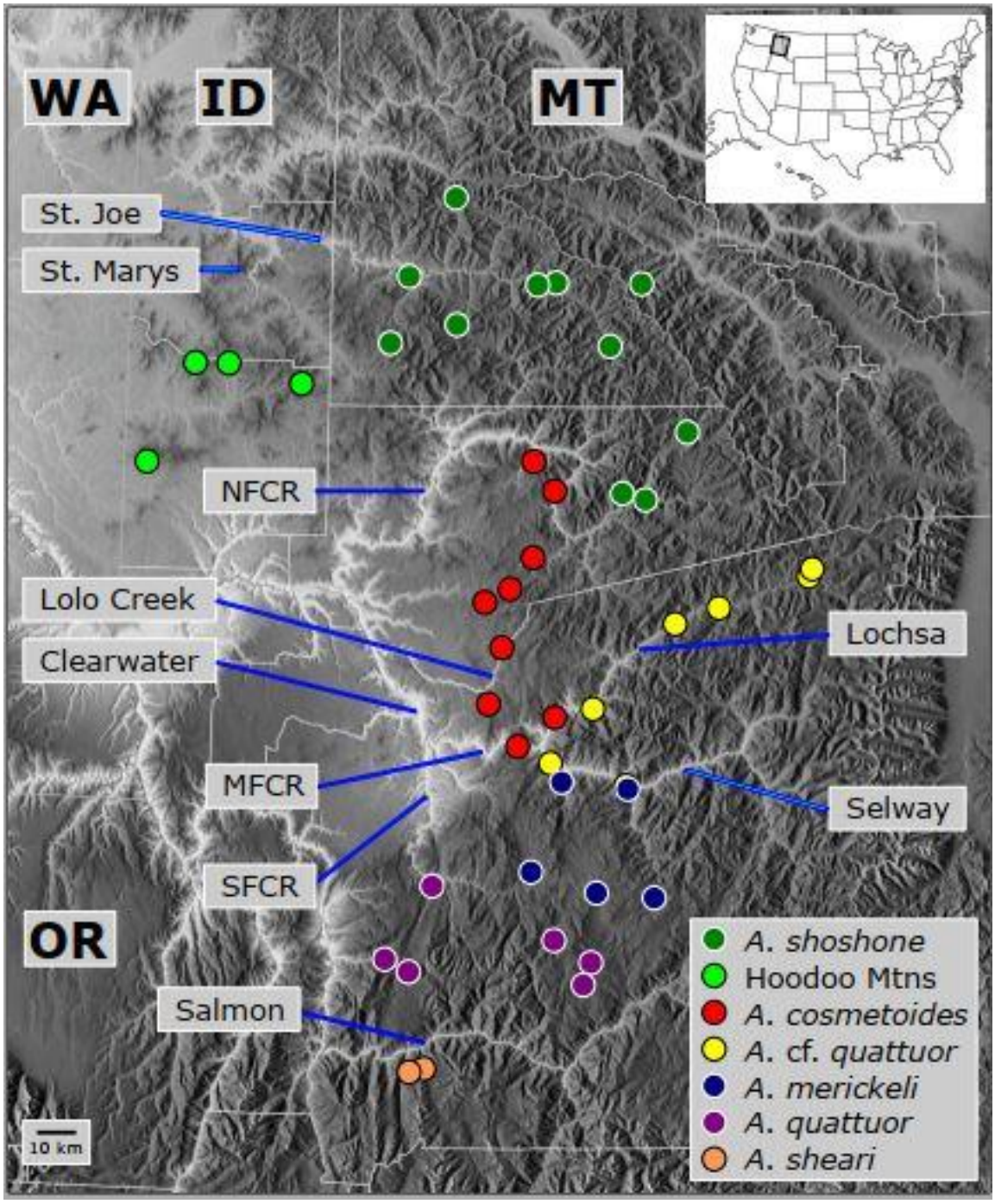


Figure 2.2. Geography.

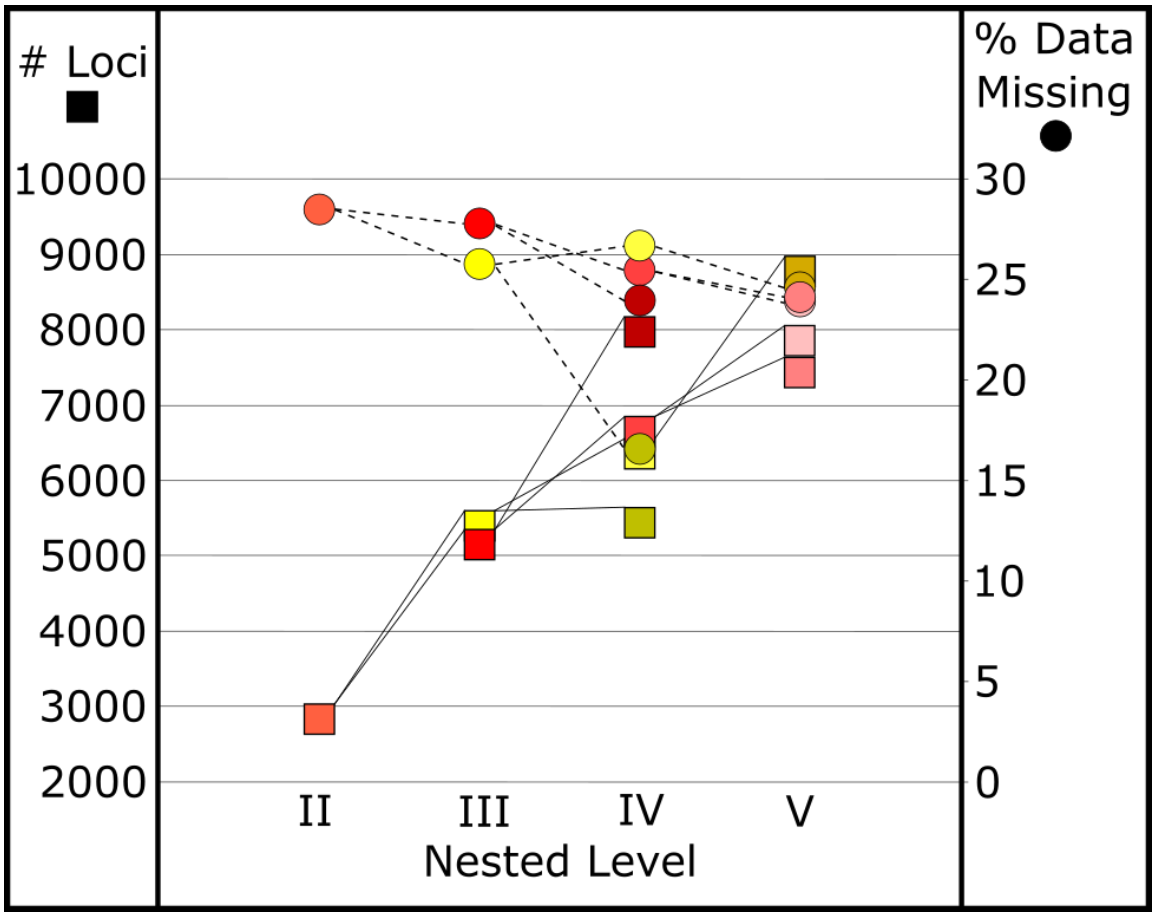


Figure 2.3. Iterative pyRAD data curation. This figure follows the nested iterations of the *Aculavella cosmetoides* and *A*. cf. *quattuor* STRUCTURE analyses with colors matching those used in Figure 2.7. The number of loci are shown using squares and solid lines, and the percent of missing data with circles and dashed lines.

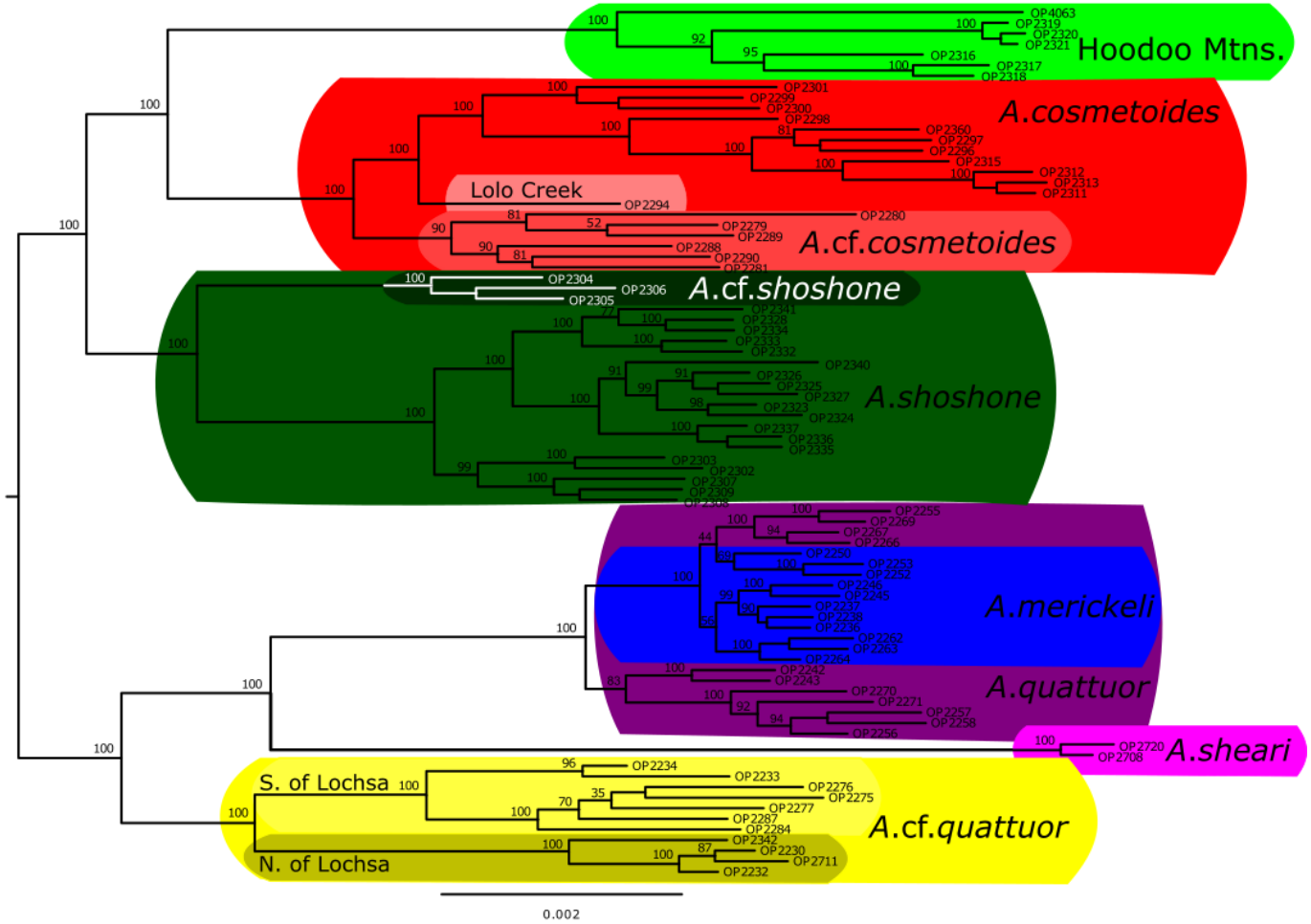


Figure 2.4. RAxML phylogeny. Bootstrap values are not shown within sampling localities when three samples from a locality form a clade.

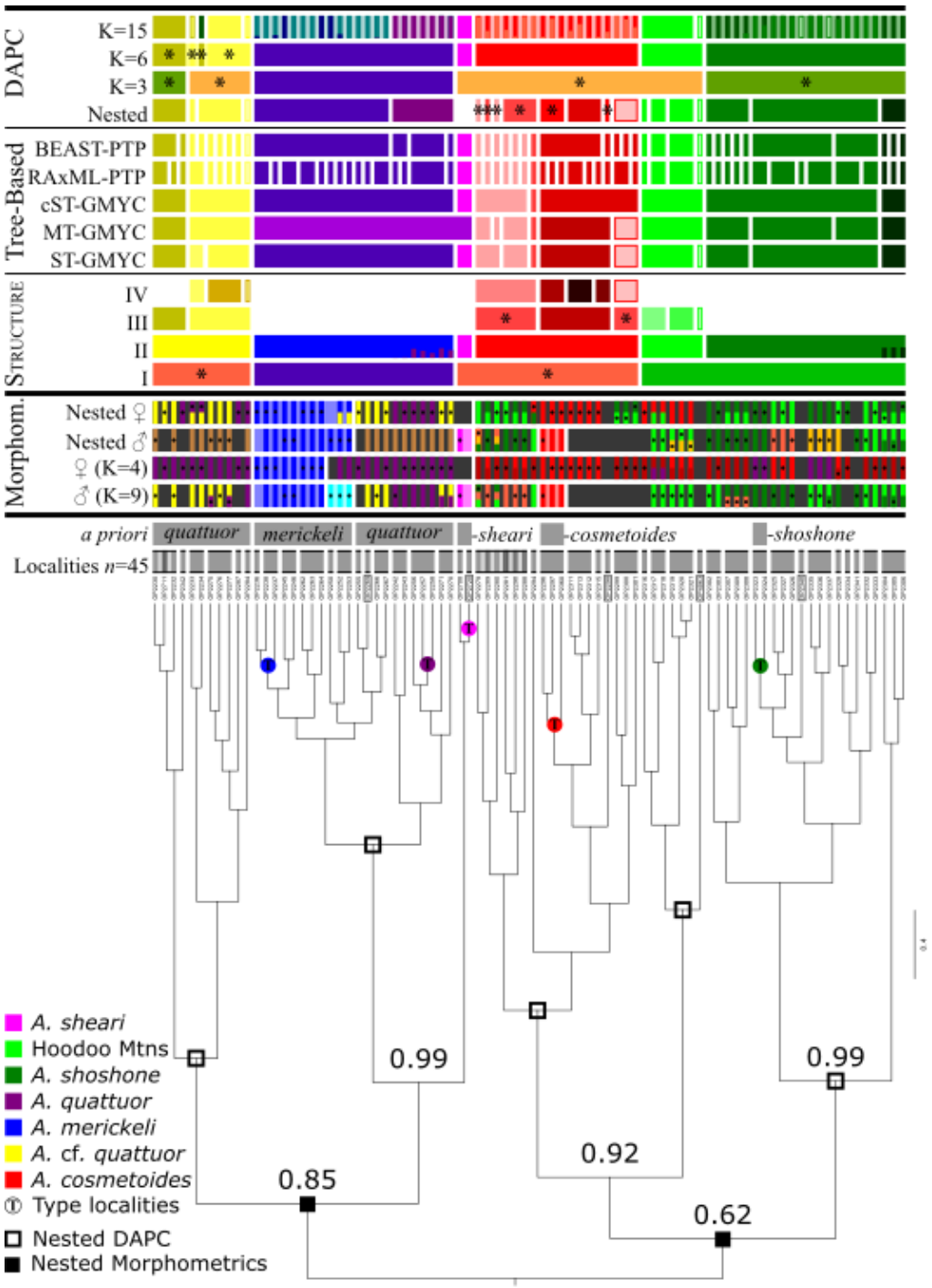


Figure 2.5. Integrative species delimitation. Node values from BEAST (pp); all others 1.0 with three within-locality exceptions. Non-monophyletic entities are indicated with an asterisk (*).

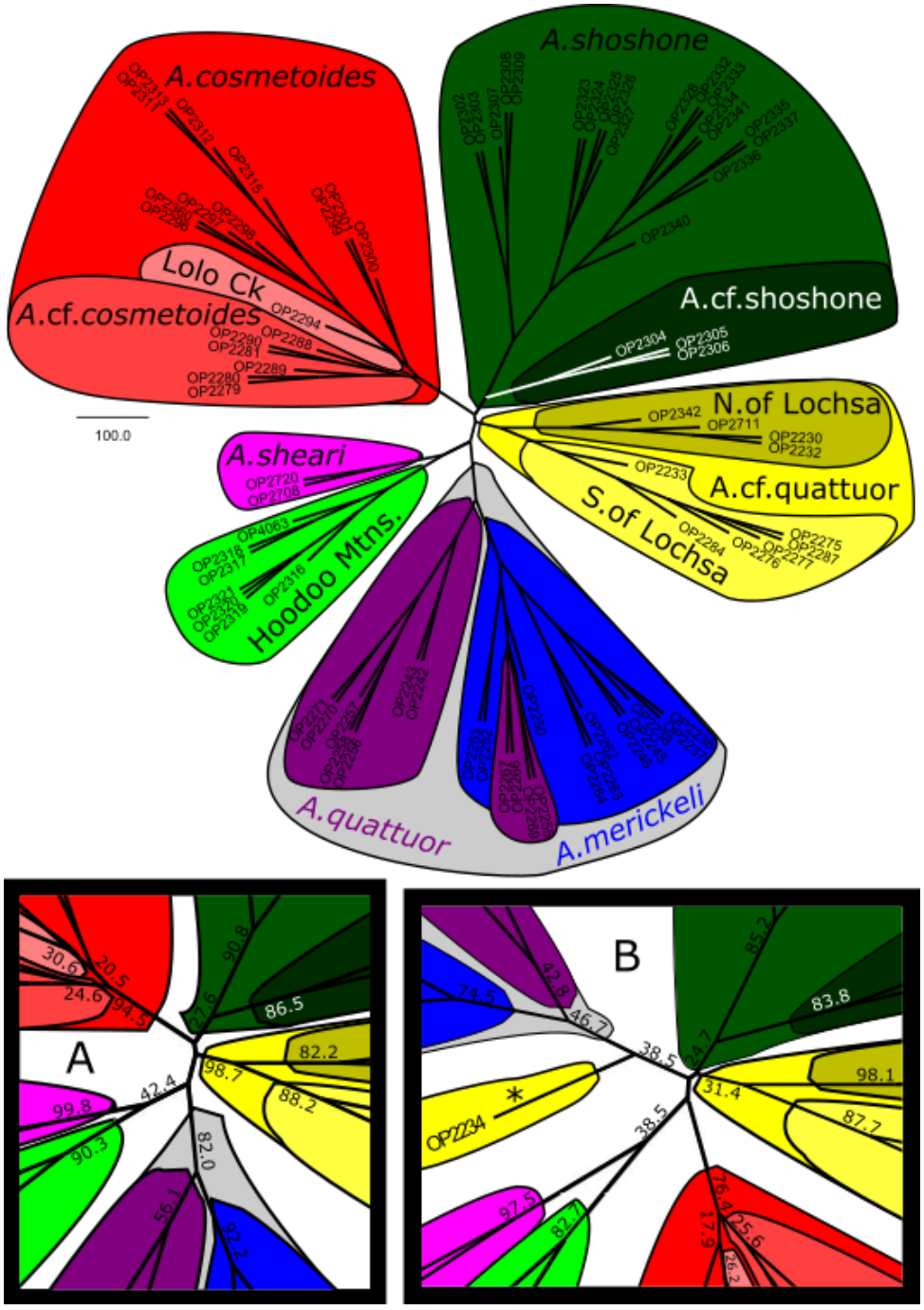


Figure 2.6. SVDQuartets phylogeny. The complete phylogeny above shows relationships inferred from SVDQuartets on the 81 terminal data set. The insets show support values for major clades and potential root nodes for the 81 terminal (A) and 82 terminal (B) matrices; inset B includes OP2234 (*), an *Acuclavella* cf. *quattuor* with a large amount of missing data.

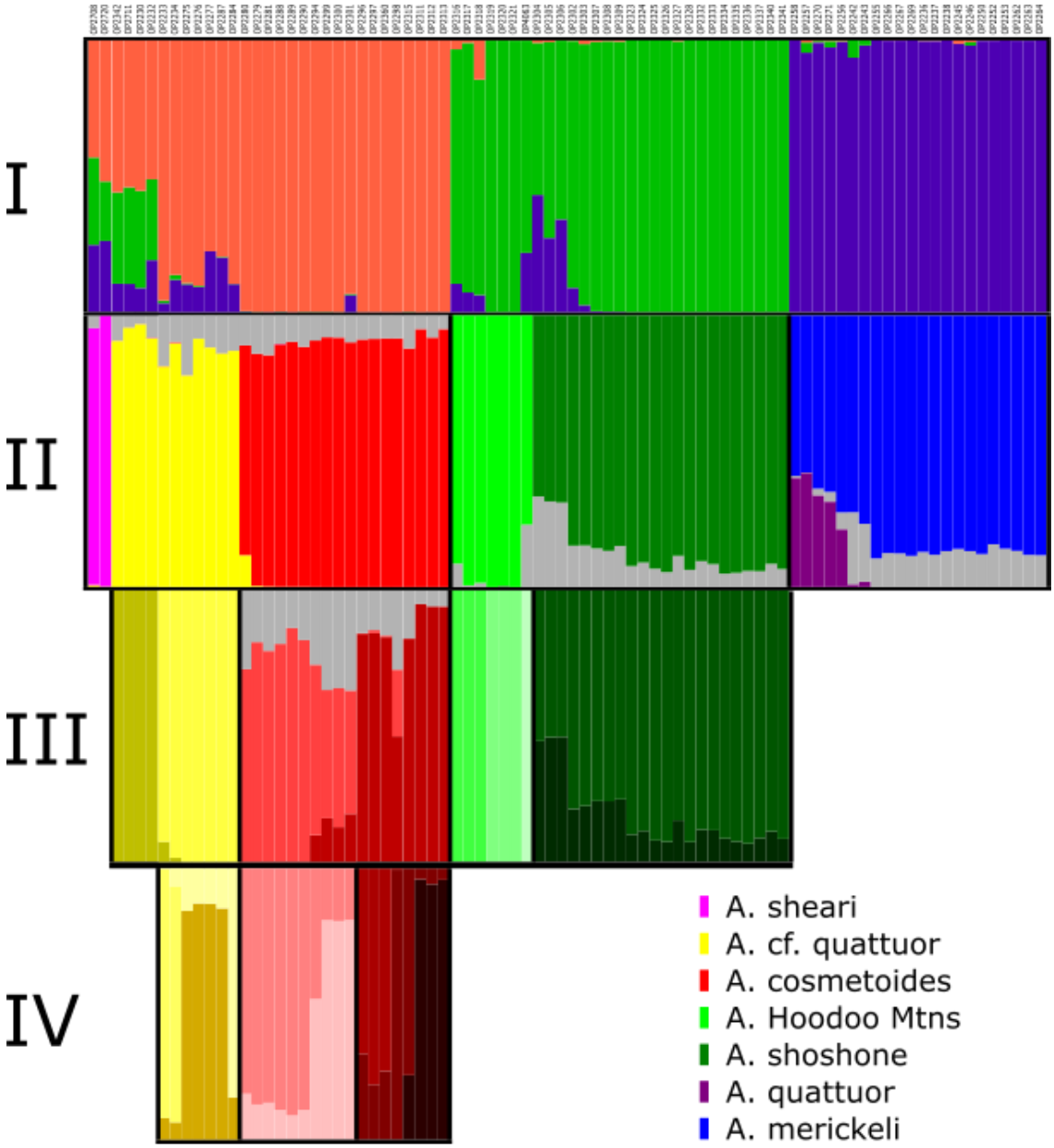


Figure 2.7. STRUCTURE results. The key corresponds to the colors in tier II.

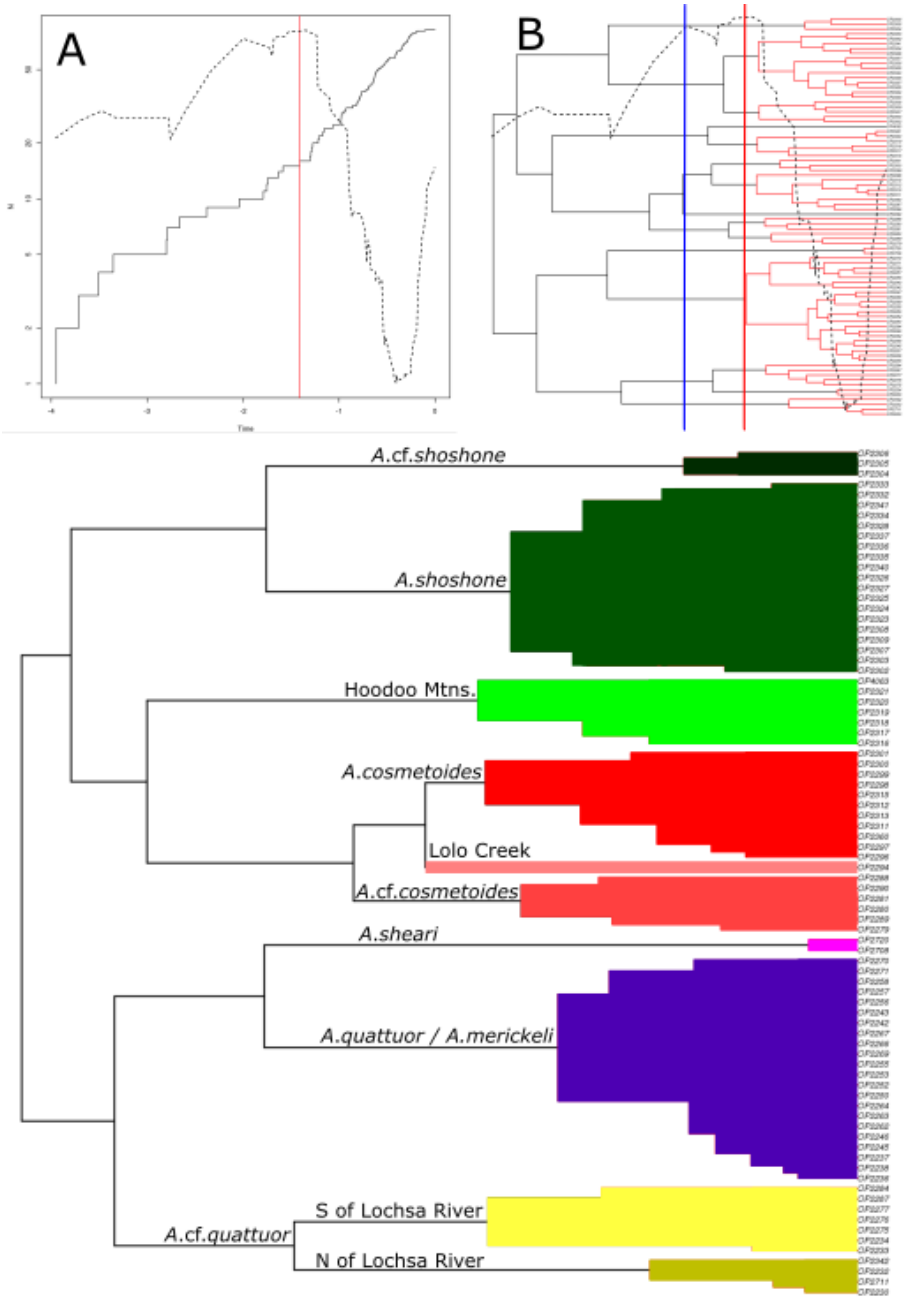


Figure 2.8. GMYC. Inset A shows the maximum likelihood inference of lineages overlaid on the lineage through time plot (note the y-axis scaling). Inset B overlays the likelihood over the input tree with the inferred most likely number of lineages (red line) and the number of putative species considered as species in this research (blue line) from the earlier island of high likelihood. Inferred lineages are labeled and demarked in the input tree below.

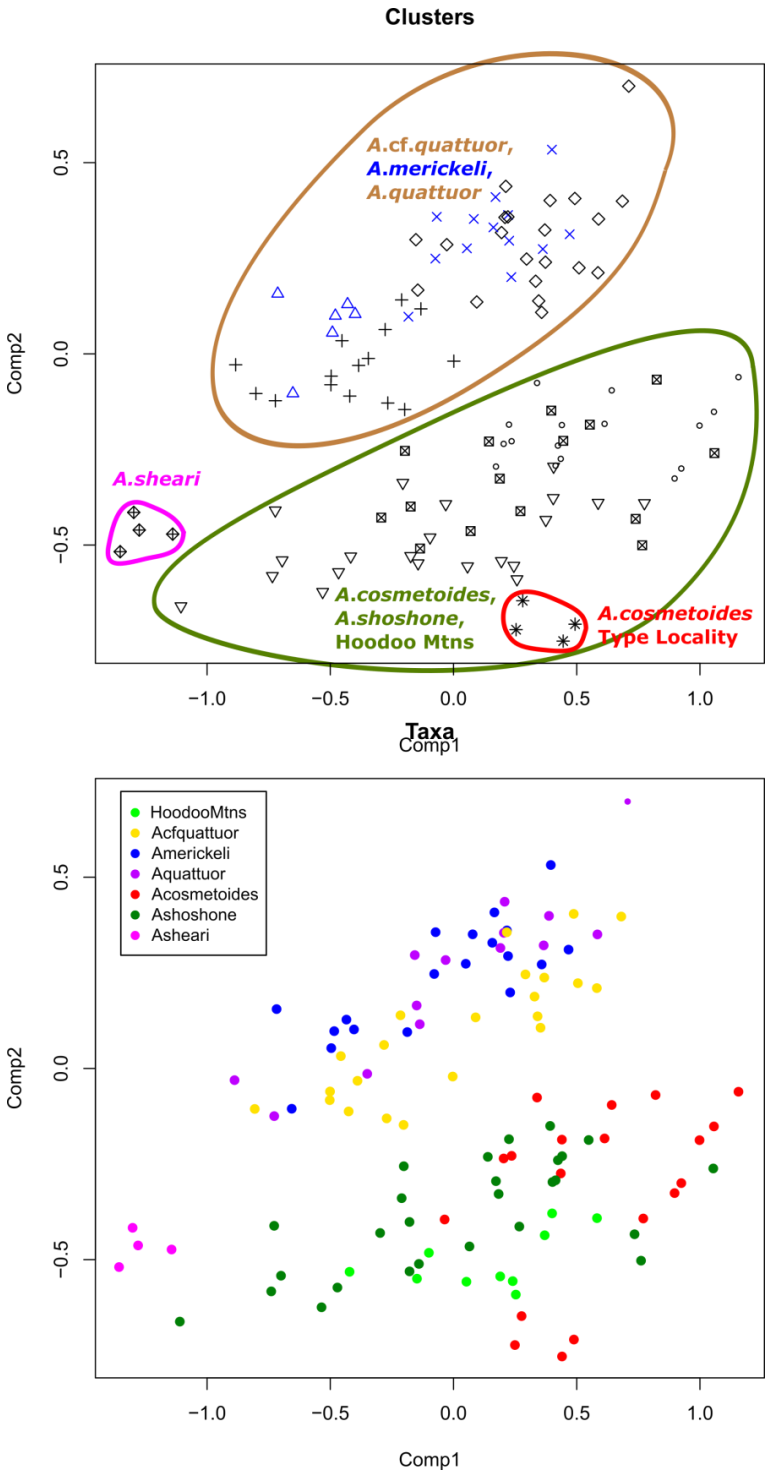


Figure 2.9. Ezard morphometrics. Morphometric analysis of the male data set using 14 characters and the *mad* robust measure of scale. For additional analyses see Appendix G.

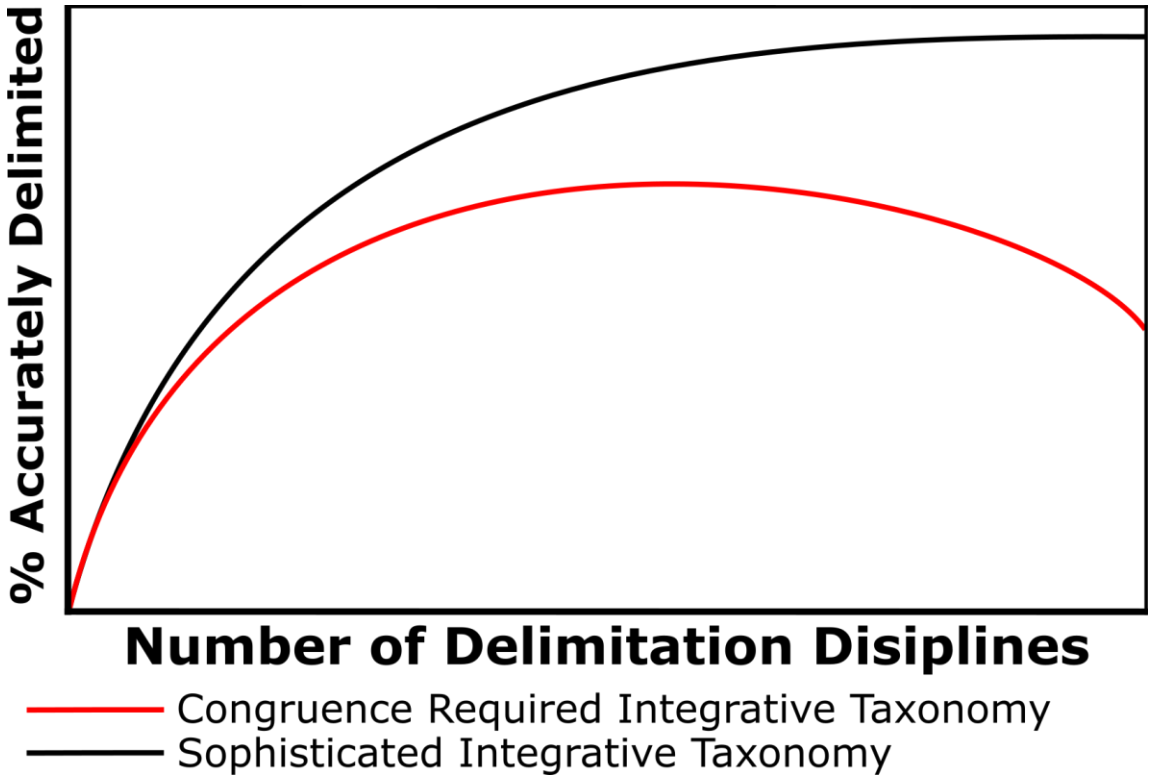


Figure 2.10. Congruence Required vs. Sophisticated Integrative Taxonomy. As more species criterion disciplines are used to delimit species, the reliability (accuracy) of delimitations decline in a congruence required integrative taxonomy, whereas accuracy improves in a sophisticated integrative taxonomy.

CHAPTER 3

A glacial refugia matrix compartmentalized by riverine barriers provides replicate tests of the down-canyon hypothesis

The disjunct inland mesic forest primarily situated in northern Idaho of northwest North America has become a model system for recent biogeography. The majority of this research has focused on an evolutionary history shared by many taxa that show either an ancient vicariance with, or a recent dispersal from, populations inhabiting coastal mesic forests. This vicariance is attributed to the rise of the Cascade Mountains and the subsequent formation of a rain shadow resulting in an intervening xeric scrub that is uninhabitable to organisms dependent on a humid forest environment. Among this research are inquiries that focused on more recent evolutionary events within the inland mesic forest that span the spectrum of biogeography and phylogeography and suggest a complex evolutionary history within this forest system. Observations of speciation, population structure, mitochondrial introgression, and range expansion suggest that this inland forest was a refuge from Pleistocene glaciation, and that during the last glacial maximum (LGM) this refuge was compartmentalized into variously isolated pockets confined to lower elevations by alpine glaciers. In this research, genomic-scale data generated using ddRADseq for a clade of poorly-dispersing ornate harvestmen (Opiliones, Dyspnoi) are used to evaluate the evolutionary consequences of the LGM. The inland species of thorn harvestmen (Ischyropsalidoidea, *Acuclavella*) form a clade endemic to this mesic forest, yet they are widespread within it as a series of allopatric

short-range endemic species. Consequently, these distinct lineages provide replicates upon which to test demographic hypotheses within the inland mesic forest during the LGM. This research finds that rivers were a major cause of refugia compartmentalization, with populations and species structured by rivers well before the LGM. Additionally, a negative correlation between elevation and heterozygosity suggests that thorn harvestmen populations were confined to lower valleys during the LGM with post-glacial colonization of higher elevations, thus supporting the down-canyon hypothesis. The combined effects of rivers and glaciers paints a picture of a complex Pleistocene glacial refugia system during the LGM. This relatively high-resolution inference of past evolutionary events within an established biogeographic model system points towards the undervalued importance of using poorly-dispersing short-range endemic taxa for biogeography and phylogeography.

I. INTRODUCTION

There have been many documented instances of diversification and population-level responses to Pleistocene climate change (e.g., Petit et al. 2003, Potter et al. 2018). Insight into evolutionary events associated with Pleistocene climate change such as diversification, range expansion, population structure, secondary contact, etc. has been revealed across the tree-of-life from mushrooms to mollusks (e.g., Burg et al. 2006, Lanier et al. 2015, Moore et al. 2015, Neiber and Hausdorf 2015, Sánchez-Ramírez et al. 2015, Perez et al. 2016a, Thesing et al. 2016, Harrington et al. 2018), and across the

globe from tropics to tundra (e.g., Haffer 1969, Zemlak et al. 2011, Poncet et al. 2013, Garzón-Orduña et al. 2014, Boyer et al. 2016). One of the well-documented phenomena associated with Pleistocene climate change are biotic elevational shifts that track habitat that moves lower into valley and foothills during peak glaciation and higher into mountains during glacial interludes (e.g., Bidegaray-Batista et al. 2016, Derkarabetian et al. 2016). Another widespread and closely related phenomenon is that after release from the LGM, species expanded their ranges out of refugia, following retreating glaciers and the progressive succession of new forest habitat (e.g., Nielson et al. 2001, Good and Sullivan 2001, Runck and Cook 2005, Hird et al. 2010, García-Vázquez and Ribera 2016, Hedin and McCormack 2017). Typically, the source populations for this expansion would come from the nearest refugial edge. Founding populations are subsequently the source populations for continued expansion, with each founding event leading to more loss of alleles and an increased homozygosity (Hewitt 1996, Cahlil and Levinton 2015). Further, this loss of genetic diversity is expected to be more prominent in poorly dispersing habitat specialists than in species with higher levels of gene flow (Jezkova et al. 2015).

A growing body of research shows that a number of arthropod lineages sharing a suite of characteristics provide remarkable resolution for biogeography and phylogeography inference. Groups such as myriapods and harvestmen are ancient lineages that were among the first invaders of terrestrial habitats (Dunlop and Garwood 2018), and thus are useful for testing hypotheses of ancient supercontinents and plate tectonics (e.g. Boyer et al. 2007, Shelley and Golovatch 2011, Giribet et al. 2012, Hedin et al. 2012, Fernández et al. 2017, Fernández et al. 2018, Oberski et al. 2018). These

groups, and other taxa such as scorpions and mygalomorph spiders, are poorly dispersing microhabitat specialists, with constrained gene flow between populations (e.g., Hedin et al. 2013, Bidegaray-Batista et al. 2016, Emata and Hedin 2016, Schwentner and Giribet 2018). Likely, viable populations are able to persist in relatively small refuges with demographic histories reflecting a finer resolution than is seen in vertebrate species (Hugall et al. 2002, Yeates et al. 2002). Thus, this suite of organisms provides valuable insight within and across a great many biogeographic systems and time scales (e.g., Ceccarelli et al. 2016, Derkarabetian et al. 2016, Xu et al. 2016, Mora et al. 2017). These types of organisms are becoming the reverie of biogeographers around the world.

Approximately 18 million years ago (MYA) the PNW housed a rich mesic warm-temperate summer-wet mixed forest of redwoods and deciduous hardwoods. This forest extended from the west slope of the Northern Rockies in what is present-day northern Idaho west across the Columbia Plateau, connecting mesic coastal forests to the interior (Leopold and Denton 1987). About 16.5 MYA, the Yellowstone Hotspot plume first surfaced near the present-day Nevada-Oregon-Idaho border and periodic massive flood basalt flows would shape the Columbia Plateau into a 63,000 square mile slab of basalt over 6,000 feet thick (Hooper 1997). About 16 MYA one of the earliest of these flows, the Grande Ronde Basalt, was emplaced as a massive and extensive horizontal sheet of flood basalt that covers about 150,00 km^2 of Washington, Oregon, and Idaho (Tolan et al. 1989). Periodic basalt flows would impound rivers, creating forest-inundating lakes where sediments would help form some of the best-documented and finest preserved Miocene floras and faunas (Russell 1900, Smith 1903, Leopold and Denton 1987, Hooper

1997). A corresponding center of volcanism further west were ancient surface volcanoes, vents of intruding magma, caused by the melting of the Juan de Fuca Plate as it subducted under the North American Plate; the same source forming modern Cascade volcanoes such as Mount Rainier and Mount Saint Helens. But these ancient volcanoes rose above a landscape of only modest relief, one that did not yet impede the inland passage of coastal precipitation (Cross and Taggart 1982, Vance et al. 1987, Mitchell and Montgomery 2006). Episodes of extended western volcanism occurred again in the mid-Miocene (11 - 9 MYA), then again across the Neogene-Quaternary boundary (5.7 - 3.4 MYA). Southern Cascade Mountain uplift initiated during this time, most strongly verified by the warping and uplift of the Grande Ronde Basalt (Smith 1903, Hammond 1979, Mitchell and Montgomery 2006). As the Cascade Mountains rose a rain shadow formed along its eastern flanks, this xeric shadow creeping east in tandem with the growing height of the range. With this reduction of rainfall, the Columbia Basin flora transitioned from wet-loving hardwoods, redwoods (*Sequoia*), and cypress (*Taxodium*) in the early Miocene, to xeric forests of oaks (*Quercus* spp.) and madrona (*Arbutus*) in the late Miocene (Chaney 1938). This trend to ever more dry-tolerate vegetation continued, and the sagebrush (*Artemisia*) dominated scrub that currently occupies eastern Washington was in place by 3 - 2 MYA (Leopold and Denton 1987, Graham 1999). This in effect isolated an inland mesic forest on the west slope of the northern Rocky Mountains. This inland forest is maintained by westerly moisture moving unimpeded over the flat and low Columbia Basin until orographic precipitation is induced by the Bitterroot Mountains (Daubenmire 1975, Franklin and Dyrness 1988, Gavin 2009). Over

two hundred species thus far have been documented to either have disjunct populations or endemic species that formed due to this isolation (e.g., amphibians: Nielson et al. 2001, Carstens et al. 2004, Steele et al. 2005; arachnids: Derkarabetian et al. 2010, Richart and Hedin 2013; insects: Barr 2011, terrestrial gastropods: Lucid et al. 2018; plants: Lorain 1988, Björk 2010, Brunsfeld et al. 2007, Gavin 2009). This interior WHZ has become the North American example of Africa's Rand Flora pattern, where numerous unrelated species show similar disjunct patterns due to the late Miocene formation of intervening deserts (Pokorny et al. 2015).

For those organisms that underwent an ancient vicariance, the interior WHZ isolation from its coastal counterpart was likely complete by about 5 MYA during the Pliocene (though see Methods: Phylogenetics and estimating divergence times). Thus, the interior mesic forest was in place during the duration of the Pleistocene glacial cycles which certainly impacted the distribution of organisms in the region. During the Pleistocene, 100,000 year Milankovitch cycles caused by cyclic variation in the earth's orbit repeatedly buried large portions of the northern hemisphere under ice caps and alpine glaciers for 90,000 years each cycle forcing habitats southward and/or to lower elevations by as much as 1,000 m down-slope (Pielou 1991, Delcourt and Delcourt 1993). The most recent of these glacial periods was probably the most extensive, reaching its maximum about 26.5 - 19 thousand years ago (KYA) and receding from the focal area approximately 15 - 13 KYA years ago (Booth 1987, Yokoyama et al. 2000, Clark et al. 2009). The poorly-dispersing endemic species of the interior WHZ likely weathered these cycles via elevational shifts, too poorly dispersing to track the southern

expansion of habitat. The majority of the interior WHZ is occupied by the Clearwater River Basin which was suggested to be a refuge for denizens of the WHZ during the LGM due to some species of plants such as the Red Alder (*Alnus rubra*) and Pacific Dogwood (*Cornus nuttallii*) that presently remain confined to Clearwater Basin (Daubenmire 1975). One of the predictions associated with the interior WHZ refuge is that endemic denizens of the Greater Clearwater Refugium (GCR) were confined to subrefugia that formed as the WHZ contracted westward and low into valleys by alpine glaciers, lower tree line, and expanded subalpine forests (e.g., Daubenmire 1975, Carstens et al. 2004, Gavin 2009). This has been termed the multiple valley refugia hypothesis (Brunsfeld et al. 2001), or the down-canyon hypothesis (DCH; Brunsfeld and Sullivan 2006). Broadly speaking, this hypothesis posits that populations within the inland WHZ were isolated by geographic features in a collection of refugial compartments within the GCR. The hypothesis most often put forward is that the GCR was compartmentalized into pockets of habitat and isolated populations in separate river canyons (e.g., Brunsfeld and Sullivan 2006). That is, that the isolating topographical features between intervening refugial compartments are the ridges of major basins. Subsequently, after retreat of the LGM populations colonized upslope following the receding alpine glaciers.

The expected phylogenetic patterns for species that were isolated into multiple refugia is multiple reciprocally monophyletic lineages within the interior WHZ refugia (Brunsfeld et al. 2001). A corollary to the DCH is that the refugial populations proximal to Pleistocene glaciation during the LGM would be the source population for genotypes

found in previously glaciated habitats (Brunsfeld et al. 2001). A few studies have found population structure within the inland WHZ which has been attributed to these populations being confined to disparate refugia during the LGM (e.g., Nielson et al. 2001). For example, evidence for structured refugia was found in the plant *Cardamine constancei*, with four documented populations including a widespread low-elevation population and three higher elevation populations confined to different basins (Brunsfeld and Sullivan 2006). Also, the Rocky Mountain Tailed-Frog (*Ascaphus montanus*) was inferred to have inhabited two inland mesic forest refugia, though only one in the Clearwater Basin with the other in the Salmon River Basin (Metzger et al. 2015). However, the few studies that have closely examined poorly dispersing invertebrates in this ecosystem have documented diversification events and allopatric species groups (e.g., the thorn harvestmen *Acuclavella*, Richart and Hedin 2013) or deep population structure (the millipede *Chonaphe armata*, Espíndola et al. 2016).

Another evolutionary phenomenon to consider, based largely on previous work on thorn harvestmen (Richart and Hedin 2013) but also scattered within the GCR literature, is the role of riverine barriers on population structure and diversification. The riverine barrier hypothesis (RBH) posits that major rivers represent barriers to gene flow which can lead to lineage diversification. This hypothesis was originally formulated by Alfred Russell Wallace to explain the distribution of related monkey species on either side of major Amazonian rivers (Wallace 1854). The majority of studies invoking the RBH continues to be from Amazonia (e.g., Boubli et al. 2015, Moraes et al. 2016), and/or is almost always associated with very large continental rivers (e.g., Dong et al. 2013, Satler

and Carstens 2016). When smaller rivers are invoked as the barriers of allopatric speciation it is often for poorly-dispersing SREs (e.g., Kane et al. 1990, Thomas and Hedin 2008, Hedin and McCormack 2017). Within the GCR, rivers have rarely been invoked to explain lineage divergence. The few GCR examples invoking rivers as barriers include the contact zone between two subspecies of the Red-tailed Chipmunk (*Tamias ruficaudus*) with the eastern *T. r. ruficaudus* and western *T. r. simulans* found on either bank down the length of the Lochsa River (Good and Sullivan 2001, Hird et al. 2010). Additionally, within the western subspecies population structure correlated to the North Fork Clearwater and the St. Joe Rivers (Hird et al. 2010). Another example involves the Coeur d'Alene Salamander (*Plethodon idahoensis*) where samples south of the Selway River and north of the Lochsa each formed a clade (Carstens et al. 2004). These clades were recovered as sister, though all of these relationships were inferred to have low support (probably owing in part to the use of a single locus). In other geographical systems, allopatry associated with riverine barriers has been reported in Opiliones (e.g., Thomas and Hedin 2008, Hedin and McCormack 2017). Richart and Hedin (2013) postulated that rivers were associated with population structure and divergence across inland thorn harvestmen species, but this hypothesis was not explicitly tested.

In this research, geographically fine-scale biogeographic inference in the interior WHZ is inferred using a clade with the thorn harvestmen genus *Acuclavella* (Opiliones, Dyspnoi, Ischyropsalidoidea). The thorn harvestmen are among the groups of organisms that show an ancient vicariance with species west of the Cascade Mountains (Richart and

Hedin 2013). Seven species of inland thorn harvestmen species are endemic to the GCR in an array of allopatric populations (Figure 3.1). Thorn harvestmen are strongly associated with perennial water features such as headwater streams, springs, and seeps (Richart and Hedin 2013). In these small perennial headwater streams sticks, branches and other large woody debris are largely immobile (Montgomery and Buffington 1997, May and Gresswell 2003). Thorn harvestmen are typically found within the channel wall under partially submerged debris. Cool stream water wicks up the woody debris, keeping it moist and dark. Thorn harvestmen sit motionless in this habitat, their black bodies remarkably obscured against the dark wet wood. These perennial headwater streams typically have a continuous canopy cover, thus little insolation which maintains a cool temperature and diminishes evaporation. In the PNW, larger streams and rivers tend to retract during the summer. The cool waters of the rivers are bracketed by hot and dry beds of cobble and bedrock that may deter dispersal attempts. Many of the *Acuclavella* populations sampled along the Bitterroot Range were from areas covered by alpine glaciers during the LGM (Figure 3.1; Fountain et al. 2007). Thus far, only a minimal effort has been made to find *Acuclavella* in areas that would indicate a northward expansion into areas covered by the Cordilleran Ice Sheet during the LGM (Figure 3.1). Here, the *a priori* prediction is that populations weathered the LGM in refugia that were down-slope from current suitable habitat, then expanded upslope tracking the shift of the WHZ after retreat of alpine glaciers. The goal of this research is to examine the combined effect of Pleistocene climate oscillations and riverine barriers on the GCR. Studies invoking both of these phylogeographic patterns are few and are comparatively vast in

their geographic scope (e.g., Dong et al. 2013, Herman and Bouzat 2016, Hedin and McCormack 2017).

II. MATERIALS AND METHODS

(1) Taxon sampling

Details of specimen collecting and curation are outlined elsewhere (Richart and Hedin 2013, Chapter 2). Here, 82 individuals from 45 northern Idaho or adjacent Montana (Bitterroot Mountains) localities were preserved in 100% EtOH and stored at -80°C before being processed for ddRADseq analyses (Appendix B). All specimens are deposited in the San Diego State University (SDSU) Museum of Biodiversity Terrestrial Arthropod Collection (TAC) molecular Opiliones (OP) collection unless otherwise noted.

(2) Molecular data collection, ddRAD-sequencing, and data processing

Details of DNA extraction, quantification, ddRADseq processing, and raw sequence processing as in Chapter 2. A brief synopsis of these methods is provided here, emphasizing parameters relevant to this research. The two restriction-site associated enzyme double-digest RADseq (ddRADseq) protocol of Peterson (et al. 2012) was used to sample the genome for single nucleotide polymorphisms (SNPs). Raw sequence reads were assembled using pyRAD 3.0.5 (Eaton and Ree 2013, Eaton 2014). Allelic diversity and heterozygosity are important in the inference of demographic histories, and

heterozygosity is reduced under stringent similarity thresholds (Ilut et al. 2014). Thus, similarity thresholds were set to 95% which has been suggested to be liberal enough to not produce underinflated values of ancestral θ and τ (Harvey et al. 2015). Analyses here were conducted on ddRADseq data sets that include curation from all inland thorn harvestmen (Appendix C).

(3) Phylogenetics and estimating divergence times

In order to test demographic responses to the LGM it was necessary to assure that lineages analyzed had diversified well before this time (at least 20,000 year ago). This assessment was undertaken via a molecular clock using multiple lines of evidence and a two-step phylogenetic process. The biogeographic split between the coastal and inland WHZ was used to calibrate divergence times within northern Idaho. Geologic evidence is not ideal for calibrating divergence analyses (Heads 2011). However, few options for calibrating the molecular clock exist, for fossils are not known for *Acuclavella* and our understanding of the rate of evolution for thousands of anonymous SNPs is not well understood.

In a synthesis of the comparative phylogeography of northwestern North America, Brunsfeld and authors cite the biogeographic split between the mesic coastal and inland forests as 5 - 2 MYA (Brunsfeld et al. 2001). This date comes from Graham (1999), who mentions the "initial appearance of the high Cascade Mountains" at about 5 MYA, the Palouse Prairie formed by 3 MYA, and the Great Basin dominated by *Artemisia* at ~3 - 2 MYA. This date of 5 - 2 MYA has become the standard in PNW

biogeography (e.g., Demboski and Sullivan 2003, Carstens et al. 2004, Brunsfeld et al. 2007, Carstens and Richards 2007, Shafer et al. 2010, Derkarabetian et al. 2016, Smith et al. 2018). However, Wilson and Pitts (2010) point out a lack of consensus among paleobiologists / paleobiogeographers, and recommend that many lines of independent evidence be used to date such events (see also Chaney 1938). For example, Mitchell and Montgomery (2006) state, "...the evidence appears conclusive that a significant climate change occurred in south- and central-eastern Washington from ~16 to 8 MY ago...". Therefore, a literature review was applied starting with Wilson and Pitts (2010) and Graham (1999). Numerous and often independent geological and paleontological lines of evidence were used to more appropriately infer the age of this biogeographical event. On average, this evidence suggests a xeric Columbia Plateau about 7 MYA, perhaps as a transition creeping eastward that spanned between 12 - 2 MYA (Table 3.1).

The organismal evidence of a xeric Columbia Basin suggests dates from about 8 - 2 MYA. Cypress forests indicative of warm-temperate swamps and hardwood forests persisted in eastern Washington until about 8 MYA and in Idaho until about 12 MYA (Leopold and Denton 1987). Columbia Basin flora transitioned from wet-loving hardwoods, redwoods (*Sequoia*), and cypress (*Taxodium*) in the early Miocene, to xeric forests of oaks (*Quercus* spp.) and madrona (*Arbutus*) in the late Miocene (Chaney 1938). Open savanna and prairie environments likely existed in the Colombia Plateau as early as the mid-Miocene, as evidenced by the diversity of fossil mammals whose modern relatives live in these habitats today; perrisodactyls such as horses and rhinoceros, and artiodactyls such as camels and merycoidontoids (Chaney and Axelrod 1959, Webb

1977). In contrast, pollen and fossil data alike suggest a rich deciduous forest vegetation during the Miocene in this area, with direct evidence of savannah habitats rare (Chaney and Axelrod 1959, Leopold and Denton 1987). Grass and Asteraceae are sporadically abundant as part of a cyclic succession, with peaks of grass pollen during the Miocene associated with an increased abundance of the Hagerman Horse (*Equus simplicidens*). During the Middle Miocene warm-temperate forests moved distinctly toward a cold-temperate fauna, and by the end of the Miocene the steppe-adapted fauna had expanded to nearly its present range (Webb 1977). In Idaho, pollen data show a transition from mesic forests to the development of local grasslands such as the Palouse Prairie and inland steppe habitat at the end of the Pliocene only about 3 MYA (Leopold and Denton 1987, Graham 1999). This trend to ever more dry-tolerate vegetation continued, and the sagebrush (*Artemisia*) dominated scrub that currently occupies eastern Washington was in place by 3 - 2 MYA (Leopold and Denton 1987, Graham 1999).

A synopsis of the geological evidence suggests that Cascade orogeny and Columbia Basin xerification occurred about 8 MYA. Cascade orogeny occurred at least more recently than ~16 MYA, as evidenced by uplifted Grande Ronde Basalt flows found high in the southern Cascades (Smith 1903, Hammond 1979, Mitchell and Montgomery 2006). The Cascade Range was probably already rising during the Miocene, with Late Miocene deposits of eroded non-marine sediment on both sides of the Cascades suggesting that uplift was underway by 15 MYA (Gresens 1987, McKee 1972, Smiley 1963). Fission track radiometric dating of uranium-bearing minerals exposed during uplift suggest a Cascade Orogeny from 12 - 5 MYA (Reiners et al. 2002). Warping and

uplift of Ellensburg and Simcoe basalt deposits on the east side of the Cascades suggest the Neogene uplift occurred in this area between 10 - 4.5 MYA (Hammond 1979). From 12 - 8 MYA a trend in eastern Washington of a decrease in oxygen isotope ratios in ancient precipitation-derived water trapped in clay mineral deposits is indicative of decreasing precipitation presumably due to the Cascade orogenic barrier to coastal moisture (Takeuchi and Larson 2005).

Another potentially confounding issue may manifest when using geological evidence to calibrate divergence analyses, which is perhaps more problematic when evaluating SRE taxa and relatively recent evolutionary events, as is done here. Thorn harvestmen populations from coastal and inland WHZ populations are reciprocally monophyletic with deep divergence, which is consistent with the predictions of the ancient vicariance hypothesis (Brunsfeld et al. 2001, Carstens et al. 2005a, Richart and Hedin 2013). However, these reciprocally monophyletic groups each subtend allopatric arrays of species - two coastal and seven inland (Shear 1986, Richart and Hedin 2013, Chapter 2). At some point in time before the isolation of the interior WHZ, thorn harvestmen must have dispersed from or have a range encompassing the Northern Rocky Mountains and coastal areas. How did this happen? Was it a rapid dispersal event that happened just before the isolation of the interior WHZ? There is no evidence for rapid dispersal in the genus. Did they exist as an array of allopatric species as is seen in modern species? If so the root node of the genus is likely older than the vicariance caused by the Columbia Basin xerification. This is a problem associated with the 'common vicariance, rare dispersal' syndrome seen in many SRE taxa (e.g., Hedin and McCormack 2017). In

the opposite direction, SREs are more prone to extinction than widely-distributed organisms (Harvey et al. 2011). Thus, it is possible that one of the lineages that diversified in response to the abiotic vicariance used for calibration is extinct, and the calibration date is being assigned to a more recent event (e.g., dispersal), resulting in estimated divergence times for related nodes to be excessively ancient (Heads 2011). Because of these reasons: 1) a much broader range of dates associated with Columbia Basin xerification than has been appreciated, and 2) problems associated with 'common vicariance, rare dispersal' SREs, the molecular clock is also used to infer lineage ages.

In most instances, with exceptions such as speciation with gene flow and when loci are fixed at the time of speciation, gene divergence predates species divergence, and genetic divergence estimates based on gene trees and concatenated matrices are overestimated (older) compared to divergences based on species trees (McCormack et al. 2011). Thus, in order to provide a prior for the rate of evolution for a chronogram estimated from ddRADseq data, a species tree analysis was conducted in *BEAST 2.4.0 (Bouckaert et al. 2014) on a two-locus data set of previously published sequences (Table 3.2; Hedin et al. 2010, Richart and Hedin 2013) from six putative *Acuclavella* lineages (two coastal, four inland). The loci comprise a 1,224 bp fragment of cytochrome oxidase I (COI) and elongation factor 1-alpha (EF1-a) which includes 690 bp of exon data bracketing an 82 bp (alignment length) intron. Partitions and substitution models were chosen with PartitionFinder 1.1.1 (Lanfear et al. 2012), allowing for partitions by codons and coding regions, using linked branch lengths, BIC model selection, a greedy search algorithm, and BEAST models of evolution. All partitions used a relaxed clock log

normal model, and resulting analyses reject a strict clock. For the COI partition a clock rate was specified based on the 3.54% divergence rate calibrated for arthropods (Papadopoulou et al. 2010). The ucldMean was set to mean: 0.0177 with a normal distribution and a sigma: set to 1.8 to include the range of dates suggested by geological and paleontological lines of evidence. Six different runs of 100 million generation sampling every 50,000 were combined in LogCombiner 2.4.0 and TreeAnnotator 2.3.2 (Rambaut and Drummond 2014) using a 10% burn-in for each run. Combined log files were visualized in Tracer 1.6.0 (Rambaut et al. 2014) and ESS values were consistently high. The posteriors from this analysis were used as the priors for the ddRADseq panel.

Analysis of 82 inland thorn harvestmen samples was conducted on a pyRAD output generated with a minimum coverage (MinCov) set to 22 (Appendix C). The unpartitioned concatenated alignment of entire reads included 4525 loci and 430,177 nucleotides with 57.3% missing data. This matrix was analyzed using BEAST 2.4.5 (Drummond and Rambaut 2007, Bouckaert et al. 2014). The best-fit model of evolution was selected by PartitionFinder 1.1.1 (Lanfear et al. 2012) using the BIC criterion with linked branch lengths and a greedy search algorithm. Six analyses of 25 million generations sampling every 1000 used a log normal relaxed clock with the root node calibrated using the estimated divergence times resulting from the *BEAST analysis (mean: 2.2, confidence interval 3.5 - 1.0 MYA) and standard deviation of 0.3 MY. One of the six runs failed to converge, the remaining five runs were combined after burn-in of 40 - 80% and subsampling from 10,000 - 20,000. Runs were combined as previous, and all ESS values were greater than 200 with the exception of the inverse of

proportionInvariant (12) and gammaShape (15). However, the parameters of most interest, ucldMean and ucldStdev were well sampled (ESS values of 2686 and 1935 respectively) and the resulting phylogeny has an identical topology as previous BEAST analyses (chapter 2) with similar support values.

(4) Riverine barriers

For the amount of riverine stream-flow a key concept is the size of the watershed or basin (USGS 2018). Basin size was calculated using USGS StreamStats v.4 (https://streamstats.usgs.gov/ss/) as the area that drains to a point on a stream (DRNAREA). The stream point was determined as the mouth of the largest stream separating the sister localities or clades on the inland BEAST time tree described above. This was almost always straightforward. On two occasions the sister localities were on opposite sides of a pass without an intervening basin; for these, basin size was recorded as 0.0 km^2. On three occasions, where a recent colonization event across a large river was apparent, the mouth was not chosen as the stream point, but rather the nearest point along the river between the two clades or localities being considered. This is justified by the very shallow node depths suggesting a post-LGM expansion. For these three instances, the large rivers were only considered for the shallowest relevant node in the time tree, with subsequent deeper clades considering intervening streams and not these within-clade large rivers. A Pearson correlation analyses assessed the significance of the relationship between phylogenetic node depth (MYA) and basin size (km^2) with the p-value calculated online (https://www.socscistatistics.com/pvalues/pearsondistribution.aspx).

(5) Expansion out of refugia - the down-canyon hypothesis

The often continental-scale founder-effect phenomenon of genetically depauperate populations that have expanded out of genetically rich refugial areas can also emerge at smaller geographical scales (e.g. Martínez-Solano et al. 2006). Many of the *Acuclavella* populations sampled along the Bitterroot Range were from areas covered by alpine glaciers during the LGM (Figure 3.1; Fountain et al. 2007, http://glaciers.us/). However, two of the species (*A. cosmetoides* and Hoodoo Mountains) do not have populations that were near alpine glaciers. Thus, as predicted by the DCH, the *a priori* expectations are that species with populations within or adjacent to areas covered by glaciers will show a decrease in genetic diversity at higher elevations. Here, Pearson correlation analyses are used to evaluate the relationship between genetic diversity and elevation. Heterozygosity is used as the metric for genetic diversity, using the output from All Inland *Acuclavella* pyRAD curation (Appendix C). Since *A. sheari* is only sampled from two populations, and this species is found in a small isolated patch of habitat on the southern edge of thorn harvestmen distribution where it has probably undergone a severe bottleneck, these samples were not included in DCH analyses.

RESULTS

(1) Phylogenetics and estimating divergence times

The *BEAST analysis calibrated with the molecular clock and with sigma set to include the range of dates suggested by geological and paleontological lines of evidence

estimated the timing of divergence between coastal and inland thorn harvestmen species to be 7.7 MYA with a 95% HPD interval of 11.5 - 4.4 MYA (Figure 3.2). The mean of 7.7 MYA is very similar to the approximate average of the dates inferred for this split using multiple independent lines of geological and paleontological evidence (Table 3.1). This suggests that the geological calibration of 5 - 2 MYA used by biogeographers to date this split is both too recent and too narrow. The root node for inland samples is inferred as 2.2 MYA (Figure 3.2) with a 95% confidence interval of 3.5 - 1.0 MYA. As expected given the prior, the SNP BEAST time tree inferred the inland thorn harvestmen root node at 2.21 MYA (Figure 3.3). Although the 95% confidence intervals inferred by divergence time estimation are very large (Figure 3.3), the most likely species divergence are all equal to or greater than 920 KYA, thus much older than the LGM (about 26.5 KYA).

An interesting endeavor would be to summarize divergence estimates inferred from a wide range of organisms showing an ancient vicariance with coastal populations using substitution rates inferred for specific loci (e.g. "known" COI rates). From this a distribution of dates could provide an average and confidence intervals associated with this ancient vicariance. This data might better inform priors for future analyses than the probably too specific and too recent 5 - 2 MYA. For example, using mtDNA, the divergence between the Van Dyke (*Plethodon vandykei*) and the Coeur d'Alene Salamander (*P. idahoensis*) was calculated to be 3.75 MYA with a 95% confidence interval encompassing 8.2 and 1.31 MYA (Carstens et al. 2004). Here, the divergence between coastal and inland populations was inferred to be 7.7 MYA (95% c.i. 11.5 - 4.4

MYA). These numbers compiled for a ~30 of the organisms showing an ancient vicariance would perhaps return a more meaningful calibration for future studies and would allow for the detection of inferred dates that significantly deviate from the distribution.

(2) Riverine barriers

A significant positive correlation was found between node depth and river basin size (Figure 3.4, Table 3.3). This strongly suggests that riverine barriers are playing a significant role in the diversification and organization of population structure within inland thorn harvestmen. This trend is supported even with the inclusion of three obvious recent dispersals across major rivers (Figure 3.4). The data series is perhaps curvilinear, with species level divergences apparently on a different trend line than within-species population structure. This infection point may be species-level signal, analogous to the inflection in a lineage-through-time plot when sampling has crossed from genealogical branching within species to diversification branching across species (e.g., Pons et al. 2006, Zhang et al. 2013).

The normal model for allopatric speciation is that a widespread species is bisected by the formation of a geographical feature which reduces gene flow and leads to speciation. For example, the Cascade Mountain orogeny dividing the coastal and inland WHZ. Here however, the bisection of a previously widespread species via riverine barriers is not likely. The Rocky Mountains formed 80 - 55 MYA as part of the Laramide Orogeny which also formed the Sierra Madre Oriental in Mexico (English and Johnston

2004). As these mountains rose, streams and rivers entrenched themselves into canyons. The major rivers of this region were in place long before the periodic Pleistocene glaciations of the last couple million years and long before the inferred age of the most recent common ancestor of extant thorn harvestmen. Thus, these rivers did not form within the range of a thorn harvestmen species causing vicariance. Rather, these rivers likely greatly inhibited gene flow once populations were established via dispersal. It is clear that *Acuclavella* species are able to disperse across large rivers as exemplified by the northward expansion of *A. shoshone* across the St. Joe River multiple times.

(3) Expansion out of refugia - the down-canyon hypothesis

As predicted under the DCH, there is a significant ($\alpha = 0.05$) decrease in heterozygosity when evaluating species whose range includes areas that contained or were adjacent to alpine glaciers (Figure 3.5 A), but not for species whose ranges were apart from glacier fields during the LGM (Figure 3.5 B). The strongest correlation was found when considering all thorn harvestmen samples regardless of whether or not they were adjacent to alpine glaciers (Figure 3.6). The only species to show a significant negative correlation between altitude and genetic diversity was *A. shoshone*, which is perhaps not surprising given the number of populations sampled that were within LGM glacier fields (Table 3.4, Figure 3.5). However, the trend line is negative for all thorn harvestmen species, though usually not significantly so (Table 3.4, Figure 3.5). *Acuclavella* cf. *quattuor* shows a strong negative correlation with an r-value of -0.4685, though the strength of this signal is not significant at $\alpha = 0.05$ (p-value 0.1546; Table

3.4); an increase of sample size would likely improve this significance value. The weakest correlations are seen in the two species that did not border alpine glaciers fields during the LGM (Table 3.4): *A. cosmetoides*, r = -0.0390, p = 0.8820; Hoodoo Mountains r = -0.2669, p = 0.5642. The lack of support within the *A. merickeli* and *A. quattuor* clade (r = -0.1615, p = 0.4741) was initially surprising, for these populations encompassed the broadest elevational range (Figure 3.5). Perhaps this is due to none of the sampled localities actually predicted to have occurred in areas covered by glaciers during the LGM (Figure 3.1). Another factor to consider is that the lowest populations may now be in marginal habitats as the WHZ tracks upslope, similar to the bottlenecks resulting in loss the of genetic diversity seen in populations and species at the trailing edge of polar expansions (e.g., Nicastro et al. 2013). Values for heterozygosity and locality elevation can be seen in Appendix H.

Phylogenetic correction analyses applied to the evaluation of genetic diversity tied to allelic diversity cannot be meaningfully applied here, both because the two species away from alpine glaciers form a clade and also because of how few terminals are in the phylogeny. The problem of too few replicates is likely to hold for the evaluation of any species group within the interior WHZ, and testing this hypothesis across a suite of unrelated poorly-dispersing SREs endemic to the inland WHZ would allow for better evaluation of the robustness of inferred patterns. Additionally, it should be noted that specimen collections were not made with testing the assumptions of the DCH under consideration. Some of the species considered (especially Hoodoo Mountain) were sampled at a very narrow elevational distribution (Figure 3.5). Future evaluations should

look for bottlenecks at the lowest margins of WHZ habitat as well as comparisons across a larger suite of poorly-dispersing SRE taxa.

(4) Biogeographical and phylogeographical conclusions and future direction

Both riverine barriers and glacial refugia are invoked to explain the biogeographical and phylogeographical patterns seen in inland thorn harvestmen species. Many of the species and populations of inland thorn harvestmen are confined to the interfluve between adjacent rivers, corroborating the riverine barrier hypotheses. Lineages that occur along the Bitterroot Mountains, which were covered by alpine glaciers during the LGM, show decreasing genetic diversity with increasing elevation, corroborating the DCH. Few studies have indicated that both of these mechanisms have influenced the evolutionary history of their system (e.g., Cazé et al. 2016, Hedin and McCormack 2017). There are three instances of thorn harvestmen species dispersing across significant riverine barriers, indicating their potential for such dispersal events, and competitive exclusion likely plays a role in maintaining rivers as species boundaries (e.g., Bohoussou et al. 2015). For example, Mediterranean trapdoor spiders *Ummidia* are likely able to balloon as young juveniles, as evidenced by observations of congeners ballooning and apparent dispersal to oceanic islands (Coyle 1983, Fisher et al. 2014). Despite this ability, strong population structure and vicariance prevails (Opatova et al. 2016). Thus, inland thorn harvestmen may be an ideal system upon which to test hypotheses of competitive exclusion (Hardin 1960, Case et al. 2005): 1) Can complete competitors co-exist? 2) Once on the other side, how much gene flow continues from the

original source population? 3) What is the extent and density of populations that have crossed riverine barriers? 4) How does their genetic diversity compare with adjacent populations of allopatric lineages? Future comparative phylogeography across poorly-dispersing SREs endemic to the inland forest could advance our understanding of the role these barriers have had on this community (*sensu* Satler and Carstens 2016).

Interestingly, the hypotheses that diversification was driven by riverine barriers (Wallace 1854) or by isolation into disparate climatic refugia (Haffer 1969) were both developed in the Amazon Basin housing the largest tropical rainforest in the world. Here, patterns from these processes are detected within the much smaller Clearwater Basin, housing in the largest inland temperate rainforest in the world (Metzger et al. 2015). Wallace first noticed that rivers were barriers with monkeys, and Haffer's hypothesized refugia with birds. Both of these observations were made possible because of the grand scale of the diversity and geography under consideration. In comparison, these phenomenon are detected here because of the fine-scale biogeographic inference possible when using poorly-dispersing SREs. The next logical step is comparative phylogeography using multiple poorly-dispersing SREs to better understand the role of these rivers as drivers of speciation. Concordance across different faunal elements would offer strong support of the riverine barrier hypothesis, as has been done for large Amazonian rivers (e.g., Boubli et al. 2015, Moraes et al. 2016). That the results presented in this research find most of their closest comparisons to research done in the tropics reinforces the notion that litter-dwelling organisms are the poor-person's rainforest (e.g., Richart and Hedin 2013).

Usually northward expansion out of Pleistocene refugia is invoked to explain patterns on a continental scale. Here, expansion to higher elevations is detected multiple times within a single mountain range. Given this, it is interesting to contemplate the extent that other phenomenon associated with post-Pleistocene climate change could be uncovered at this high resolution. Refugia typically house higher biodiversity as well as high genetic diversity within species. Do other inland WHZ endemic SREs such as terrestrial gastropods, millipedes, and other harvestmen species occur at the same elevations that house higher genetic diversity within the thorn harvestmen? Another phenomenon of climate-induced polar range expansion is a trailing edge, where populations become isolated, suffer severe bottlenecks, and have high extinction rates (Hewitt 2000, Gavin 2009). Do these processes manifest at the lowest levels within these canyons such as along the xeric to mesic transition on the southern slope of the Salmon River?

Currently, the major rivers that currently have their headwaters below the crest of the Bitterroot Mountains housed alpine glaciers during the LGM. Consequently, perhaps appropriate habitat did not occur in headwater areas where dispersal across rivers would have been more easily accomplished. Within many headwater stream basins a continuous canopy cover provides a steady supply of woody debris that regularly spans the channel, providing abundant thorn harvestmen microhabitat. Such areas are unlikely to impede dispersal from one side to the other. Downstream, as streams coalesce forming larger channels and rivers, gene flow is commensurately more likely to occur parallel to the watercourses rather than across it. Thus, although divergence events appear correlated

with riverine barriers, it is possible they were reinforced during the LGM due to isolation into separate refugia. Future research should assess whether different species come into contact in headwater regions and exchange genetic material (e.g., Weir et al. 2015).

For these harvestmen, glacial refugia during the LGM were compartmentalized by riverine barriers. This suggests a more complex GCR, a model system for the study of glacial refugia, than has previously been put forward. This highlights the exceptional utility of using poorly-dispersing SREs taxa for biogeographical and phylogeographical inquiries. Most authors have suggested that the GCR was compartmentalized into isolated basins, and for frogs and understory plants that very well may be how they were structured. Thorn harvestmen however likely do not have the dispersal capabilities of any organism that has been used to evaluate the evolutionary history of the GCR. Alternative basin refugia likely structured historic populations of thorn harvestmen, as appears to be the case for the ridge separating the NFCR from the Lochsa, but for all other species it is the rivers that are the landscape feature associated with isolation. What allows for this insight is the use of poorly-dispersing short-range endemics for biogeography and phylogeography inquiries.

DISCUSSION

Interior Western Hemlock Zone biogeography

The results reported here provide important insight into the evolutionary history of a model system for recent biogeography. With this picture we can put some of the

previous empirical research in this system into better context. The multiple valley refugia hypothesis, and its prediction of multiple reciprocally monophyletic lineages within the GCR was first put forward by Brunsfeld and colleagues in 2001. However, though previous empirical research has uncovered population structure within the GCR, their results have not resulted in well-supported and reciprocally monophyletic groups (e.g., Nielson et al. 2001, Carstens et al. 2004, Brunsfeld and Sullivan 2006, Hird et al. 2010, Metzger et al. 2015). Other studies have inferred a single refugial population during the LGM (Tsai and Carstens 2013). Phylogeographic analyses of mtDNA sequences has suggested that the Idaho Giant Salamander (*Dicamptodon aterrimus*) took refuge from the LGM south of the Salmon River (Carstens et al. 2005b). Similarly, research on the Rocky Mountain Tailed Frog has inferred that other glacial refugia existed south of the study area presented here in the East Fork of the South Fork Salmon River (Nielson et al. 2006) or higher in the Salmon River drainage (Metzger et al. 2015). The remnant forests in these mountainous uplands should be the target of future arthropod surveys. The discovery of Shear's Thorn Harvestmen (*Acuclavella sheari*) streamside in Grand Fir (*Abies grandis*) forests south of the Salmon River (Richart and Hedin 2013), along with the unique amphibian populations, suggest that this area likely accommodates many isolated short-range endemic species as is expected at the trailing edge of a habitat shift to higher latitudes.

The Coeur d'Alene Salamander (*Plethodon idahoensis*) has been shown to have strong population structure associated with the Lochsa and Selway Rivers (Carstens et al. 2004). Unfortunately in the *Plethodon* study samples were only analyzed north of the

Lochsa or south of the Selway, and not the intervening habitat between these two rivers, thus the alternative role of rivers vs. basins (or simply geographic distance) as a causal mechanisms for this structure cannot be evaluated. This research suggests that sampling the interfluve between the Lochsa and Selway rivers may be an important key to understanding the evolutionary history of the denizens of the interior WHZ. Others have hypothesized that different river basins may house populations that diverged in disparate refugia due to genetic drift and limited gene flow (e.g., Brunsfeld et al. 2001). This may indeed be true for organisms that disperse relatively easily across these northern Idaho rivers, but in thorn harvestmen rivers themselves limit gene flow, likely resulting in diversification, rather than the ridges that partition different basins. Sister lineages of *Hemphillia* jumping-slugs are also inferred to have diversified within the mesic forests of northern Idaho, and both of these species (*H. camelus* and *H. skadei*) show genetic structure within the inland WHZ (Lucid et al. 2018). However, these two species were not surveyed or sampled south of the North Fork Clearwater River, and a third endemic jumping-slug (*H. danielsi*) was not surveyed (or sampled) north of the Selway / Middle Fork Clearwater River, and not surveyed/sampled south of the South Fork Clearwater, so evaluation of riverine barriers on phylogenetic structure of inland jumping slugs and comparison to *Acuclavella* is not yet possible.

Many species that were likely confined to refugia in the GCR were able to greatly expand their ranges northward (e.g., Carstens et al. 2004, Metzger et al. 2015). Other species, especially those endemic to the inland mesic forest, do not have current ranges that expand north into areas covered by glaciers during the LGM, even though it is likely

suitable habitat (e.g., Daubenmire 1975, Brunsfeld and Sullivan 2006, Gavin 2009, Björk 2010, Shafer et al. 2010). Understory plants endemic to, or with disjunct populations within the GCR, have had more difficulty expanding northward into glaciated regions than have their coastal counterparts, with animal-dispersed species having more easily expanded north than plants with other dispersal mechanisms (Gavin 2009). Likewise, a review of glacial expansion in the PNW found a positive and significant relationship between occupancy of multiple refugia and dispersal ability (Shafer et al. 2010). Thorn harvestmen are not currently known to extend into areas once covered by the continental ice sheet though surveys north of currently known localities are not extensive (Richart and Hedin 2013, Figure 3.1). There are however organisms such as terrestrial gastropods that are thought to be poorly dispersing that have greatly expanded northward into British Columbia (Burke 2013) such as the Magnum Mantleslug (*Magnipelta mycophaga*), Pale Jumping-Slug (*Hemphillia camelus*), and the Smoky Taildropper (*Prophysaon humile*). How species richness of poorly-dispersing litter dwellers vary between glaciated and unglaciated regions of the interior WHZ is an intriguing question. Perhaps in some of these poorly-dispersing lineages, gene flow from refugial populations is so low as to be non-existent, and at the expanding margin genetic diversity is so low that inbreeding depression plays a role in slowing northward expansion.

Poorly Dispersing Terrestrial Invertebrates as Model-Organisms for Biogeography and Phylogeography

A suite of terrestrial organisms possess a combination of traits making them ideal for studies of biogeography. They are among the earliest colonizers of terrestrial landscapes, thus suitable for testing ancient hypotheses. They are very poorly dispersing, likely much less vagile than a vertebrate or plant representatives from the landscape of interest. They often have strict habitat requirements, usually seeking cool environs such as rock outcroppings, adjacent seeps and headwater streams, or under large woody debris. The combination of a limited ability to disperse with a specialized microhabitat requirement can lead to pervasive population genetic subdivision and evolutionary divergence (Keith and Hedin 2012). This collection of organisms includes representatives in many disparate clades, including for example millipedes, harvestmen, and mygalamorph spiders. Of vertebrate taxa, salamanders most often possess these traits (e.g., Slatkin 1981, Larson et al. 1984). Comparative phylogeography of lizards and frogs endemic to the Australian wet tropics found both congruent and unique signatures of different Pleistocene refugia, whereas a land snail was inferred to have occupied all of the Pleistocene refugia indicated by different vertebrate species and reflected finer scale evolutionary processes (Hugall et al. 2002). Faunal studies have shown that flightless insects have much higher levels of regional and subregional endemism that do vertebrates (Yeates et al. 2002). Some denizens of the northern Idaho WHZ are endemic to single basins within this geographic zone (e.g., the terrestrial gastropods *Securicauda hermani* Leonard et al. 2011, *Kootenaia burkei* Leonard et al. 2003), and this is also true for some *Acuclavella* species (Richart and Hedin 2013). The diversity of poorly-dispersing SREs such as terrestrial gastropods and flightless insects is a strong predictor of vertebrate

diversity but not vice versa, highlighting the importance of SREs as surrogates for conservation prioritization (Moritz et al. 2001). Phylogeographic analyses of the Rocky Mountain Tailed-Frog (*Ascaphus montanus*) within the inland mesic forest found that the majority of sampled populations, and all of the populations encompassing the area discussed in this research, show a distinctive lack of genetic structure and minimal variation (Nielson et al. 2001). This is in stark contrast to the thorn harvestmen, although both taxa are obligate denizens of headwater streams.

Poorly-dispersing SRE taxa are very useful for many spatial and temporal levels of terrestrial biogeography questions (Figure 3.7). However, they are not well-suited for all biogeographic questions. As we have seen with SREs still confined to the GCR, they often disperse too slowly to test hypotheses regarding the rapid colonization of an area, especially on a continental scale (Lester et al. 2007). For example, when studying the colonization of tundra and taiga after the LGM (Figure 3.7 a). Another important exception are questions under the purview of landscape genetics, when the evolutionary events are so recent that a better disperser is more likely to be informative. For example, when evaluating the efficacy of a restoration corridor on connecting two habitat patches (Figure 3.7 b), a landscape geneticist would want to use organisms that disperse farther than SRE organisms, unless the hypotheses to be tested were extremely localized.

In a comparative biogeographic analysis of the WHZ, Carstens and collaborators (2005a) found that more vagile mammals and plants show a recent dispersal from coastal to inland forests, whereas the poorly dispersing amphibian species show an ancient vicariance. Though some signal of structure was uncovered, none of the amphibians

showed reciprocally monophyletic groups as predicted by the multiple valley refugia hypothesis. The only species thus far inferred to show such phylogenetic patterns are the thorn harvestmen here and perhaps the xystodesmid millipede *Chonaphe armata* (Espíndola et al. 2016, supplemental material). When it comes to showing structure and diversification at finer spatial scales and for more recent evolutionary events, poorly-dispersing invertebrate SREs are to amphibians what amphibians are to voles and willows. Evidence suggests that glacial refugia (areas of stable habitat through Pleistocene glacial cycles) predict areas of genetic diversity (Carnaval et al. 2009). Some of the SRE taxa still endemic to the areas that provided refuge from the LGM must have difficulty maintaining viable populations under current climatic conditions, especially at the trailing edge (Gavin 2009, Barbosa et al. 2017). The inclusion of multiple SRE taxa in comparative phylogeography would allow for a historical template that would allow inquiry into vertebrate evolutionary history. For example, why were some refugia sufficient for only some species to persist and why does this vary across microrefugia?

Postglacial expansion in the Northern Hemisphere typically follows a pattern of decreased genetic diversity with increasing latitude, as populations expand northwards following retreating glacial ice sheets. Here, testing biogeographic and phylogeographic hypotheses using poorly dispersing terrestrial invertebrates allowed resolution within basins of populations expanding to higher elevations following the retreat of alpine glaciers. By using a poorly-dispersing SRE, a clearer picture has been painted of the structure and location within a model system of compartmentalized glacial refugium. This was accomplished in large part simply by choosing an organism more appropriate

for the questions at hand. Organisms with these characteristics will generally be able to survive and persist in small pockets of habitat compared to organisms farther up the trophic level. Unlike vertebrates, they are less likely to be capable of moving from pocket to pocket, but must persist where they are. If you want to study isolation and diversification caused by Pleistocene glaciations, don't choose winged birds as your study organism (e.g., Klicka and Zink 1997). Biogeographers interested in the geographic history of a region should almost never choose birds, or winged insects, or organisms that disseminate via Aeolian dispersal such as conifers, ballooning spiders, and gilled mushrooms; the biogeographer needs to choose an appropriate system, preferable a suite of poorly dispersing organisms such as millipedes, short-legged harvestmen, springtails, and gastropods.

CONCLUSIONS

The inland temperate rainforest has been a model of recent biogeography with a complex evolutionary history. By using genomic scale data on the poorly-dispersing *Acuclavella* species endemic to this inland forest, this complexity has become focused with unprecedented resolution. This forest was heavily impacted by the LGM by alpine glaciers and downward habitat shifts. For the first time, this research shows that riverine barrier played a large role in structuring populations and species within this forest. Further, this research provides the best evidence to date for the down-canyon hypothesis. The combined effects of rivers and glaciers paints a picture of a complex Pleistocene

glacial refugia system during the LGM. This relatively high-resolution inference of past evolutionary events within an established biogeographic model system points towards the undervalued importance of using poorly-dispersing short-range endemic taxa for biogeography and phylogeography.

REFERENCES

Barbosa, S., J. Paupério, J. S. Herman, C. M. Ferreira, R. Pita, H. M. Vale-Gonçalves, J. A. Cabral, J. A. Garrido-García, R. C. Soriguer, P. Beja, A. Mira, et al. (2017). Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole. Molecular Ecology 26:951–967.

Barr, C. B. (2011). *Bryelmis* Barr (Coleoptera: Elmidae: Elminae), a New Genus of Riffle Beetle with Three New Species from the Pacific Northwest, U.S.A. The Coleopterists Bulletin 65:197–212.

Bidegaray-Batista, L., A. Sánchez-gracia, G. Santulli, L. Maiorano, A. Guisan, A. P. Vogler, and M. A. Arnedo (2016). Imprints of multiple glacial refugia in the Pyrenees revealed by phylogeography and palaeodistribution modelling of an endemic spider. Molecular Ecology 25:2046–2064.

Björk, C. R. (2010). Distribution patterns of disjunct and endemic vascular plants in the interior wetbelt of northwest North America. This paper is one of a selection of papers published as part of the special Schofield Gedenkschrift. Botany 88:409–428.

Bohoussou, K. H., R. Cornette, B. Akpatou, M. Colyn, J. Kerbis Peterhans, J. Kennis, R. Šumbera, E. Verheyen, E. N'Goran, P. Katuala, and V. Nicolas (2015). The phylogeography of the rodent genus *Malacomys* suggests multiple Afrotropical Pleistocene lowland forest refugia. Journal of Biogeography 42:2049–2061.

Booth, D. B. (1987). Timing and processes of deglaciation along the southern margin of the Cordilleran ice sheet. In North America and Adjacent Oceans During the Last Deglaciation (W. F. Ruddiman and H. E. Wright, Editors). Geological Society of America, Boulder, Colorado 80301, pp. 71–90.

Boubli, J. P., C. Ribas, J. W. Lynch Alfaro, M. E. Alfaro, M. N. F. da Silva, G. M. Pinho, and I. P. Farias (2015). Spatial and temporal patterns of diversification on the Amazon: A test of the riverine hypothesis for all diurnal primates of Rio Negro and Rio Branco in Brazil. Molecular Phylogenetics and Evolution 82, Part B:400–412.

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 Comput Biol 10:e1003537.

Boyer, S. L., R. M. Clouse, L. R. Benavides, P. Sharma, P. J. Schwendinger, I. Karunarathna, and G. Giribet (2007). Biogeography of the world: a case study

from cyphophthalmid Opiliones, a globally distributed group of arachnids. Journal of Biogeography 34:2070–2085.

Boyer, S. L., T. M. Markle, C. M. Baker, A. M. Luxbacher, and K. H. Kozak (2016). Historical refugia have shaped biogeographical patterns of species richness and phylogenetic diversity in mite harvestmen (Arachnida, Opiliones, Cyphophthalmi) endemic to the Australian Wet Tropics. Journal of Biogeography 43:1400–1411.

Brunsfeld, S. J., T. R. Miller, and B. C. Carstens (2007). Insights into the Biogeography of the Pacific Northwest of North America: Evidence from the phylogeography of *Salix melanopsis*. Systematic Botany 32:129–139.

Brunsfeld, S. J., and J. Sullivan (2006). A multi-compartmented glacial refugium in the northern Rocky Mountains: Evidence from the phylogeography of *Cardamine constancei* (Brassicaceae). Conservation Genetics 6:895–904.

Brunsfeld, S., J. Sullivan, D. Soltis, and P. Soltis (2001). Comparative phylogeography of northwestern North America: a synthesis. In Integrating Ecology and Evolution in a Spatial Context (J. Silvertown and J. Antonovics, Editors). Blackwell Publishing, Williston, Vermont, pp. 319–339.

Burg, T. M., A. J. Gaston, K. Winker, and V. L. Friesen (2006). Effects of Pleistocene glaciations on population structure of North American chestnut-backed chickadees. Molecular Ecology 15:2409–2419.

Burke, T. E. (2013). Land snails and slugs of the Pacific Northwest. Oregon State University Press.

Cahill, A. E., and J. S. Levinton (2015). Genetic differentiation and reduced genetic diversity at the northern range edge of two species with different dispersal modes. Molecular Ecology: n/a-n/a.

Carnaval, A. C., M. J. Hickerson, C. F. B. Haddad, M. T. Rodrigues, and C. Moritz (2009). Stability Predicts Genetic Diversity in the Brazilian Atlantic Forest Hotspot. Science 323:785–789.

Carstens, B. C., S. J. Brunsfeld, J. R. Demboski, J. M. Good, and J. Sullivan (2005a). Investigating the Evolutionary History of the Pacific Northwest Mesic Forest Ecosystem: Hypothesis Testing Within a Comparative Phylogeographic Framework. Evolution 59:1639–1652.

Carstens, B. C., J. D. Degenhardt, A. L. Stevenson, and J. Sullivan (2005b). Accounting for coalescent stochasticity in testing phylogeographical hypotheses: modelling

Pleistocene population structure in the Idaho giant salamander *Dicamptodon aterrimus*. Molecular Ecology 14:255–265.

Carstens, B. C., A. L. Stevenson, J. D. Degenhardt, and J. Sullivan (2004). Testing Nested Phylogenetic and Phylogeographic Hypotheses in the *Plethodon vandykei* Species Group. Systematic Biology 53:781–792.

Case, T. J., R. D. Holt, M. A. McPeek, and T. H. Keitt (2005). The community context of species' borders: ecological and evolutionary perspectives. Oikos 108:28–46.

Cazé, A. L. R., G. Mäder, T. S. Nunes, L. P. Queiroz, G. de Oliveira, J. A. F. Diniz-Filho, S. L. Bonatto, and L. B. Freitas (2016). Could refuge theory and rivers acting as barriers explain the genetic variability distribution in the Atlantic Forest? Molecular Phylogenetics and Evolution 101:242–251.

Ceccarelli, F. S., A. A. Ojanguren-Affilastro, M. J. Ramírez, J. A. Ochoa, C. I. Mattoni, and L. Prendini (2016). Andean uplift drives diversification of the bothriurid scorpion genus *Brachistosternus*. Journal of Biogeography 43:1942–1954.

Chaney, R. W. (1938). The Deschutes flora of eastern Oregon. Miocene and Pliocene floras of western North America: Carnegie Institution of Washington Publication 476:185–216.

Chaney, R. W., and D. I. Axelrod (1959). Miocene Floras of the Columbia Plateau: Composition and interpretation. Systematic considerations, by Ralph W. Chaney and Daniel I. Axelrod. Part I.. Part II.

Clark, P. U., A. S. Dyke, J. D. Shakun, A. E. Carlson, J. Clark, B. Wohlfarth, J. X. Mitrovica, S. W. Hostetler, and A. M. McCabe (2009). The Last Glacial Maximum. Science 325:710–714.

Coyle, F. A. (1983). Aerial Dispersal by Mygalomorph Spiderlings (Araneae, Mygalomorphae). The Journal of Arachnology 11:283–286.

Cross, A. T., and R. E. Taggart (1982). Causes of Short-Term Sequential Changes in Fossil Plant Assemblages: Some Considerations Based on a Miocene Flora of the Northwest United States. Annals of the Missouri Botanical Garden 69:676–734.

Daubenmire, R. (1975). Floristic Plant Geography of Eastern Washington and Northern Idaho. Journal of Biogeography 2:1–18.

Delcourt, P. A., and H. R. Delcourt (1993). Paleoclimates, paleovegetation, and paleofloras during the late Quaternary. Flora of North America 1:71–94.

Demboski, J. R., and J. Sullivan (2003). Extensive mtDNA variation within the yellow-pine chipmunk, *Tamias amoenus* (Rodentia: Sciuridae), and phylogeographic inferences for northwest North America. Molecular Phylogenetics and Evolution 26:389–408.

Derkarabetian, S., D. B. Steinmann, and M. Hedin (2010). Repeated and Time-Correlated Morphological Convergence in Cave-Dwelling Harvestmen (Opiliones, Laniatores) from Montane Western North America. PLoS ONE 5:e10388.

Derkarabetian, S., M. Burns, J. Starrett, and M. Hedin (2016). Population genomic evidence for multiple Pliocene refugia in a montane-restricted harvestman (Arachnida, Opiliones, *Sclerobunus robustus*) from the southwestern United States. Molecular Ecology 25:4611–4631.

Dong, L., G. Heckel, W. Liang, and Y. Zhang (2013). Phylogeography of Silver Pheasant (*Lophura nycthemera* L.) across China: aggregate effects of refugia, introgression and riverine barriers. Molecular Ecology 22:3376–3390.

Drummond, A. J., and A. Rambaut (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7:214.

Dunlop, J. A., and R. J. Garwood (2018). Terrestrial invertebrates in the Rhynie chert ecosystem. Phil. Trans. R. Soc. B 373:20160493.

Eaton, D. A. R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics:btu121.

Eaton, D. A. R., and R. H. Ree (2013). Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Systematic Biology 62:689–706.

Emata, K. N., and M. Hedin (2016). From the mountains to the coast and back again: Ancient biogeography in a radiation of short-range endemic harvestmen from California. Molecular Phylogenetics and Evolution 98:233–243.

English, J. M., and S. T. Johnston (2004). The Laramide Orogeny: What Were the Driving Forces? International Geology Review 46:833–838.

Espíndola, A., M. Ruffley, M. L. Smith, B. C. Carstens, D. C. Tank, and J. Sullivan (2016). Identifying cryptic diversity with predictive phylogeography. Proc. R. Soc. B 283:20161529.

Fernández, R., G. D. Edgecombe, and G. Giribet (2018). Phylogenomics illuminates the backbone of the Myriapoda Tree of Life and reconciles morphological and molecular phylogenies. Scientific Reports 8:83.

Fernández, R., P. P. Sharma, A. L. Tourinho, and G. Giribet (2017). The Opiliones tree of life: shedding light on harvestmen relationships through transcriptomics. Proc. R. Soc. B 284:20162340.

Fisher, J. R., D. M. Fisher, M. J. Skvarla, and A. P. G. Dowling (2014). Pre-ballooning in *Ummidia* Thorell 1875 (Araneae: Ctenizidae) from the Interior Highlands, USA: second account from the region and review of mygalomorph ballooning. Journal of Arachnology 42:318–321.

Fountain, A. G., M. Hoffman, K. Jackson, H. Basagic, T. Nylen, and D. Percy (2007). Digital outlines and topography of the glaciers of the American West.

Franklin, J. F., and C. T. Dyrness (1988). Natural vegetation of Oregon and Washington. Oregon State University Press Corvallis.

García-Vázquez, D., and I. Ribera (2016). The origin of widespread species in a poor dispersing lineage (diving beetle genus *Deronectes*). PeerJ 4:e2514.

Garzón-Orduña, I. J., J. E. Benetti-Longhini, and A. V. Z. Brower (2014). Timing the diversification of the Amazonian biota: butterfly divergences are consistent with Pleistocene refugia. Journal of Biogeography 41:1631–1638.

Gavin, D. G. (2009). The coastal-disjunct mesic flora in the inland Pacific Northwest of USA and Canada: refugia, dispersal and disequilibrium. Diversity and Distributions 15:972–982.

Giribet, G., P. P. Sharma, L. R. Benavides, S. L. Boyer, R. M. Clouse, D. Bivort, B. L, D. Dimitrov, G. Y. Kawauchi, J. Murienne, and P. J. Schwendinger (2012). Evolutionary and biogeographical history of an ancient and global group of arachnids (Arachnida: Opiliones: Cyphophthalmi) with a new taxonomic arrangement. Biological Journal of the Linnean Society 105:92–130.

Good, J. M., and J. Sullivan (2001). Phylogeography of the red-tailed chipmunk (*Tamias ruficaudus*), a northern Rocky Mountain endemic. Molecular Ecology 10:2683–2695.

Graham, A. (1999). Late Cretaceous and Cenozoic history of North American vegetation: north of Mexico. Oxford University Press New York.

Gresens, R. L. (1987). Early Cenozoic geology of central Washington State: I. Summary of sedimentary, igneous, and tectonic events. Summary of sedimentary, igneous, and tectonic events: Washington Division of Earth Resources Bulletin 77:169–177.

Haffer, J. (1969). Speciation in Amazonian forest birds. Science 165:131–137.

Hammond, P. E. (1979). A tectonic model for evolution of the Cascade Range.

Hardin, G. (1960). The Competitive Exclusion Principle. Science 131:1292–1297.

Harrington, S. M., B. D. Hollingsworth, T. E. Higham, and T. W. Reeder (2018). Pleistocene climatic fluctuations drive isolation and secondary contact in the red diamond rattlesnake (*Crotalus ruber*) in Baja California. Journal of Biogeography 45:64–75.

Harvey, M. S., M. G. Rix, V. W. Framenau, Z. R. Hamilton, M. S. Johnson, R. J. Teale, G. Humphreys, and W. F. Humphreys (2011). Protecting the innocent: studying short-range endemic taxa enhances conservation outcomes. Invertebrate Systematics 25:1–10.

Harvey, M. S., B. Y. Main, M. G. Rix, and S. J. B. Cooper (2015). Refugia within refugia: in situ speciation and conservation of threatened *Bertmainius* (Araneae : Migidae), a new genus of relictual trapdoor spiders endemic to the mesic zone of south-western Australia. Invertebrate Systematics 29:511–553.

Heads, M. (2011). Old Taxa on Young Islands: A Critique of the Use of Island Age to Date Island-Endemic Clades and Calibrate Phylogenies. Systematic Biology 60:204–218.

Hedin, M., S. Derkarabetian, M. McCormack, C. Richart, and J. W. Shultz (2010). The phylogenetic utility of the nuclear protein-coding gene EF-1α for resolving recent divergences in Opiliones, emphasizing intron evolution. Journal of Arachnology 38:9–20.

Hedin, M., and M. McCormack (2017). Biogeographical evidence for common vicariance and rare dispersal in a southern Appalachian harvestman (Sabaconidae, *Sabacon cavicolens*). Journal of Biogeography 44: 1665-1678.

Hedin, M., J. Starrett, S. Akhter, A. L. Schönhofer, and J. W. Shultz (2012). Phylogenomic Resolution of Paleozoic Divergences in Harvestmen (Arachnida, Opiliones) via Analysis of Next-Generation Transcriptome Data. PLoS ONE 7:e42888.

Hedin, M., J. Starrett, and C. Hayashi (2013). Crossing the uncrossable: novel trans-valley biogeographic patterns revealed in the genetic history of low-dispersal mygalomorph spiders (Antrodiaetidae, Antrodiaetus) from California. Molecular Ecology 22:508–526.

Herman, T. A., and J. L. Bouzat (2016). Range-wide phylogeography of the four-toed salamander: out of Appalachia and into the glacial aftermath. Journal of Biogeography 43:666–678.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. Nature 405:907–913.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. Biological Journal of the Linnean Society 58:247–276.

Hird, S., N. Reid, J. Demboski, and J. Sullivan (2010). Introgression at differentially aged hybrid zones in red-tailed chipmunks. Genetica 138:869–883.

Hooper, P. R. (1997). The Columbia River Flood Basalt Province: Current Status. In Large Igneous Provinces: Continental, Oceanic, and Planetary Flood Volcanism (J. J. Honey and M. F. Coffin, Editors). American Geophysical Union, pp. 1–27.

Hugall, A., C. Moritz, A. Moussalli, and J. Stanisic (2002). Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). Proceedings of the National Academy of Sciences 99:6112–6117.

Ilut, D. C., M. L. Nydam, and M. P. Hare (2014). Defining Loci in Restriction-Based Reduced Representation Genomic Data from Nonmodel Species: Sources of Bias and Diagnostics for Optimal Clustering. BioMed Research International 2014:e675158.

Jezkova, T., B. R. Riddle, D. C. Card, D. R. Schield, M. E. Eckstut, and T. A. Castoe (2015). Genetic consequences of postglacial range expansion in two codistributed rodents (genus *Dipodomys*) depend on ecology and genetic locus. Molecular Ecology 24:83–97.

Kane, T. C., T. C. Barr, and G. E. Stratton (1990). Genetic patterns and population-structure in Appalachian *Trechus* of the *vandykei* group (Coleoptera, Carabidae). Brimleyana:133–150.

Keith, R., and M. Hedin (2012). Extreme mitochondrial population subdivision in southern Appalachian paleoendemic spiders (Araneae: Hypochilidae: *Hypochilus*), with implications for species delimitation. Journal of Arachnology 40:167–181.

Klicka, J., and R. M. Zink (1997). The Importance of Recent Ice Ages in Speciation: A Failed Paradigm. Science 277:1666–1669.

Larson, A., D. B. Wake, and K. P. Yanev (1984). Measuring Gene Flow Among Populations Having High Levels of Genetic Fragmentation. Genetics 106:293–308.

Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. Molecular Biology and Evolution 29:1695–1701.

Lanier, H. C., R. Massatti, Q. He, L. E. Olson, and L. L. Knowles (2015). Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). Molecular Ecology 24:3688–3705.

Leonard, W. P., L. Chichester, J. Baugh, and T. Wilke (2003). *Kootenaia burkei*, a new genus and species of slug from northern Idaho, United States (Gastropoda: Pulmonata: Arionidae). Zootaxa 355.

Leonard, W. P., L. Chichester, C. H. Richart, and T. A. Young (2011). *Securicauda hermani* and *Carinacauda stormi*, two new genera and species of slug from the Pacific Northwest of the United States (Gastropoda: Stylommatophora: Arionidae), with notes on *Gliabates oregonius* Webb 1959. Zootaxa 2746:43–56.

Leopold, E. B., and M. F. Denton (1987). Comparative Age of Grassland and Steppe East and West of the Northern Rocky Mountain. Annals of the Missouri Botanical Garden 74:841–867.

Lester, S. E., B. I. Ruttenberg, S. D. Gaines, and B. P. Kinlan (2007). The relationship between dispersal ability and geographic range size. Ecology Letters 10:745–758.

Lorain, C. C. (1988). Floristic history and distribution of coastal disjunct plants of the northern Rocky Mountains. Moscow, ID: University of Idaho.

Lucid, M. K., A. Rankin, A. Espíndola, L. Chichester, S. Ehlers, L. Robinson, and J. Sullivan (2018). Taxonomy and biogeography of *Hemphillia* (Gastropoda: Pulmonata: Arionidae) in North American rainforests, with description of a new species (Skade's jumping-slug, *Hemphillia skadei* sp. nov.). Canadian Journal of Zoology:305–316.

Martínez-Solano, I., J. Teixeira, D. Buckley, and M. García-París (2006). Mitochondrial DNA phylogeography of *Lissotriton boscai* (Caudata, Salamandridae): evidence for old, multiple refugia in an Iberian endemic. Molecular Ecology 15:3375–3388.

May, C. L., and R. E. Gresswell (2003). Large wood recruitment and redistribution in headwater streams in the southern Oregon Coast Range, U.S.A. Canadian Journal of Forest Research 33:1352–1362.

McCormack, J. E., J. Heled, K. S. Delaney, A. T. Peterson, and L. L. Knowles (2011). Calibrating Divergence Times on Species Trees Versus Gene Trees: Implications for Speciation History of *Aphelocoma* Jays. Evolution 65:184–202.

McKee, B. (1972). Cascadia; the geologic evolution of the Pacific Northwest.

Metzger, G., A. Espindola, L. P. Waits, and J. Sullivan (2015). Genetic Structure across Broad Spatial and Temporal Scales: Rocky Mountain Tailed Frogs (*Ascaphus montanus*; Anura: Ascaphidae) in the Inland Temperate Rainforest. Journal of Heredity 106:700–710.

Mitchell, S. G., and D. R. Montgomery (2006). Polygenetic topography of the Cascade Range, Washington State, USA. Am J Sci 306:736–768.

Montgomery, D. R., and J. M. Buffington (1997). Channel-reach morphology in mountain drainage basins. GSA Bulletin 109:596–611.

Moore, J.-S., R. Bajno, J. D. Reist, and E. B. Taylor (2015). Post-glacial recolonization of the North American Arctic by Arctic char (*Salvelinus alpinus*): genetic evidence of multiple northern refugia and hybridization between glacial lineages. Journal of Biogeography 42:2089–2100.

Mora, E., A. Paspati, A. E. Decae, and M. A. Arnedo (2017). Rafting spiders or drifting islands? Origins and diversification of the endemic trap-door spiders from the Balearic Islands, Western Mediterranean. Journal of Biogeography 44:924–936.

Moraes, L. J. C. L., D. Pavan, M. C. Barros, and C. C. Ribas (2016). The combined influence of riverine barriers and flooding gradients on biogeographical patterns for amphibians and squamates in south-eastern Amazonia. Journal of Biogeography 43:2113–2124.

Moritz, C., K. S. Richardson, S. Ferrier, G. B. Monteith, J. Stanisic, S. E. Williams, and T. Whiffin (2001). Biogeographical concordance and efficiency of taxon indicators for establishing conservation priority in a tropical rainforest biota. Proceedings of the Royal Society of London B: Biological Sciences 268:1875–1881.

Neiber, M. T., and B. Hausdorf (2015). Phylogeography of the land snail genus *Circassina* (Gastropoda: Hygromiidae) implies multiple Pleistocene refugia in the western Caucasus region. Molecular Phylogenetics and Evolution 93:129–142.

Nicastro, K. R., G. I. Zardi, S. Teixeira, J. Neiva, E. A. Serrão, and G. A. Pearson (2013). Shift happens: trailing edge contraction associated with recent warming trends threatens a distinct genetic lineage in the marine macroalga *Fucus vesiculosus*. BMC Biology 11:6.

Nielson, M., K. Lohman, C. H. Daugherty, F. W. Allendorf, K. L. Knudsen, and J. Sullivan (2006). Allozyme and Mitochondrial DNA Variation in the Tailed Frog (Anura: *Ascaphus*): The Influence of Geography and Gene Flow. Herpetologica 62:235–258.

Nielson, M., K. Lohman, and J. Sullivan (2001). Phylogeography of the tailed frog (*Ascaphus truei*): implications for the biogeography of the Pacific Northwest. Evolution; International Journal of Organic Evolution 55:147–160.

Oberski, J. T., P. P. Sharma, K. R. Jay, M. J. Coblens, K. A. Lemon, J. E. Johnson, and S. L. Boyer (2018). A dated molecular phylogeny of mite harvestmen (Arachnida: Opiliones: Cyphophthalmi) elucidates ancient diversification dynamics in the Australian Wet Tropics. Molecular Phylogenetics and Evolution 127:813–822.

Opatova, V., J. E. Bond, and M. A. Arnedo (2016). Uncovering the role of the Western Mediterranean tectonics in shaping the diversity and distribution of the trap-door spider genus *Ummidia* (Araneae, Ctenizidae). Journal of Biogeography 43:1955–1966.

Papadopoulou, A., I. Anastasiou, and A. P. Vogler (2010). Revisiting the Insect Mitochondrial Molecular Clock: The Mid-Aegean Trench Calibration. Molecular Biology and Evolution 27:1659–1672.

Perez, M. F., I. a. S. Bonatelli, E. M. Moraes, and B. C. Carstens (2016). Model-based analysis supports interglacial refugia over long-dispersal events in the diversification of two South American cactus species. Heredity 116:550–557.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7:e37135.

Petit, R. J., I. Aguinagalde, J.-L. de Beaulieu, C. Bittkau, S. Brewer, R. Cheddadi, R. Ennos, S. Fineschi, D. Grivet, M. Lascoux, A. Mohanty, et al. (2003). Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. Science 300:1563–1565.

Pielou, E. C. (1991). After the Ice Age: the Return of Life to Glaciated North America. University of Chicago Press, Chicago.

Pokorny, L., R. Riina, M. Mairal, A. S. Meseguer, V. Culshaw, J. Cendoya, M. Serrano, R. Carbajal, S. Ortiz, M. Heuertz, and I. Sanmartín (2015). Living on the edge: timing of Rand Flora disjunctions congruent with ongoing aridification in Africa. Frontiers in Genetics 6.

Poncet, V., F. Munoz, J. Munzinger, Y. Pillon, C. Gomez, M. Couderc, C. Tranchant-Dubreuil, S. Hamon, and A. de Kochko (2013). Phylogeography and niche modelling of the relict plant *Amborella trichopoda* (Amborellaceae) reveal multiple Pleistocene refugia in New Caledonia. Molecular Ecology 22:6163–6178.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler (2006). Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Systematic Biology 55:595–609.

Potter, S., A. T. Xue, J. G. Bragg, D. F. Rosauer, J. E. Roycroft, and C. Moritz (2018). Pleistocene climatic changes drive diversification across a tropical savanna. Molecular Ecology 27:520–532.

Rambaut A., A. J. Drummond (2014). TreeAnnotator v2. 1.2. Edinburgh: University of Edinburgh, Institute of Evolutionary Biology.

Rambaut A., M. A. Suchard, D. Xie, A. J. Drummond (2014). Tracer v1. 6. Computer program and documentation distributed by the author. Available at: http://beast.bio.ed.ac.uk/Tracer

Reiners, P. W., T. A. Ehlers, J. I. Garver, S. G. Mitchell, D. R. Montgomery, J. A. Vance, and S. Nicolescu (2002). Late Miocene exhumation and uplift of the Washington Cascade Range. Geology 30:767–770.

Richart, C., and M. Hedin (2013). Three new species in the harvestmen genus Acuclavella (Opiliones, Dyspnoi, Ischyropsalidoidea), including description of male *Acuclavella quattuor* Shear, 1986. ZooKeys 311:19–68.

Runck, A. M., and J. A. Cook (2005). Postglacial expansion of the southern red-backed vole (*Clethrionomys gapperi*) in North America. Molecular Ecology 14:1445–1456.

Russel, I. C. (1900). A preliminary paper on the geology of the Cascade Mountains in northern Washington: US Geol. Survey. 20th Ann. Rept., pt 2:83–210.

Sánchez-Ramírez, S., R. E. Tulloss, L. Guzmán-Dávalos, J. Cifuentes-Blanco, R. Valenzuela, A. Estrada-Torres, F. Ruán-Soto, R. Díaz-Moreno, N. Hernández-Rico, M. Torres-Gómez, H. León, and J.-M. Moncalvo (2015). In and out of refugia: historical patterns of diversity and demography in the North American Caesar's mushroom species complex. Molecular Ecology 24:5938–5956.

Satler, J. D., and B. C. Carstens (2016). Phylogeographic concordance factors quantify phylogeographic congruence among co-distributed species in the *Sarracenia alata* pitcher plant system. Evolution 70:1105–1119.

Schwentner, M., and G. Giribet (2018). Phylogeography, species delimitation and population structure of a Western Australian short-range endemic mite harvestman (Arachnida: Opiliones: Pettalidae: *Karripurcellia*). Evolutionary Systematics 2:81–87.

Shafer, A. B. A., C. I. Cullingham, S. D. Côté, and D. W. Coltman (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology 19:4589–4621.

Shear, W. A. (1986). A cladistic analysis of the opilionid superfamily Ischyropsalidoidea, with descriptions of the new family Ceratolasmatidae, the new genus *Acuclavella*, and four new species. American Museum novitates; no. 2844.

Shelley, R. M., and S. I. Golovatch (2011). Atlas of Myriapod Biogeography. I. Indigenous Ordinal and Supra-Ordinal Distributions in the Diplopoda: Perspectives on Taxon Origins and Ages, and a Hypothesis on the Origin and Early Evolution of the Class. Insecta Mundi 158:1–134.

Slatkin, M. (1981). Estimating levels of gene flow in natural populations. Genetics 99:323–335.

Smiley, C. J. (1963). The Ellensburg flora of Washington. University of California Press.

Smith, G. O. (1903). Geology and physiography of central Washington. In Contributions to the Geology of Washington. Series C, Systematic Geology and Paleontology 64:3–39.

Smith, M. L., M. Ruffley, A. M. Rankin, A. Espíndola, D. C. Tank, J. Sullivan, and B. C. Carstens (2018). Testing for the presence of cryptic diversity in tail-dropper slugs (*Prophysaon*) using molecular data. Biological Journal of the Linnean Society 124:518–532.

Steele, C. A., B. C. Carstens, A. Storfer, and J. Sullivan (2005). Testing hypotheses of speciation timing in *Dicamptodon copei* and *Dicamptodon aterrimus* (Caudata: Dicamptodontidae). Molecular Phylogenetics and Evolution 36:90–100.

Takeuchi, A., and P. B. Larson (2005). Oxygen isotope evidence for the late Cenozoic development of an orographic rain shadow in eastern Washington, USA. Geology 33:313–316.

Thesing, B. D., R. D. Noyes, D. E. Starkey, and D. B. Shepard (2016). Pleistocene climatic fluctuations explain the disjunct distribution and complex phylogeographic structure of the Southern Red-backed Salamander, *Plethodon serratus*. Evolutionary Ecology 30:89–104.

Thomas, S. M., and M. Hedin (2008). Multigenic phylogeographic divergence in the paleoendemic southern Appalachian opilionid *Fumontana deprehendor* Shear (Opiliones, Laniatores, Triaenonychidae). Molecular Phylogenetics and Evolution 46:645–658.

Tsai, Y.-H. E., and B. C. Carstens (2013). Assessing model fit in phylogeographical investigations: an example from the North American sandbar willow *Salix melanopsis*. Journal of Biogeography 40:131–141.

U.S. Geological Survey, 2012, The StreamStats program, online at http://streamstats.usgs.gov.

Vance, J. A., G. A. Clayton, J. M. Mattinson, C. W. Naeser, and J. E. Schuster (1987). Early and middle Cenozoic stratigraphy of the Mount Rainier-Tieton River area, southern Washington Cascades. Selected papers on the geology of Washington: Washington Division of Geology and Earth Resources Bulletin 77:269–290.

Wallace, A. R. (1854). On the monkeys of the Amazon. Journal of Natural History 14:451–454.

Webb, S. D. (1977). A history of savanna vertebrates in the New World. Part I: North America. Annual Review of Ecology and Systematics 8:355–380.

Wei, S.-J., L.-J. Cao, Y.-J. Gong, B.-C. Shi, S. Wang, F. Zhang, X.-J. Guo, Y.-M. Wang, and X.-X. Chen (2015). Population genetic structure and approximate Bayesian computation analyses reveal the southern origin and northward dispersal of the oriental fruit moth *Grapholita molesta* (Lepidoptera: Tortricidae) in its native range. Molecular Ecology 24:4094–4111.

Wilson, J. S., and J. P. Pitts (2010). Illuminating the lack of consensus among descriptions of earth history data in the North American deserts: A resource for biologists. Progress in Physical Geography 34:419–441.

Xu, X., F. Liu, J. Chen, H. Ono, I. Agnarsson, D. Li, and M. Kuntner (2016). Pre-Pleistocene geological events shaping diversification and distribution of primitively segmented spiders on East Asian margins. Journal of Biogeography 43:1004–1019.

Yeates, D. K., P. Bouchard, and G. B. Monteith (2002). Patterns and levels of endemism in the Australian Wet Tropics rainforest: evidence from flightless insects. Invertebrate Systematics 16:605–619.

Yokoyama, Y., K. Lambeck, P. D. Deckker, P. Johnston, and L. K. Fifield (2000). Timing of the Last Glacial Maximum from observed sea-level minima. Nature 406:713–716.

Zemlak, T. S., S. J. Walde, E. M. Habit, and D. E. Ruzzante (2011). Climate-induced changes to the ancestral population size of two Patagonian galaxiids: the influence of glacial cycling. Molecular Ecology 20:5280–5294.

Zhang, J., P. Kapli, P. Pavlidis, and A. Stamatakis (2013). A general species delimitation method with applications to phylogenetic placements. Bioinformatics 29:2869–2876.

Table 3.1. Evidence for Columbia Basin xerification.

Citation	Page #	Evidence	Time
Smiley 1963	206	fossil flora transition	5 MYA
Axelrod 1948	129	fossil flora transition	7 - 5 MYA
Webb 1977	369	distribution mammal fossils	10 - 7 MYA
Leopold & Denton 1987	845	fossil flora transition	8 - 4.5 MYA
Leopold & Denton 1987	848	pollen samples	8.5 - 5 MYA
Leopold & Denton 1987	849	pollen samples	4 - 3 MYA
Graham 1999	261	Palouse Prairie formation	3 MYA
Graham 1999	268	floral community	3 - 2 MYA
Chaney 1938	387	fossil flora transition	≥ 11 MYA
Chaney & Axelrod 1959	53	faunal fossils	~ 12 MYA
McKee 1972	186	basalt uplift	8 - 5 MYA
McKee 1972	278	volcanic sediment deposits	8 - 5 MYA
Hammond 1979	227	basalt uplift	10 - 4.5 MYA
Graham 1999	268	Cascade orogeny	5 MYA
Gresens 1987	176	volcanic sediment deposits	≤ 15 MYA
Reiners et al. 2002	*769	Fission track dating	10 - 5 MYA
Takeuchi & Larson 2005	**315	ancient water isotope ratios	12 - 8 MYA

*taken from Figure 3 for the samples near Mount Rainier

**transition samples as shown in Figure 3

Notes: Paleontological (above the line) and geological (below) evidence for the timing of the xerification of the Columbia Basin and isolation of the inland WHZ.

Table 3.2. Thorn harvestmen species and GenBank numbers used for *BEAST.

Species	Distribution	SDSU TAC#	COI	EF1a
A. makah	WA, Olympic Peninsula	OP2345	GQ870647	GQ872168
		OP2715	KF181737	KF181770
		OP2716	KF181738	KF181771
A. leonardi	WA, Southern Cascades	OP2347	GQ870648	GQ872169
		OP2714	KF181728	KF181766
A. sheari	ID, S of Salmon River	OP2708	KF181741	KF181775
		OP2709	KF181742	KF181776
		OP2720	KF181743	KF181777
A. quattuor	ID, S of SF Clearwater Rv	OP2242	KF181739	KF181773
		OP2257	KF181740	KF181774
A. merickeli	ID, S of Selway River	OP2237	KF181733	KF181767
		OP2250	KF181735	KF181768
		OP2251	GQ870644	GQ872165
A. cf. *quattuor*	ID, N of Selway River	OP2230	KF181730	KF181759
		OP2233	GQ870643	GQ872164
		OP2275	KF181731	KF181760
		OP2284	KF181732	KF181761

Table 3.3. The BEAST time-tree node depth and taxa span, the latitude and longitude in decimal degrees used to measure basin size, and the result of this measurement.

Depth (MYA)	TaxaSpan	Lat_River	Long_Riv	BasinSize (km^2)
0.19	OP2720 to OP2708	45.3996	-116.0106	2.69256
0.18	OP2328 to OP2334	47.22495	-115.57899	934.03353
0.16	OP2255 to OP2269	45.614	-115.4488	0
0.33	OP2246 to OP2236	46.06041	-115.33984	30.5502
0.37	OP2250 to OP2252	45.76255	-115.2382	18.40779
0.34	OP2255 to OP2266	45.62063	-115.51326	21.79938
0.49	OP2323 to OP2327	47.24975	-116.02117	369.21729
0.37	OP2271 to OP2258	45.62036	-116.06738	210.4857
0.5	OP2250 to OP2266	45.79379	-115.401	358.47294
0.6	OP2250 to OP2236	45.80864	-115.47523	237.07473
0.56	OP2332 to OP2341	47.15078	-115.407	71.94831
0.61	OP2335 to OP2326	47.2489	-115.8464	105.76065
0.49	OP2340 to OP2327	47.25118	-116.02771	2303.25207
0.35	OP2315 to OP2311	46.73983	-115.54559	2807.14914
0.74	OP2303 to OP2308	46.79026	-115.22043	24.36249
0.59	OP2342 to OP2230	46.49304	-114.85854	69.59232
0.49	OP2316 to OP2318	46.94868	-116.67298	179.49537
0.47	OP2287 to OP2275	46.22624	-115.43271	45.54051
0.46	OP2264 to OP2245	45.9944	-115.6389	0
0.36	OP2323 to OP2325	47.24875	-116.01906	369.11373
0.36	OP2341 to OP2334	47.20186	-115.51701	59.23632
0.53	OP2360 to OP2311	46.70679	-115.55823	122.12313
0.63	OP2284 to OP2275	46.05691	-115.31403	124.71213
0.73	OP2242 to OP2258	45.69978	-116.3165	84.78975
0.85	OP2333 to OP2327	47.22665	-115.70528	1026.38316
0.8	OP2280 to OP2281	46.1654	-115.5891	71.32695
0.69	OP2319 to OP2318	46.9828	-116.6706	95.68944
0.92	OP2242 to OP2236	45.8235	-115.88978	291.7803
1.04	OP2303 to OP2327	46.983	-115.3666	198.39507
0.92	OP2234 to OP2275	46.2524	-115.3998	113.73477
1	OP4063 to OP2318	46.9152	-116.8845	625.39884
0.69	OP2298 to OP2315	46.5641	-115.62162	81.16515
1.45	OP2342 to OP2275	46.14297	-115.59808	3057.01353
0.8	OP2301 to OP2315	46.62286	-115.99189	229.23006
1.42	OP2304 to OP2327	46.71569	-115.25989	1499.47113

Table 3.3 (continued). The BEAST time-tree node depth and taxa span, the latitude and longitude in decimal degrees used to measure basin size, and the result of this measurement.

Depth (MYA)	TaxaSpan	Lat_River	Long_Riv	BasinSize (km^2)
1.03	OP2294 to OP2315	46.37215	-116.16864	628.63509
1.28	OP2280 to OP2315	46.37215	-116.16864	628.63509
1.67	OP2288 to OP2294	46.50347	-116.32361	6325.39302
1.9	OP2304 to OP2315	46.50347	-116.32361	6325.39302
1.98	OP2230 to OP2708	46.14018	-115.59574	5203.26864
2.21	OP2720 to OP2279	46.14679	-115.97885	8825.79744
1.45	OP2720 to OP2230	45.4162	-116.3143	9061.5

Table 3.4. Data relevant to the Down Canyon Hypothesis

Taxa	**r-value**	**p-value**	**n**
All inland *Acuclavella*	-0.2616	**0.0179**	82
Bordering Glaciers	-0.2759	**0.0422**	55
Not Bordering Glaciers	-0.2213	0.2884	25
A. merickeli + *A. quattuor*	-0.1615	0.4741	22
A. cf. *quattuor*	-0.4685	0.1546	11
A. shoshone	-0.4480	**0.0370**	22
A. cosmetoides	-0.0390	0.8820	18
Hoodoo Mountains	-0.2669	0.5642	7

Notes: Results from the Pearson correlation analyses between homozygosity and elevation and sample sizes (n). Significant results **($\alpha = 0.05$)** are emboldened.

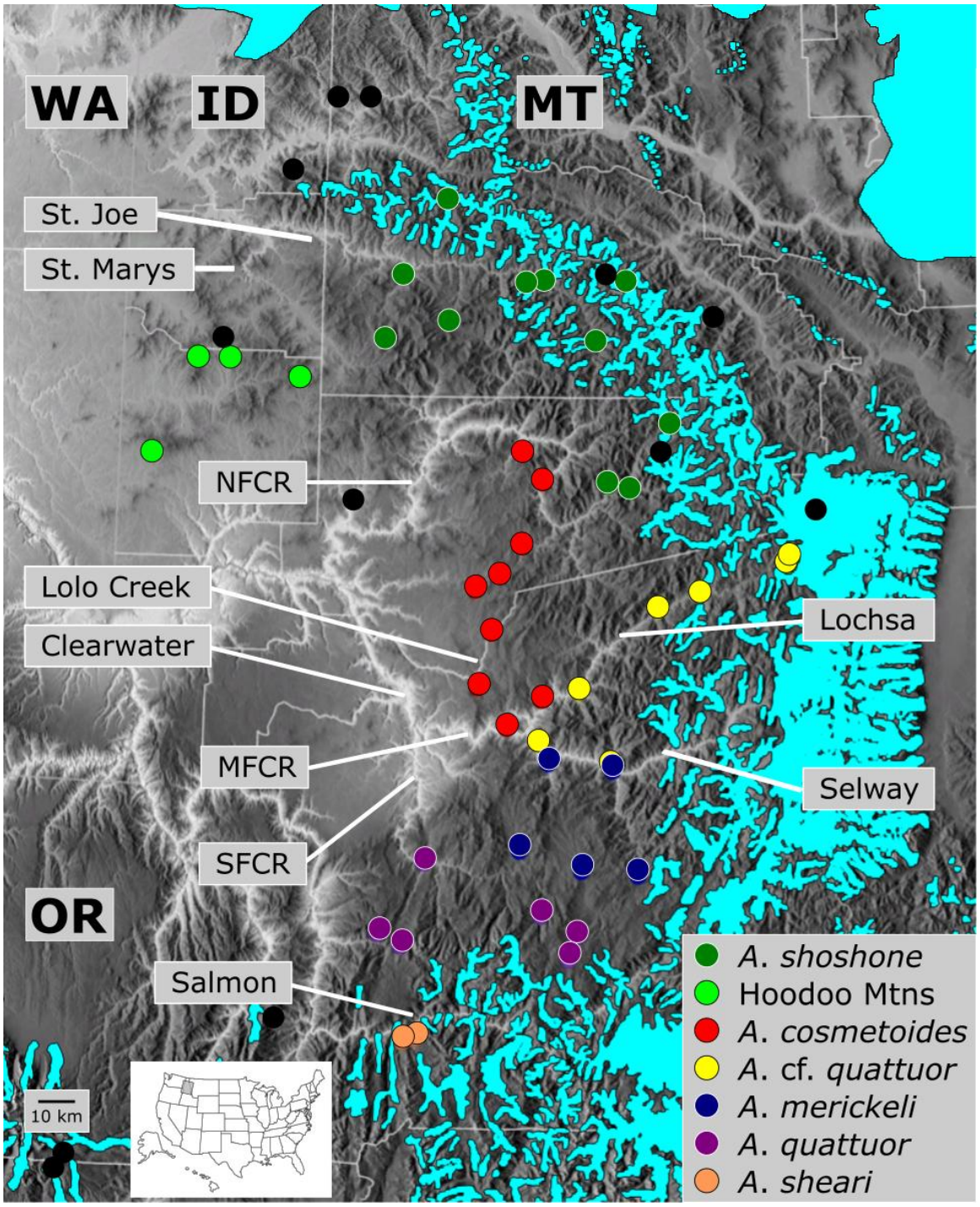


Figure 3.1. Geographic distribution of inland *Acuclavella* species showing the location of alpine glaciers at the LGM (Fountain et al. 2007) in turquoise, major rivers (NFCR, MFCR, SFCR are North, Middle, and South Fork Clearwater River), and negative search localities as black circles.

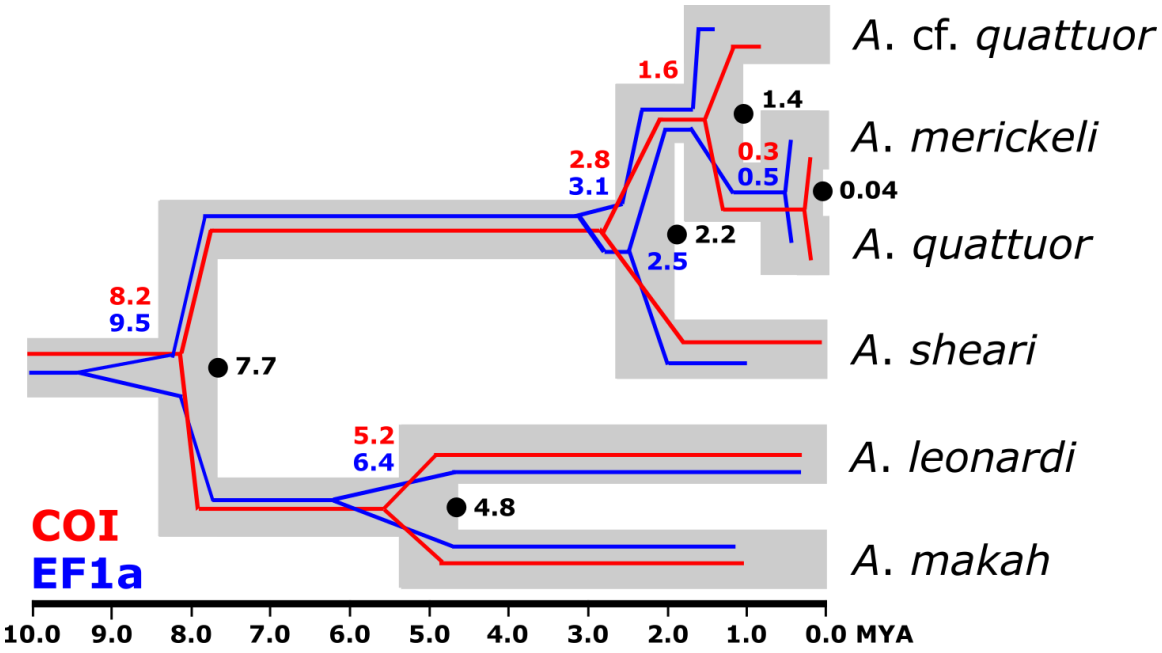


Figure 3.2. *BEAST showing the inference of the split between coastal and inland *Acuclavella* at 7.7 MYA, and the root of inland *Acuclavella* at 2.2 MYA. Red and blue values are the estimated divergence times for COI and EF1a respectively.

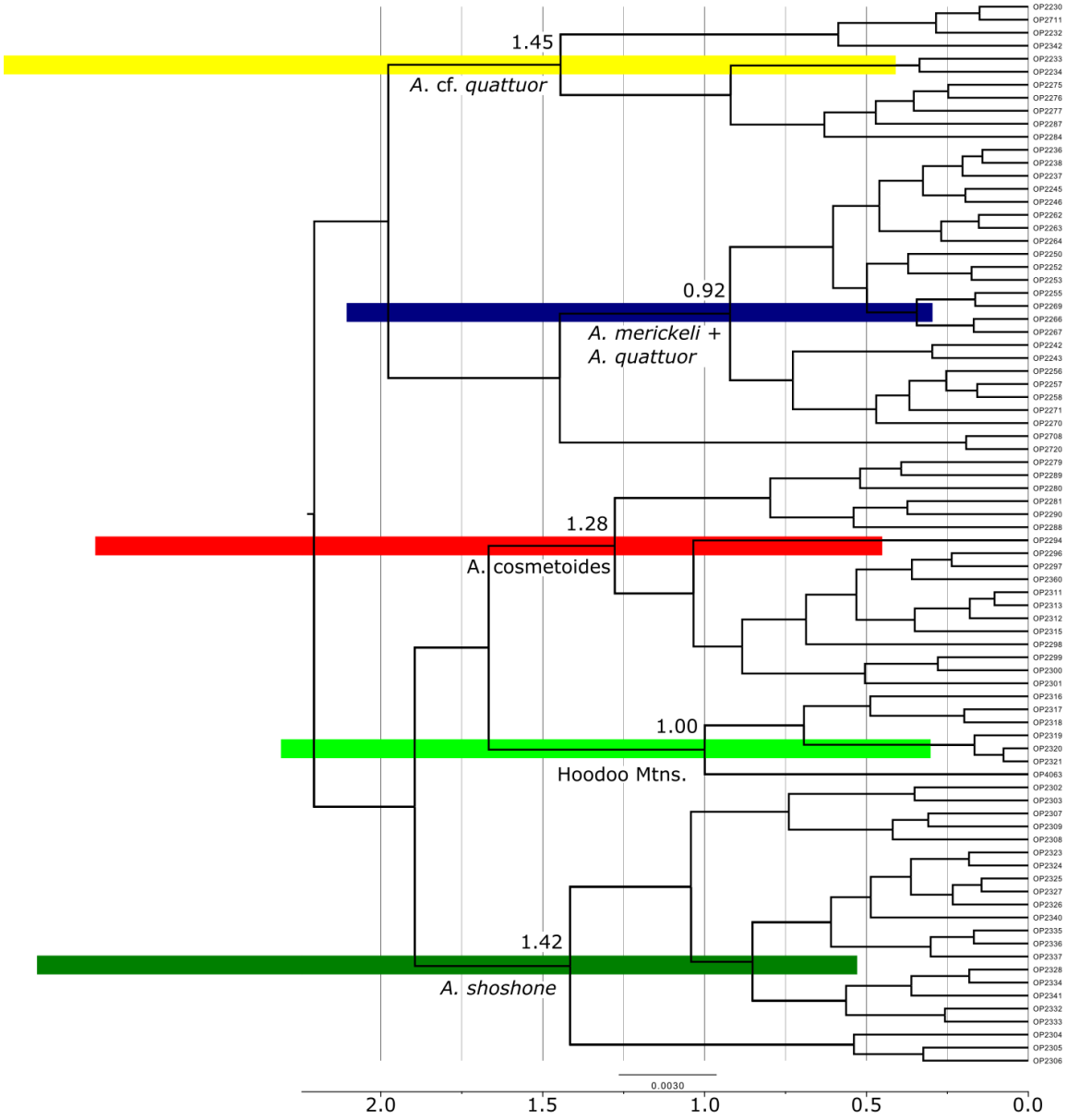


Figure 3.3. BEAST divergence dating.

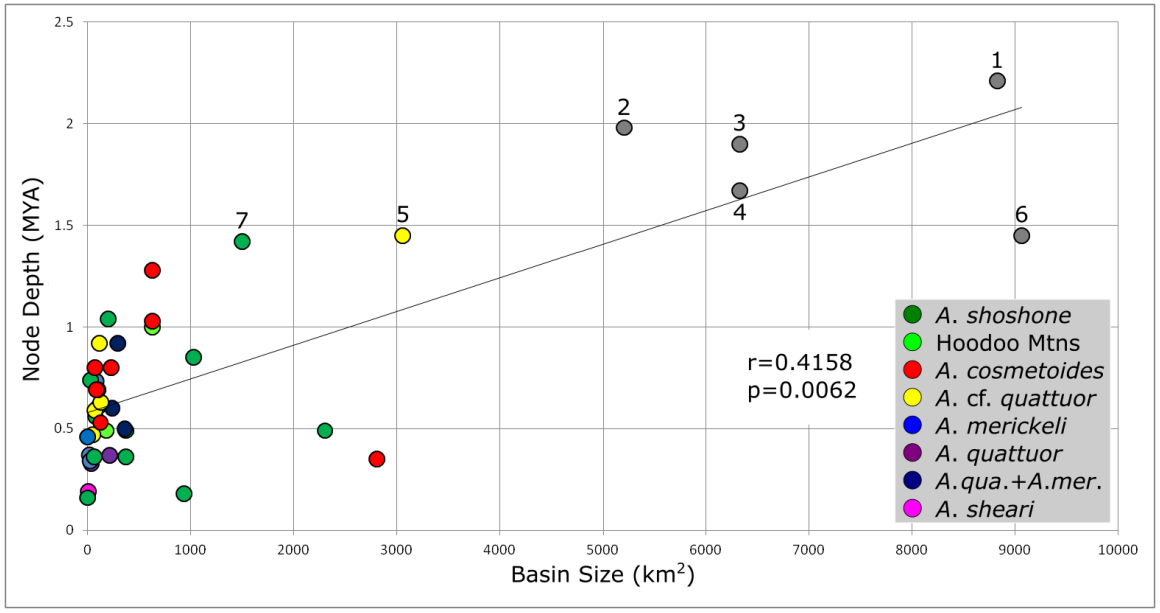


Figure 3.4. The correlation between the depth BEAST time tree nodes and the size of the rivers (basin size) that separate them. Gray points in the data series (1-4, 6) represent nodes above the species-level: 1) root node and MFCR, 2) *A*. cf. *quattuor* and the Selway River, 3) *A*. *shoshone* and the NFCR, 4) Hoodoo Mountains from *A*. *cosmetoides* and the NFCR, 6) *A*. *sheari* and the Salmon River. Data series point 5) is *A*. cf. *quattuor* north and south of the Lochsa River, and 7) is *A*. cf. *shoshone* and the NFCR further up-basin. The three shallow-node points well-below the trend line represent recent dispersal events across major rivers.

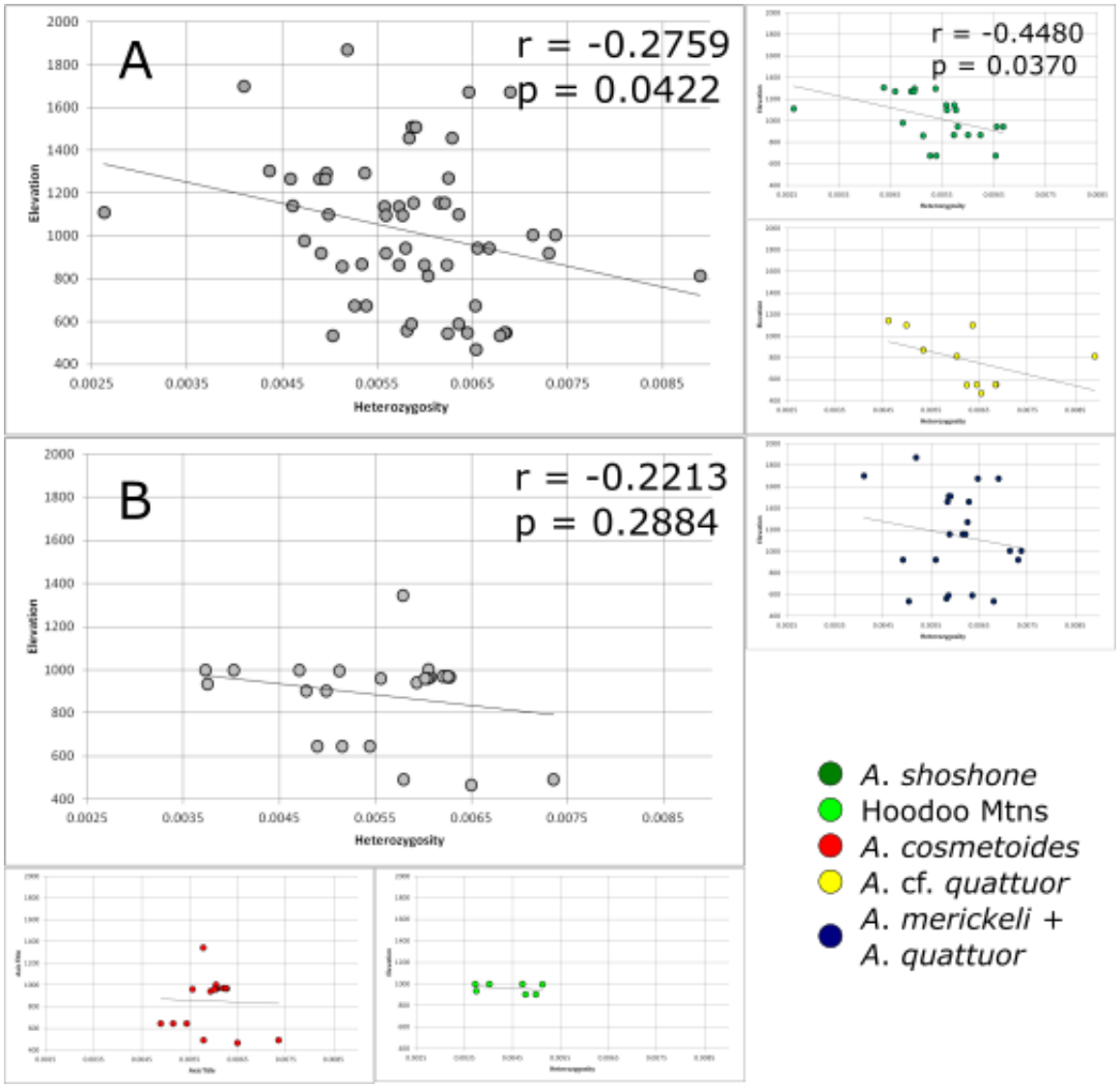


Figure 3.5. The correlation between genetic diversity (homozygosity) and elevation in: A) species that had populations within or adjacent to Pleistocene glacier fields, but not in B) species that did not. Though a negative trend line is consistently recovered in all analyses, the only significant within-species negative correlation is in *A. shoshone*.

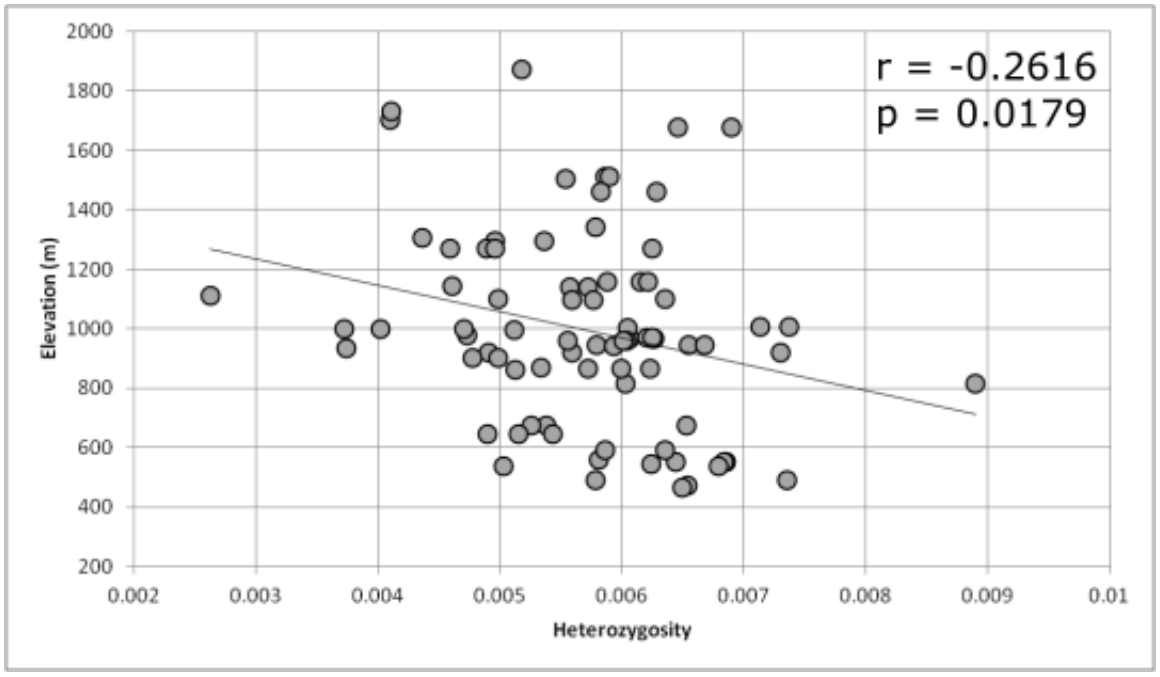


Figure 3.6. The correlation between genetic diversity (homozygosity) and elevation for all inland *Acuclavella*.

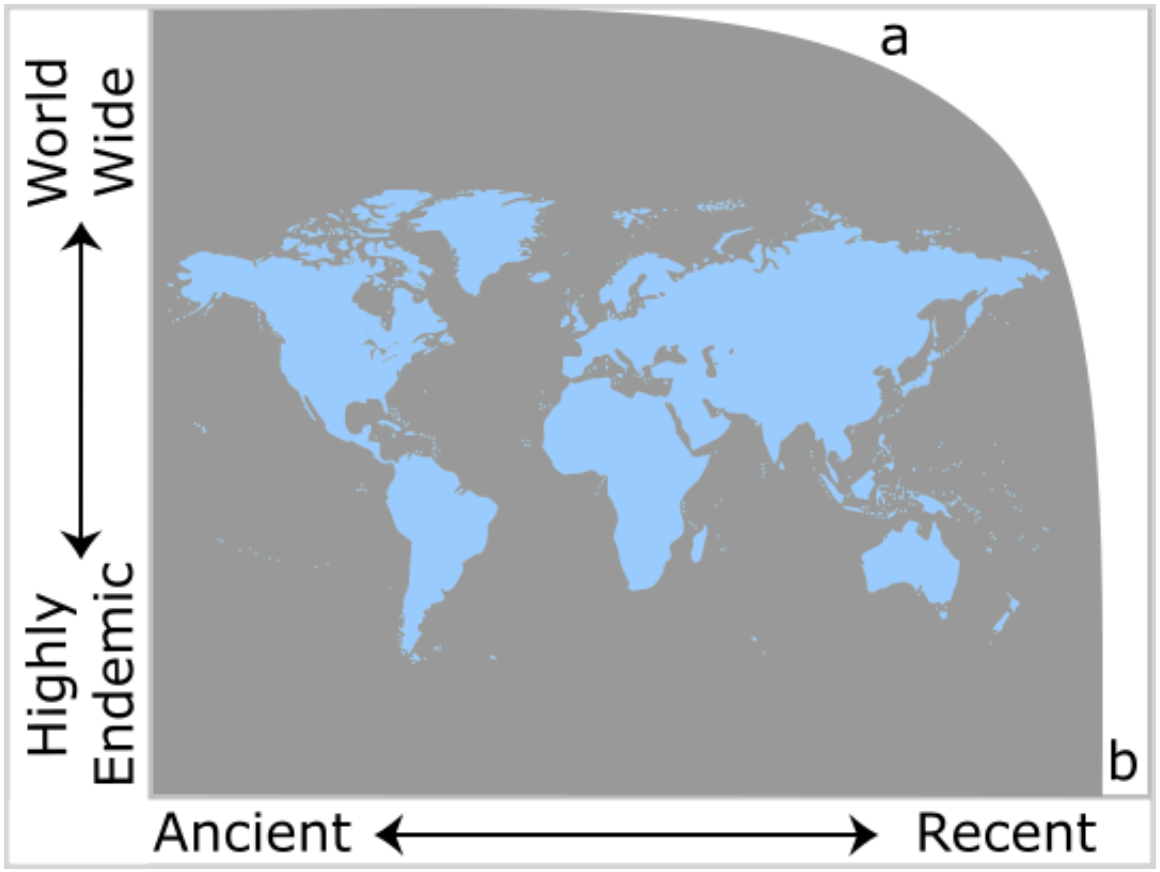


Figure 3.7. Ancient lineages of poorly-dispersing SREs are ideally suited to test the vast majority of biogeographic scenarios. Exceptions are a) the rapid colonization of an area such as the tundra during the retreat of the LGM, and b) for landscape genetics, for example, when testing the response of organisms to a stream buffer or a corridor connecting reserves.

CONCLUSIONS

This dissertation lays the ground work for a research program that seeks to highlight the important contribution that the study and conservation of leaf-litter dwelling short-range endemic taxa can confer to society. Many new species have been and will continue to be described from the leaf litter of the temperate rain forests of western North America. Discovery, description, and organization via taxonomy is the necessary first step towards furthering the contributions these taxa can make. They are important ecologically, for example they are primary decomposers, they are integral to forest food webs, and they are likely important fungal dispersers. Many of these taxa, such as the harvestmen focal to this research, belong to clades that are among the earliest terrestrial invaders. Thus, since they are poorly dispersing, they have high biogeographical and phylogeographical utility to test hypotheses across a very broad temporal spectrum of evolutionary and geological events. Many of these taxa such as harvestmen and millipedes, produce defensive chemicals which may be of economic value as pesticides. The incredible diversity of locomotive mechanisms are potentially inspirational to future engineers. Species are already immensely important hypotheses to many downstream users, thus these hypotheses must be clearly demarked and organized. This dissertation takes perspectives and necessitated the understanding of ancient phylogenetics, recent population genetics, as well as their confluence at speciation. Time to search for opportunities and seize the moment.

Appendix A. Phylogenetic trees in Newick format

Appendix A.1. Transcriptome Panel 672 loci, 5 taxa.

MP-EST 672 Loci Phylogeny

(Ortholasma:9.0,(Trogulus:9.0,(Hesperonemastoma:9.0,(Acuclavella:9.0,Sabacon:9.0)71.00:0.0673798450704)100.00:3.20657204)100.00:7.0)100.00;

STAR 672 Loci Phylogeny

(Hesperonemastoma:2.71591889877,(Ortholasma:6.02319444516,Trogulus:3.97680555484)100.00:1.25798735055,(Acuclavella:2.69792038713,Sabacon:2.61512276812)69.00:0.0665814780428)

RAxML 672 Loci Phylogeny

((Trogulus:0.31713712622990508638,(Hesperonemastoma:0.44375325746620147838,(Sabacon:0.33392801346094569404,Acuclavella:0.35099390210620096120)97:0.06946422000160606958)100:0.33130866506340439503):0.16427692636329516152,Ortholasma:0.16427692636329516152);

RAxML 420 Loci Hidden Support Analysis

((Trogulus:0.31985234982208721188,((Hesperonemastoma:0.46180003756829723915,Sabacon:0.31844207503686428051)100:0.08351420444972977586,Acuclavella:0.31906069277264104134)100:0.33053939116846636770):0.16291699221664157626,Ortholasma:0.16291699221664157626);

Appendix A.2. Expanded Panel 14 loci, 14 taxa.

Locus 008 RAxML gene tree

(((HesperonemastomaB:0.05781404450440949527,HesperonemastomaA:0.04623648317960571408)100:0.13646575437065960879,((AcuclavellaB:0.00000597380564149358,(CeratolasmaA:0.00645069528565072724,CeratolasmaB:0.03489437338924013948)99:0.10172080215902343237,AcuclavellaA:0.01979548888560083916)55:0.01967292854718188239)100:0.15333298240375309307,SabaconA:0.17077290103798645204)92:0.14021602380665323939)100:0.24045632055223806200,(Trogulus:0.442540106973679125 97,Ortholasma:0.16961603948574210943)100:0.24045632055223806200);

Locus 017 RAxML gene tree

(((AcuclavellaB:0.06527131812911972364,IschyropsalisA:0.09120171274957343532)91:0.3130007827717307678 0,(SabaconA:0.08317703721141189721,((HesperonemastomaA:0.22612429949576021149,HesperonemastomaB:0.037824688750473865 75)100:1.26024484620205390328,SabaconB:0.011160103424893021 15)21:0.043650005603864491 80)29:0.043512851890823454 89)100:0.10655530676319972805,(Trogulus:0.06309650267756343467,Ortholasma:0.06401296545219956058)100:0.10655530676319972805);

Locus 049 RAxML gene tree

((((HesperonemastomaA:0.06174400977729906093,HesperonemastomaB:0.06003864177092328031)100:0.29357594220433225196,(TaracusA:0.07982576448902034549,TaracusB:0.026363242975137247 26)97:0.11833409748764397995)29:0.03629984438057595741,((AcuclavellaA:0.01024678763076031750,(AcuclavellaB:0.01334001238089785844,CeratolasmaA:0.09773995023104187219)58:0.01655572857449489482)100:0.24997775362507404506,(SabaconB:0.08176973250131666648,SabaconA:0.11089655549910519550)100:0.22880034928296116847)40:0.032533434906699805 48)100:0.15589877581719985278,(Trogulus:0.15774554706559187900,Ortholasma:0.80915126751655253035)100:0.15589877581719985278);

Locus 069 RAxML gene tree

(((HesperonemastomaA:0.71937248114738971960,(TaracusA:0.3589166676724961369 0,(SabaconB:0.27631939617866746461,SabaconA:0.24532230131455520783)46:0.0716974030056246464 1)21:0.07196126546945523506)80:0.14354240346101870829,(CeratolasmaB:0.11840011773809339113,(AcuclavellaA:0.01339935224273456141,AcuclavellaB:0.05475098196553396246)93:0.14683246942518313194)98:0.28565668387271536677)97:0.13037786808636778191,(Trogulus:0.21117533693046405818,Ortholasma:0.43831828480924212021)97:0.13037786808636778191);

Locus 156 RAxML gene tree

((((AcuclavellaA:0.02186749452456136431,AcuclavellaB:0.00585307732745875969)89:0.06064426946261591150,CeratolasmaB:0.04125525919966671617)95:0.16038845882395741382,(((TaracusA:0.01789369832489284481,TaracusB:0.06467914681976212610)100:0.46094714845415718507,(HesperonemastomaB:0.07767806127319891452,HesperonemastomaA:0.06829272503867085675)98:0.17781578593953059619)89:0.21879602039075565778,SabaconA:0.35795396071230345392)49:0.07825948462053422539)100:0.33180824416842036317,(Trogulus:0.36709636630440234395,Ortholasma:0.23610636196925238739)100:0.33180824416842036317);

Locus 281 RAxML gene tree

((AcuclavellaB:0.27520388798204131753,(((HesperonemastomaB:0.00726146229263554058,HesperonemastomaA:0.14348030813094081459)100:0.22605859882401094230,(TaracusB:0.02934040536084490253,TaracusA:0.04300217355540306152)100:0.2321491

9472036779635)74:0.13097483380380303686,(SabaconA:0.12855623861343579151,Sa
baconB:0.15003055413205806934)94:0.12492949106314847618)71:0.09139918167936
986193)100:0.15848676120891200259,(Trogulus:0.29105933868560679434,Ortholasma
:0.23042749809809562334)100:0.15848676120891200259);

Locus 300 RAxML gene tree

((SabaconA:0.25018567923217721605,(((TaracusA:0.01288511762958802383,Taracus
B:0.14239519122817712038)100:0.43959460679656620075,HesperonemastomaA:1.115
51491689033843713)44:0.16327827513638282220,((CeratolasmaA:0.02655601681556
859356,CeratolasmaB:0.045051406160805877 73)94:0.20845705802068956181,(Acucla
vellaA:0.00136831884518079394,AcuclavellaB:0.01585920060326828640)90:0.155806
94828203628388)69:0.18770156287190162270)39:0.09493584344334241032)98:0.200
74177282700644831,(Ortholasma:0.36388537046451885359,Trogulus:0.368408291004
84029368)98:0.20074177282700644831);

Locus 334 RAxML gene tree

((((TaracusA:0.12329131536048561213,TaracusB:0.13041122041639308105)97:0.2456
5326985016455730,(HesperonemastomaB:0.14925772242101220666,Hesperonemastom
aA:0.17025945164125749764)100:0.588776849127363610 67)35:0.08838301682215114
752,(SabaconA:0.74306594515692636627,AcuclavellaB:0.566538966029826229 76)28:
0.09495995496577083750)100:0.24482738303456563767,(Trogulus:0.47957878324506
725276,Ortholasma:0.45809796580518674647)100:0.24482738303456563767);

Locus COI RAxML gene tree

((Trogulus:2.90566235155673036061,(((SabaconB:1.03949630636825540719,SabaconA
:0.72580372647443569534)50:0.42860259079511364710,(HesperonemastomaA:0.5821
0879950023730611,(HesperonemastomaB:0.36130041498008025336,TaracusB:0.95015
641139312922991)61:0.21011685183426101475)100:0.904386204778614311 02)31:0.2
6874008749396033346,((CeratolasmaA:0.72496048806012147736,(AcuclavellaA:0.598
23400000820725619,AcuclavellaB:0.56931326833058726322)57:0.1186335881469422
3431)91:0.45272915346169795292,(IschyropsalisA:0.30844040258552513079,Ischyrop
salisB:0.66731683921289153272)84:0.64265989833812875442)53:0.292032587819458
51119)17:0.49141121181272084950):0.59402940232195478032,Ortholasma:0.5940294
0232195478032);

Locus EF1a RAxML gene tree

((((SabaconB:0.08149278141999426184,SabaconA:0.09930662825552367956)50:0.047
51995625225352571,((HesperonemastomaB:0.09642047926920173373,Hesperonemasto
maA:0.16489292092902876896)89:0.05670145922529670757,(TaracusB:0.0521336533
1049985701,TaracusA:0.07740405478861704658)100:0.191241927901239838 76)43:0.0
6179921335204217187)32:0.03945290498443915911,((IschyropsalisB:0.018391534672

84129415,IschyropsalisA:0.027010659969705190 51)98:0.05189028597554867089,((AcuclavellaA:0.01942147370607493623,AcuclavellaB:0.00134094067527473965)100:0.07919663547181929331,(CeratolasmaA:0.01804724308393024224,CeratolasmaB:0.01391331178767637490)99:0.05174357021115311150)99:0.04045920360891640255)99:0.11502089915438475087)87:0.07975957900718352866,(Trogulus:0.324861900693817817 75,Ortholasma:0.140122351540415163 16)87:0.07975957900718352866);

Locus polII RAxML gene tree

(((HesperonemastomaA:1.57858932586511802043,SabaconA:1.2967035420940991041 8)46:0.38447861719024317706,((IschyropsalisA:0.29112284080384376317,(AcuclavellaB:0.22043034127858152083,CeratolasmaA:0.27888259750311106266)99:0.80846510350890854468)68:0.35031351471543081821,TaracusB:0.88689514316665019056)43:0.1371373858079104335 3)97:0.48124858725647734481,(Ortholasma:1.328976172302024 36112,Trogulus:0.710684122231292958 46)97:0.48124858725647734481);

Locus r18S RAxML gene tree

(((CeratolasmaA:0.00195454398639851451,CeratolasmaB:0.00185984613780147017)100:0.00802616133792437292,((SabaconA:0.00178071019135267418,SabaconB:0.00591684671897571915)100:0.01005977246333684093,(((TaracusB:0.0000010303204936627 2,TaracusA:0.0000010303204936627 2)92:0.00180605225703031396,(HesperonemastomaA:0.0000010303204936627 2,HesperonemastomaB:0.0000010303204936627 2)98:0.005885279000232316 11)54:0.00188793840614184544,(IschyropsalisA:0.0198038880445 8587879,(AcuclavellaB:0.00575225458811981071,AcuclavellaA:0.00000103032049366 272)92:0.00196155184669587245)51:0.00472768097276583810)12:0.00174072543615 857653)45:0.0000010303204936627 2)98:0.01259185801590672824,(Ortholasma:0.0074 0135343742243969,Trogulus:0.008570888819601328 24)98:0.01259185801590672824)

Locus r28S RAxML gene tree

(((TaracusA:0.07835584886152616169,(HesperonemastomaA:0.027811371929464659 41,HesperonemastomaB:0.01960819105041602603)100:0.05418687556559304008)100:0.10714868297737101255,((SabaconB:0.04827840592013879495,SabaconA:0.05148866 758802764926)100:0.08046368419780947079,((IschyropsalisA:0.016713021260661188 59,IschyropsalisB:0.00436109214417934512)100:0.03407755615649212672,((AcuclavellaB:0.02077887262384393516,AcuclavellaA:0.00530403030917979152)84:0.00680230 541044292945,CeratolasmaA:0.02279297034069643901)64:0.01155347892829227015)95:0.04897528575339116230)43:0.01829086655753124624)100:0.07905640068638922 213,(Ortholasma:0.14155765684169932528,Trogulus:0.05148324046403470605)100:0.07905640068638922213);

Locus Wnt2 RAxML gene tree

(((TaracusB:0.04358751921901943965,TaracusA:0.03960114513655649260)80:0.06817634033271495475,((HesperonemastomaB:0.06754797076358839969,HesperonemastomaA:0.11670541537689539668)97:0.11957719018590859394,((CeratolasmaB:0.01947248539993706035,((AcuclavellaB:0.00154594005907381505,AcuclavellaA:0.00867974819203037526)85:0.04250788251906752191,IschyropsalisA:0.13753215386186701252)40:0.048546628642518192 82)94:0.15294253510673402885,(SabaconB:0.09372740380533395310,SabaconA:0.11134716993066146040)100:0.15480587161740533286)53:0.09513723514571637008)74:0.032186316001187952 05):0.35989275003397774855,Trogulus:0.35989275003397774855);

Concatenated RAxML phylogeny run 1

((((HesperonemastomaB:14.06948344789248395159,HesperonemastomaA:22.26705053069550643841)100:49.60998128217789115979,(TaracusA:11.54031106237080805954,TaracusB:10.93183326755467454916)100:33.704538426287179220 25)100:26.627906310267910328 09,((SabaconA:24.94986179351228372525,SabaconB:23.63065767783087167686)100:32.06855723684248715699,((IschyropsalisB:5.45756855971684284867,IschyropsalisA:8.82369520397149109669)100:18.39025723694578573486,((AcuclavellaB:4.53484658152330588621,AcuclavellaA:4.30233332937969947807)100:10.20571981813143835893,(CeratolasmaA:2.98291348160444735882,CeratolasmaB:5.86324437407421417134)100:13.331095214398816395 37)99:9.694488793542346272 15)100:32.37169178778008671316)68:7.162862052354530106 87)100:32.80515539090488630336,(Orthol asma:58.79955969197761334044,Trogulus:49.87954579757406037288)100:32.80515539090488630336);

Concatenated RAxML phylogeny run 2

((((HesperonemastomaB:14.06948344789248395159,HesperonemastomaA:22.26705053069550643841)100:49.60998128217789115979,(TaracusA:11.54031106237080805954,TaracusB:10.93183326755467454916)100:33.704538426287179220 25)100:26.627906310267910328 09,((SabaconA:24.94986179351228372525,SabaconB:23.63065767783087167686)100:32.06855723684248715699,((IschyropsalisB:5.45756855971684284867,IschyropsalisA:8.82369520397149109669)100:18.39025723694578573486,((AcuclavellaB:4.53484658152330588621,AcuclavellaA:4.30233332937969947807)100:10.20571981813143835893,(CeratolasmaA:2.98291348160444735882,CeratolasmaB:5.86324437407421417134)100:13.331095214398816395 37)99:9.694488793542346272 15)100:32.37169178778008671316)68:7.162862052354530106 87)100:32.80515539090488630336,(Orthol asma:58.79955969197761334044,Trogulus:49.87954579757406037288)100:32.80515539090488630336);

Concatenated RAxML phylogeny run 3

((((HesperonemastomaB:14.06948344789248395159,HesperonemastomaA:22.26705053069550643841)100:49.60998128217789115979,(TaracusA:11.54031106237080805954,TaracusB:10.93183326755467454916)100:33.704538426287179220 25)100:26.6279063

1026791032809,((SabaconA:24.94986179351228372525,SabaconB:23.63065767783087167686)100:32.06855723684248715699,((IschyropsalisB:5.45756855971684284867,IschyropsalisA:8.82369520397149109669)100:18.390257236945785 73486,((AcuclavellaB:4.53484658152330588621,AcuclavellaA:4.302333329379699478 07)100:10.20571981813143835893,(CeratolasmaA:2.982913481604447358 82,CeratolasmaB:5.863244374074214171 34)100:13.331095214398816395 37)99:9.694488793542346272 15)100:32.37169178778008671316)68:7.162862052354530106 87)100:32.805155390904886303 36,(Ortholasma:58.799559691977613340 44,Trogulus:49.879545797574060372 88)100:32.805155390904886303 36);

Concatenated MrBayes phylogeny

(Ortholasma:0.3770732617807573,Trogulus:0.3304396485003028,(((HesperonemastomaA:0.1524050526879162,HesperonemastomaB:0.09399006205575164)100:0.3070794552950473,(TaracusA:0.07876908141364744,TaracusB:0.07318725714855423)100:0.2137294784164536)100:0.1615533578766353,((SabaconA:0.1688439411314099,SabaconB:0.1620669212044138)100:0.2003203601558221,(((AcuclavellaA:0.03773869731433995,AcuclavellaB:0.04193446464500469)100:0.07437045773734205,(CeratolasmaA:0.02486118841128121,CeratolasmaB:0.03896826506646361)100:0.09492510064404786)100:0.07154255548723766,(IschyropsalisA:0.07528312908313238,IschyropsalisB:0.05690802975804672)100:0.1362915833627291)100:0.1918451064901203)100:0.05092437936269988)100:0.3909099636364343);

*BEAST species tree

((((Sabacon2:0.08082843423805999,Sabacon1:0.08082843423805999)0.9811:0.12142006379721995,(((Ceratolasma1:0.010894854343758975,Ceratolasma2:0.010894854343758975)0.9805:0.032165695024493,(Acuclavella2:0.004545539261849016,Acuclavella1:0.004545539261849016)0.9898:0.038515010106402015)0.9725:0.04389753287128299,(Ischyropsalis2:0.019094852979719995,Ischyropsalis1:0.019094852979719995)0.9847:0.06786322925981497)0.9615:0.11529041579574498)0.8207:0.01737815306191004,(Hesperonemastoma2:0.04009138737664,Hesperonemastoma1:0.04009138737664)0.9843:0.12380742014669,(Taracus2:0.022035089035289945,Taracus1:0.022035089035289945)0.9829:0.14186371848804005)0.9616:0.055727843573859986)0.9753:0.08167940366473991,(Ortholasma1:0.17819295422634704,Trogulus2:0.17819295422634704)0.9749:0.12311310053558291)1;

Concatenated RAxML Reduced Panel phylogeny

((Ortholasma:0.99258964906713587073,Trogulus:1.7094422983319950137 8),(SabaconA:1.85270684985780786924,(AcuclavellaB:1.664194001260155042 49,HesperonemastomaA:3.360264719944723310 62)52:0.32143449058520717765)100:2.45409113934446843075):0.99258964906713587073;

Concatenated MrBayes Reduced Panel phylogeny

(Ortholasma:0.2925759218444314,Trogulus:0.2622511260776455,(HesperonemastomaA:0.4585668068842176,(SabaconA:0.2855766873832775,AcuclavellaB:0.24903450674 1622)49:0.04643007470520056)100:0.3097482215135367);

Concatenated *BEAST Reduced Panel phylogeny

((Hesperonemastoma1:0.30224581610609,(Acuclavella2:0.23269153586910002,Sabacon1:0.23269153586910002)0.9777:0.06955428023698995)0.9881:0.08943091528872399,(Ortholasma1:0.21993913634538997,Trogulus2:0.21993913634538997)0.989:0.171737 59504942404)1;

Concatenated unpartitioned PhyML translated AA phylogeny

(Trogulus:0.0624218,Ortholasma:0.0705466,((SabaconA:0.0289665,SabaconB:0.014059 1)100:0.0224142,(((HesperonemastomaA:0.0324936,HesperonemastomaB:0.01708111)100:0.0672445,(TaracusA:0.0200726,TaracusB:0.0163463)100:0.0310589)100:0.0252002,(IschyropsalisB:0.0050992,((AcuclavellaA:0.0043280,AcuclavellaB:0.0047011)82:0.0066907,(CeratolasmaA:0.0000001,(IschyropsalisA:0.0175554,CeratolasmaB:0.0000001)41:0.0019397)54:0.0066142)52:0.0152594)94:0.0249230)52:0.0131327)100:0.0925769);

Appendix B. Specimens and localities. Locality in decimal degrees (dd), sex, personal identification number, SDSU Biodiversity Museum Opiliones Molecular Collection number, and indication of inclusion in morphometric and genomic analyses.

Species	dd lat	dd long	Sex	CHR	OP	Morph	ddRAD
			m	CHR2070.1		x	
			f	CHR2070.2		x	
			m	CHR2070.3		x	
			m	CHR2070.4		x	
			f	CHR2070.5		x	
			m	CHR2070.6		x	
			m	CHR2070.7		x	
A.cf.quattuor	46.5389	-114.6762	m	CHR2070.8		x	
			m	CHR2070.9		x	
			m	CHR2073	OP2230	x	x
			m	CHR2074	OP2231	x	
			m	CHR2075	OP2232	x	x
			f	CHR2445.0		x	
			f	CHR2445.1		x	
			f	CHR2445.2		x	
			f	CHR2076.0		x	
			f	CHR2076.1		x	
			f	CHR2076.2		x	
A.cf.quattuor	46.4292	-115.1335	f	CHR2082	OP2233	x	x
			f	CHR2083	OP2234	x	x
			f	CHR2084	OP2235	x	
			m	CHR2192.0		x	
			m	CHR2192.1		x	
			m	CHR2192.2		x	
			m	CHR2192.3		x	
A.cf.quattuor	46.2311	-115.4161	m	CHR2192.4		x	
			f	CHR2192.5		x	
			m	CHR2196	OP2275	x	x
			m	CHR2197	OP2276	x	x
			m	CHR2198	OP2277	x	x
A.cf.quattuor	46.0848	-115.4955	f	CHR2221	OP2283	x	

Species	dd lat	dd long	Sex	CHR	OP	Morph	ddRAD
A.cf.quattuor	46.0498	-115.3013	f	CHR2227.0		x	
			f	CHR2227.1		x	
			f	CHR2227.2		x	
			f	CHR2229	OP2284	x	x
			m	CHR2230	OP2285	x	
			m	CHR2231	OP2286	x	
A.cf.quattuor	46.1025	-115.5555	f	CHR2237	OP2287	x	x
A.cf.quattuor	46.4678	-114.9854	f	CHR2446	OP2342	x	x
A.cf.quattuor	46.5563	-114.6635	f	CHR3286	OP2711	x	x
A.cosmetoides	46.2101	-115.5442	m	CHR2200.0		x	
			m	CHR2200.1		x	
			m	CHR2200.2		x	
			m	CHR2200.3		x	
			m	CHR2200.4		x	
			f	CHR2200.5		x	
			m	CHR2202	OP2278	x	
			m	CHR2203	OP2279	x	x
			f	CHR2204	OP2280	x	x
A.cosmetoides	46.1396	-115.6660	m	CHR2207.0		x	
			f	CHR2207.1		x	
			m	CHR2209	OP2281	x	x
			f	CHR2210	OP2282	x	
A.cosmetoides	46.2395	-115.7691	m	CHR2241	OP2288	x	x
			m	CHR2242	OP2289	x	x
			m	CHR2243	OP2290	x	x
			f	CHR2245	OP2292	x	
			m	CHR2248.0		x	
			m	CHR2248.1		x	
			m	CHR2248.2		x	
			f	CHR2248.3		x	
			f	CHR2248.4		x	
			f	CHR2248.5		x	
A.cosmetoides	46.3721	-115.7235	m	CHR2257	OP2293	x	
			f	CHR2258	OP2294	x	x
			f	CHR2259	OP2295	x	

Species	dd lat	dd long	Sex	CHR	OP	Morph	ddRAD
A.cosmetoides	46.4767	-115.7809	m	CHR2264.0		x	
			m	CHR2264.1		x	
			m	CHR2264.2		x	
			f	CHR2264.3		x	
			m	CHR2266	OP2296	x	x
			f	CHR2267	OP2297	x	x
			f	CHR2268	OP2360	x	x
A.cosmetoides	46.5105	-115.6941	f	CHR2274	OP2298		x
A.cosmetoides	46.5836	-115.6164	f	CHR2280.0		x	
			f	CHR2280.1		x	
			f	CHR2280.2		x	
			f	CHR2282	OP2299	x	x
			f	CHR2283	OP2300	x	x
			f	CHR2284	OP2301	x	x
A.cosmetoides	46.7395	-115.5444	f	CHR2321	OP2311	x	x
			f	CHR2322	OP2312	x	x
			f	CHR2323	OP2313	x	x
			f	CHR2324	OP2314	x	
A.cosmetoides	46.8093	-115.6156	f	CHR2328	OP2315	x	x
Hoodoo Mtns	47.0351	-116.6725	m	CHR1409		x	
			f	CHR1410		x	
			f	CHR1411		x	
			f	CHR1427	OP1645	x	
			f	CHR1444	OP1634	x	
			m	CHR2341.0		x	
			m	CHR2341.1		x	
			f	CHR2341.2		x	
			m	CHR2343	OP2317	x	x
			m	CHR2344	OP2318	x	x
Hoodoo Mtns	46.9873	-116.4199	f	CHR2338	OP2316	x	x
Hoodoo Mtns	47.0352	-116.7896	m	CHR2348.0		x	
			m	CHR2348.1		x	
			f	CHR2348.2		x	
			m	CHR2351	OP2319	x	x
			m	CHR2352	OP2320	x	x
			m	CHR2353	OP2321	x	x
			f	CHR2354	OP2322	x	
Hoodoo Mtns	46.8014	-116.9495	u	CHR4653	OP4063		x

Species	dd lat	dd long	Sex	CHR	OP	Morph	ddRAD
			f	CHR2096	OP2236	x	x
			m	CHR2100.0		x	
			m	CHR2100.1		x	
			f	CHR2100.2		x	
A.merickeli	46.0385	-115.2943	f	CHR2101	OP2237	x	x
			f	CHR2102	OP2238	x	x
			m	CHR2103	OP2239	x	
			m	CHR2104	OP2240	x	
			f	CHR2108	OP2241	x	
			m	CHR2121.0		x	
			m	CHR2121.1		x	
			m	CHR2121.2		x	
			m	CHR2121.3		x	
			f	CHR2121.4		x	
A.merickeli	46.0559	-115.5195	f	CHR2121.5		x	
			f	CHR2121.6		x	
			m	CHR2122	OP2245	x	x
			m	CHR2123	OP2246	x	x
			m	CHR2124	OP2247	x	
			f	CHR2125	OP2248	x	
			f	CHR2126	OP2249	x	
A.merickeli	45.7939	-115.4006	m	CHR2130	OP2250	x	x
			m	CHR2134	OP2251	x	
			m	CHR2135	OP2252	x	x
A.merickeli	45.7853	-115.2026	m	CHR2136	OP2253	x	x
			m	CHR2137	OP2254	x	
			m	CHR2140.1		x	
			f	CHR2140.2		x	
			m	CHR2161.0		x	
			m	CHR2161.1		x	
			f	CHR2161.2		x	
			f	CHR2161.3		x	
A.merickeli	45.8437	-115.6178	f	CHR2161.4		x	
			f	CHR2161.5		x	
			f	CHR2163	OP2262	x	x
			f	CHR2164	OP2263	x	x
			f	CHR2165	OP2264	x	x
			f	CHR2166	OP2265	x	

Species	dd lat	dd long	Sex	CHR	OP	Morph	ddRAD
A.quattuor	45.8120	-115.9530	m	CHR2117	OP2242	x	x
			f	CHR2118	OP2243	x	x
			m	CHR2119	OP2244	x	
A.quattuor	45.5791	-115.4431	f	CHR2142	OP2255	x	x
A.quattuor	45.6385	-116.1096	m	CHR2146.1		x	
			f	CHR2146.2		x	
			m	CHR2146.3		x	
			m	CHR2146.4		x	
			m	CHR2146.5		x	
			m	CHR2146.6		x	
			m	CHR2146.7		x	
			f	CHR2148	OP2256	x	x
			f	CHR2149	OP2257	x	x
			f	CHR2150	OP2258	x	x
			f	CHR2152	OP2260	x	
			f	CHR2153	OP2261	x	
A.quattuor	45.6853	-115.5427	m	CHR2171	OP2266	x	x
			f	CHR2172	OP2267	x	x
A.quattuor	45.6099	-116.0300	m	CHR2180.0		x	
			m	CHR2180.1		x	
			m	CHR2180.2		x	
			f	CHR2180.3		x	
			f	CHR2180.4		x	
			f	CHR2182	OP2270	x	x
			f	CHR2183	OP2271	x	x
			f	CHR2184	OP2272	x	
			f	CHR2185	OP2273	x	
			f	CHR2186	OP2274	x	
A.quattuor	45.6338	-115.4177	m	CHR2176.0		x	
			m	CHR2176.1		x	
			f	CHR2178	OP2268	x	
			f	CHR2179	OP2269		x
A.sheari	45.3902	-115.9877	m	CHR3236.0		x	
			f	CHR3236.1		x	
			f	CHR3248.0	OP2710	x	
A.sheari	45.3723	-116.0220	m	CHR3253.0		x	
			m	CHR3253.1		x	
			m	CHR3254	OP2708	x	x

Species	dd lat	dd long	Sex	CHR	OP	Morph	ddRAD
A.sheari	45.3822	-115.9779	f	CHR3404	OP2720		x
A.shoshone	46.8776	-115.0889	m	CHR2293	OP2302	x	x
			f	CHR2294	OP2303	x	x
A.shoshone	46.7192	-115.2317	f	CHR2302	OP2304	x	x
			m	CHR2303	OP2305	x	x
			f	CHR2304	OP2306	x	x
			m	CHR2309.0		x	
			m	CHR2309.1		x	
			f	CHR2309.2		x	
			f	CHR2309.3		x	
A.shoshone	46.7331	-115.3111	m	CHR2312.0		x	
			m	CHR2312.1		x	
			m	CHR2312.2		x	
			f	CHR2312.3		x	
			f	CHR2312.4		x	
			m	CHR2313	OP2307	x	x
			m	CHR2314	OP2308	x	x
			m	CHR2315	OP2309	x	x
			f	CHR2316	OP2310	x	
A.shoshone	47.0861	-116.1129	f	CHR1492	OP1648	x	
			m	CHR2359.0		x	
			f	CHR2359.1		x	
			f	CHR2362	OP2323	x	x
			f	CHR2363	OP2324	x	x
A.shoshone	47.2436	-116.0500	m	CHR2368.0		x	
			m	CHR2368.1		x	
			f	CHR2368.2		x	
			f	CHR2368.3		x	
			m	CHR2370	OP2325	x	x
			m	CHR2371	OP2326	x	x
			f	CHR2372	OP2327	x	x
A.shoshone	47.2236	-115.6072	m	CHR2379.0		x	
			f	CHR2379.1		x	
			f	CHR2379.2		x	
			f	CHR2381	OP2328	x	x
			f	CHR2382	OP2329	x	

Species	dd lat	dd long	Sex	CHR	OP	Morph	ddRAD
A.shoshone	47.2834	-115.7738	f	CHR2385	OP2330	x	
			m	CHR2386	OP2331	x	
A.shoshone	47.0802	-115.3548	m	CHR2389	OP2332	x	x
			f	CHR2390	OP2333	x	x
A.shoshone	47.2122	-115.5484	f	CHR2403.0		x	
A.shoshone	47.2296	-115.5406	f	CHR2411	OP2334	x	x
A.shoshone	47.1296	-115.8819	m	CHR2414.0		x	
			m	CHR2414.1		x	
			m	CHR2414.2		x	
			m	CHR2414.3		x	
			f	CHR2414.4		x	
			m	CHR2416	OP2335	x	x
			m	CHR2417	OP2336	x	x
			m	CHR2418	OP2337	x	x
			m	CHR2419	OP2338	x	
			f	CHR2420	OP2339	x	
A.shoshone	47.4304	-115.8913	f	CHR2437	OP2340		x
A.shoshone	47.2279	-115.2464	m	CHR2441	OP2341	x	x
totals:	48 total localities			233 total individuals		228	82

Footnote: The localities represented with morphological data are also represented with ddRADseq data with three exceptions. First, a northern locality (47.2834°, -115.7738°) that is surrounded by samples all relatively closely related in phylogenetic and population genetic analyses. Second, a locality (46.0848°, -115.4955°) found along the north shore of the Selway River, which is bracketed to the east and west along the north shore by localities containing specimens of similar appearance. Third, a locality (45.3902°, -115.9877°) very near other *A. sheari* samples, with only three total known localities for this species all in a small area south of the Salmon River.

Appendix C. pyRAD curated data sets.

Data Set	*N*	MinCov	Coverage	# of loci final data set	# of loci in all samples	# unlinked SNPs	Analyses
All Inland Acuclavella	82	22	26.8%	4525	13	4474	RAxML BEAST SVDQuartets DAPC Structure I
Acfquattuor	11	6	54.5%	5918	202	5388	Structure III DAPC nested
AcfquattuorA	7	4	57.1%	7670	3010	6352	Structure IV
AcfquattuorB	4	2	50.0%	10792	4967	5442	Structure IV
Ashoshone	22	11	50.0%	5552	852	5158	Structure III DAPC nested
Asheari, Acfquattuor, Acosmetoides	31	16	51.6%	2905	46	2829	Structure II
Hoodoo Mtns, Ashoshone	29	15	51.7%	4200	432	4021	Structure II
Acosmetoides	18	9	50.0%	5560	788	5125	Structure III DAPC nested
AcosmetoidesB	10	5	50.0%	7687	1498	6645	Structure IV
AcosmetoidesA	8	4	50.0%	10250	2990	7977	Structure IV
Hoodoo Mtns	7	4	57.1%	7394	2595	5237	Structure III DAPC nested
AcosmetoidesA2	2	2	100.0%	6493	6493	4042	Structure V
AcosmetoidesB1	6	3	50.0%	9891	2729	7427	Structure V
AcosmetoidesB2	4	2	50.0%	12538	4090	7863	Structure V
AcfquattuorA2	4	2	50.0%	13997	4378	8778	Structure V
Asheari	2	2	100.0%	6391	6391	1640	Structure III
Aquattuor, Americkeli	22	11	50.0%	5984	1220	5386	Structure II DAPC nested

Appendix D. Statistics from sequencing an pyRAD demultiplexing, quality control, and within-sample clustering of 95% including raw reads and reads passed loci counts, the percent of reads that passed quality control, the number of loci recovered, estimated heterozygosity, and error rate.

Sample	# Raw reads	# Reads passed	% Passed	# of Loci	Est. heterozyg.	Error rate
OP2230	1304700	1223169	0.938	66483	0.00498	0.00029
OP2232	601193	554646	0.923	57307	0.00636	0.00098
OP2233	1326221	1243547	0.938	69745	0.00603	0.00030
OP2234	79045	72743	0.920	22374	0.00890	0.00091
OP2236	1735431	1620095	0.934	86977	0.00581	0.00034
OP2237	2136596	1985701	0.929	75177	0.00503	0.00029
OP2238	701188	658856	0.940	54272	0.00679	0.00065
OP2242	2521134	2276471	0.903	115189	0.00737	0.00050
OP2243	1592258	1469271	0.923	100159	0.00713	0.00076
OP2245	1715038	1602615	0.934	115253	0.00586	0.00074
OP2246	1216074	1130528	0.930	70727	0.00635	0.00059
OP2250	3693117	3442221	0.932	129308	0.00625	0.00032
OP2252	1835410	1705391	0.929	112009	0.00586	0.00065
OP2253	1173267	1095904	0.934	69188	0.00590	0.00057
OP2255	1089712	1024520	0.940	52306	0.00518	0.00052
OP2256	1244641	1106561	0.889	113132	0.00730	0.00103
OP2257	1218357	1111526	0.912	62903	0.00559	0.00050
OP2258	2031273	1845512	0.909	74190	0.00491	0.00040
OP2262	1005657	950656	0.945	59487	0.00615	0.00059
OP2263	1456655	1365664	0.938	70795	0.00621	0.00054
OP2264	1002237	922111	0.920	101623	0.00588	0.00086
OP2266	1928007	1810144	0.939	75505	0.00628	0.00019
OP2267	1525097	1440689	0.945	79190	0.00583	0.00029
OP2269	894197	837576	0.937	48881	0.00410	0.00039
OP2270	2903412	2784978	0.959	144256	0.00690	0.00017
OP2271	1170186	1083187	0.926	65764	0.00646	0.00042
OP2275	1324466	1183784	0.894	91268	0.00644	0.00065
OP2276	1648657	1550143	0.940	77597	0.00685	0.00032
OP2277	1590877	1474116	0.927	85337	0.00683	0.00047
OP2279	1189459	1069244	0.899	84009	0.00735	0.00098
OP2280	710244	667182	0.939	64263	0.00579	0.00045
OP2281	1832044	1718659	0.938	97004	0.00649	0.00024
OP2284	1979199	1854258	0.937	106889	0.00624	0.00026

Sample	# Raw reads	# Reads passed	% Passed	# of Loci	Est. heterozyg.	Error rate
OP2287	1297277	1212137	0.934	91430	0.00654	0.00049
OP2288	601211	565433	0.940	54286	0.00607	0.00065
OP2289	1084539	1003132	0.925	71559	0.00625	0.00037
OP2290	745094	693654	0.931	61286	0.00627	0.00073
OP2294	1232996	1135361	0.921	71427	0.00604	0.00058
OP2296	1165511	1100375	0.944	69771	0.00604	0.00037
OP2297	676306	638968	0.945	50850	0.00555	0.00045
OP2298	663430	619980	0.935	55719	0.00578	0.00070
OP2299	1578829	1476470	0.935	90983	0.00624	0.00026
OP2300	1078071	1004953	0.932	70734	0.00620	0.00063
OP2301	1969267	1819030	0.924	98050	0.00625	0.00053
OP2302	607981	571750	0.940	44397	0.00557	0.00055
OP2303	1246075	1157670	0.929	88091	0.00573	0.00040
OP2304	1408632	1300209	0.923	82558	0.00600	0.00072
OP2305	1529634	1435171	0.938	87404	0.00623	0.00046
OP2306	645610	610959	0.946	46091	0.00572	0.00075
OP2307	1149063	1046070	0.910	71483	0.00655	0.00041
OP2308	1324604	1229064	0.928	81103	0.00580	0.00056
OP2309	1242568	1159996	0.934	74645	0.00667	0.00050
OP2311	1122921	1054603	0.939	74930	0.00515	0.00031
OP2312	1830918	1723221	0.941	78943	0.00489	0.00024
OP2313	1035508	963750	0.931	74006	0.00543	0.00065
OP2315	1456852	1339657	0.920	102280	0.00593	0.00075
OP2316	571897	535065	0.936	40637	0.00512	0.00060
OP2317	1037762	956811	0.922	62781	0.00498	0.00042
OP2318	2136593	1976762	0.925	110414	0.00477	0.00047
OP2319	599838	548699	0.915	51375	0.00402	0.00090
OP2320	1391369	1277912	0.918	82262	0.00471	0.00052
OP2321	1126411	1053454	0.935	76835	0.00373	0.00026
OP2323	1796940	1618763	0.901	83841	0.00537	0.00061
OP2324	1124265	1053503	0.937	55772	0.00496	0.00034
OP2325	941299	860119	0.914	63203	0.00538	0.00060
OP2326	1299863	1199208	0.923	68853	0.00526	0.00038
OP2327	813433	730953	0.899	68586	0.00653	0.00089
OP2328	1329070	1256138	0.945	63723	0.00513	0.00029
OP2332	1029221	960507	0.933	66412	0.00576	0.00022
OP2333	1473311	1362930	0.925	84386	0.00559	0.00054

Sample	# Raw reads	# Reads passed	% Passed	# of Loci	Est. heterozyg.	Error rate
OP2334	933889	877664	0.940	56374	0.00473	0.00043
OP2335	2461053	2361300	0.959	92868	0.00489	0.00013
OP2336	1210852	1125391	0.929	70413	0.00496	0.00058
OP2337	1998765	1883830	0.942	105627	0.00459	0.00031
OP2340	1377100	1290604	0.937	91756	0.00263	0.00030
OP2341	1212908	1137387	0.938	59708	0.00436	0.00023
OP2342	969779	902981	0.931	65168	0.00533	0.00044
OP2360	1015724	958953	0.944	63621	0.00601	0.00030
OP2708	902648	802288	0.889	75143	0.00554	0.00090
OP2711	1161511	1074688	0.925	68191	0.00461	0.00035
OP2720	867768	819602	0.944	60911	0.00411	0.00054
OP4063	1360328	1255989	0.923	72525	0.00375	0.00048

Appendix E. Statistics from pyRAD with between sample clustering of 95% (n=82), minimum coverage 22 samples, and minimum depth of 6 reads, including the number and frequency of polymorphisms, count of shared loci, and count of sampled unlinked SNPs.

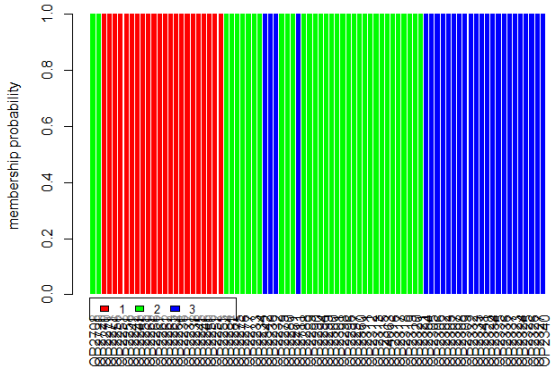
Sample	# Polymorphs.	Freq. of polymorphs.	Loci shared	Sampled unlinked SNPs
OP2230	3794	0.00329	1920	1895
OP2232	3858	0.00425	1683	1666
OP2233	4952	0.00447	1790	1770
OP2234	533	0.00556	127	123
OP2236	4025	0.00357	2077	2049
OP2237	4061	0.00355	2125	2097
OP2238	3607	0.00405	1738	1713
OP2242	7628	0.00434	2881	2842
OP2243	5352	0.00397	2312	2277
OP2245	11627	0.00407	2796	2758
OP2246	4091	0.00394	2033	2003
OP2250	8277	0.00389	2750	2706
OP2252	11043	0.00395	2780	2743
OP2253	3585	0.00357	1968	1941
OP2255	3001	0.00291	2017	1991
OP2256	12076	0.00494	2665	2620
OP2257	3839	0.00362	1910	1884
OP2258	4189	0.00362	2071	2047
OP2262	3617	0.00371	1900	1875
OP2263	3954	0.00357	2074	2047
OP2264	8744	0.00394	2622	2583
OP2266	4431	0.00366	2288	2260
OP2267	4079	0.00365	2145	2116
OP2269	3003	0.00308	1916	1892
OP2270	5564	0.00411	2107	2074
OP2271	4357	0.00413	1902	1878
OP2275	6116	0.00438	1599	1581
OP2276	6569	0.00489	2002	1980
OP2277	5490	0.00446	1916	1896
OP2279	6528	0.00488	2058	2039
OP2280	4727	0.00431	1731	1716
OP2281	6818	0.00482	2222	2195
OP2284	6298	0.00452	2132	2106

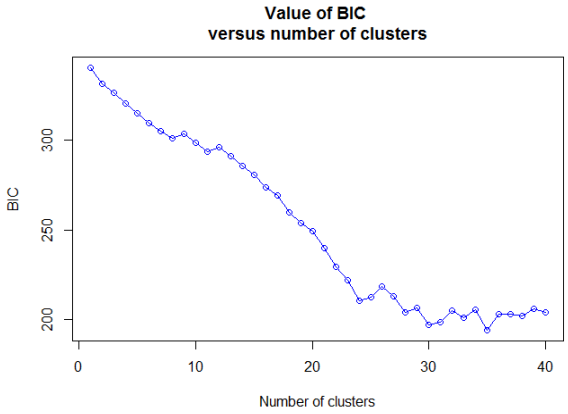
Sample	# Polymorphs.	Freq. of polymorphs.	Loci shared	Sampled unlinked SNPs
OP2287	13059	0.00523	2456	2429
OP2288	4404	0.00466	1552	1536
OP2289	4490	0.00417	1830	1812
OP2290	4505	0.00454	1697	1682
OP2294	4723	0.00405	1902	1881
OP2296	6137	0.00460	1956	1933
OP2297	4138	0.00442	1448	1476
OP2298	4376	0.00439	1581	1566
OP2299	6501	0.00470	2079	2052
OP2300	4663	0.00434	1737	1717
OP2301	6246	0.00433	2360	2331
OP2302	2914	0.00390	1448	1438
OP2303	4330	0.00381	2053	2036
OP2304	5037	0.00422	2231	2208
OP2305	6310	0.00461	2148	2129
OP2306	3808	0.00445	1612	1599
OP2307	4599	0.00441	1956	1930
OP2308	4850	0.00415	2152	2132
OP2309	5744	0.00431	2299	2279
OP2311	4652	0.00362	1773	1747
OP2312	6002	0.00369	2138	2109
OP2313	3937	0.00350	1653	1632
OP2315	13116	0.00464	2407	2380
OP2316	2858	0.00404	997	990
OP2317	2713	0.00321	1191	1183
OP2318	5327	0.00325	1952	1928
OP2319	2004	0.00261	1171	1161
OP2320	3064	0.00283	1592	1569
OP2321	2568	0.00244	1482	1471
OP2323	4809	0.00348	2330	2304
OP2324	3419	0.00341	1827	1815
OP2325	3425	0.00348	1812	1794
OP2326	3848	0.00338	2085	2067
OP2327	4823	0.00441	2027	2014
OP2328	3986	0.00350	2057	2041
OP2332	3622	0.00370	1802	1785
OP2333	3878	0.00348	2028	2008

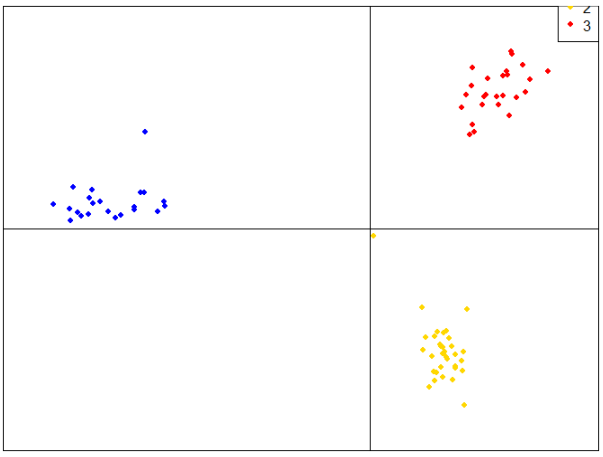
Sample	# Polymorphs.	Freq. of polymorphs.	Loci shared	Sampled unlinked SNPs
OP2334	3236	0.00327	1898	1885
OP2335	4045	0.00328	1990	1970
OP2336	3764	0.00323	2053	2033
OP2337	10479	0.00338	2795	2767
OP2340	4790	0.00184	2596	2569
OP2341	2911	0.00287	1990	1975
OP2342	3486	0.00350	1730	1714
OP2360	4959	0.00453	1723	1701
OP2708	3712	0.00351	1454	1440
OP2711	3563	0.00318	1823	1799
OP2720	1772	0.00198	1426	1412
OP4063	2450	0.00221	1637	1621

Appendix F. Structure-like, the BIC per number of clusters, and plot graphic outputs from DAPC analyses for **A**) *K*=3, **B**) *K*=6, **C**) *K*=15, and analyses on subclades **D**) *A*. *shoshone*, **E**) Hoodoo Mts., **F**) *A*. *cosmetoides*, **G**) *A*. cf. *quattuor*, and **H**) *A*. *quattuor* + *A*. *merickeli*.

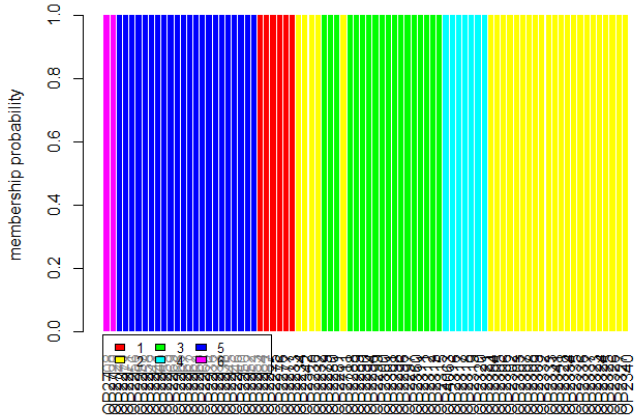
Appendix F.1. DAPC results for all inland *Acuclavella* - *K*=3.

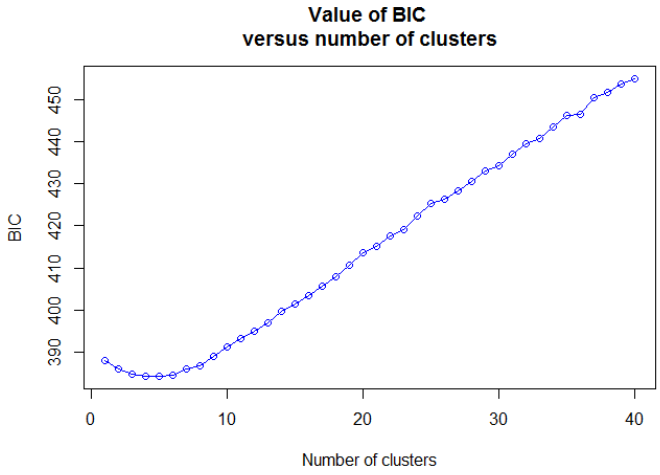


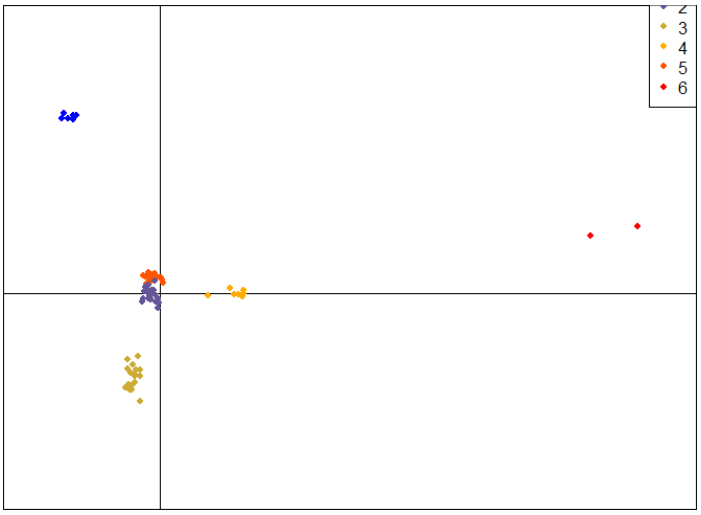




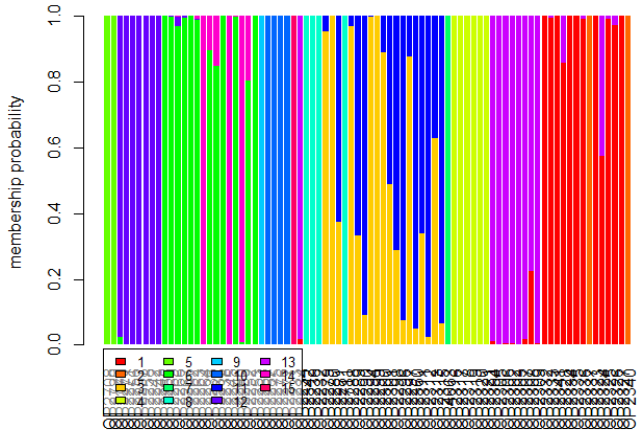
Appendix F.2. DAPC results for all inland *Acuclavella* - *K*=6.

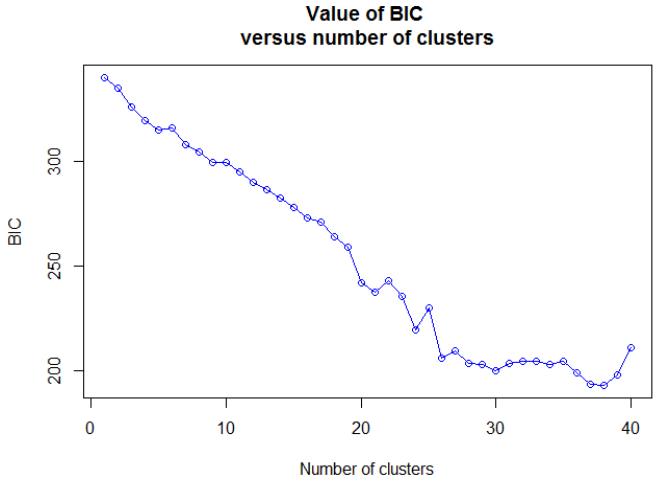


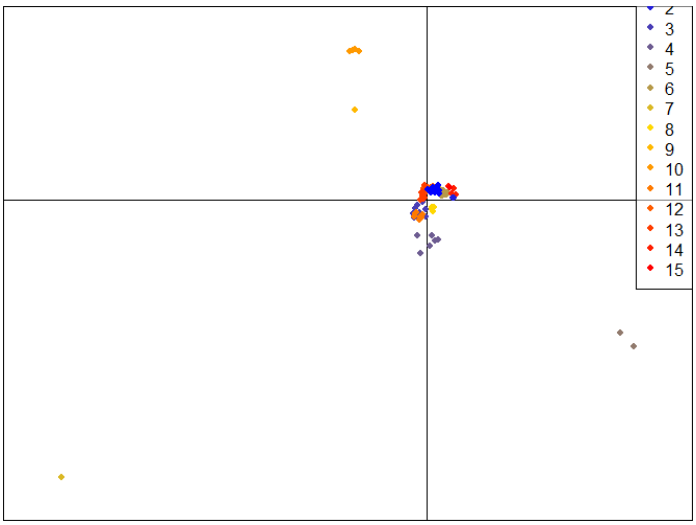




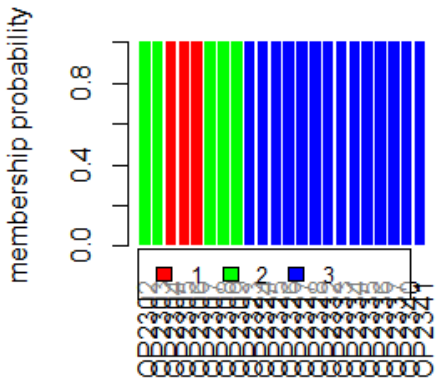
Appendix F.3. DAPC results for all inland *Acuclavella* - *K*=15.

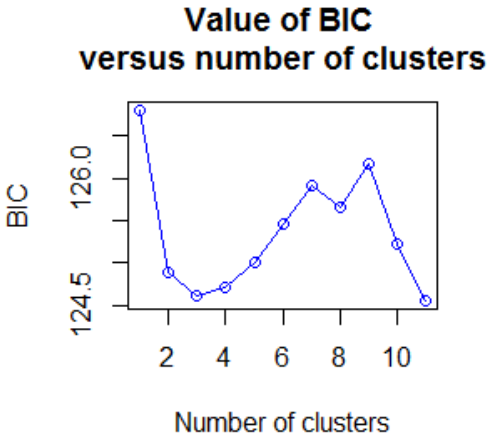


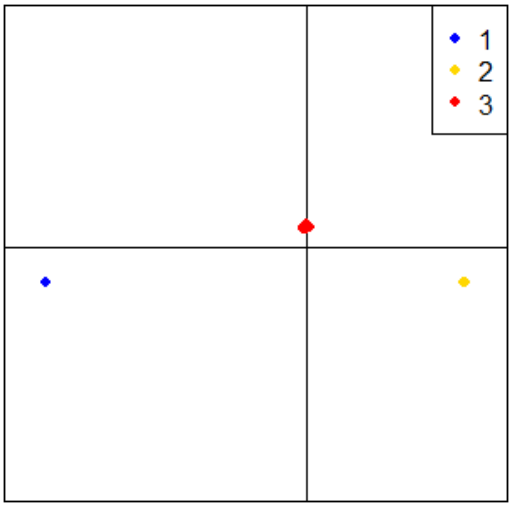




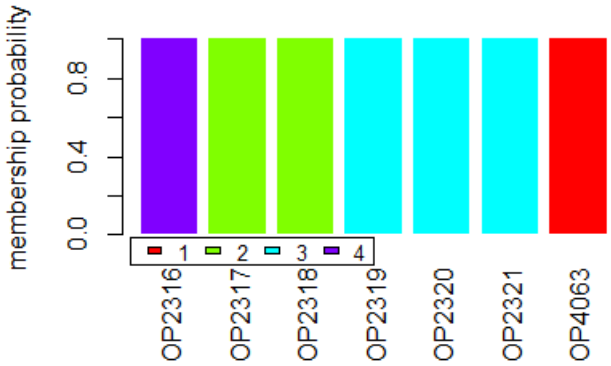
Appendix F.4. DAPC results for *A*. *shoshone*.

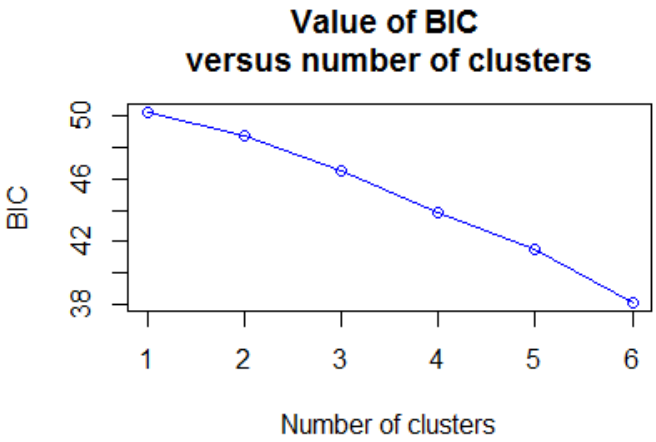


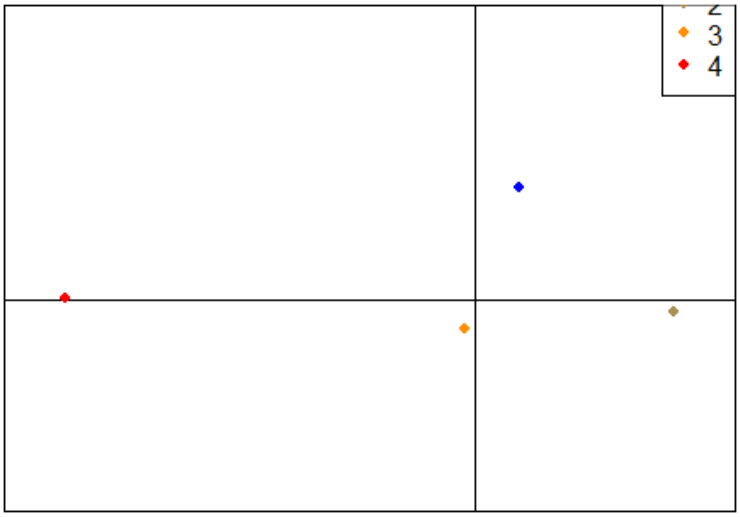




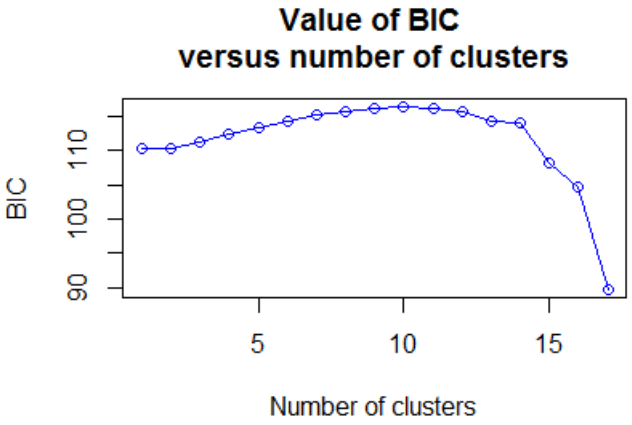
Appendix F.5. DAPC results for Hoodoo Mountains clade.

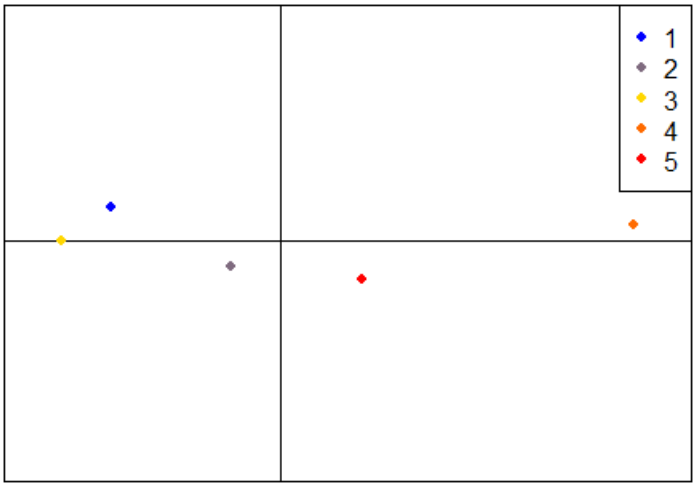




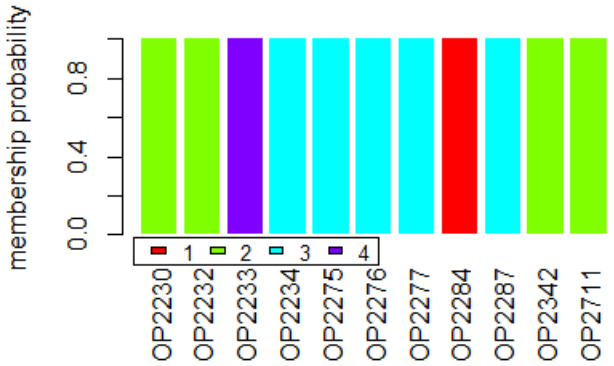


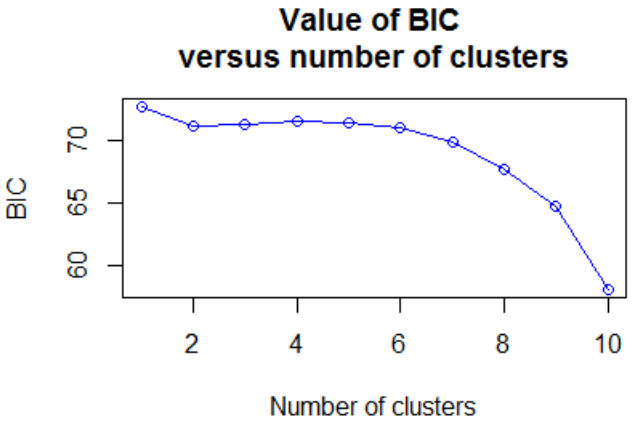
Appendix F.6. DAPC results for *A. cosmetoides*.

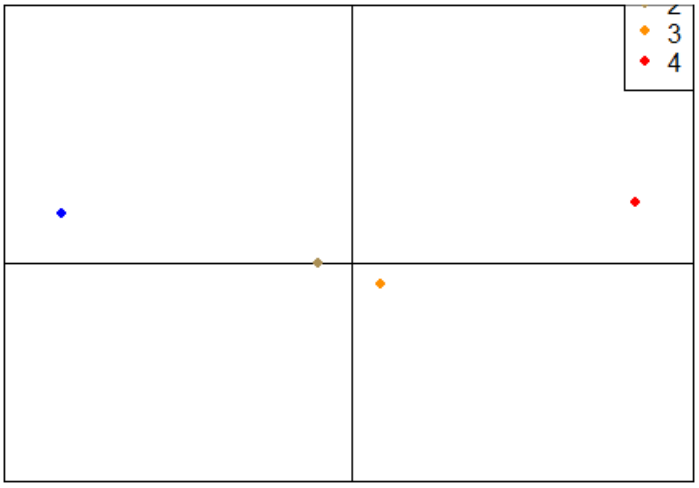




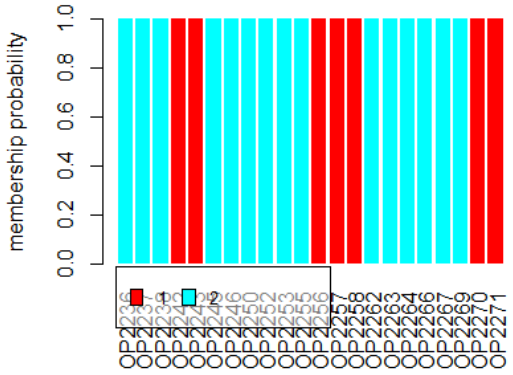
Appendix F.7. DAPC results for *A*. cf. *quattuor*.

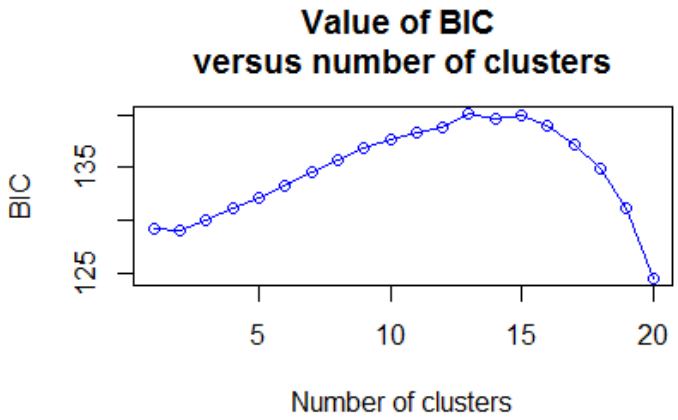


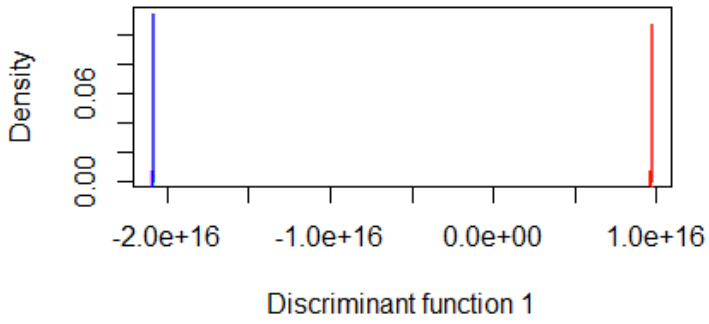




Appendix F.8. DAPC results for *A. quattuor* + *A. merickeli*.

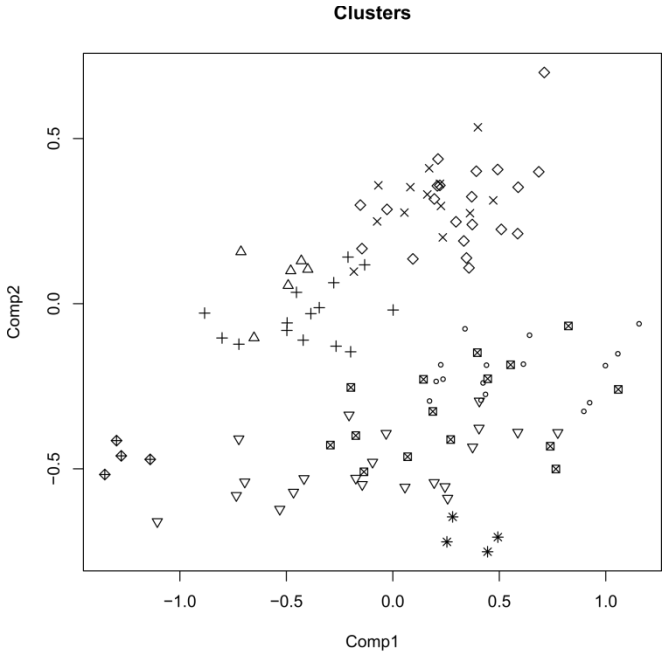


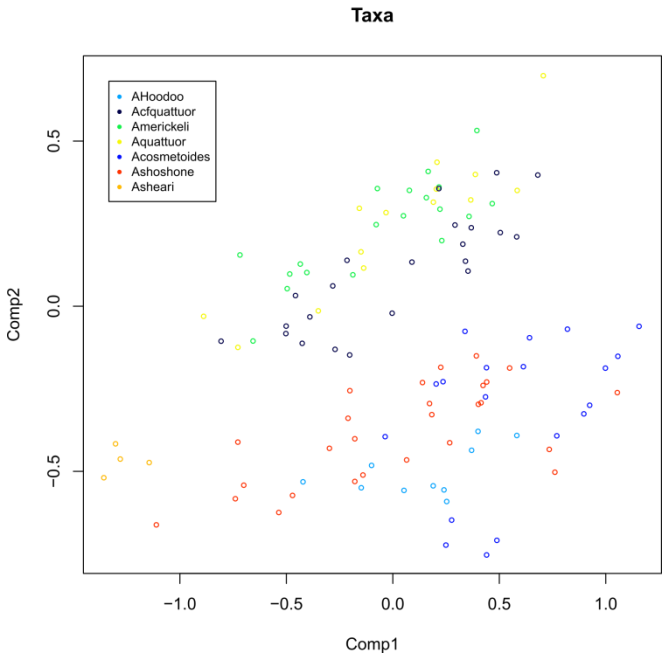




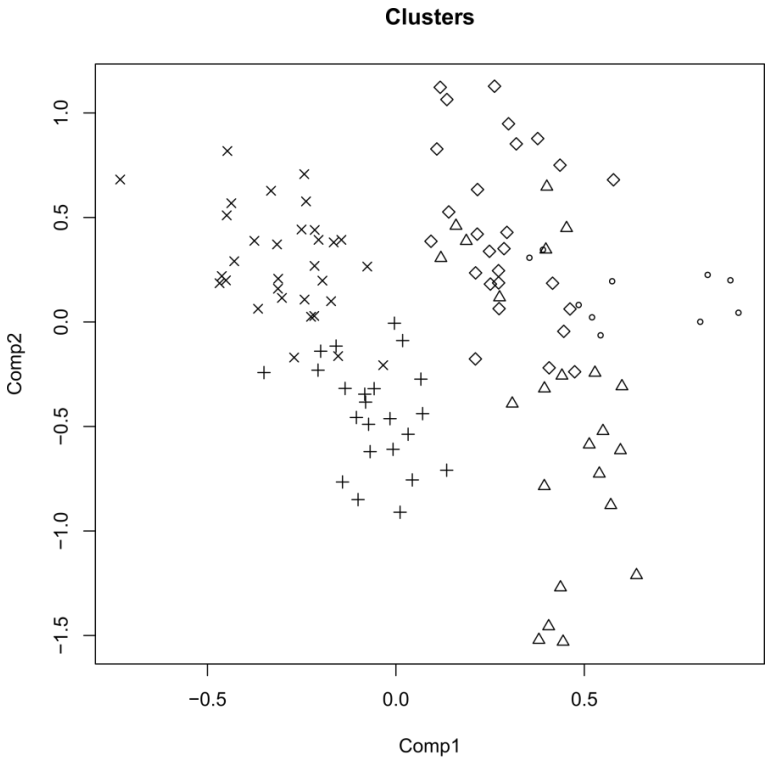
Appendix G. Morphometrics**.** Output graphics from morphometric analyses. For each analyses a Clusters graphic shows different symbols for each cluster, and a Taxa graphic shows where *a priori* identified individuals or major clades are found in morphometric space.

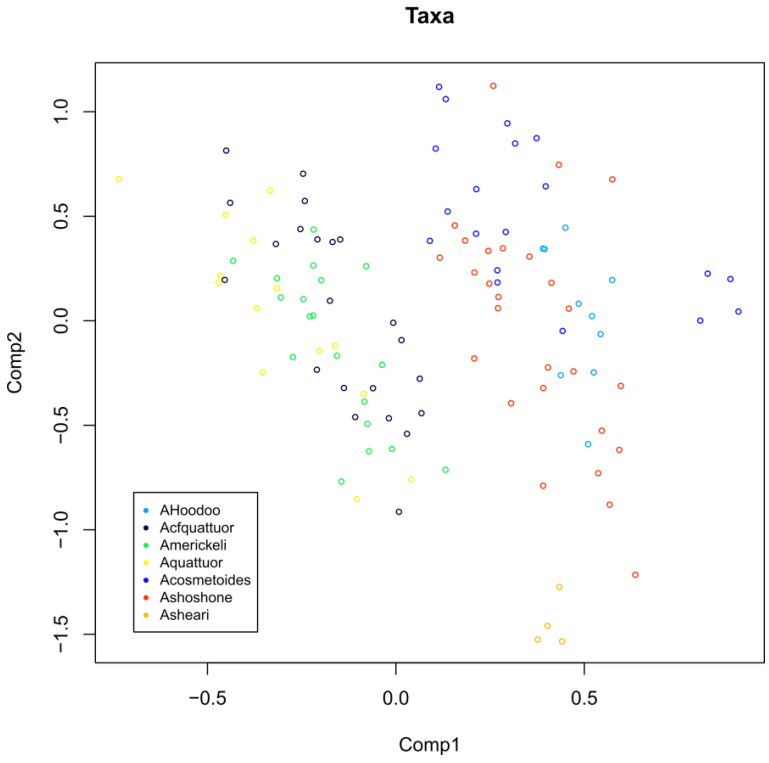
Appendix G.1. All inland *Acuclavella* males, *mad* estimator, 14 characters, *K*=9.



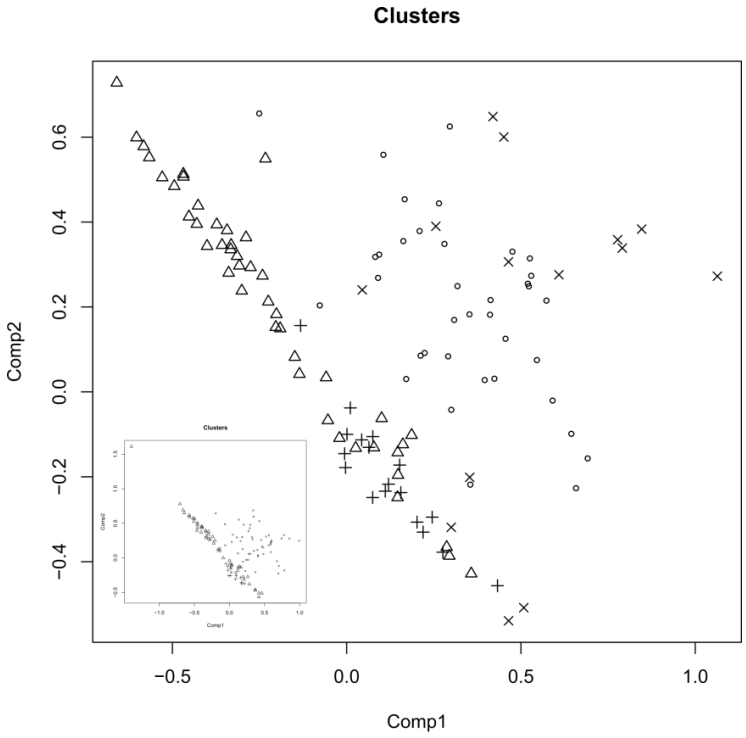


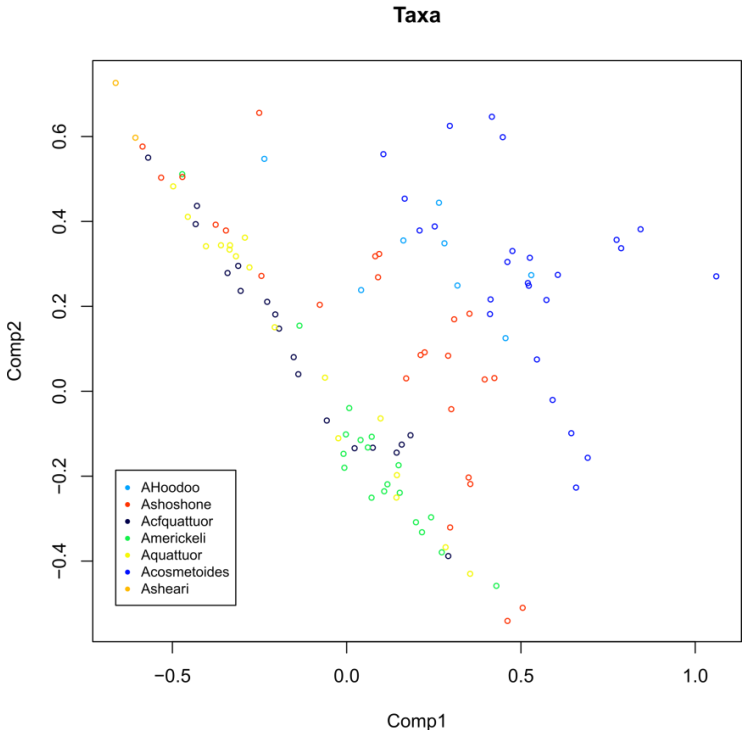
Appendix G.2. All inland *Acuclavella* males, *qn* estimator, 14 characters, *K*=5.



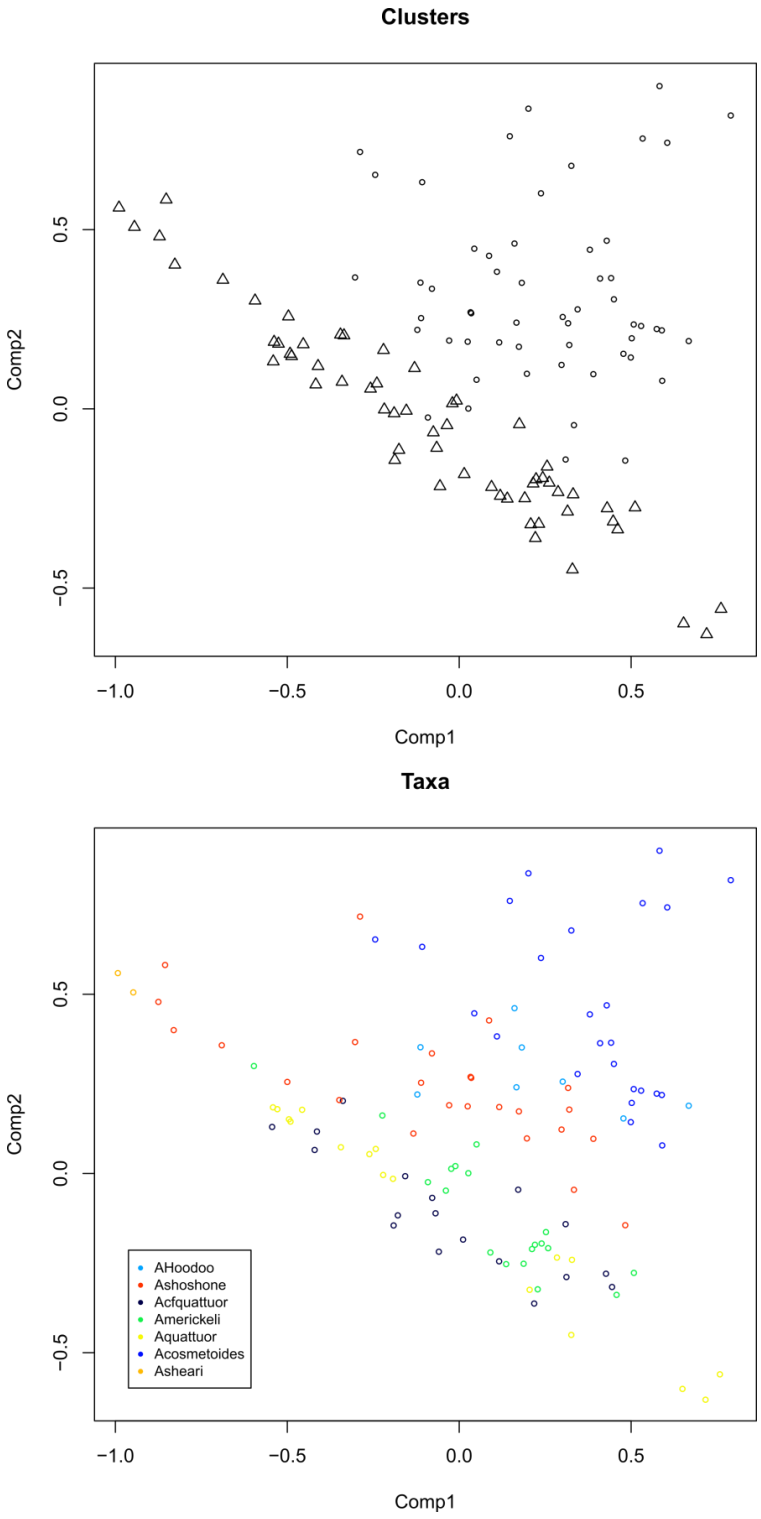


Appendix G.3. All inland *Acuclavella* females, *mad* estimator, 12 characters, *K*=4, outlier removed. Clusters inset shows analysis with outlier.

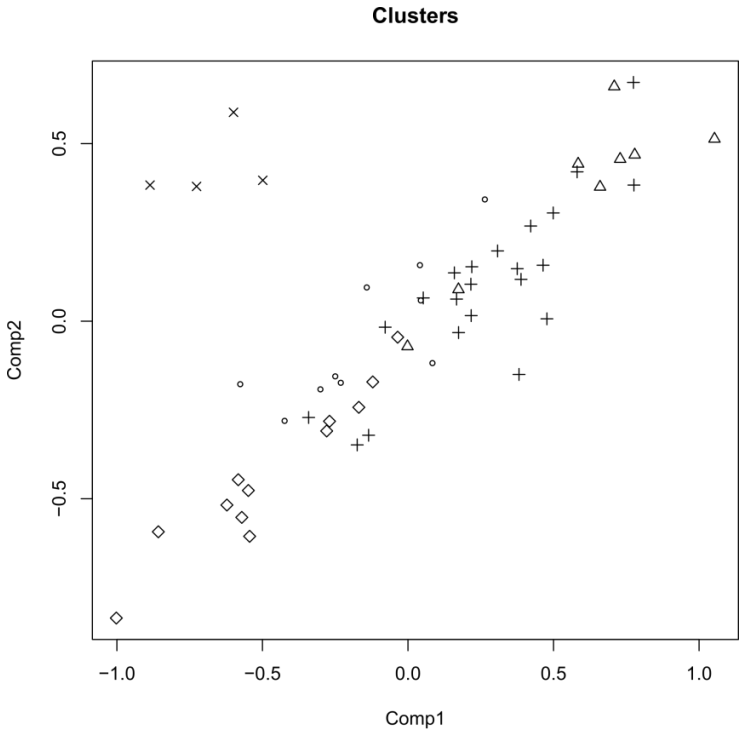


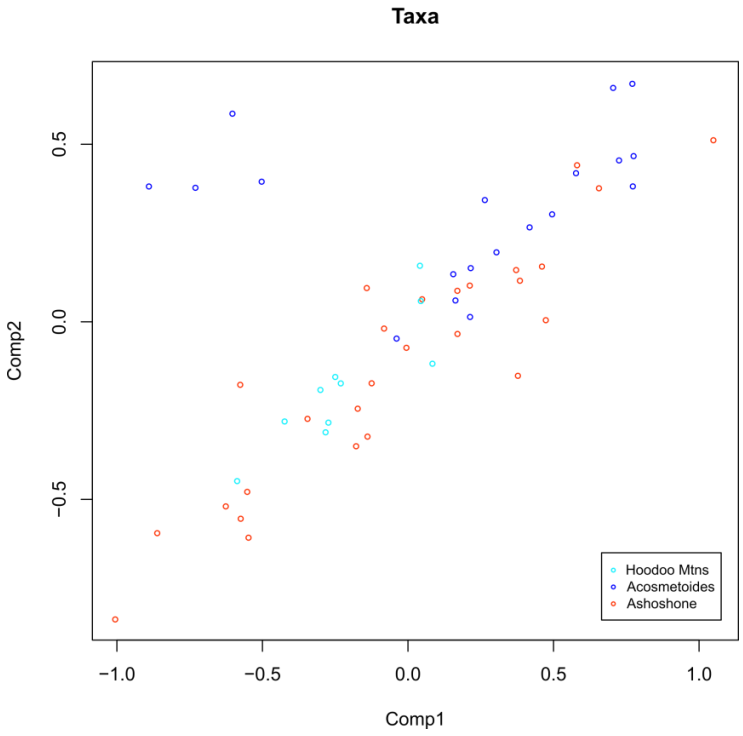


Appendix G.4. All inland *Acuclavella* females, *qn* estimator, 12 characters, *K*=2.

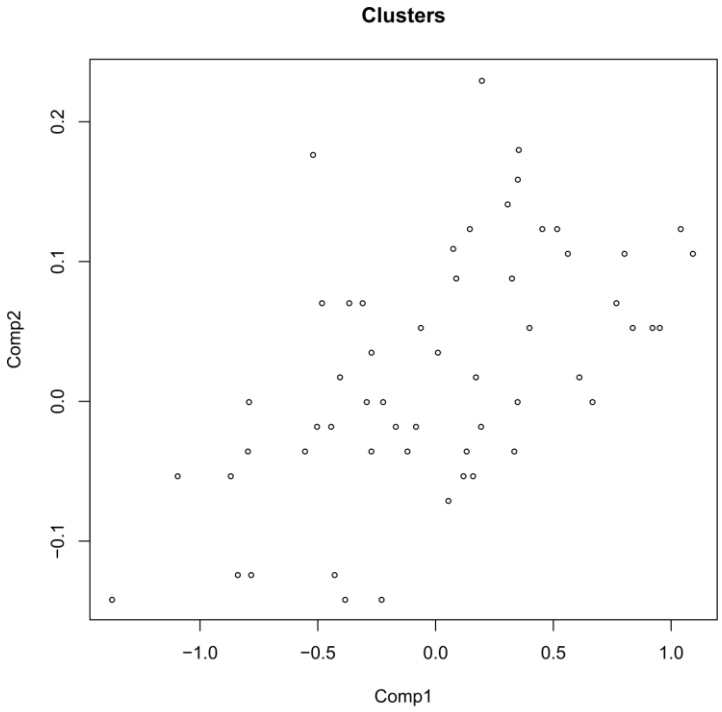


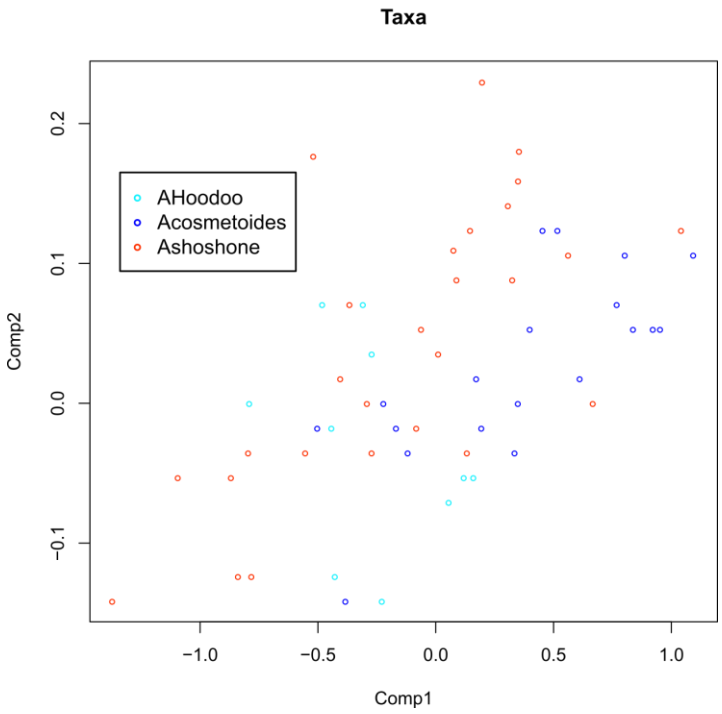
Appendix G.5. *Acuclavella cosmetoides*, *A. shoshone*, and Hoodoo Mountain clade, males, *mad* estimator, 12 characters, *K*=5.



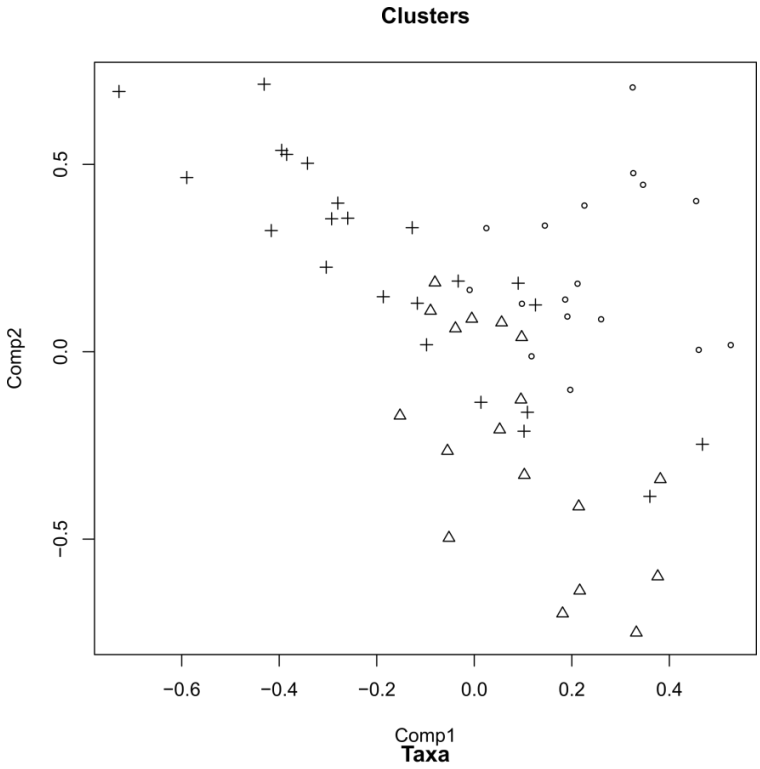


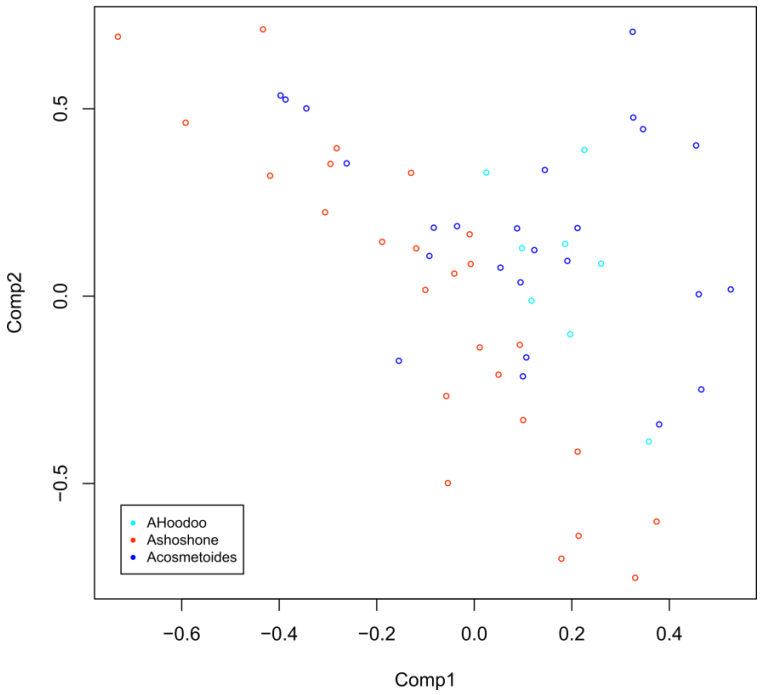
Appendix G.6. *Acuclavella cosmetoides*, *A. shoshone*, and Hoodoo Mountain clade, males, *qn* estimator, 14 characters, *K*=1.



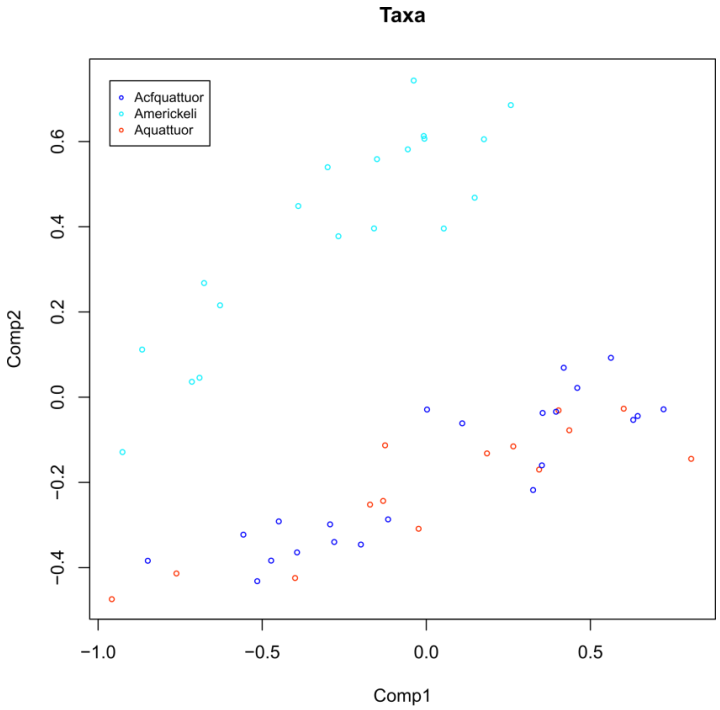


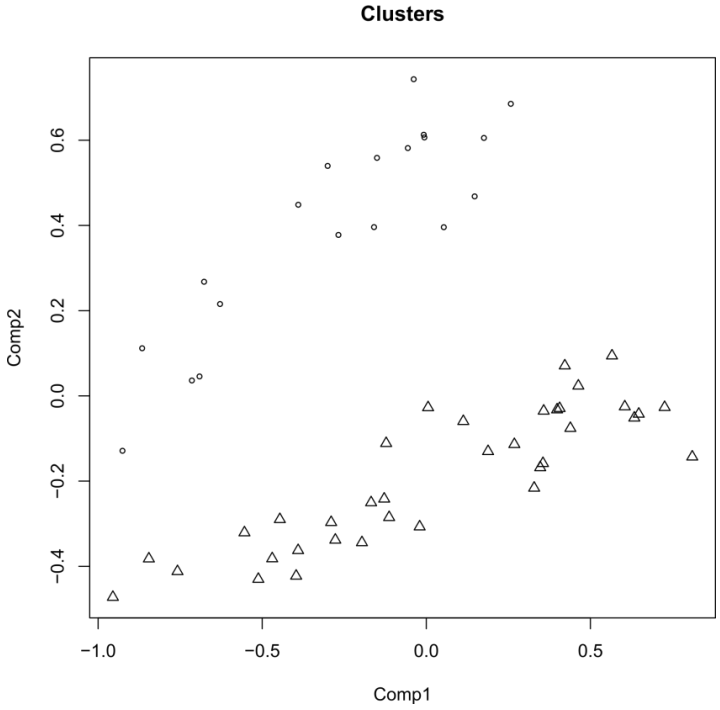
Appendix G.7. *Acuclavella cosmetoides*, *A. shoshone*, and Hoodoo Mountain clade, females, *mad* estimator, 12 characters, *K*=3.



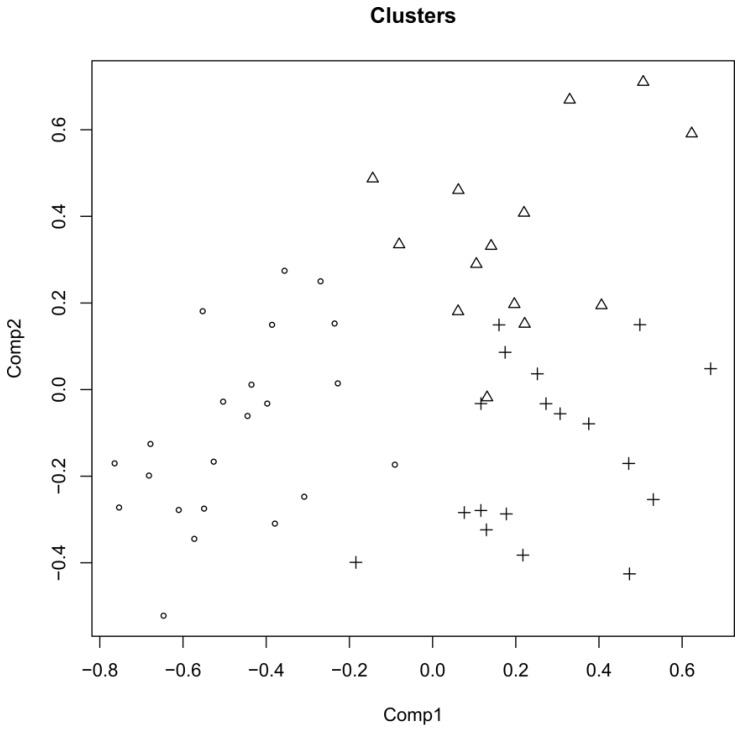


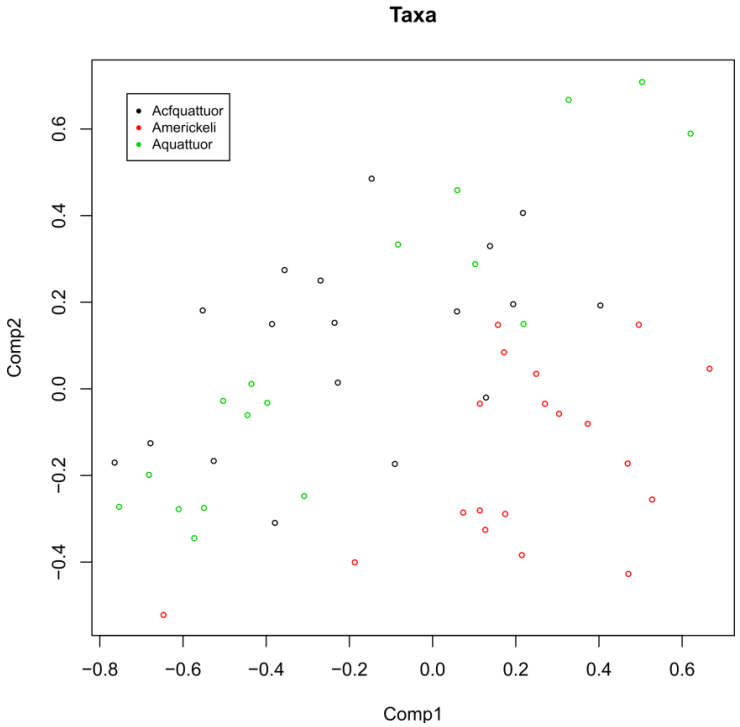
Appendix G.8. *Acuclavella merickeli*, *A*. *quattuor*, and *A*. cf. *quattuor*, males, *mad* estimator, 10 characters, *K*=2.





Appendix G.9. *Acuclavella merickeli*, *A*. *quattuor*, and *A*. cf. *quattuor*, females, *mad* estimator, 12 characters, *K*=3.





Appendix H. Inferred heterozygosity and locality elevation.

Specimen	Species	Heterozygosity	Elevation (m)
OP2230	cfquattuor	0.00498384	1100
OP2232	cfquattuor	0.00635616	1100
OP2233	cfquattuor	0.0060307	813
OP2234	cfquattuor	0.00889536	813
OP2275	cfquattuor	0.00644212	550
OP2276	cfquattuor	0.00684931	550
OP2277	cfquattuor	0.00683318	550
OP2284	cfquattuor	0.0062356	545
OP2287	cfquattuor	0.00653728	470
OP2342	cfquattuor	0.00533385	870
OP2711	cfquattuor	0.00460922	1143
OP2236	merickeli	0.00581395	560
OP2237	merickeli	0.00503284	535
OP2238	merickeli	0.0067925	535
OP2245	merickeli	0.00585871	590
OP2246	merickeli	0.00634981	590
OP2250	merickeli	0.00625	1270
OP2252	merickeli	0.00586218	1510
OP2253	merickeli	0.00589982	1510
OP2262	merickeli	0.0061508	1156
OP2263	merickeli	0.00620975	1156
OP2264	merickeli	0.00587588	1156
OP2242	quattuor	0.00737279	1005
OP2243	quattuor	0.00713115	1005
OP2255	quattuor	0.0051794	1870
OP2256	quattuor	0.00730188	920
OP2257	quattuor	0.00559033	920
OP2258	quattuor	0.00490629	920
OP2266	quattuor	0.00628211	1460
OP2267	quattuor	0.00583105	1460
OP2269	quattuor	0.00409773	1700
OP2270	quattuor	0.00689869	1675
OP2271	quattuor	0.00646214	1675
OP2708	sheari	0.00553542	1504
OP2720	sheari	0.00411298	1730
OP2302	shoshone	0.0055715	1140

Specimen	Species	Heterozygosity	Elevation (m)
OP2303	shoshone	0.00572649	1140
OP2304	shoshone	0.00599724	865
OP2305	shoshone	0.00623452	865
OP2306	shoshone	0.00572238	865
OP2307	shoshone	0.00654991	944
OP2308	shoshone	0.00579631	944
OP2309	shoshone	0.00667461	944
OP2323	shoshone	0.0053668	1295
OP2324	shoshone	0.00495852	1295
OP2325	shoshone	0.0053835	675
OP2326	shoshone	0.00526185	675
OP2327	shoshone	0.00652762	675
OP2328	shoshone	0.00512883	860
OP2332	shoshone	0.00576231	1096
OP2333	shoshone	0.00558703	1096
OP2334	shoshone	0.00473433	977
OP2335	shoshone	0.00489079	1269
OP2336	shoshone	0.00495563	1269
OP2337	shoshone	0.00458759	1269
OP2340	shoshone	0.00262697	1110
OP2341	shoshone	0.00436406	1305
OP2279	cosmetoides	0.00735272	490
OP2280	cosmetoides	0.00578636	490
OP2281	cosmetoides	0.00649478	465
OP2288	cosmetoides	0.00606726	965
OP2289	cosmetoides	0.00625195	965
OP2290	cosmetoides	0.00627007	965
OP2294	cosmetoides	0.00604496	1000
OP2296	cosmetoides	0.00603904	960
OP2297	cosmetoides	0.00555451	960
OP2298	cosmetoides	0.00578391	1342
OP2299	cosmetoides	0.00624433	968
OP2300	cosmetoides	0.00620098	968
OP2301	cosmetoides	0.00624923	968
OP2311	cosmetoides	0.00515189	645
OP2312	cosmetoides	0.00489462	645
OP2313	cosmetoides	0.00543426	645
OP2315	cosmetoides	0.00592903	940
OP2360	cosmetoides	0.00601228	960

Specimen	Species	Heterozygosity	Elevation (m)
OP2316	hoodoo	0.00512092	993
OP2317	hoodoo	0.00498478	900
OP2318	hoodoo	0.00477496	900
OP2319	hoodoo	0.00401864	997
OP2320	hoodoo	0.00470721	997
OP2321	hoodoo	0.00372636	997
OP4063	hoodoo	0.00374551	934