AN ABSTRACT OF THE DISSERTATION OF

John S. Sproul for the degree of Doctor of Philosophy in Integrative Biology presented on June 14, 2018

Title: Delimiting Species in the Mountaintops: New Trails for Studying Genomic Architecture and Sequencing Historical Specimens Uncover Hidden Diversity within the *breve* Species Group of *Bembidion* (Coleoptera: Carabidae)

Abstract approved: _____

David R. Maddison

Species are one of the foundational units upon which entire fields of scientific inquiry are built. Discovering and documenting the planet's biodiversity remains one of the grand challenges of science. A proper conceptualization of species provides a critical framework for diverse fields such as biophysics, biochemistry, agriculture and pharmacology, and for all of comparative biology. The need to advance the knowledge of biodiversity, and improve the methods used to study that diversity is the central aim of this dissertation. The work presented herein centers on investigating patterns of diversity in the *Bembidion breve* species group, a group of small ground beetles (Coleoptera: Carabidae) found at high elevations in the mountains of western North America. In undertaking to discover and document cryptic species present in this group, a set of challenges and opportunities led to research projects that develop the *breve* group as an ideal system for developing innovative approaches for conducting molecular studies in diverse non-model groups.

In addressing challenges associated with the identification of a 100-year-old type specimens in the *breve* group, Chapter 2 presents a methodological study designed to optimize sample preparation protocols for next-generation of small-bodied specimens with degraded DNA. Compare the library preparation success of several library

preparation protocols on low DNA input from several old specimens, including type specimens, ranging from 58–159 years in age. I also test the effect of enzymatic repair on library success and use several metrics of sequencing success to evaluate the effect of the various treatments. I demonstrate that excellent library preparation and sequencing success can be obtained using as little as 1 ng of degraded input DNA. I recommend simple library preparation protocol modifications that can be used to optimize sample preparation success of challenging museum specimens, and make recommendations for preserving valuable DNA of rare or unique specimens.

In Chapter 3, the I present the species delimitation and taxonomic revision of *breve* species group. In a group that has consisted of two recognized species for the last several decades, the I use molecular, morphological, and geographic data as evidence that at least nine species are present in the group and provides identification tools and species distribution data. I present a novel, sequence-based approach to identifying a 100-year-old type specimen that used evidence of copy number variation within the ribosomal DNA (rDNA) cistron (the rDNA region that encodes for 18S and 18S ribosomal RNA genes) to confirm the identity the type specimen of *Bembidion lividulum*, a specimen for which degradation evident in DNA sequence data prevented unambiguous identification using analysis of gene trees.

In Chapter 4, I further investigate interesting patterns of copy number variation within the rDNA cistron first reported in Chapter 3. I describe a simple method detecting signatures of genomic architecture using copy number variation profiles of the rDNA cistron to produce "rDNA profiles". I investigate the pattern of signatures in rDNA profiles among and within species of the *breve* group. I show that rDNA profiles hold excellent signal at the species level in a challenging species group, and is variable across a broader taxonomic group in the *Bembidion* subgenus *Plataphus*. I demonstrate that this approach is compatible with phylogenomic data generation workflows, and use fluorescence *in situ* hybridization (FISH) to corroborate patterns seen in rDNA profiles. These results highlight the potential value of methods that incorporate signal of genomic architecture and in species delimitation/phylogenetics, and how the patterns observed in those studies provide natural synergy with studies on genome evolution.

DELIMITING SPECIES IN THE MOUNTAINTOPS: NEW TRAILS FOR STUDYING GENOMIC ARCHITECTURE AND SEQUENCING HISTORICAL SPECIMENS UNCOVER HIDDEN DIVERSITY WITHIN THE *BREVE* SPECIES GROUP OF *BEMBIDION* (COLEOPTERA: CARABIDAE)

by
John S. Sproul

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 14, 2018
Commencement June 2019

Doctor of Philosophy dissertation of John S. Sproul presented on June 14, 2018

APPROVED:

Major Professor, representing Integrative Biology

Head of the Department of Integrative Biology

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

John S. Sproul, Author

ACKNOWLEDGEMENTS

I thank the Graduate School and Department of Integrative Biology at Oregon State University for accepting me to the program and providing support through teaching assistantships and tuition remission throughout my degree. I'm grateful to the Integrative Biology department office staff, Tara Bevandich, Traci Durrell-Khalife, Trudy Powell, Tresa Bowlin Salleng, Jane Van Order, and Torri Givigliano for their patience and support in helping me navigate my degree and for making the department a better place to be through their excellent service. I have especially appreciated interactions with Tara Bevandich through my role as department Safety Coordinator and have benefited from her always-available support and excellent knowledge of the department, university, and many other practical things. I am indebted to the department head Dr. Virginia Weis for her leadership in making the department a great environment for graduate study, and allowing me to serve as the department Safety Coordinator throughout my degree.

I thank the Graduate Student Committee in the department, other department leadership and office staff for supporting my research through ZoRF funds. I thank Harold and Leona Rice for their endowment to Oregon State University, from which many of my projects have received support. I also thank Paul and Mary Roberts for their generous funding of the Paul and Mary Roberts Fellowship which provided further financial support to my graduate research success. I am grateful to the National Science Foundation and anonymous reviewers for their generous funding of a Doctoral Dissertation Improvement Grant (DDIG-1702080) which greatly enhanced the scope of this dissertation.

I thank the good folks at the Center for Genomics Research and Biocomputing for their excellent support of the many sequencing projects that were part of my graduate research. I especially thank Mark Dasenko for always being willing to share his expertise and serve as a sounding board for experimental design and protocol modifications.

I am indebted to Drs. Tiffany Garcia, Aaron Liston, David Lytle, and Eli Meyer for their willingness to volunteer many precious hours and serve on my graduate committee. I appreciate their guidance and flexibility throughout my degree, and for

sharing their expertise and perspective during committee meetings, email exchanges, and individual meetings. I also thank Dr. Meyer and Dr. Liston for their excellent instruction during my coursework in their bioinformatics and comparative genomics courses respectively.

I am grateful for thousands of interactions with hundreds of undergraduate students in the courses I have taught (A&P and Biology Lab, Advanced Writing, and Discovering Insect Species). They have provided me an invaluable forum for developing my teaching philosophy and learning about myself. I thank them for their patience with me, and for occasional kind words that have fueled my dedication to teach. I am particularly thankful to twelve fantastic undergraduate researchers who have worked alongside me in the Maddison Lab: Austin Baker, Kaitlyn Traynor, Tiffany Soto, Joseph Dubie, Regina Kurapova, Kalyn Hansen, Chris Cohen, Ana Vasconcelos, Estany Campbell, Danielle Mendez, Tiana Week, and Lindsey Barton. Each brought skills and energy to the lab that made it a more diverse, enjoyable place to be. I owe them a debt of gratitude for their patience with unnecessarily longwinded explanations and anecdotes. I would especially like to thank Danielle Mendez, Tiana Week, Lindsey Barton, and Ana Vasconcelos for their direct contributions to either my dissertation chapters, or side projects in which I have been heavily invested. I have been lucky to have their support and comradery during many months of lab work.

I thank Dr. Devon Quick for her teaching mentorship as my supervisor through five years of teaching assistantship assignments in Anatomy and Physiology lab courses. She has been endlessly generous in supporting my growth as a teacher, and in supporting me in my other roles as a student, husband, and father. Her dedication to student success inspires me as an educator, as has her dedication to using her sphere of influence to provide graduate students, teaching interns, and lecture assistants with growth opportunities within her course. Watching her empower students at all levels with responsibilities, and her trust, has helped shaped my values as I move forward in mentoring roles. I thank her for all she has taught me about human anatomy. I thank her for supporting my career growth through writing several letters of support, and for always taking an interest in my family and life outside of teaching and research.

I am indebted to my fellow graduate students in the Maddison Lab: Kojun Kanda, James Pflug, Antonio Gomez, and Olivia Boyd. Each brought great energy to the lab and enriched my experience as a graduate student with their personality traits and kindness. They enriched my research outcomes with stimulating discussions and generous feedback. They have been kind to listen through increasingly frequent soap box rants and idealistic musings. They have been continually inclusive of me in lab activities and have always made it easy to be myself, for which I am grateful. I admire so much about each of their personalities, and will miss their daily companionship.

I am grateful to my parents, Janet and Greg Reed, for their unflagging support. So many of the good outcomes in my life are a direct result of their love, good examples, and the foundation of principles they taught me in my youth. They continue to teach me and inspire me to work hard and make them proud. I am thankful for their generosity in supporting our family in timely ways financially, which has frequently relieved school expenses, helped with travel costs, enabled more presents under the Christmas tree for the kids, or just covered a trip to Chipotle. I am indebted to them for managing our rental property and relieving so much of that stress and enabling monthly cash flow and the outcome its recent sale. I'm indebted in a similar way to my older sister Alisa Affleck. She and her family have been so supportive and encouraging on many levels. Her example has always been a light to me. I am grateful for my father Steve Sproul whose love and presence I have felt at key crossroads before, and during this experience. I am grateful to my in-laws, Vickie and Mike Twede. They have been so selfless in their support of me even though my study has created more physical distance between their daughter and grandchildren. Similar to my parents, their support on many levels including timely visits, a steady stream of surprise packages with clothing and diapers, travel support, friendship, and emotional support has lightened our load and added bright spots to our family experience that has meant the world to us. I'm also indebted to Steve Wainwright for continuing to be a mentor and friend as he has since my childhood. His praise and acceptance has always been an important part of my self-esteem. He has added invaluable diversity of perspective to my grant proposal planning, helped polish manuscripts and proposals, and volunteered timely emotional and financial support.

Many other family and friends deserve much thanks: Liz's siblings, our grandparents and extended families. In addition, our adoptive family of Corvallis friends too numerous to list have created such a great support network, especially for Liz and the kids, that has been critical to maintaining the balance and contentment in our family.

I am thankful to my major advisor, Dr. David Maddison. I feel so fortunate that he was willing to take a chance on me as a student. He has dedicated an astounding amount of energy and resources to my success as a student and researcher, much more than I deserve. I am grateful for the countless hours he spent in training sessions, scheduled and unscheduled meetings, exchanging emails, collecting and processing specimens, previewing presentations, reviewing manuscripts, supporting grant proposals, writing letters, and many other activities in support of my success and development. I am grateful for his generosity in sharing his study system (*Bembidion*) with me, for sharing his past discoveries in the *breve* group, and for enabling such added depth and breadth to the quality of my projects through his intellectual brilliance and generous financial backing. I am grateful to him for sharing his excellent standards of scientific rigor and his example of dedication to his work. I thank him for his willingness to support me in my roles outside the lab and encouraging family involvement in my research. I have so appreciated his listening ear, willingness to communicate (and patience with my communication style), and commitment to maintaining positive working relationships with myself, and everyone in the lab. I thank him for his patience with my deliberate learning style, poor grammar, and frequent derailing of lab meetings with strange commentary. Most of all, I thank him for committing the time and energy to be so fully engaged in my program – I have been more fortunate than most graduate students in this way. I do not take it for granted that David was directly alongside me at many of my most important moments of discovery throughout my program, adding energy, extending resources, and teaching me the next steps. I will always be indebted to him for his all-in mentorship.

Finally, I thank my wife Elizabeth, and children, George, Pearl, and Johnny. Getting to know George, Pearl, and Johnny has been the thrill of my life. Their presence in my life has been a daily reservoir of motivation for me during this degree. George and Pearl have been unforgettable companions in adventure both during fieldwork associated

with my research, and as we've explored the Pacific Northwest. Johnny is still figuring out how to adventure, but he has two excellent mentors. George's sense of humor, maturity, and love of his siblings has been so important to the peace in our home. Pearl's endless positivity and enthusiasm for life has kept us active and happy. Johnny's antics and chatter have brought us together smiling every day. I am so grateful to George Pearl, and Johnny for their unconditional patience and support. I thank my family for joining me on many long, challenging, and beautiful collecting trips. This dissertation has taken us to the tops of the mountains throughout western North America. Liz and the kids have climbed to field sites throughout the Sierra Nevada and Trinity Alps in California. The Cascades, Steens, and Wallowa Mountains of Oregon. The Bitterroot, Mission, and Anaconda ranges in Montana, and throughout Glacier National Park. The Wasatch and Uintah ranges in Utah. I will always treasure the dozens of miles on the trail and thousands of miles in the car they shared with me during my research.

Liz's sacrifice in behalf of this degree has been immeasurable. She embraced the experience from the beginning with her characteristic brightness that has made the experience joyful despite the often challenging, and sometimes overwhelming circumstances. She has embraced the realities of raising a young family on a student salary in a privileged town, insufficient cash budgets, and foregone wants and needs. She has embraced six years in an ever-shrinking two-bedroom house that is too cold in the winter and too hot in the summer, a master bedroom shared with George and Pearl for 18 months of sleep training with a new baby. She has embraced the 15-hour drive from her close-knit family and the challenge of building relationships in a new place. She has embraced all the stressors that come with a PhD program, weeks of camping fieldwork, and solo parenting during my travel, months-worth of hard weeks leading up to program and proposal deadlines, and the uncertainty that comes with post-PhD job hunting. I thank her for her companionship. I am grateful for hours of listening, encouragement, editing manuscripts and emails, for her patience when stress ran high. I thank her for approaching this experience knowing that these years will make up fondest memories of our lives, and insisting that we make that hope a reality by staying balanced and finding ways to enjoy each day. I thank her for being my friend and not just my partner in

responsibilities. I thank her for laughing at my jokes, for her gentleness, and strength. I thank her for lifting me in my low points, and letting me do the same for her. I thank her for expecting the best of me and helping me keep my priorities straight. I thank her for her loyalty to our relationship, and her quickness to forgive. I thank her for the thousands of hours she spends loving our children and taking the lead on so many responsibilities in our home. I thank her for trusting me and loving me despite my occasional eyebrow dandruff. Her friendship, laughter, love and endless support has been my greatest source of strength these last six years.

I am grateful for the mossy forests of the Pacific Northwest, firewood collecting permits, and wood that needed splitting and stacking. For rhododendrons and azaleas in the spring, for berries in the summer, for months of changing colors in the fall, and the contrast of lichen-draped oaks against dark, rain-drenched Douglas Fir all winter long.

"*I long for weight and strength*

To feel the earth as rough

To all my length."

-Robert Frost

CONTRIBUTION OF AUTHORS

Lindsey M. Barton contributed to data analysis and figure production of repetitive DNA patterns presented in Chapter 4 and will be an author on the manuscript that is in preparation.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF FIGURES (CONTINUED)

Figure Page

LIST OF FIGURES (CONTINUED)

Figure Page

LIST OF TABLES

LIST OF TABLES (CONTINUED)

CHAPTER 1: INTRODUCTION

OVERVIEW

Species are one of the foundational units upon which entire fields of scientific inquiry are built. Discovering and documenting the planet's biodiversity remains one of the grand challenges of science. A proper conceptualization of species provides a critical framework for diverse fields such as biophysics, biochemistry, agriculture and pharmacology, and for all of comparative biology (Mayr, 1982; de Queiroz, 2005). Most of what is known in biological sciences today can eventually be traced back to the process of discovering of new species, gaining a deeper understanding of their characteristics, and eventually using the patterns observed to generate and test hypotheses. The discovery of morphological variation in Galapagos fauna, further understanding patterns of that variation across the landscape, and delimiting the observed forms into species concepts were fundamental steps to the development of Darwin's theory of evolution through natural selection (Darwin, 1859). Over a century later, the discovery of *Thermus aquaticus* (Brock & Freeze, 1969) in boiling geothermal pools provided molecular biologists a blueprint for enzymes functional at high temperatures that enabled polymerase chain reaction (Holland *et al.*, 1991) and the subsequent explosion of knowledge resulting from DNA sequencing technology.

Well-characterized species with desirable traits such as short generation time, or small genome size serve as model systems for hypothesis testing in science. *Drosophila melanogaster*, *Mus muscula*, *Escherichia coli*, *Caenorhabditis elegans*, and others model organisms drive major scientific advancements, including much of what is known in human health. Outside of model organisms, countless species serve as models that advance knowledge in biological questions ranging from climate

change, to agricultural advancements, to cancer research. In this way, all fields of biology benefit from diverse models afforded by the diversity of species on the planet. Discovering and documenting the millions of yet unknown species is critical to the continued acceleration of scientific knowledge and preservation of the planet's biodiversity.

Improving our ability to identify and study biodiversity is the theme that unites the present work. This dissertation presents a microcosm of how species discovery and documentation provide a platform for additional scientific inquiry. The *Bembidion breve* species group serves as the model system for this work.

Bembidion is a genus of small ground beetles (Coleoptera: Carabidae) containing approximately 1200 species world-wide (Maddison, 2012). The *breve* species group is in the subgenus *Plataphus* which contains approximately 85 species that are associated with river shores, pond shores, marshlands, wet forests, and damp alpine habitats. Species in the *breve* group are found at high elevation in the mountains of western North America, and most often reside along the edges of small streams and patches of melting snow in alpine and sub-alpine settings. For several decades, there have been two species names recognized in the *breve* group, *B. breve* (Motschulsky) and *B. laxatum* Casey. Previous work by carabidologist Carl Lindroth noted high levels of intraspecific variation, distinctive geographic forms, and the need for a thorough revision (Lindroth, 1963). More recent investigation into the group by David Maddison noted distinctive molecular forms from the Sierra Nevada in California, and that also showed morphological variation corroborating molecular

patterns, and strongly suggested the existence of additional species. This set the stage for the present work.

I began studying the *breve* group with the expectation that a taxonomic revision of the species would serve as one dissertation chapter, and that would be the extent of its role in my dissertation. However, a series of challenges and opportunities encountered while studying the group redirected the course of my dissertation to focus on developing of the *breve* group as a model for improving studies of biodiversity in non-model groups. The following sections provide a brief overview of the three projects that have arisen from my journey into understanding species boundaries in the *breve* group.

SEQUENCING HISTORICAL SPECIMENS

An initial challenge that arose in revising the taxonomy of the *breve* group was that assigning names to species concepts within the *breve* group required assessing the potential validity of junior synonyms to determine which previously existing names should be applied to species concepts identified in the group. Geographic locality of critical type specimens, or dissections of male genitalia, allowed for the confident placement of most existing names. However, two century-old type specimens (those of *Bembidion saturatum* Casey and *B. lividulum* Casey) are female, and are from geographic localities in which several candidate species occur in sympatry. Arising out of the need to address this challenge, my first research chapter focuses on improving methods for obtaining DNA sequences from challenging museum specimens with low amounts of degraded DNA.

DNA sequencing technology has greatly accelerated our ability to identify species not previously known to science. Several studies in recent years have highlighted the value of partnering next-generation sequencing technology with studies with natural history collections in order to accelerate understanding of biodiversity (Wandeler, Hoeck, & Keller, 2007; Staats *et al.*, 2013; Kanda *et al.*, 2015). Advances in short-read sequencing potentially make available for molecular study the millions of specimens housed in the world's museums, which were not available for molecular study under Sanger sequencing technology studies which is sensitive to DNA degradation that occurs in specimens not preserved for molecular work.

Despite advances that allow DNA sequencing of old museum specimens, sequencing small-bodied, historical specimens can be challenging and unreliable as many contain only small amounts of fragmented DNA. Ideal sample preparation protocols for such historical specimens fall in between existing guidelines for ancient DNA that is thousands to hundreds of thousands of years old (Gansauge & Meyer, 2013; Bennett *et al.*, 2014), and protocols optimized for high-quality DNA, which are often used for historical specimens despite being optimized for higher-quality samples. Dependable methods to sequence such specimens are especially critical if the specimens are unique. This chapter outlines a study designed to sequence small-bodied (3–6 mm) historical specimens (including nomenclatural types) of beetles that have been housed, dried, in museums for 58–159 years, and for which few or no suitable replacement specimens exist. To better understand ideal approaches of sample preparation and produce preparation guidelines, I compared different library

preparation protocols using low amounts of input DNA (1–10 ng). I also explored low-cost optimizations designed to improve library preparation efficiency and sequencing success of historical specimens with minimal DNA, such as enzymatic repair of DNA. This chapter not only addresses the need to obtain sequences from challenging type sequences needed for the revision of the *breve* group, but also seeks to present general guidelines to facilitate more economical use of valuable DNA, and enable more consistent results in projects that aim to sequence challenging, irreplaceable historical specimens.

TAXONOMIC REVISION OF THE *BREVE* GROUP

This chapter presents the delimitation and formal revision of the *breve* species group. Considering multiple lines of evidence is critical when delimiting species (de Queiroz, 2007; Padial *et al.*, 2010; Carstens *et al.*, 2013), and integrating evidence from a combination of molecular, morphological, ecological, behavioral, geographic, and other data sources is a common theme in recent studies seeking to elucidate species boundaries (Domingos *et al.*, 2014; Huang & Knowles, 2015; Firkowski *et al.*, 2016; Ojanguren-Affilastro *et al.*, 2016; Papakostas *et al.*, 2016). I approached the delimitation of *breve* group using evidence taken from a combination of molecular, morphological and geographic data sources.

Sequencing of *breve* group type specimens in the previously introduced project provided initial evidence as to the identification of female type specimens critical to taxonomic decisions. This project provides an additional line of evidence that corroborates the previous assignment by using evidence of genomic architecture in the form of ribosomal DNA (rDNA) copy number variation within the rDNA

cistron (the region of rDNA containing 18S and 28S rRNA genes and their spacer regions). This approach, which relies on sequences of multi-copy genes, show promising rates of recovery in projects attempting to sequence old specimens (Wandeler *et al.*, 2007; Guschanski *et al.*, 2013; Staats *et al.*, 2013; Burrell, Disotell, & Bergey, 2015; Kanda *et al.*, 2015). The promising initial results of this approach to measuring aspects of genomic architecture using signal from the rDNA cistron, and the framework of breve group species produced by this chapter provided context for the subsequent chapter which further investigates patterns of rDNA genomic architecture differences in the *breve* group.

LOW-COST GENOMIC ARCHITECTURE IN NON-MODEL GROUPS

This chapter further investigates evidence that variation in ribosomal DNA (rDNA) copy number is present among closely related species in the *breve* species group. Following the Illumina sequencing of the type specimen of *B. lividulum*, I was examining the pattern of coverage depth across the ribosomal DNA cistron, and noted that although the region containing the 18S ribosomal RNA gene showed very poor coverage, the 28S rRNA gene just a few thousand bases downstream showed >400X coverage. Subsequent analysis of fresh specimens of that species showed a similar signature of copy number variation. Sequencing specimens of two other species, showed a species-specific signature of copy number variation for each species measured.

In this chapter, I investigated the rDNA copy number variation present within and among breve group species. I use a simple method of visualizing evidence of differences in the genomic architecture of the rDNA cistron by mapping low-

coverage genomic reads to a reference sequence and visualizing variation present in the resulting read pileups. I map these "rDNA profiles" onto the species tree generated as part of the species delimitation chapter introduced previously. I validate patterns seen in sequence-based approaches through cytogenetic mapping of rDNA using fluorescent *in situ* hybridization (FISH), and tested whether cluster-based analysis of repetitive DNA corroborated patterns seen in rDNA profiles using RepeatExplorer (Novák *et al.*, 2013). I also tested for variation in rDNA profiles at broader taxonomic scales in the *Bembidion* subgenus *Plataphus*, the subgenus containing the *breve* group.

This chapter highlights the potential value of methods that incorporate signal of genomic architecture and in species delimitation/phylogenetics, and how the patterns observed in those studies provide natural synergy with studies on genome evolution.

REFERENCES

Bennett EA, Massilani D, Lizzo G, Daligault J, Geigl EM, Grange T. 2014. Library construction for ancient genomics: single strand or double strand. *BioTechniques* 56: 289–90, 292–6, 298.

Brock TD, Freeze H. 1969. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of Bacteriology* 98: 289–297.

Burrell AS, Disotell TR, Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* 79: 35–44.

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. *Molecular Ecology* 22: 4369–4383.

Darwin C. 1859. *On the Origin of Species*. Routledge.

Domingos FM, Bosque RJ, Cassimiro J, Colli GR, Rodrigues MT, Santos MG, Beheregaray LB. 2014. Out of the deep: cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods. *Molecular Phylogenetics and Evolution* 80: 113–124.

Firkowski CR, Bornschein MR, Ribeiro LF, Pie MR. 2016. Species delimitation, phylogeny and evolutionary demography of co-distributed, montane frogs in the southern Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution* 100: 345–360.

Gansauge MT, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* 8: 737–748.

Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gilissen E, Sonet G, Nagy ZT. 2013. Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology* 62: 539–554.

Holland PM, Abramson RD, Watson R, Gelfand DH. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'—-3'exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences* 88: 7276–7280.

Huang JP, Knowles LL. 2015. 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.

Kanda K, Pflug JM, Sproul JS, Dasenko MA, Maddison DR. 2015. Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. *PLoS One* 10: e0143929.

Lindroth CH. 1963. The ground beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska, Part 3. *Opuscula Entomologica Supplementum*: 201–408.

Maddison DR. 2012. Phylogeny of Bembidion and related ground beetles (Coleoptera: Carabidae: Trechinae: Bembidiini: Bembidiina). *Molecular Phylogenetics and Evolution* 63: 533–576.

Mayr E. 1982. *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29: 792–793.

Ojanguren-Affilastro AA, Mattoni CI, Ochoa JA, Ramírez MJ, Ceccarelli FS, Prendini L. 2016. Phylogeny, species delimitation and convergence in the South American bothriurid scorpion genus Brachistosternus Pocock 1893: integrating morphology, nuclear and mitochondrial DNA. *Molecular Phylogenetics and Evolution* 94: 159–170.

Padial JM, Miralles A, De la Riva I, Vences M. 2010. The integrative future of taxonomy. *Frontiers in Zoology* 7: 1.

Papakostas S, Michaloudi E, Proios K, Brehm M, Verhage L, Rota J, Pe√±a C, Stamou G, Pritchard VL, Fontaneto D. 2016. Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: evidence from a rotifer cryptic species complex. *Systematic Biology* 65: 508–524.

de Queiroz K. 2005. Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences* 102: 6600–6607.

de Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886.

Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One* 8: e69189.

Wandeler P, Hoeck PE, Keller LF. 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* 22: 634–642.

CHAPTER 2: SEQUENCING HISTORICAL SPECIMENS: SUCCESSFUL PREPARATION OF SMALL SPECIMENS WITH LOW AMOUNTS OF DEGRADED DNA

John S. Sproul
David R. Maddison

Molecular Ecology Resources
https://onlinelibrary.wiley.com/journal/17550998
Volume 17, Issue 6 (Pages 1183-1201)

ABSTRACT

Despite advances that allow DNA sequencing of old museum specimens, sequencing small-bodied, historical specimens can be challenging and unreliable as many contain only small amounts of fragmented DNA. Dependable methods to sequence such specimens are especially critical if the specimens are unique. We attempt to sequence small-bodied (3–6 mm) historical specimens (including nomenclatural types) of beetles that have been housed, dried, in museums for 58–159 years, and for which few or no suitable replacement specimens exist. To better understand ideal approaches of sample preparation and produce preparation guidelines, we compared different library preparation protocols using low amounts of input DNA (1–10 ng). We also explored low-cost optimizations designed to improve library preparation efficiency and sequencing success of historical specimens with minimal DNA, such as enzymatic repair of DNA. We report successful sample preparation and sequencing for all historical specimens despite our low-input DNA approach. We provide a list of guidelines related to DNA repair, bead handling, reducing adapter dimers, and library amplification. We present these guidelines to facilitate more economical use of valuable DNA, and enable more consistent results in projects that aim to sequence challenging, irreplaceable historical specimens.

INTRODUCTION

The millions of historical specimens housed in the world's museums are a vast repository of biological data. Advances in DNA sequencing technology have opened doors to obtaining sequence data for many historical specimens despite their not having been intentionally preserved for DNA research, creating a resurgence of interest in museum specimens and their potential contribution to molecular investigations (Besnard *et al.*, 2014, 2015; Burrell, Disotell, & Bergey, 2015). Several recent studies have leveraged sequence data from historical specimens to address biological questions that could not be addressed as adequately, or at all, without historical specimens. For example, investigations into historical population genomics (Bi *et al.*, 2013) and phylogenetic placement of enigmatic lineages (Kanda *et al.*, 2015), testing phylogeographic hypotheses for conservation (Carmi *et al.*, 2016), and delimitating cryptic species and confirming their taxonomic status (Hind *et al.*, 2015; Lindstrom *et al.*, 2015; McCormack, Tsai, & Faircloth, 2015) have all been bolstered or made possible through sequencing historical specimens.

Even with technical advances, working with historical specimens is challenging because of their relatively poor-quality DNA. Two primary challenges faced by molecular studies with historical specimens are DNA degradation (including fragmentation and base damage) and low amounts of total DNA. Both of these factors, as well as challenges related to sample contamination, contribute to wide variation in sequencing success, and failure to proceed with a subset of otherwise desirable specimens (Kanda *et al.*, 2015; Blaimer *et al.*, 2016; Lim & Braun, 2016). While total DNA quantity is less often an issue for organisms whose body mass is

measured in grams and kilograms (Mason *et al.*, 2011; Bi *et al.*, 2013; Guschanski *et al.*, 2013; Besnard *et al.*, 2015; Hofreiter *et al.*, 2015), DNA quantity is a central limitation to preparing smaller-bodied organisms for sequencing, in particular the millions of small arthropods which dominate organismal diversity of museum collections (Staats *et al.*, 2013; Maddison & Cooper, 2014; Kanda *et al.*, 2015; McCormack *et al.*, 2015; Blaimer *et al.*, 2016; Prosser *et al.*, 2016).

Sequencing small, historical specimens may be especially challenging if a specimen is unique, or nearly so, with no alternative specimens available for study should the first specimen fail. Studies that include historical specimens commonly rely on unique or rare specimens. For example, rare or difficult to collect specimens have added invaluable data that affected the ecological conservation status of species (Wandeler, Hoeck, & Keller, 2007; Carmi *et al.*, 2016). Sequencing nomenclatural types (the name bearers of a scientific name), and placing them into species delimited using fresh specimens, may be necessary to establish to which species a name belongs (Hind *et al.*, 2015; Lindstrom *et al.*, 2015; Mutanen *et al.*, 2015). (The latter happens to be the nature of research that inspired our study.) Or it may be that a unique specimen will need to be sequenced to infer its phylogenetic position (Kanda *et al.*, 2015). Methods used to extract and prepare DNA for sequencing must both be more or less guaranteed to work, and, in many cases, allow for preservation of DNA for future study.

Ideal sample preparation protocols for such historical specimens fall in between existing guidelines for ancient DNA that is thousands to hundreds of thousands of years old (Gansauge & Meyer, 2013; Bennett *et al.*, 2014), and

protocols optimized for high-quality DNA, which are often used for historical specimens despite being optimized for higher-quality samples. Establishing best practices for these specimens is needed to help investigators make decisions on feasibility of research projects, resource allocation, and to improve results.

In the present study our goal was to obtain sequences from 16 small-bodied historical specimens (including type specimens) for which few or no suitable replacement specimens exist, in the context of ongoing projects in species discovery and delimitation of carabid beetles. Specimen ages range from 58 to 159 years, and many specimens had little starting DNA (less than 10 ng). In order to better understand ideal approaches of sample preparation for specimens with minimal DNA, we intentionally limited DNA input to 1–10 ng per specimen. Thus, our study samples are at the upper limits of specimen age and the lower end of DNA amounts examined in recent studies that report successful sequencing of historical specimens collected in the last two centuries (Wandeler *et al.*, 2007; Tin, Economo, & Mikheyev, 2014; McCormack *et al.*, 2015; Blaimer *et al.*, 2016). In addition, we attempted to generate libraries using several alternative methods (e.g., single-stranded versus double-stranded library kits; with or without enzymatic repair of DNA), focusing on techniques designed to improve library preparation efficiency and sequencing success of historical specimens with minimal DNA. We sequenced most of the libraries thus produced, and used various measures to explore the extent and accuracy of resulting sequences. We are hopeful that our success at generating and sequencing low-input libraries, and the guidelines for optimizing protocols presented herein, will facilitate more consistent positive outcomes for projects attempting to

sequence challenging specimens, as well as encourage more economical use of valuable DNA from irreplaceable specimens.

MATERIALS AND METHODS

A graphical summary of our methods is presented in Fig. 1.1. We extracted DNA from 16 small carabid beetle specimens (Table 1, Fig. 2.2; Table 2.1). For each specimen we attempted to generate sequencing libraries with low amounts (1–10 ng) of DNA using different protocols, then used various measures as indicators of success. Our methods are described below, with more details provided in Appendix 1.

Specimen selection

We selected historical specimens from the genera *Lionepha* Casey and *Bembidion* Latreille (Table 2.2). The nine *Bembidion* specimens include members of the *breve* group of subgenus *Plataphus* Motschulsky, the *obscuripenne* group of the *Ocydromus* Clairville complex, as well as subgenera *Trepanedoris* Netolitzky, *Notaphus* Dejean, and *Odontium* LeConte. They are either nomenclatural types or other unique specimens critical to resolving questions in carabid taxonomy. They range in size from 3.1–5.4 mm, and have been housed dried and pinned in the following museums for 58 to 159 years: National Museum of Natural History, Washington (USNM); Museum of Comparative Zoology (MCZ); Carnegie Museum of Natural History (CMNH); The Natural History Museum, London (BMNH); California Academy of Sciences (CAS); Oregon State Arthropod Collection (OSAC); Illinois Natural History Survey (INHS); University of Arizona Insect Collection (UAIC). The methods by which these specimens were collected and killed are not known.

In addition, we selected 54 specimens (Table 2.3) to provide a context for examining accuracy of sequences obtained from historical specimens. These fresh

specimens were collected intentionally for DNA studies, and those measured contain large quantities of long DNA fragments; we will refer to them as “context specimens”.

DNA extraction

We extracted DNA from historical specimens in a laminar flow hood, sterilized with UV light before each use, in a clean room designed to minimize contamination from non-target DNA. We prepared specimens following Kanda *et al.* (2015) and extracted DNA using a QIAamp DNA Micro kit (Qiagen) using the optional carrier RNA. We extracted DNA from context specimens in a standard molecular lab with DNeasy Blood and Tissue kits (Qiagen) following the manufacturer’s recommended protocol and stored eluted DNA in Buffer AE (Qiagen) at -20° F.

We measured total DNA in extractions using a Qubit Fluorometer (Life Technologies) with a Quant-iT dsDNA HS Assay Kit with 2 µl of sample, and DNA fragment length distributions with a 2100 Bioanalyzer (Agilent Technologies) using the High Sensitivity DNA Analysis Kit and 1 µl of sample.

DNARepair

We treated an aliquot of DNA from 15 historical specimens with enzymes (NEBNext® FFPE DNA Repair Mix, New England BioLabs) designed to repair nicks and damaged bases prior to library preparation as described in Appendix 1, in order to compare yield of libraries constructed with repaired DNA to that of unrepaired DNA.

Library preparation

We used two commercially available kits for library preparation. For a subset of specimens for which there was sufficient DNA, and that spanned a broad range of ages, we produced two to four libraries using different approaches so that we could compare results. We used 1–10 ng of input DNA (unrepaired or repaired) for each preparation for each of the 46 total libraries.

The first kit (NEBNext® DNA Ultra II, New England BioLabs) requires double-stranded DNA (dsDNA) as input. This kit is representative of several commercial kits having a simple protocol with minimal hands-on time, and relatively low cost per sample. This kit accommodates low amounts of input DNA (as little as 500 pg of high-quality DNA) and is optimized to produce high yields with minimal amplification, when DNA quality is high. Despite being optimized for high throughput of high-quality DNA, this kit, and similar kits from other manufacturers (e.g., Kapa Hyper kits, Kapa Biosystems) are used widely in studies cited herein that report successful sequencing of historical specimens. We refer to our basic use of this kit as the "*dsDNA*" protocol (referenced as DS and DSRep in tables and figures).

The second kit (Accel-NGS® 1S Plus, Swift Biosciences) is optimized for low input DNA (as little as 10 pg of high-quality DNA), and employs adapter ligation technology designed to improve library conversion efficiency in degraded DNA samples by conducting adapter ligation on single-stranded DNA. This kit has a straightforward protocol that requires more preparation time than the *dsDNA* kit, and costs roughly twice as much per sample. We call this protocol "*ssDNA"* (referenced as SS and SSRep in tables and figures).

After sequencing several *ssDNA* and *dsDNA* libraries, we repeated *dsDNA* libraries on four specimens using a modified protocol hereafter called the "*dsDNA Mod*" protocol (referenced as DSM and DSMRep in tables and figures), designed to minimize adapter dimers in libraries (see Appendix 1). Following sequencing of these libraries, we evaluated consistency of the modified protocol across additional samples by selecting four more specimens (specimens **2**, **3**, **8**, and **10**) to prepare with the modified protocol.

We determined library concentration with real-time PCR using KAPA SYBR® FAST qPCR Kits (KAPA Biosystems) and an Applied Biosystems® 7500 Real-Time PCR System (ThermoFisher Scientific).

Sequencing and assembly

We sequenced 28 of the 46 dual-indexed libraries, including multiple libraries for most specimens (Table 2.4), on an Illumina HiSeq 3000 operated by the Oregon State University Center for Genome Research and Biocomputing. Five HiSeq 3000 lanes were run, two of which were devoted to this study; the remaining three lanes contained samples from other studies. Each library was allocated 0.1 of a lane. All lanes yielded a total of 769 to 779 million reads. Demultiplexing of sequencing reads was performed using CASAVA version 1.8 (Illumina). We trimmed reads and conducted *de novo* and reference-based assemblies in CLC Genomic Workbench version 8.5.1 (CLC Bio) using parameters described in Appendix 1. For reference-based assemblies we used reference sequences from species that were closely related to, but not conspecific with, historical specimens, as summarized in Table 2.5.

Assessing recovery of gene targets from sequenced libraries

We measured recovery of multi-copy target regions (the mitochondrial genome and nuclear rDNA complex) from both reference-based and *de novo* assemblies. We tested recovery of 67 low-copy-number nuclear protein-coding genes previously used in arthropod phylogeny (Regier *et al.*, 2008) using a reference-based assembly approach similar to Kanda *et al.* (2015). In brief, we extracted target sequences from *de novo* assemblies by making BLAST databases of historical specimen contigs, which we probed with BLAST query sequences of target genes from reference specimens. We obtained targets from reference-based assemblies through mapping reads to reference sequences in CLC Genomics Workbench.

Assessing accuracy of sequences recovered from historical specimen libraries

We analyzed sequences obtained from historical specimens phylogenetically to assess their placement at two levels: broad-scale phylogeny of subtribe Bembidiina, and fine-scale phylogeny among closely related species. The broad-scale phylogenetic placement was examined within a five-gene dataset of 162 Bembidiina species (plus 13 outgroups) derived primarily from Maddison (2012), with additions as outlined by Kanda *et al.* (2015); genes examined include 28S rDNA, 18S rDNA, COI 5' (the barcode region of cytochrome oxidase I), CAD (a portion of the carbamoyl phosphate synthetase domain of the *rudimentary* gene), and Topo (topoisomerase I). In this analysis we tested the placement of historical sequences in their predicted clades. To examine placement at a much finer scale, we analyzed historical sequences in the context of a seven-gene-fragment data set (the five gene fragments mentioned above plus COI 3' (the remainder of COI), and COII

(cytochrome oxidase II plus a small portion of tRNA-Leu)) from 54 DNA-preserved, context specimens that include species to which historical specimens are predicted to belong, or to which they are very close, based on morphological data. For example, three context specimens belong to *Lionepha casta* (Casey), to which we predict specimens **11** and **15** also belong. Details of context specimens and their sequences are summarized in Tables 2.3 and 2.6, with our predictions for historical specimen placement relative to context specimens in Table 2.7 and Appendix 1, along with sequencing methods.

Our methods for obtaining sequences of the seven genes from historical specimens, and our phylogenetic methods are provided in Appendix 1. In short, we used the same approach for extracting the seven target genes from historical specimen reads as was used for the mtGenome and rDNA complex sequences. We analyzed multiple sequences from each historical specimen library, with sequences obtained through reference-based assembly, *de novo* assembly with 50 bases trimmed from each end, and *de novo* assembly without trimming. We processed and aligned most target sequences and context sequences in Mesquite (Maddison & Maddison 2016, 2017) supplemented by MAFFT version 7.130b (Katoh & Standley, 2013) for alignment of ribosomal genes, determined appropriate models of evolution in jModelTest version 2.1.10 (Darriba *et al.*, 2012) and PartitionFinder 1.1.1 (Lanfear *et al.*, 2012), and performed tree building in GARLI 2.0 (Zwickl, D.J., 2006).

For historical specimens **3** and **7**, we further explored sequence accuracy by aligning ref-based and *de novo* sequences to the complete mtGenome and rDNA

complex (*de novo* assembled) of several context specimens as described in Appendix 1.

RESULTS

DNA quantity and the effect of enzymatic repair

DNA extractions from study specimens yielded 5.3 ng to 231.3 ng of DNA per specimen (Table 2.2). The five specimens from the 1800s included the extreme values in extraction yield.

Enzymatic repair of DNA and the associated bead cleanup resulted in an average loss of 41.3% (SD = 18.0%) of starting DNA by mass (Table 2.8). Fragment sizes ranged from 600 bases in specimen **5** to being too short to visualize (or too dilute) in specimen **7**. For most specimens, the majority of fragments were less than 200 bases, with maximum fragment sizes between 200–400 bases.

Success of library preparation

We successfully generated sequencing libraries for all historical specimens using only 1–10 ng of DNA input in each library (3.58 ng per specimen on average). All our comparative trials produced libraries with enough DNA to be sequenced, regardless of whether the *dsDNA* or *ssDNA* protocol was used, resulting in a total of 46 sequenceable libraries across the 16 specimens (Tables 2.4 and 2.9). Libraries were considered successful if preparation yielded a concentration of at least 2.0 nM, which is the minimum required for the Illumina HiSeq 3000. Library concentrations ranged from 3.55–33.06 nM (Table 2.9).

Libraries prepared with enzyme-repaired DNA showed higher yields than libraries produced with unrepaired DNA, given the same number of amplification cycles, in both *ssDNA* and *dsDNA* approaches. Repaired *ssDNA* libraries produced an average of 29.2% more library ($p=0.015$, paired T-test) than unrepaired libraries, and

repaired *dsDNA* libraries produced 38.9% more (p=0.004), for an average yield increase of 36.3% across all libraries (Fig. 2.3; Table 2.10, Supporting Information).

Bioanalyzer traces of libraries showed the presence of unwanted small fragments (<180 bases; presumably primer dimers and adapter dimers) in 38 of 46 libraries. Bead cleanup of those libraries resulted in an average loss of 48.7% (SD=13.6%) of library per cleanup (Table 2.11). Although a single bead cleanup was generally sufficient to remove unwanted small fragments, two libraries required a second cleanup, and for four libraries (LIB0161, LIB0162, LIB0166, LIB0172) three cleanups were not sufficient. We made new libraries to replace LIB0161, LIB0162, and LIB0166 using the *dsDNA Mod* protocol with enzymatic repair. We did not attempt to redo LIB0172.

Trimming and assembly

Reads removed through trimming of adapters and low-quality bases ranged from 1.03% to 38.5% of total reads for a library (Table 2.12). The great majority of reads removed (> 99.8% in every library) were due to adapter sequences in reads. Libraries generated with the *ssDNA* protocol lost 1.86% (SD = 0.8%) of reads to trimming on average. This contrasts with libraries generated with the *dsDNA* protocol, which lost an average of 22.59% (SD = 14.0%) of reads. Reads lost to trimming in *dsDNA Mod* were reduced to 6.20% (SD = 4.26%) on average relative to *dsDNA*. Fig. 2.4 shows example traces of adapter content in reads for each preparation protocol. N50 of *de novo* assemblies of sequenced libraries ranged from 177 to 969 (Table 2.12).

Recovery of mtGenome and rDNA complex sequences

We recovered >9,000 bases of the mtGenome and >5,000 bases of the rDNA complex from reference-based assemblies from all 28 low-input DNA libraries sequenced. We recovered at least 4,000 bases of the mtGenome from *de novo* assemblies in 25 of the 28 libraries sequenced, and at least 3,000 bases of the rDNA complex from 22 of the 28 libraries sequenced (Fig. 2.5; Table 2.13).

Average coverage in reference-based assemblies ranged from 4.4–15,263x for the mtGenome, and 5.4–1,339x for the rDNA complex (Table 2.13). Libraries from specimens **7** and **9** (each 98 years old) and specimen **1** (159 years old) showed the poorest recovery of multi-copy targets. Libraries from two of the oldest specimens, specimen **2** (159 years old) and specimen **4** (136 years old) had among the highest rates of recovery of one or both multi-copy targets with both specimens showing greater than 99% recovery of the mtGenome; specimen **4** had greater than 86% recovery of the rDNA complex. *Lionepha* specimens **1**, **2**, **3**, **11**, **15**, and **16** showed somewhat reduced rates of recovery in ref-based assembly of the rDNA complex, likely as a result of extensive rDNA indels between species of *Lionepha* (Maddison and Sproul, unpublished).

Recovery of 67 low-copy nuclear protein coding genes

Fig. 2.6 shows a heat map summarizing recovery in reference-based assemblies of 67 low-copy nuclear protein-coding genes from libraries. Average recovery across all genes ranged from 0.5%–64.7% of target bases recovered. Specimens of similar age varied widely in recovery success. Libraries from the five specimens >125 years old ranged from 0.7–34.1% average recovery, the four 97 or

98-year-old specimens ranged from 0.5–51.8% recovery, and the three specimens 58–77 years old ranged from 23.3%–64.7% average recovery.

Comparison of target recovery in dsDNA vs ssDNA protocols

The *dsDNA* and *ssDNA* protocols did not differ significantly in many of the sequencing success metrics. The *ssDNA* protocol retained 16.8% more reads ($p=0.004$, paired T-test). No other differences were statistically significant; however, the *dsDNA* protocol showed a non-significant 10.4% increase in recovery of rDNA target sequences, and a 6.5% increase in the number of unique reads that mapped to the mtGenome and rDNA complex. The remaining three metrics showed less than 1.6% difference between protocols. Fig. 2.7 summarizes relative success metrics for *dsDNA* and *ssDNA* sequencing.

Accuracy of sequences recovered: Phylogenetic analysis and multi-copy sequence comparison

Our broad-scale phylogenetic analysis placed sequences from all 28 historical specimen libraries with their near relatives in the five-gene concatenated analysis (Fig. 2.8). Single-gene analyses also accurately assigned sequences from all historical specimen libraries to group (Figs. 2.9–2.11) with two exceptions: for specimen **1**, the ref-based sequence from LIB0153 was sister to the rest of *Lionepha* in 28S (Fig. 2.9, arrow), and all sequences from two of specimen **7**'s three libraries occurred on long branches outside of the *breve* group in COI 5' (Fig. 2.10, arrows). There were several instances in which the *de novo* sequences with untrimmed ends occur on longer branches than reference-based or default (trimmed) *de novo* sequences from the same library (e.g., specimen **4** LIB0168 and specimen **5** LIB0149 and LIB0186 in Fig. 2.8).

Our fine-scale analysis of closely related species showed that sequences obtained from most historical specimens were either identical to sequences obtained from putatively conspecific specimens with high-quality DNA, or showed branch length variation consistent with what might be expected from intraspecific variation (Figs. 2.12–2.16; predictions outlined in Table 2.7). In some cases, however, sequences from historical specimen libraries occurred on long branches within their predicted clade. In particular, sequences from multiple libraries for specimens **1**, **5** and **7** occur on long branches. This may be due to sequence degradation, sequence and assembly errors related to low coverage, or both.

For *Lionepha* and the *breve* group (specimens **1–3, 11, 7, 9, 15, 16**), we have sufficient background data to be confident that our context specimens include representatives of species to which the historical specimens belong, as explained in Appendix 1. We were able to unambiguously assign seven of eight historical specimens in these groups to species based on their consistent placement across several gene trees (Figs. 2.12 and 2.13, Tables 2.14 and 2.15). Available sequences of specimens **1–3** and **16** of *Lionepha* were placed in a clade with *Lionepha chintimini* (Erwin and Kavanaugh), or *Lionepha chintimini* + *Lionepha* "Carson Spur" in all seven gene trees, with a single exception being the reference-based sequence of Topo for *Lionepha* specimen **1** LIB0153, which was on a long branch and sister to the clade containing *Lionepha* "Carson Spur" and *Lionepha chintimini*. All sequences analyzed from *Lionepha* specimens **11** and **15** libraries were recovered in a clade with *Lionepha casta* or *L. casta* + *L.* "Bitterroots" (Fig. 2.12 and Table 2.14).

In the *breve* group, specimen **9** was placed with *Bembidion* "University Peak" in all five gene trees for which data were available. Sequences from specimen **7** showed the least consistent placement of any specimen, especially sequences from LIB0182. Sequences from LIB0178 and LIB0179 reference-based assemblies were placed in a clade with *B.* "Ebbets Pass" in COII, where they were identical to two of the three context specimens. Outside of these and one other sequence (LIB0178 Ref-Based in COI 5'), the sequences occur on long branches, and fail to form consistent clades among themselves, and not with a context species (Fig. 2.13 and Table 2.15). Although we were unable to confidently assign specimen **7** to a species using gene trees due to variable placement, we were able to assign it to *B.* "Ebbets Pass" through comparison of the complete mtGenome and rDNA complex to those of context specimens *B.* "Ebbets Pass" 4161 and *B. laxatum* Casey 4918. Specimen **7** matched the former at 335 sites that distinguish the two candidate species, compared to 44 at which it matched *B. laxatum* (Fig. 2.17). This method proved even more successful at assigning specimen **3** (Fig. 2.18).

DISCUSSION

In this study, we were consistently able to generate and successfully sequence libraries from 1–10 ng of DNA (most had 1–5.5 ng) for small, dried insects 58–159 years old. We are unaware of another study that demonstrates consistent success with such low input using commercial kits. The extent and accuracy of sequences obtained from a large subset of libraries validate our sample preparation methods, as we had good recovery of multi-copy genes, and for several specimens, low-copy genes that are effective at placing historical specimens to species. This result implies that low DNA content should not be viewed as an insurmountable barrier to sequencing unique historical specimens. A number of studies report abandoning a subset of specimens for which too little input DNA was available (i.e., below 10 ng in Kanda *et al*. 2015, below 5 ng in Blaimer *et al*. 2016), a practice not supported by our results.

Although taking a low-input approach preserves valuable DNA, it compromises library diversity. Effective sequence assembly depends upon sequencing a diverse pool of DNA fragments that overlap at a given locus. Historical specimens may already have a reduced pool of fragments that are of a suitable length for sequencing due to effects of DNA degradation, and sampling a small fraction of those suitable fragments may further reduce the diversity of fragments going into the library, which potentially compromises assembly quality. Despite this, we successfully recovered multi-copy regions from all specimens (Fig. 2.5) and substantial fractions of nuclear protein-coding targets for several specimens (Fig. 2.6). Our low-coverage, whole genome sequencing approach is well suited to studies for which recovery of multi-

copy genes will be sufficient, and for which relatively low numbers of specimens are being targeted for sequencing.

If recovery of many nuclear protein genes is critical, studies using a targeted sequencing approach such as hybrid capture have reported successful recovery of many low-copy genes for some specimens (McCormack *et al.*, 2015; Blaimer *et al.*, 2016). However, the variable nature of sample quality among historical specimens presents challenges to steps that involve multiplexed hybridization and sequencing reactions, and can lead to notable underrepresentation of reads from a subset of libraries in multiplexed hybrid capture and sequencing (Hawkins *et al.*, 2015). Whether or not whole-genome shotgun sequencing or a hybrid capture approach is employed, library preparation is needed as the first step, and our sample preparation guidelines (see Appendix 2) may be applied to improve success of challenging samples.

Sequence accuracy

Sequences generated from our historical specimens were generally of sufficient quality for consistent phylogenetic placement, including at fine scales of evolutionary divergence (Figs. 2.8–2.16, Table S5, Supporting Information). In most cases, sequences were identical or extremely similar to those from context specimens. This validates our approach as a strategy for projects attempting to assign specimens to species, as is sought in several studies that include historical specimens (Hind *et al.*, 2015; Kanda *et al.*, 2015; McCormack *et al.*, 2015; Carmi *et al.*, 2016). Only two specimens (**1** and **7**) showed sequences notably divergent from context specimens, although in almost all cases they were still closer to their putative relatives than to

other species. It is likely the observed sequence difference is due to low DNA quality or degradation, and sequencing and assembly errors that persist due to low coverage. This is consistent with poor sequence recovery and coverage for these specimens (Figs. 2.5 and 2.6; Table 2.13).

Although the sequence quality of specimen **7** prevented our assigning it to species through gene tree analysis, we were able to do so through comparison of complete mtGenome and rDNA complex sequences with sequences we obtained from fresh specimens of candidate species (Fig. 2.17). This approach of comparing thousands of bases of multi-copy regions, which we also successfully used for specimen **3** (Fig. 2.18), allowed us to see past the poor sequence quality of specimen **7** that clouded results of individual gene trees, and place to species an irreplaceable, but challenging historical specimen.

Preserving DNA from irreplaceable historical specimens

Our findings provide a reference for planning DNA preservation from irreplaceable samples. Because we consistently produced successful libraries with less than 10 ng of DNA, we did not need to use all the DNA of a specimen. Using a small fraction of available DNA may allow for multiple attempts should the first attempt fail. More importantly, it allows for archiving a large fraction of DNA. A decision to extract and sequence DNA from an historical specimen must be made in anticipation of future research and technological advances, and not just the goals of one project. Subjecting a specimen to an inefficient library preparation attempt could result in the loss of all or most DNA from that specimen. Studies that seek to obtain sequences from irreplaceable specimens should include as a primary objective

preservation of the majority of DNA for archival purposes. We utilized an average of 13% (and as little as 2.1%) (Table 2.4) of available DNA as library preparation input, which demonstrates that library preparation success and DNA conservation are not mutually exclusive, even for specimens with minimal DNA. Continued discussion is needed within and among institutions housing historical specimens to establish guidelines related to permission and archival requirements of studies that attempt to sequence unique, historical specimens. In addition, an ongoing evaluation of DNA extraction protocols is needed to establish best practices as methods improve.

Optimizing protocols for low input

Many protocol optimizations exist for ancient DNA obtained from specimens thousands to hundreds of thousands of years old (Pääbo *et al.*, 2004; Gansauge & Meyer, 2013; Bennett *et al.*, 2014; Cruz-Dávalos *et al.*, 2016). Ancient DNA approaches are inappropriate or unnecessary for historical specimens collected in recent centuries; however, for specimens containing low DNA content, unmodified protocols for fresh DNA may also be inappropriate. Existing studies from which guidelines can be gleaned are biased towards specimens for which tissue is plentiful (see reviews by Wandeler *et al.*, 2007; Burrell *et al.*, 2015), or for RAD-seq data (Tin *et al*., 2014), rather than small specimens for which specific gene targets are sought, as in the current study.

The various protocols we used successfully generated low-input libraries, in part through modifications to the kit manufacturers' recommended protocols (see Appendix 2). For both *ssDNA* and *dsDNA* protocols, we found it necessary to increase amplification cycle number. We observed increased library yield with both

protocols when we repaired DNA prior to library construction (Fig. 2.3; Table 2.10). The greatest disadvantage of DNA repair was DNA loss associated with bead cleanup following repair (Table 2.8); however, updated protocols are available that eliminate the bead cleanup (personal communication with New England Biolabs) and may prove useful for low-DNA historical specimens.

A comparison of sequencing success showed only minor differences between protocols, except for reads retained following adapter trimming, for which the *ssDNA* protocol performed significantly better (Fig. 2.7). This is not surprising given that the *ssDNA* protocol is specialized for low-input samples, and includes up to three bead cleanups in the steps prior to library amplification, compared to a single cleanup step in the *dsDNA* protocol. Others have noted the challenges posed by adapter dimers in historical specimen libraries (Burrell *et al.,* 2015; Tin *et al.,* 2014) and discuss the need for elimination of small fragments with a bead cleanup prior to amplification. Although the manufacturer's *dsDNA* protocol includes a bead cleanup step, we found that several samples still contained adapter sequences in up to 38.5% of reads (Table 2.4). We found that much of the disparity in adapter reads retained between kits was eliminated by adding a second cleanup step to the *dsDNA* protocol prior to library amplification (Fig. 2.4 and Table 2.4).

We had expected to see more striking differences between protocols than was measured, in part as the *ssDNA* protocol is designed for low-input, degraded DNA, whereas the *dsDNA* protocol is not. The lack of distinction we saw may in part be due to our relatively low sample sizes, but it is surely also caused by our unexpectedly widespread success with the *dsDNA* protocol.

Based upon our results, we suggest the following optimizations designed to increase library success: (1) enzymatic repair of DNA as a strategy for improving library yield for challenging specimens if DNA quantity is sufficient; (2) pelleting beads on the side of tubes (rather than in ring) during bead cleanups, avoiding protocol-recommended fixed drying times of beads prior to elution, and maximizing the volume of eluate transferred to later steps; (3) diluting adapters beyond the manufacturer's protocol, if needed, to reduce adapter dimers in libraries and sequencing reads; (4) two bead cleanups prior to library amplification for low-DNA-input, high-amplification-cycle-number libraries; (4) adding 2–5 amplification cycles beyond manufacturer's recommendations. These guidelines are described in more detail in Appendix 2. In principle, these optimizations can be applied to a variety of commercial kits, and for libraries destined for both whole genome shotgun sequencing and hybrid capture.

Conclusions

Despite advances in DNA sequencing technology that allow sequencing many historical specimens, the variable nature of DNA quality and quantity acquired from these specimens can leave investigators unsure of how to proceed with important samples that have low DNA quantity and quality. We show that taking a low-input approach can allow investigators to preserve extracted DNA while still obtaining sequences from irreplaceable historical specimens. We suggest optimizations related to DNA repair, bead handling, reducing adapter dimers, and library amplification, which we used to consistently generate successful sequencing libraries from small-bodied organisms with low amounts of degraded DNA, including specimens dating

back to the mid 1800s. We hope our guidelines facilitate more confident initiation of projects among investigators wanting to sequence historical specimens for the first time, and increase the consistency of positive outcomes of libraries in labs and core facilities already sequencing historical specimens. Establishing accessible, cost-effective sample preparation guidelines that increase likelihood of successful preparation and sequencing of historical specimens will further increase the utility of the vast resources in museum collections.

ACKNOWLEDGEMENTS

We thank James Pflug, Kojun Kanda, and Jeff Oliver for help with data analysis. We thank Kojun Kanda for reviewing the manuscript. We thank David Kavanaugh, Wendy Moore, Estève Boutaud, Antonio Gomez, Olivia Boyd, Pietro Brandmayr and Jose Serrano for helpful discussions related to our findings. We are indebted to the technical support staff from New England Biolabs and Swift Biosciences for several hours of discussion as we implemented optimizations that deviated from their published protocols. We thank Terry Erwin (USNM), Philip Perkins (MCZ) Beulah Garner and Max Barclay (BMNH), Robert Davidson (CMNH), David Kavanaugh (CAS), Christopher Dietrich (INHS), and Wendy Moore (UAIC) for providing historical specimens. Thanks to Robert Davidson for providing information about the collection date for the paratype of *Bembidion ulkei*. For help in collecting context specimens, we thank K.W. Will, J.N. Caira, A. Gill, J.R. LaBonte, J.H.A. Maddison, A.E. Arnold, K. Kanda, K.T. Eldredge, W.Moore, S.D. Schoville, G.J. Binford, M. Lahti, and D.H. Kavanaugh.

This work was funded in part by the Harold E. and Leona M. Rice Endowment Fund at Oregon State University, as well as National Science Foundation grant DEB-1258220 to DRM.

REFERENCES

Bates HW. 1878 On new genera and species of geodephagous Coleoptera from Central America. *Proceedings of the Scientific Meetings of the Zoological Society of London* 1878, 587-609.

Bennett EA, Massilani D, Lizzo G, Daligault J, Geigl EM, Grange T. 2014. Library construction for ancient genomics: single strand or double strand. *BioTechniques* 56: 289–90, 292–6, 298.

Besnard G, Bertrand JA, Delahaie B, Bourgeois YX, Lhuillier E, Thébaud C. 2015. Valuing museum specimens: high-throughput DNA sequencing on historical collections of New Guinea crowned pigeons (Goura). *Biological Journal of the Linnean Society*.

Besnard G, Christin PA, Malé PJG, Lhuillier E, Lauzeral C, Coissac E, Vorontsova MS. 2014. From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition. *Journal of Experimental Botany* 65: 6711–6721.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* 22: 6018–6032.

Blaimer BB, Lloyd MW, Guillory WX, Brady SG. 2016. Sequence Capture and Phylogenetic Utility of Genomic Ultraconserved Elements Obtained from Pinned Insect Specimens. *PLoS One* 11: e0161531.

Burrell AS, Disotell TR, Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* 79: 35–44.

Carmi O, Witt CC, Jaramillo A, Dumbacher JP. 2016. Phylogeography of the Vermilion Flycatcher species complex: multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. *Molecular Phylogenetics and Evolution*.

Cruz-Dávalos DI, Llamas B, Gaunitz C, Fages A, Gamba C, Soubrier J, Librado P, Seguin-Orlando A, Pruvost M, Alfarhan AH. 2016. Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular Ecology Resources*.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.

Erwin TL, Kavanaugh DH. 1981. Systematics and zoogeography of Bembidion Latreille: 1. The carlhi and erasum groups of western North America (Coleoptera: Carabidae, Bembidiini). *Entomologica Scandinavica Supplement* 15: 33–72.

Gansauge MT, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* 8: 737–748.

Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gilissen E, Sonet G, Nagy ZT. 2013. Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology* 62: 539–554.

Hawkins MT, Hofman CA, Callicrate T, McDonough MM, Tsuchiya MT, Gutiérrez V, Helgen KM, Maldonado JE. 2015. In-solution hybridization for mammalian mitogenome enrichment: pros, cons and challenges associated with multiplexing degraded DNA. *Molecular Ecology Resources*.

Hind KR, Miller KA, Young M, Jensen C, Gabrielson PW, Martone PT. 2015. Resolving cryptic species of Bossiella (Corallinales, Rhodophyta) using contemporary and historical DNA. *American Journal of Botany* 102: 1912–1930.

Hofreiter M, Paijmans JL, Goodchild H, Speller CF, Barlow A, Fortes GG, Thomas JA, Ludwig A, Collins MJ. 2015. The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays* 37: 284–293.

Kanda K, Pflug JM, Sproul JS, Dasenko MA, Maddison DR. 2015. Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. *PLoS One* 10: e0143929.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–80.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.

Lim HC, Braun MJ. 2016. High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. *Molecular Ecology Resources* 16: 1204–1223.

Lindroth CH. 1963. The ground beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska, Part 3. *Opuscula Entomologica Supplementum*: 201–408.

Lindstrom SC, Gabrielson PW, Hughey JR, Macaya EC, Nelson WA. 2015. Sequencing of historic and modern specimens reveals cryptic diversity in Nothogenia (Scinaiaceae, Rhodophyta). *Phycologia* 54: 97–108.

Maddison DR. 2012. Phylogeny of Bembidion and related ground beetles (Coleoptera: Carabidae: Trechinae: Bembidiini: Bembidiina). *Molecular Phylogenetics and Evolution* 63: 533–576.

Maddison DR, Cooper KW. 2014. Species delimitation in the ground beetle subgenus Liocosmius (Coleoptera: Carabidae: Bembidion), including standard and next-generation sequencing of museum specimens. *Zoological Journal of the Linnean Society* 172: 741–770.

Maddison WP, Maddison DR. 2016. Mesquite: a modular system for evolutionary analysis. version 3.10. http://mesquiteproject.org

Maddison WP, Maddison DR. 2017. Mesquite: a modular system for evolutionary analysis. Version 3.2. Available: http://mesquiteproject.org.

Mason VC, Li G, Helgen KM, Murphy WJ. 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research* 21: 1695–1704.

McCormack JE, Tsai WL, Faircloth BC. 2015. Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*.

Mutanen M, Kekkonen M, Prosser SW, Hebert PD, Kaila L. 2015. One species in eight: DNA barcodes from type specimens resolve a taxonomic quagmire. *Molecular Ecology Resources* 15: 967–984.

Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. 2004. Genetic analyses from ancient DNA. *Annual Reviews in Genetics* 38: 645–679.

Prosser SW, deWaard JR, Miller SE, Hebert PD. 2016. DNA barcodes from century-old type specimens using next-generation sequencing. *Molecular Ecology Resources* 16: 487–497.

Regier JC, Shultz JW, Ganley ARD, Hussey A, Shi D, Ball B, Zwick A, Stajich JE, Cummings MP, Martin JW, Cunningham CW. 2008. Resolving Arthropod Phylogeny: Exploring Phylogenetic Signal within 41 kb of Protein-Coding Nuclear Gene Sequence. *Systematic Biology* 57: 920–938.

Sheffield N, Song H, Cameron S, Whiting M. 2008. A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. *Molecular Biology and Evolution* 25: 2499–2509.

Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One* 8: e69189.

Tin MMY, Economo EP, Mikheyev AS. 2014. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLoS One* 9: e96793.

Wandeler P, Hoeck PE, Keller LF. 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* 22: 634–642.

Zwickl, D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Unpublished thesis, The University of Texas.

DATA ACCESSIBILITY

Sequence assembly.fas files are deposited in DRYAD (entry doi: 10.5061/dryad.31g05).

Sequence read files are deposited in NCBI Sequence Read Archive (accessions SAMN06276505 – SAMN06276522). DNA sequences included in phylogenetic analyses are deposited in Genbank (accessions KY246628 - KY246894).

Matrices and results from phylogenetic analysis are deposited in DRYAD (entry doi: 10.5061/dryad.31g05).

AUTHOR CONTRIBUTIONS

Conceived and designed experiments: JSS and DRM. Performed experiments: JSS. Analyzed the data: JSS and DRM. Contributed reagents/materials/analysis tools: DRM. Wrote the manuscript: JSS and DRM.

FIGURES

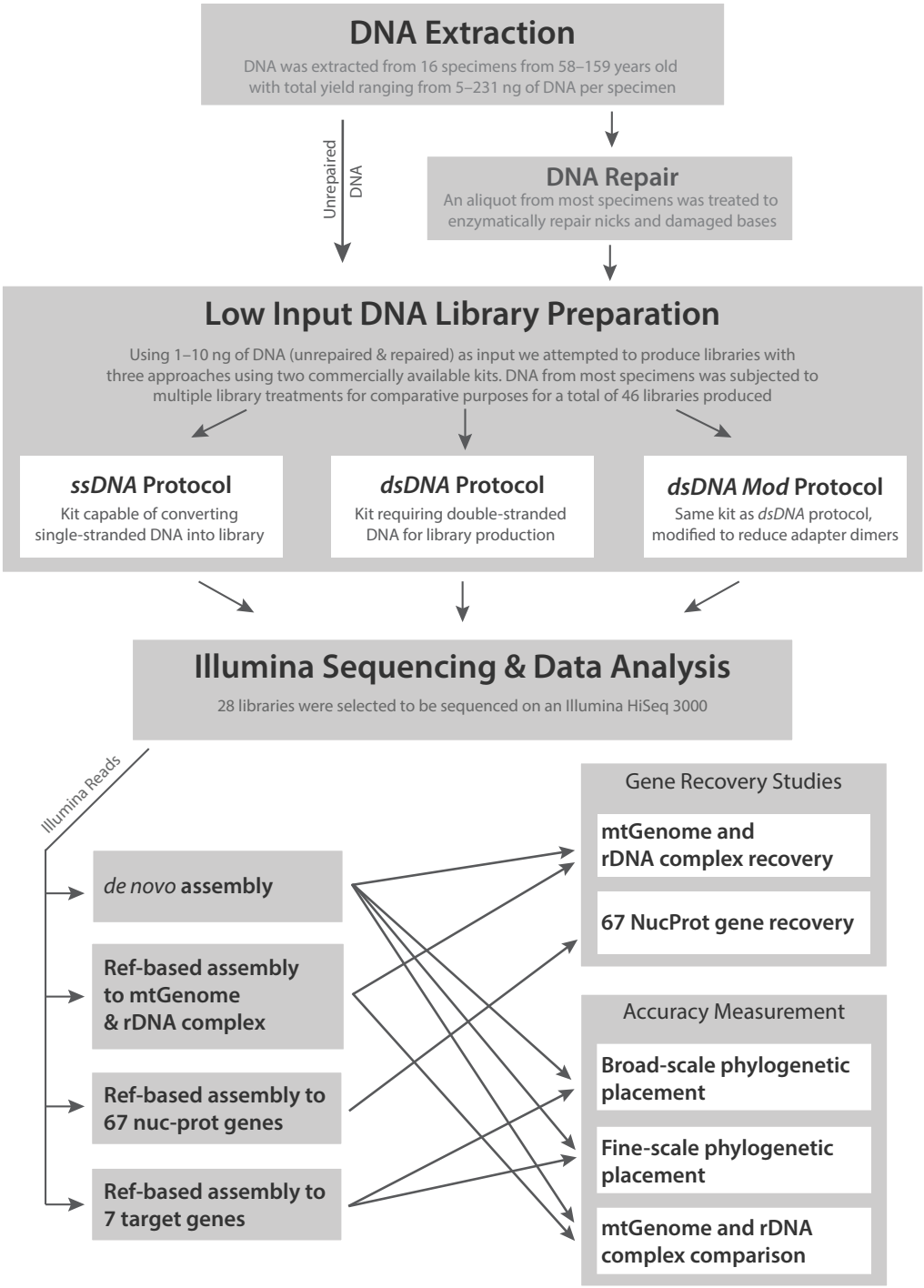


Fig. 2.1. Flowchart overview of our methodological approach.

Fig. 2.2. Habitus images of historical specimens.
Habitus images of specimens 7 (top) and 3 (bottom) after DNA extraction. Ruler on left shows 1 mm increments.

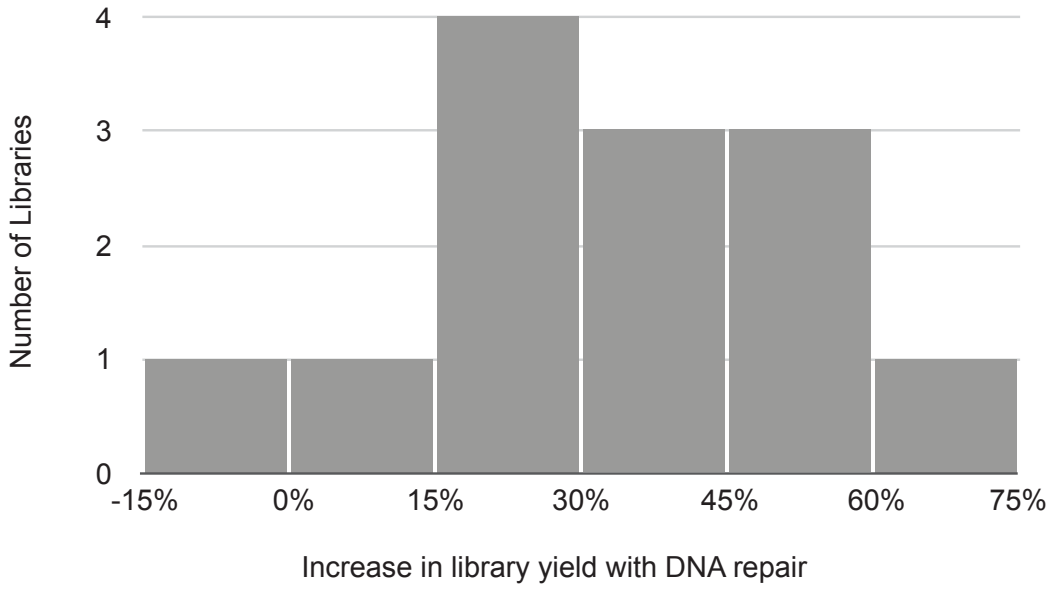


Fig. 2.3. Histogram comparing library yields of repaired vs non-repaired libraries. Distribution of differences in yield between repaired and non-repaired libraries.

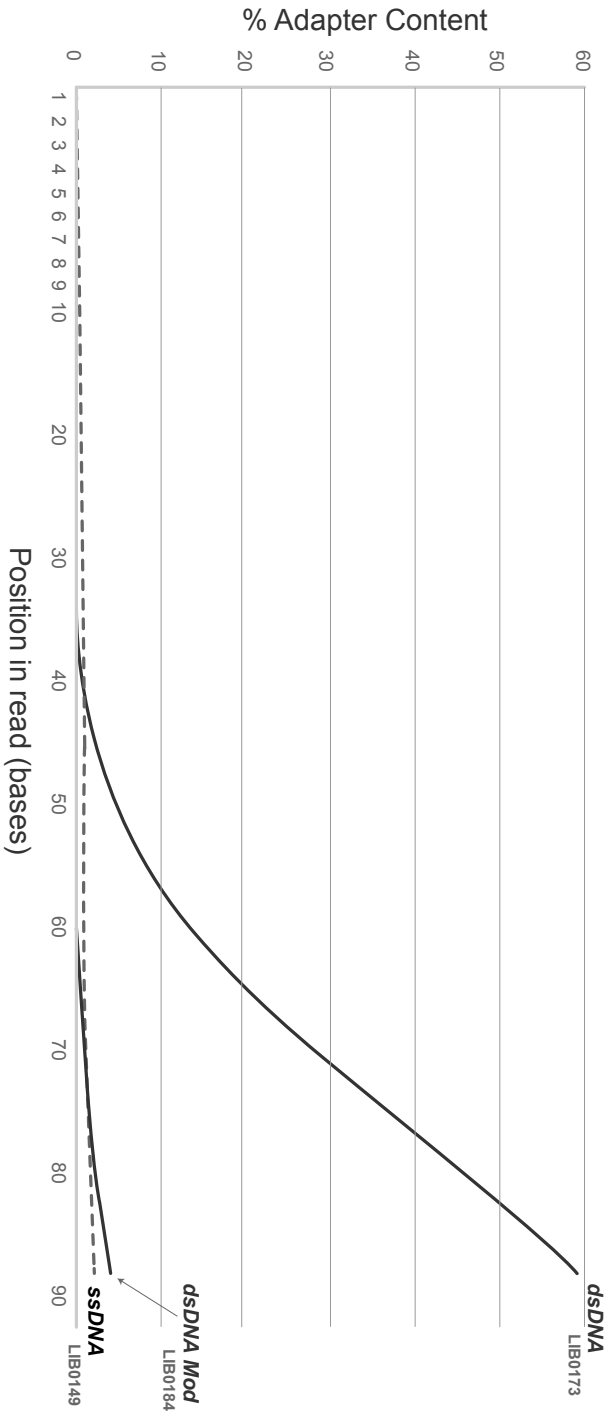


Fig. 2.4. Comparison of adapter sequences in reads for *ssDNA*, *dsDNA*, and *dsDNA Mod* libraries. Traces show the percentage of reads with adapter sequence present at given position in the read for each of three libraries. These three traces were representative of general patterns seen in libraries generated with these preparation protocols.

Specimen	Library	Age	Protocol	Mitochondrial Genome		rDNA Complex	
				de novo	ref-based	*de novo*	ref-based
1	LIB0153	159	SSRep	4107	11759	322	6159
1	LIB0154	159	SS	1607	9854	165	5864
1	LIB0184	159	DSMRep	4002	12382	2922	7298
2	LIB0213	159	DSMRep	14489	14889	3228	7216
3	LIB0210	157	DSM	6798	12847	7196	7408
4	LIB0167	136	SS	13276	14818	14244	12896
4	LIB0168	136	SSRep	12821	14819	13503	12896
4	LIB0173	136	DSRep	10728	14816	13438	13521
5	LIB0149	126	SSRep	13879	14391	4682	12419
5	LIB0150	126	SS	13556	14392	4793	12309
5	LIB0185	126	DSMRep	13443	14388	6855	12543
5	LIB0186	126	DSM	9636	14348	4188	12485
7	LIB0178	98	SSRep	2478	10167	1935	12855
7	LIB0179	98	SS	1517	10072	1295	12099
7	LIB0182	98	DSRep	11948	14743	3300	13458
9	LIB0177	98	SSRep	12652	14689	2174	12060
9	LIB0181	98	DSRep	12764	14702	9218	13594
11	LIB0176	98	DSRep	12540	14689	6435	8300
12	LIB0171	97	SSRep	14593	14228	10671	7582
12	LIB0175	97	DSRep	13587	14218	11233	7692
13	LIB0147	77	SSRep	13966	14396	11777	12376
13	LIB0157	77	DSRep	14097	14384	11204	12479
15	LIB0155	64	SSRep	14827	14889	8027	7473
15	LIB0164	64	DSRep	13195	14804	7775	7653
16	LIB0180	58	SSRep	14830	14893	7558	7725
16	LIB0183	58	DSRep	14831	14887	7521	7751

Fig. 2.5. Recovery of mtGenome and rDNA-complex targets. Cell darkness in *de novo* and ref-based columns corresponds to the percentage of target fragment length recovered, with black corresponding to 100% recovery and white to 0%. Numbers in shaded cells indicate total bases recovered. For protocol abbreviations, see Table 2.4.

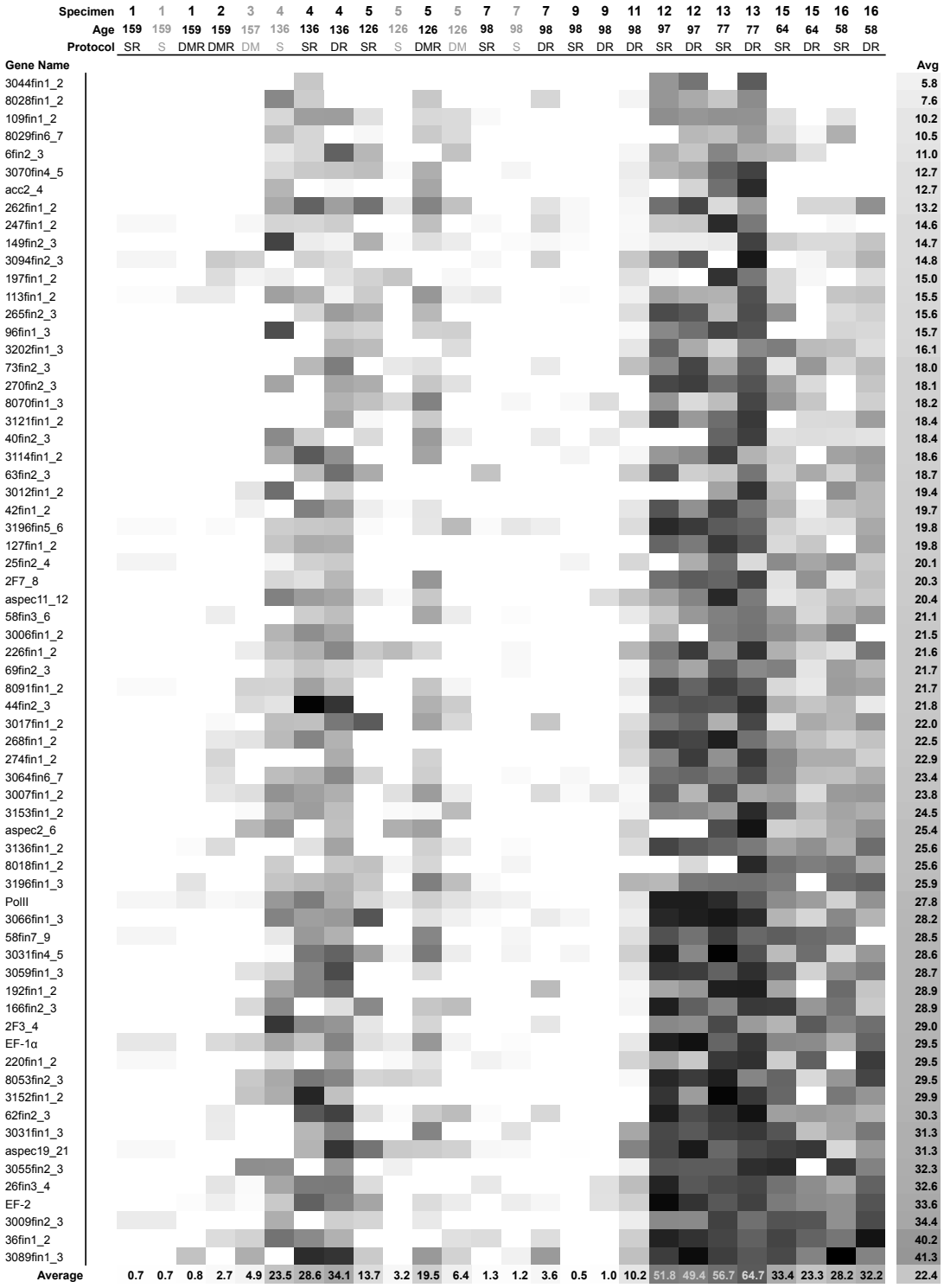


Fig. 2.6. Recovery of 67 low-copy nuclear protein-coding genes. Cell darkness corresponds to percentage of the target fragment length recovered, with black cells

Fig. 2.6. (Continued)

corresponding to 100% recovery and white to 0%. Gene fragments are ordered by average recovery as measured from reference-based assemblies. Gene abbreviations are those used in Regier *et al*. (2008). At the top of each column in the heatmap is the specimen number, then specimen age in years, and the library preparation protocol abbreviation (S: SS protocol, SR: SSRep, DR: DSRep, DMR: DSMRep, DR: DSRep). Libraries prepared with unrepaired DNA are indicated with column titles in gray text.

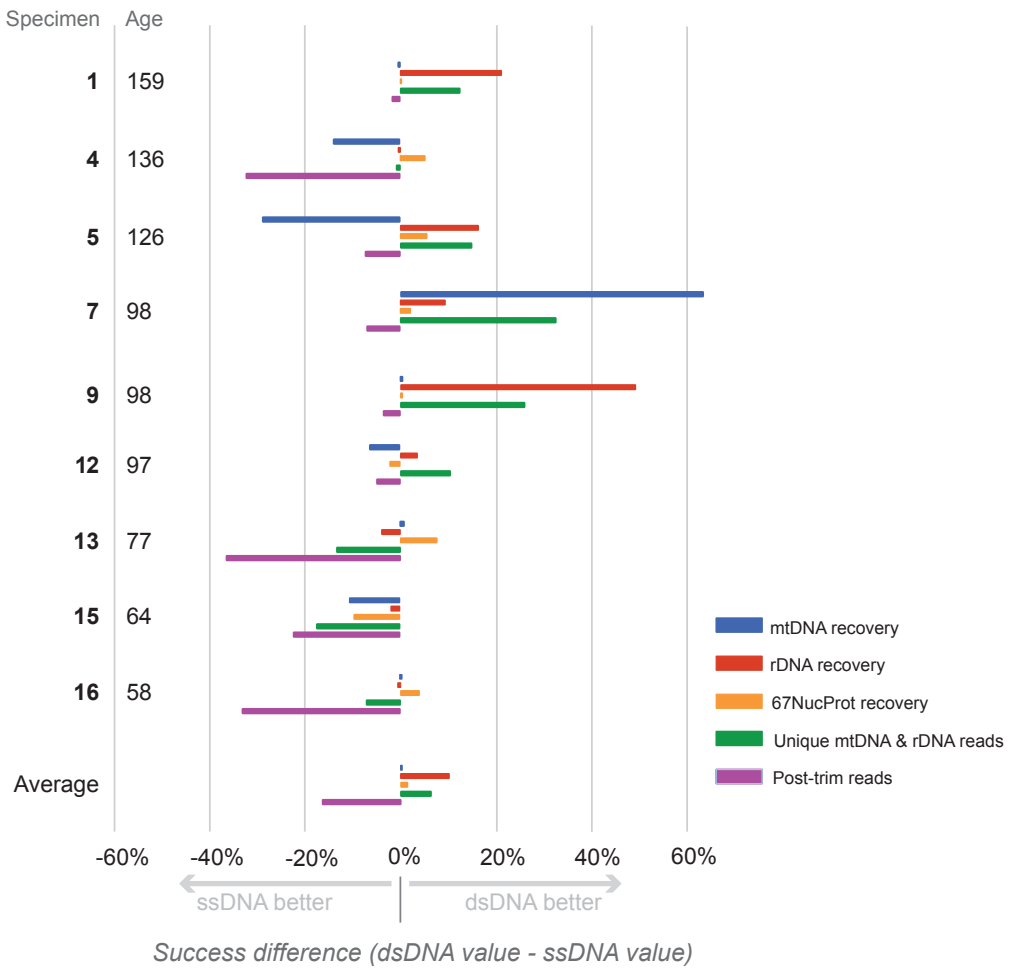


Fig. 2.7. Comparison of sequencing success metrics from *ssDNA* and *dsDNA* libraries. Colored bars for each specimen show *dsDNA* library sequencing success relative to the *ssDNA* library for that specimen, across five metrics of success. mtDNA recovery: the percent of the mtGenome target length recovered from *de novo* assemblies. rDNA recovery: the percent of the rDNA complex target length recovered from *de novo* assemblies. 67NucProt recovery: the average percent of target length recovered across all 67 low-copy nuclear-protein coding genes. Plotted values were calculated by subtracting the value of a success metric for the *ssDNA* library from the *dsDNA* library value of the same specimen.

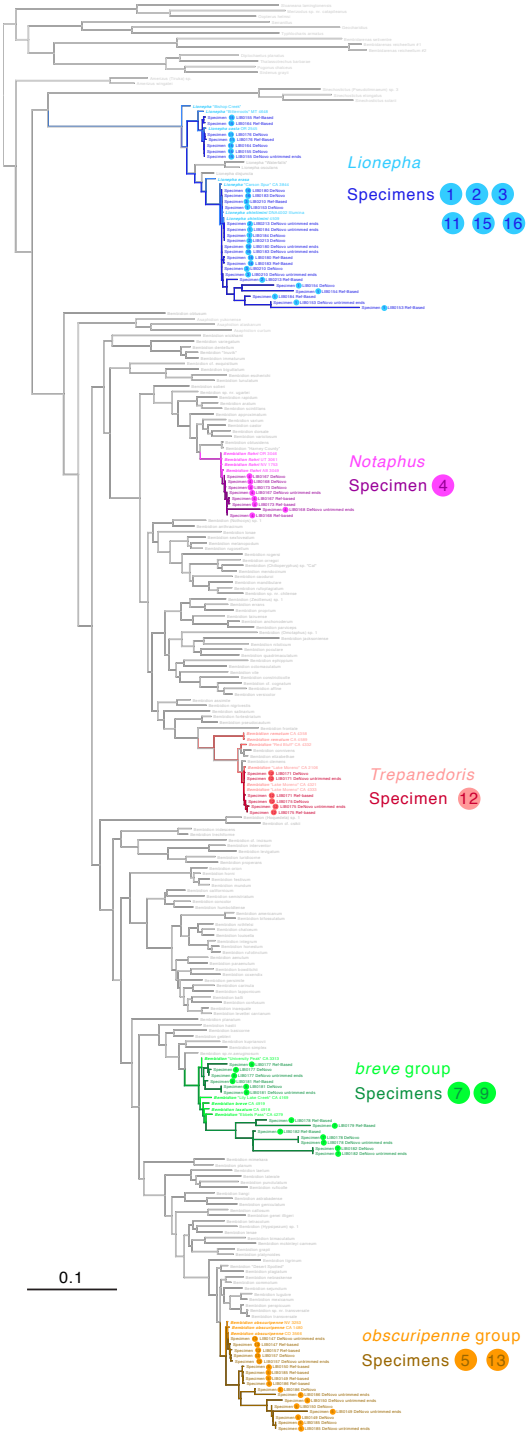


Fig. 2.8. Broad-scale phylogenetic placement of historical specimens. Five-gene concatenated maximum likelihood tree of historical specimens in the context of the phylogeny of subtribe Bembidiina with historical specimen sequences shown in dark color shades and candidate species sequences in light color shades. Branch length is shown proportional to relative divergence, as estimated by GARLI.

Fig. 2.9. Maximum Likelihood gene trees for 28S and 18S. Bembidiina as a whole shown, showing placement of context specimens (dark colors) and historical specimen sequences (light colors). Branch length is shown proportional to relative divergence, as estimated by GARLI.

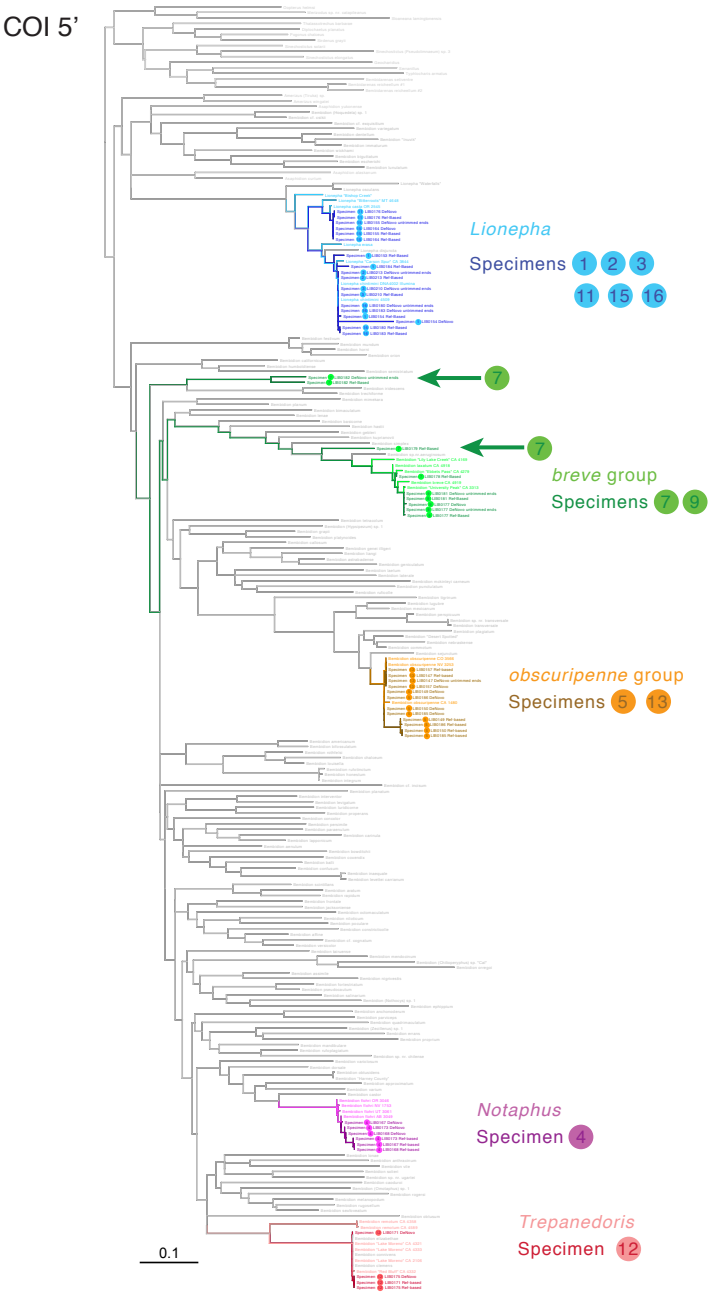


Fig. 2.10. Maximum Likelihood gene trees for COI 5'. Bembidiina as a whole shown, showing placement of context specimens (dark colors) and historical specimen sequences (light colors). Branch length is shown proportional to relative divergence, as estimated by GARLI.

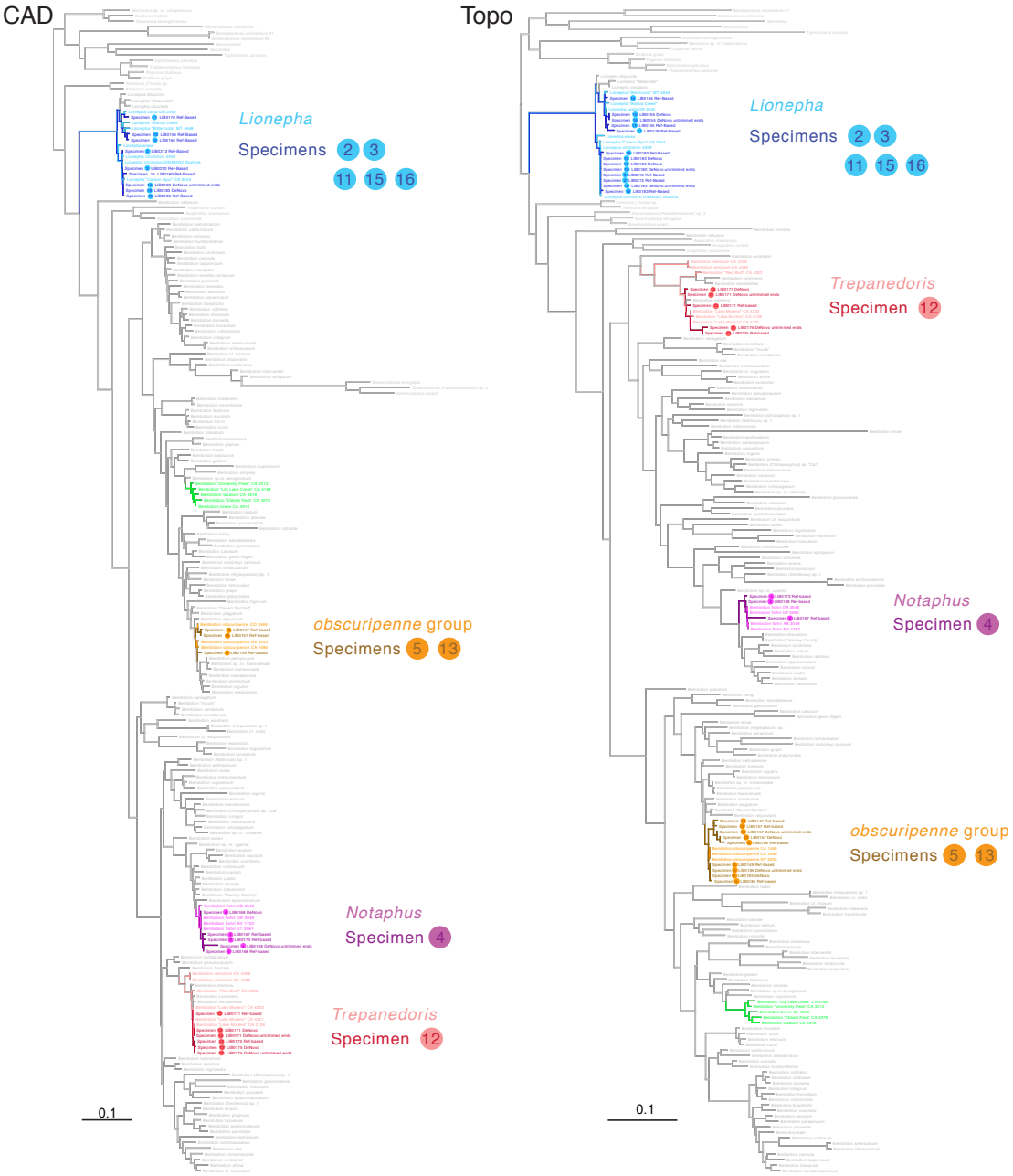


Fig. 2.11. Maximum Likelihood gene trees for CAD and Topo. Bembidiina shown as a whole, showing placement of context specimens (dark colors) and historical specimen sequences (light colors). Branch length is shown proportional to relative divergence, as estimated by GARLI.

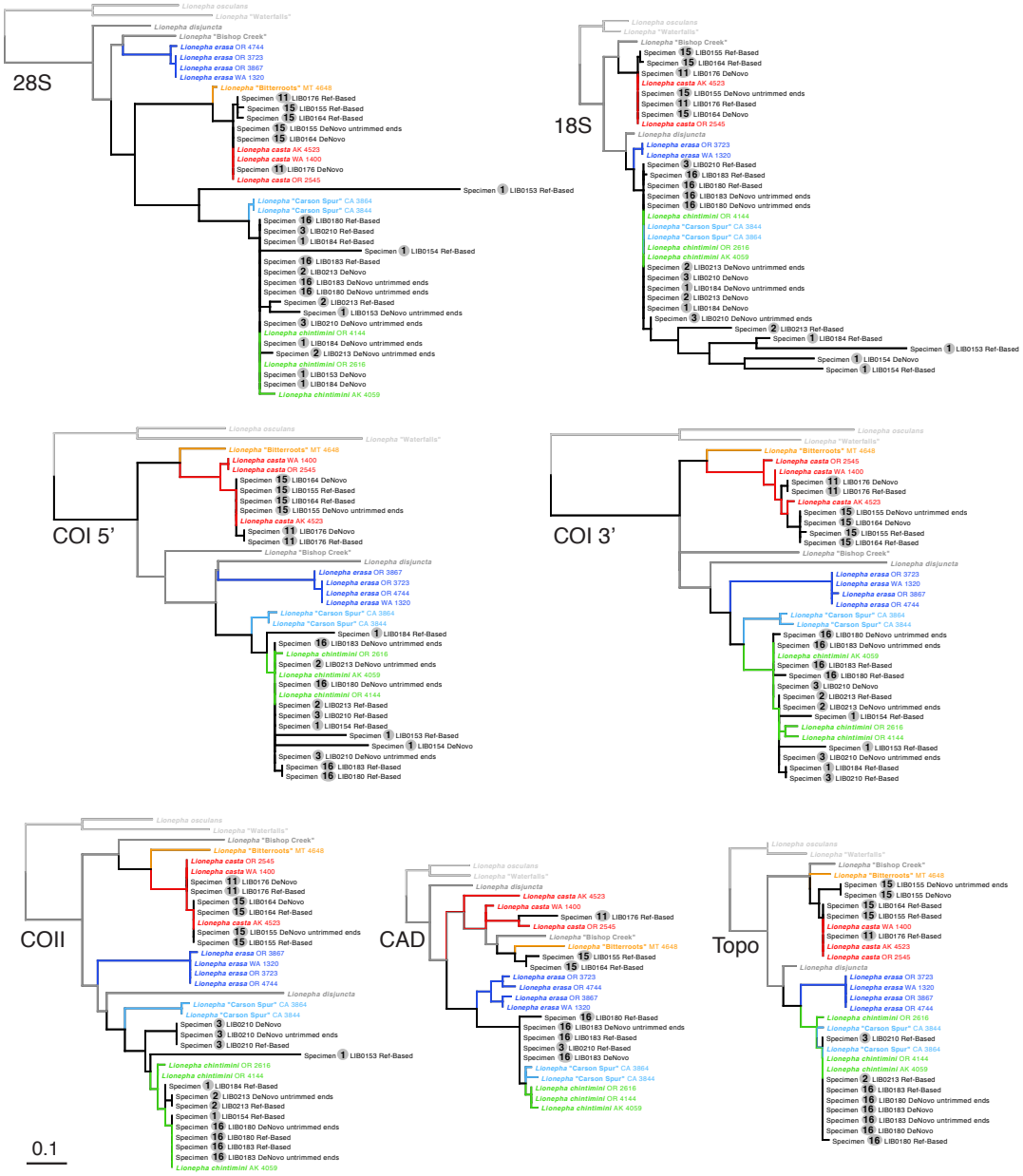


Fig. 2.12. Fine-scale phylogenetic placement of *Lionepha* specimens. Maximum likelihood gene trees with *Lionepha* historical specimen sequences shown in black text, candidate species in colored text, and others in grey. Branch length is shown proportional to relative divergence, as estimated by GARLI.

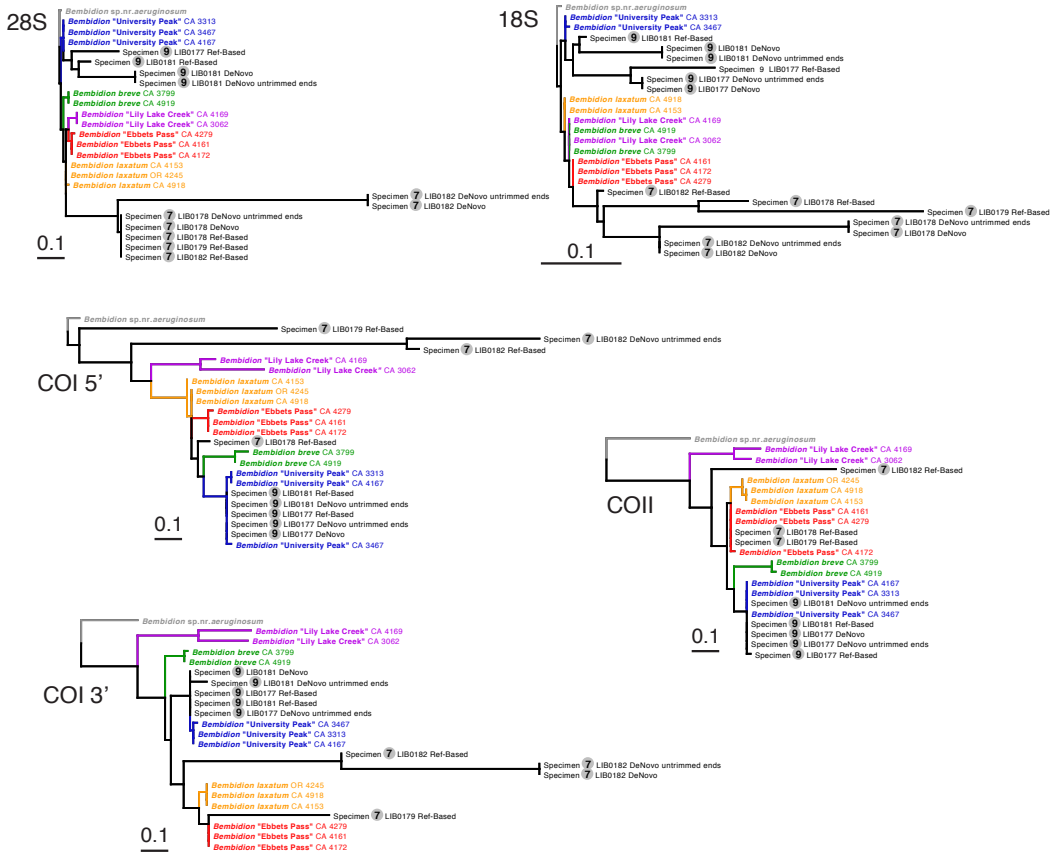


Fig. 2.13. Fine-scale phylogenetic placement of *breve* group specimens. Maximum likelihood gene trees with *breve* group historical specimen sequences shown in black text, candidate species in colored text, and others in grey. Branch length is shown proportional to relative divergence, as estimated by GARLI.

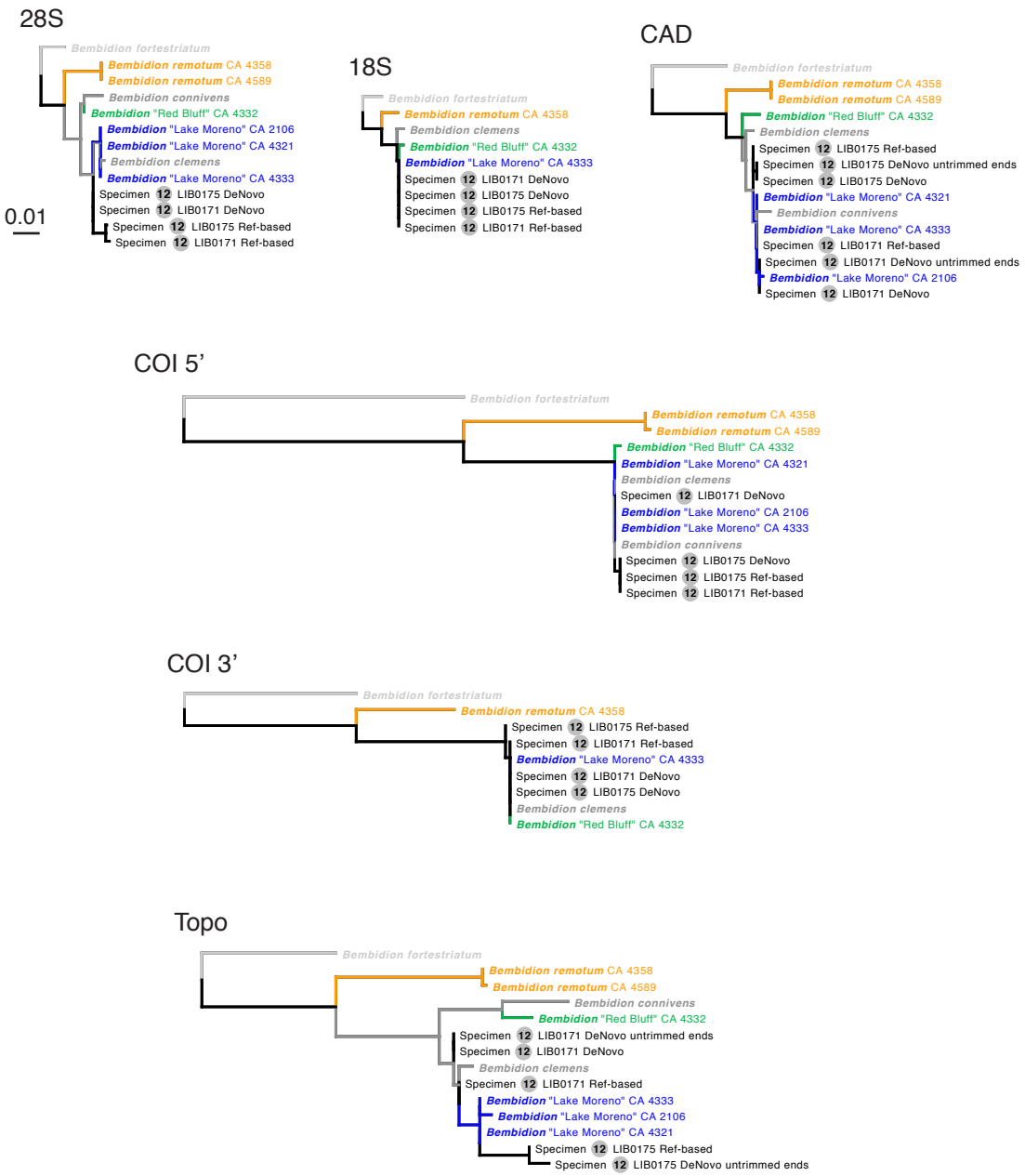


Fig. 2.14. Maximum Likelihood gene trees for *Trepanedoris* specimens. Branch length is shown proportional to relative divergence, as estimated by GARLI.

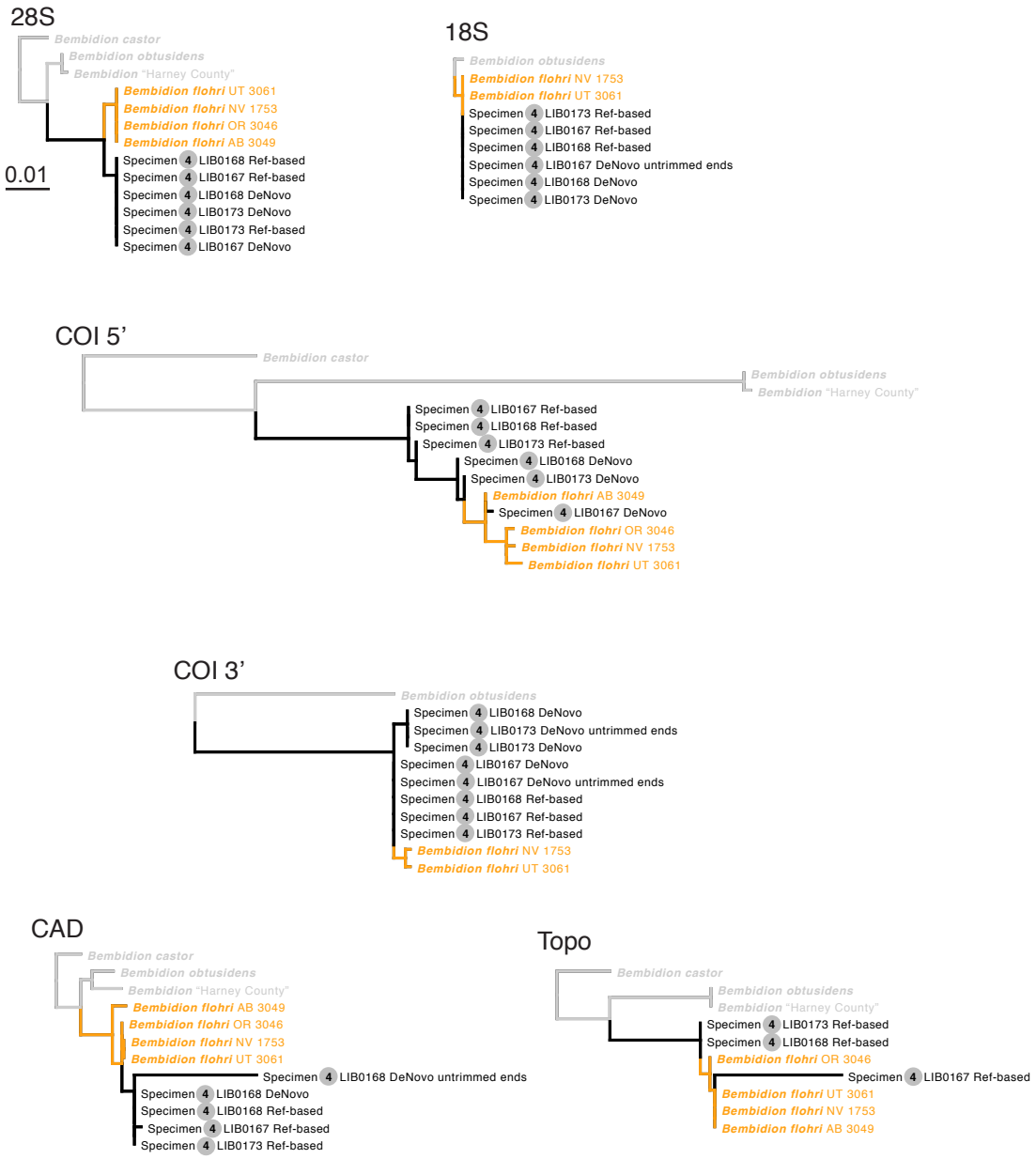


Fig. 2.15. Maximum Likelihood gene trees for *Notaphus* specimens. Branch length is shown proportional to relative divergence, as estimated by GARLI.

Fig. 2.16. Maximum Likelihood gene trees for *obscuripenne* group specimens. Branch length is shown proportional to relative divergence, as estimated by GARLI.

			match "Ebbets Pass" 4161	match *laxatum* 4918	unique base
LIB0178	rDNA complex	*de novo*	2	0	1
		ref-based	123	10	14
	mtGenome	*de novo*	2	3	13
		ref-based	19	10	116
LIB0179	rDNA complex	*de novo*	0	0	9
		ref-based	164	10	75
	mtGenome	*de novo*	3	4	1
		ref-based	24	7	240
LIB0182	rDNA complex	*de novo*	24	1	19
		ref-based	99	95	146
	mtGenome	*de novo*	20	15	1100
		ref-based	23	31	523
Total LIB0178 + LIB0179 + LIB0182			503	186	2257
Total LIB0178 + LIB0179			335	44	468

Fig. 2.17. MtGenome and rDNA complex analysis for specimen 7.
Assemblies from each of three libraries were compared to complete mtGenome and rDNA complex sequences from context specimens of both candidate species (*Bembidion* "Ebbets Pass", and *B. laxatum*). 'match "Ebbets Pass" 4161' indicates the number of sites at which the historical specimen matches the "Ebbets Pass" specimen among sites that distinguish that "Ebbets Pass" from *B. laxatum*. The reverse is true for "match *laxatum* 4918". "unique bases" are based for which the historical specimen matches neither candidate.

			match *chintimini* 4144	match *erasa* 3723	unique base
LIB0210	rDNA complex	*de novo*	552	2	2
		ref-based	110	9	25
	mtGenome	*de novo*	147	5	18
		ref-based	312	24	38
		Total	1121	40	83

Fig. 2.18. MtGenome and rDNA complex analysis for specimen 3. Assemblies from LIB0210 compared to complete mtGenome and rDNA complex sequences from context specimens of both candidate species (*Lionepha chintimini* and *L. erasa*). See caption of Fig. 2.17 for additional explanation.

TABLES

Table 2.1. Explanation of historical specimen ages. Minimum age used was extraction year minus last estimated collecting year. Some specimens differ in minimum age not because of a difference in collection year, but a difference in extraction year.

Specimen	Collection Date	Minimum Age	Explanation of Collection Date Estimate
1	1853–1857	159	The type series was provided by George Suckley (LeConte 1859), presumably during his travels through Oregon as naturalist for the governor of Washington Territory (Cooper & Suckley 1859) during 1853-1857.
2	1853–1857	159	See explanation for specimen 1.
3	1853–1857	157	See explanation for specimen 1. Extracted two years earlier.
4	≤1878	136	This species was described by Bates in 1878, and thus the specimen must have been collected in 1878 or before.
5	1871–1890	126	Specimens came from the Ulke collection, and were likely collected before 1871, but there is a possibility they could have been collected as late as 1890 (R.L. Davidson, pers. comm.)
6	≤1900	106	The specimen is from the Andreas Bolter Collection; this collection was bequeathed to INHS in 1900
7	≤1918	98	This species was described by Casey in 1918, and thus the specimen must have been collected in 1918 or before.
8	≤1918	98	See explanation for specimen 7.
9	≤1918	98	See explanation for specimen 7.
10	≤1918	98	See explanation for specimen 7.
11	≤1918	98	See explanation for specimen 7.
12	≤1918	97	See explanation for specimen 7. Extracted one year earlier.
13	1939	77	Collecting date on specimen label is 1939.

Table 2.1. (Continued)

14	<1940?	76?	Collected by Douglas K. Duncan, of Globe, AZ. Specimen from Pima Co., AZ. Based upon a broad literature search, Duncan appeared to be an active insect collector in the 1930s, but there is no evidence of specimens from him after 1940.
15	1952	64	Collecting date on specimen label is 23 September 1952.
16	1958	58	Collecting date on specimen label is 24-25 June 1958.

Cooper JG, Suckley G (1859) *The natural history of Washington Territory, with much relating to Minnesota, Nebraska, Kansas, Oregon, and California, between the thirty-sixth and fourth-ninth parallels of latitude, being those parts of the final reports on the survey of the Northern Pacific railroad route, containing the climate and physical geography, with full catalogues and descriptions of the plants and animals collected from 1853 to 1857.* Baillière Brothers, New York.

LeConte JL (1859) Catalogue of the Coleoptera of Fort Tejon, California. *Proceedings of the Academy of Natural Sciences of Philadelphia* **11**, 69-90.

Table 2.2. Historical specimens included in the study. Group: Genus (*Lionepha*), subgenus of *Bembidion* (other name beginning with an upper case letter), or species group (lower case). **#**: D.R.Maddison DNA voucher number. **Age**: Minimum number of years between the last possible collecting date and extraction date (see Table S1, Supporting Information). **Museum**: Museum in which the specimen has been housed. **Description**: For type specimens, reference for description of species name. **DNA**: Total DNA in the extraction.

Specimen	Group	Description	Description	#	Age	Museum	DNA (ng)
1	*Lionepha*	*Bembidium erasum* paralectotype 1	LeConte (1859)	4892	159	MCZ	39.84
2	*Lionepha*	*Bembidium erasum* paralectotype 2	LeConte (1859)	4915	159	MCZ	48.75
3	*Lionepha*	*Bembidium erasum* lectotype	LeConte (1859)	4241	157	MCZ	5.34
4	*Notaphus*	*Bembidium flohri* syntype	Bates (1878)	4185	136	BMNH	7.65
5	*obscuripenne*	*Bembidion ulkei* paratype	Lindroth (1963)	4869	126	CMNH	231.30
6	*Notaphus*	A *Bembidion lecontei* Csiki from S. California		4879	106	INHS	14.28
7	*breve*	*Bembidion lividulum* lectotype	Casey (1918)	4889	98	USNM	7.99
8	*Lionepha*	*Bembidion probatum* paralectotype	Casey (1918)	4938	98	USNM	149.4
9	*breve*	*Bembidion saturatum* lectototype	Casey (1918)	4890	98	USNM	7.22
10	*Trepanedoris*	*Bembidion scenicum* paralectotype	Casey (1918)	4932	98	USNM	29.34
11	*Lionepha*	*Bembidion brumale* lectotype	Casey (1918)	4893	98	USNM	12.00
12	*Trepanedoris*	*Bembidion disparile* lectotype	Casey (1918)	4764	97	USNM	13.44
13	*obscuripenne*	A *Bembidion* from Kenosha Pass, CO		4868	77	CAS	91.80
14	*Odontium*	A *Bembidion gilae* Lindroth from Pima Co., AZ		4888	76?	UAIC	50.07
15	*Lionepha*	A *Lionepha* from Ketchikan, AK		4894	64	OSAC	37.35
16	*Lionepha*	*Bembidion lindrothellus* holotype	Erwin & Kavanaugh (1981)	4891	58	MCZ	19.12

Table 2.3. Context specimens examined for fine-scale phylogenetic study. #: D.R. Maddison DNA voucher numbers.

Taxon	#	Locality
Lionepha		
Lionepha "Bishop Creek"	3568	USA: California: Inyo Co., South Fork Bishop Creek, 2835m, 37.1843°N 118.5585°W
Lionepha "Bitterroots"	4648	U.S.A., Montana, Ravalli County, Lost Horse Creek, 17.1 miles W of Highway 93 on Lost Horse Road 46.14142°N 114.48584°W
Lionepha "Carson Spur"	3844	USA: California: Amador Co., Oyster Lake, Silver Lake Cpgd, 2205m, 38.6711°N 120.1186°W
Lionepha "Carson Spur"	3864	USA: California: Tehama Co., Nanny Creek, Lassen NF, 1585m, 40.3696°N 121.5607°W
Lionepha "Waterfalls"	3782	USA: Oregon: Benton Co., Marys Peak Rd, Alder Creek Falls, 700m, 44.4746°N 123.5286°W
Lionepha casta Casey	1400	USA: Washington: Kittitas Co., Wenatchee National Forest, Taneum Creek Campground, 745m 47.1093°N 120.8567°W
Lionepha casta Casey	2545	USA: Oregon: Benton Co., Marys Peak, Alder Creek Falls, 700m, 44.4745°N 123.5282°W
Lionepha casta Casey	4523	USA: Alaska: Prince of Wales Island, Coffman Cove, logging road 3030-300, 55.98217°N 132.79002°W
Lionepha chintimini Erwin & Kavanaugh	2616	USA: Oregon: Benton Co., Marys Peak, 1110m, 44.5124°N 123.5554°W
Lionepha chintimini Erwin & Kavanaugh	4059	USA: ALASKA: Thompson Pass el. 796 m, 61.13731°N 145.74487°W
Lionepha chintimini Erwin & Kavanaugh	4144	USA: Oregon: Hood River Co., Cold Spring Creek below Tamanawas Falls, 968m, 45.403°N 121.5868°W
Lionepha disjuncta Lindroth	1896	Canada: British Columbia: Summit Creek 28 km E Kootenay Pass, 750m, 49.1421°N 116.7349°W
Lionepha erasa LeConte	1320	USA: Washington: Kittitas Co., Wenatchee National Forest, Taneum Creek Campground, 745m 47.1093°N 120.8567°W
Lionepha erasa LeConte	3723	USA: Oregon: Linn Co., Lost Prairie Campground on route 20, 1025m, 44.4037°N 122.0750°W
Lionepha erasa LeConte	3867	USA: Oregon: Klamath Co., Odell Creek near Davis Lake, Deschutes NF, 1340m, 43.5879°N 121.8564°W
Lionepha erasa LeConte	4744	USA: Oregon: Crook Co., just E of Lonesome Spring, Ochoco NF, 1596m 44.3958°N 120.0379°W
Lionepha osculans Casey	1401	USA: California: Placer Co., Rainbow, Hampshire Rocks Campground, 1860m 39.311°N 120.500°W

Table 2.3 (Continued)

Taxon	#	Locality
***Bembidion breve* group**		
Bembidion “Ebbetts Pass”	4161	USA: California: Alpine Co., pond below Ebbetts Pass, 2671m, 38.5445°N 119.8115°W
Bembidion “Ebbetts Pass”	4172	USA: California: Sierra Co., creek above Tamarack Lake, 2065m, 39.607°N 120.6568°W
Bembidion “Ebbetts Pass”	4279	USA: California: Tuolumne Co., snow field near Sonora Pass, 2908m, 38.3322°N 119.6500°W
Bembidion “Lily Lake Creek”	3062	USA: California: El Dorado Co., Lily Lake, 2000m, 38.8739°N 120.0808°W
Bembidion “Lily Lake Creek”	4169	USA: California: Sierra Co., creek above Tamarack Lake, 2065m, 39.607°N 120.6568°W
Bembidion "University Peak"	3313	USA: California: Inyo Co. 1.5 km NE University Peak, 3240 m. 36.76030°N, 118.35450°W
Bembidion "University Peak"	3467	USA: California: Lassen Co., Silver Lake, 1975m 40.494°N 121.162°W
Bembidion "University Peak"	4167	USA: California: Fresno Co., Kaiser Pass Meadow, 2783m, 37.2948°N 119.1006°W
Bembidion aeruginosum	3890	Russia, Altai Republic, Krasnaya Mountain, 50.0939°N 85.2279°E 1786m
Bembidion breve Motschulsky	3799	USA: California: Tehama Co., Nanny Creek, Lassen NF, 1584m, 40.3696°N 121.5612°W
Bembidion breve Motschulsky	4919	USA: California: El Dorado Co., Lily Lake, 2000m 38.8743°N 120.0801°W
Bembidion laxatum Casey	4918	USA: California: Alpine Co., Sonora Pass, 2900m 38.3323°N 119.6401°W
Bembidion laxatum Casey	4153	USA: California: Mono Co., snow field above Ellery Lake, 2901m, 37.9345°N 119.2318°W
Bembidion laxatum Casey	4245	USA: Oregon: Harney Co., Steens Mts., snowfield at Kiger Gorge, 2618m, 42.7152°N 118.5786°W
Bembidion (Trepanedoris)		
Bembidion " Lake Moreno "	4321	USA: California: San Luis Obispo Co., San Simeon St Pk, San Simeon Creek, 6m 35.5953°N 121.1228°W
Bembidion " Lake Moreno "	4333	USA: California: Ventura Co., Rancho Neuvo, Los Padres NF, 1063m 34.6952°N 119.3985°W
Bembidion "Lake Moreno"	2106	USA: California: San Diego Co., Lake Moreno, 32.6914°N 116.5244°W
Bembidion "Red Bluff"	4332	USA: California: Sutter Co., Feather River S of Nicolaus, 6m 38.8825°N 121.6131°W
Bembidion clemens Casey	2105	USA: Arizona: Yavapai Co., Prescott, Granite Creek near Watson Lake, 34.5819°N 112.4260°W
Bembidion connivens LeConte	2107	USA: California: Marin Co., Nicasio Reservoir
Bembidion fortestriatum Motschulsky	2098	Canada: British Columbia: km 10, Blowdown Road, 50.3713°N 122.2001°W

Table 2.3 (Continued)

Taxon	#	Locality
Bembidion remotum Casey	4358	USA: California: San Luis Obispo Co., San Simeon St. Park, 6m, 35.5955°N 121.1258°W
Bembidion remotum Casey	4589	USA: California: Riverside Co., Bautista Creek E of Hemet. 33.6515°N 116.8174°W, 805m
***Bembidion obscuripenne* group**		
Bembidion commotum Casey	2136	USA: California: Alpine Co., Sonora Pass, 2900m, 38.3323°N 119.6401°W
Bembidion nebraskense LeConte	2501	USA: Colorado: Fremont Co., Arkansas River, 11 km W Parkdale, 1815m, 38.457°N 105.4866°W
Bembidion obscuripenne Blaisdell	1480	USA: California: Tuolumne Co., Sonora Pass, 2925m 38.3301°N 119.6363°W
Bembidion obscuripenne Blaisdell	3253	USA: Nevada: Lander Co., rt722 at Campbell Creek 2050m, 39.2653°N 117.6864°W
Bembidion obscuripenne Blaisdell	3566	USA: Colorado: Ouray Co., Uncompahgre River, 5.6 km N Ouray, 2237m, 38.0679°N 107.6952°W
Bembidion sejunctum Casey	1817	USA: Washington: Pacific Co., Cape Disappointment
Bembidion sp. nr. *transversale*	3205	USA: Oregon: Benton Co., Corvallis, Willamette River, 60m, 44.5491°N 123.2451°W
Bembidion (Notaphus)		
Bembidion "Harney Lake"	3357	USA: Oregon: Lake County, 1.5km N of Musser Reservoir, 1419m, N43.5614 W 120.0052
Bembidion castor Lindroth	2043	Canada: Nova Scotia: Wentworth Provincial Park, Wallace River, 45.6271°N 63.5581°W
Bembidion flohri Bates	1753	USA: Nevada: Lyon Co., Carson River near Weeks, 390m, 39.2866°N 119.2778°W
Bembidion flohri Bates	3046	USA: Oregon: Harney Co., Harney Lake, NE corner, 1237m, 43.275°N 119.0902°W
Bembidion flohri Bates	3049	CANADA: Alberta: Birch Lake, 640m, 53.362°N 111.5231°W
Bembidion flohri Bates	3061	USA: Utah: Salt Lake Co., Great Salt Lake Marina, 1280m, 40.7482°N 112.1856°W
Bembidion obtusidens Fall	2042	Canada: Alberta: Burbank, junction of Red Deer and Blindman Rivers, 52.3542°N 113.7556°W

Table 2.4. Libraries generated and sequenced for historical specimens. Input: DNA quantity used as input for library preparation. **%Used**: Percent of total DNA available used as input for each library. **DSRep**: Libraries prepared with the *dsDNA* or *dsDNA Mod* protocols and enzyme-repaired DNA. **SSRep**: Libraries prepared with the *ssDNA* protocol and enzyme-repaired DNA. **DS**: Libraries prepared with the *dsDNA* or *dsDNA Mod* protocols and unrepaired DNA. **SS**: Libraries prepared with the *ssDNA* protocol and unrepaired DNA. In the last four columns, bold text indicates libraries that were sequenced, and non-bold text those not sequenced.

Specimen	Age	Input (ng)	%Used	DSRep	SSRep	DS	SS
1	159	5.3	13.4	**LIB0184***	**LIB0153**	LIB0165	**LIB0154**
2	159	6.1	12.5	**LIB0213***	-	LIB0214*	-
3	157	2.4	45.0	-	-	**LIB0210***	-
4	136	1.0	13.1	**LIB0173**	**LIB0168**	-	**LIB0167**
5	126	5.5	2.4	**LIB0185***	**LIB0149**	**LIB0186***	**LIB0150**
6	106	1.0	7.0	LIB0174	LIB0170	-	LIB0169
7	98	1.0	12.5	**LIB0182**	**LIB0178**	-	**LIB0179**
8	98	3.1	2.1	**LIB0217***		-	-
9	98	1.5	20.8	**LIB0181**	**LIB0177**	-	-
10	98	4.4	15.0	**LIB0216***		-	-
11	98	2.0	16.7	**LIB0176**	LIB0172†	-	-
12	97	2.5	18.6	**LIB0175**	**LIB0171**	-	-
13	77	10.0	10.9	**LIB0157**	**LIB0147**	LIB0158	LIB0148
14	76?	5.4	10.8	LIB0159	LIB0151	LIB0160	LIB0152
15	64	4.1	11.0	**LIB0164**	**LIB0155**	LIB0163	LIB0154
16	58	2.0	10.5	**LIB0183**	**LIB0180**	-	-

* library made with the dsDNA Mod protocol
† library not sequenced due to persistence of adapter dimer fragments in the library after three cleanups.

Table 2.5. Species used as a source of reference sequences for each historical specimen. Reference used: species from which query sequences were derived for reference-based assembly and as query sequences for probing *de novo* assemblies using BLAST. **#**: D.R.Maddison DNA voucher number of reference specimen. **Relationship**: relationship of the reference specimen to the historical specimen.

Specimen	Reference used	#	Relationship
1	*Lionepha* "Waterfalls"	3782	Same genus, but in different species group
2	*Lionepha* "Waterfalls"	3782	Same genus, but in different species group
3	*Lionepha* "Waterfalls"	3782	Same genus, but in different species group
4	*Bembidion castor*	4233	Same subgenus, but in different species group
5	*Bembidon* sp.nr. *transversale*	3205	Same subgenus, but in different species group
7	*Bembidion aeruginosum*	3890	Sister group to the remaining *breve* group members
9	*Bembidion aeruginosum*	3890	Sister group to the remaining *breve* group members
11	*Lionepha* "Waterfalls"	3782	Same genus, but in different species group
12	*Bembidion castor*	4233	Same subgeneric series, but different subgenus
13	*Bembidon* sp.nr. *transversale*	3205	Same subgenus, but in different species group
15	*Lionepha* "Waterfalls"	3782	Same genus, but in different species group
16	*Lionepha* "Waterfalls"	3782	Same genus, but in different species group

Table 2.6. Sequences examined for context specimens. The "#" column contains D.R. Maddison DNA voucher numbers. GenBank accession numbers from previous studies are listed in cells in the table; cells with check marks indicate sequences newly acquired for this study.

	#	28S	18S	COI 5'	COI 3'	COII	CAD	Topo
Lionepha								
Lionepha "Bishop Creek"	3568	✔	✔	✔	✔	✔	✔	✔
Lionepha "Bitterroots"	4648	✔		✔	✔	✔	✔	✔
Lionepha "Carson Spur"	3844	✔	✔	✔	✔	✔	✔	✔
Lionepha "Carson Spur"	3864	✔	✔	✔	✔	✔	✔	✔
Lionepha "Waterfalls"	3782	✔	✔	✔	✔	✔	✔	✔
Lionepha casta	1400	JN170467		JN171140	✔	✔	JN170947	JN171319
Lionepha casta	2545	✔	✔	✔	✔	✔	✔	✔
Lionepha casta	4523	✔	✔	✔	✔	✔	✔	✔
Lionepha chintimini	4059	KU233784	✔	KU233833	✔	✔	KU233943	KU234069
Lionepha chintimini	2616	✔	✔	✔	✔	✔	✔	✔
Lionepha chintimini	4144	✔	✔	✔	✔	✔	✔	✔
Lionepha disjuncta	1896	JN170468	JN170252	JN171141	✔	✔	JN170948	JN171320
Lionepha erasa	1320	JN170469	JN170253	JN171142	✔	✔	JN170949	JN171321
Lionepha erasa	3723	✔	✔	✔	✔	✔	✔	✔
Lionepha erasa	3867	✔		✔	✔	✔	✔	✔
Lionepha erasa	4744	✔		✔	✔	✔	✔	✔
Lionepha osculans	1401	JN170470	JN170254	JN171143	✔	✔	JN170950	JN171322
***Bembidion breve* group**								
Bembidion "Ebbets Pass"	4161	✔	✔	✔	✔	✔	✔	✔
Bembidion "Ebbets Pass"	4279	✔	✔	✔	✔	✔	✔	✔
Bembidion "Ebbets Pass"	4172	✔	✔	✔	✔	✔	✔	✔
Bembidion "Lily Lake Creek"	4169	✔	✔	✔	✔	✔	✔	✔
Bembidion "Lily Lake Creek"	3062	✔	✔	✔	✔	✔	✔	✔
Bembidion "University Peak"	3467	✔	✔	✔	✔	✔	✔	✔
Bembidion "University Peak"	3313	✔	✔	✔	✔	✔	✔	✔
Bembidion "University Peak"	4167	✔		✔	✔	✔	✔	✔

Table 2.6. (Continued)

	#	28S	18S	COI 5'	COI 3'	COII	CAD	Topo
Bembidion aeruginosum	3890	✔	✔	✔	✔	✔	✔	✔
Bembidion breve	3799	✔	✔	✔	✔	✔	✔	✔
Bembidion breve	4919	✔	✔	✔	✔	✔	✔	✔
Bembidion laxatum	4245	✔		✔	✔	✔	✔	✔
Bembidion laxatum	4918	✔	✔	✔	✔	✔	✔	✔
Bembidion laxatum	4153	✔	✔	✔	✔	✔	✔	✔
Bembidion (Trepanedoris)								
Bembidion "Lake Moreno"	4321	✔		✔			✔	✔
Bembidion "Lake Moreno"	4333	✔	✔	✔	✔		✔	✔
Bembidion "Lake Moreno"	2106	✔		✔			✔	✔
Bembidion "Red Bluff"	4332	✔	✔	✔	✔		✔	✔
Bembidion clemens	2105	JN170315	✔	JN171015	✔		JN170781	JN171197
Bembidion connivens	2107	JN170327		JN171025			JN170794	JN171207
Bembidion fortestriatum	2098	JN170341	JN170174	JN171036	✔		JN170808	JN171217
Bembidion remotum	4358	✔	✔	✔	✔		✔	✔
Bembidion remotum	4589	✔		✔			✔	✔
***Bembidion obscuripenne* group**								
Bembidion commotum	2136	GU454737	JN170164	GU454767			JN170783	JN171199
Bembidion nebraskense	2501	JN170389		✔			JN170861	✔
Bembidion obscuripenne	1480	✔	✔	✔	✔		✔	✔
Bembidion obscuripenne	3566	✔	✔	✔	✔		✔	✔
Bembidion obscuripenne	3253	✔	✔	✔	✔		✔	✔
Bembidion sejunctum	1817	GU454738	JN170231	GU454768			JN170913	JN171290
Bembidion sp. nr. *transversale*	3205	KU233791	KU233691	KU233841	✔		KU233979	KU234076
Bembidion (Notaphus)								
Bembidion "Harney County"	3357	✔		✔			✔	✔
Bembidion castor	2043	JN170307		JN171008			JN170773	JN171190
Bembidion flohri	3061	✔	✔	✔	✔		✔	✔

Table 2.6. (Continued)

	#	28S	18S	COI 5'	COI 3'	COII	CAD	Topo
Bembidion flohri	1753	JN170340	JN170173	JN171035	✔		JN170807	JN171216
Bembidion flohri	3046	✔		✔			✔	✔
Bembidion flohri	3049	✔		✔			✔	✔
Bembidion obtusidens	2042	✔	✔	✔	✔		✔	✔

Table 2.7. Prediction of placement of each historical specimen relative to context species. Predictions are based on available morphological characters and geographic distributions, which we have analyzed in ongoing projects on taxonomy and species delimitation within Bembidiina, as documented in Supplementary Methods.

Specimen	Prediction relative to context species
1	Belongs to *Lionepha erasa, L. chintimini, L*. "Bitterroots", or *L*. "Carson Spur"
2	Belongs to *Lionepha erasa, L. chintimini, L*. "Bitterroots", or *L*. "Carson Spur"
3	Belongs to *Lionepha erasa, L. chintimini, L*. "Bitterroots", or *L*. "Carson Spur"
4	Belongs to or is close to northern populations of *Bembidion flohri*
5	Close to, but distinct from, *Bembidion obscuripenne*
7	Belongs to *Bembidion* "Ebbets Pass" or *B. laxatum*
9	Belongs to *Bembidion breve* or *B.* "University Peak", or *B.* "Lily Lake Creek"
11	Belongs to *Lionepha casta*
12	*Belongs to Bembidion* "Lake Moreno", *B.* "Red Bluff", or *B. remotum*
13	Belongs to or is close to *Bembidion obscuripenne*
15	Belongs to *Lionepha casta*
16	Belongs to or is close to *Lionepha erasa* or *L. chintimini*

Table 2.8. Total input DNA lost during DNA repair. Extract Conc: Concentration of DNA extraction. **Input Vol**: Volume of extraction input to DNA repair. **Input**: Total DNA input for repair. **Rep Conc**: Concentration of repaired DNA following bead cleanup. **Rep DNA**: Total DNA following repair and cleanup. **Loss**: Percentage of input DNA lost due to bead cleanup of repaired DNA. **Avg. Corr. Loss**: For specimens represented by more than one repair event, we averaged the loss for each repair event and used that average in the calculation of the overall average.

Specimen	Extract Conc (ng/µl)	Input Vol (µl)	Input (ng)	Rep Conc (ng/µl)	Rep DNA (ng)	Loss (%)
1	0.498	40	19.9	0.356	11.4	42.8
1	0.498	22	11.0	0.117	5.9	46.6
2	0.650	28	18.2	0.127	6.1	66.5
4	0.085	62	5.3	0.087	2.8	47.2
5	2.570	7	18.0	0.393	6.5	64.0
5	2.570	15	38.6	0.385	12.3	68.0
5	2.570	15	38.6	0.385	12.3	68.0
5	2.570	8.6	22.0	0.112	5.6	74.6
6	0.168	30	5.0	0.104	3.1	38.1
6	0.168	34	5.7	0.128	4.1	28.3
7	0.094	60	5.6	0.154	4.9	12.6
8	1.660	6	10.0	0.062	3.1	68.9
9	0.085	58	4.9	0.151	4.8	2.0
10	0.326	25	8.2	0.088	4.4	46.0
11	0.150	64	9.6	0.191	6.1	36.3
12	0.192	52	10.0	0.216	6.9	30.8
13	1.020	25	25.5	0.902	14.9	41.6
13	1.020	15	15.3	0.128	7.1	53.6
14	0.589	40	23.6	0.36	11.5	51.1
16	0.415	40	16.6	0.274	8.8	47.2
16	0.239	63	15.1	0.24	7.7	49.0
					Avg. Corr. Loss	**41.3**
					SD	**18.0**

Table 2.9. Library preparation details. Protocol: See Table 2 caption. **Amp Cycles**: Number of PCR cycles used during library amplification. **Library Yield**: Library yield following a single post-amplification bead cleanup. **Extra Cleanups**: Number of additional bead cleanups required to remove unwanted small fragments from the library. **Post Clean**: Library mass remaining following all bead cleanups. **Frag Size**: Average library fragment size in bases. **Conc**: Final library concentration.

Library	Specimen	Protocol	Input DNA (ng)	DNA Repair	Amp Cycles	Library Yield (ng)	Extra Cleanups	Post Clean (ng)	Frag Size	Conc (nM)
LIB0147	**13**	SSRep	10.0	yes	18	297	-	-	292	21.59
LIB0148	**13**	SS	10.0	no	18	293	-	-	296	22.69
LIB0149	**5**	SSRep	5.5	yes	18	82	1	51	307	13.53
LIB0150	**5**	SS	5.5	no	18	30	1	16	296	6.24
LIB0151	**14**	SSRep	5.4	yes	18	303	-	-	357	15.98
LIB0152	**14**	SS	5.4	no	18	179	-	-	357	16.39
LIB0153	**1**	SSRep	5.3	yes	18	39	1	22	336	7.17
LIB0154	**1**	SS	5.3	no	18	21	1	11	318	3.55
LIB0155	**15**	SSRep	4.1	yes	18	269	1	132	350	19.09
LIB0156	**15**	SS	4.1	no	18	187	1	101	338	17.11
LIB0157	**13**	DSREP	10.0	yes	20	522	2	95	291	32.22
LIB0158	**13**	DS	10.0	no	20	357	2	66	281	33.06
LIB0159	**14**	DSREP	5.4	yes	20	481	1	270	265	31.59
LIB0160	**14**	DS	5.4	no	20	444	1	219	260	32.67
LIB0161	**5**	DS	5.5	no	20	223	3	69	-	-
LIB0162	**5**	DSREP	5.5	yes	20	339	3	130	-	-

Table 2.9. (Continued)

Library	Specimen	Protocol	Input DNA (ng)	DNA Repair	Amp Cycles	Library Yield (ng)	Extra Cleanups	Post Clean (ng)	Frag Size	Conc (nM)
LIB0163	**15**	DS	4.1	no	19	554	1	260	280	26.50
LIB0164	**15**	DSREP	4.1	yes	19	708	1	398	299	25.07
LIB0165	**1**	DS	5.3	no	20	212	1	47	-	-
LIB0166	**1**	DSREP	5.3	yes	20	252	3	83	-	-
LIB0167	**4**	SS	1.0	no	20	241	1	138	312	20.16
LIB0168	**4**	SSREP	1.0	yes	20	349	1	185	310	22.57
LIB0169	**6**	SS	1.0	no	20	53	1	33	361	9.08
LIB0170	**6**	SSREP	1.0	yes	20	66	1	46	368	12.83
LIB0171	**12**	SSREP	2.5	yes	19	240	-	-	336	19.25
LIB0172	**11**	SSREP	2.0	yes	19	85	3	42	-	-
LIB0173	**4**	DSREP	1.0	yes	18	479	-	-	255	23.20
LIB0174	**6**	DSREP	1.0	yes	20	360	1	170	300	17.73
LIB0175	**12**	DSREP	2.5	yes	18	476	1	171	312	10.84
LIB0176	**11**	DSREP	2.0	yes	20	589	-	171	268	24.51
LIB0177	**9**	SSREP	1.5	yes	19	99	1	58	335	13.98
LIB0178	**7**	SSREP	1.0	yes	19	102	1	55	331	15.06
LIB0179	**7**	SS	1.0	no	19	66	1	32	327	10.26
LIB0180	**8**	SSREP	2.0	yes	19	223	-	-	295	21.05
LIB0181	**9**	DSREP	1.5	yes	19	247	1	78	301	19.43
LIB0182	**7**	DSREP	1.0	yes	19	357	1	100	295	17.40
LIB0183	**8**	DSREP	2.0	yes	18	470	-	-	242	22.76

Table 2.9. (Continued)

Library	Specimen	Protocol	Input DNA (ng)	DNA Repair	Amp Cycles	Library Yield (ng)	Extra Cleanups	Post Clean (ng)	Library	Specimen
LIB0184	**1**	DSREP	5.3	yes	18	123	1	61	319	15.20
LIB0185	**5**	DSREP	5.5	yes	18	325	1	127	281	15.44
LIB0186	**5**	DSM	5.5	no	18	63	1	24	282	10.82
LIB0210	**3**	DSM	2.4	no	19	263	1	155	379	16.24
LIB0213	**2**	DSMREP	6.1	yes	18	115	1	111	355	11.08
LIB0214	**2**	DSM	6.1	no	18	67	1	37	357	8.94
LIB0216	**10**	DSMREP	4.4	yes	18	447	1	222	300	17.52
LIB0217	**8**	DSMREP	3.1	yes	18	368	1	163	282	19.73

Table 2.10. Comparison of library yield for repaired vs non-repaired DNA libraries. **Protocol**: Indicates whether the ssDNA protocol (SS) or dsDNA protocol (DS) was used. **No Rep Yield**: The concentration of the library constructed with unrepaired input DNA quantified through qPCR. **Rep Yield**: The concentration of the library constructed with enzyme-repaired input DNA quantified through qPCR. **Increase Repair**: Percent increase (or decrease) observed in repaired DNA libraries.

Specimen	Protocol	No Rep Yield (nM)	Rep Yield (nM)	Increase Repair (%)
1	DS	3.55	7.17	50.5
2	DS	8.94	20.8	57.0
4	SS	60.5	90.95	33.5
5	SS	6.24	15.12	58.7
5	DS	10.82	37.18	70.9
6	SS	9.08	12.83	29.2
7	SS	10.26	16.97	39.5
13	SS	159.72	155.69	-2.6
13	DS	42.26	57.09	26.0
14	SS	35.49	58.69	39.5
14	DS	118.83	139.08	14.6
15	SS	34.64	48.64	28.8
16	DS	106.5	144.44	26.3
			Average:	**36.3**
			SD:	**19.7**

Table 2.11. DNA lost during additional post-amplification bead cleanups. Loss for libraries that required a single additional cleanup. **Pre-cleanup**: DNA in library prior to additional bead cleanup. **Post-cleanup:** DNA in library after additional bead cleanup. **Loss**: Percentage of library lost due to bead cleanup.

Specimen	Protocol	Pre-cleanup (ng)	Post-cleanup (ng)	Loss (%)
1	SSRep	39	22	42.7
1	SS	21	11	45.1
1	SSRep	269	132	50.9
1	DS	212	47	77.8
1	SS	241	138	42.7
1	SSRep	349	185	47.0
1	DSRep	123	61	50.5
2	DSMRep	115	111	3.2
2	DSM	67	37	44.1
3	DSM	263	155	41.0
5	SSRep	82	51	38.3
5	SS	30	16	46.3
5	DSRep	325	127	60.9
5	DSM	63	24	61.8
6	SS	53	33	37.1
6	SSRep	66	46	30.6
6	DSRep	360	170	52.8
7	SSRep	102	55	46.0
7	SS	66	32	51.9
7	DSRep	357	100	72.0
8	DSMRep	368	163	55.7
9	SSRep	99	58	41.7
9	DSRep	247	78	68.6
10	DSMRep	447	222	50.4
12	DSRep	476	171	64.1
14	DSRep	481	270	44.0
14	DS	444	219	50.7
15	SS	187	101	45.9
15	DS	554	260	53.1
15	DSRep	708	398	43.7
			Average	48.7
			SD	13.6

Table 2.12. Sequencing, trimming, and *de novo* assembly statistics of sequenced libraries. Total Reads: Number of Illumina reads obtained. **Reads Removed**: Reads trimmed by quality and adapter trimming. % **Removed**: Percent of total reads removed during trimming. **n50**: n50 score for *de novo* assemblies.

Specimen	Library	Protocol	Total Reads	Reads Removed	% Removed	n50
1	LIB0153	SSRep	67344348	695956	1.03	243
1	LIB0154	SS	76331360	1526692	2.00	242
1	LIB0184	DSMRep	84059938	2592830	3.08	249
2	LIB0213	DSMRep	82987522	1548367	1.87	*
3	LIB0210	DSM	88050314	1906487	2.17	244
4	LIB0167	SS	87230858	2464984	2.83	332
4	LIB0168	SSRep	89253324	1965943	2.20	366
4	LIB0173	DSRep	66227728	22978500	34.70	307
5	LIB0149	SSRep	77126890	1500420	1.95	306
5	LIB0150	SS	81239380	2730639	3.36	346
5	LIB0185	DSMRep	81228656	7765800	9.56	289
5	LIB0186	DSM	57540524	6944203	12.07	341
6	LIB0174	DSRep	66806098	4201849	6.29	271
7	LIB0178	SSRep	85615410	911103	1.06	439
7	LIB0179	SS	105140366	1574670	1.50	442
7	LIB0182	DSRep	84597754	7140350	8.44	320
8	LIB0217	DSMRep	59115150	5989023	10.13	177
9	LIB0177	SSRep	86810992	905068	1.04	268
9	LIB0181	DSRep	77311484	3751598	4.85	270
10	LIB0216	DSMRep	66750624	3000560	4.50	179
11	LIB0176	DSRep	68072568	18724810	27.51	264
12	LIB0171	SSRep	75182682	1028710	1.37	446
12	LIB0175	DSRep	68586306	4603150	6.71	353
13	LIB0147	SSRep	76163428	1385267	1.82	412
13	LIB0157	DSRep	71476720	27521629	38.50	324

Table 2.12. (Continued)

Specimen	Library	Protocol	Total Reads	Reads Removed	% Removed	n50
15	LIB0155	SSRep	80888332	1115472	1.38	365
15	LIB0164	DSRep	80374544	19324716	24.04	969
16	LIB0180	SSRep	74710298	2077776	2.78	322
16	LIB0183	DSRep	65513092	23586300	36.00	299

*not recorded

Table 2.13. Target recovery from the mtGenome and rDNA complex from *de novo* and reference-based assemblies. Bases mtDNA DN: Number of mtGenome bases recovered from *de novo* assemblies. **Bases mtDNA Ref**: Number of mtGenome bases recovered from reference-based assemblies. **Unique Reads mtDNA**: Number of unique reads mapped to the mtGenome reference sequence. **Avg Depth mtDNA**: Average depth of sequencing coverage for recovered bases. **mtDNA Ref**: Percent of mtGenome bases recovered from reference-based assemblies relative to the number of bases in the reference sequence. **mtDNA DN**: Percent of mtGenome bases recovered from *de novo* assemblies relative to the number of bases in reference sequence. **Bases rDNA DN**: Number of rDNA complex bases recovered from *de novo* assemblies. **Bases rDNA Ref**: Number of rDNA complex bases recovered from reference-based assemblies. **Unique Reads rDNA**: number of unique reads mapped to the rDNA complex reference sequence. **Avg Depth rDNA**: Average depth of sequencing coverage for recovered bases. **rDNA Ref**: Percent of rDNA complex bases recovered from reference-based assemblies relative to the number of bases in reference sequence. **rDNA DN**: Percent of rDNA complex bases recovered from *de novo* assemblies relative to the number of bases in reference sequence.

Specimen	Library	Protocol	Bases mtDNA DN	Bases mtDNA Ref	Unique Reads mtDNA	Avg Depth mtDNA	mtDNA Ref (%)	mtDNA DN (%)	Bases rDNA Ref	Bases rDNA Ref	Unique Reads rDNA	Avg Depth rDNA	rDNA Ref (%)	rDNA DN (%)
1	LIB0153	SSRep	4107	11,760	1,394	5	78.9	27.6	322	6,159	909	5	50.9	2.7
1	LIB0154	SS	1607	9,848	1,587	5	66.1	10.8	165	5,864	1,077	6	48.4	1.4
1	LIB0184	DSMRep	4002	12,383	967	4	83.1	26.9	2,922	7,279	1,770	11	60.1	24.1
2	LIB0213	DSMRep	14489	14,889	6,874	39	99.9	97.2	3,228	7,216	2,191	14	59.6	26.7
3	LIB0210	DSM	6798	12,844	1,022	6	86.2	45.6	7,196	7,408	7,578	47	61.2	59.4
4	LIB0167	SS	13276	14,818	218,293	1,141	99.4	89.6	14,244	12,896	57,519	285	82.8	91.4
4	LIB0168	SSRep	12821	14,819	254,389	1,328	99.4	86.5	13,503	12,896	84,956	421	82.8	86.7
4	LIB0173	DSRep	10728	14,816	32,744	171	99.4	72.4	13,438	13,521	89,249	468	86.8	86.3
5	LIB0149	SSRep	13879	14,391	121,918	642	96.6	95.2	4,682	12,419	62,299	361	94.4	35.6
5	LIB0150	SS	13556	14,394	43,457	228	96.6	93.0	4,793	12,309	29,109	164	93.6	36.4
5	LIB0185	DSMRep	13443	14,388	97,895	557	96.6	92.2	6,855	12,543	89,351	558	95.3	52.1
5	LIB0186	DSM	9636	14,341	30,005	164	96.2	66.1	4,188	12,485	31,504	185	94.9	31.8

Table 2.13. (Continued)

Specimen	Library	Protocol	Bases mtDNA DN	Bases mtDNA Ref	Unique Reads mtDNA	Avg Depth mtDNA	mtDNA Ref (%)	mtDNA DN (%)	Bases rDNA Ref	Bases rDNA Ref	Unique Reads rDNA	Avg Depth rDNA	rDNA Ref (%)	rDNA DN (%)
7	LIB0179	SS	1517	10,071	1,349	6	67.6	10.2	1,295	12,099	10,354	52	85.1	9.1
7	LIB0178	SSRep	2478	10,166	1,709	8	68.2	16.7	1,935	12,855	11,167	59	90.4	13.6
7	LIB0182	DSRep	11948	14,737	3,438	17	98.9	80.5	3,300	13,458	98,419	496	94.7	23.2
9	LIB0177	SSRep	12652	14,686	2,750	14	98.6	85.3	2,174	12,060	1,248	7	84.8	15.3
9	LIB0181	DSRep	12764	14,695	4,981	26	98.6	86.0	9,218	13,594	8,610	42	95.6	64.8
11	LIB0176	DSRep	12540	14,688	23,918	74	98.6	84.2	6,435	8,300	9,608	62	68.5	53.1
12	LIB0171	SSRep	14593	14,231	157,246	813	95.5	98.5	10,671	7,582	94,746	476	48.7	68.5
12	LIB0175	DSRep	13587	14,219	53,096	299	95.4	91.7	11,233	7,692	114,640	634	49.4	72.1
13	LIB0147	SSRep	13966	14,396	265,751	1,427	96.6	95.8	11,777	12,376	208,124	1,228	94.1	89.5
13	LIB0157	DSRep	14097	14,384	24,417	132	96.5	96.7	11,204	12,479	216,205	1,339	94.9	85.2
15	LIB0155	SSRep	14827	14,890	43,032	222	99.9	99.5	8,027	7,473	74,857	482	61.7	66.3
15	LIB0164	DSRep	13195	14,804	6,883	32	99.3	88.6	7,775	7,653	60,232	414	63.2	64.2
16	LIB0180	SSRep	14830	14,895	2,907,045	15,263	100.0	99.5	7,558	7,725	47,554	307	63.8	62.4
16	LIB0183	DSRep	14831	14,887	302,505	1,746	99.9	99.5	7,521	7,751	68,343	474	64.0	62.1

Table 2.14. Phylogenetic placement of *Lionepha* historical sequences. Large, bold-faced numbers in cells indicate the number of gene fragments for which all libraries of a specimen are placed in a clade with the candidate group listed on the left. Small numbers indicate the number of gene fragments for which a minority of sequences for that specimen were in a different place in the tree. For example, for specimen **1**, of the seven gene fragments examined, in four (28S, COI 5', COI 3', and COII) all sequences were in a clade with *Lionepha chintimini* (Fig. 9), except for the sequence from one library for one gene fragment (28S), which was in a clade with both *L. chintimini* and *L.* "Carson Spur" (and thus the small "+1" in the bottom row). In addition, for the last gene (18S), all sequences were in a clade with both *L. chintimini* and *L.* "Carson Spur" (and thus the large, bold "1" in the bottom row). An explanation of predicted placement is given is Table S5 (Supporting Information).

	Specimens					
Candidate group	**1**	**2**	**3**	**11**	**15**	**16**
L. casta				**7**	**6**	
L. "Bitterroots"					**1**	
L. erasa						
L. chintimini	**4**	**4**	**4**			**4**
L. "Carson Spur"						
L. chintimini + *L.* "Carson Spur"	**1**+1	**2**	**3**			**3**

Table 2.15. Phylogenetic placement of *breve* group historical sequences. See Table 5 for explanation.

	Specimens	
Candidate group	**7**	**9**
B. laxatum		
B. "Ebbets Pass"	**2**+1	
B. laxatum + *B.* "Ebbets Pass"	**1**	
B. breve		
B. "Lily Lake Creek"		
B. "University Peak"		**5**
B. breve + *B.* "University Peak"	1	
B. laxatum + *B.* "Ebbets Pass" + *B. breve* + *B.* "University Peak"	1	
B. laxatum + *B.* "Ebbets Pass" + *B.* "Lily Lake Creek"	**1**	
Outside all candidates	**1**	

Table 2.16. Library preparation details of context specimens for which Illumina data was acquired. #: D.R. Maddison DNA voucher numbers. **Amp Cycles**: Number of PCR cycles used during library amplification. **Library Yield**: Total library yield. **Extra Cleanup**: Number of extra bead cleanups required to remove unwanted small fragments. **Post Cleanup**: Total library mass remaining following additional bead cleanup. **Frag Size**: Average library fragment. **Lib Conc**: Final library concentration.

Specimen	#	Input DNA (ng)	DNA Repair	Amp. Cycles	Library Yield (ng)	Extra Cleanups	Post Cleanup (ng)	Mean Frag Size	Lib Conc (nM)
Bembidion "University Peak"	3467	49.8	no	6	177	1	122	396	11.32
Bembidion “Lily Lake Creek”	4169	50.3	no	7	377	-	-	390	15.60
Bembidion aeruginosum	3890	50.0	no	6	99	-	-	347	15.72
Bembidion breve	4919	45.8	no	6	365	-	-	362	16.80
Bembidion laxatum	4918	50.8	no	6	290	1	178	406	8.77
Bembidion obscuripenne	3253	48.3	no	7	629	-	-	387	15.60
Lionepha "Bitterroots"	4646	46.4	no	6	240	-	-	409	14.41
Lionepha "Carson Spur"	3864	48.1	no	7	626	-	-	409	15.67
Lionepha casta	4523	49.4	no	7	223	-	-	346	15.27
Lionepha chintimini	4144	49.1	no	7	940	-	-	392	16.10
Lionepha chintimini	4059	51.0	no	6	124	1	83	425	9.37
Lionepha erasa	3723	40.7	no	7	444	-	-	380	16.51

Table 2.17. Models of evolution used in phylogenetic analysis. Models of evolution selected for each gene fragment.

	28S	18S	COI 5'	COI 3'	COII	CAD	Topo
Lionepha	TVMef+I+G	TVMef+I	TIM1+G	TIM1+I+G	TrN+G	HKY+G	TrN+G
B. breve group	TPM2	TPM3+I	TPM1uf+I	TIM2+I	TrN+G	HKY+I	TrN+I
B. (Trepanedoris)	K80+I	K80	TPM1uf+G	TIM1+I		HKY+I	TIM3+I
B. obscuripenne group	K80+I	TPM2	TIM1+I	TIM2		HKY+I	HKY+I
B. (Notaphus)	JC	JC	TPM2uf+I	HKY		HKY	HKY
Bembidiina	TVM+I+G	GTR+I+G	GTR+I+G			TVM+I+G	TrN+I+G

CHAPTER 3: CRYPTIC SPECIES IN THE MOUNTAINTOPS: SPECIES DELIMITATION AND TAXONOMY OF THE *BREVE* SPECIES GROUP (CARABIDAE: *BEMBIDION*) AIDED BY GENOMIC ARCHITECTURE OF A CENTURY-OLD TYPE SPECIMEN

John S. Sproul
David R. Maddison

Zoological Journal of the Linnean Society
https://academic.oup.com/zoolinnean
(volume/issue not assigned) zlx076

ABSTRACT

The *breve* species group includes closely related *Bembidion* Latreille ground beetles commonly found at high elevation in the mountains of western North America. For several decades, the group has been considered to consists of two species. Here we present evidence from morphological, molecular and geographic data that the group contains nine species: *Bembidion ampliatum*, *B. breve*, *B. geopearlis*, *B. laxatum*, *B. lividulum*, *B. oromaia*, *B. saturatum*, *B. testatum*, and *B. vulcanix*. We describe three species (*B. geopearlis, B. oromaia*, and *B. vulcanix*) as new, and resurrect four previously synonymized names (*B. ampliatum*, *B. lividulum*, *B. saturatum*, and *B. testatum*). Species diversity is highest throughout the Cascades in Oregon and Washington, and Sierra Nevada of California, where up to seven species can occur in sympatry. We resolved challenging nomenclatural issues through analysis of sequences obtained from century-old type specimens by using a novel application of rDNA copy number analysis — an approach that may prove useful for other historical specimens.

INTRODUCTION

The *breve* species group is a complex of closely related *Bembidion* Latreille ground beetles (Figs. 3.1–3.3) of the subgenus *Plataphus* Motschulsky, which are common in the mountains of western North America. Members of the species group are dark-bodied, medium-sized *Bembidion*, and live along margins of receding snow patches and shorelines of streams and lakes.

For several decades, the *breve* group has been considered to consist of two species, called *B. breve* (Motschulsky) and *B. laxatum* Casey. However, the last worker to treat the group formally noted high levels of intraspecific variation, distinctive geographic forms, and the need for a thorough revision (Lindroth, 1963). As we gathered and sequenced *breve* group material, we noted distinctive molecular forms from the Sierra Nevada in California that appeared to have subtle correlations with external morphological structures. We then examined male genitalia, and found morphological variation that corroborated molecular patterns, and strongly suggested the existence of additional species. Through these early findings of distinct forms in California, and because Lindroth's work targeted only the northern geographic range of the group (primarily Canada and Alaska), we began a concentrated sampling effort throughout the western United States, and set out to delimit species and revise the taxonomy of group.

We faced a challenge in assigning names to some of the species, as two critical, century-old type specimens (those of *Bembidion saturatum* Casey and *B. lividulum* Casey) are female, and lack male genitalic characters required for confident morphological identification. As part of another project, these types were sequenced

with next-generation sequencing and assigned to species concepts (Sproul & Maddison, 2017). Although the assignment of the *B. saturatum* lectotype was unambiguous across the trees of several genes, the sequences obtained from the *B. lividulum* lectotype were of poor quality (likely due to degradation), such that placement of that specimen was less certain. We provide an additional line of evidence that corroborates the previous assignment (Sproul & Maddison, 2017) of the *B. lividulum* type. We use evidence of genomic architecture in the form of ribosomal DNA (rDNA) copy number variation within the rDNA cistron (the region of rDNA containing 18S, 5.8S, and 28S rRNA genes and their spacer regions). Given that our approach relies on sequences of multi-copy genes, which show promising rates of recovery in projects attempting to sequence old specimens (Wandeler, Hoeck, & Keller, 2007; Guschanski *et al.*, 2013; Staats *et al.*, 2013; Burrell, Disotell, & Bergey, 2015; Kanda *et al.*, 2015; Sproul & Maddison, 2017), it has potential broader application, which we are presently investigating.

Here we present species delimitation and taxonomic revision of the *breve* group using morphological, molecular, and geographic data. We present multiple lines of evidence that indicate the group contains at least nine species, six of which were previously described, and three of which we describe as new. We then synthesize our results and observations in species descriptions, identification tools, and distribution maps.

METHODS

We examined over 2000 specimens of the *breve* group that are from or will be deposited in the collections listed below. Each collection's listing begins with the code used in the text.

BMNH The Natural History Museum, London

CAS California Academy of Sciences, San Francisco

CSCA California State Collection of Arthropods, Sacramento

EMEC Essig Museum Entomology Collection, University of California, Berkeley

MNHN Muséum National d'Histoire Naturelle, Paris

OSAC Oregon State Arthropod Collection, Oregon State University, Corvallis

UAM University of Alaska Museum, University of Alaska, Fairbanks

UASM E.H. Strickland Entomological Museum, University of Alberta, Edmonton

USNM National Museum of Natural History, Smithsonian Institution, Washington, DC

ZMH Zoological Museum, University of Helsinki, Helsinki

ZMMU Zoological Museum, Moscow State University, Moscow

Taxon sampling and DNA extraction for molecular analysis

We sampled for molecular study 141 specimens from 94 localities across the known range of the species group. We included five specimens of *B. (Plataphus) aeruginosum* (Gebler) sampled from five localities in Russia as outgroups, as this species is the sister group of a monophyletic *breve* group in a broader analysis of bembidiines (Maddison et al., unpublished). Sample localities are summarized in

Table 3.1. Specimens were collected into 95–100% ethanol and then stored at -20º C until DNA was extracted. We extracted DNA from most specimens (those preserved in ethanol) using the Qiagen® DNeasy™ (Qiagen) extraction kits and the manufacturer's recommended protocol. We extracted DNA from a few additional pinned specimens in a clean room with minor protocol modifications as described in Kanda *et al*. (2015).

PCR amplification and Sanger DNA sequencing

We amplified portions of five genes via Polymerase Chain-Reaction (PCR): **28S**: ~1100 bases of nuclear ribosomal subunit 28S; **COI**: 659 bases of mitochondrial protein-coding cytochrome *c* oxidase subunit I; **CAD**: 730 bases of nuclear protein-coding carbamoyl phosphate synthetase domain of the rudimentary gene; **Topo**: 740 bases nuclear protein-coding gene topoisomerase I; **MSP**: 930 bases of nuclear protein-coding gene muscle-specific protein 300. We conducted PCR on an Eppendorf Mastercycler ProS using TaKaRa Ex Taq with thermal profiles and PCR primers as explained in Maddison (2012) (for 28S, COI, CAD and Topo), and Maddison and Cooper (2014) (for MSP). PCR products were purified and sequenced at the University of Arizona's Genomic and Technology Core Facility using either a 3730 or 3730 XL Applied Biosystems automatic sequencer.

Assembly, alignment, and molecular analysis

Following sequencing, we assembled chromatograms using Phred (Green & Ewing, 2002) and Phrap (Green, 1999) via the Chromaseq package in Mesquite v3.2 (Maddison & Maddison, 2016, Maddison & Maddison 2017). Final sequence editing was conducted manually in Chromaseq. We aligned sequences from protein-coding

genes in Mesquite; no insertion or deletion events need be presumed in the history of the sequences examined. The ribosomal gene (28S) was aligned in MAFFT 7.130b (Katoh & Toh, 2008) with the G-INS-I algorithm as implemented in Mesquite v3.2 (Maddison & Maddison, 2017). Following alignment, data matrices for each gene were prepared for downstream analysis using Mesquite. We performed model selection for all genes using jModelTest v2.1.4 (Darriba *et al.*, 2012), and identified optimal models using the Bayesian Information Criterion. We inferred gene trees for each locus through Maximum Likelihood (ML) analysis across 100 search replicates in GARLI v2.0 (Zwickl, 2006).

We inferred the species tree using STACEY v1.2.2 (Jones, 2017), as implemented in BEAST v2.4.5 (Bouckaert *et al.*, 2014). STACEY uses a multi-species coalescent approach (Yang, 2002; Rannala & Yang, 2003; Degnan & Rosenberg, 2009; Edwards, 2009) similar to *BEAST (Heled & Drummond, 2010) to simultaneously infer gene trees and the species tree, except that it does not require *a priori* assignment of individuals to species, or guide trees. We set all individuals as separate species in order to view the clustering of individuals in the species trees and set the epsilon value to $1*10^{-4}$. We set the CollapseWeight parameters to 0.5 and 10 with a Beta prior and ran the first replicate run for 1 billion iterations, followed by two additional replicate runs for 500 million iterations, logging every 10,000th iteration. We evaluated sampling sufficiency using ESS values in Tracer v1.6, combined independent runs using LogCombiner after excluding 10% of trees as burn in, and summarized output trees using TreeAnnotator (Bouckaert *et al.*, 2014).

rDNA copy number variation analysis of the Bembidion lividulum Casey lectotype

We compared the signature of rDNA copy number variation observed in the lectotype of *Bembidion lividulum* Casey to those of newly sequenced specimens. There are three species to which the lectotype could belong based on morphological evidence and preliminary analysis of sequences presented in Sproul and Maddison (2017). We sequenced two ethanol-preserved specimens from each of the three candidate species (specimens DNA3593, DNA4149, DNA4165, DNA4245, DNA4918, DNA5032). The extraction and sequencing methods used for the *B. lividulum* lectotype are described in Sproul and Maddison (2017).

We prepared the newly sampled specimens for sequencing with NEBNext® DNA Ultra II Library Prep Kits (New England Biolabs) using the manufacturer's recommended protocol. The dual-indexed samples were then pooled and sequenced on an Illumina HiSeq 3000 maintained by the Oregon State University Center for Genome Research and Biocomputing, and for each sample we allocated approximately 1/30 of a 150 paired-end lane. Demultiplexing was performed using CASAVA version 1.8 (Illumina). Paired-end reads were imported into CLC Genomic Workbench version 8.5.1 (CLC Bio, referred to below as CLC GW), with failed reads removed during import. We trimmed and excluded adapter sequences from reads in CLC GW. We then mapped reads to a reference sequence of *Bembidion aeruginosum* using the 'Map Reads to Reference' tool, and visualized pileups of mapped reads in CLC GW. To reduce the chance of mapping spurious reads, the mismatch penalty was increased to four and the length and similarity fraction cutoffs were both increased to 0.85. The resulting rDNA signatures were then compared to that of the *B.*

lividulum lectotype, which we obtained through the same methods. We estimated the maximum number of rDNA copies for any point across the rDNA cistron by dividing the maximum read coverage depth of the rDNA cistron by the average coverage depth of 67 putatively single-copy nuclear protein coding genes (Regier *et al.*, 2008) that we mapped from the same set of reads, using the same parameters.

Morphological methods

Basic methods for studying adult structures, and terms used, are given in Maddison (1993). "Dorsal setae" of the elytra refer to setae ed3 and ed5, with "dorsal punctures" referring to the regions of the elytra at which those setae are attached.

We performed genitalic dissections on all male specimens whose DNA was extracted. The genitalia were cleared in KOH and mounted in Euparal on small cards, which were then pinned under the specimen. We studied external and genitalic structures of all specimens using a Leica M165C stereo microscope with an LED ring light, and a Leica DM5500B compound microscope. With the exception of six specimens whose genitalia were damaged or lost, we examined male genitalia of all male DNA voucher specimens, and recorded to which of several genitalic forms each belonged. This was done without active consideration of other data, thus allowing us to test how well genitalic forms corroborated molecular patterns. In addition to studying male genitalia from DNA vouchers, we examined genitalia from 63 additional specimens from key geographic localities.

We photographed habitus, pronotum and elytral microsculpture of DNA vouchers using a Leica Z6 Apo lens and DMC4500 camera. We photographed male genitalia with a Leica Z6 Apo lens and DMC4500 camera, or a Leica DM5500

compound microscope and DFC 425 C camera. We generated stacks of images at various focal depths using Leica Application Suite v4.8 (Leica Microsystems), and merged the stacks using Zerene Stacker (Zerene Systems) with either the PMax or DMap algorithms.

RESULTS

Molecular results

GenBank accession numbers for newly acquired DNA sequences are KY950685 – KY951331. GenBank Accession numbers for previously published sequences are given in Table 3.2. Models of evolution used for gene tree and STACEY analyses are listed in Table S2.

Most inferred species were monophyletic in the majority of maximum likelihood gene trees (Figs. 3.4–3.5). Six species were monophyletic in at least three of five gene trees, and three species were monophyletic in two or fewer gene trees (Table 3.4). Sequences from species pairs *B. ampliatum* + *B. laxatum* and *B. saturatum* + *B. vulcanix* showed the lowest differentiation among inferred species, and were interdigitated in multiple gene trees for both species pairs (Figs. 3.4–3.5).

Specimens of *B. saturatum* showed genetic structure in COI consistent with geographic patterns (Fig. 3.4C), with specimens from the Ruby Mountains in Nevada (specimens 5040, 3588, and 5039) showing a notably divergent haplotype (with a unique base at 10 sites) that renders the inferred species paraphyletic in COI. We also note genetic variants for *B. oromaia* (specimen 3886) and *B. testatum* (specimen 3166) in 28S (Fig. 3.5B). For both species, we sampled a single specimen from outside the Sierra Nevada (Trinity Alps in northwest California for *B. oromaia*, and Mount Ashland in southern Oregon for *B. testatum*) that differed by at least two bases in 28S (all other sequences were identical in 28S for both species), but lacked notable differences in the remaining genes.

The STACEY analysis revealed nine major clades with high Bayesian posterior probability (BPP) support (>94%) that corresponded to the nine distinct genitalic forms (Fig. 3.6). There are some clades found within these nine clades, but none of those corresponded to distinct morphological forms. Two of these clades had BPP of at least 85%: the two *B. saturatum* from the Steens Mountains, Oregon, grouped together with BPP of 85%, and two of the northern Sierra Nevada *B. breve* grouped with BPP of 90%. No other within-species clades in the STACEY analysis had BPP greater than 70%, and none of the within-species clades show evident morphological distinctiveness.

Our analysis of rDNA copy number variation within the rDNA cistron corroborated the previous assignment of the *Bembidion lividulum* Casey (1918) lectotype made by Sproul and Maddison (2017). The type specimen showed dramatic inflation of copy number in the second internal transcribed spacer (ITS2) and 28S gene regions (Fig. 3.7G). We observed this same signature of copy number inflation in both fresh specimens sequenced (from distinct geographic localities) of the species herein referred to as *B. lividulum* (Fig. 3.7E–F). We observed less region-specific copy number variation overall in specimens of the other two candidate species (*B. laxatum* and *B. ampliatum*), and different signatures in visualized read pileups (Fig. 3.7A–D).

Genitalic dissections of DNA vouchers revealed nine distinct genitalic forms (Figs. 3.8–3.10). Overlaying these forms on the STACEY tree showed that each form corresponds to one of the major clades recovered in the STACEY analysis (Fig. 3.6). We found shape and location of sclerite “St” (Lindroth, 1963) and length of the

flagellum to be diagnostic characters that corroborated molecular evidence (Fig. 3.10), with a single exception. Specimen DNA3321 showed a unique genitalic form that made it difficult to place as either *Bembidion ampliatum* or *B. laxatum*. The weight of molecular evidence, including the STACEY results, places this specimen with *B. ampliatum* (but note this specimen's placement within *B. laxatum* in the CAD gene tree of Fig. 3.4B). As this is our only specimen assigned to that species from the Cascades, additional sampling is needed to confirm the distinctiveness and meaning of morphological and molecular patterns.

In addition to genitalic characters, characters in the pronotum (Figs. 3.11–3.12), elytra (Figs. 3.1–3.3), and the extent of microsculpture (Fig. 3.13) corroborated patterns in gene trees and male genitalia, but are subject to more intraspecific variation than genitalic characters. The remaining results of the morphological investigation are presented in SPECIES DESCRIPTION AND IDENTIFICATION below.

DISCUSSION

Species delimitation

We consider as species separately evolving metapopulation lineages (de Queiroz, 2007). We delimit these lineages using evidence provided by patterns of gene trees, morphological data, and geographic data that suggest a lack of gene flow between entities so delimited, but presence of gene flow within such entities. Given patterns in individual gene trees (Figs. 3.4–3.5), results of the STACEY analysis (Fig. 3.6), and corroborating morphological characters, with many apparent species pairs so delimited found in sympatry (Table 3.5), we find multiple independent lines of evidence supporting the nine species recognized herein.

The weakest support for species we consider distinct is for the separation of *Bembidion ampliatum* from *B. laxatum*, and for *B. vulcanix* from *B. saturatum.* In both cases, the specimens within each pair are interdigitated in multiple gene trees (e.g., Figs. 3.4B, 3.4C, and 3.5B). Despite the lack of reciprocal monophyly for these species in individual gene trees, the multi-gene tree inference in STACEY recovered each species as monophyletic with high support (Fig. 3.6). Morphological evidence further corroborates these groupings; in particular, evidence from the male genitalia supports the STACEY analysis (Figs. 3.8–3.10). Finally, geographic data provide additional support for the separation of *B. ampliatum* and *B. laxatum,* as the two forms are microsympatric in the Sierra Nevada (Table 3.5). We sequenced both species from each of two localities (see localities for DNA vouchers 4160, 5123, 5125, 5127, and 5130 in *B. ampliatum*, and 4918, 4153, 5124, 5126, and 5128 in *B. laxatum* in Table 3.1). The fact that morphological and genetic distinctiveness are

maintained by individuals in microsympatry is additional evidence that validates their separation.

Finally, the morphological evidence, in particular male genitalic characters and extensive sympatry, provide evidence that differences inferred with genetic data are not simply due to population structure within the same species, which is a core concern of any study delimiting species using standard coalescent analyses (Knowles & Carstens, 2007; Carstens *et al.*, 2013; Sukumaran & Knowles, 2017).

Assigning primary type specimens to species concepts

We were able to confidently assign most primary type specimens to species concept using morphological analysis. We assigned primary types of *Bembidion adumbratum* Casey, *B. ampliatum* Casey, *B. improvisum* Casey, *B. laxatum* Casey, B. *lividulum* Casey, *B. testatum* Casey, *Notaphus incertus* Motschulsky, and *Peryphus brevis* Motschulsky using male genitalic characters and external structure. Female types of *Bembidion blanditum* Casey and *B. raineri* Hatch could be assigned confidently based upon external structure and geography. The primary type of *Peryphus tetraglyptus* Mannerheim was not examined; however, only *B. breve* is known from the latitudes at which it was collected, and it likely belongs as a synonym under *B. breve.*

Our efforts to assign morphologically ambiguous type specimens to species concept benefitted from the analysis of DNA sequences from the century-old primary type specimens of *Bembidion saturatum* Casey and *B. lividulum* Casey. Although Sproul and Maddison (2017) unambiguously placed the former specimen through gene-tree analysis, the latter showed evidence of sequence degradation that prevented

its fine-scale placement. Sproul and Maddison (2017) analyzed large fractions of the mitochondrial genome and rDNA and used sequence similarity at distinguishing sites as evidence to assign the *B. lividulum* lectotype to species.

In further exploring sequences obtained from the *B. lividulum* lectotype, we noted a striking pattern of read depth variation across the rDNA cistron. Although the 18S rRNA gene showed poor coverage (~0–10X coverage depth) across its length, the 28S rRNA gene, just a few thousand bases downstream, showed greater than 400X coverage, which is equivalent to a copy number of over 80,000 based on the average coverage of putatively single-copy nuclear protein-coding genes obtained from the same specimen (Fig. 3.7G).

In sequencing fresh specimens belonging to the candidate species we call herein *B. lividulum*, we observed the same dramatic signature of rDNA copy number seen in the *B. lividulum* Casey lectotype (Fig. 3.7E–G); in contrast, specimens of the other candidate species each showed a different, species-specific signal (Fig. 3.7A–D). This result confirms the previous assignment of Sproul and Maddison (2017) and strengthens our taxonomic conclusions. Although several cytogenetic studies document region-specific copy number variation the rDNA cistron between closely related species (Raskina *et al.*, 2008), we are unaware of any studies that have measured this signal via high-throughput sequencing as a line of evidence for assigning specimens to species concept. In the present application, this signature was a valuable source of data because it allowed us to detect a clear signal in the sequence data of a specimen that was too degraded to allow for placement through analysis of individual sequences. We have additional studies underway to explore the broader

potential of rDNA copy number variation as a tool for species delimitation and taxonomy.

SPECIES DESCRIPTION AND IDENTIFICATION

Characteristics of the breve group

The *breve* group belongs to *Bembidion* subgenus *Plataphus* (Lindroth, 1963; Maddison, 2012). Historically the group was placed within subgenus *Plataphodes* Gangblauer (e.g., Lindroth, 1963), a name now considered a junior synonym of *Plataphus* (Maddison, 2012). Distributed across high-elevation western North America (Figs. 3.14–3.17), most species have notably broad prothoraces basally compared to other *Plataphus*, with dark forebodies (dark brown to black, in some specimens with a metallic aeneous or blue hue) and elytra either the same color as the forebody, or dark brown to dark reddish brown. Legs and antennae are dark, similar in color to forebody and elytra. All species in the group show re-curvature of the lateral bead at the base of the elytron (Fig. 3.12B), a character used to define the former subgenus *Plataphodes*. They are distinguishable from other *Plataphus* by the broad width of the base of the pronotum relative to elytral width, combined with the re-curved elytral bead (Fig. 3.12B).

In sub-alpine habitats in western North America, particularly along creeks, many dark-bodied bembidiines may be present, including those of the genus *Lionepha* Casey, as well as other members of *Bembidion* subgenus *Plataphus*, and members of the *Ocydromus* complex of *Bembidion*. Two *Bembidion* commonly found alongside species of the *breve* group in alpine habitats are *B. (Plataphus) complanulum* (Mannerheim) and *B*. (*Ocydromus* complex) *commotum* Casey. *Bembidion complanulum* is smaller, flatter, and narrower than any *breve* group, and *B. commotum* lacks microsculpture on the elytra. Members of the *breve* group are easily

confused with *B*. *manningense* Lindroth and other similar *Plataphus*; *breve* group members can be distinguished by having a broader pronotum basally, and being more convex with less pronounced elytral striae.

Identification of species

Identifying specimens to species within the *breve* group is challenging, especially if only external structure is used, and several species pairs or trios can be difficult to separate. Although differences in external structure discussed herein generally hold within a species, notable intraspecific variation (e.g., in pronotal shape) is common. The external characters presented in the following key can aid in identification; however, examination of internal sac details of cleared male genitalia, or DNA sequences, may be needed to place some specimens with certainty.

1 Elytra notably convex, with lateral margins strongly rounded, resulting in inflated appearance (Fig. 3.2D); striae 3 and 4 partially disappeared or very weak (Fig. 3.2D). Pronotum sinuate laterally (Fig. 3.11G, see also Fig. 3.11I). Hindbody dark brown or reddish brown and generally slightly paler than the forebody. Male genitalia with darkened patch of scales apically (Fig. 3.9H, see inset) and lacking sclerite "St" (Figs. 3.9G–H and 3.10G). 4.6–5.1 mm ..... *B. testatum*

– Elytra flat or convex, or with lateral margins somewhat rounded, but not strongly rounded laterally and therefore lacking inflated appearance, striae weak, or not. Pronotum sinuate laterally, or not. Male genitalia lacking darkened patch of scales apically 2

2 (1) Pronotum with hind angles near 90° (Fig. 3.11D–F, I), or slightly obtuse (Fig. 3.11H) . 3

– Pronotum with hind angles strongly obtuse (Fig. 3.11A–B) 8

3 (2) Pronotum broad (Fig. 3.2A–C), with broad lateral explanation and broad base (Fig. 3.11D–F). Body convex, with a stout appearance. Smaller, most specimens less than 4.8 mm (although can be up to 5 mm) 4

– Pronotum broad or not (Fig. 3.11H, I). Body convex or somewhat flattened, but greater than 4.8 mm (females > 5.1 mm) 7

4 (3) Male genitalia with expanded sclerite "St" as in Fig. 3.10E. Montana, southeastern British Columbia, southwestern Alberta, northeastern Oregon (and likely northern Idaho) (Fig. 3.16B). 4–4.5 mm ..... *B. geopearlis*

– Male genitalia with sclerite "St" as in Fig. 3.10C–D, F. California, Nevada, western and southeastern Oregon, western Washington, western and southeastern British Columbia, north to Alaska 5

5 (4) Elytra with lateral margins somewhat rounded and narrowed at shoulder, dorsal punctures strongly foveate and striae often pronounced (Fig. 3.1C). Pronotum fairly large relative to elytral length, broad lateral explanation generally lacking (specimens from the Sierra Nevada may have a broad lateral explanation). Male genitalia with medium-length flagellum and diamond-shaped sclerite "St" (Figs. 3.8E–F and 3.10C). Smaller, 3.7–4.5 mm ..... *B. breve*

– Elytra fairly parallel-sided, broad at shoulder, often tapering towards apex, dorsal punctures weakly foveate. Male genitalia with elongate taper in apical third and long, sinuate flagellum (Fig. 3.9A–B, E–F) 6

6 (5) Male genitalia lacking large arcuate sclerite "St" (Figs. 3.9A–B and 3.10D). Sierra Nevada in California, Nevada, southeastern Oregon (Fig. 3.16A). 4–4.6 mm ..... *B. saturatum*

– Male genitalia with a large, arcuate sclerite “St” (Figs. 3.9E–F and 3.10F). Northwestern California, and the Cascades from Oregon to southern British Columbia (Fig. 3.16A). 4.1–5.0 mm, most ≤ 4.8 mm ..... *B. vulcanix*

7 (3) Pronotum narrow and strongly sinuate laterally (Figs. 3.3 and 3.11I). Legs and antennae elongate and slender. Elytra widest behind middle. Forebody and hind body generally unicolorous black (in rare specimens, elytral disc brown or reddish brown), in some specimens with a bluish hue. 5.2–6 mm ..... *B. oromaia*

– Pronotum broad (at least as broad as one elytron); in some specimens slightly sinuate laterally (Figs. 3.1D and 3.11H). Legs and antennae less elongate and slender. Elytra more or less parallel-sided, widest near middle and generally tapering towards apex. Elytral microsculpture often granulate in females (Fig. 3.13F). Forebody and hindbody unicolorous dark brown (occasionally black) with aeneous hue. Male genitalia not bent basally, flagellum short and weakly sinuate. 4.9–5.6 mm ..... *B. laxatum*

8 (2) Male genitalia as in Figs. 3.8C–D and 3.10B. Elytral dorsal punctures weakly foveate (especially in specimens from California and Oregon) (Fig. 3.1B). Pronotum relatively flat with a broad lateral explanation (Fig. 3.11B), weak laterobasal carina and shallow basal fovea (Fig. 3.12C). 4.0–5.3 mm, most ≥ 4.5 mm ..... *B. ampliatum*

– Elytral dorsal punctures strongly foveate (Fig. 3.1A, C). Pronotum with strong laterobasal carina, basal fovea somewhat deeper with less broad lateral explanation (Figs. 3.11A, G and 3.12A). 9

9 (8) Body parallel-sided and relatively flat (Fig. 3.1A). Male genitalia as in Fig. 3.8A–B and 3.10A. 3.7–4.9 mm, most ≥ 4.1 mm ..... *B. lividulum*

– Pronotum and elytra quite convex. Elytra with somewhat rounded lateral margin (Fig.

3.1C). Male genitalia as in Figs. 3.8E–F and 3.10C. Smaller, 3.7–4.5 mm.....*B. breve*

Species accounts

BEMBIDION LIVIDULUM CASEY

(FIGS. 3.1A, 3.8A–B, 3.10A, 3.11A, 3.12A, 3.13B, 3.14A)

Bembidion lividulum Casey, 1918: 25. Lectotype female, designated by Lindroth (1975: 117), in USNM, labeled 'Placer Co. CAL.' [white paper], 'CASEY bequest 1925' [white paper], 'TYPE USNM 36830' [red paper], 'lividula Csy.' [white paper, hand written], 'LECTOTYPE saturatum Csy. By C.H. Lindroth' [white paper, partly handwritten]. Type locality: Placer Co., California. Examined, including DNA sequences. Extracted DNA is deposited at the USNM and associated with the GUID of the type specimen: ark:/65665/3fa4e0e6d-4705-4d96-b32b-83af093df729.

Nomenclatural notes: This is the species referred to as '*Bembidion* "Ebbets Pass"' in Sproul & Maddison (2017), and *B. breve* (specimen 1930) in Maddison (2012).

Diagnosis: A shiny, medium-sized, parallel-sided, relatively flat species with strongly foveate dorsal punctures (Fig. 3.1A). Forebody and elytra black or dark brown often with aeneous hue. Pronotum widest anterior to middle with obtuse hind angles; not sinuate laterally; laterobasal carina strong (Figs. 3.11A and 3.12A). Elytra parallel-sided; dorsal punctures strongly foveate. Elytral disc commonly with uneven surface (caused by a depression in elytral surface) in basal third anterior to dorsal punctures

(see pattern of shadows in the basal third of elytral disc in Fig. 3.1A). Microsculpture weakened in males (Fig. 3.13B) causing shiny appearance and making foveate dorsal punctures easily visible without magnification. Male genitalia with flagellum sinuate and moderately long; sclerite "St" slender, commonly hourglass-shaped (Figs. 3.8A–B and 3.10A).

Comparison with similar species: Most easily confused with *B. ampliatum* and *B. breve*. Can be distinguished from the former by having a slightly more convex pronotum and stronger laterobasal carina, more strongly foveate dorsal punctures (particularly where they co-occur with *B. ampliatum* in California) and unevenness in the basal third of the elytral disc anterior to the dorsal punctures, and a longer, more sinuate flagellum and more slender sclerite "St". Distinguished from the latter by having more parallel-sided elytra, the pronotum widest anterior to middle, and a slender sclerite "St". May also be confused with *B. laxatum* from which it is distinguished by having a smaller body size, weaker microsculpture, a narrower pronotum relative to elytral width, and sclerite "St" lacking U-shaped recurvature.

Geographic distribution: From southern British Columbia along the Cascade Range to the southern Sierra Nevada. East through Idaho to the Rocky Mountains in Montana and southeastern British Columbia (Fig. 3.14A).

Habitat: Known from a wider range of environments than most other species in the group. It is often extremely abundant in the damp soil below receding snow patches

on alpine slopes. Also present along the shoreline of streams and lakes at high elevation. Common along the shorelines of moderate-sized rivers at somewhat lower elevation with increasing latitude (e.g., the Pacific Northwest and Alaska).

Geographic variation: This species is fairly variable across its range and within populations. Notable size variation is common in multiple locations of the Oregon Cascades and in Montana, with very small females in some populations (those sequenced do not show obvious differences in the genes examined).

BEMBIDION AMPLIATUM CASEY

(FIGS. 3.1B, 3.8C–D, 3.10B, 3.11B, 3.12D, 3.15A)

Bembidion ampliatum Casey, 1918: 24. Lectotype male, designated by Lindroth (1975: 117), in USNM, labeled ‘Col’ [white paper], [male symbol, hand drawn on white paper], ‘CASEY bequest 1925’ [white paper], ‘TYPE USNM 36828’ [red paper], ‘ampliatum Csy.’ [white paper, handwritten], ‘LECTOTYPE ampliatum Csy. By C.H. Lindroth’ [white paper, partly handwritten]. Type locality: Colorado. Examined, including genitalia.

Bembidion improvisum Casey, 1918: 24. Lectotype male, designated by Lindroth (1975: 117), in USNM, labeled ‘Col’ [white paper], [male symbol, hand drawn on white paper], ‘CASEY bequest 1925’ [white paper], ‘TYPE USNM 36832’ [red paper], ‘improvisum Csy.’ [white paper, hand written], ‘LECTOTYPE improvisum Csy. By C.H. Lindroth’ [white paper, partly handwritten]. Type locality: Colorado. Examined, including genitalia.

Diagnosis: This medium-sized species is parallel-sided with a distinctly flat pronotum. Forebody and hindbody black or very dark brown, often with aeneous or metallic hue. Pronotum relatively flat; hind angles obtuse; widest anterior to middle with broad lateral explanation and weak laterobasal carina (as the basal fovea beside it is not as deep, and thus the carina does not stand out so prominently); not sinuate laterally (Figs. 3.11B and 3.12C). Elytra are more or less parallel-sided and broadly rounded at apex; elytral disc with smooth appearance (particularly in California and Oregon), in part due to weak striae and weakly foveate dorsal punctures (Fig. 3.1B) which are not easily observed without magnification. Microsculpture meshes moderately etched in females, but weakly or at least unevenly etched in most males, often partially disappeared (Fig. 3.13D). Male genitalia with the ventral portion of sclerite "St" angled anteriorly such that the ventral extremity is even with, or anterior to, the dorsal extremity; flagellum short and somewhat sinuate (Figs. 3.8C–D and 3.10B).

Comparison with similar species: Most easily confused with *B. lividulum* from which it can be distinguished by having a flatter pronotum with weak laterobasal carina (especially reliable in California), less pronounced elytral striae with weakly foveate dorsal punctures, and a smooth elytral disc in the basal third anterior to the dorsal punctures (though often less so in specimens east of California and Oregon), and the male genitalia having a more laterally expanded sclerite "St" with the ventral portion angled anteriorly. May also be confused with *B. laxatum* from which it can be

separated through the weakened elytral microsculpture both sexes, by having a flatter pronotum with obtuse hind angles and weakened laterobasal carina, and lacking the U-shaped recurvature in sclerite “St” of the male genitalia.

Geographic distribution: Throughout the Sierra Nevada and White Mountains in California and north throughout the eastern mountains of Oregon (Steens and Wallowas) to Alberta. Easterly through the Great Basin and Rocky Mountains to Montana, Wyoming, and Colorado, south to New Mexico and Arizona. The only *breve* group species known from Utah, Wyoming, Colorado, New Mexico, and Arizona (Fig. 3.15A). A single questionable specimen is reported from Washington (see note below on geographic variation).

Habitat: Open alpine slopes, commonly below patches of melting snow.

Geographic variation: Specimens east of California and Oregon (e.g., Great Basin and Rocky Mountains) are generally smaller-bodied, especially males, and show more intraspecific variation in external structures. Notable within-population size variation has been observed in the Manti La Sal Mountains of eastern Utah, and the Sangre de Cristo Mountains of New Mexico. The single specimen from Washington reported herein (DNA3321) is a morphological and molecular outlier that we doubtfully include within this species.

BEMBIDION BREVE (MOTSCHULSKY)

(FIGS. 3.1C, 3.8E–F, 3.10C, 3.11C, 3.13A, 3.14B)

Peryphus brevis Motschulsky, 1845: 28. Lectotype male, designated by Bousquet and Larochelle (1993: 16), in ZMMU labeled 'Plataphus brevis Motsch [illegible]' [green paper, handwritten], 'B. breve Mtsch spec. interto proxima cfr. et. tetraglyptum dt. Netolitzky' [white paper, handwritten; 'interto' is likely a mispelling of 'incerto'], [red rectangle], 'LECTOTYPE Peryphus brevis Motschulsky Des. by Y. Bousquet' [red paper, partly handwritten]. Type locality: Sitka, Alaska. Examined, including genitalia.

Notaphus incertus Motschulsky, 1845: 350. Lectotype male, designated herein, in ZMMU, labeled 'Sitka' [green paper, handwritten], 'Plataphus incertus Motsc Am.b.[illegible] Sitka' [green paper, handwritten], 'B. incertum Mts spec.[illegible] det Netolitsky' [white paper, handwritten], 'LECTOTYPE Notaphus incertus Mtsch. designated Sproul & Maddison 2014' [red and white paper, partly handwritten]. Type locality: Sitka, Alaska. Examined, including genitalia.

Peryphus tetraglyptus Mannerheim, 1853: 151. Lectotype male, designated by Lindroth (1963: 273), in ZMH.

Bembidion blanditum Casey, 1918: 23. Lectotype female, designated by Lindroth (1975: 116), in USNM, labeled 'Metlakatla B. Col. Keen' [white paper], 'CASEY bequest 1925' [white paper], 'blandita Csy.' [white paper, handwritten], 'TYPE USNM 36829' [red paper], 'LECTOTYPE blanditum Csy. By C.H. Lindroth'

[white paper, partly handwritten]. Type locality: Metlakatla, British Columbia. Examined.

Nomenclatural notes: Lindroth's concept of *"Bembidion incertum",* as a widespread species distributed from the Pacific states to Colorado, included this species, *B. ampliatum*, *B. lividulum*, and *B. saturatum*.

Diagnosis: A small-bodied, convex species with strongly foveate dorsal punctures. Forebody dark brown, hindbody dark brown or reddish brown, often lighter than forebody; forebody and hindbody commonly with a metallic hue. Pronotum with fairly rounded lateral margin, widest at middle or just anterior to middle; laterobasal carina strong but often short (not proceeding far anteriorly) due to convexity of pronotum; basal fovea deep; hind angles slightly obtuse (Fig. 3.11C). Elytral striae generally pronounced and dorsal punctures strongly foveate; elytra fairly short relative to length of pronotum (Fig. 3.1C). Microsculpture with meshes moderately etched in both sexes (Fig. 3.13A) (but note geographic variation below). Male genitalia with medium-length flagellum; sclerite "St" more or less diamond-shaped (Figs. 3.8E–F and 3.10C).

Comparison with similar species: Most easily confused with *B. lividulum*, *B. saturatum,* and *B. vulcanix*. Distinguished from *B. lividulum* by having more convex pronotum and elytra with somewhat rounded lateral margin of elytra, pronotum widest closer to middle, and male genitalia with a longer, more sinuate flagellum and

broader sclerite “St”. Distinguished from *B. saturatum* and *B. vulcanix* by having a slightly smaller body size, more rounded lateral margin of elytra that narrows at the shoulder and more strongly foveate dorsal punctures, with pronotum larger relative to elytra and with a short laterobasal carina, and male genitalia with a shorter flagellum.

Geographic distribution: The northernmost species, ranging from the Aleutian Islands south along the coastal mountains of British Columbia, and in the Cascades of Oregon and Washington, south throughout California in the Sierra Nevada (Fig. 3.14B). Also known from one locality in Yoho National Park in eastern British Columbia.

Habitat: Most common on small, subalpine creeks in the southern part of its range. In the north, it occurs along creeks, rivers, or open slopes at high elevation.

Geographic variation: Some specimens are larger-bodied in the Sierra Nevada with broader pronota relative to northern localities. Notable intraspecific variation (in body size, pronotum shape, forebody and elytral coloration) is present in northern populations (e.g., southeast Alaska). Microsculpture shape and intensity is variable; in particular, some specimens may have notably transverse meshes (e.g., Queen Charlotte Islands, BC), or less deeply etched meshes such that portions of cells are partially disappeared in males (e.g., Snoqualmie Pass, WA). Some individuals from Yoho National Park are notably small in size.

BEMBIDION LAXATUM CASEY

(FIGS. 3.1D, 3.8G–H, 3.10H, 3.11H, 3.12B, 3.13E–F, 3.15B)

Bembidion laxatum Casey, 1918: 24. Lectotype male, designated by Lindroth (1975: 117), in USNM, labeled 'Placer Co. CAL.' [white paper], 'CASEY bequest 1925', 'TYPE USNM 36833' [red paper], 'laxata Csy.' [white paper, handwritten], 'LECTOTYPE laxatum Csy. By C.H. Lindroth' [white paper, partly handwritten]. Type locality: Placer Co., California. Examined, including genitalia.

Bembidion adumbratum Casey, 1918: 26. Lectotype male, designated by Lindroth (1975: 117), in USNM, labeled 'Placer Co. CAL.' [white paper], 'Oct.' [white paper], 'CASEY bequest 1925' [white paper], 'TYPE USNM 36827' [red paper], 'adumbrata Csy.' [white paper, handwritten], 'LECTOTYPE adumbratum Csy. By C.H. Lindroth' [white paper, partly handwritten]. Type locality: Placer Co., California. Examined, including genitalia.

Bembidion rainieri Hatch, 1950: 97. Holotype female in USNM, labeled 'Mt. Rainier, WASH. Sunrise Park Sept. 6, 1934 M. H. Hatch' [white paper], 'ADP 115747' [white paper], 'Bembidion (Plataphodes) laxatum Csy. M. Hatch-1969' [white paper, handwritten], 'TYPE Bembidion (Trechonepha) rainieri 1948.-M. H. Hatch' [red paper, hand written]. Examined. Type locality: Sunrise Park, Mount Rainier Pierce Co., Washington.

Diagnosis: Large and heavy-bodied, this parallel-sided species is characteristic in its dull appearance, particularly females. Forebody and hindbody dark brown, rarely black, typically with aeneous, or bluish, metallic hue. Pronotum broad and somewhat

convex; widest near middle with hind angles near 90° or slightly obtuse (Fig. 3.11H); laterobasal carina strong (Fig. 3.12B). Elytra widest near middle, often narrowing towards apex; elytral striae somewhat pronounced, and dorsal punctures moderately foveate (Fig. 3.1D). Microsculpture strongly etched in males; very strongly etched, often granulate in females (Fig. 3.13E–F). Male genitalia with short, weakly sinuate flagellum; sclerite "St" with U-shaped recurvature ventrally (Figs. 3.8G–H and 3.10H).

Comparison with similar species: Most easily confused with *B. oromaia* from which it is distinguished by having a wider pronotum relative to width of the elytra, the elytra being widest near middle, having larger protarsomeres in males, less slender, elongate legs and antennae, and by various characters in the male genitalia. May also be confused with *lividulum* and *B. ampliatum*. It can be distinguished from both by its larger, more convex body, duller appearance due to strong microsculpture in both sexes, broadened pronotum relative the elytra with hind angles near 90°, and the U-shaped recurvature of sclerite "St" in the male genitalia.

Geographic distribution: Throughout the Sierra Nevada in California, north along the Cascade Range to Washington and southern British Columbia (Fig. 3.15B).

Habitat: Open alpine slopes commonly below patches of melting snow or along small alpine creeks.

Geographic variation: Sclerite "St" in males from California localities is fairly elongate (Fig. 3.9G–H). Sclerite "St" in the few Washington and British Columbia specimens available for examination appear slightly less elongate and more broadened than California specimens.

Bembidion saturatum Casey

(Figs. 3.2A, 3.9A–B, 3.10D, 3.11D, 3.16A)

Bembidion saturatum Casey, 1918: 24. Lectotype female, designated by Lindroth (1975: 117), in USNM, labeled 'Placer co. Cal.' [white paper], 'CASEY bequest 1925' [white paper], 'TYPE USNM 36831' [red paper], 'saturata Csy.' [white paper, handwritten], 'LECTOTYPE saturatum Csy. By C.H. Lindroth' [white paper, partly handwritten]. Type locality: Placer Co., California. Examined, including DNA sequences.

Nomenclatural notes: This is the species referred to as '*Bembidion* "University Peak"' in Sproul & Maddison (2017). Extracted DNA is deposited at the USNM and associated with the GUID of the type specimen: ark:/65665/380c4cce2-3007-4a2d-9d40-5865c9760b4f.

Diagnosis: This small-bodied convex species has a broad pronotum and stout appearance (Fig. 3.2A). Forebody and hindbody dark brown, some specimens with an aeneous reflection. Pronotum very broad basally with hind angles near 90°; lateral explanation broad (Fig. 3.11D), laterobasal carina somewhat weak as the basal fovea beside it is quite shallow, and thus the carina does not stand out so prominently,

although carina may still extend far anteriorly. Elytra long relative to length of pronotum; tapered apically in some specimens; dorsal punctures weakly foveate (Fig. 3.2A). Microsculpture meshes strongly etched in females and moderately etched in males (except for populations in eastern Oregon and Nevada where microsculpture is notably weakened). Male genitalia with long sinuate flagellum (Figs. 3.9A–B and 3.10D); sclerite "St" either lacking (Fig. 3.9A), or with a hint of sclerotization in the position of sclerite "St" (Fig. 3.9B).

Comparison with similar species: Most easily confused with *B. vulcanix* and *B. geopearlis*, from which it is extremely difficult to separate using only external structures. It is most easily distinguished using characters in the male genitalia and geography. It is separated from both by lacking sclerite "St" (but see note on geographic variation below), and by having a southern and southeastern distribution (Fig. 3.16). May also be confused with *B. breve,* from which it is distinguished by a slightly larger body with more parallel-sided elytra, weakly foveate dorsal punctures, and longer, more sinuate flagellum of the male genitalia.

Geographic distribution: From Washington south to the southern Sierra Nevada, east to the Ruby Mountains in Nevada and Steens Mountains in Oregon (Fig. 3.16A).

Habitat: A variety of damp environments at or below the tree line including damp meadows, along lakeshores or small creeks, or on open alpine slopes below patches of melting snow.

Geographic variation: Populations from the Steens Mountains in eastern Oregon and the Ruby Mountains in Nevada (the "eastern form") are smaller-bodied, shinier and more black than dark brown, with slight sclerotization in the position of sclerite "St" evident (e.g., Fig. 3.9B), but not nearly so expanded as in *B. geopearlis* (Figs. 3.9C–D and 3.10E) or *B. vulcanix* (Figs. 3.9E–F and 3.10F). A slight patch of sclerotization has also been observed occasionally in individuals from California (Fig. 3.9B).

BEMBIDION GEOPEARLIS SPROUL & MADDISON *SP.NOV.*

(FIGS. 3.2B, 3.9C–D, 3.10E, 3.11E, 3.16B)

HOLOTYPE male (in OSAC) here designated, labeled: 'David R. Maddison DNA4727 DNA Voucher' [pale green paper], 'USA: Montana: Glacier N.P., east slope Clements Mtn., 2129m, 48.692°N 113.7292°W, 112 Aug 2015. JSS 2015.107-1 ['-1', handwritten]. J.S. Sproul & family' [white paper], 'HOLOTYPE *Bembidion geopearlis* Sproul + Maddison 2017' [partly handwritten, red paper], 'Oregon State Arthropod Collection OSAC_0002000000 [matrix code]' [printed on both sides of white paper]. Genitalia mounted in Euparal on small card labeled 'DNA4727' beneath the specimen; extracted DNA stored separately. GenBank accession numbers for DNA sequences of the holotype are: KY950786 (28S); KY950914 (CAD); KY951044 (COI), KY951174 (MSP), KY951301 (Topo).

Type locality: USA: Montana: Glacier National Park, east slope Clements Mountain near Logan Pass, 2129 meters, 48.69204°N 113.72920°W.

Paratypes: 38 paratypes from the following localities, specimens deposited in OSAC and USNM: USA: Montana: Glacier National Park, east slope Clements Mountain near Logan Pass, 2129 meters, 48.69204°N 113.72920°W (25); USA: Montana: Glacier Co., Glacier National Park, Iceberg Lake (2); USA: Montana: Glacier Co., Glacier National Park, Logan Pass (1); USA: Montana: Glacier National Park (1); USA: Montana: Mineral Co., Hoodoo Creek, 1780m (1); USA: Montana: Missoula Co., inlet to Heart Lake, 1891m, 47.3801°N 113.8501°W (1); USA: Montana: Ravalli Co., Lost Horse Creek, 1760m, 46.1417°N 114.4863°W (1); USA: Montana: Ravalli Co., Lost Horse Creek, 1660m, 46.1402°N 114.4371°W (1); USA: Montana: Flathead Co., Glacier National Park, Sperry Chalets (1); Canada: British Columbia: Akamina Pass, 1740m, 49.0261°N 114.0611°W (3); Canada: Alberta: Waterton Lakes National Park, Cameron Lake, 5440m (1).

Derivation of specific epithet: The name "geopearlis" is derived from an informal combination of letters taken from the names of JSS's children 'George' (*geo*) and 'Pearl' (*pearl*) with an ending (*is*) to make the name euphonic. The name recognizes the contribution of George (age 9) and Pearl (age 7) to the present work. They have accompanied JSS on over 6,500 miles of high-intensity collecting road trips, hiked over 30 miles to high elevation habitats in California, Oregon, and Montana, and helped collect hundreds of specimens including members of the type series of this species. JSS is indebted to them for their companionship and support. The

components of the name also reference the organism: ‘geo’ evokes something of the earth; thereby, ‘earth pearl’ or ‘a precious thing from the earth’ is descriptive of these seldom-collected ground beetles.

Diagnosis: A small, convex species with a notably broad prothorax (Fig. 3.2B). Forebody dark brown or black; hind body dark brown or reddish brown. Pronotum variable but generally very broad basally with a broad lateral explanation; hind angles near 90°; sinuate laterally in some specimens (Fig. 3.11E). Elytra widest behind middle; often but not always broadly rounded at apex; dorsal punctures weakly foveate. Microsculpture meshes strongly etched in females and moderately etched in males. Genitalia somewhat parallel-sided in outer shape having a short taper towards the apex; flagellum sinuate and moderately long; sclerite “St” expanded (Figs. 3.9C–D and 3.10E).

Comparison with similar species: Most similar in appearance to *B. saturatum, B. vulcanix* and *B. breve*. Most reliably distinguished from the all three with male genitalic characters and by its eastern geographic distribution. In particular, distinguished from *B. saturatum* and *B. vulcanix* by having elytra widest behind middle resulting in a more rounded apex of the elytra, and by the apical half of the aedeagus being less curved and elongate. Distinguished in male genitalia from *B. saturatum* by the presence of an expanded sclerite “St”. Distinguished from *B. vulcanix* by having a less expanded, non-arcuate sclerite “St”. Distinguished from *B.*

breve by having a broader pronotum, more weakly foveate elytral dorsal punctures, the shape and position of sclerite "St".

Geographic distribution: Known only from the Rocky Mountains of Montana, the Glacier/Waterton National Parks area of southern Alberta and British Columbia, and the Wallowa and Blue Mountains in northeastern Oregon (Fig. 3.16B).

Habitat: Collected in abundance on open slopes above the tree line at the type locality. Small series or singletons have also been collected on small creeks or depressions with damp soil below the timberline.

Geographic variation: none noted.

BEMBIDION VULCANIX SPROUL AND MADDISON *SP.NOV.*

(FIGS. 3.2C, 3.9E–F, 3.10F, 3.11F, 3.16A)

HOLOTYPE male (in OSAC) here designated, labeled: 'David R. Maddison DNA4615 DNA Voucher' [pale green paper], 'USA: Oregon: Deschutes Co., Stream east of Todd Lake, 1952m, 44.0282°N 121.6709°W, 24.v.2015. JSS.2015.030-1 ['-1', handwritten]. J.S., E.C., G.S., & P.E. Sproul' [white paper], 'HOLOTYPE *Bembidion vulcanix* Sproul + Maddison 2017' [partly handwritten, red paper], 'Oregon State Arthropod Collection OSAC_0002000001 [matrix code]' [printed on both sides of white paper]. Genitalia mounted in Euparal on small card labeled 'DNA4615'

beneath the specimen; extracted DNA stored separately. GenBank accession numbers for DNA sequences of the holotype are: KY950767 (28S), KY950895 (CAD), KY951025 (COI), KY951155 (MSP), KY951282 (Topo).

Type locality: USA: Oregon: Deschutes Co., Deschutes National Forest, stream east of Todd Lake, 1952m, 44.0282°N 121.6709°W.

Paratypes: 72 paratypes from the following localities, specimens deposited in BMNH, CAS, CSCA, EMEC, MNHN, and USNM: USA: Oregon: Deschutes Co., stream east of Todd Lake, 1952m, 44.0282°N 121.6709°W (57); USA: Oregon: Deschutes Co., Creek below Little Three Creek Lake, 2018m, 44.1057°N 121.6347°W (6); USA: Oregon: Deschutes Co., NE Todd Lake, Deschutes NF road 370, 2067m, 44.038°N 121.6718°W (1); USA: Oregon: Deschutes Co., E Todd Lake, Deschutes NF road 370, 1976m, 44.0306°N 121.6683°W (1); USA: Oregon: Hood River Co., Mt. Hood, Hood River Meadow Ski Area, 5300 ft. (2); USA: Washington: Whatcom Co., Bagley Lakes, Mt Baker, Snoqualmie NF, 1326m, 48.8528°N 121.6886°W (1); USA: Washington: Whatcom Co., Bagley Lakes, Mt Baker, Snoqualmie NF, 1290m, 48.8534°N 121.6948°W (1); USA: Washington: Pierce Co., Mt. Rainier, Tipsoo Lake (1); USA: Washington: Pierce Co., Mt. Rainier, Yakima Park (1); Canada: British Columbia: Garibaldi Provincial Park, S. slope Black Tusk (1).

Derivation of specific epithet: Informally derived by combining the two Latin words *Vulcanis*, the blacksmith god of fire and volcanoes from Roman mythology, and *nix*,

meaning snow. The name references snow-covered volcanoes of the Cascade Range where this species can be commonly found at high elevation.

Diagnosis: This is convex species is recognized by a broad pronotum basally, elongate elytra, and distinctive male genitalia. Forebody and hindbody dark brown in some specimens aeneous. Pronotum very broad basally; hind angles near 90°; lateral explanation broad; laterobasal carina fairly weak due to shallow adjacent basal fovea, but may extend far anteriorly nearly parallel to lateral margin of pronotum (Fig. 3.11F). Elytra long relative to length of pronotum and somewhat bullet-shaped in that they are parallel-sided, and tapering toward apex (although not readily obvious in Fig. 3.2C); dorsal punctures weakly foveate; striae often pronounced (Fig. 3.2C). Microsculpture strongly etched in females and moderately etched in males. Aedeagus is strongly curved, with an elongate taper towards the narrow apex; sclerite "St" in apical half, large and arcuate; flagellum long and sinuate (Figs. 3.9E–F and 3.10F).

Comparison with similar species: Most easily confused with *B. saturatum* and *B. geopearlis*. It is separated from both by having an expanded, arcuate sclerite "St" in the apical half (which is absent in the former, and not as expanded, arcuate, or apically positioned in the latter), and by its northwestern distribution (Fig. 3.16). May also be confused with *B. breve* from which it is distinguished by its generally larger body size, more parallel-sided elytra, weakly foveate dorsal punctures, and longer, more sinuate shape of the flagellum and large of sclerite "St" in the male genitalia.

Habitat: Along the shorelines of subalpine small creeks or lakes.

Geographic distribution: From southwestern British Columbia and the Olympic Peninsula in Washington, south along the Cascades of Washington and Oregon to the Trinity Alps in northwestern California (Fig. 3.16A).

Geographic variation: None noted.

Bembidion testatum Casey

(Figs. 3.2D, 3.9G–H, 3.10G, 3.11G, 3.17A)

Bembidion testatum Casey, 1918: 30. Lectotype male, designated by Erwin (1984: 174), in USNM, labeled 'Ca[a vertical line crossed by two shorter horizontal lines]' [white paper], (male symbol) [hand drawn on white paper], 'CASEY bequest 1925' [white paper], 'TYPE USNM 36842' [red paper], 'LECTOTYPE [male symbol] Bembidion testatum Csy. By Erwin '77' [white paper, partly handwritten]. Examined. Type locality: Lake Tahoe, California.

Nomenclatural notes: This is the species referred to as '*Bembidion* "Lily Lake Creek"' in Sproul & Maddison (2017).

Diagnosis: This fairly large-bodied species is most easily recognized by the inflated appearance of the elytra and narrow pronotum relative to width of the elytra (Fig. 3.2D). Forebody dark brown or black; hind body from dark brown to reddish brown, often paler than forebody. Pronotum narrow relative to the width of the elytra;

strongly sinuate laterally (Figs. 3.2D and 3.11G); hind angles near 90° with basal fovea deeply excavated (Fig. 3.11G). Elytra notably convex with strongly rounded lateral margin resulting in an inflated appearance; dorsal punctures weakly foveate, elytral striae weak with striae three and four partially disappeared in some specimens. Microsculpture strongly etched in females and moderately etched in males. Male genitalia with darkened patch of membranes apically; sclerite "St" lacking; flagellum sinuate and moderately long; ostial flag with an abbreviated sinuation not nearing the ventral surface of the aedeagus, and not extending far anteriorly (Figs. 3.9G–H and 3.10G).

Comparison with similar species: Can be confused with *B. saturatum* and *B. breve* in the northern Sierra Nevada. Distinguished from both by a larger body size, more convex elytra, and a darkened patch of scale-like structures apically in the male genitalia (see inset in Fig. 3.9H). Further distinguished from the former by having a narrower pronotum at the base. Distinguished from the latter by having very weakly foveate dorsal punctures and weak striae of the elytra.

Geographic distribution: Known from the Sierra Nevada and Trinity Alps in California, as well as a single locality in southern Oregon (Fig. 3.17A).

Habitat: Appears to be restricted to small, subalpine creeks.

Geographic variation: Some specimens from the Trinity Alps in northwestern California have a slightly longer flagellum in male genitalia. The single individual we

sampled from Mount Ashland, Oregon has four distinctive bases in 28S (three of which are ambiguities in DNA4173, but non-ambiguous and different in all other specimens), but is not notably separated in the other genes. The Oregon specimen also has less sinuate later margins of the pronotum than typical California specimens.

BEMBIDION OROMAIA SPROUL & MADDISON

SP.NOV.

(FIGS. 3.3, 3.9I, 3.10I, 3.11I, 3.13C, 3.17B)

HOLOTYPE male (in OSAC) herein designated, labeled 'David R. Maddison DNA4250 DNA Voucher' [pale green paper], 'USA: California: Tulare Co., snow field above Emerald Lake, 2851m, 36.5959°N 118.6756°W, 21.vi.2014. JSS 2014.064-12 ['-12', handwritten]. J.S. Sproul & Family' [white paper], 'HOLOTYPE *Bembidion oromaia* Sproul + Maddison 2017' [partly handwritten, red paper], 'Sequoia and Kings Canyon National Parks SEKI 23092' [green paper], 'Oregon State Arthropod Collection OSAC_0002000002 [matrix code]' [printed on both sides of white paper]. Genitalia mounted in Euparal on small card labeled 'DNA4250' beneath the specimen; extracted DNA stored separately. GenBank accession numbers for DNA sequences of the holotype are KY950760 (28S), KY950889 (CAD), KY951019 (COI), KY951149 (MSP), KY951276 (Topo).

Type locality: USA: California: Tulare Co., snowfield above Emerald Lake, 2851m, 36.5959°N 118.6756°W.

Paratypes: 32 specimens from the following localities deposited in CAS, OSAC, and USNM: USA: California: Tulare Co., Upper East Fk. Kaweah River, 2812m, 36.4189°N 118.5927°W (12); USA: California: Tulare Co., snow field above Emerald Lake, 2851m, 36.5959°N 118.6756°W (6); USA: California: Tulare Co., snowfield below White Chief Lake, 2912m, 36.417°N 118.5941°W (3); USA: California: Tulare Co., Lower Franklin Lake, 36.4203°N 118.5614°W (1); USA: California: Tuolumne Co., Blue Canyon Creek, 2750m, 38.3151°N 119.6613°W (2); USA: California: Tuolumne Co., stream draining N. face Leavitt Peak, 2930m, 38.3098°N 119.6619°W (2); USA: California: Tuolumne Co., Deadman Creek, 2700m, 38.3188°N 119.6634°W (1); USA: California: Tuolumne Co., Hwy. 108, stream SE of Chipmunk Flat, 2440m (1); USA: California: Tuolumne Co., Deadman Creek at junction with Blue Canyon Creek, 2665m, 38.3174°N 119.6652°W (1); USA: California: Mono Co., snow field above Ellery Lake, 37.9345°N 119.2318°W (2); USA: California: Mono Co., H. M. Hall Natural Area, Lee Vining Creek, 3020m, 37.9591°N 119.2838°W (1).

Derivation of specific epithet: Derived from Greek roots, *'oro'* meaning 'mountain', and *'maia'* meaning 'good mother', or 'caregiver'. Thus, the name connotes "good mother of the mountain" or "caregiver of the mountains". The name recognizes the contribution of Elizabeth C. Sproul to the present work. Mother of George and Pearl mentioned above, Elizabeth has spent hundreds of hours traveling to *breve* group localities, guiding young legs up steep trails, waiting at trailheads, recording locality data, and collecting and processing specimens in support of JSS's dissertation

research, and to enable positive associations for George and Pearl. Her steady support has added tremendous physical and emotional energy to the sampling efforts and overall scope of this work. The specific epithet also references the beetles. Distributed at higher elevation than any other members of the group, their elegant appearance inspires imagery of a maternal caregiver high in mountains.

Diagnosis: This large-bodied black species with long, slender appendages is the most distinctive member of the group (Fig. 3.3). Forebody and hindbody black, commonly with greenish or bluish metallic lustre. Pronotum narrow relative to elytral width and sinuate laterally; hind angles near 90° (Fig. 3.11I). Elytra widest behind middle with microsculpture meshes strongly etched in females and moderately etched in males (Fig. 3.13C). Legs and antennae black, long and slender; first protarsomeres in males quite small (not illustrated as Fig. 3.3 is a female). Male genitalia distinctive with outer shape having the appearance of being bent basally; flagellum long and sinuate; sclerite "St" elongate, bi-lobal and heavily sclerotized (Figs. 3.9I and 3.10I).

Comparison with similar species: Most easily confused with *B. laxatum* from which it is distinguished by having a pronotum which is narrower relative to the elytra, and more sinuate laterally, longer more slender legs and antennae, smaller first protarsomeres in males, opaque black coloration, and by various male genitalic characters.

Geographic distribution: Throughout the Sierra Nevada, north to Mount Lassen and the Trinity Alps in California (Fig. 3.17A).

Habitat: Open alpine slopes under fairly large rocky substrate on soil, commonly below patches of melting snow. May also be present along small alpine creeks.

Geographic variation: The single specimen we sampled from the Trinity Alps in California has two distinctive bases in 28S (a gene for which all other specimens have identical sequences), but is not notably distinctive in other genes.

REFERENCES

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

Burrell AS, Disotell TR, Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* 79: 35–44.

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. *Molecular Ecology* 22: 4369–4383.

Casey TL. 1918. A review of the North American Bembidiinae. *Memoirs on the Coleoptera. VIII. The New Era Printing Company, Lancaster, Pennsylvania*: 1–223.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

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

Erwin TL. 1984. Studies of the tribe Bembidiini (Coleoptera: Carabidae): lectotype designations and species group assignments for Bembidion species described by Thomas L. Casey and others. *The Pan-Pacific Entomologist* 60: 165–197.

Ganglbauer L. 1891. Die Käfer von Mitteleuropa. Die Käfer der österreichisch-ungarischen Monarchie, Deutschlands, der Schweiz, sowie des französischen und italienischen Alpengebietes. Ester Band. Familienreihe Caraboidea. *Carl Gerold's Sohn*: iii, 557.

Green P. 1999. Phrap. Version 0.990329. Available: http://phrap.org.

Green P, Ewing B. 2002. Phred. Version 0.020425c. Available: http://phrap.org.

Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gilissen E, Sonet G, Nagy ZT. 2013. Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology* 62: 539–554.

Hatch MH. 1950. Studies on the Coleoptera of the Pacific Northwest. II: Carabidae: Bembidiini. *The Pan-Pacific Entomologist* 26: 97–106.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27: 570–580.

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

Kanda K, Pflug JM, Sproul JS, Dasenko MA, Maddison DR. 2015. Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. *PLoS One* 10: e0143929.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9: 286–298.

Knowles LL, Carstens BC. 2007. Delimiting species without monophyletic gene trees. *Systematic Biology* 56: 887–895.

Lindroth CH. 1963. The ground beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska, Part 3. *Opuscula Entomologica Supplementum*: 201–408.

Lindroth CH. 1969. The ground beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska, Part 6. *Opuscula Entomologica Supplementum* 34: 945–1192.

Maddison D. 1985. Chromosomal diversity and evolution in the ground beetle genus Bembidion and related taxa (Coleoptera: Carabidae: Trechitae). *Genetica* 66: 93–114.

Maddison DR. 1993. Systematics of the holarctic beetle subgenus Bracteon and related Bembidion (Coleoptera: Carabidae). *Bulletin of the Museum of Comparative Zoology* 153: 143–299.

Maddison DR. 2012. Phylogeny of Bembidion and related ground beetles (Coleoptera: Carabidae: Trechinae: Bembidiini: Bembidiina). *Molecular Phylogenetics and Evolution* 63: 533–576.

Maddison DR, Cooper KW. 2014. Species delimitation in the ground beetle subgenus Liocosmius (Coleoptera: Carabidae: Bembidion), including standard and next-generation sequencing of museum specimens. *Zoological Journal of the Linnean Society* 172: 741–770.

Maddison DR, Maddison WP. 2014. Chromaseq: a Mesquite module for analyzing sequence chromatograms. Version 1.1. Available: http://mesquiteproject.org/packages/chromaseq.

Maddison WP, Maddison DR. 2017. Mesquite: a modular system for evolutionary analysis. Version 3.2. Available: http://mesquiteproject.org.

Mannerheim CG. 1853. Dritter Nachtrag zur Kaefer-Fauna der nord-amerikanischen Laender des russischen Reiches. *Bulletin de la Société Impériale des Naturalistes de Moscou* 26: 95–273.

Motschoulsky V. 1845. Remarques sur la collection de coléoptères Russes de Victor de Motschoulsky. *Bulletin de la Société Impériale des Naturalistes de Moscou* 18: 3–127.

Motschulsky V. 1864. Enumération des nouvelle espèces de coléoptères rapportés de ses voyages. 4-éme article. *Bulletin de la Société Impériale des Naturalistes de Moscou* 37: 171–240.

de Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886.

Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164: 1645–1656.

Raskina O, Barber J, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* 120: 351–357.

Regier JC, Shultz JW, Ganley ARD, Hussey A, Shi D, Ball B, Zwick A, Stajich JE, Cummings MP, Martin JW, Cunningham CW. 2008. Resolving Arthropod Phylogeny: Exploring Phylogenetic Signal within 41 kb of Protein-Coding Nuclear Gene Sequence. *Systematic Biology* 57: 920–938.

Sproul JS, Maddison DR. 2017. Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources* 17: 1183–1201.

Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One* 8: e69189.

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

Wandeler P, Hoeck PE, Keller LF. 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* 22: 634–642.

Yang Z. 2002. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics* 162: 1811–1823.

Zwickl, D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Unpublished thesis, The University of Texas.

DATA ACCESSIBILITY

Sequence assembly files are deposited in Dryad. doi:10.5061/dryad.10qj1. DNA sequences included in phylogenetic analyses are deposited in Genbank [KY950685–KY951331]. Sequence read files are deposited in NCBI Sequence Read Archive [SRR5514451–SSR5514456]. Matrices and results from phylogenetic analysis are deposited in Dryad. doi:10.5061/dryad.10qj1

FIGURES

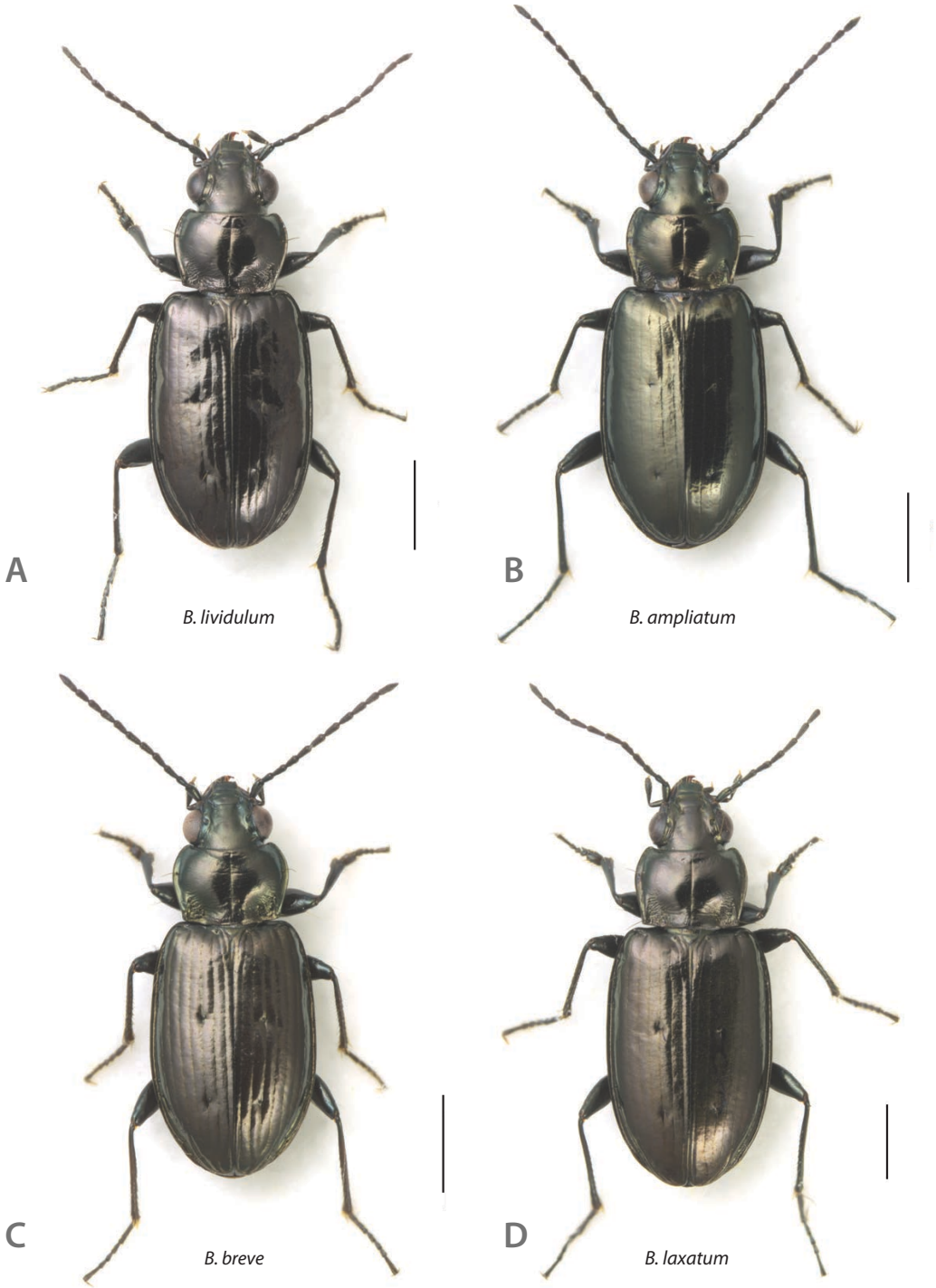


Figure 3.1. Habitus of *breve* group adults. (A) *Bembidion lividulum*, USA: California: Tuolumne Co., Deadman Creek; DRM voucher V100961; (B) *B. ampliatum*, USA: Oregon: Harney Co., Steens Mts., snowfield at Kiger Gorge, DRM voucher V100963; (C) *B. breve*, USA: Washington: King Co, Snoqualmie River at Alpental, DRM voucher V100955; (D) *B. laxatum*, USA: California: Mono Co., Mammoth Mountain Ski Area, DRM voucher V101128. Scale bar is 1 mm.

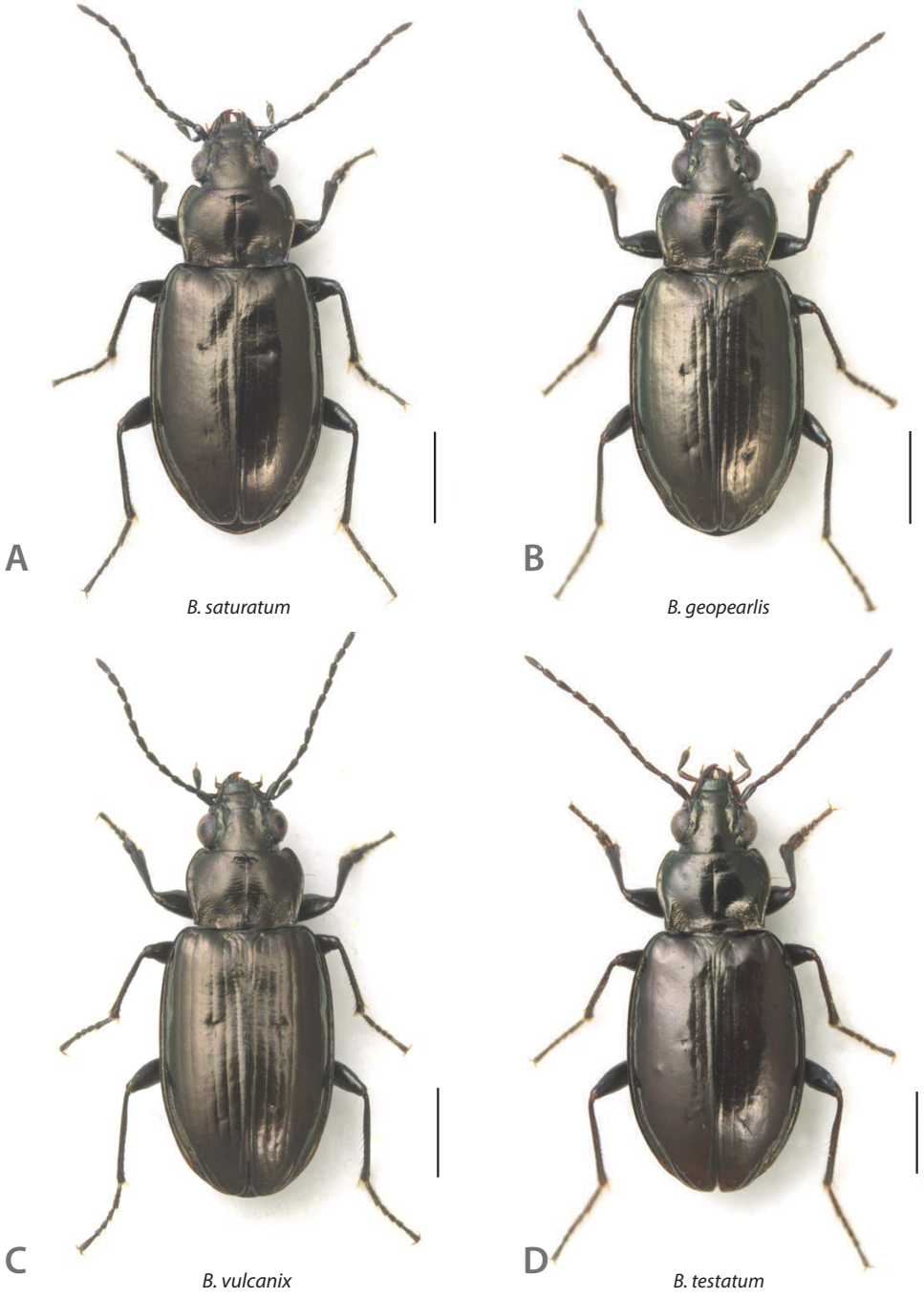


Figure 3.2. Habitus of *breve* group adults. (A) *B. saturatum*, USA: California: Fresno Co., Kaiser Pass Meadow, DRM voucher V100942; (B) *Bembidion geopearlis*, USA: Montana: Glacier Co., Glacier N.P., east slope Clements Mtn., DRM voucher V100957; (C) *B. vulcanix,* USA: Oregon: Deschutes Co., Stream east of Todd Lake, DRM voucher V101129; (D) *B. testatum*, USA: California: El Dorado Co., Lily Lake, DRM voucher DNA3062. Scale bar is 1 mm.

Figure 3.3. Habitus of *Bembidion oromaia*. USA: California: Tulare Co., snowfield below White Chief Lake, DRM voucher V100959. Scale bar is 1 mm.

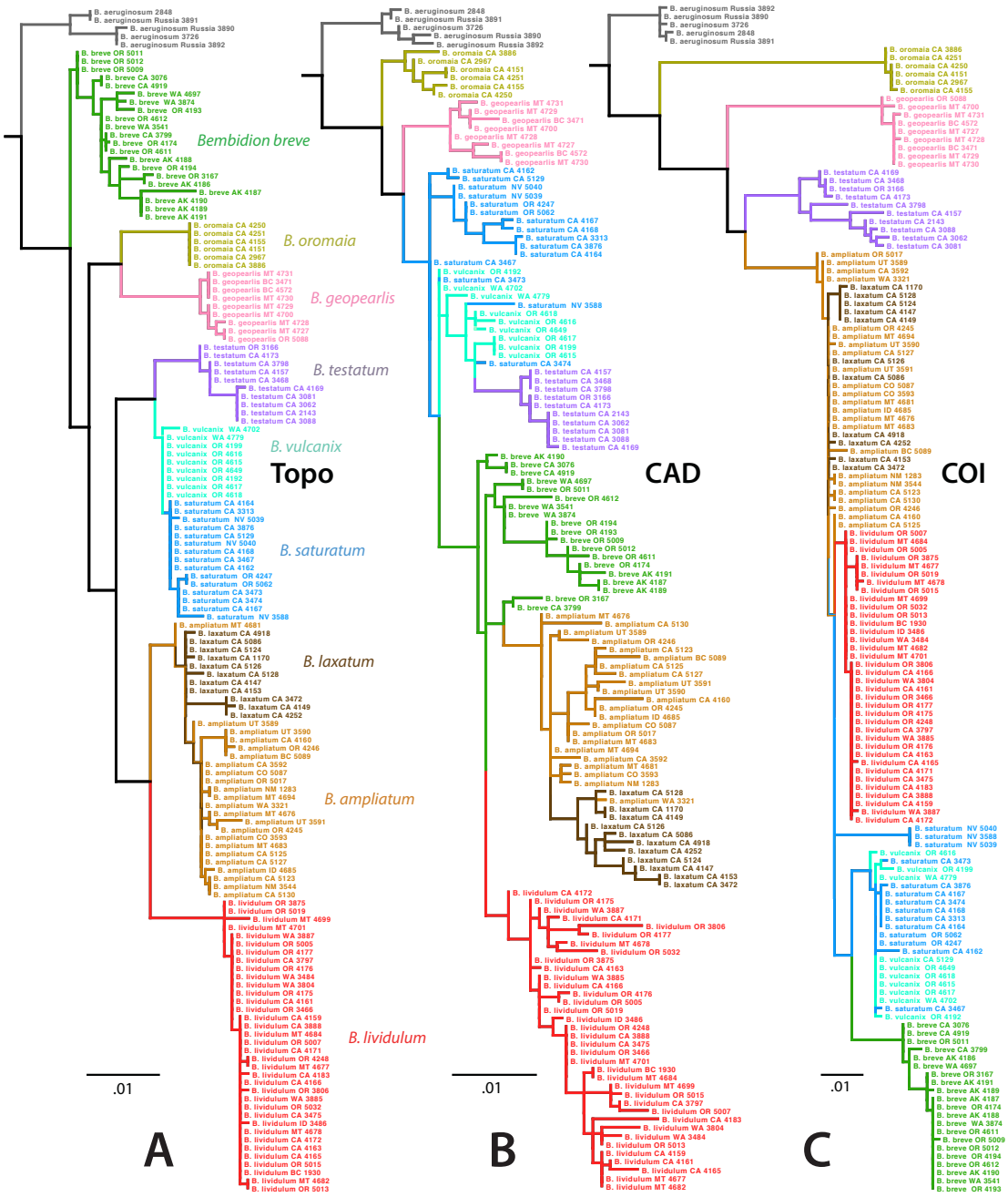


Figure 3.4. Maximum likelihood trees for three genes, (A) Topo, (B) CAD, and (C) COI. Branch length shown is proportional to relative divergence with scale bars indicating 0.01 units. Outgroup taxa are shown in gray.

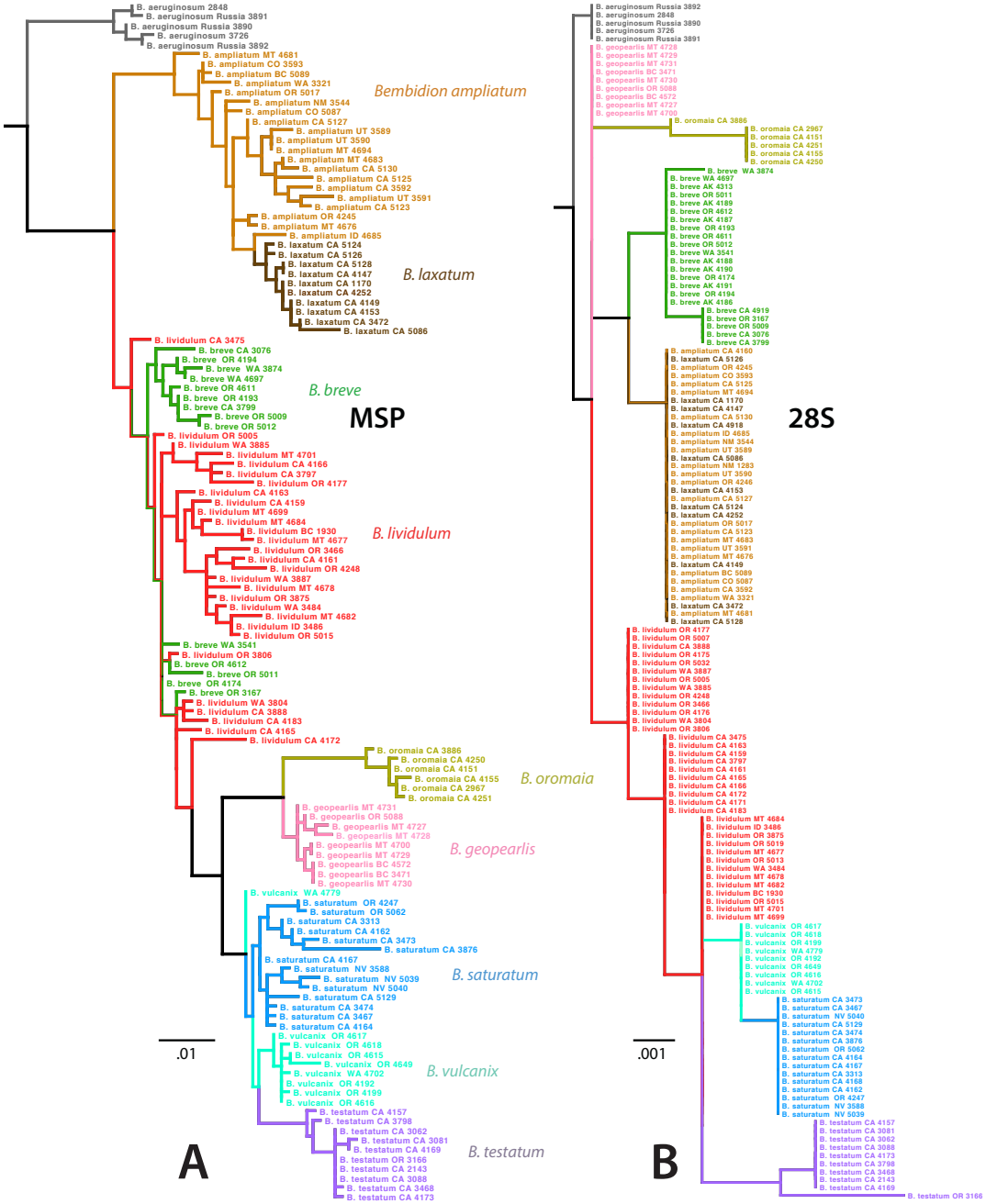


Figure 3.5. Maximum likelihood trees for two genes, (A) MSP and (B) 28S. Branch length shown is proportional to relative divergence with scale bars indicating 0.01 units for MSP, and 0.001 units in 28S. Outgroup taxa are shown in gray.

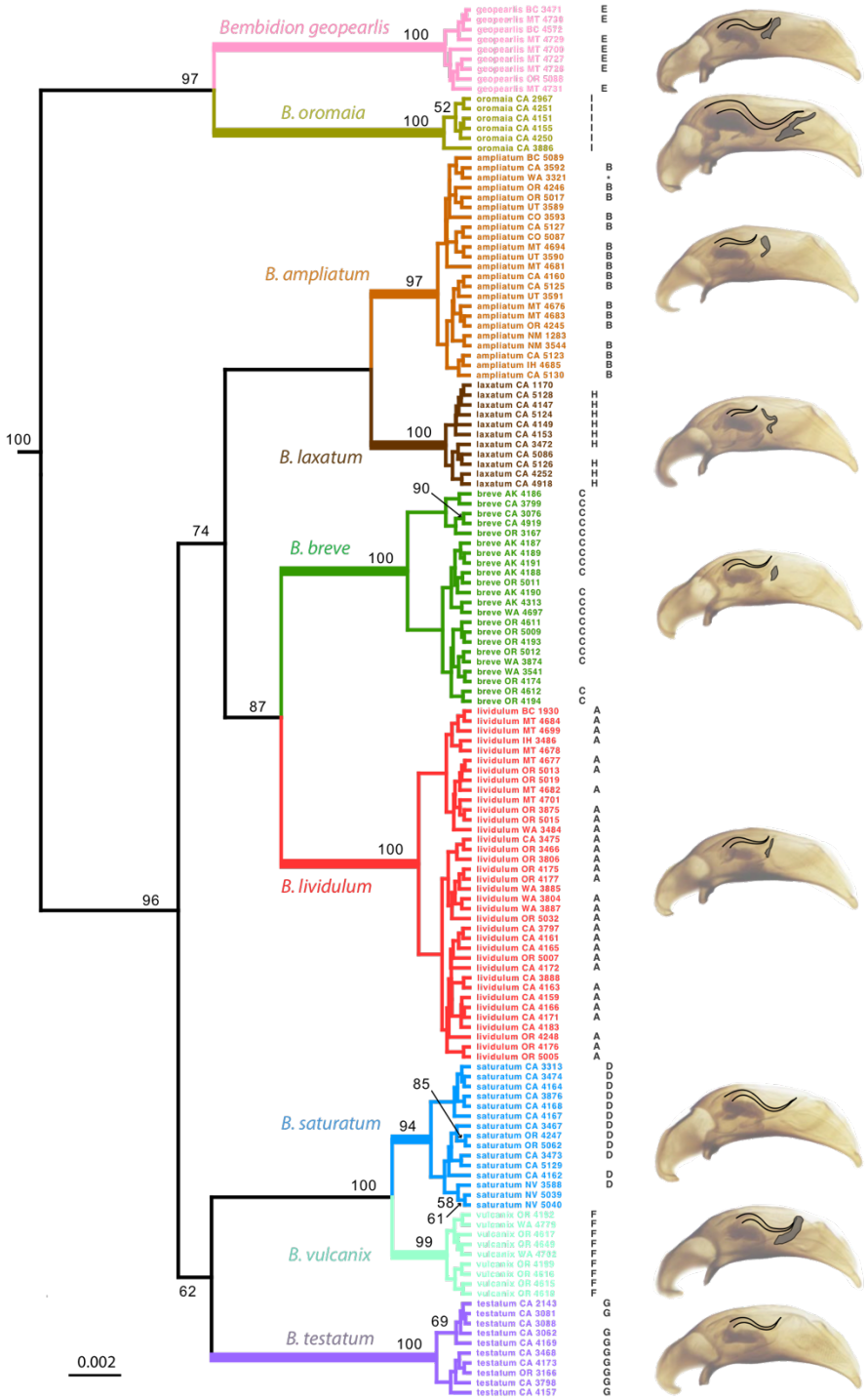


Figure 3.6. Species tree inferred by STACEY using five loci. Terminal taxa (and subtending branches) are colored according to species as inferred in this study based on all evidence; branches giving rise to each inferred species' clade are thickened. Posterior probabilities are shown above all branches with at least 50% support. Letters in columns to the right of most terminals indicate the coding of male genitalic characters assigned independent from the molecular data, and correspond to the letters in Fig. 3.10 (e.g., a specimen with an "E" corresponds to the genitalic form illustrated by Fig. 3.10E). Most terminals lacking a letter code are female; the

Figure 3.6. (Continued)

remaining six were uncoded males due to the genitalia being damaged or lost. Branch length shown is proportional to relative divergence with scale bars indicating 0.01 units

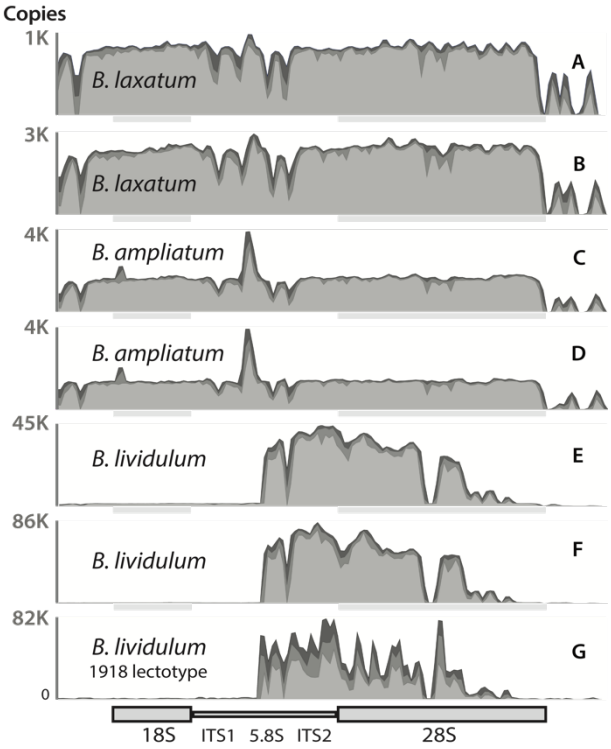


Figure 3.7. Plots of copy number variation across the rDNA cistron. (A) *B. laxatum*, California: Alpine Co., Sonora Pass, 2900m 38.3323°N 119.6401°W, DRM voucher DNA4918; (B) *B. laxatum*, California: Tulare Co., snowfield below White Chief Lake, 2912m, 36.417°N 118.5941°W, DRM voucher DNA4149; (C) *B. ampliatum*, Oregon: Harney Co., Steens Mts., snowfield at Kiger Gorge, 2618m, 42.7152°N 118.5786°W, DRM voucher DNA4245; (D) *B. ampliatum*, Colorado: Mesa Co., Grand Mesa, route 65 at FS100, 3243m, 39.0316°N 108.0561°W, DRM voucher DNA3593; (E) *B. lividulum*, California: Fresno Co., Kaiser Pass Meadow, 2783m, 37.2948°N 119.1006°W, DRM voucher DNA4165; (F) *B. lividulum*, Oregon: Klamath Co., Crater Lake NP, Sun Notch, 2163m, 42.9009°N 122.0988°W, DRM voucher DNA5032; (G) *Bembidion lividulum* Casey 1918 lectotype, California: Placer Co. "Copies" indicates the maximum copy number observed for any portion of the rDNA cistron. For example, there are approximately 1,000 copies of most of the cistron present in the specimen shown in A; for the specimen shown in F, there are almost 86,000 copies for a region from 5.8S through part of 28S. The position of mapped reads relative to the boundaries of rRNA genes within the rDNA cistron is provided along the bottom of the figure.

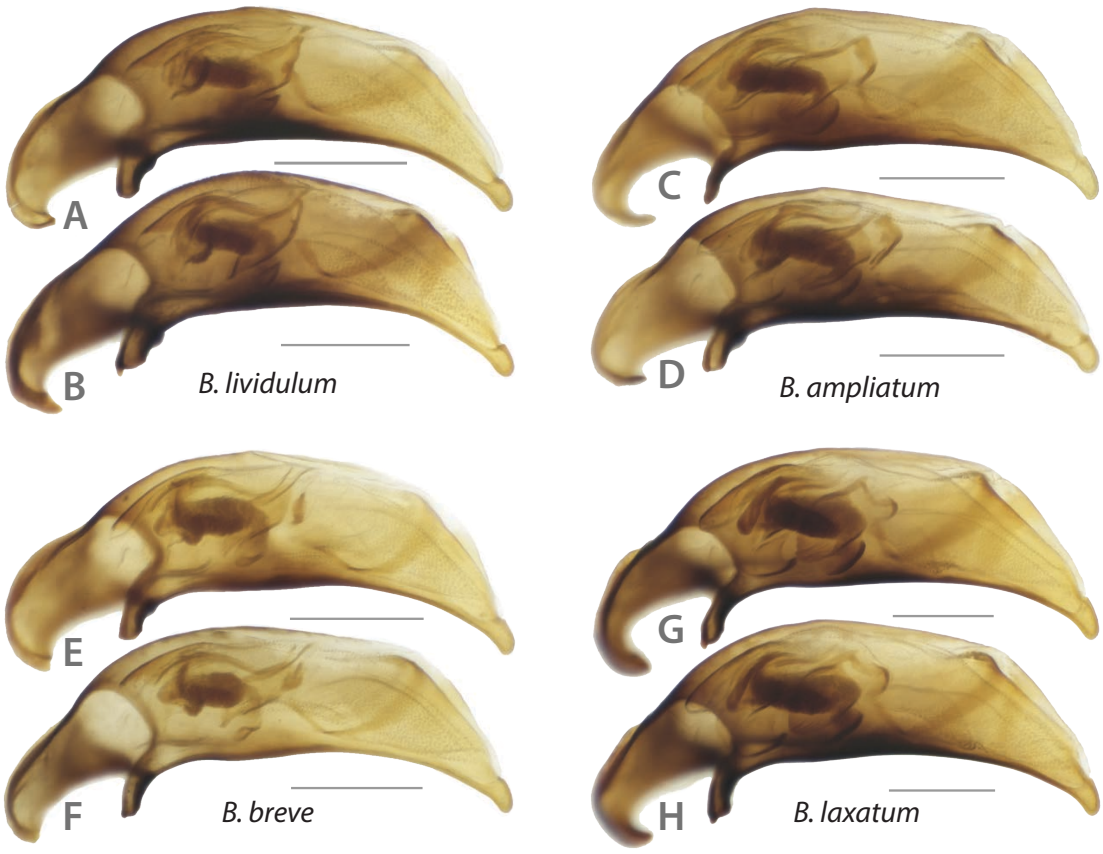


Figure 3.8. Male genitalia, left side. (A) *B. lividulum*, USA: California: Alpine Co., pond below Ebbetts Pass, DRM voucher DNA4161; (B) *Bembidion lividulum*, USA: Montana: Ravalli Co., snow above Bailey Lake, DRM voucher DNA4684; (C) *B. ampliatum*, USA: Utah: San Juan Co., Geyser Pass Rd nr Horse Ck, La Sal Mtns, DRM voucher DNA3590; (D) *B. ampliatum*, USA: California: Mono Co., snow field above Ellery Lake, DRM voucher DNA4160; (E) *B. breve*, USA: Alaska: Juneau, Heintzleman Ridge, DRM voucher DNA4187; (F) *B. breve*, USA: Washington: Whatcom Co., Bagley Lakes, Mt Baker, Snoqualmie NF, DRM voucher DNA4697;

Figure 3.8. (Continued)

(G) *B. laxatum*, USA: California: Fresno Co., South Fork Kings River, DRM voucher DNA4147; (H) *B. laxatum*, USA: California: Tulare Co., snowfield below White Chief Lake, DRM voucher DNA4149. Scale bar is 0.25 mm.

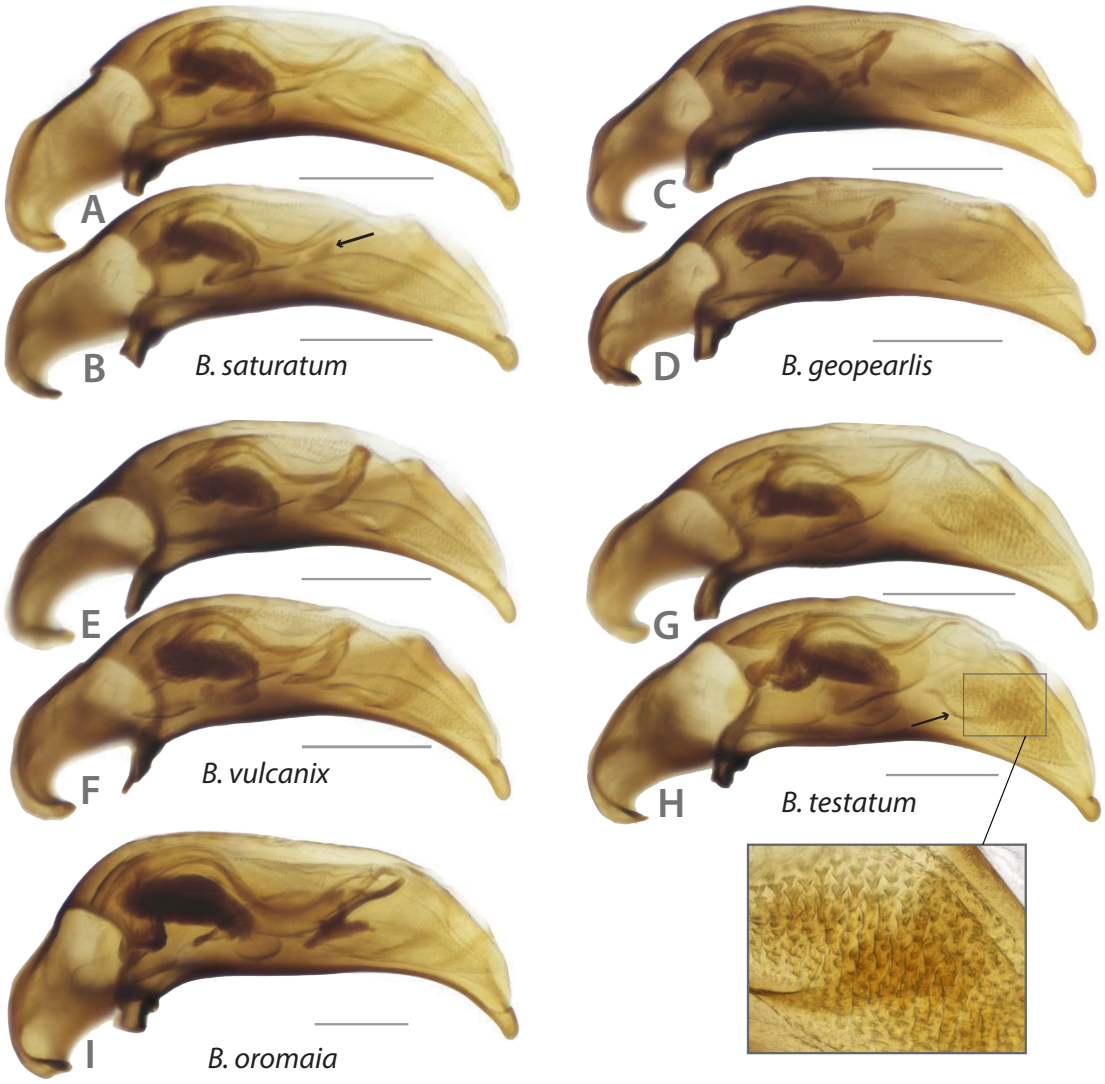


Figure 3.9. Male genitalia, left side. (A) *B. saturatum*, USA: California: Fresno Co., Kaiser Pass Meadow, DRM voucher DNA4168; (B) *B. saturatum*, arrow indicates the region that is occasionally lightly sclerotized in some specimens, USA: California: Tulare Co., snow field above Emerald Lake, DRM voucher DNA4162; (C) *Bembidion geopearlis*, USA: Montana: Glacier Co., Glacier N.P., east slope Clements Mtn., DRM voucher DNA4728; (D) *B. geopearlis*, USA: Montana: Glacier Co., Glacier N.P., east slope Clements Mtn., DRM voucher DNA4730; (E) *B. vulcanix*, USA: Oregon: Deschutes Co., Stream east of Todd Lake, DRM voucher DNA4617;

Figure 3.9. (Continued)

(F) *B. vulcanix*, USA: Oregon: Deschutes Co., Stream east of Todd Lake, DRM voucher DNA4615 (G) *B. testatum*, arrow indicates abbreviated ostial flag, USA: California: Tulare Co., 2.5 km N Sherman Pass, Sequoia NF, DRM voucher DNA3798; (H) *B. testatum* with closeup of distinctive scales in apical third that cause a darkened patch unique to this species, USA: California: Trinity Co., Canyon Creek, DRM voucher DNA4173; (I) *B. oromaia*, USA: California: Tulare Co., Upper East Fk. Kaweah River, DRM voucher DNA4251. Scale bar is 0.25 mm.

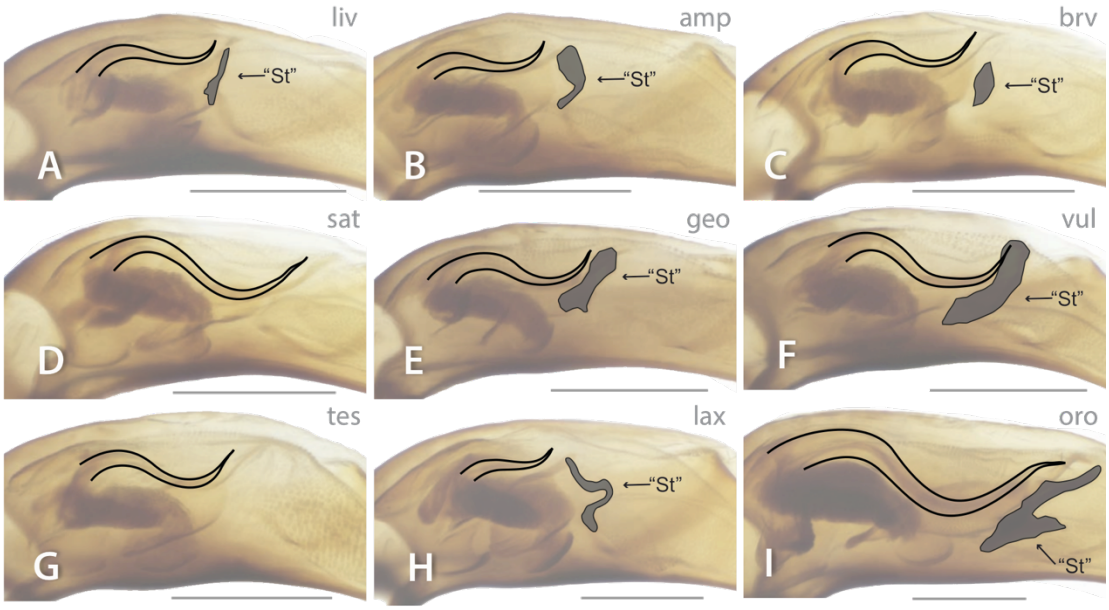


Figure 3.10. Male genitalia (left side) with emphasis added to sclerite "St" and flagellum shape. (A) *Bembidion lividulum*, see 'A' in Fig. 3.8 caption; (B) *B. ampliatum*, see 'C' in Fig. 3.8 caption; (C) *Bembidion breve*, see 'E' in Fig. 3.8 caption; (D) *B. saturatum*, see 'A' in Fig. 3.9 caption; (E) *B. geopearlis*, see 'D' in Fig. 3.9 caption; (F) *B. vulcanix*, see 'E' in Fig. 3.9; (G) *B. testatum*, see 'G' in Fig. 3.9; (H) *B. laxatum*, see 'G' in Fig. 3.8; (I) *B. oromaia*, see 'I' in Fig. 3.9. Scale bar is 0.25 mm.

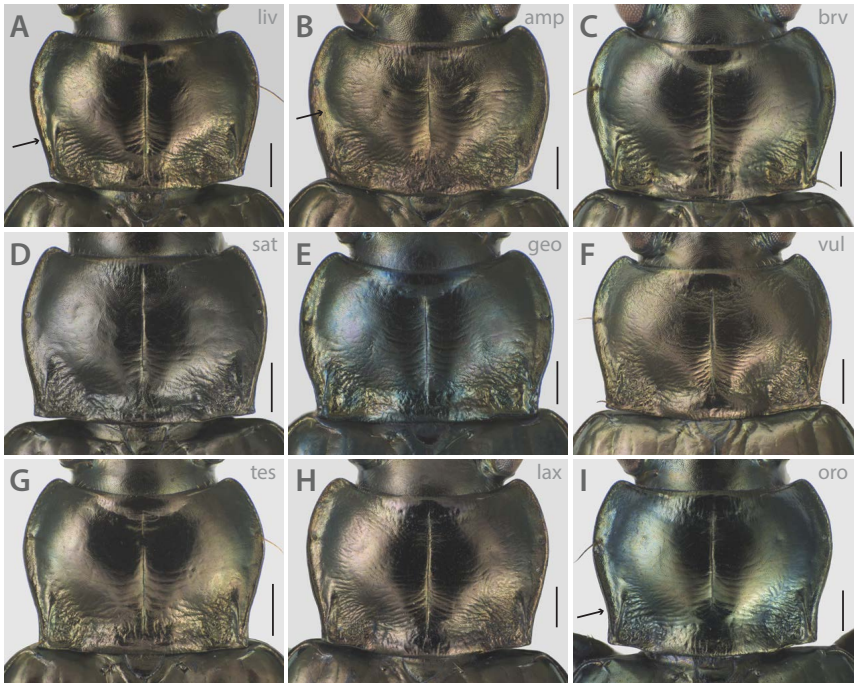


Figure 3.11. Pronota of the *breve* group species. (A) *Bembidion lividulum*, arrow indicates the non-sinuate lateral border, compare with 'H' and 'I' below, USA: California: Fresno Co., Kaiser Pass Meadow, DRM voucher DNA4165; (B) *B. ampliatum* indicating the broad lateral explanation, USA: New Mexico: Santa Fe Co., Santa Fe Ski Basin, DRM voucher DNA3544; (C) *B. breve*, USA: Washington: King Co, Snoqualmie River at Alpental, DRM voucher V100955; (D) *B. saturatum*, USA: California: Tulare Co., outlet of Emerald Lake, DRM voucher DNA4164; (E) *B. geopearlis*, USA: Montana: Ravalli Co., Lost Horse Creek, DRM voucher DNA4700; (F) *B. vulcanix*, USA: Oregon: Deschutes Co., Stream east of Todd Lake, DRM voucher V101130 (G) *B. testatum*, USA: California: Sierra Co., creek above Tamarack Lake, DRM voucher DNA4169; (H) *B. laxatum*, USA: California: Tulare Co., snow field above Emerald Lake, DRM voucher DNA4252; (I) *B. oromaia*, arrow indicates the strongly sinuate lateral border, compare with 'A' and 'B' above, USA: California: Tulare Co., snowfield below White Chief Lake, DRM voucher V100959. Scale bar is 0.25 mm. Note: we removed forelegs and painted the background in Photoshop in order to reduce visual complexity around the margins of the pronotum.

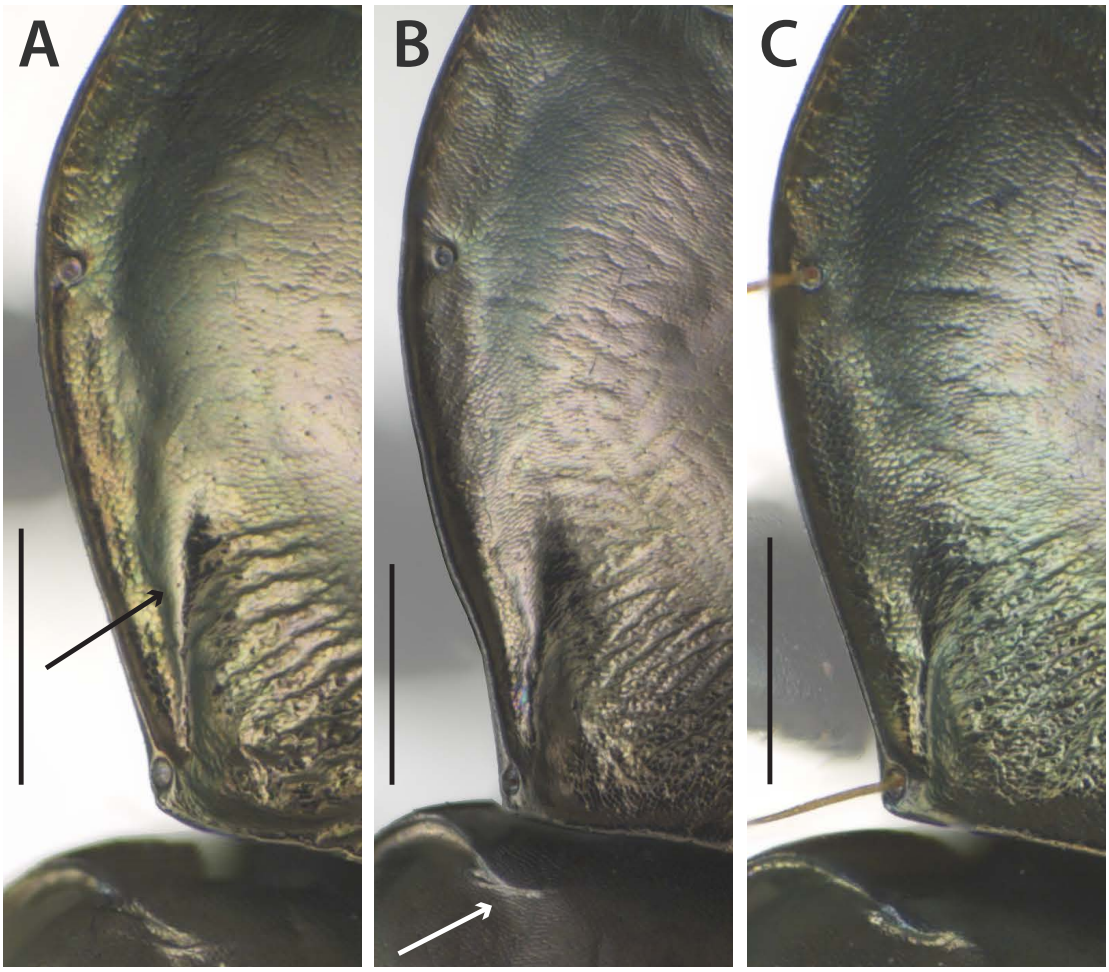


Figure 3.12. Laterobasal carina of pronota. Black arrow indicates the location of the carina, white arrow indicates the re-curved elytral bead typical of *breve* group members and some other *Plataphus*. (A) *Bembidion lividulum*, USA: California: Fresno Co., Kaiser Pass Meadow, DRM voucher DNA4165; (B) *B. laxatum,* USA: California: Tulare Co., snow field above Emerald Lake, DRM voucher DNA4252; (C) *B. ampliatum*, the small shadow along the right margin of the carina is due to the shallower basal fovea relative to 'A' and 'B', USA: California: Mono Co., snow field above Ellery Lake, DRM voucher DNA4160. Scale bar is 0.25 mm.

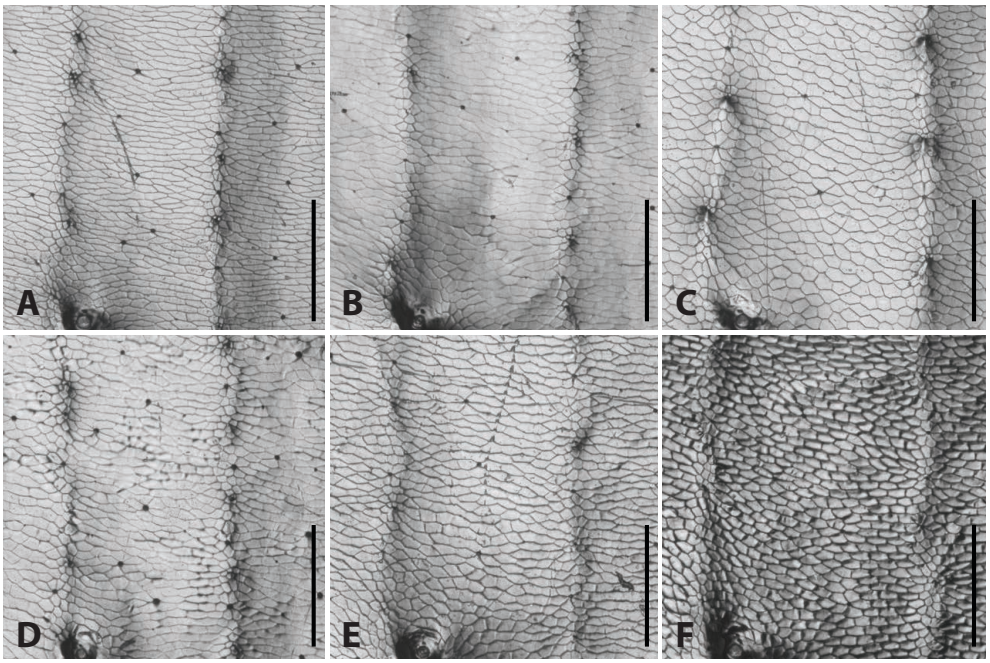


Figure 3.13. Elytral microsculpture. (A) *Bembidion breve*, USA: Alaska: Dall Isl., karst shallow pond, male, DRM voucher DNA4188; (B) *B. lividulum*, male, USA: California: Tuolumne Co., Deadman Creek, DRM voucher V100961; (C) *B. oromaia,* male, USA: California: Tulare Co., Upper East Fk. Kaweah River, DRM voucher V101127; (D) *B. ampliatum*, male, USA: Colorado: Mesa Co., Grand Mesa, route 65 at FS100, DRM voucher DNA3593; (E) *B. laxatum*, male, USA: California: Tulare Co., snowfield below White Chief Lake, DRM voucher DNA4149; (F) *B. laxatum*, female, USA: California: Mono Co., pond above Tioga Lake, DRM voucher V101126. Scale bar is 0.1 mm.

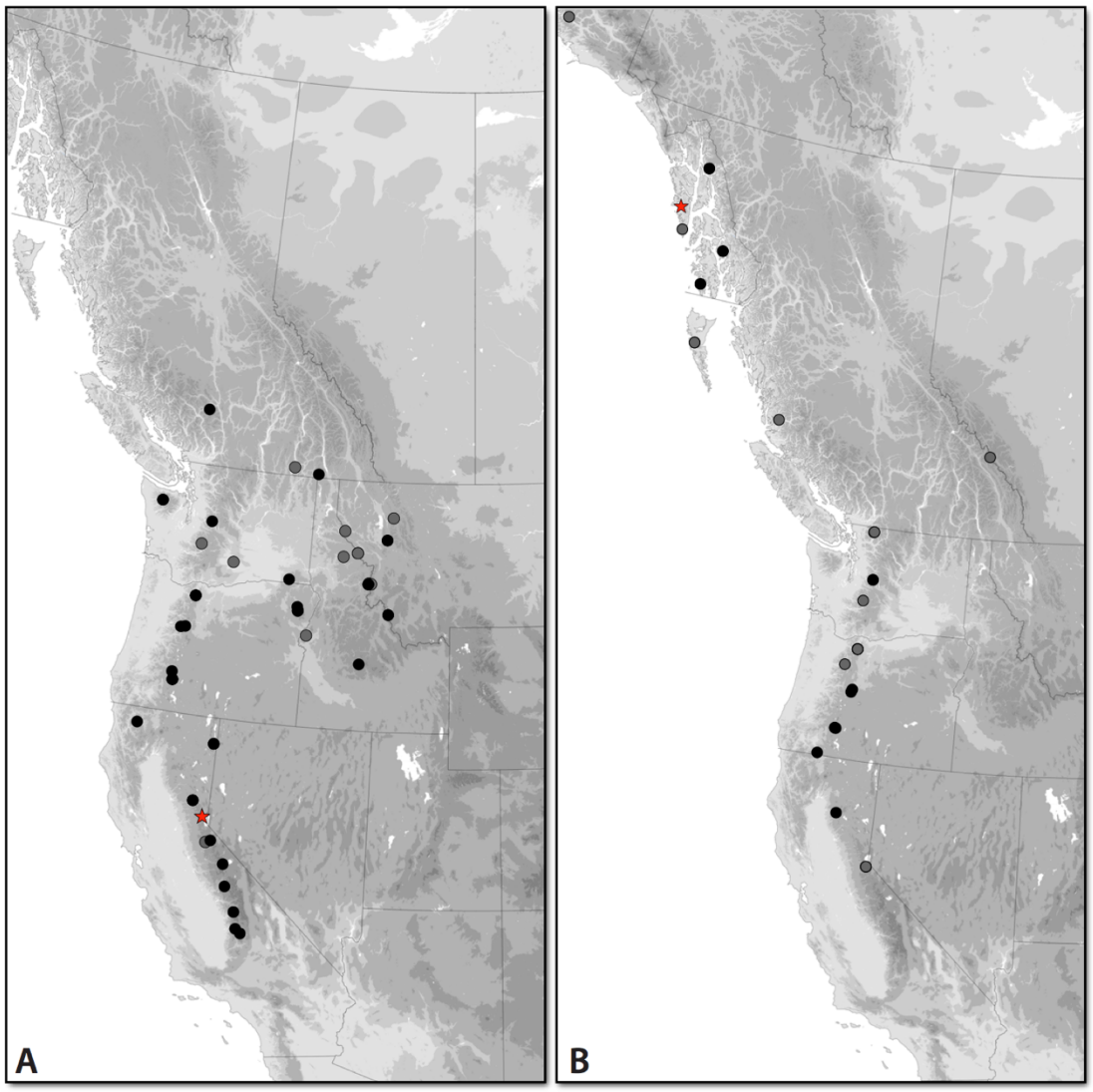


Figure 3.14. Distributions of *Bembidion lividulum* (A) and *B. breve* (B). Black circles indicate specimens identified through analysis of DNA (male and female) and genitalic characters (males only). Gray circles indicate specimens identified using male genitalic characters. Red stars indicate the type localities estimated using label data. The locality for a specimen of *B. breve* specimen (DNA4313) from the Aleutian Islands is not shown.

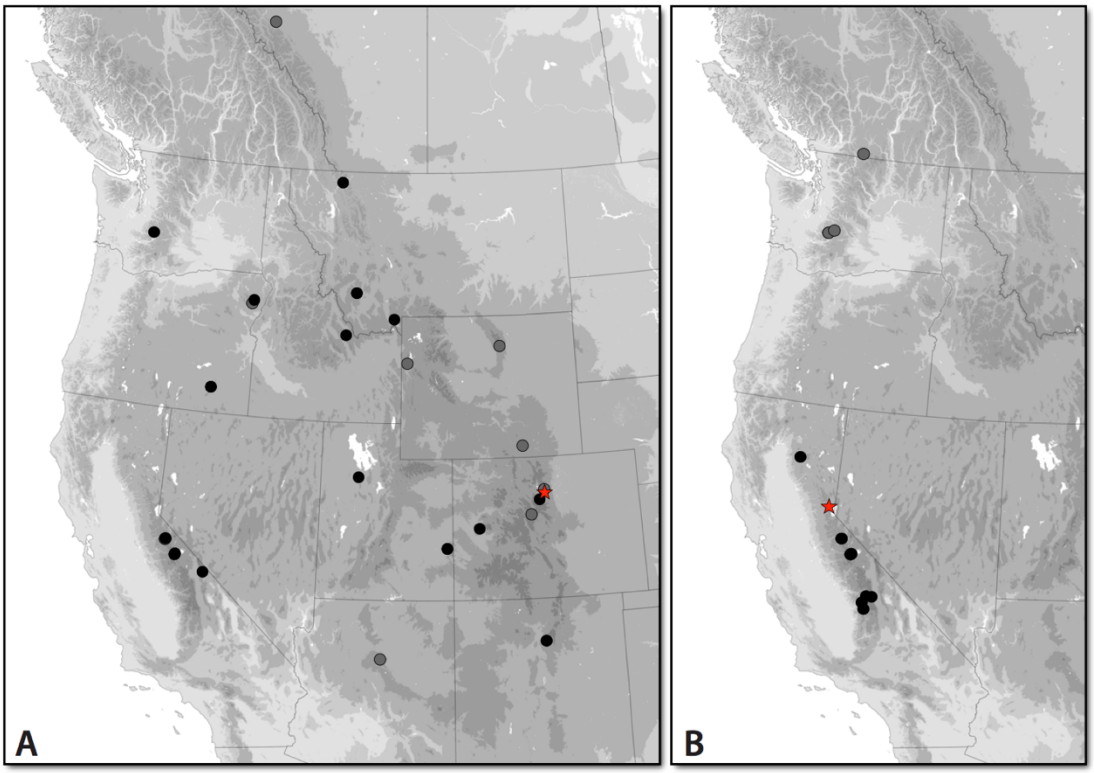


Figure 3.15. Distributions of *Bembidion ampliatum* (A) and *B. laxatum* (B). Black circles indicate specimens identified through analysis DNA (male and female) and genitalic characters (males only). Gray circles indicate specimens identified using male genitalic characters. Red stars indicate type localities estimated using label data.

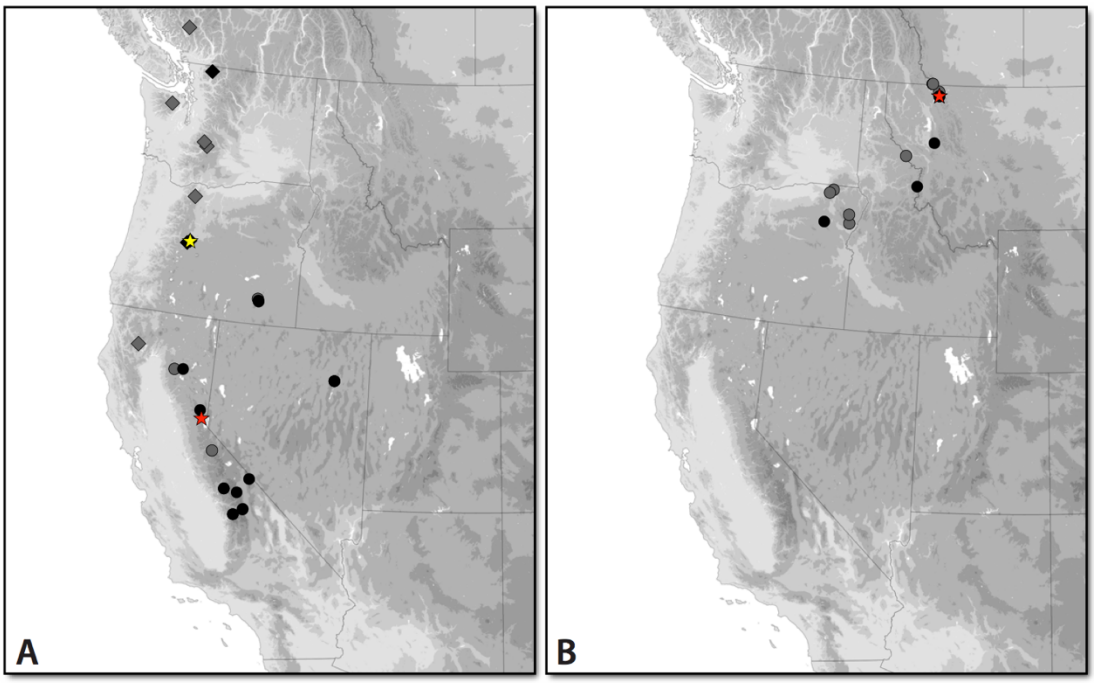


Figure 3.16. Distributions of *B. saturatum* (indicated by circles) and *B. vulcanix* (indicated by diamonds) (A) and *Bembidion geopearlis* (B). Black shapes indicate specimens identified through analysis DNA (male and female) and genitalic characters (males only). Gray shapes indicate specimens identified using male genitalic characters. Red stars indicate type localities of *B. saturatum* (A) and *B. geopearlis* (B), the latter being estimated using label data. The yellow star indicates the type locality of *B. vulcanix* (A).

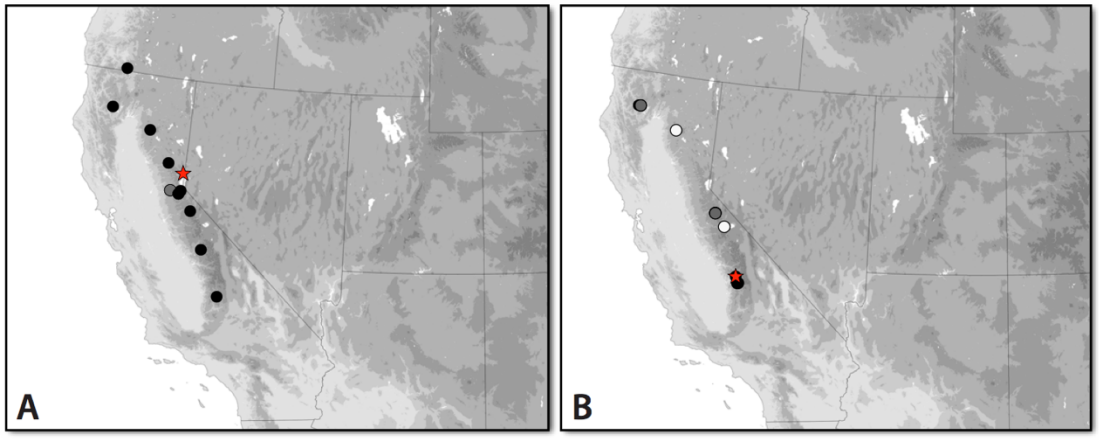


Figure 3.17. Distributions of *Bembidion testatum* (A) and *B. oromaia* (B). Black circles indicate specimens identified through analysis DNA (male and female) and genitalic characters (males only). Gray circles indicate specimens identified using male genitalic characters. White circles indicate specimens identified with external morphological characters. Red stars indicate type localities (estimated using label data in the case of *B. testatum*).

TABLES

Table 3.1. Localities of *breve* group specimens examined in molecular studies. Four digit numbers in the first column are DRM voucher numbers for DNA extracted specimens.

***Bembidion aeruginosum* (Gebler)**

3890	Russia: Altai Republic, Krasnaya Mountain, 1786m, 50.09393°N 85.22787°E
3726	Russia: Altai Republic, Krasnaya Mountain, north slope, 1688m, 50.14306° N 85.23018°E
3892	Russia: Altai Republic, Mukhor Cherga, 8m, 51.32917°N 85.30889°E
2848	Russia: Irkutsk, Khamar-Daban Mts., 3 km S peak Chersky, 1900m, 51°29.14'N 103°36.50'E
3891	Russia: Krasnoyarskij Krai, Kulumys Ridge, 1547m, 52.87291°N 93.24260°E

***Bembidion ampliatum* Casey**

5089	Canada: British Columbia: Downtown Road, km 17, 50.5303°N 122.2712°W
4160, 5127	USA: California: Mono Co., snow field above Ellery Lake, 2901m, 37.9345°N 119.2318°W
5123, 5125	USA: California: Mono Co., snow field near Tioga Pass, 3116m, 37.9123°N 119.2472°W
3592	USA: California: Mono Co., White Mtns, N Fk Crooked Creek, 3060m, 37.5064°N 118.1628°W
5130	USA: California: Tuolumne Co., snow field near Sonora Pass, 2908m, 38.3322°N 119.65°W
3593	USA: Colorado: Mesa Co., Grand Mesa, route 65 at FS100, 3243m, 39.0316°N 108.0561°W
4685	USA: Idaho: Lemhi Co., Meadow Lake Creek, Lemhi Mtns., 2735m, 44.4393°N 113.3144°W
4681, 4683	USA: Montana: Beaverhead Co., Canyon Creek, Pioneer Mtns., 2219m, 45.6273°N 112.9375°W
4676	USA: Montana: Gallatin Co., Timber Creek on Forest Road 985, 2109m, 44.914°N 111.3593°W
4694	USA: Montana: Glacier Co., Glacier N.P., east slope Clements Mtn., 2131m, 48.6907°N 113.7308°W
1283	USA: New Mexico: Santa Fe Co., Santa Fe Ski Basin, 3260m, 35.79°N 105.797°W
3544	USA: New Mexico: Santa Fe Co., Santa Fe Ski Basin, 3286m, 35.7889°N 105.7953°W
4245, 4246	USA: Oregon: Harney Co., Steens Mts., snowfield at Kiger Gorge, 2618m, 42.7152°N 118.5786°W

Table 3.1. (Continued)

5017	USA: Oregon: Wallowa Co., S of Mount Howard, 2525m, 45.255°N 117.1769°W
3589, 3590	USA: Utah: San Juan Co., Geyser Pass Rd nr Horse Ck, La Sal Mtns, 3012m, 38.4801°N 109.2646°W
3591	USA: Utah: Tooele Co., Stansbury Mtns, Deseret Peak, 2660m, 40.4715°N 112.621°W
3321	USA: Washington: Thurston Co., Mt Rainer, Summerland trail (approx. 1580m, 46.8695°N 121.6522°W)
4306	USA: Wyoming: Johnston Co., Powder River Pass, 2957m, 44.1504°N 107.0801°W
5087	USA: Colorado: Clear Creek Co., Mountain W of Berthoud Pass, 3620m, 39.8021°N 105.7868°W
***Bembidion breve* (Motschuylsky)**	
4186, 4188	USA: Alaska: Dall Isl., karst shallow pond, 574m, 54.99884°N, 133.01418°W
4190	USA: Alaska: Etolin Isl., 2944m, 56.13769°N, 132.32967°W
4189	USA: Alaska: Etolin Isl., alpine pond, 2454m, 56.14282°N, 132.33691°W
4187, 4191	USA: Alaska: Juneau, Heintzleman Ridge, 826m, 58.41904°N, 134.4422°W
4313	USA: Alaska: Umnak Island, Crater Creek E. of Ogmak Caldera, (approx. 54m, 53.5476°N, 167.9876°W)
3076, 4919	USA: California: El Dorado Co., Lily Lake, 2000m, 38.8743°N 120.0801°W
3799	USA: California: Tehama Co., Nanny Creek, Lassen NF, 1584m, 40.3696°N 121.5612°W
4611	USA: Oregon: Deschutes Co., Creek below Little Three Creek Lake, 2018m, 44.1057°N 121.6347°W
4612	USA: Oregon: Deschutes Co., Stream east of Todd Lake, 1952m, 44.0282°N 121.6709°W
4174, 4194	USA: Oregon: Hood River Co., Hood River Meadows Ski Area, 1586m, 45.3254°N 121.6625°W
4193	USA: Oregon: Hood River Co., Mitchell Creek, 1627m, 45.3276°N 121.6661°W
3167	USA: Oregon: Jackson Co., Mt Ashland Campground, Klamath NF, 2040m, 42.0756°N 122.714°W
5009	USA: Oregon: Klamath Co., Castle Crest Trailhead, Crater Lake NP, 1942m, 42.8907°N 122.132°W

Table 3.1. (Continued)

5012	USA: Oregon: Klamath Co., Munson Creek, Crater Lake NP, 1981m, 42.8987°N 122.1343°W
5011	USA: Oregon: Klamath Co., Vidae Falls, Crater Lake NP, 1980m, 42.8832°N 122.0993°W
3541	USA: Washington: King Co., Snoqualmie Pass, Alpental Ski Area, 950m, 47.4449°N 121.4245°W
3874	USA: Washington: King Co., Snoqualmie R., 949m, 47.4446°N, 121.4247°W
4697	USA: Washington: Whatcom Co., Mt. Baker, Snoqualmie NF, 1291m, 48.8544°N 121.6969°W
4303	Canada: British Columbia, Rivers Inlet, Nickaqueet River, 300m, 54.65°N 127.267°W
***Bembidion geopearlis* Sproul and Maddison, sp. nov.**	
3471, 4572	Canada: British Columbia: Akamina Pass, 1740m, 49.0261°N 114.0611°W
4572	Canada: British Columbia: Akamina Pass, 1740m, 49.0261°N 114.0611°W
†4727, 4728, 4729, 4730	USA: Montana: Glacier Co., Glacier N.P., east slope Clements Mtn., 2129m, 48.692°N 113.7292°W
4305	USA: Montana: Mineral Co., Hoodoo Creek, 1780m, 46.9788°N 115.0347°W
4731	USA: Montana: Missoula Co., inlet to Heart Lake, 1891m, 47.3801°N 113.8501°W
4700	USA: Montana: Ravalli Co., Lost Horse Creek, 1760m, 46.1417°N 114.4863°W
5088	USA: Oregon: Baker County, Blue Mountains, Anthony Lake, 185m, 44.96122°N 118.23200°W
***Bembidion laxatum* Casey**	
5086	USA: CA: Lassen Co. Lassen National Park, Helen Lake, 2506m, 40.46740°N 121.50860°W
1170, 4918	USA: California: Alpine Co., Sonora Pass, 2900m, 38.3323°N 119.6401°W
4147	USA: California: Fresno Co., South Fork Kings River, 1564m, 36.7888°N 118.5529°W
3472	USA: California: Inyo Co. Lower lake at base of Dragon Peak, 3445m, 36.78650°N 118.36340°W

Table 3.1. (Continued)

4153, 5128	USA: California: Mono Co., snow field above Ellery Lake, 2901m, 37.9345°N 119.2318°W
5124, 5126	USA: California: Mono Co., snow field above Tioga Lake, 2963m, 37.9184°N 119.2542°W
4252	USA: California: Tulare Co., snow field above Emerald Lake, 2851m, 36.5959°N 118.6756°W
4149	USA: California: Tulare Co., snowfield below White Chief Lake, 2912m, 36.417°N 118.5941°W
***Bembidion lividulum* Casey**	
4308	Canada: Alberta: Wild Hay River, 1433m, (approx. 53.4736°N 118.2309°W)
1930	Canada: British Columbia: Summit Creek 5.5 km E of Kootenay Pass, 1420m, 49.0956°N 116.9998°W
3888	USA: CA: Siskiyou Co., Marble Mtn. Wilderness, S. of Black Marble Mtn,, 1917m, 41.5668°N 123.2068°W
4161	USA: California: Alpine Co., pond below Ebbetts Pass, 2671m, 38.5445°N 119.8115°W
4165	USA: California: Fresno Co., Kaiser Pass Meadow, 2783m, 37.2948°N 119.1006°W
4159, 4163	USA: California: Modoc Co., Emerson Creek, 1770m, 41.2635°N 120.1393°W
4183	USA: California: Mono Co., snow field above Tioga Lake, 2963m, 37.9184°N 119.2542°W
4171, 4172	USA: California: Sierra Co., creek above Tamarack Lake, 2065m, 39.607°N 120.6568°W
3797	USA: California: Tulare Co., 2.5 km N Sherman Pass, Sequoia NF, 2608m, 36.0096°N 118.3678°W
4166	USA: California: Tulare Co., outlet of Emerald Lake, 2814m, 36.599°N 118.677°W
3475	USA: California: Tulare Co., Rte 190, waterfall 21.5 mi E Springville, 1935m, 36.1292°N 118.5564°W
3486	USA: Idaho: Blaine Co., Galena Summitt, 2680m, 43.8728°N 114.7176°W
4677, 4678	USA: Montana: Beaverhead Co., creek crossing at Miner Lk. Rd., 2182m, 45.2976°N 113.6176°W
4304	USA: Montana: Mineral Co., Hoodoo Creek, 1780m, 46.9788°N 115.0347°W

Table 3.1. (Continued)

4699, 4701	USA: Montana: Missoula Co., Glacier Creek, 1484m, 47.3811°N 113.7948°W
4656	USA: Montana: Ravalli Co., Lost Horse Creek, 1513m, 46.1340°N 114.39418°
4682, 4684	USA: Montana: Ravalli Co., snow above Bailey Lake, 2290m, 46.1264°N 114.5206°W
4655	USA: Montana: Ravalli Co., stream at Schumacher Campground, 1998m, 46.1518°N 114.4986°W
4654	USA: Montana: Sanders Co., Prospect Creek, 1091m, 47.5754°N 115.64034°W
4248	USA: Oregon: Douglas Co., Silent Creek at Hwy 4795, 1587m, 43.1267°N 122.161°W
4175, 4176, 4177	USA: Oregon: Hood River Co., Hood River Meadows Ski Area, 1586m, 45.3254°N 121.6625°W
5005, 5007, 5032	USA: Oregon: Klamath Co., Sun Notch, Crater Lake NP, 2163m, 42.9009°N 122.0988°W
3806	USA: Oregon: Linn Co., Lost Lake, 1220m, 44.4336°N 121.9006°W
3466	USA: Oregon: Linn Co., Lost Prairie Campground on route 20, 1025m, 44.4037°N 122.075°W
3875	USA: Oregon: Wallowa Co., Lostine River at Two Pan Trailhead, 1706m, 45.2503°N 117.3768°W
5015	USA: Oregon: Wallowa Co., Lostine River Valley, 1483m, 45.3485°N 117.4152°W
5019	USA: Oregon: Wallowa Co., Lostine River, 1840m, 45.2378°N 117.3803°W
5013	USA: Oregon: Wallowa Co., Lostine River, Two Pan Trailhead, 1728m, 45.249°N 117.3763°W
3885, 3887	USA: WA: Jefferson Co., Olympic NP, Mt. Olympus, 1541m, 47.82465°N 123.68221°W
3484	USA: Washington: Columbia Co., Blue Mountains, .6 mi W Blue Ski Area, 1415m, 46.0949°N 117.8651°W
3804	USA: Washington: King Co, Snoqualmie Pass, 954 m, 47.4451°N 121.4245°W
***Bembidion oromaia* Sproul and Maddison, sp. nov.**	
3886	USA: CA: Trinity Co., Trinity Mountain Wilderness, Below Grizzly Lake, 2083m, 41.01509°N 123.04890°W

Table 3.1. (Continued)

2967	USA: California: Tulare Co., Lower Franklin Lake, 36.42029°N 118.56135°W
4151, †4250	USA: California: Tulare Co., snow field above Emerald Lake, 2851m, 36.5959°N 118.6756°W
4251	USA: California: Tulare Co., Upper East Fk. Kaweah River, 2812m, 36.4189°N 118.5927°W
4155	USA: California: Tuolumne Co., Deadman Creek, 2700m, 38.3188°N 119.6634°W
***Bembidion saturatum* Casey**	
4167, 4168	USA: California: Fresno Co., Kaiser Pass Meadow, 2783m, 37.2948°N 119.1006°W
3313	USA: California: Inyo Co. 1.5 km NE University Peak, 3240 m. 36.76030°N, 118.35450°W
3474	USA: California: Inyo Co. Onion Valley, near Independence Creek, 2929m, 36.77670°N 118.34590°W
3876	USA: California: Inyo Co., North Lake Campground, 2842m, 37.2266°N 118.6284°W
3467	USA: California: Lassen Co., Silver Lake, 1975m, 40.494°N 121.162°W
3473	USA: California: Mono Co., White Mountains Perry Aiken Canyon, 3292m, 37.6398°N 118.2278°W
5129	USA: California: Nevada Co. snow field, NW Carpenter Ridge, 2546 m, 39.41470°N 120.31440°W
4164	USA: California: Tulare Co., outlet of Emerald Lake, 2814m, 36.599°N 118.677°W
4162	USA: California: Tulare Co., snow field above Emerald Lake, 2842m, 36.5961°N 118.6752°W
3588, 5039, 5040	USA: Nevada: Elko Co., Lamoille Canyon, Ruby Mtns, 2697m, 40.6017°N 115.378°W
5062	USA: Oregon: Harney Co., N. slope Big Indian Gorge, Steens Mts., 2720 m, 42°40.00N 118° 35.33'W
4247	USA: Oregon: Harney Co., Steens Mountains, creek below summit, 2754m, 42.6408°N 118.5749°W
***Bembidion vulcanix* Sproul and Maddison, sp. nov.**	
4649	USA: Oregon: Deschutes Co., Creek below Little Three Creek Lake, 2018m, 44.1057°N 121.6347°W

Table 3.1. (Continued)

4199	USA: Oregon: Deschutes Co., E Todd Lake, Deschutes NF road 370, 1976m, 44.0306°N 121.6683°W
4192	USA: Oregon: Deschutes Co., NE Todd Lake, Deschutes NF road 370, 2067m, 44.038°N 121.6718°W
†4615, 4616, 4617, 4618	USA: Oregon: Deschutes Co., stream east of Todd Lake, 1952m, 44.0282°N 121.6709°W
4702	USA: Washington: Whatcom Co., Mt Baker, Snoqualmie NF, 1290m, 48.8534°N 121.6948°W
4779	USA: Washington: Whatcom Co., Mt Baker, Snoqualmie NF, 1326m, 48.8528°N 121.6886°W
***Bembidion testatum* Casey**	
3062	USA: California: El Dorado Co., Lily Lake, 2000m, 38.8739°N 120.0808°W
3081, 3088	USA: California: El Dorado Co., Strawberry Creek at Sciots Camp, 1760m, 38.7835°N 120.1463°W
4157	USA: California: Fresno Co., creek below Kaiser Pass, 2722m, 37.2865°N 119.1009°W
3468	USA: California: Shasta Co., Lassen National Park, nr Bumpass Hell. 2600m, 40.458°N 121.502°W
4169	USA: California: Sierra Co., creek above Tamarack Lake, 2065m, 39.607°N 120.6568°W
4173	USA: California: Trinity Co., Canyon Creek, 1440m, 40.949°N 123.0179°W
3798	USA: California: Tulare Co., 2.5 km N Sherman Pass, Sequoia NF, 2608m, 36.0096°N 118.3678°W
2143	USA: California: Tuolumne Co., Route 108 at Sonora Pass. 2870m , 38°20'10"N 119°38'44"
3166	USA: Oregon: Jackson Co., Mt Ashland Campground, Klamath NF, 2040m, 42.0756°N 122.714°W

†Indicates primary type specimens

Table 3.2. GenBank accession numbers for previously published sequences.

Species	#	28S	CAD	COI	Topo
Bembidion aeruginosum	3890	KY246723	KY246804	KY246764	KY246845
Bembidion breve	3799	KY246719	KY246800	KY246760	KY246841
Bembidion breve	4919	KY246741	KY246822	KY246782	KY246863
Bembidion laxatum	4153	KY246725	KY246806	KY246766	KY246847
Bembidion laxatum	4245	KY246730	KY246811	KY246771	KY246852
Bembidion laxatum	4918	KY246740	KY246821	KY246781	KY246862
Bembidion lividulum	1930	JN170300	JN170766	JN171002	JN171185
Bembidion lividulum	4161	KY246726	KY246807	KY246767	KY246848
Bembidion lividulum	4172	KY246729	KY246810	KY246770	KY246851
Bembidion saturatum	3313	KY246712	KY246793	KY246753	KY246834
Bembidion saturatum	3467	KY246714	KY246795	KY246755	KY246836
Bembidion saturatum	4167	KY246727	KY246808	KY246768	KY246849
Bembidion testatum	3062	KY246710	KY246791	KY246751	KY246832
Bembidion testatum	4169	KY246728	KY246809	KY246769	KY246850

Table 3.3. Optimal substitution models for phylogenetic analysis estimated in jModelTest. Differences between the model used in Garli and STACEY are because not all models are available in STACEY, and thus the next best fitting model that is available was chosen.

	28S	COI	Topo	MSP	CAD
Garli	K80+I	HKY+I	TrN+I	TPM3uf+I+G	TPM1uf+I+G
STACEY	JC+I	HKY+I	HKY+I	HKY+I+G	GTR+I

Table 3.4. Monophyly of inferred *breve* group species. #Gene Trees: indicates the number gene trees in which the species in monophyletic; STACEY Tree: indicates if the species was monophyletic in the STACEY topology.

Species	#Gene Trees	STACEY Tree
Bembidion geopearlis	5	yes
B. oromaia	5	yes
B. testatum	5	yes
B. breve	3	yes
B. lividulum	3	yes
B. saturatum	3	yes
B. laxatum	2	yes
B.ampliatum	0	yes
B. vulcanix	0	yes

Table 3.5. Spatial relationships among *breve* group species. M: indicates specimens are microsympatric (that is, collected at the same locality); S: indicates specimens that are broadly sympatric, but not known to occur in at the same locality; cells with numbers indicate the distance in kilometers between the nearest confirmed localities for non-sympatric species.

	B. ampliatum	*B. breve*	*B. geopearlis*	*B. lividulum*	*B. laxatum*	*B. oromaia*	*B. saturatum*	*B. testatum*
B. breve	S							
B. geopearlis	M	273 km						
B. lividulum	M	M	M					
B. laxatum	M	S	289 km	M				
B. oromaia	M	S	587 km	M	M			
B. saturatum	M	M	258 km	M	M	S		
B. testatum	S	M	484 km	M	S	S	M	
B. vulcanix	M	M	273 km	M	M	S	141 km	S

CHAPTER 4:

LOW-COST GENOMIC ARCHITECTURE IN NON-MODEL GROUPS

ABSTRACT

The structure and organization of the components within genomes (e.g., genes, gene families, repetitive elements, chromosome number and organization, etc.) holds information above and beyond what can be seen by an analysis of the sequences in the absence of their genomic context. Development of low-cost approaches to measure aspects of genomic architecture in non-model groups is needed, both to improve studies in phylogenetics, species delimitation, and genome evolution. In this study, we investigate evidence that substantial copy number variation of ribosomal DNA (rDNA), easily measured through low-coverage genome sequencing, is present across members of the *Bembidion breve* species group, a group of small ground beetles. We observe dramatic variation in copy number across the rDNA cistron (including in rRNA 18S and 28S gene regions) which shows species-specific signatures when sequencing reads are mapped to, and visualized against, a reference sequence. The pattern of these "rDNA profiles" varies across and is stable within putative species. We validate patterns seen in our sequence-based approach of generating rDNA profiles using fluorescence *in situ* hybridization (FISH). We also conduct cluster analysis of repetitive DNA to corroborate patterns seen in rDNA profiles and investigate patterns of variation in the repetitive components of the genome outside of rDNA. Our findings are consistent with many cytogenetic studies that document rDNA mobilization as a driver of genomic variation, and identify the *breve* species group as a potential model for studying rDNA mobilization and rapid genome evolution. Our methods may be useful in many groups as rDNA mobilization is a well-documented phenomenon; in addition, our approach could easily be adapted

to detect genomic architecture variation of other DNA repeats by substituting the reference sequence with a different repeat sequence. Our results highlight the potential value of methods that incorporate the signal of repetitive genomic architecture in studies on species delimitation and phylogenetics, and how the patterns observed in those studies can enhance studies on genome evolution. We argue that phylogenetic studies mapping measures of genomic architecture variation onto the phylogeny will be a critical intermediate strategy in moving toward an era when information contained in whole genomes can be used to both resolve phylogenetic relationships and understand the mechanisms that underlie patterns of diversification across the tree of life.

INTRODUCTION

Whole genomes are an emergent property of an organism in that the structure and organization of the genome's individual parts (e.g., the architecture of genes, gene families, repetitive elements, and their arrangement on chromosomes, referred to here as "genomic architecture") holds information above and beyond what can be seen by an analysis of the sequences in the absence of their genomic context (Ogura *et al.*, 2009; Stukenbrock, 2013; Soria-Carrasco *et al.*, 2014). However, whole genome assemblies are unavailable in most groups. The advent of reduced-representation sequencing approaches such as hybrid enrichment (Gnirke *et al.*, 2009; Faircloth *et al.*, 2012; Lemmon, Emme, & Lemmon, 2012) enable genome-scale analysis in these groups in the absence of whole-genome assemblies. These methods take the approach of sequencing a set of putatively homologous target loci, and eliminating most non-target DNA prior to sequencing. This strategy dramatically reduces sequencing cost, pairs well with existing sequence analysis tools, and are proving to be powerful tools in connecting the branches of the tree of life (McCormack *et al*., 2017).

However, data resulting from these methods are decoupled from their genomic context, which could otherwise add valuable signal (Stukenbrock, 2013, Dodsworth *et al*., 2015). For example, phylogenetic analysis of many independent loci may fail to separate specimens from two recently diverged lineages due to incomplete lineage sorting; however, discovery of fixed differences in chromosome number present in one lineage could make their distinctiveness immediately apparent (Maddison, 2008). In this way, the emergent property of genomic architecture can

concentrate and enhance the signal beyond what can be seen by knowing all of the details, and can allow us to see the forest in spite of all of the diverse gene trees. Development of low-cost approaches to measure signal of genomic architecture in non-model groups is needed, both to improve studies in phylogenetics and species delimitation, and increase synergy with studies on genome evolution.

Traditionally, genomic architecture variation has been studied using cytogenetic approaches such as karyotyping, G-banding, *in situ* hybridization, and comparative genomic hybridization (Pinkel *et al.*, 1988; Kallioniemi *et al.*, 1992; Speicher & Carter, 2005). Although these methods remain excellent tools, they typically require freshly available tissue and a specific set of cytogenetic or molecular skills not available to many. The genomic era now enables sequence-based study of genomic architecture through comparative genomic methods such as analysis of synteny, which is a powerful lens for both detecting variation in genomic architecture and understanding mechanisms that give rise to that variation (Consortium, 2002, 2004; Krzywinski *et al.*, 2009). However, despite the ever-increasing availability of genomic resources, comparative analysis of DNA sequences in their whole genomic context remains on the distant horizon in groups spanning diverse and non-model lineages. For example, in depth study of the genomic architecture across genes and gene families will not be possible without a nearby reference sequence. However, recent studies demonstrate that aspects of genomic architecture can be mined from the repetitive genome (Dodsworth *et al*., 2015; Sander-Lower *et al*., 2017; Sproul and Maddison, 2017). Development of additional sequence-based approaches for comparing genome architecture that can be easily and inexpensively measured from

any specimen has potential to add clarifying signal to studies in species delimitation, phylogenetics, and genome evolution.

In this study, we follow up on preliminary evidence that substantial variation in genomic architecture may be present across members of the *Bembidion breve* species group, a group of small ground beetles (Fig. 4.1). In a previous study that examined the recovery success of ribosomal DNA (rDNA) genes from a 100 year-old-type specimen from this group, the authors noted poor recovery of the 18S gene, which ranged from ~0–10X coverage depth (Sproul & Maddison, 2017a). The same specimen, however, showed >400X coverage in the 28S gene just a few thousand bases downstream of 18S. Subsequent sequencing of fresh conspecifics confirmed that the intergenic spacer (IGS) and some 28S regions had > 100-fold copy number increase relative to the 18S gene. Sequencing two additional species in the group showed each species had a unique signature of copy number variation across the rDNA cistron (the tandemly repeated region of rDNA containing 18S and 28S genes). The copy number differences between species were sufficiently large as to suggest major differences in this aspect of genomic architecture between closely related species. For example, in one specimen of *Bembidion laxatum*, 0.6% of all reads obtained through whole genome shotgun sequencing mapped to the rDNA complex, while for a *Bembidion lividulum*, an astounding 16.9% of all genomic reads obtained mapped to the rDNA cistron - the vast majority mapping to the inflated region of IGS+28S. These preliminary data led to the present study in which we conduct a more thorough investigation of variation in patterns of rDNA copy number within the *breve* group.

Nuclear rDNA is expected to occur in tandem arrays within the nucleolar organizing regions of euchromatin (loosely packed, gene-rich DNA), with clusters often appearing on more than one chromosome (McClintock, 1934; White, 1977; Schwarzacher & Wachtler, 1993). However, numerous studies document the transfer of rDNA fragments from euchromatic nucleolar organizing regions into heterochromatin (tightly packed, gene-poor, repeat-rich DNA) where they can undergo extensive multiplication, and subsequent loss (McClintock, 1934; White, 1977; Schwarzacher & Wachtler, 1993; Martins *et al.*, 2006; Raskina *et al.*, 2008; Nguyen *et al.*, 2010; Cioffi & Bertollo, 2012; Iwata-Otsubo *et al.*, 2016).

Mobilization of rDNA clusters has been frequently documented using cytogenetic methods in many groups including plants (Raskina, Belyayev, & Nevo, 2004; Qi *et al.*, 2015; Ding *et al.*, 2016; Wang *et al.*, 2016), fish (Martins *et al.*, 2006; Da Silva, Busso, & Parise-Maltempi, 2012; Symonová *et al.*, 2013, 2017), protists (Gong *et al.*, 2013), insects (Cabral-de-Mello, Moura, & Martins, 2010; Cabral-de-Mello *et al.*, 2011; Nguyen *et al.*, 2010; Panzera *et al.*, 2012; Palacios-Gimenez & Cabral-de-Mello, 2015), bivalves (Pérez-García *et al.*, 2014), and mammals (Sotero-Caio *et al.*, 2015), and is regarded as strong evidence of substantial changes to genomic architecture (Jiang & Gill, 1994; Raskina *et al.*, 2004, 2008). A number of studies further suggest that spreading of such multicopy gene families into heterochromatic regions may be mediated by retrotransposon activity (Dimitri *et al.*, 1997; Dimitri & Junakovic, 1999; Symonová *et al.*, 2013; de Bello Cioffi *et al.*, 2015), and ectopic recombination (Nguyen *et al.*, 2010).

We investigate patterns of rDNA copy number variation in the *breve* group at two levels: the variation across specimens within putative species, and the variation among putative species. We focus our efforts on sequence-based evidence derived from genome skimming (low-coverage whole genome sequencing) data, but also validate patterns seen in sequence-based approaches through cytogenetic mapping of rDNA using fluorescent *in situ* hybridization (FISH). We tested whether cluster-based analysis of repetitive DNA corroborated patterns seen in rDNA profiles. We also tested for variation in rDNA profiles at broader taxonomic scales in *Bembidion* subgenus *Plataphus*, the subgenus containing the *breve* group. As part of our investigation in the *breve* group, we outline a simple approach to visualizing differences in genomic architecture of rDNA by mapping reads to a reference and comparing the signatures resulting from copy number variation across specimens. As researchers studying molecular patterns in non-model groups find new ways to measure patterns of genomic architecture, this will both enhance studies in those disciplines, as well as promote new synergy with studies on comparative genomics and help identify new model systems for studying genome evolution.

METHODS

Overview

We investigated patterns of copy number variation (CNV) in the ribosomal cistron (hereafter referred to as rDNA profiles") across a framework of species recently delimited using evidence from molecular, morphological, and geographic data in Sproul and Maddison (2017a). For each of the nine recognized *breve* group species, we selected 3-8 specimens from across the species' geographic ranges to test whether signatures observed in rDNA profiles were consistent with putative species boundaries, and stable within species. We generated rDNA profiles by obtaining low-coverage whole-genome sequencing data and mapped reads for each specimen to a 14K-base reference sequence of the rDNA cistron of *Bembidion aeruginosum*. We chose *B. aeruginosum*, as our phylogenetic studies (see below) indicate that it is the sister group of the remaining *breve* group species. An overview of methods used to generate and display rDNA profiles shown here is provided in Figs. 4.2 and 4.3. We screened for contaminants as an alternative explanation to variation in rDNA profiles, conducted parameter sensitivity analysis for generating profiles (Fig. 4.4 and Table 4.1), studied the effect of reference choice (Fig. 4.5), compared profiles obtained from males and females, explored stability of profiles across varying read depth, tested whether profiles could be obtained from targeted sequencing approaches (i.e., hybrid capture), and searched patterns correlated with geography or phylogenetic patterns within species.

We tested the assumption that regions showing inflated copy number in rDNA profiles represent differences in genomic architecture with comparative fluorescence *in situ* hybridization (FISH) with two *breve* group species. We further validated patterns observed in rDNA profiles, and tested for additional differences in genomic architecture by conducting analysis of repetitive genomic elements using RepeatExplorer (Novák, Neumann, & Macas, 2010; Novák *et al.*, 2013).

We tested the hypothesis that rDNA CNV was of broader taxonomic significance by generating rDNA profiles for many species of the subgenus *Plataphus*, the clade that contains the *breve* group. Our methods are explained in more detail below, and in Appendix 3.

rDNA profile variation in the breve species group

Taxon sampling and DNA extraction, library preparation, and sequencing

We generated low-coverage whole-genome sequencing data from 42 breve group specimens selected from the taxon sampling of Sproul and Maddison (2017a), with 3–8 specimens chosen per species (Table 4.2). We obtained Illumina reads for six additional specimens (3593, 4149, 4165, 4245, 4918, and 5032) from Sequence Read Archive (SRA SRR5514451–SSR5514456). Reads obtained from SRA were generated previously by the authors using the same library preparation and sequencing protocols reported herein for newly sequenced specimens (Sproul & Maddison, 2017b,a). DNA extraction protocols are provided in Sproul and Maddison (2017a).

We prepared DNA extractions for Illumina sequencing by sonicating genomic DNA using a Diagenode Bioruptor Pico Sonicator for 10 minutes, and prepared dual-indexed libraries using NEBNext DNA Ultra II Library Prep Kits (New England Biolabs) following the manufacturer's recommended protocol. We attempted to maximize library evenness by standardizing input DNA quantities between 10–50 ng (except for five samples, three having less than 10 ng of available DNA, and two for which maximizing input was desirable due to their role in other projects), and by using a consistent number of library amplification cycles for a given amount of input (Table 4.3). The libraries were then pooled and sequenced on one of four 150 base paired-end lanes on an Illumina HiSeq 3000 maintained by the Oregon State University Center for Genome Research and Biocomputing. We allocated as much as 1/4 and as little as 1/52 of a lane per sample. Lane pairings are provided in Table 4.3.

rDNA profile generation and phylogenetic mapping

Demultiplexing of Illumina reads was performed using CASAVA v1.8 (Illumina). We imported paired-end reads into CLC Genomic Workbench v9.5.3 (CLC Bio, referred to below as CLC GW), with failed reads removed during import. We trimmed and excluded adapter sequences from reads in CLC GW. We randomly downsampled trimmed reads to 10 million per specimen, such that downstream analyses for all samples had a standardized number of input reads. We mapped trimmed reads to a 14K-base reference sequence of the rDNA cistron obtained from a *de novo* assembly of reads from *Bembidion aeruginosum* using the 'Map Reads to Reference' tool in CLC GW (match score=3, mismatch=4, insertion cost=3, deletion

cost=3, length fraction=0.85, similarity fraction=0.85). We chose read mapping parameters following a sensitivity analysis in which we repeated read mapping analysis across a range of parameter settings using four representative samples. Additional methods used in the parameter sensitivity analysis, screening mapped reads for contaminants and assembly artifacts, and obtaining the rDNA reference sequence are provided in Appendix 3. Following read mapping, we removed duplicate mapped reads in CLC GW.

We visualized the pattern of coverage depth resulting from read mapping by generating graphs of the read pileups in CLC GW (this can alternatively be accomplished by generating a histogram (in R, for example) of read coverage from a BAM file resulting from read mapping in CLC GW or other programs). We estimated the maximum number of rDNA copies for any point across the rDNA cistron by dividing the maximum read coverage depth of the rDNA cistron by the average coverage depth of 67 putatively single-copy nuclear protein-coding genes (Regier et al., 2008) which we mapped from the same set of reads (Sproul & Maddison, 2017a). We used copy number estimates to apply a color ramp to rDNA profiles in Illustrator such that the color of all rDNA profiles shown herein indicates copy number relative to the same absolute scale.

We studied the effect of reference choice on rDNA profiles by mapping reads to the standard reference of *B. aeruginosum* and an alternative approach in which the reads were mapped to a reference sequence obtained from *de novo* assembly of reads from the same species being mapped. Because CNV could be due to differences in rDNA clusters on sex chromosomes, and our initial taxon sampling was strongly

biased towards males, we included in the final dataset rDNA profiles for at least one female of each species that showed rDNA regions with increased copy number (i.e., greater than a 2-fold increase in copy number of any rDNA region) to determine whether profiles were stable in both sexes. We also tested the stability of rDNA profiles across varying numbers of input reads by generating profiles for the same four specimens using 10M, 5M, and 1M as input into read mapping.

We tested whether rDNA profiles could be conveniently obtained as part of a hybrid capture sequencing project by simply spiking an unenriched library of each sample into a sequencing run that included enriched libraries for the same samples. We conducted solution-based hybrid enrichment of nine breve group specimens (Table 4.3). Prior to sequencing, we pooled enriched libraries with equal weights of the original unenriched libraries from each sample. We sequenced enriched+unenriched pooled libraries on the same 150 paired-end lane on an Illumina HS 3000 at Oregon State University. The resulting reads were processed in the same way as samples not subject to enrichment, except that they were downsampled to 20M reads, instead of 10M reads, to account for the fact that approximately half the reads were expected to have come from molecules belonging to the enriched libraries. Importantly, the bait set used to enrich libraries lacked baits targeting rDNA regions such that, any reads contributing to rDNA profiles should not be affected by enrichment, and would therefore produce effectively equivalent rDNA profiles as those produced from a dedicated sequencing run of unenriched whole genomic libraries. We enriched libraries using a MYcroarray (Ann Arbor, MI) custom bait set designed to target approximately 1200 loci from carabid beetles (Maddison et al., in

prep). We performed target capture of libraries following the manufacturer's recommended protocol with additional details provided in Appendix 3.

We mapped rDNA profiles obtained for all *breve* group specimens to the tree used to infer species boundaries by Sproul and Maddison (2017a) in order to determine the extent of rDNA profile variation within the species group, and whether distinctive features in rDNA profiles within a species showed stable variation across individuals sampled from diverse geographic localities.

Cytogenetic mapping of ribosomal DNA

We performed fluorescence *in situ* hybridization (FISH) experiments with two *breve* group species, *Bembidion lividulum* and *B. testatum*. We designed fluorescent probes to target two regions of the rDNA cistron, a region in 18S that lacks copy number inflation in *B. lividulum* and has minor (approximately two-fold) inflation in *B. testatum*, and a region in 28S that had marked copy number inflation in *B. lividulum*, but no inflation in *B. testatum*. For each ribosomal target (e.g., 18S and 28S), we synthesized two non-overlapping ~500-base probes (Table 4.4) which we fluorescently labeled, pooled by locus, and hybridized to chromosome squashes of *B. lividulum* and *B. testatum*. We prepared chromosome squashes of testis tissue such that two squashes were available from each individual. This design allowed us to compare the pattern of FISH signals for both 18S and 28S in the same individual. With few exceptions, both 18S and 28S hybridizations were conducted for the same individual, in the same FISH experiment (i.e., the same batch of chemicals, incubation duration, and wash conditions).

We performed tissue dissection and fixation following Larracuente & Ferree (2015), and conducted FISH using protocols that combined steps from Larracuente (2017) and Symonová *et al*., (2015), and confirmed results using multiple probe synthesis and post-hybridization wash strategies. Additional details on FISH methods are provided in Appendix 3.

Cluster analysis of repetitive DNA

We further validated patterns observed in rDNA profiles, and tested for additional differences in genomic architecture by conducting analysis of repetitive genomic elements using RepeatExplorer (Novák *et al.*, 2010). RepeatExplorer uses short-read sequence data to generate graph-based clusters (Blondel *et al.*, 2008) of assembled repeats, and annotates clusters using public databases (Jurka *et al.*, 2005; Marchler-Bauer *et al.*, 2010). We conducted cluster analysis on all *breve* group specimens for which we generated rDNA profiles, except for eight specimens for which sequences were generated from a pool of both hybrid enriched and unenriched libraries (Table 4.3), as the inclusion of enriched loci violates the assumptions of the analysis. Prior to cluster analysis we estimated genomic coverage and downsampled reads to 0.25x coverage in CLC GW, and conducted clustering using the RepeatExplorer Galaxy-based web server (Novák *et al.*, 2013). RepeatExplorer output orders clusters based on genome proportion. We analyzed the top 100 clusters (i.e., the 100 clusters most abundant in the genome) and grouped clusters into six repetitive DNA categories: Class I transposable elements (TEs), Class II TEs, rDNA, simple repeats, unknown clusters, and unknown clusters containing rDNA hits from

BLAST and RepeatMasker databases. We generated pie charts to visualize variation in repetitive DNA across specimens. For each sample, we plotted the total number of clusters containing rDNA, the genome proportion of rDNA, and the number of clusters that contained shared hits for rDNA and any other repeat category. For two species, *B. lividulum* and *B. laxatum*, we also plotted the abundance of hits from superfamilies of Class I and Class II TEs.

Testing for rDNA profile variation across Bembidion (Plataphus)

We tested for profile variation across a broader taxonomic scope by sampling 31 species across the subgenus *Plataphus* (Table 4.5). We inferred the *Plataphus* phylogeny in IQ-TREE (Nguyen *et al.*, 2014) as implemented in Mesquite v3.2 (Maddison & Maddison, 2017) using a six gene dataset. Methods used for sequencing, alignment, and phylogenetic analysis are provided in Appendix 3.

We generated rDNA profiles for *Plataphus* species using the same methods described above, except that we relaxed the stringency of read mapping parameters slightly (length fraction=0.80 and similarity fraction=80) following the results of our parameter sensitivity analysis (Fig. 4.4). We selected multiple taxa from each major clade in the subgenus, including several pairs of closely related species.

RESULTS

rDNA profile variation in the breve species group

Ribosomal DNA profiles generated from the *breve* species group showed species-specific signatures of variation across the group (Figs. 4.6). Profiles from six of nine species showed regions with notably inflated copy number (i.e., 2–100+ fold CN increase) relative to the rest of the rDNA cistron (Fig. 4.6). The position of inflated regions within the rDNA cistron, the magnitude of inflation (indicated by the color ramp of profiles in figures), and variation in maximum copy number across the inflated region (i.e., the shape of the inflated region), all contributed to the distinctive characteristics of rDNA profiles for most species in the group. Profiles for three species (*B. laxatum, B. oromaia* and, *B.vulcanix*) lacked regions of notable variation within the rDNA cistron (Fig. 4.6) Sister relationships were frequently defined by striking differences in rDNA profile variation. For example, rDNA profiles from *Bembidion geopearlis, B. breve*, and *B. testatum,* all had regions within the rDNA cistron that showed at least a 10-fold increase (averaged across all individuals sampled) in maximum copy number relative to the same region in their sister taxon (Fig. 4.7 and Table 4.3).

Species specific signatures observed in rDNA profiles were highly stable across multiple individuals of each species sampled from various geographic localities (with one notable exception discussed in more detail below) (Fig. 4.7 and Figs. 4.8–25). Across individuals sampled for *Bembidon saturatum*, two specimens had profiles that lacked the CN inflation in IGS and 28S that was observed in the

remaining specimens of that species (Figs. 4.7 and 4.9). The two distinctive *B. saturatum* profiles were from specimens that were geographically isolated (Steens Mountains, OR and Ruby Mountains, NV) from the remaining specimens, which were all collected in the Sierra Nevada, CA. Both specimens from which rDNA profiles were generated were identified by Sproul and Maddison (2017a) to have morphological traits distinctive from Sierra Nevada populations. Profiles from these specimens were similar to those of *B. laxatum, B. oromaia*, and *B. vulcanix* in that they lacked any regions of notable (i.e., greater than 2-fold) copy number inflation.

Although rDNA profiles for *Bembidion lividulum* all showed striking copy number inflation in the same region of IGS and 28S, we noted minor variation in profile shape (i.e., copy number) within the inflated region which showed patterns consistent with phylogenetic position and geographic locality of the specimens sampled (Fig. 4.8 and 4.17). We noted minor variation in other species (e.g., *B. saturatum*) that did not show obvious correlation with phylogenetic or geographic patterns (Fig. 4.9).

Profiles generated from males and females showed consistent patterns regardless of sex (Figs. 4.8–16). Profiles generated from the same specimen using 10 million, 5 million, and 1 million reads were all nearly identical in shape (Fig. 4.26). We also found that profiles obtained from a combined pool of hybrid enrichment samples with the unenriched libraries were consistent with profiles obtained from unenriched samples. The average total fraction of reads mapping to the 14K-base reference sequence ranged from an average of 1.07% (StDev=0.2%, n=7) in *B. ampliatum* to 14.7% (StDev=2.8%, n=8) in *B. lividulum*.

Cytogenetic mapping of ribosomal DNA

Patterns observed in FISH experiments corroborated our assumption that regions showing inflated copy in rDNA profiles are due to mobilization of rDNA throughout the genome, and contribute to differences in genomic architecture between species. In both species, hybridization with probes targeting rDNA regions lacking copy number inflation (18S in *B. lividulum*, and 28S in *B. testatum*) produced two FISH signals in interphase cells, whereas hybridization with probes designed in regions with copy number inflation (marked 28S inflation in *B. lividulum*, and slight 18S inflation in *B. testatum*) showed more than two bright FISH signals (4–5 loci in *B. testatum*, and many loci in *B. lividulum*) (Figs. 4.27 and 4.28). In *B. lividulum*, sufficient tissue and replicate squashes were available to confirm the distribution of FISH signals on condensed, well-spread chromosomes. Uninflated 18S rDNA mapped to two chromosomes, while markedly inflated 28S rDNA showed FISH signals on portions of all 24 chromosomes (Figure 4.27). The pattern of FISH signals on condensed chromosomes suggests that much of the 28S rDNA is concentrated in heterochromatic regions of chromosomes, and frequently absent on euchromatic tails (Figure 4.29).

Experiments in which we hybridized 28S probes synthesized from *B. lividulum* DNA to *B. testatum* chromosomes produced the same patterns as 28S probes synthesized from *B. testatum* DNA, as did hybridization of *B. testatum*-based 28S probes to *B. lividulum* chromosomes. FISH results in *B. lividulum* corroborate our evidence seen in overall reads that mapped to the rDNA cistron and the extreme patterns observed in rDNA profiles for this species suggesting that rDNA (or the

remnants of the rDNA cistron that have mobilized throughout the genome) makes up a surprisingly large fraction of the *B. lividulum* genome (Fig. 4.6–7, Table 4.3).

Cluster analysis of repetitive DNA

Cluster analysis in Repeat Explorer corroborated general patterns observed in rDNA profiles. *Bembidion lividulum* had an average of 21 clusters containing rDNA hits whereas none of the species that lacked regions of notable inflation in rDNA profiles had more than 4 clusters with rDNA. Plotting genome proportions for major repeat categories, as well as total hits for TE superfamilies in *B. lividulum* vs *B. laxatum* revealed notable variation between the species in the repetitive genome outside of rDNA. *Bembidion lividulum* showed variation in major repeat categories and superfamilies of Class I and Class II TEs that followed geographic and phylogenetic patterns similar to rDNA profiles of the same specimens.

rDNA profile variation across Bembidion (Plataphus)

We found evidence in rDNA profiles that rDNA mobilization is widely distributed across the subgenus *Plataphus* (Fig. 4.30). Although the majority of species in the subgenus showed one to two peaks of slight (approximately 1.5 to 2-fold) inflation within the IGS region of the rDNA cistron (similar to those observed in *Bembidion ampliatum and B. laxatum* in the *breve* group (Fig 4.6)), but they otherwise showed little variation between species. However, approximately one in three species sampled showed distinctive variation relative to the common profile form just described, and in many cases, there appeared to be phylogenetic signal underlying that variation. Those species with variable profiles showed either dramatic

inflation in the IGS peak beyond that of typical rDNA profiles (e.g., *Bembidion gordoni* and near relatives), or they showed additional regions of inflation in either 18S or 28S regions (e.g., *B. haruspex* and *B. sp.nr.sierracola*) (Fig. 4.30). Although the remaining species in the subgenus were generally less variable across species than the breve group, there were three instances where two species that are difficult to tell apart using morphological and or molecular data showed variation from one another (*B. sp.nr.sierracola* + *B. sierracola*, *B. sp.nr.curtulatum "*Idaho*"* + *B. sp.nr.curtulatum "*Bay Area*"*, and *B. kupranovii* #1 + *B. kupranovii* #2 BC and AK). *B. haruspex* had 6% plus total reads that mapped to the rDNA cistron (Table 4.6).

DISCUSSION

Our results demonstrate that rDNA profiles have excellent potential as a low-cost measure of genomic architecture variation in groups with limited genomic resources. Our findings from both sequence-based analysis and cytogenetic mapping of rDNA corroborate cytogenetic studies in numerous groups across phyla that suggest that the mobilization (*i.e.,* movement of a portion of the rDNA cistron to a new location), and subsequent expansion (*i.e.,* duplication) of rDNA is an important driver of genomic architecture differentiation among closely related taxa (Raskina *et al.*, 2008; Panzera *et al.*, 2012; Gong *et al.*, 2013; Symonová *et al.*, 2013; Sember *et al.*, 2015). We show through rDNA profiles that this variation can be stable within putative species (cite Figs. 4.7 and 4.8-16) and can provide cleaner signal than individual gene trees, multi-gene analyses (as in the tree shown by Fig. 4.7), and morphological characters (Sproul & Maddison, 2017a). We show that rDNA profiles can be obtained with very low sequencing coverage depth (Fig. 4.26) in the absence of high quality reference genomes, and can be generated within the workflow of phylogenomic projects. Using this, or similar methods with other repetitive loci, to incorporate sequence-based signal from genomic architecture has potential to enhance studies on species delimitation, phylogenomics, and genome evolution.

rDNA profiles as a species delimitation tool

Delimiting the nine putative species in the *breve* group required several years of focused effort to synthesize evidence from multi-gene datasets, external morphological structures, internal reproductive structures, and geographic data (e.g.,

maintained molecular differences among putative species that occur in microsympatry). In the present example, signatures in rDNA profiles showed variation consistent with putative species boundaries that were stable across multiple individuals sampled per species (Fig. 4.7). These signatures of variation allowed for unambiguous identification of all putative species in a challenging group, for which other single lines of evidence (e.g., morphological differences alone) frequently have insufficient signal for separating species pairs or trios. This finding demonstrates that rDNA profiles (or similar measures of locus-specific genomic architecture variation) have excellent potential as a species delimitation tool in non-model groups.

The only species concept in which we observed discordant rDNA profile shapes across specimens was *B. saturatum*. The two distinctive *B. saturatum* profiles (Fig. 4.9–10) were from specimens collected at geographically isolated localities (Fig. 4.10) (Steens Mountains, OR and Ruby Mountains, NV) known to be regions of endemism in other plant and animal groups (Hershler & Sada, 2002; Houston, Shiozawa, & Riddle, 2010; Sproul *et al.*, 2014, 2015). Both specimens from which rDNA profiles were generated were identified by Sproul and Maddison (2017a) to have morphological traits distinctive from Sierra Nevada populations (e.g., differences in male genitalia, pronotal and elytral shape, and microsculpture intensity), but were not formally recognized as separate for a lack of clear evidence from gene trees, and for having sampled few individuals from few localities. In this example, rDNA profiles add additional evidence that cryptic species may be present within *B. saturatum* specimens, and they do so with more clarity than coalescent

analysis of several genetic loci, which did not corroborate morphological data (Sproul & Maddison, 2017a).

We observed notable intraspecific variation in rDNA profiles in *Bembidion lividulum* and *B. saturatum*. The intraspecific profile variation, and general estimates of repetitive genome content in *B. lividulum*, showed patterns consistent with phylogenetic and geographic signal (Fig. 4.8 and 4.17). This signal may be due to the fact that *B. lividulum* has highly structured populations across its range (evidenced by consistent structure across gene trees) and rDNA profiles are reflecting that structure. This conclusion is consistent with fine-scale studies of rDNA structure in yeast that have shown that polymorphism and copy number variation in rDNA show phylogenetic signal and geographic signal across strains of putative species (West *et al.*, 2014; James *et al.*, 2016). An alternative hypothesis is that fine-scale variation in *B. lividulum* profiles is detecting fully differentiated, recently diverged lineages across which gene flow does not occur (i.e., separate species). This is also plausible given that variation in rDNA profiles corresponds to phylogenetic position within major clades of *B. lividulum* (Fig. 4.8) and we have observed slight differences in morphological patterns (e.g., body size variation and distinctive male genitalia in the northern Rocky Mountains). In this case, patterns in rDNA profiles provide signal that prompts a directed re-evaluation of other lines of evidence and collecting of additional specimens to challenge the current species concept. This finding also highlights the need for future exploration of quantitative methods for analyzing fine-scale differences among rDNA profiles.

Our finding evidence of variation in rDNA profiles across the subgenus *Plataphus* is consistent with other studies showing that rDNA mobilization can be common across broad taxonomic groups (Nguyen *et al.*, 2010; Cabral-de-Mello *et al.*, 2011; Sember *et al.*, 2015; Wang *et al.*, 2016), and broadens the potential relevance of rDNA profiles as tools in species delimitation. Whether or not rDNA profiles will be informative within a group will depend on (1) whether or not mobilization of rDNA has occurred in the evolutionary history of the lineage being studied, and (2) whether the event was recent enough to allow detection of inflated rDNA regions using sequence similarity, as mobilized rDNA pseudogene sequences that have escaped concerted evolution are expected to degrade over time (Wang *et al.*, 2016). For those groups in which rDNA lacks interesting signal, similar principles to those presented here with rDNA profiles may be employed with other repetitive loci such as sequences of specific transposable elements.

rDNA profiles as a model for studying genomic architecture signal in other repetitive loci

Although our approach here focuses on copy number differences in the rDNA cistron, which will be relevant for some groups, similar methods can be used with any repetitive locus that contributes to genomic architecture variation. We found that general characteristics of the repetitive genomes outside of rDNA showed signal between species that corroborated patterns seen in rDNA profiles (Figs. 4.8–4.16). Our strategy of mapping whole-genomic reads to a reference sequence to create a visual profile for comparisons can be applied to any loci that hold signal of genomic

architecture differences among samples. Analyzing low-coverage reads with a tool such as RepeatExplorer (Novák *et al.*, 2010) (one of the few analytical tools that allows for study of the repetitive DNA without requiring a reference genome or library of repeats) can facilitate identification of candidate repeats (i.e., specific transposable elements, or satellites) to serve as reference sequences. Such approaches may be a valuable tool for groups in which rapid evolution of the repetitive genome is already a well-documented driver of differentiation among lineages (Vitte & Bennetzen, 2006; Feschotte & Pritham, 2007; Raskina *et al.*, 2008; Cioffi & Bertollo, 2012), and can assist in identifying new groups for study.

Further development of low-cost, sequence-based measures of genomic architecture is especially valuable in studying specimens that are otherwise difficult to study, such as those that have large genomes, lack genomic resources, and historical specimens with poor quality DNA. Analyzing rDNA sequence of a historical specimen provided the context in which the authors discovered the rDNA signal in the *breve* group. The pattern of rDNA inflation in *B. lividulum* allowed the authors to identify a 100-year-old female type specimen for which the DNA was too degraded to provide sufficiently clean signal in analysis of individual gene sequences, and the female specimen lacked the male genitalic characters otherwise critical to studying the group (Sproul & Maddison, 2017a).

rDNA or other repetitive DNA profiles paired with phylogenomics

Development of reduced representation sequencing methods such as hybrid capture (Gnirke *et al.*, 2009; Faircloth *et al.*, 2012; Lemmon, Emme, & Lemmon, 2012) enable cost-effective generation of genome-scale molecular data in groups with limited genomic resources. However, because these methods are designed to sequence a set of target loci in the genome and thereby eliminate non-target DNA from final datasets (both through pre-sequencing sample preparation and post-sequencing bioinformatics analyses that filter out repetitive loci), data resulting from this non-random reduction of the genome eliminates signal that may be present in the repetitive genome. We show that directly sequencing the unenriched library on the same sequencing lane as the enriched library is a viable approach to generating rDNA profiles in parallel with targeted enrichment phylogenomics.

This strategy for also generating whole-genome shotgun (or genome skimming) data as part of target enriched phylogenomic studies adds no additional sample preparation cost, given that a whole genomic library must be produced as the first step in preparing a sample for hybrid enrichment, and not all of that library need be subjected to enrichment. Thus, the only cost associated with generating genome skimming data in parallel with target enrichment data is the cost of sequencing. Given that analyses presented herein were conducted using 10 million reads per sample which provided excessive coverage for extracting the signal we report herein (Fig 4.26), we estimate the total cost of generating sequence data for repetitive DNA profiles can be as low as $5–50 per sample (depending on genome size) for sequencing projects using an Illumina HiSeq 3000 (this cost estimate assumes the genomic library is already in hand as part of another project, and the estimate is based

on current sequencing costs at Oregon State University). For a project that includes multiple lanes of sequencing, the whole-genomic libraries can be pooled and sequenced on lanes separate from the enriched libraries derived from the same specimens such that reads resulting from each category of library can be maintained separate, despite the fact that both categories of reads will share multiplexing indices.

Many studies have highlighted the utility of genome skimming for harvesting sequences of multi-copy genes such as rDNA, and organellar genomes for phylogenetic analysis (Straub *et al.*, 2012; Kanda *et al.*, 2015; Richter *et al.*, 2015). Several recent studies demonstrate creative ways to extract more depth of signal from genome skimming data by moving beyond the analysis of individual sequences and considering signals that capture signal of genome-scale variation (Dodsworth *et al.*, 2015; Denver *et al.*, 2016; Lower *et al.*, 2017). Our results support findings that characteristics of the repetitive genome itself can directly add signal to phylogenetic studies (West *et al.*, 2014; Dodsworth *et al.*, 2015). Such data can hold important signal in recent radiations of species which occur on short branches for which the signal to noise ratio is poor in classic phylogenetic studies.

We argue that phylogenetic studies that map measures of genomic architecture variation onto the phylogeny will be a critical intermediate strategy that can hasten the arrival of an era when information contained in whole genomes can be used to both resolve phylogenetic relationships and understand the mechanisms that underlie patterns of diversification across the tree of life.

rDNA profiles as a complement to studies in genome evolution

Although our original goal in studying the rDNA profiles was to assess their utility in the context of species delimitation studies, the pattern of variation they revealed immediately stimulated research questions related to genome evolution. Our results corroborate many studies cited herein that document mobilization rDNA as an important driver of genome evolution across plants, animals, and fungi. Our finding of extreme rDNA mobilization in *B. lividulum* is consistent with other studies that show rDNA movement can result in dramatic, rapid restructuring of genomic architecture over short time periods (Raskina *et al.*, 2008; Panzera *et al.*, 2012; Symonová *et al.*, 2013; Ding *et al.*, 2016), and identifies the *breve* group as a future model for studies that explore the mechanisms underlying rapid genome evolution, or the effect of rDNA mobilization on chromosome pairing and genome evolution (McKee, Habera, & Vrana, 1992), and its role in speciation (Raskina *et al.*, 2004). In this way, the use of rDNA profiles or similar measures of genome architecture differences has potential to increase synergy between fields of phylogenetics, molecular cytogenetics and comparative genomics.

Although our sequence-based approach lacks fine-scale details (such as signals within chromosomes) provided by cytogenetic mapping techniques, our approach has the advantage that it can be applied to any specimen for which DNA sequences can be obtained, including specimens with old and ancient DNA (Fig. 4.31). Because DNA sequencing projects can be designed for increasingly high throughput, rDNA profiles or similar methods have excellent potential as a tool for efficiently surveying patterns of genomic architecture across phyla, which has

potential to identify new model systems for study and efficiently direct efforts of higher resolution, but more costly approaches to studying genome evolution.

Future directions

We are hopeful that the simple methods presented herein will be directly useful in future studies on species delimitation and genome evolution, and that our findings will stimulate development of additional approaches to measuring low-cost genomic architecture signal from sequence data. In general, our test data showed notable variation and clean signal in rDNA profiles, such that differences between putative species were quite obvious in most cases. We do not present a quantitative method for comparing moderate or minor variation among profiles. Such an approach may not be necessary in cases where variation across profiles is striking, but would be useful in discerning patterns when variation is less obvious.

We urge researchers generating genomic libraries for phylogenomic approaches such as target capture to consider the potential value of allocating a fraction of a lane to a multiplexed pool of whole-genomic libraries in order to map easily obtained metrics of genome evolution across their phylogeny.

ACKNOWLEDGEMENTS

We are indebted to Barbara Taylor, Amanda Larracuente, Radka Symonova, and James Strother for their support and advice in optimizing FISH protocols. We thank Anne-Marie Girard-Pohjanpelto for help with fluorescence imaging. We thank Aaron Liston, Eli Meyer, David Lytle, Tiffany Garcia for input on the scientific approach and the manuscript.

We thank Olivia Boyd for her valuable input on figure design. We are grateful to Danielle Mendez for her help with library quantitation. We thank Elizabeth Sproul and Steven Wainwright for editorial input on the writing. We thank Olivia Boyd, William Cresko, Dee Denver, Michael Freitag, Antonio Gomez, Kojun Kanda, David Kavanaugh, Wendy Moore, James Pflug, José Serrano, and Kipling Will for many stimulating discussions that improved the scope work presented here. We thank Elizabeth, George, and Pearl Sproul, as well as Janet and Greg Reed for their help collecting specimens used for FISH.

REFERENCES

de Bello Cioffi M, Bertollo LAC, Villa MA, de Oliveira EA, Tanomtong A, Yano CF, Supiwong W, Chaveerach A. 2015. Genomic organization of repetitive DNA elements and its implications for the chromosomal evolution of channid fishes (Actinopterygii, Perciformes). *PLoS One* 10: e0130199.

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008: P10008.

Cabral-de-Mello DC, Cabrero J, López-León MD, Camacho JPM. 2011. Evolutionary dynamics of 5S rDNA location in acridid grasshoppers and its relationship with H3 histone gene and 45S rDNA location. *Genetica* 139: 921–931.

Cabral-de-Mello DC, Moura R, Martins C. 2010. Chromosomal mapping of repetitive DNAs in the beetle Dichotomius geminatus provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. *Heredity* 104: 393–400.

Cioffi M, Bertollo L. 2012. Chromosomal distribution and evolution of repetitive DNAs in fish. *Repetitive DNA*. Karger Publishers, 197–221.

Consortium MGS. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520.

Consortium ICGS. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.

Da Silva E, Busso A, Parise-Maltempi PP. 2012. Characterization and Genome Organization of a Repetitive Element Associated with the Nucleolus Organizer Region in Leporinuselongatus (Anostomidae: Characiformes). *Cytogenetic and Genome Research* 139: 22–28.

Denver DR, Brown AMV, Howe DK, Peetz AB, Zasada IA. 2016. Genome Skimming: A Rapid Approach to Gaining Diverse Biological Insights into Multicellular Pathogens. *PLOS Pathogens* 12: e1005713.

Dimitri P, Arcà B, Berghella L, Mei E. 1997. High genetic instability of heterochromatin after transposition of the LINE-like I factor in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* 94: 8052–8057.

Dimitri P, Junakovic N. 1999. Revising the selfish DNA hypothesis: new evidence on accumulation of transposable elements in heterochromatin. *Trends in Genetics* 15: 123–124.

Ding XL, Xu TL, Wang J, Luo L, Yu C, Dong GM, Pan HT, Zhang QX. 2016. Distribution of 45S rDNA in Modern Rose Cultivars (Rosa hybrida), Rosa rugosa,

and Their Interspecific Hybrids Revealed by Fluorescence in situ Hybridization. *Cytogenetic and Genome Research* 149: 226–235.

Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoël M, Weiss-Schneeweiss H, Leitch AR. 2015. Genomic Repeat Abundances Contain Phylogenetic Signal. *Systematic Biology* 64: 112–126.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*: sys004.

Feschotte C, Pritham EJ. 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics* 41: 331–368.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.

Gong J, Dong J, Liu X, Massana R. 2013. Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist* 164: 369–379.

Hershler R, Sada DW. 2002. Biogeography of Great Basin aquatic snails of the genus Pyrgulopsis. *Smithsonian Contributions to the Earth Sciences* 33: 255–276.

Houston DD, Shiozawa DK, Riddle BR. 2010. Phylogenetic relationships of the western North American cyprinid genus Richardsonius, with an overview of phylogeographic structure. *Molecular Phylogenetics and Evolution* 55: 259–273.

Iwata-Otsubo A, Radke B, Findley S, Abernathy B, Vallejos CE, Jackson SA. 2016. Fluorescence In Situ Hybridization (FISH)-Based Karyotyping Reveals Rapid Evolution of Centromeric and Subtelomeric Repeats in Common Bean (Phaseolus vulgaris) and Relatives. *G3: Genes, Genomes, Genetics* 6: 1013–1022.

James SA, West C, Davey RP, Dicks J, Roberts IN. 2016. Prevalence and Dynamics of Ribosomal DNA Micro-heterogeneity Are Linked to Population History in Two Contrasting Yeast Species. *Scientific Reports* 6: 28555.

Jiang J, Gill BS. 1994. New 18S- 26S ribosomal RNA gene loci: chromosomal landmarks for the evolution of polyploid wheats. *Chromosoma* 103: 179–185.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 110: 462–467.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258: 818–821.

Kanda K, Pflug JM, Sproul JS, Dasenko MA, Maddison DR. 2015. Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. *PLoS One* 10: e0143929.

Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research* 19: 1639–1645.

Larracuente AM. 2017. FISH in Drosophila. *Fluorescence In Situ Hybridization (FISH)*. Springer, 467–472.

Larracuente AM, Ferree PM. 2015. Simple method for fluorescence DNA in situ hybridization to squashed chromosomes. *Journal of visualized experiments: JoVE*.

Lemmon AR, Emme SA, Lemmon EM. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*: sys049.

Lower SS, Johnston JS, Stanger-Hall KF, Hjelmen CE, Hanrahan SJ, Korunes K, Hall D. 2017. Genome Size in North American Fireflies: Substantial Variation Likely Driven by Neutral Processes. *Genome Biology and Evolution* 9: 1499–1512.

Maddison DR. 2008. Systematics of the North American beetle subgenus Pseudoperyphus (Coleoptera: Carabidae: Bembidion) based upon morphological, chromosomal, and molecular data. *Annals of Carnegie Museum* 77: 147–193.

Maddison WP, Maddison DR. 2017. Mesquite: a modular system for evolutionary analysis. Version 3.2. Available: http://mesquiteproject.org.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR. 2010. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* 39: D225–D229.

Martins C, Ferreira IA, Oliveira C, Foresti F, Galetti Jr PM. 2006. A tandemly repetitive centromeric DNA sequence of the fish Hoplias malabaricus (Characiformes: Erythrinidae) is derived from 5S rDNA. *Genetica* 127: 133–141.

McClintock B. 1934. The relation of a particular chromosomal element to the development of the nucleoli in Zea mays. *Zeitschrift für Zellforschung und mikroskopische Anatomie* 21: 294–326.

McCormack JE, Tsai WL, Faircloth BC. 2015. Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*.

McKee BD, Habera L, Vrana JA. 1992. Evidence that intergenic spacer repeats of Drosophila melanogaster rRNA genes function as X-Y pairing sites in male meiosis, and a general model for achiasmatic pairing. *Genetics* 132: 529–544.

Nguyen P, Sahara K, Yoshido A, Marec F. 2010. Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica* 138: 343–354.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Novák P, Neumann P, Macas J. 2010. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11: 378.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29: 792–793.

Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K. 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic Escherichia coli. *Proceedings of the National Academy of Sciences* 106: 17939–17944.

Palacios-Gimenez OM, Cabral-de-Mello DC. 2015. Repetitive DNA chromosomal organization in the cricket Cycloptiloides americanus: a case of the unusual X1X20 sex chromosome system in Orthoptera. *Molecular Genetics and Genomics* 290: 623–631.

Panzera Y, Pita S, Ferreiro M, Ferrandis I, Lages C, Pérez R, Silva A, Guerra M, Panzera F. 2012. High dynamics of rDNA cluster location in kissing bug holocentric chromosomes (Triatominae, Heteroptera). *Cytogenetic and Genome Research* 138: 56–67.

Pérez-García C, Hurtado NS, Morán P, Pasantes JJ. 2014. Evolutionary dynamics of rDNA clusters in chromosomes of five clam species belonging to the family Veneridae (Mollusca, Bivalvia). *BioMed Research International* 2014.

Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, Gray J. 1988. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proceedings of the National Academy of Sciences* 85: 9138–9142.

Qi X, Zhang F, Guan Z, Wang H, Jiang J, Chen S, Chen F. 2015. Localization of 45S and 5S rDNA sites and karyotype of Chrysanthemum and its related genera by fluorescent in situ hybridization. *Biochemical Systematics and Ecology* 62: 164–172.

Raskina O, Barber J, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* 120: 351–357.

Raskina O, Belyayev A, Nevo E. 2004. Quantum speciation in Aegilops: molecular cytogenetic evidence from rDNA cluster variability in natural populations. *Proceedings of the National Academy of Sciences of the United States of America* 101: 14818–14823.

Richter S, Schwarz F, Hering L, Böggemann M, Bleidorn C. 2015. The Utility of Genome Skimming for Phylogenomic Analyses as Demonstrated for Glycerid Relationships (Annelida, Glyceridae). *Genome Biology and Evolution* 7: 3443–3462.

Schwarzacher HG, Wachtler F. 1993. The nucleolus. *Anatomy and Embryology* 188: 515–536.

Sember A, Bohlen J, Šlechtová V, Altmanová M, Symonová R, Ráb P. 2015. Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. *BMC Evolutionary Biology* 15.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *science* 344: 738–742.

Sotero-Caio CG, Volleth M, Hoffmann FG, Scott L, Wichman HA, Yang F, Baker RJ. 2015. Integration of molecular cytogenetics, dated molecular phylogeny, and model-based predictions to understand the extreme chromosome reorganization in the Neotropical genus Tonatia (Chiroptera: Phyllostomidae). *BMC Evolutionary Biology* 15: 220.

Speicher MR, Carter NP. 2005. The new cytogenetics: blurring the boundaries with molecular biology. *Nature Reviews Genetics* 6: 782.

Sproul JS, Houston D, Davis N, Barrington E, Oh SY, Evans RP, Shiozawa DK. 2014. Comparative phylogeography of codistributed aquatic insects in western North America: insights into dispersal and regional patterns of genetic structure. *Freshwater Biology* 59: 2051–2063.

Sproul JS, Houston DD, Nelson CR, Evans RP, Crandall KA, Shiozawa DK. 2015. Climate oscillations, glacial refugia, and dispersal ability: factors influencing the genetic structure of the least salmonfly, Pteronarcella badia (Plecoptera), in Western North America. *BMC Evolutionary Biology* 15: 279.

Sproul JS, Maddison DR. 2017a. Cryptic species in the mountaintops: species delimitation and taxonomy of the Bembidion breve species group (Coleoptera:

Carabidae) aided by genomic architecture of a century-old type specimen. *Zoological Journal of the Linnean Society*: zlx076.

Sproul JS, Maddison DR. 2017b. Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources* 17: 1183–1201.

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Stukenbrock EH. 2013. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytologist* 199: 895–907.

Symonová R, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Rábová M, Ráb P. 2013. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evolutionary Biology* 13: 1.

Symonová R, Ocalewicz K, Kirtiklis L, Delmastro GB, Pelikánová Š, Garcia S, Kovařík A. 2017. Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (Esox sp.). *BMC Genomics* 18: 391.

Symonová R, Sember A, Majtánová Z, Ráb P. 2015. Characterization of fish genomes by GISH and CGH. *Fish Cytogenetic Techniques. Ray-Fin Fishes and Chondrichthyans. CCR Press: Boca Raton*: 118–131.

Vitte C, Bennetzen JL. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proceedings of the National Academy of Sciences* 103: 17638–17643.

Wang W, Ma L, Becher H, Garcia S, Kovarikova A, Leitch IJ, Leitch AR, Kovarik A. 2016. Astonishing 35S rDNA diversity in the gymnosperm species Cycas revoluta Thunb. *Chromosoma* 125: 683–699.

West C, James SA, Davey RP, Dicks J, Roberts IN. 2014. Ribosomal DNA Sequence Heterogeneity Reflects Intraspecies Phylogenies and Predicts Genome Structure in Two Contrasting Yeast Species. *Systematic Biology* 63: 543–554.

White MJD. 1977. *Animal cytology and evolution*. Cambridge University Press, London.

FIGURES

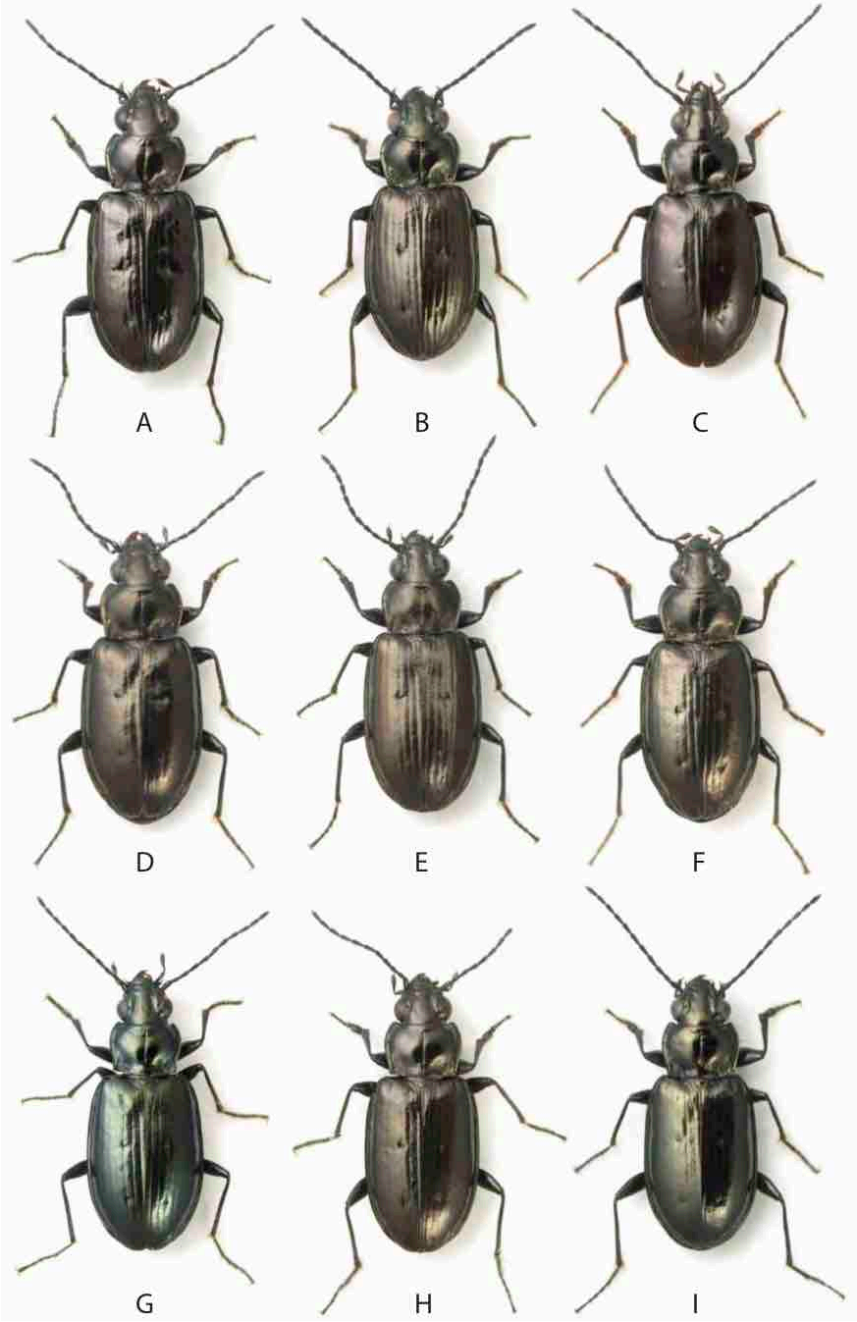


Figure 1. Images of nine species in *breve* species group of *Bembidion* (Carabidae). A, *Bembidion lividulum*. B, *B. breve*. C, *B. testatum*. D, *B. saturatum*. E, *B. vulcanix*. F, *B. geopearlis*. G, *B. oromaia*. H, *B. laxatum*. I, *B. ampliatum*.

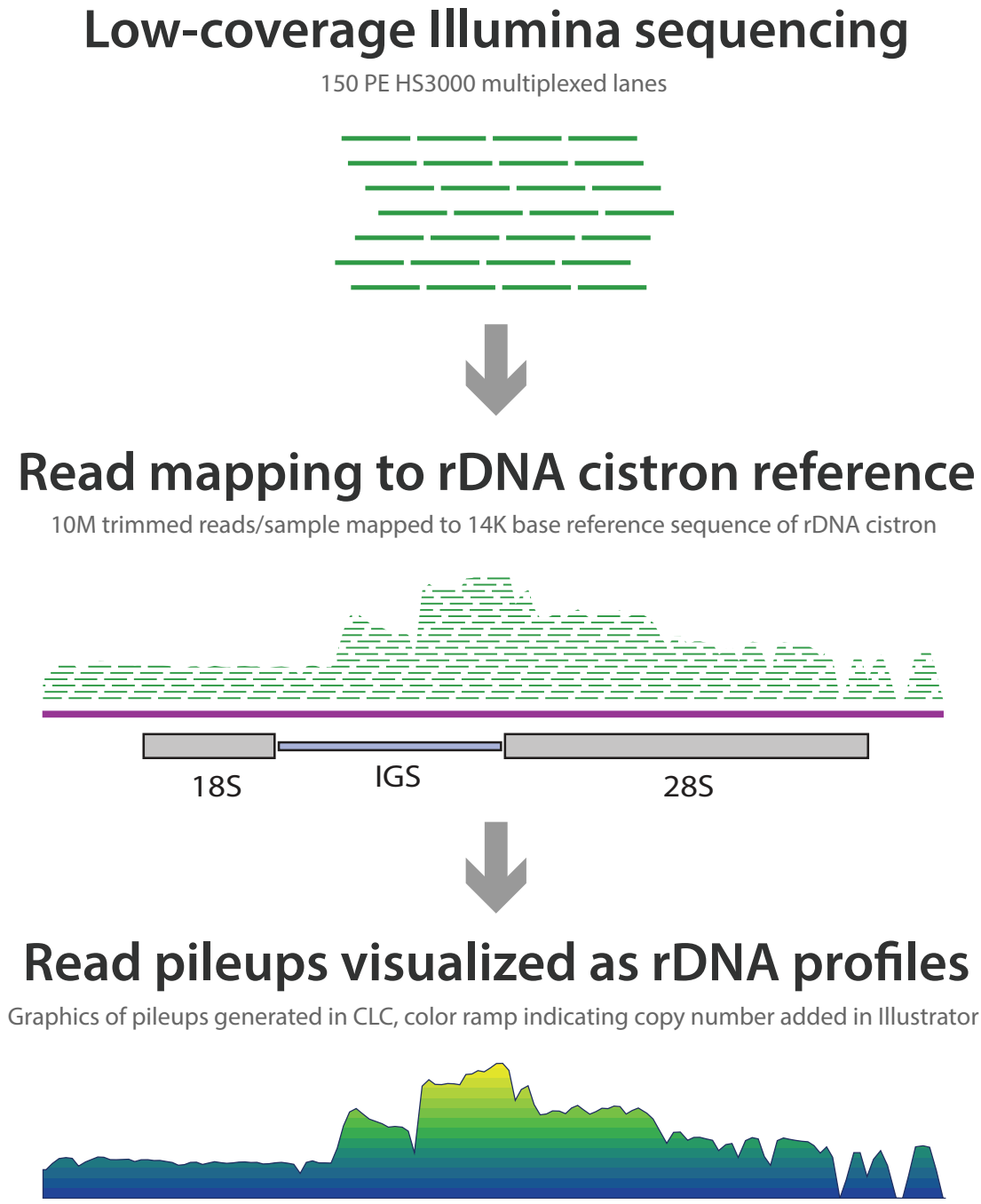


Figure 4.2. Flowchart illustrating the steps used to generate rDNA profiles from short-read sequencing data.

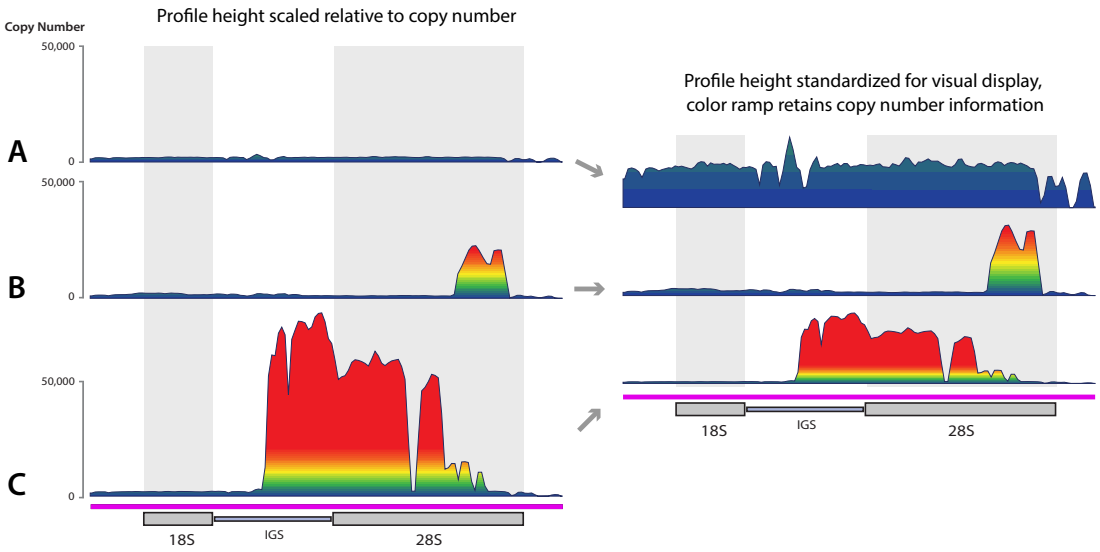


Figure 4.3. A comparison rDNA profiles shown with relative vs fixed scales on the y-axis. Three specimens (*Bembidion laxatum*, 5086; *B. testatum*, 5157; *B. lividulum*, 3486) with profiles on the left scaled relative to 50,000 copies, and the same profiles on the right constrained to the same maximum height. We use the latter scaling strategy throughout the paper to simplify visual display of rDNA profiles. To emphasize differences in copy number that are less apparent in profiles that are scaled to a uniform height, we applied a color ramp to all rDNA profiles when scaled by copy number such that any region with >20K copies = red, >15K copies = orange, >10K copies = yellow, >5K copies = green, and <2.5K copies = blue.

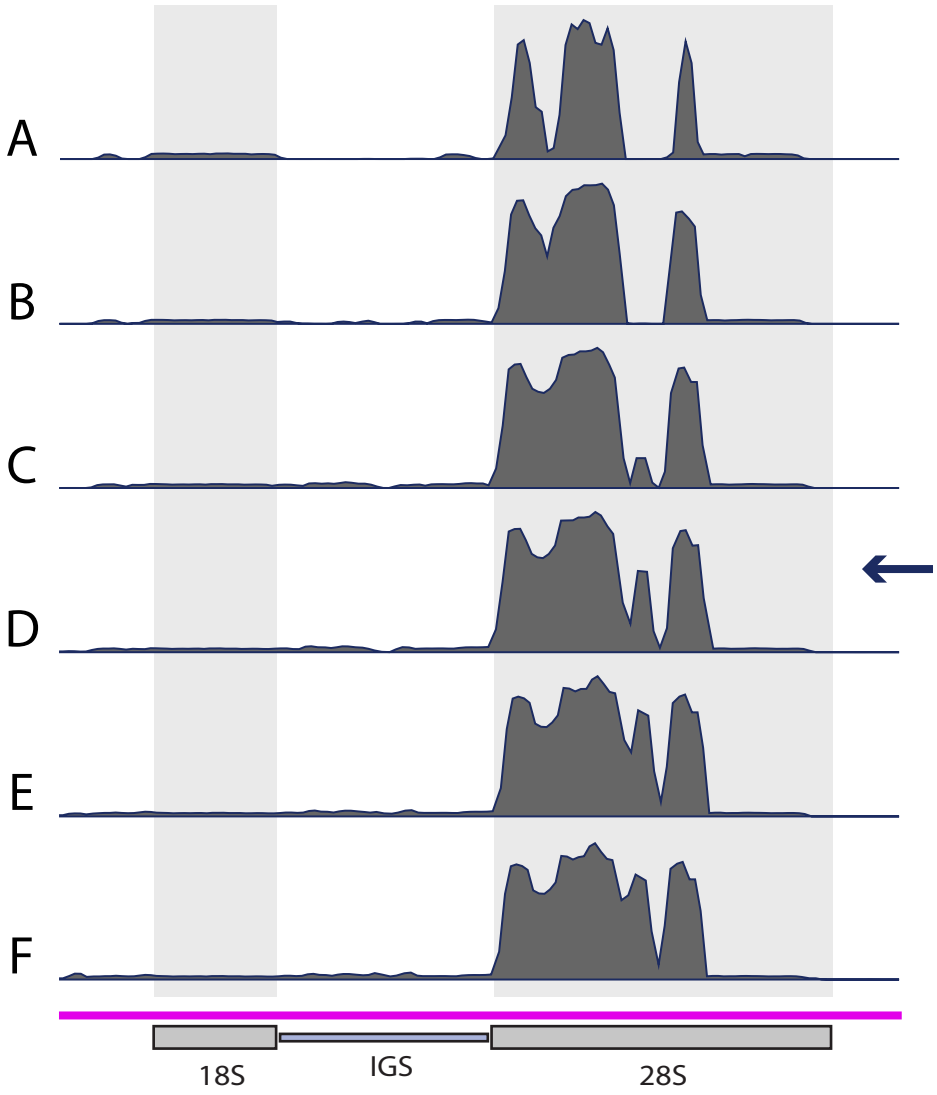


Figure 4.4. Parameter sensitivity analysis. An example of profiles generated for the same specimen (*Bembidion sp.nr.curtulatum "Idaho"* 2145) using read mapping parameters of varying stringency with A being the most stringent and F being the least stringent. Parameter settings used are described in the text. The settings selected for final analysis are indicated by the arrow.

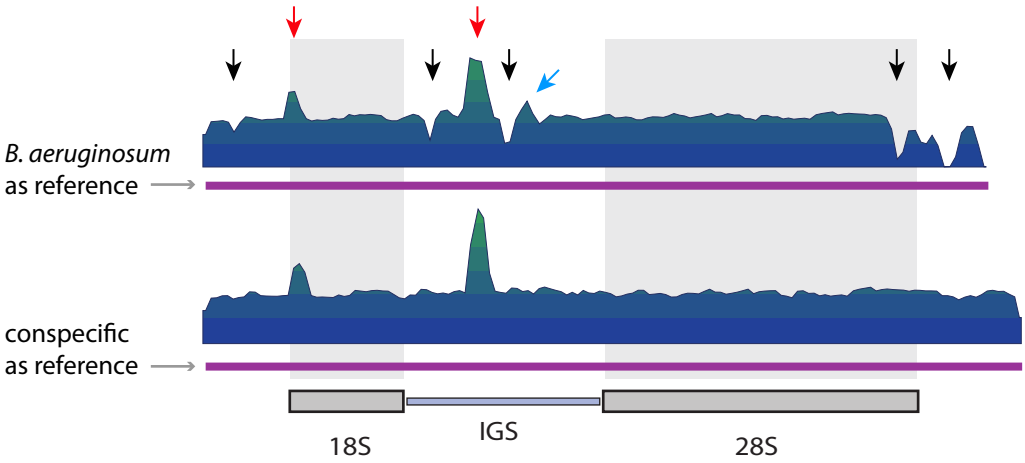


Figure 4.5. Reference sequence sensitivity analysis. An example of rDNA profiles illustrating the difference between a profile generated by mapping reads to a reference sequence derived from a closely related taxon compared to a profile generated by mapping reads to a reference sequence derived from a conspecific sample. Red arrows indicate peaks due to CN variation that are present regardless of reference choice. Black arrows indicate valleys that are artifacts that can be present when mapping reads to a non-conspecific reference, and presumably arise due to sequence divergence (including indels) between the reference sequence and the reads being mapped. The blue arrow indicates a small peak, an artifact of read mapping, that can be appear adjacent to a large valley. The position of rRNA genes relative to rDNA profiles is shown along the bottom for reference.

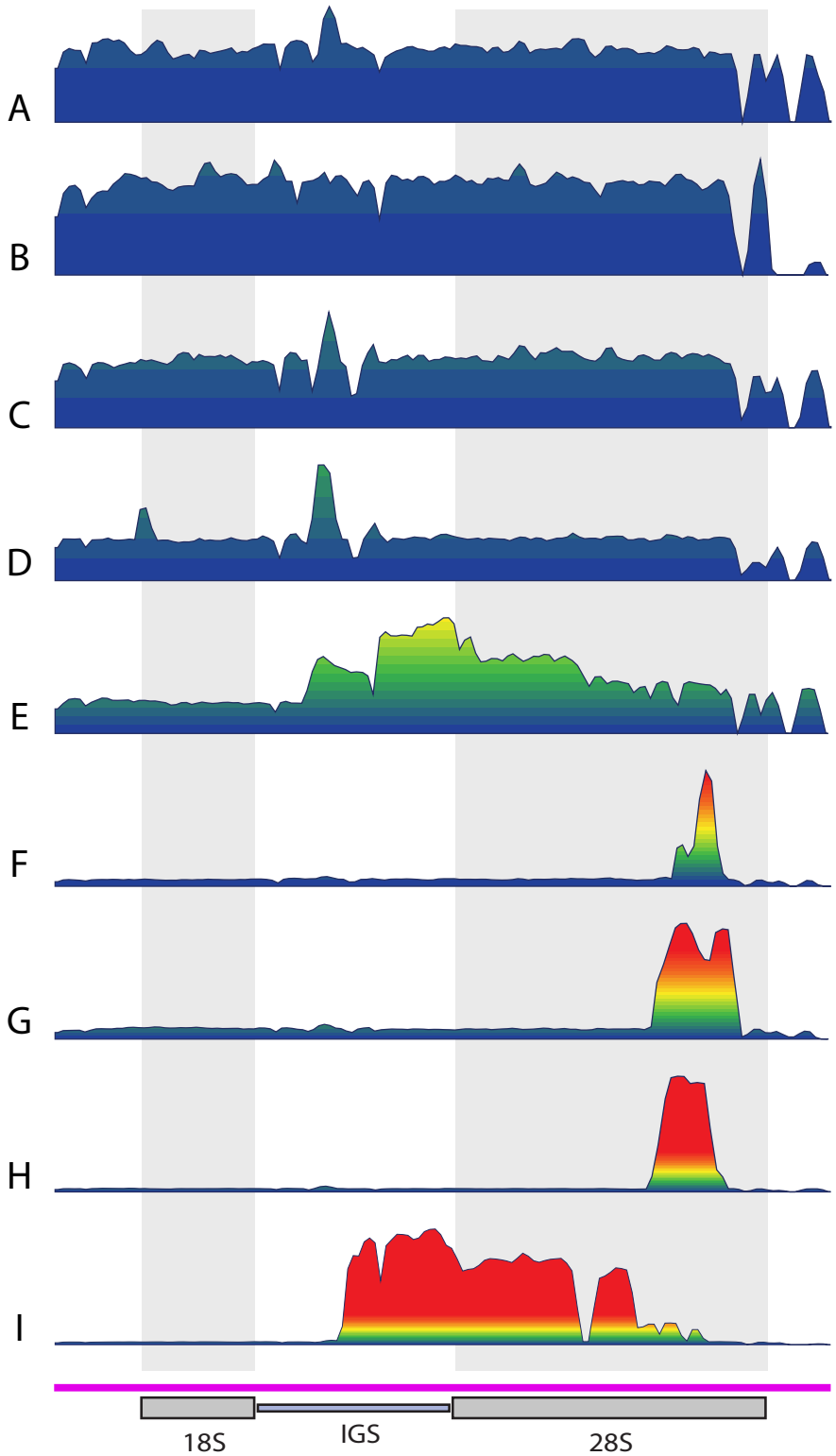
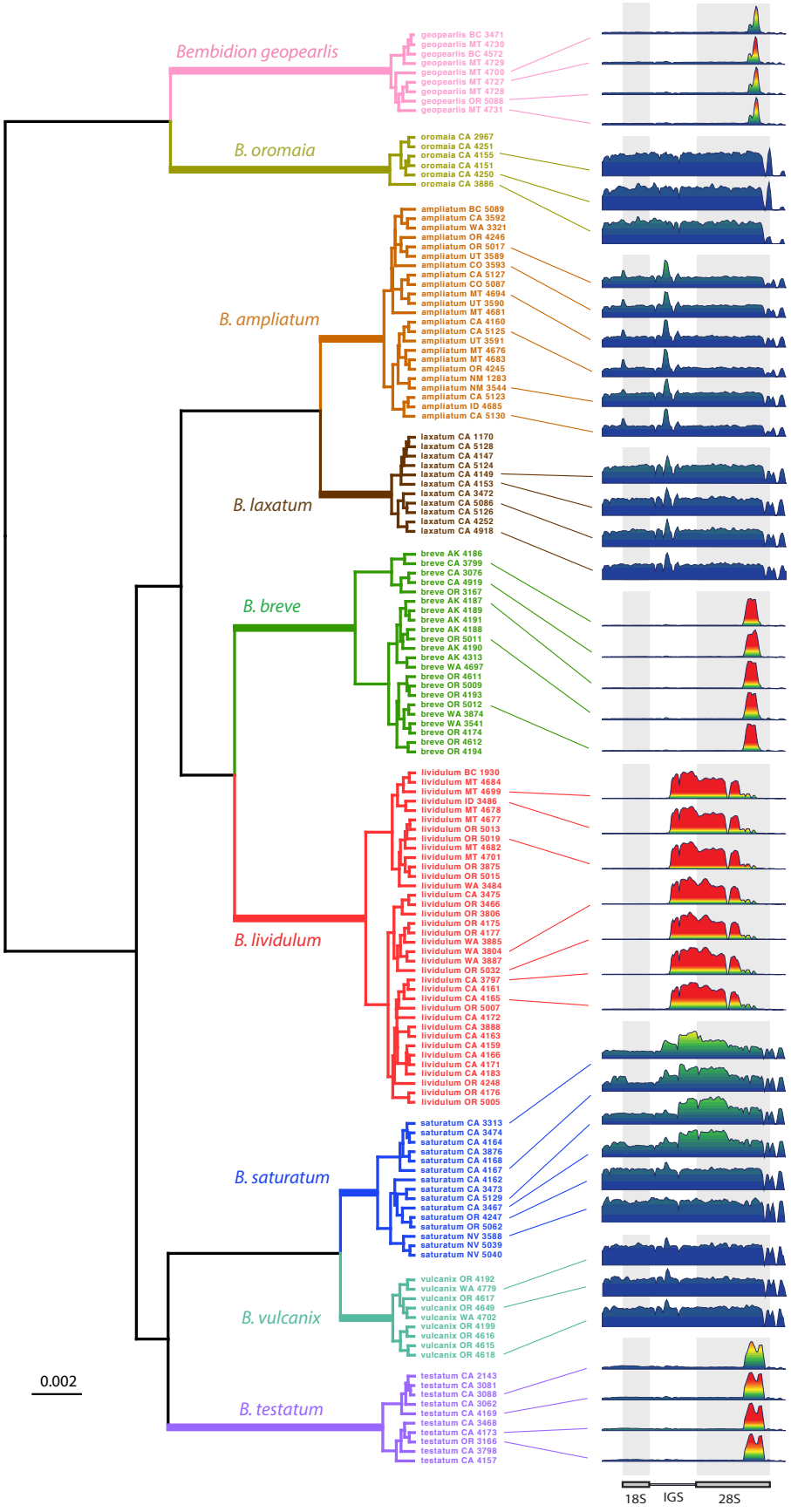


Figure 4.6. rDNA profiles from nine *breve* group species as follows: A, *Bembidion vulcanix* (4649); B, *B. oromaia* (4250); C, *B. laxatum* (5086); D, *B. ampliatum* (4245); E, *B. saturatum* (3313); F, *B. geopearlis* (4731); G, *B. testatum* (4169); H, *B. breve* (4187); I, *B. lividulum* (3486). The position of rRNA genes relative to rDNA profiles is shown along the bottom for reference.

Figure 4.7. The tree used to infer species boundaries of the *breve* species group (adapted from Sproul and Maddison (2017a)) with rDNA profiles. Terminal taxa are colored by inferred species. rDNA profiles for several specimens of each species are shown of the right of the terminals. One to two profiles for some species (e.g., *Bembidion lividulum* and *B. ampliatum*) were excluded to facilitate visual display; however, all profiles not shown corroborate patterns evident in the figure. All profiles generated are shown in Figs. 4.8–16. Branch length shown is proportional to relative divergence with scale bars indicating 0.01 units.


Bembidion geopearlis
geopearlis BC 3471
geopearlis MT 4730
geopearlis BC 4572
geopearlis MT 4729
geopearlis MT 4700
geopearlis MT 4727
geopearlis MT 4728
geopearlis OR 5088
geopearlis MT 4731
B. oromaia
oromaia CA 2967
oromaia CA 4251
oromaia CA 4155
oromaia CA 4151
oromaia CA 4250
oromaia CA 3886
B. ampliatum
ampliatum BC 5089
ampliatum CA 3592
ampliatum WA 3321
ampliatum OR 4246
ampliatum OR 5017
ampliatum UT 3589
ampliatum CO 3593
ampliatum CA 5127
ampliatum CO 5087
ampliatum MT 4694
ampliatum UT 3590
ampliatum MT 4681
ampliatum CA 4160
ampliatum CA 5125
ampliatum UT 3591
ampliatum MT 4676
ampliatum MT 4683
ampliatum OR 4245
ampliatum NM 1283
ampliatum NM 3544
ampliatum CA 5123
ampliatum ID 4685
ampliatum CA 5130
B. laxatum
laxatum CA 1170
laxatum CA 5128
laxatum CA 4147
laxatum CA 5124
laxatum CA 4149
laxatum CA 4153
laxatum CA 3472
laxatum CA 5086
laxatum CA 5126
laxatum CA 4252
laxatum CA 4918
B. breve
breve AK 4186
breve CA 3799
breve CA 3076
breve CA 4919
breve OR 3167
breve AK 4187
breve AK 4189
breve AK 4191
breve AK 4188
breve OR 5011
breve AK 4190
breve AK 4313
breve WA 4697
breve OR 4611
breve OR 5009
breve OR 4193
breve OR 5012
breve WA 3874
breve WA 3541
breve OR 4174
breve OR 4612
breve OR 4194
B. lividulum
lividulum BC 1930
lividulum MT 4684
lividulum MT 4699
lividulum ID 3486
lividulum MT 4678
lividulum MT 4677
lividulum OR 5013
lividulum OR 5019
lividulum MT 4682
lividulum MT 4701
lividulum OR 3875
lividulum OR 5015
lividulum WA 3484
lividulum CA 3475
lividulum OR 3466
lividulum OR 3806
lividulum OR 4175
lividulum OR 4177
lividulum WA 3885
lividulum WA 3804
lividulum WA 3887
lividulum OR 5032
lividulum CA 3797
lividulum CA 4161
lividulum CA 4165
lividulum OR 5007
lividulum CA 4172
lividulum CA 3888
lividulum CA 4163
lividulum CA 4159
lividulum CA 4166
lividulum CA 4171
lividulum CA 4183
lividulum OR 4248
lividulum OR 4176
lividulum OR 5005
B. saturatum
saturatum CA 3313
saturatum CA 3474
saturatum CA 4164
saturatum CA 3876
saturatum CA 4168
saturatum CA 4167
saturatum CA 4162
saturatum CA 3473
saturatum CA 5129
saturatum CA 3467
saturatum OR 4247
saturatum OR 5062
saturatum NV 3588
saturatum NV 5039
saturatum NV 5040
B. vulcanix
vulcanix OR 4192
vulcanix WA 4779
vulcanix OR 4617
vulcanix OR 4649
vulcanix WA 4702
vulcanix OR 4199
vulcanix OR 4616
vulcanix OR 4615
vulcanix OR 4618
B. testatum
testatum CA 2143
testatum CA 3081
testatum CA 3088
testatum CA 3062
testatum CA 4169
testatum CA 3468
testatum CA 4173
testatum OR 3166
testatum CA 3798
testatum CA 4157
0.002
18S
IGS
28S

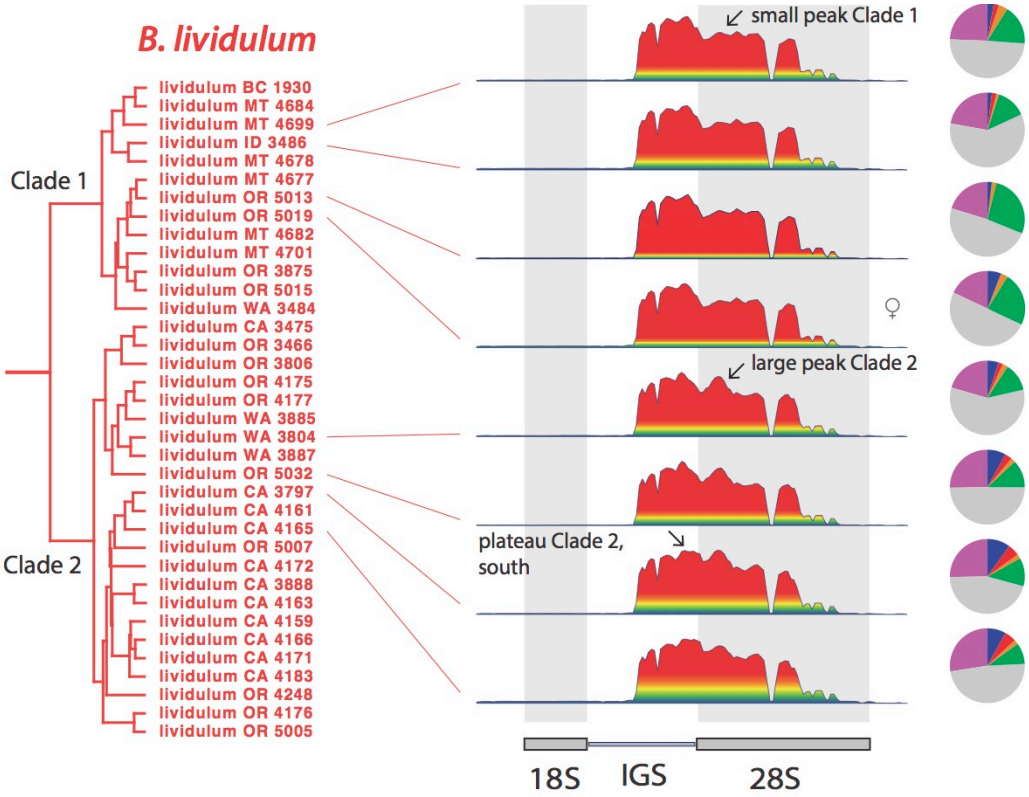


Figure 4.8. The *Bembidion lividulum* clade taken from the species tree in Fig. 4.7 with rDNA profiles. Features in rDNA profiles that appear to show phylogenetic signal are indicated with arrows and text. Profiles generated from female specimens are indicated by the female symbol. Pie charts indicate the fraction of clusters in the top 100 clusters that belong to each of six categories of repetitive DNA: Class I TEs in blue, Class II TEs in red, simple repeats in green, rDNA in orange, and unknown with rDNA hits in pink.

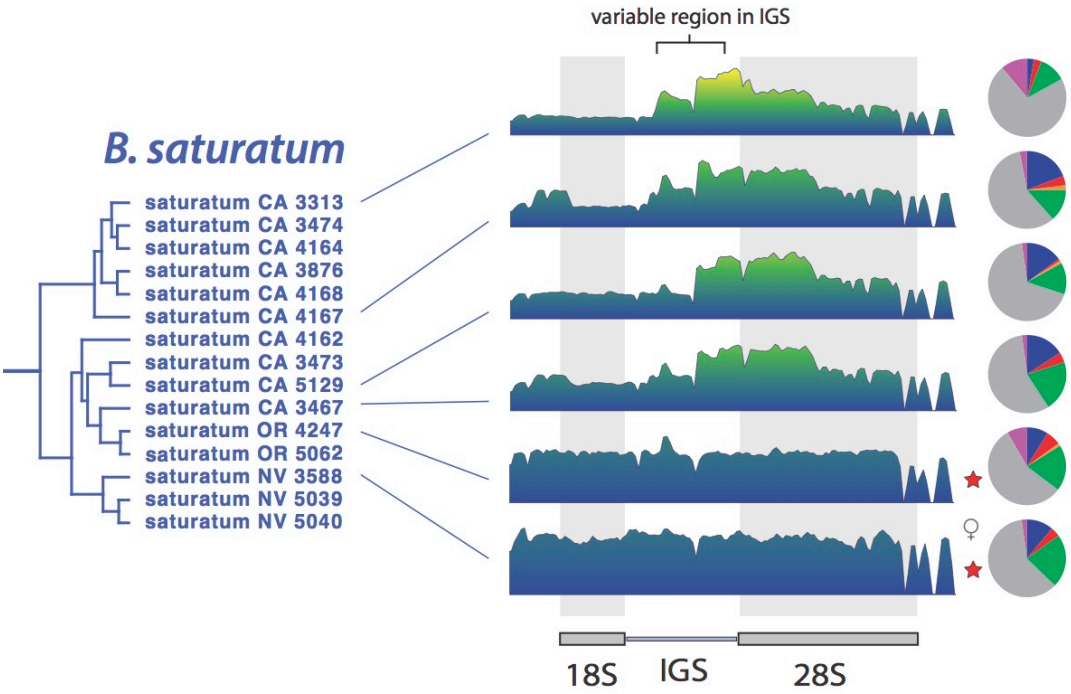


Figure 4.9. The *Bembidion saturatum* clade taken from the species tree in Fig. 4.7 with rDNA profiles. Profiles generated from female specimens are indicated by the female symbol. Pie charts show the fraction of repetitive DNA in each of six categories explained further in the Fig. 4.8 caption.

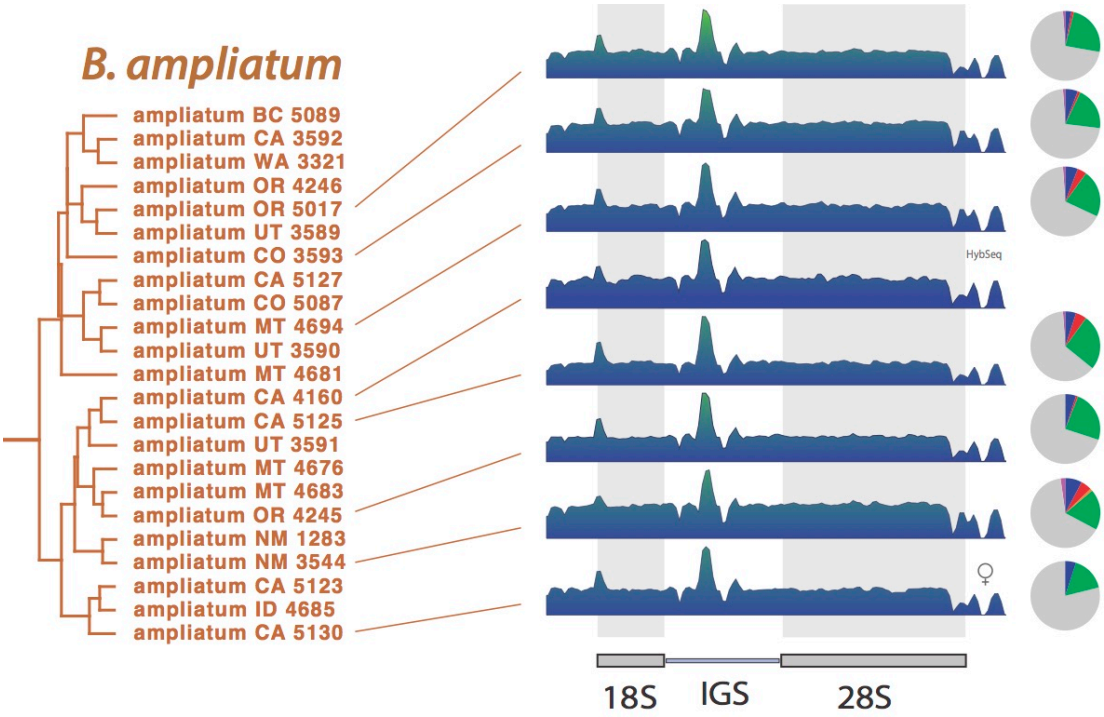


Figure 4.10. The *Bembidion ampliatum* clade taken from the species tree in Fig. 4.7 with rDNA profiles. Profiles generated from female specimens are indicated by the female symbol, and those generated from hybrid enrichment sequencing are annotated with "HybSeq". Pie charts show the fraction of repetitive DNA in each of six categories explained further in the Fig. 4.8 caption.

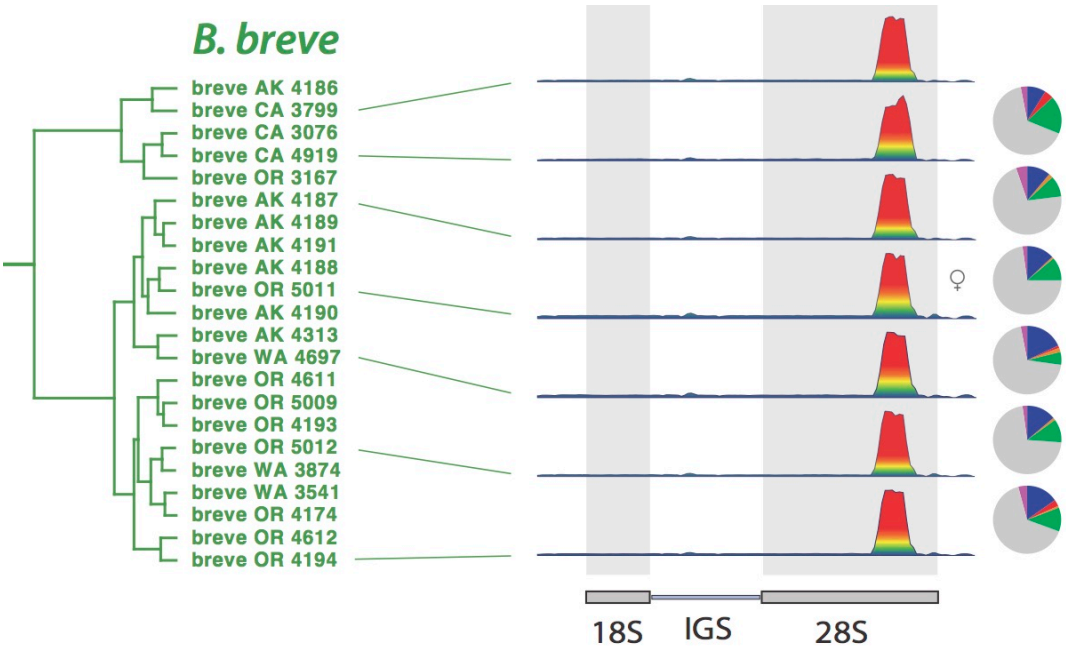


Figure 4.11. The *Bembidion breve* clade taken from the species tree in Fig. 4.7 with rDNA profiles. Profiles generated from female specimens are indicated by the female symbol. Pie charts show the fraction of repetitive DNA in each of six categories explained further in the Fig. 4.8 caption.

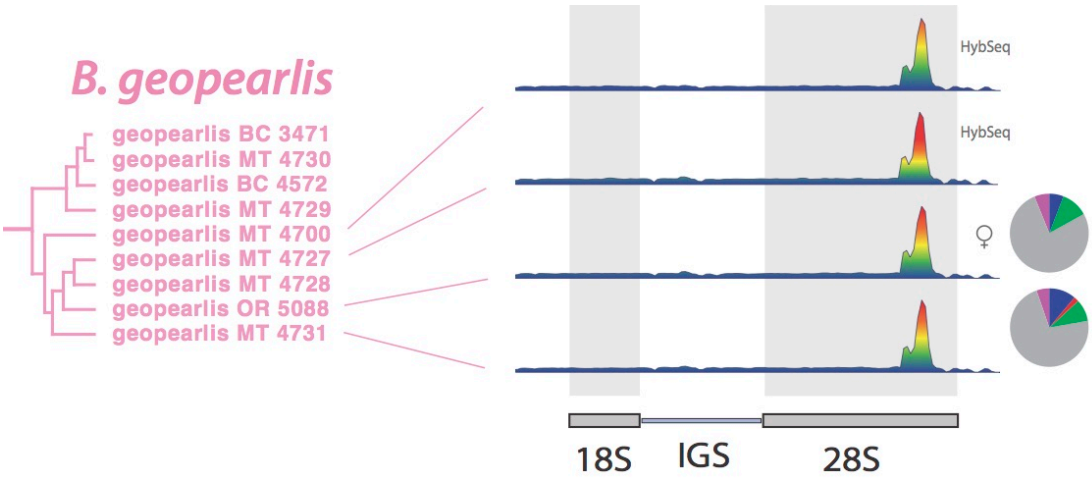


Figure 4.12. The *Bembidion geopearlis* clade taken from the species tree in Fig. 4.7 with rDNA profiles. Profiles generated from female specimens are indicated by the female symbol, and those generated from hybrid enrichment sequencing are annotated with "HybSeq". Pie charts show the fraction of repetitive DNA in each of six categories explained further in the Fig. 4.8 caption.

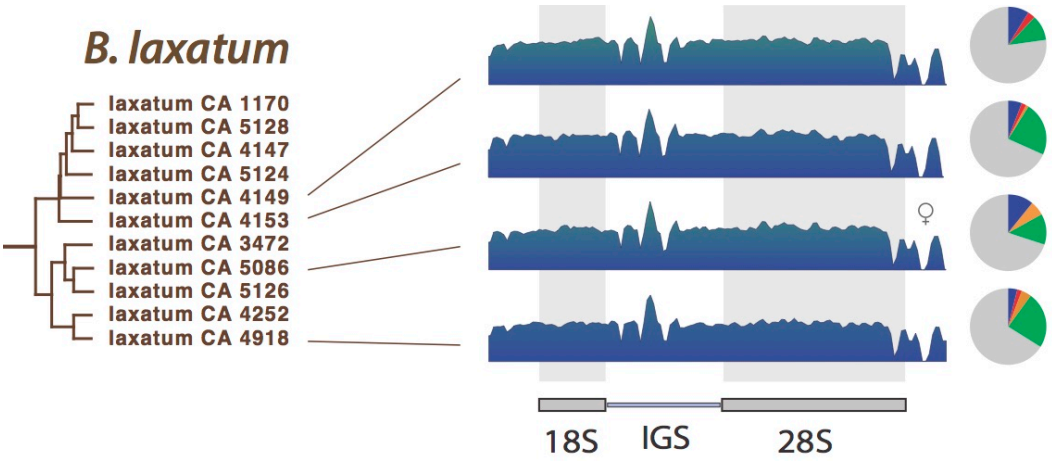


Figure 4.13. The *Bembidion laxatum* clade taken from the species tree in Fig. 4.7 with rDNA profiles. Profiles generated from female specimens are indicated by the female symbol. Pie charts show the fraction of repetitive DNA in each of six categories explained further in the Fig. 4.8 caption.

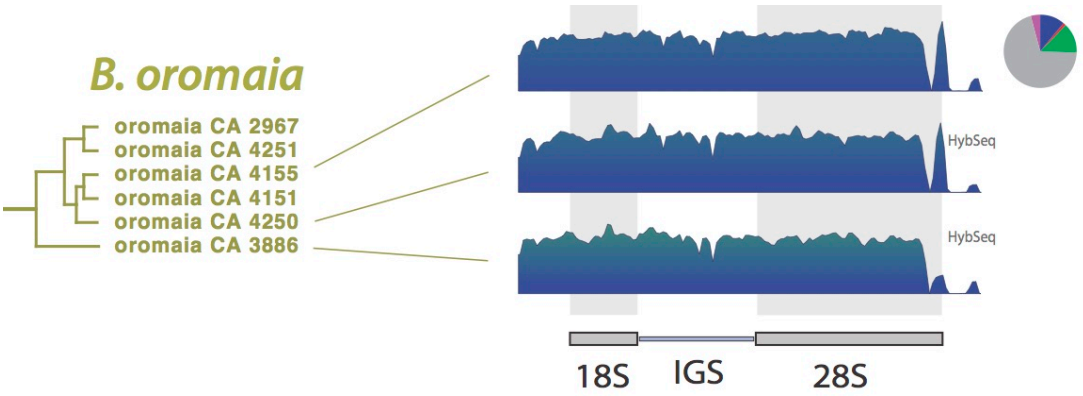


Figure 4.14. The *Bembidion oromaia* clade taken from the species tree in Fig. 4.7 with rDNA profiles. Profiles generated hybrid enrichment sequencing are annotated with "HybSeq". Pie charts show the fraction of repetitive DNA in each of six categories explained further in the Fig. 4.8 caption.

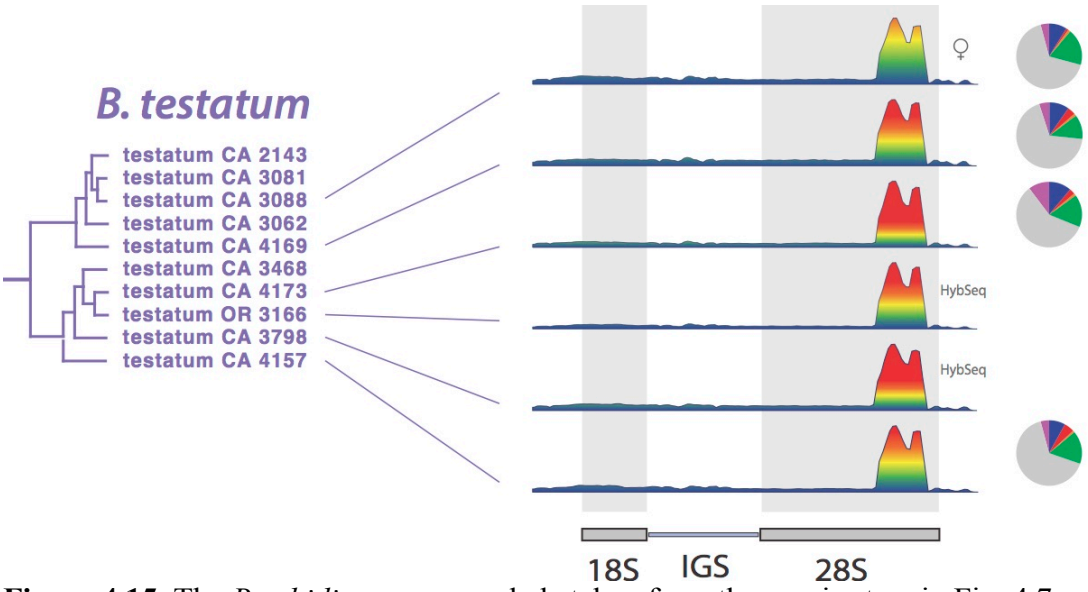


Figure 4.15. The *Bembidion testatum* clade taken from the species tree in Fig. 4.7 with rDNA profiles. Profiles generated from female specimens are indicated by the female symbol, and those generated from hybrid enrichment sequencing are annotated with "HybSeq". Pie charts show the fraction of repetitive DNA in each of six categories explained further in the Fig. 4.8 caption.

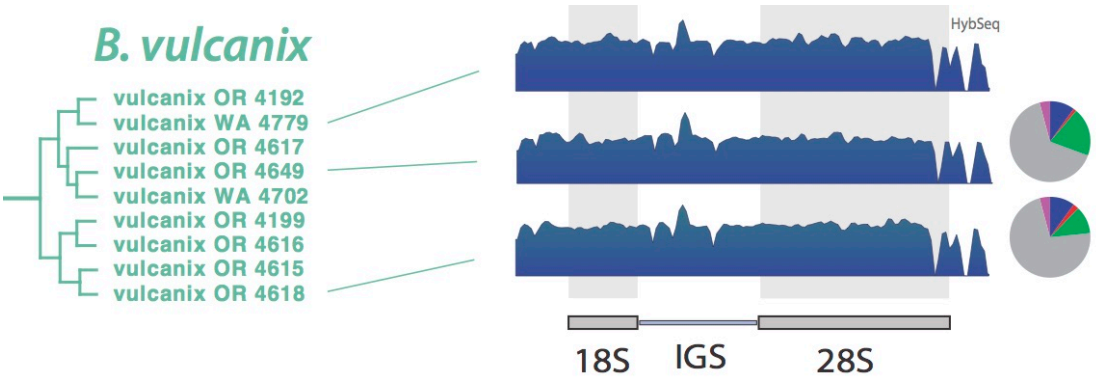


Figure 4.16. The *Bembidion vulcanix* clade taken from the species tree in Figure? with rDNA profiles. Profiles generated hybrid enrichment sequencing are annotated with "HybSeq". Pie charts show the fraction of repetitive DNA in each of six categories explained further in the Fig. 4.8 caption.

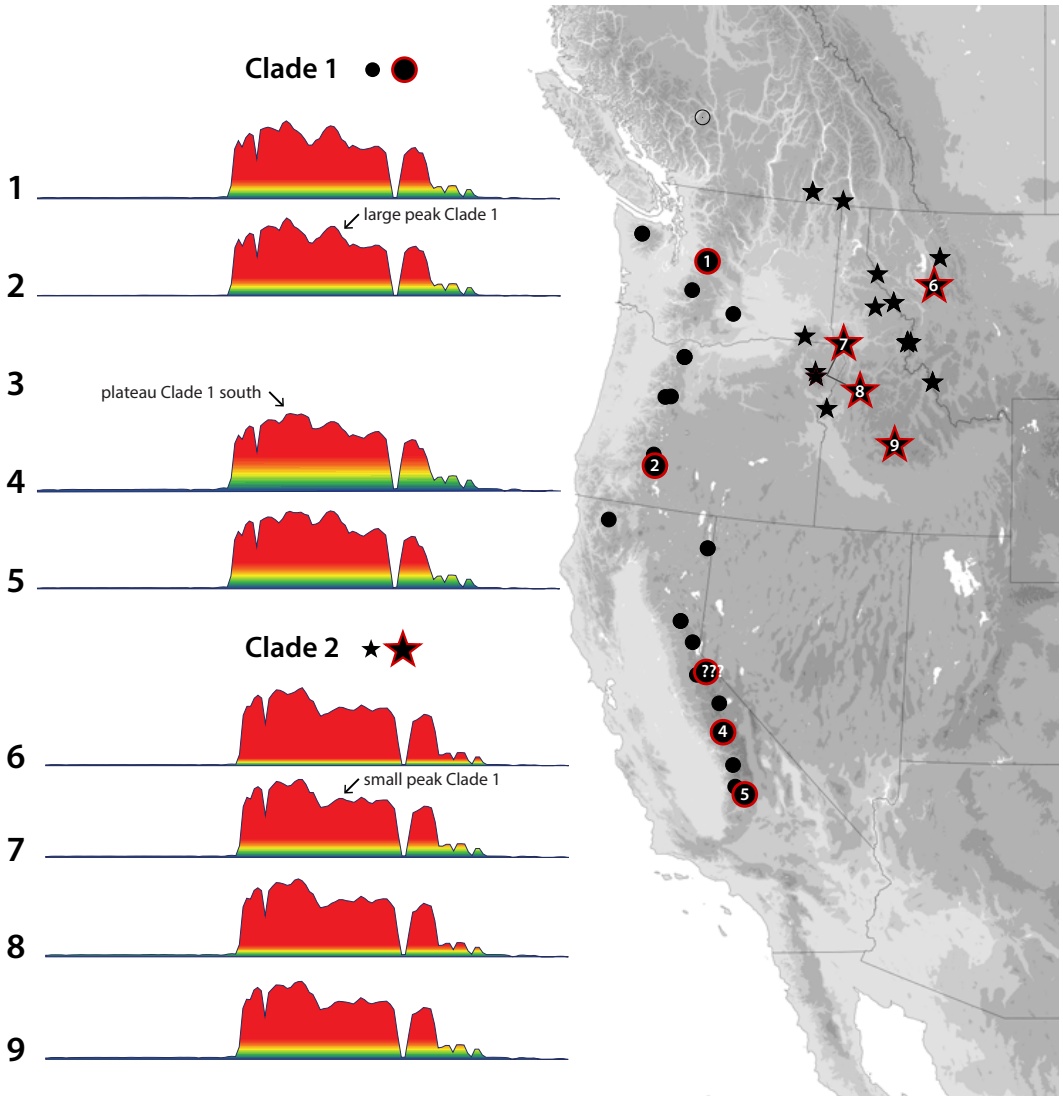


Figure 4.17. Species distribution map for *Bembidion lividulum* showing the geographic sampling of rDNA profiles. Localities from with specimens belonging to Clade 1 are shown by circles, while localities with specimens belonging to Clade 1 are shown by stars. Circles and stars outlined in red indicate localities from which we obtained rDNA profiles, with numbers in shapes corresponding to rDNA profiles shown on the left of the figure. Features in rDNA profiles that appear to show geographic signal are indicated with arrows and text.

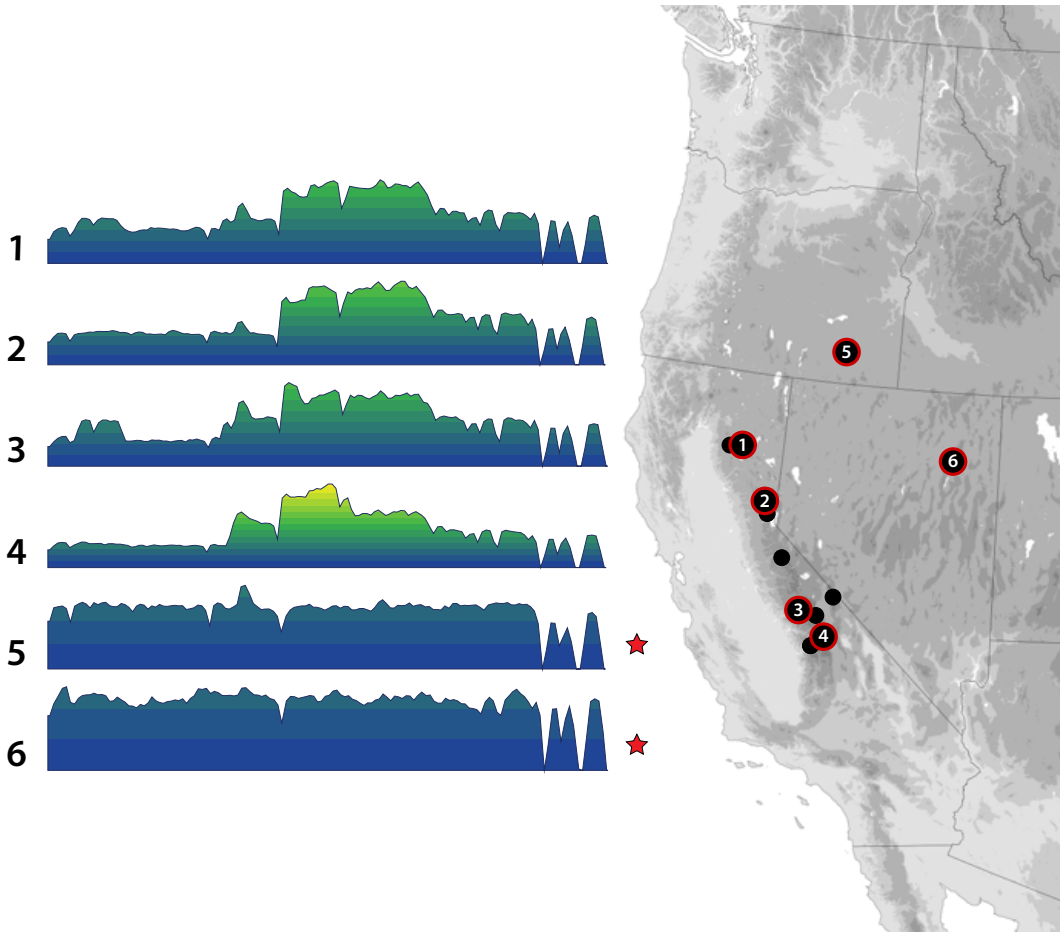


Figure 4.18. Species distribution map for *Bembidion saturatum* showing the geographic sampling of rDNA profiles. Confirmed localities of the species are shown by either small black circles, larger red-outlined black circles, or black stars with a red outline. Red-outlined circles and stars indicate those localities from which we obtained rDNA profiles. Localities containing specimens identified by Sproul and Maddison (2017a) to be morphologically distinctive are shown by red-outlined stars. Numbers in circles and stars correspond to rDNA profiles shown on the left of the figure.

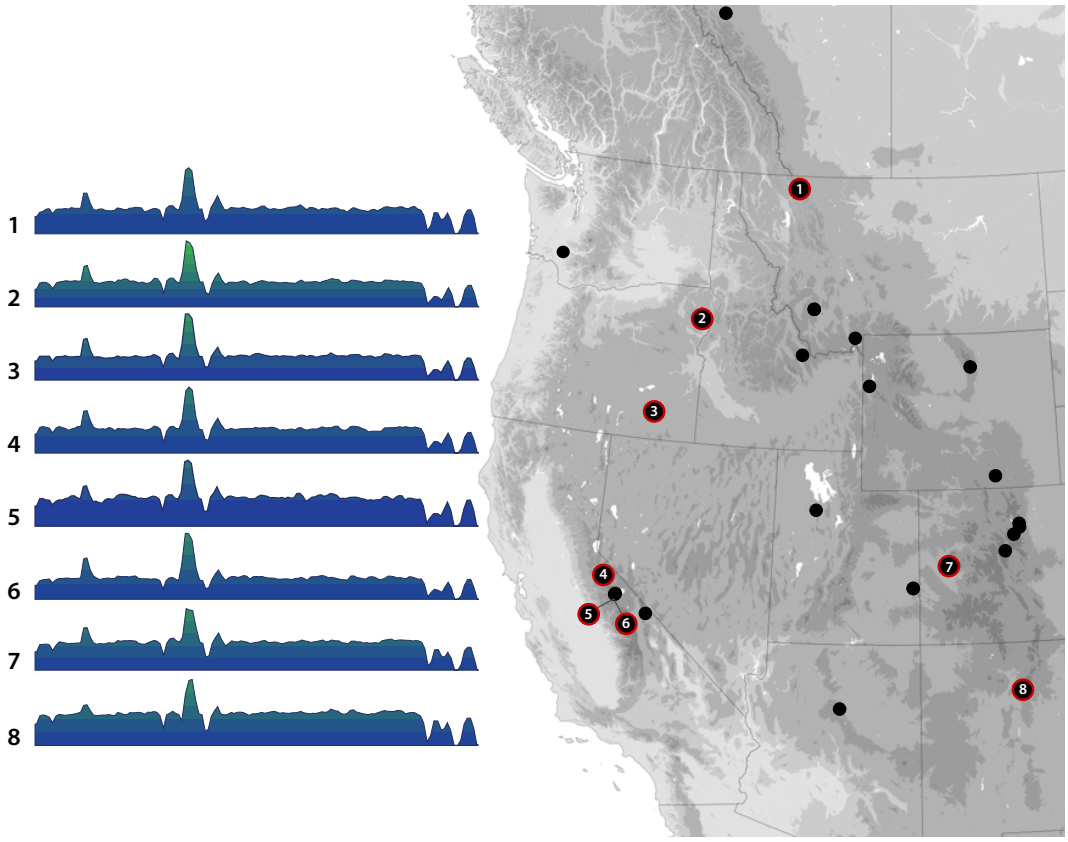


Figure 4.19. Species distribution map for *Bembidion ampliatum* showing the geographic sampling of rDNA profiles. Confirmed localities of the species are shown by either small black circles, or larger red-outlined black circles, the latter indicating those localities from which we obtained rDNA profiles. Numbers in red-outlined circles correspond to rDNA profiles shown on the left of the figure.

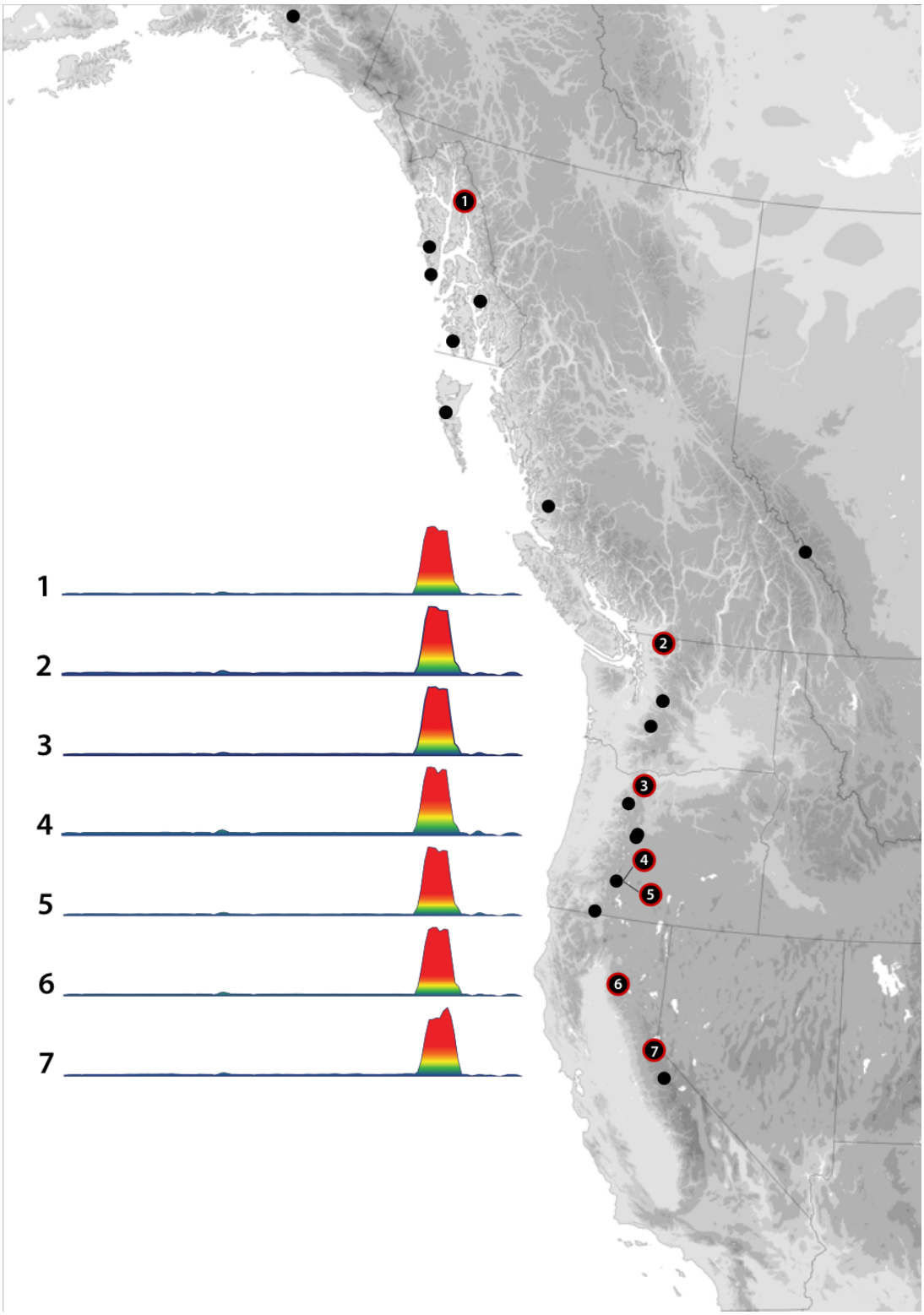


Figure 4.20. Species distribution map for *Bembidion breve* showing the geographic sampling of rDNA profiles. See the figure caption for Fig. 4.19 or additional explanation.

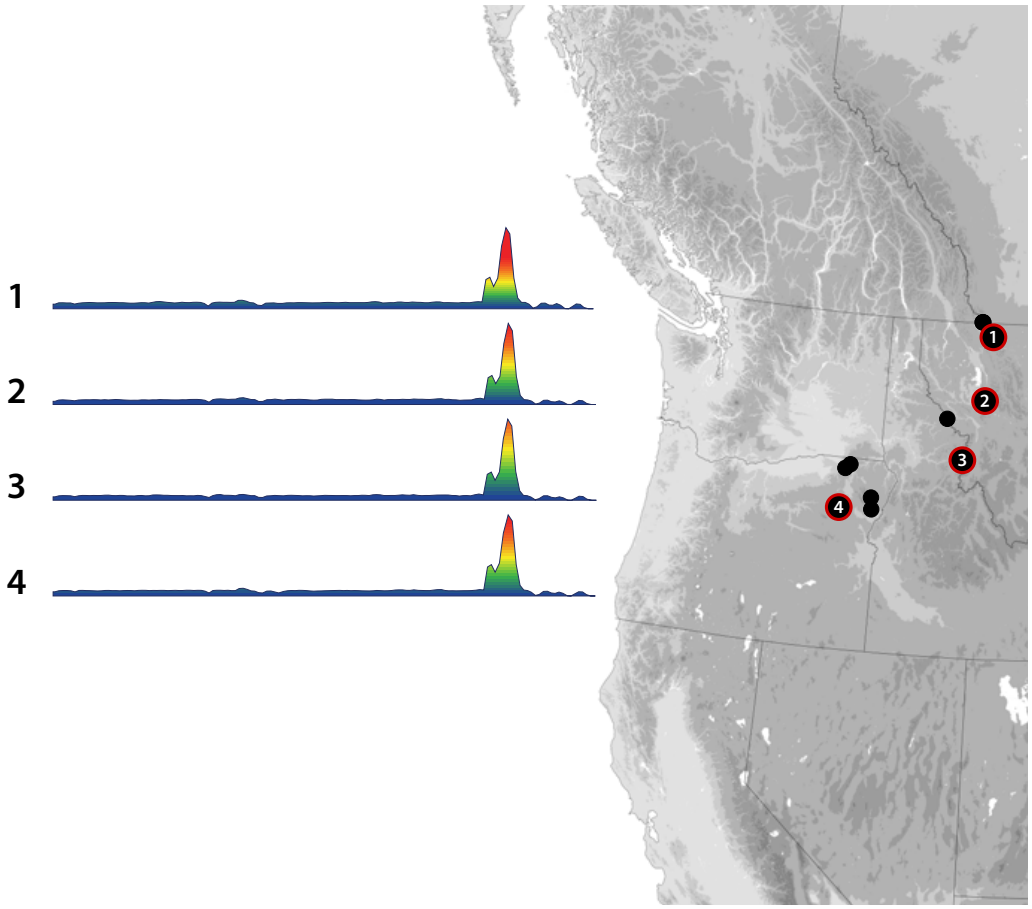


Figure 4.21. Species distribution map for *Bembidion geopearlis* showing the geographic sampling of rDNA profiles. See the figure caption for Fig. 4.19 for additional explanation.

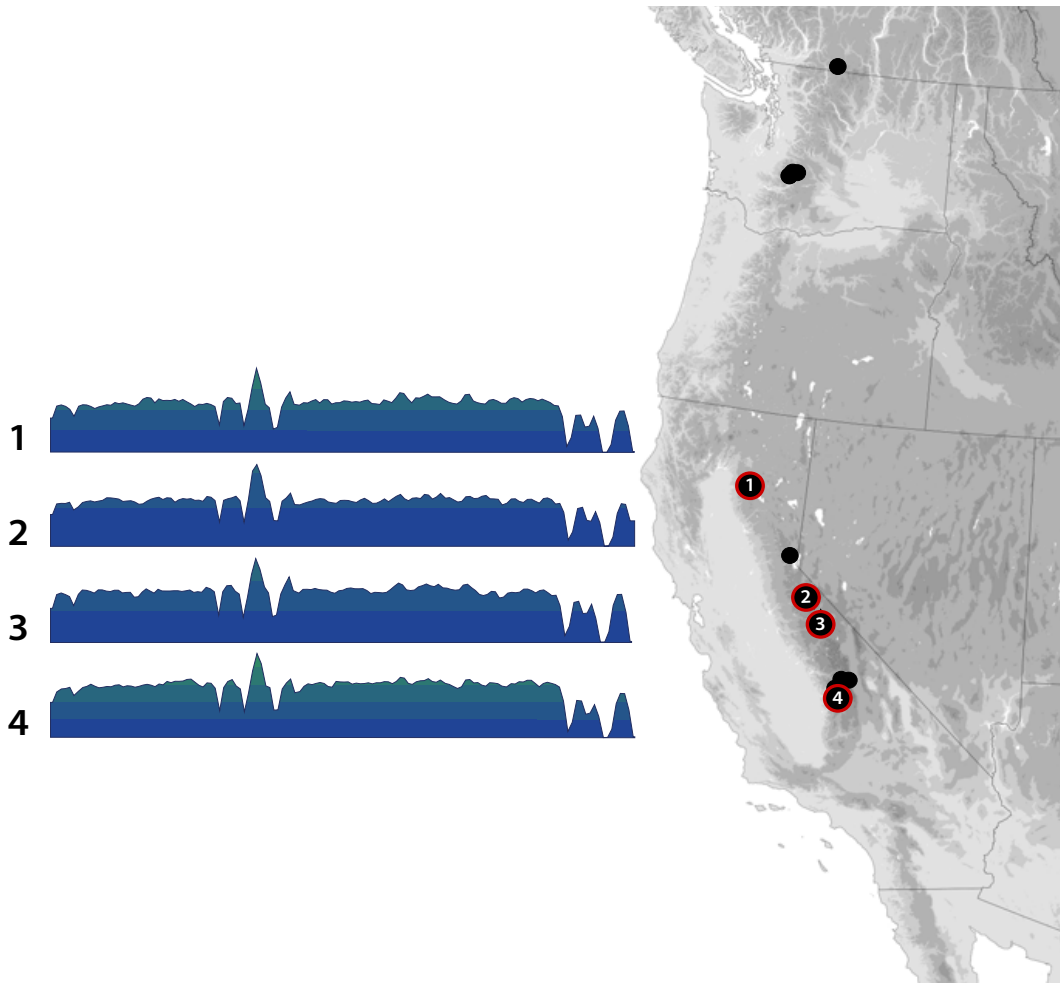


Figure 4.22. Species distribution map for *Bembidion laxatum* showing the geographic sampling of rDNA profiles. See the figure caption for Fig. 4.19 for additional explanation.

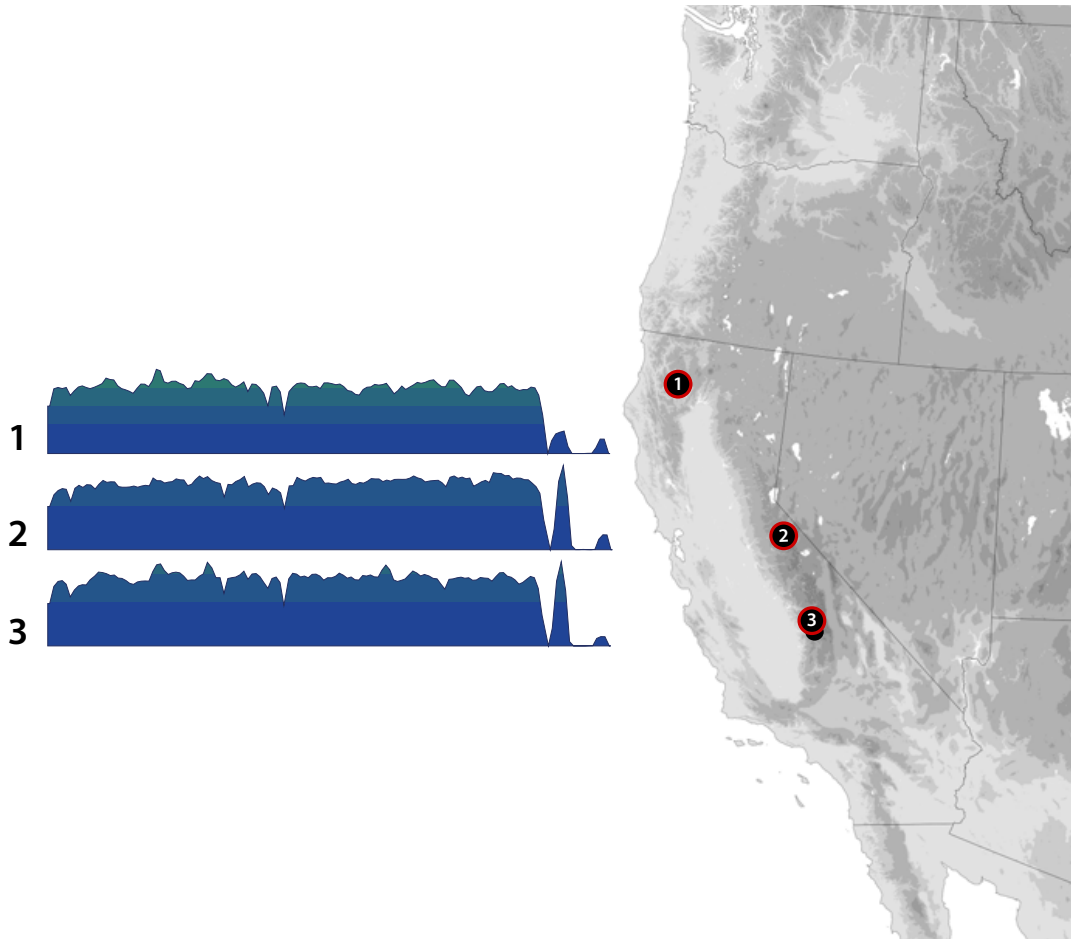


Figure 4.23. Species distribution map for *Bembidion oromaia* showing the geographic sampling of rDNA profiles. See the figure caption for Fig. 4.19 for additional explanation.

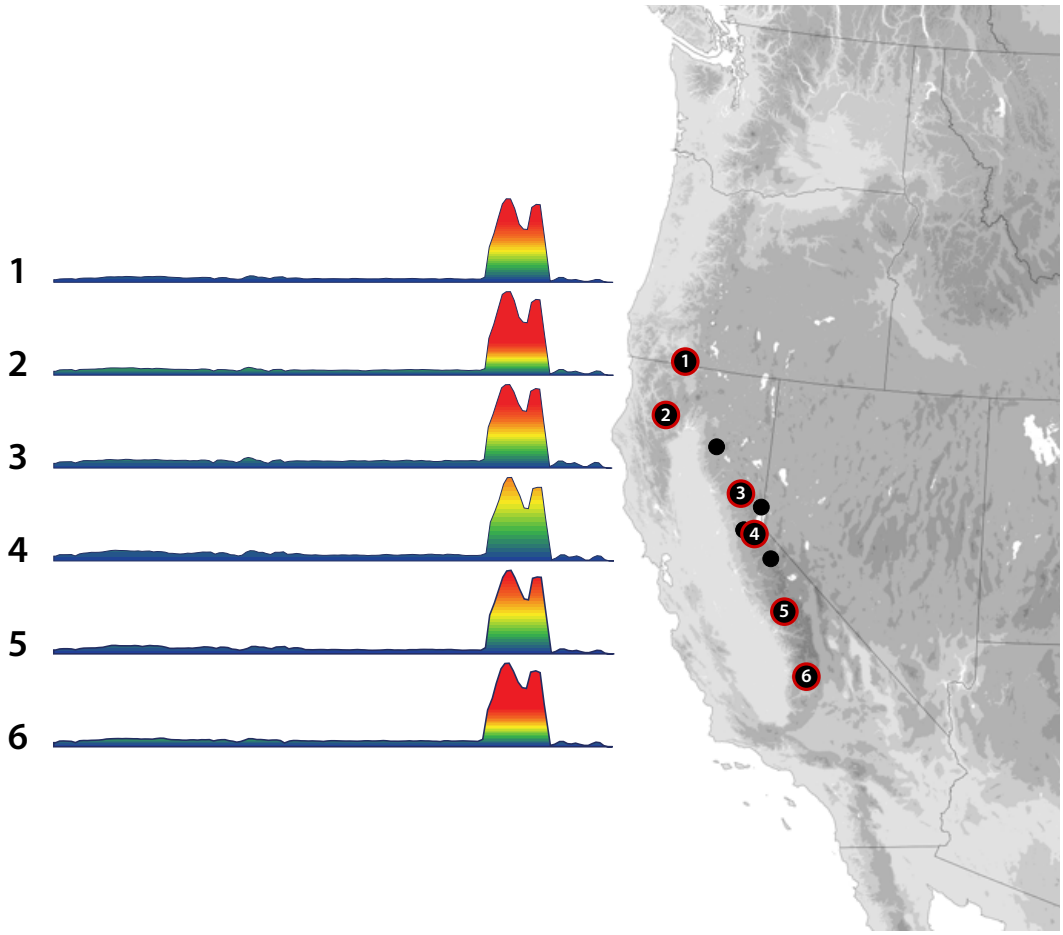


Figure 4.24. Species distribution map for *Bembidion testatum* showing the geographic sampling of rDNA profiles. See the figure caption for Fig. 4.19 for additional explanation.

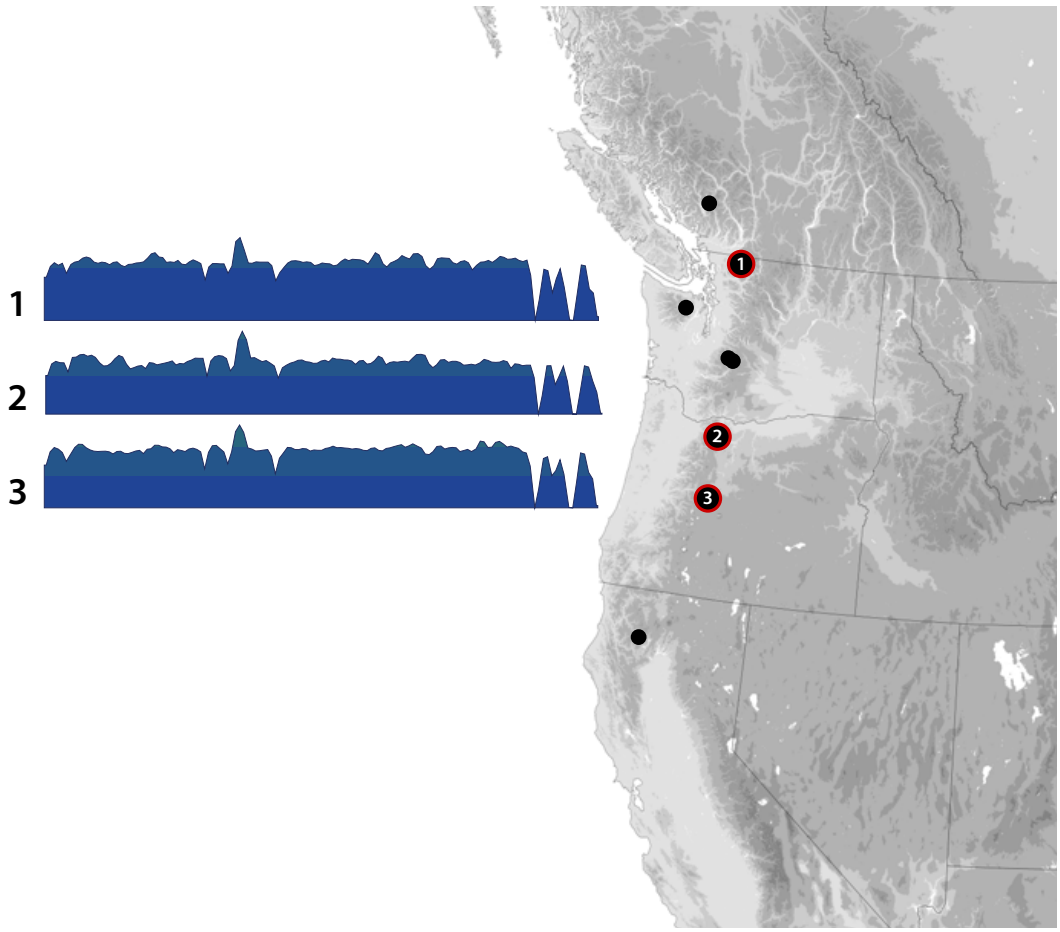


Figure 4.25. Species distribution map for *Bembidion vulcanix* showing the geographic sampling of rDNA profiles. See the figure caption for Fig. 4.19 for additional explanation.

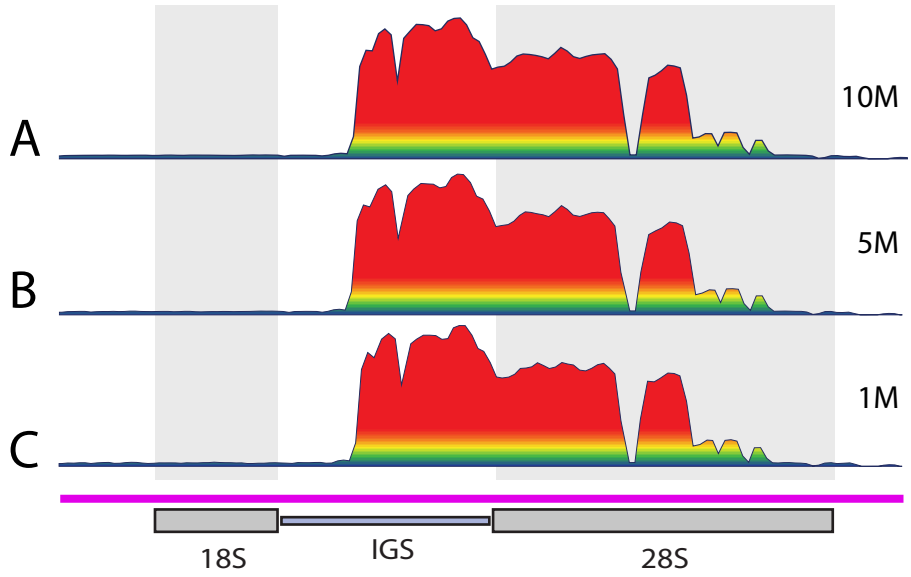


Figure 4.26. rDNA profiles obtained from the same specimen (*B. lividulum* 3486) using 10 million (A), 5 million (B), and 1 million reads (C).

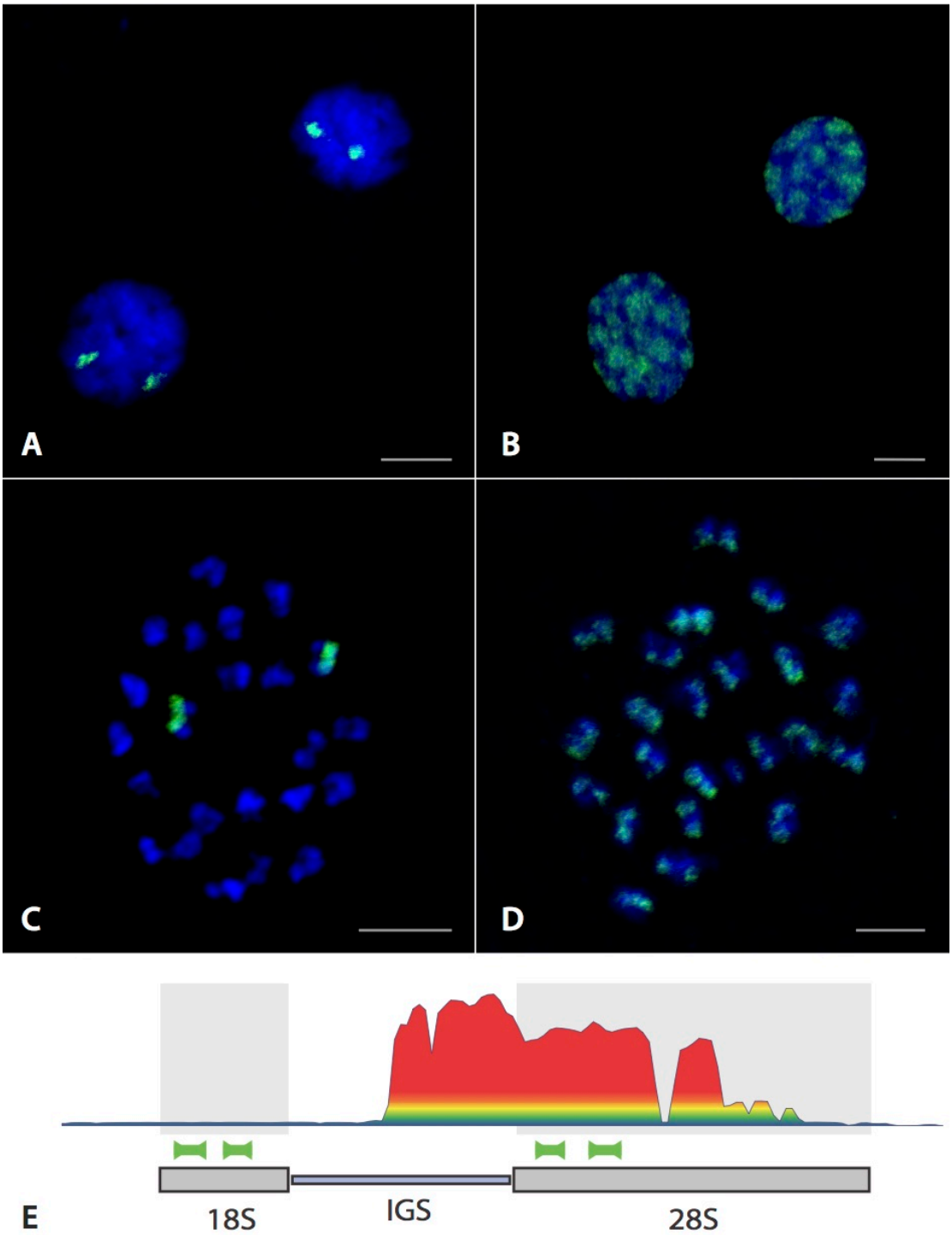


Figure 4.27. FISH signals obtained by cytogenetic mapping of rDNA in *B. lividulum*. A, FISH signals resulting from hybridization of 18S probes to interphase nuclei. B, FISH signals resulting from hybridization of 28S probes to interphase nuclei. C, FISH signals resulting from hybridization of 18S probes to condensed chromosomes. D, FISH signals resulting from hybridization of 28S probes to condensed chromosomes. E, a rDNA profile for *B. lividulum* indicating the location of 18S and 28S FISH probes with green boxes.

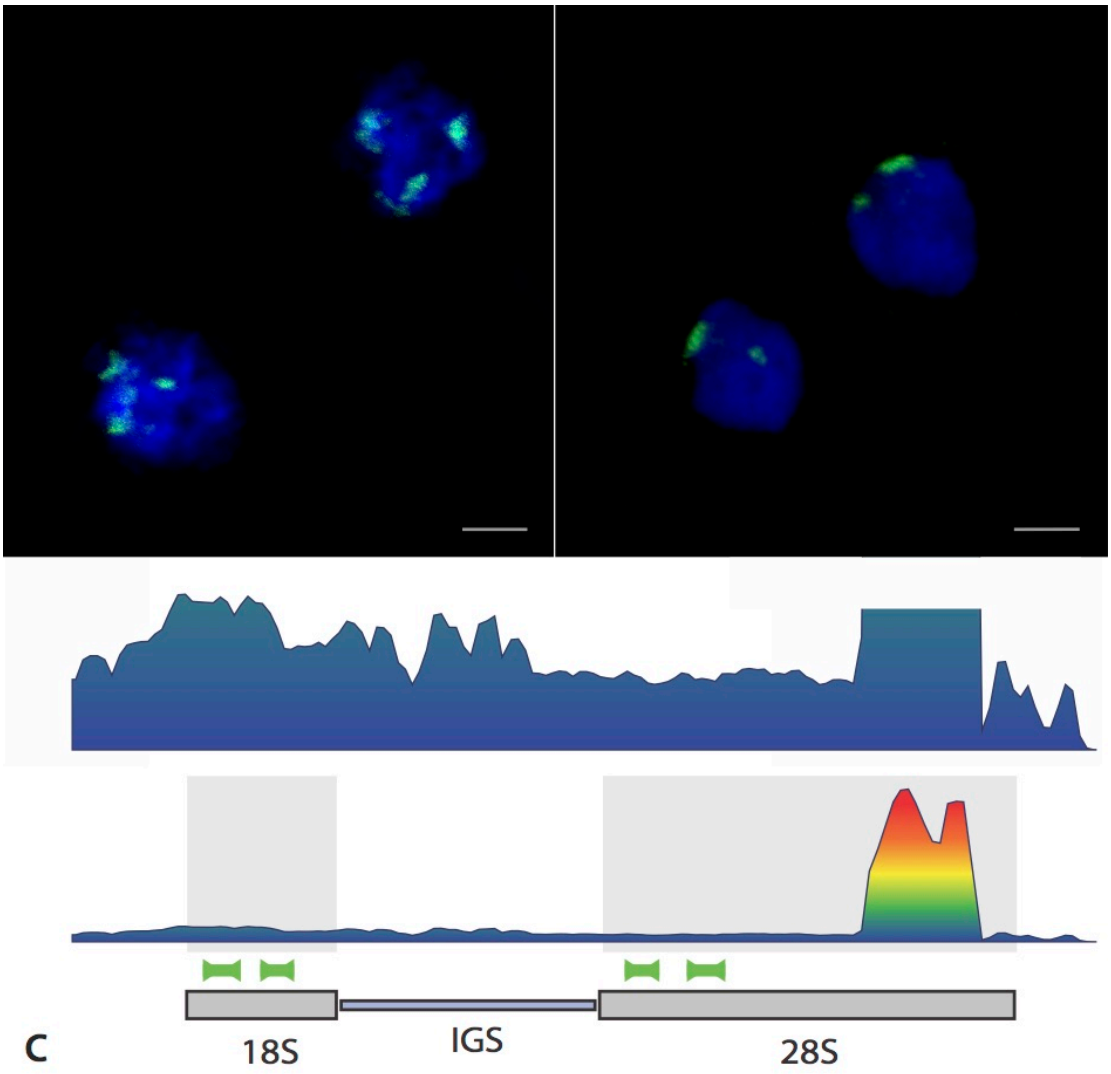


Figure 4.28. FISH signals obtained by cytogenetic mapping of rDNA in *B. testatum*. FISH signals resulting from hybridization of 18S probes to interphase nuclei. B, FISH signals resulting from hybridization of 28S probes to interphase nuclei. E, a rDNA profile for *B. lividulum* indicating the location of 18S and 28S FISH probes with green boxes.

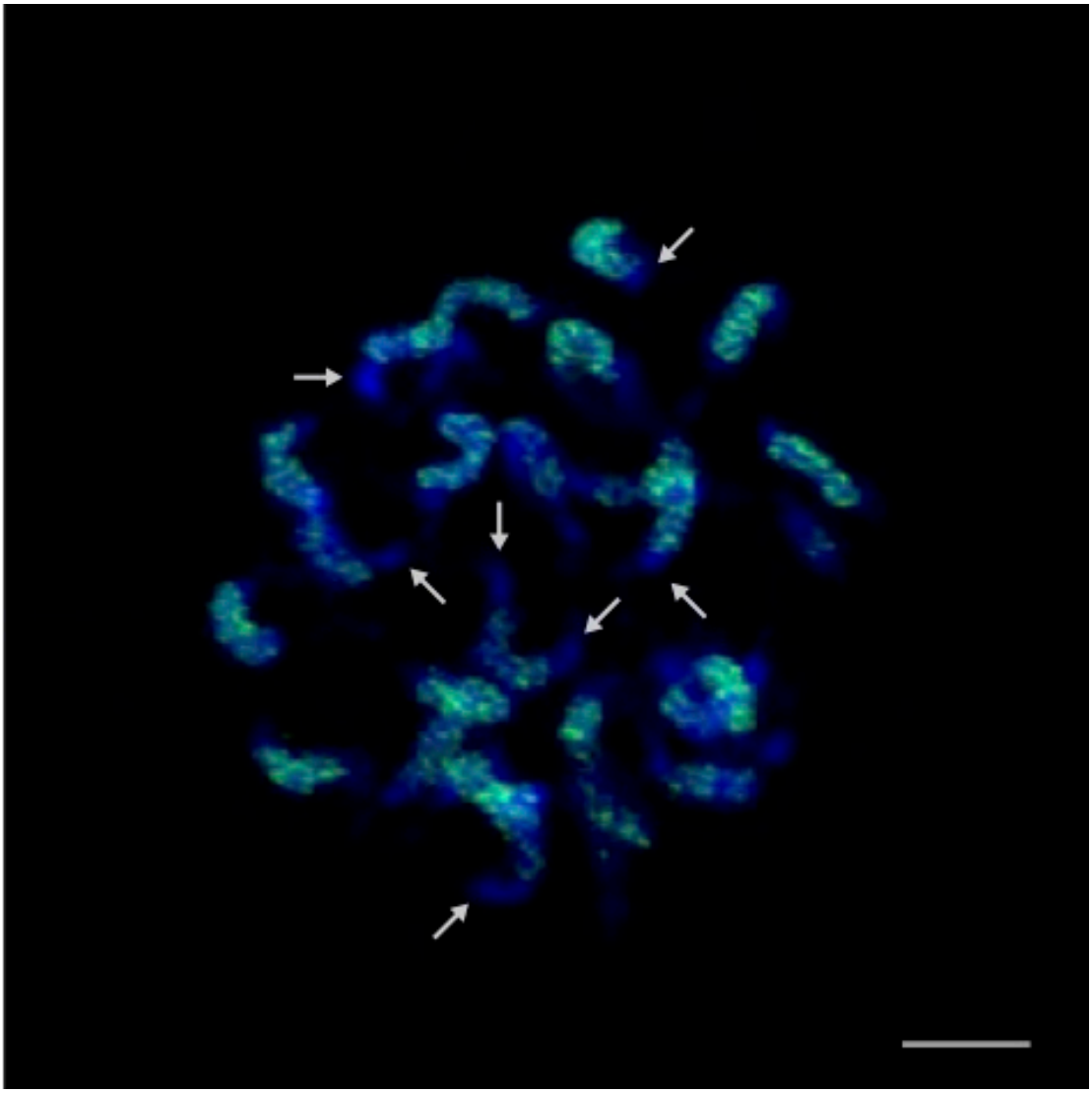


Figure 4.29. FISH signals resulting from hybridization of 28S probes to condensed chromosomes in *B. lividulum* with arrows indicating several euchromatic tails that are free from FISH signals.

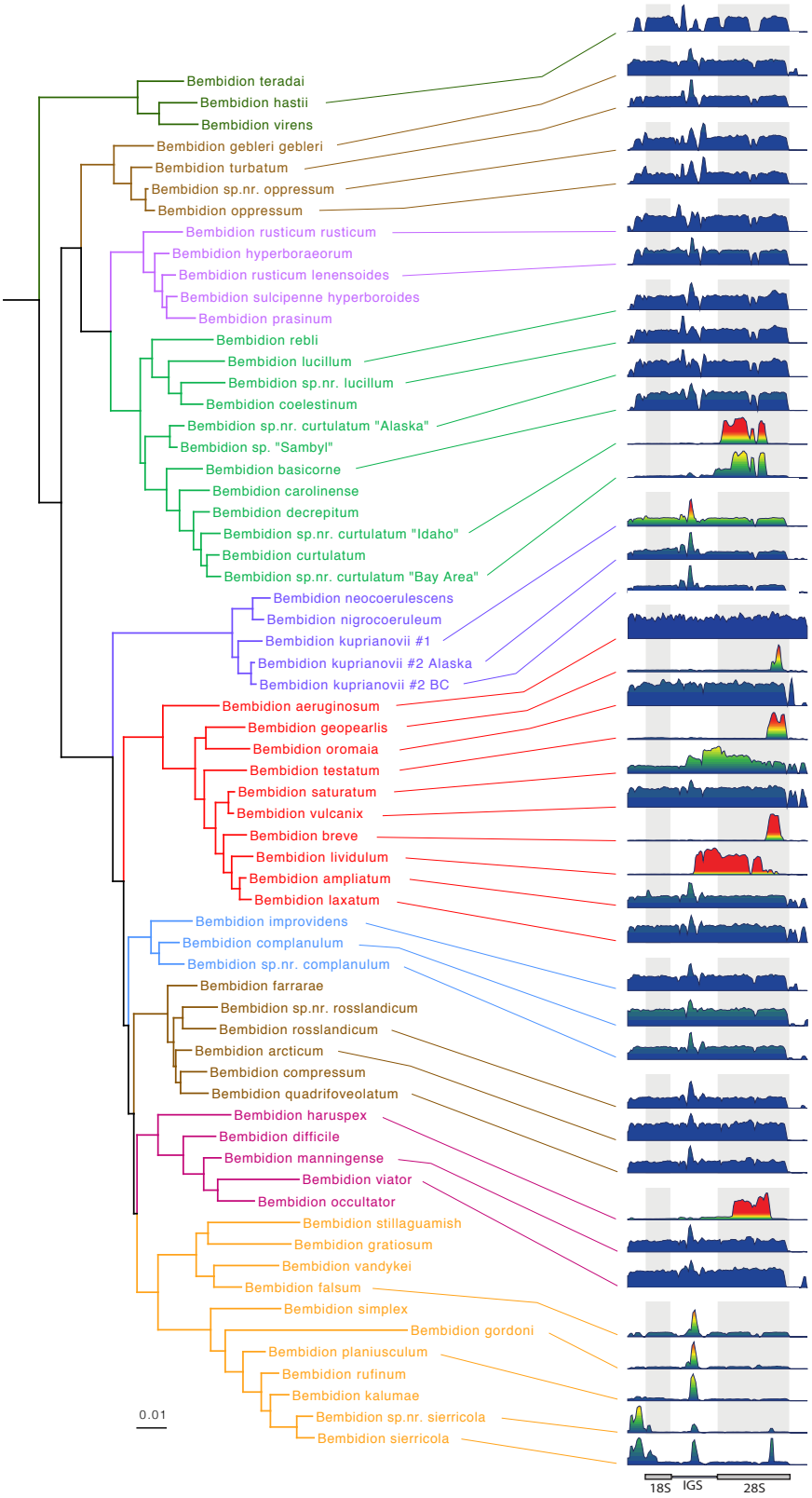


Figure 4.30. Maximum Likelihood tree of *Bembidion* subgenus *Plataphus*, the subgenus containing the *breve* species group with rDNA profiles for many species.

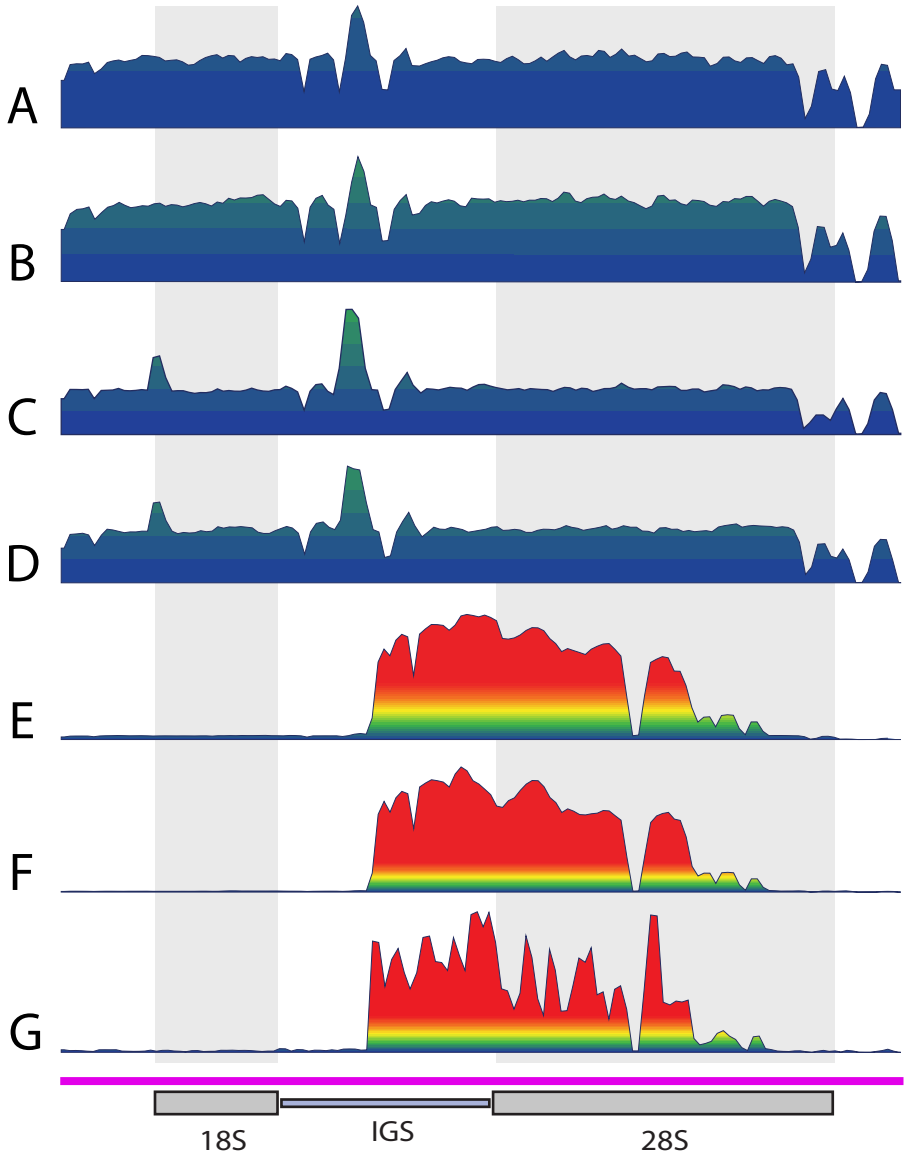


Figure 4.31. rDNA profile of 100-year-old type specimen (*B. lividulum* Casey), G, alongside rDNA profiles of DNA-preserved specimens of *Bembidion laxatum*, A–B; *B. ampliatum*, C–D; *B. lividulum* EF. Adapted from Sproul and Maddison (2017)

TABLES

Table 4.1. Settings used for parameter sensitivity analysis in CLC GW.

	MatchScore	MMCost	InsertCost	DelCost	LenFrac	SimFrac
Trial 1	1	1	1	1	0.4	0.5
Trial 2	1	1	1	1	0.5	0.6
Trial 3	2	2	2	2	0.6	0.75
Trial 4	2	2	3	3	0.5	0.75
Trial 5	2	2	3	3	0.5	0.8
Trial 6	3	4	3	3	0.8	0.8
Trial 7	3	4	3	3	0.85	0.85
Trial 8	3	4	3	3	0.9	0.9
Trial 9	3	4	3	3	0.95	0.95

Table 4.2. Specimens and locality data for the *breve* group.

Species	#	Locality
Bembidion ampliatum Casey	4160	USA: California: Mono Co., snow field above Ellery Lake, 2901m, 37.9345°N 119.2318°W
	5125	USA: California: Mono Co., snow field near Tioga Pass, 3116m, 37.9123°N 119.2472°W
	5130	USA: California: Tuolumne Co., snow field near Sonora Pass, 2908m, 38.3322°N 119.65°W
	3593	USA: Colorado: Mesa Co., Grand Mesa, route 65 at FS100, 3243m, 39.0316°N 108.0561°W
	4694	USA: Montana: Glacier Co., Glacier N.P., east slope Clements Mtn., 2131m, 48.6907°N 113.7308°W
	3544	USA: New Mexico: Santa Fe Co., Santa Fe Ski Basin, 3286m, 35.7889°N 105.7953°W
	4245	USA: Oregon: Harney Co., Steens Mts., snowfield at Kiger Gorge, 2618m, 42.7152°N 118.5786°W
	5017	USA: Oregon: Wallowa Co., S of Mount Howard, 2525m, 45.255°N 117.1769°W
Bembidion breve (Motschuylsky)	4187	USA: Alaska: Juneau, Heintzleman Ridge, 826m, 58.41904°N, 134.4422°W
	4919	USA: California: El Dorado Co., Lily Lake, 2000m, 38.8743°N 120.0801°W
	3799	USA: California: Tehama Co., Nanny Creek, Lassen NF, 1584m, 40.3696°N 121.5612°W
	4194	USA: Oregon: Hood River Co., Hood River Meadows Ski Area, 1586m, 45.3254°N 121.6625°W
	5012	USA: Oregon: Klamath Co., Munson Creek, Crater Lake NP, 1981m, 42.8987°N 122.1343°W
	5011	USA: Oregon: Klamath Co., Vidae Falls, Crater Lake NP, 1980m, 42.8832°N 122.0993°W
	4697	USA: Washington: Whatcom Co., Mt. Baker, Snoqualmie NF, 1291m, 48.8544°N 121.6969°W

Table 4.2. (Continued)

Species	**#**	**Locality**
Bembidion geopearlis Sproul and Maddison	4727	USA: Montana: Glacier Co., Glacier N.P., east slope Clements Mtn., 2129m, 48.692°N 113.7292°W
	4731	USA: Montana: Missoula Co., inlet to Heart Lake, 1891m, 47.3801°N 113.8501°W
	4700	USA: Montana: Ravalli Co., Lost Horse Creek, 1760m, 46.1417°N 114.4863°W
	5088	USA: Oregon: Baker County, Blue Mountains, Anthony Lake, 185m, 44.96122°N 118.23200°W
Bembidion laxatum Casey	5086	USA: CA: Lassen Co. Lassen National Park, Helen Lake, 2506m, 40.46740°N 121.50860°W
	4918	USA: California: Alpine Co., Sonora Pass, 2900m, 38.3323°N 119.6401°W
	4153	USA: California: Mono Co., snow field above Ellery Lake, 2901m, 37.9345°N 119.2318°W
	4149	USA: California: Tulare Co., snowfield below White Chief Lake, 2912m, 36.417°N 118.5941°W
Bembidion lividulum Casey	4165	USA: California: Fresno Co., Kaiser Pass Meadow, 2783m, 37.2948°N 119.1006°W
	3797	USA: California: Tulare Co., 2.5 km N Sherman Pass, Sequoia NF, 2608m, 36.0096°N 118.3678°W
	3486	USA: Idaho: Blaine Co., Galena Summitt, 2680m, 43.8728°N 114.7176°W
	4699	USA: Montana: Missoula Co., Glacier Creek, 1484m, 47.3811°N 113.7948°W
	5032	USA: Oregon: Klamath Co., Sun Notch, Crater Lake NP, 2163m, 42.9009°N 122.0988°W
	5019	USA: Oregon: Wallowa Co., Lostine River, 1840m, 45.2378°N 117.3803°W
	5013	USA: Oregon: Wallowa Co., Lostine River, Two Pan Trailhead, 1728m, 45.249°N 117.3763°W

Table 4.2. (Continued)

Species	#	Locality
	3804	USA: Washington: King Co, Snoqualmie Pass, 954 m, 47.4451°N 121.4245°W
Bembidion oromaia Sproul and Maddison	3886	USA: CA: Trinity Co., Trinity Mountain Wilderness, Below Grizzly Lake, 2083m, 41.01509°N 123.04890°W
	4250	USA: California: Tulare Co., snow field above Emerald Lake, 2851m, 36.5959°N 118.6756°W
	4155	USA: California: Tuolumne Co., Deadman Creek, 2700m, 38.3188°N 119.6634°W
Bembidion saturatum Casey	4167	USA: California: Fresno Co., Kaiser Pass Meadow, 2783m, 37.2948°N 119.1006°W
	3313	USA: California: Inyo Co. 1.5 km NE University Peak, 3240 m. 36.76030°N, 118.35450°W
	3467	USA: California: Lassen Co., Silver Lake, 1975m, 40.494°N 121.162°W
	5129	USA: California: Nevada Co. snow field, NW Carpenter Ridge, 2546 m, 39.41470°N 120.31440°W
	3588	USA: Nevada: Elko Co., Lamoille Canyon, Ruby Mtns, 2697m, 40.6017°N 115.378°W
	4247	USA: Oregon: Harney Co., Steens Mountains, creek below summit, 2754m, 42.6408°N 118.5749°W
Bembidion vulcanix Sproul and Maddison	4649	USA: Oregon: Deschutes Co., Creek below Little Three Creek Lake, 2018m, 44.1057°N 121.6347°W
	4618	USA: Oregon: Deschutes Co., stream east of Todd Lake, 1952m, 44.0282°N 121.6709°W
	4779	USA: Washington: Whatcom Co., Mt Baker, Snoqualmie NF, 1326m, 48.8528°N 121.6886°W
	3088	USA: California: El Dorado Co., Strawberry Creek at Sciots Camp, 1760m, 38.7835°N 120.1463°W

Table 4.2. (Continued)

Species	#	Locality
Bembidion testatum Casey	4157	USA: California: Fresno Co., creek below Kaiser Pass, 2722m, 37.2865°N 119.1009°W
	4173	USA: California: Trinity Co., Canyon Creek, 1440m, 40.949°N 123.0179°W
	3798	USA: California: Tulare Co., 2.5 km N Sherman Pass, Sequoia NF, 2608m, 36.0096°N 118.3678°W
	3166	USA: Oregon: Jackson Co., Mt Ashland Campground, Klamath NF, 2040m, 42.0756°N 122.714°W

Note: values inside parentheses are estimated from the data provided on locality labels.

Table 4.3. Library preparation and sequencing statistics for the *breve group*. #, the Maddison Lab DNA extraction number. **Library**, the Maddison Lab library preparation number. **Genome/HybSpike,** indicates whether the data were obtained through whole genome sequencing or a hybrid capture approach. **Input DNA**, nanograms of DNA used as input for library preparation. **Amp cycles**, number of cycles during library amplification. **% Map rDNA**, the percentage of all reads that mapped to the reference sequence of the rDNA cistron. **% cov Regier**, the average coverage of 67 single-copy nuclear protein coding genes used to convert coverage depth to copy number. **Maximum read depth**, the value of coverage depth for the region of the rDNA cistron with the highest coverage. **Max copy number**, the value of the copy number for the region of the rDNA cistron with the highest coverage.

Sample ID	#	Library	Genome/ HybSpike	Input DNA	Amp cycles	% Map rDNA	% cov Regier	Maximum read depth	Max copy number
B. ampliatum	5017	LIB0247	Genome	24.9	7	1.3%	0.63	3876	**6,154.99**
B. ampliatum	4245	LIB0249	Genome	50.0	6	1.0%	0.69	3169	**4,599.86**
B. ampliatum	3593	LIB0250	Genome	49.3	6	1.2%	0.77	3196	**4,156.41**
B. ampliatum	4160	LIB0303	HybSpike	49.0	6	1.1%	0.96	3072	3,200.00
B. ampliatum	3544	LIB0324	Genome	50.1	7	1.4%	0.82	3367	**4,081.54**
B. ampliatum	4694	LIB0328	Genome	50.1	7	0.8%	0.99	2566	2,580.45
B. ampliatum	5125	LIB0332	Genome	50.1	7	0.9%	0.88	3251	3,697.68
B. ampliatum	5130	LIB0344	Genome	50.2	7	0.9%	0.82	2671	3,273.28
B. breve	4919	LIB0209	Genome	45.8	6	**2.4%**	0.76	36568	**47,989.50**
B. breve	3799	LIB0251	Genome	49.1	6	**2.9%**	0.63	43749	**69,516.42**
B. breve	5012	LIB0252	Genome	49.2	6	**2.9%**	0.67	43750	**65,091.25**
B. breve	4194	LIB0253	Genome	49.5	6	**2.8%**	0.83	42452	**51,171.65**
B. breve	4187	LIB0327	Genome	49.7	7	**3.0%**	0.75	45620	**61,109.13**
B. breve	4697	LIB0329	Genome	50.0	7	**2.5%**	0.79	34073	**42,941.94**

Table 4.3. (Continued)

Sample ID	#	Library	Genome/ HybSpike	Input DNA	Amp cycles	% Map rDNA	% cov Regier	Maximum read depth	Max copy number
B. breve	5011	LIB0330	Genome	50.0	7	**2.8%**	0.86	34266	**39,980.55**
B. geopearlis	4700	LIB0305	HybSpike	50.8	6	1.1%	0.95	17037	**17,933.68**
B. geopearlis	4727	LIB0306	HybSpike	50.3	6	1.8%	0.64	21616	**33,775.00**
B. geopearlis	4731	LIB0321	Genome	50.1	7	1.2%	0.78	17700	**22,661.32**
B. geopearlis	5088	LIB0322	Genome	50.0	7	1.7%	0.75	22185	**29,706.75**
B. laxatum	4918	LIB0208	Genome	50.8	6	0.8%	0.98	1750	1,785.47
B. laxatum	4149	LIB0248	Genome	49.9	6	1.9%	0.93	3610	3,883.39
B. laxatum	4153	LIB0309	Genome	498.0	5	1.0%	0.90	1974	2,197.89
B. laxatum	5086	LIB0331	Genome	2.6	10	1.2%	0.70	2345	3,358.96
B. lividulum	5032	LIB0243	Genome	50.0	6	**14.8%**	0.70	61968	**87,989.40**
B. lividulum	5013	LIB0244	Genome	23.1	7	**19.1%**	0.37	76611	**206,239.23**
B. lividulum	4165	LIB0245	Genome	52.3	6	**11.2%**	0.89	41553	**46,604.98**
B. lividulum	4699	LIB0246	Genome	51.3	6	**16.1%**	0.72	64243	**89,160.34**
B. lividulum	3804	LIB0317	Genome	50.0	7	**13.3%**	0.64	53352	**83,275.75**
B. lividulum	3486	LIB0323	Genome	50.0	7	**13.2%**	0.64	50966	**80,202.48**
B. lividulum	3797	LIB0325	Genome	50.1	7	**12.0%**	0.72	45202	**62,445.20**
B. lividulum	5019	LIB0342	Genome	10.0	8	**17.8%**	0.51	69727	**137,221.86**
B. oromaia	4250	LIB0304	HybSpike	50.3	6	1.3%	0.82	2005	2,445.12
B. oromaia	3886	LIB0308	HybSpike	501.1	5	1.6%	0.66	2484	3,763.64

Table 4.3. (Continued)

Sample ID	#	Library	Genome/ HybSpike	Input DNA	Amp cycles	% Map rDNA	% cov Regier	Maximum read depth	Max copy number	
B. oromaia	4155	LIB0318	Genome	50.0	7		0.9%	0.82	1380	1,674.22
B. saturatum	3467	LIB0204	Genome	49.8	6		**3.4%**	1.01	7156	**7,107.67**
B. saturatum	3313	LIB0254	Genome	49.0	6		**3.6%**	0.84	9206	**10,959.52**
B. saturatum	4247	LIB0255	Genome	49.8	6		1.5%	0.85	2462	2,892.39
B. saturatum	4167	LIB0320	Genome	49.9	7		**2.8%**	0.95	5832	**6,153.63**
B. saturatum	3588	LIB0340	Genome	50.1	7		1.3%	0.84	1852	2,196.05
B. saturatum	5129	LIB0343	Genome	50.4	7		**3.1%**	0.96	6004	**6,234.25**
B. testatum	4169	LIB0202	Genome	50.3	7		**2.9%**	0.80	22156	**27,593.82**
B. testatum	3166	LIB0300	HybSpike	50.8	6		**2.4%**	0.79	23455	**29,689.87**
B. testatum	3798	LIB0301	HybSpike	50.0	6		**4.4%**	0.78	36673	**47,016.67**
B. testatum	4157	LIB0319	Genome	50.0	7		**2.1%**	0.81	18674	**23,016.43**
B. testatum	4173	LIB0326	Genome	21.8	7		**3.3%**	0.54	29285	**54,111.23**
B. testatum	3088	LIB0339	Genome	50.1	7		1.9%	0.98	15023	**15,377.71**
B. vulcanix	4618	LIB0256	Genome	49.1	6		1.2%	0.86	2002	2,318.56
B. vulcanix	4779	LIB0307	HybSpike	49.7	6		1.2%	1.01	1982	1,962.38
B. vulcanix	4649	LIB0341	Genome	25.0	8		1.2%	1.19	2196	1,847.86

Table 4.4. Primers used to amplify FISH probes.

Gene	Primer	Dir.	Sequence	Source
18S	breve.18S.F1.2	F	GGATAACTGTGGTAATTCTAGAGC	1
	breve.18S.R1.2	R	ACACAGATTCAACTACGAGC	1
	breve.18S.F2.2	F	CCTGAATACTGTGTGCATGG	1
	breve.18S.R2.1	R	CACCGAATCAAGAAAGAGCTC	1
28S	28SsF1	F	GAAACCGTTCAGGGGTAAACCTGAG	2
	28SsR2	R	CTCCACCGYRGGCCGTARATGGC	2
	breve.28SsF3	F	GGGAGATTCAATTGCTTTACCG	1
	breve.28SsR3	R	TTTATCCCAATGACTCGCGC	1
	breve.28SsF4	F	GGCAATGTAGTGTTTAGGAGAGC	1
	breve.28S.R4	R	CTTRATGCTCAACGGGTCAC	1
	breve.28SsF5	F	GTGTGAACAGAGGGAAGATGG	1
	breve.28S.R5	R	CTGTCTTAAGTTACCAACGCCT	1

Gene: gene name. Primer: published name of primer. Dir: direction of primer. Source: (1) this study, (2) Kanda *et al* 2015.

Table 4.5. Specimen and locality data for subgenus *Plataphus*.

B. manningense Lindroth	1419	USA: Montana: Missoula Co., Lolo Creek, 1200m, 46.7673°N 114.4654°W
B. haruspex Casey	1476	USA: Oregon: Lincoln Co., Cape Perpetua Campground on route 101 S of Yachats, 25m, 44.2809°N 124.1014°W
B. hastii C.R. Sahlberg	1703	Russia: Murmausu Area, Kandalaksha City, White Sea Coast, Palkina Guba Gulf
B. basicorne Notman	1911	USA: West Virginia: Pocahontas Co., N Fork Cherry River, 1010m, 38.1926°N 80.3537°W
B. complanulum Mannerheim	2083	Canada: British Columbia: Downtown Road, km 17, 50.5303°N 122.2712°W
B. improvidens Casey	2085	Canada: British Columbia: Downtown Road, km 17, 50.5303°N 122.2712°W
B. kuprianovii #1	2101	Canada: Alberta: Edmonton, 53.53°N 113.513°W
B. sp.nr.curtulatum "Idaho"	2145	USA: Idaho: Idaho Co., Lochsa River, 40.3 mi NE Lowell on route 12, 840m, 46.4507°N 115.0825°W
B. gordoni Lindroth	2358	USA: Montana: Gallatin Co., Bridger Creek E of Bozeman, 1495m, 45.7077°N 110.9743°W
B. rusticum lenensoides Lindroth	2380	USA: Alaska: Bear River at Nome-Council road, 45m, 64.8649°N 163.6917°W
B. arcticum Lindroth	2384	USA: Alaska: Solomon River at mouth of East Fork, 45m, 64.6924°N 164.2828°W
B. cascadia	2603	USA: Oregon: Benton Co., Marys Peak, 730m, 44.4981°N 123.5644°W
B. planiusculum Mannerheim	2604	USA: Oregon: Lane Co., Tenmile Creek, mi 4.7 on Tenmile Ck Rd, 75m, 44.2211°N 124.0365°W
B. viator Casey	3023	USA: Oregon: Lincoln Co., Neotsu, 5m, 45.0018°N 123.9779°W
B. sierricola Casey	3089	USA: California: El Dorado Co., Glen Alpine Creek, 2075m, 38.8752°N 120.0969°W
B. sp.nr.oppressum "Sierras"	3192	USA: California, Tuolumne Co., Deadman Creek at junction with Blue Canyon Creek, 2665 m, 38.31741°N 119.66522°W
B. lucillum Bates	3195	Japan: Kanagawa Pref., Nakatsu River nr. Yadoriki, 290m, 35.4015°N 139.1367°E
B. sp.nr. lucillum	3198	Japan, Aomori Pref., Oirase River along rt 102
B. sp.nr.curtulatum "Alaska"	3257	USA: Alaska: Niukluk River at Council, 30m, 64.8921°N 163.6807°W

Table 4.5. (Continued)

B. quadrifoveolatum Mannerheim	3326	CANADA: Alberta: Mt Edith Cavell, Jasper NP, 1765m, 52.6835°N 118.0529°W
B. rusticum rusticum Casey	3501	USA: Vermont: Windsor Co., Ottauquechee River, Bridgewater, 250m, 43.5878°N 72.6377°W
B. rosslandicum Lindroth	3604	CANADA: BC: Mt Baldy ski area, 1760m, 49.152°N 119.233°W
B. sp.nr.curtulatum "Bay Area"	3613	USA: California: San Mateo Co., Portola State Park, Pescadero Ck, 110m, 37.2511°N 122.2189°W
B. gebleri gebleri	3639	
B. falsum Blaisdell	3655	USA: California: Del Norte Co., Smith River, 12.7 km E Crescent City, 29m, 41.8206°N 124.1056°W
B. kuprianovii #2	3740	USA: Alaska: Chatanika River, 200m, 65.1399°N 147.4635°W
B. kuprianovii #2	3741	Canada: British Columbia: Whistler, Rainbow, Madely Trail, 50.1341°N 122.998°W
B. oppressum Casey	3828	USA: California: Sonoma Co., Russian River, Monte Rio, 2m, 38.4664°N 123.0118°W
B. aeruginosum Gebler	3890	Russia, Altai Republic, Krasnaya Mountain, 1786m, 50.09393°N 085.22787°E
B. sp.nr.complanulum	4994	USA: Oregon: Wallowa Co., S of Mount Howard, 2525m, 45.255°N 117.1769°W

Table 4.6. Genbank accession numbers for *Plataphus* specimens. TBD indicates sequences pending GenBank submission

Species	#	28S	CAD	wg	ArgK	Topo	COI
Bembidion bimaculatum	1281	TBD	TBD	TBD	TBD	TBD	TBD
B. californicum	-	EF648832	EF649386	EF649471	EF648692	TO_ADD	EF649109
B. chalceum	-	EF648892	EF649431	EF649548	EF648737	EU677650	EF649200
B. festivum	-	TBD	TBD	TBD	TBD	TBD	TBD
B. mimekara	1366	TBD	TBD	TBD	TBD	TBD	TBD
B. planum	1423	TBD	TBD	TBD	TBD	TBD	TBD
B. properans	-	TBD	TBD	TBD	TBD	TBD	TBD
B. punctulatum	1713	TBD	TBD	TBD	TBD	TBD	TBD
B. transversale	2157	EU677688	EU677541	EU677667	EU677517	EU677639	GU454797
B. genei illigeri	1484	TBD	TBD	TBD	TBD	TBD	TBD
B. geniculatum	1756	TBD	TBD	TBD	TBD	TBD	TBD
B. iridescens	-	TBD	TBD	TBD	TBD	TBD	TBD
B. planatum	-	GU556086	TBD	GU556035	TBD	TBD	TBD
B. hastii	1703	TBD	TBD	TBD	TBD	TBD	TBD
B. virens	2825	TBD	TBD	TBD	TBD	TBD	TBD
B. teradai	3834	TBD	TBD	TBD	TBD	TBD	TBD
B. turbatum	1417	TBD	TBD	TBD	TBD	TBD	TBD
B. sp.nr. oppressum	3192	TBD	TBD	TBD	TBD	TBD	TBD
B. gebleri gebleri	3639	TBD	TBD	TBD	TBD	TBD	TBD

Table 4.6. (Continued)

Species	#	28S	CAD	wg	ArgK	Topo	COI
B. oppressum	3828	TBD	TBD	TBD	TBD	TBD	TBD
B. kuprianovii #1	2101	TBD	TBD	TBD	TBD	TBD	TBD
B. neocoerulescens	2556	TBD	TBD	TBD	TBD	TBD	TBD
B. nigrocoeruleum	2571	TBD	TBD	TBD	TBD	TBD	TBD
B. kuprianovii #2 Alaska	3740	TBD	TBD	TBD	-	TBD	TBD
B. kuprianovii #2 BC	3741	TBD	TBD	TBD	-	TBD	TBD
B. rufinum	1434	TBD	TBD	TBD	TBD	TBD	TBD
B. gordoni	2358	TBD	TBD	TBD	TBD	TBD	TBD
B. simplex	1921	TBD	TBD	TBD	TBD	TBD	TBD
B. stillaguamish	1438	TBD	TBD	TBD	TBD	TBD	TBD
B. gratiosum	1340	TBD	TBD	TBD	TBD	TBD	TBD
B. sp.nr. sierricola	2603	TBD	TBD	TBD	TBD	TBD	TBD
B. planiusculum	2604	TBD	TBD	TBD	TBD	TBD	TBD
B. vandykei	2606	TBD	TBD	TBD	TBD	TBD	TBD
B. sierricola	3089	TBD	TBD	TBD	TBD	TBD	TBD
B. falsum	3655	TBD	TBD	TBD	TBD	TBD	TBD
B. kalumae	4235	TBD	TBD	TBD	TBD	TBD	TBD
B. aeruginosum	2848	TBD	TBD	TBD	TBD	TBD	TBD
B. ampliatum	3593	TBD	TBD	TBD	-	TBD	TBD
B. breve	3076	TBD	TBD	TBD	TBD	TBD	TBD

Table 4.6. (Continued)

Species	#	28S	CAD	wg	ArgK	Topo	COI
B. geopearlis	3471	TBD	TBD	TBD	TBD	TBD	TBD
B. laxatum	1170	TBD	TBD	TBD	TBD	TBD	TBD
B. lividulum	1930	TBD	TBD	TBD	TBD	TBD	TBD
B. oromaia	2967	TBD	TBD	TBD	TBD	TBD	TBD
B. saturatum	3313	TBD	TBD	TBD	TBD	TBD	TBD
B. testatum	3062	TBD	TBD	TBD	TBD	TBD	TBD
B. vulcanix	4615	TBD	TBD	TBD	-	TBD	TBD
B. basicorne	1911	TBD	TBD	TBD	TBD	TBD	TBD
B. decrepitum	2144	TBD	TBD	TBD	TBD	TBD	TBD
B. carolinense	2089	TBD	TBD	TBD	TBD	TBD	TBD
B. sp.nr. curtulatum "Idaho"	2145	TBD	TBD	TBD	TBD	TBD	TBD
B. curtulatum	2572	TBD	TBD	TBD	TBD	TBD	TBD
B. rebli	3119	TBD	TBD	TBD	TBD	TBD	TBD
B. lucillum	3195	TBD	TBD	TBD	TBD	TBD	TBD
B. sp.nr. lucillum	3198	TBD	TBD	TBD	TBD	TBD	TBD
B. sp.nr. curtulatum "Alaska"	3257	TBD	TBD	TBD	TBD	TBD	TBD
B. coelestinum	3690	TBD	TBD	TBD	TBD	TBD	TBD
B. sp.nr. curtulatum "Bay Area"	3613	TBD	TBD	TBD	TBD	TBD	TBD
B. sp. "Sambyl"	3691	TBD	TBD	TBD	TBD	TBD	TBD
B. farrarae	2084	TBD	TBD	TBD	TBD	TBD	TBD

Table 4.6. (Continued)

Species	#	28S	CAD	wg	ArgK	Topo	COI
B. sp.nr. rosslandicum	1356	TBD	TBD	TBD	TBD	TBD	TBD
B. arcticum	2384	TBD	TBD	TBD	TBD	TBD	TBD
B. compressum	2385	TBD	TBD	TBD	TBD	TBD	TBD
B. quadrifoveolatum	3326	TBD	TBD	TBD	TBD	TBD	TBD
B. rosslandicum	3604	TBD	TBD	TBD	TBD	TBD	TBD
B. manningense	1419	TBD	TBD	TBD	TBD	TBD	TBD
B. viator	3023	TBD	TBD	TBD	TBD	TBD	TBD
B. occultator	3172	TBD	TBD	TBD	TBD	TBD	TBD
B. difficile	3905	TBD	TBD	TBD	TBD	TBD	TBD
B. rusticum rusticum	1302	TBD	TBD	TBD	TBD	TBD	TBD
B. hyperboraeorum	2382	TBD	TBD	TBD	TBD	TBD	TBD
B. rusticum lenensoides	2380	TBD	TBD	TBD	TBD	TBD	TBD
B. sulcipenne hyperboroides	2383	TBD	TBD	TBD	TBD	TBD	TBD
B. prasinum	3269	TBD	TBD	TBD	TBD	TBD	TBD
B. complanulum	2083	TBD	TBD	TBD	TBD	TBD	TBD
B. haruspex	1476	TBD	TBD	TBD	TBD	TBD	TBD
B. improvidens	2085	TBD	TBD	TBD	TBD	TBD	TBD
B. sp.nr. complanulum	4994	TBD	TBD	TBD	-	TBD	TBD

Table 4.7. Models and partitioning schemes used in phylogenetic of *Bembidion* subgenus *Plataphus*

Model	Gene.CodonPosition
GTR+F+I+G4	CAD.1 + wg.1 + Topo.1
TPM3u+F+I+G4	CAD.2 + wg.2 + Topo.2
JC+I	ArgK.1 + ArgK.2
TN+F+I+G4	COI.1
K3Pu+F+I+G4	COI.2
K3P+I+G4	CAD.3
HKY+F+G4	Wg.3 + Topo.3
TPM2u+F+I+G4	ArgK.3
TVM+F+G4	COI.3
TVMe+I+G4	28S

Table 4.8. Library preparation and sequencing statistics for the specimens of subgenus *Plataphus*.

Sample ID	DNA #	Library	Genome/ HybSpike	Input DNA	Amp cycles	% Map rDNA	% coverage Regier	Maxumim read depth	Max copy number
B. aeruginosum	3890	LIB0206	Genome	50.0	6	0.6%	0.95	771	813.75
B. turbatum	1417	LIB0345	Genome	10.0	8	1.1%	1.40	4054	2,900.13
B. manningense	1419	LIB0346	Genome	25.0	8	0.9%	0.98	1687	1,720.96
B. haruspex	1476	LIB0347	Genome	25.0	8	**6.6%**	0.53	40694	**77,286.65**
B. hastii	1703	LIB0348	Genome	25.3	8	0.5%	1.65	1789	1,081.62
B. basicorne	1911	LIB0349	Genome	10.0	8	0.9%	0.74	1927	2,591.91
B. complanulum	2083	LIB0350	Genome	4.2	9	1.8%	0.71	3602	**5,044.82**
B. improvidens	2085	LIB0351	Genome	5.0	9	1.0%	1.15	2259	1,959.35
B. kupranovii 1	2101	LIB0352	Genome	25.1	8	**2.7%**	0.41	11665	**28,451.22**
B. sp.nr.curtulatum	2145	LIB0353	Genome	25.1	8	**4.8%**	0.73	31137	**42,536.89**
B. rusticum lenensoides	2380	LIB0354	Genome	25.0	8	0.8%	0.83	2215	2,677.71
B. arcticum	2384	LIB0355	Genome	10.0	8	0.9%	1.45	1933	1,333.47
B. cascadia	2603	LIB0356	Genome	10.1	8	0.6%	0.73	9558	**13,050.25**
B. planiusculum	2604	LIB0357	Genome	10.1	8	0.5%	0.60	7482	**12,542.47**
B. viator	3023	LIB0358	Genome	25.2	8	**2.0%**	0.95	4728	**4,983.14**
B. sierracola	3089	LIB0359	Genome	25.1	8	0.7%	0.87	4549	**5,232.75**
B. sp.nr.oppressum	3192	LIB0360	Genome	25.0	8	0.6%	1.16	1881	1,627.91
B. lucillum	3195	LIB0361	Genome	25.2	8	0.6%	0.96	1637	1,711.63
B. sp.nr. lucillum	3198	LIB0362	Genome	10.1	8	0.5%	0.93	1410	1,516.13

Table 4.8. (Continued)

Sample ID	DNA #	Library	Genome/ HybSpike	Input DNA	Amp cycles	% Map rDNA	% coverage Regier	Maxumim read depth	Max copy number
B. gordoni	2358	LIB0363	Genome	50.2	8	0.5%	0.37	7123	**19,440.50**
B. sp.nr.curatulatum	3257	LIB0364	Genome	50.3	8	0.6%	1.39	1363	980.58
B. quadrifoveolatum	3326	LIB0365	Genome	50.3	8	0.5%	0.77	1480	1,915.11
B. rusticum rusticum	3501	LIB0366	Genome	50.4	8	0.7%	1.25	1599	1,281.66
B. rosslandicum	3604	LIB0367	Genome	50.2	8	0.3%	0.92	1080	1,171.37
B. sp.nr.curtulatum	3613	LIB0368	Genome	20.1	8	1.7%	0.75	9351	**12,419.43**
B. gebleri gebleri	3639	LIB0369	Genome	50.5	8	0.6%	1.27	1790	1,414.95
B. falsum	3655	LIB0370	Genome	50.8	8	1.5%	0.82	13782	**16,733.85**
B. kupranovii 2	3740	LIB0371	Genome	50.4	8	1.2%	0.95	4423	**4,655.79**
B. kupranovii 2	3741	LIB0372	Genome	50.7	8	1.1%	1.04	6019	**5,774.91**
B. oppressum	3828	LIB0373	Genome	50.1	8	0.3%	1.13	1053	931.75
B. sp.nr.complanulum	4994	LIB0374	Genome	50.3	8	2.4%	1.33	6526	**4,910.71**

CHAPTER 5: CONCLUSION

ADVANCING KNOWLEDGE IN THE *BREVE* GROUP

The present work advances knowledge in a challenging species group for which little was known previously. It resurrects, or describes for the first time seven species in a group previously thought to contain only two. Species distribution maps and phylogenetic trees and reveal interesting patterns of sympatry and biogeography across the landscape. The use of Illumina sequencing with type specimens clarifies the group's previously complex taxonomic history.

Patterns of variation in rDNA profiles presented in Chapter 4 corroborate the species concepts delimited in Chapter 3. They also provide a new line of evidence that cryptic species are likely present within *B. saturatum*, and suggest that re-evaluation of geography-specific morphological patterns in *B. lividulum* may provide evidence for the presence of additional cryptic diversity. Ribosomal DNA profiles and fluorescence *in situ* hybridization (FISH) results provide evidence of notable genomic architecture differences in the repetitive genome among several species in the group.

THE BREVE GROUP AS A MODEL TO ADVANCE KNOWLEDGE IN SCIENCE, AND FUTURE DIRECTIONS

The pairing of short-read sequencing technology with the wealth of biodiversity data held in historical collections shows increasing promise as a tool for studying biodiversity (Wandeler, Hoeck, & Keller, 2007; Staats *et al.*, 2013; Kanda *et al.*, 2015). In Chapter 2, the *breve* group served as the impetus for an in-depth study on improving sample preparation and sequencing results of small-bodied historical specimens. The sample preparation guidelines presented herein have potential to not only improve sequencing outcomes in other studies attempting to sequence small-

bodied arthropods (which are the most abundant category of specimens held in historical collections), but also demonstrate that obtaining sequencing success and preservation of valuable DNA extractions are not mutually exclusive.

The cryptic diversity documented in the *breve* group is concentrated in high elevation habitats that are currently imperiled as global climates continue to warm, and has potential to serve as a model on future studies in climate change.

The in-depth investigation of rDNA profile variation presented in Chapter 4 documents extensive mobilization of rDNA in the *breve* group. This finding supports many other studies in diverse organisms that document similar mobilization of rDNA (Raskina *et al.*, 2008; Symonová *et al.*, 2013). This work presents the most extensive sampling of individuals and species showing rDNA mobilization of which I am aware and identifies the *breve* group as an ideal model for future study of this biological phenomenon. The dramatic copy number inflation observed rDNA profiles observed in *B. lividulum* which were corroborated by FISH signals further identify the *breve* group as potential model for studying rapid genome evolution. Future work is needed to elucidate mechanisms that cause the spread of rDNA throughout the genome, test for the involvement of retrotransposons, and test whether this process is directly involved in reproductive isolation and speciation. In addition, future work is needed to allow for quantitative comparison of minor variation that may be present in rDNA profiles.

Cluster analysis of repetitive DNA in the *breve* group (Chapter 4) represents the first genome-scale characterization of repetitive DNA content in carabid beetles. These results will be potentially valuable to efforts underway to study genome size

evolution in carabids (Maddison *et al.*, unpublished). The finding that profiles of repetitive DNA can hold phylogenetic signal support the findings of other studies (Dodsworth *et al.*, 2015). The simple methods outlined in Chapter 4 for visualizing signal of variation for repetitive loci such as rDNA has potential to add clarifying signal to other studies in phylogenomics, species delimitation, and genome evolution with minimal cost. As studies in phylogenomics find new ways to map patterns of genome evolution in non-model groups across phylogenies, we will come closer to understanding genome-scale mechanisms that have contributed to the diversification of life on the planet.

REFERENCES

Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoël M, Weiss-Schneeweiss H, Leitch AR. 2015. Genomic Repeat Abundances Contain Phylogenetic Signal. *Systematic Biology* 64: 112–126.

Kanda K, Pflug JM, Sproul JS, Dasenko MA, Maddison DR. 2015. Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. *PLoS One* 10: e0143929.

Raskina O, Barber J, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* 120: 351–357.

Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One* 8: e69189.

Symonová R, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Rábová M, Ráb P. 2013. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evolutionary Biology* 13: 1.

Wandeler P, Hoeck PE, Keller LF. 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* 22: 634–642.

APPENDICES

APPENDIX I

DNA repair

Prior to preparation of some libraries, we treated an aliquot of extractions with enzymes designed to repair nicks, gaps, and damaged bases in double-stranded DNA. We diluted extraction aliquots to 50 µl and treated DNA with NEBNext® FFPE DNA Repair Mix (New England BioLabs) following the manufacturer's protocol, except that we reduced elution volume to 33 µl (32.5 µl of which was retained) to accommodate the input volumes required by downstream library preparation protocols. We quantified 2µl of repaired DNA using a Qubit Fluorometer (Life Technologies) with a Quant-iT dsDNA HS Assay Kit before proceeding with library preparation.

Library preparation

We prepared 46 libraries from the 16 historical specimens. For most specimens, we generated libraries with each of two commercially available kits using the same amount of input DNA. Protocols used with each kit are described in the paragraphs that follow. We also conducted side-by-side comparisons within a given protocol using the same amount of repaired and unrepaired DNA as input. The libraries constructed for each specimen are summarized in Table 2.4. We determined how many libraries to construct with DNA from each historical specimen based on DNA availability (with the goal of preserving at least half of the total DNA for archival

purposes), and the potential research value of sequences that would be obtained from multiple libraries.

dsDNA libraries

We used the NEBNext® DNA Ultra II kit (New England BioLabs) to generate *dsDNA* libraries. We followed manufacturer's protocol with four minor modifications developed during early experimentation with the kit, as follows. (1) During both bead cleanup steps we reduced the elution volume by 1.5 µl and transferred all but 0.5 µl of the eluate to downstream steps, instead of transferring all but 2 µl as the protocol suggests. (2) During bead cleanup steps we pelleted beads on the side of the tube and reduced drying time with the lid open to less than one minute (elution buffer added when alcohol around pellet had evaporated and the pellet began changing from a glossy sheen to a dull appearance, and before the pellet began turning a lighter brown color). (3) We increased the number of PCR amplification cycles to 18-20 cycles. (4) We substituted the NEBNext® Ultra II Q5® Master Mix (New England BioLabs) that is included in the kit with NEBNext® High-Fidelity 2X PCR Master Mix (New England BioLabs) based on early results that provided some evidence that the latter performed better with degraded DNA template.

dsDNA Mod libraries

We prepared *dsDNA Mod* libraries using the *dsDNA* approach described above, but with three additional modifications to the protocol designed to reduce the quantity of adapter dimers in the final library, as follows. (1) We diluted the Illumina

adapter 1:25 even if samples had over 5 ng of input DNA (the protocol recommends a 1:10 dilution for greater than 5 ng of input). (2) We conducted two bead cleanups after adapter ligation, before proceeding to library amplification (the protocol calls for a single bead cleanup step). Both cleanups used a 0.9X ratio of beads to sample - the same ratio used in the pre-amplification cleanup of the standard *dsDNA* protocol. (3) We used a maximum of 18 cycles during PCR amplification of libraries, two less than the maximum number of cycles used in the standard *dsDNA* protocol.

ssDNA libraries

We generated *ssDNA* libraries using the Accel-NGS® 1S Plus kit (Swift Biosciences) following the manufacturer's protocol except that we increased the number of PCR amplification cycles to 18-20 cycles based on early experimentation with the kit.

Context specimen libraries

We constructed libraries for 12 context specimens using the NEBNext® DNA Ultra II kit (New England Biolabs) using the manufacturer's recommended protocol. Library details for each specimen are provided in Table 2.16.

Post-library cleanup

Following library preparation, we quantified libraries using a Qubit Fluorometer (Life Technologies) with a Quant-iT dsDNA HS Assay Kit with 2 µl of sample. We then diluted an aliquot of each library to a concentration of 1 ng/µl and

bioanalyzed 1 µl of the dilution on a 2100 Bioanalyzer (Agilent Technologies) using the High Sensitivity DNA Analysis Kit. We visualized bioanalysis traces to detect the presence of small fragments (<180 bases), presumably representing adapter dimers, which require removal prior to sequencing to avoid a large fraction of adapter reads in the sequencing output.

Most libraries (38 of 46) generated with both kits required additional cleanup to eliminate small fragments. We excluded small fragments from these libraries using bead-based size selection with Agencourt Ampure XP beads (Beckman Coulter, Inc.) by diluting libraries to 50 µl with 0.1X TE and adding 42.5 µl (or a 0.85X ratio of beads to sample) and then followed the post-amplification bead cleanup steps in the manufacturer's protocol, except that we eluted DNA with 33 µl 0.1X TE buffer and transferred 32.5 ul to a new tube for quantification and bioanalysis. For most libraries, the above cleanup removed undesired small fragments; however, for six libraries, a single cleanup did not eliminate all the unwanted fragments. For these libraries the cleanup process was repeated using a 0.83X ratio of beads. Five libraries required a third cleanup in which we used a 0.8X ratio of beads.

Illumina Sequencing

We selected 28 of the 46 libraries, and sequenced them on an Illumina HiSeq 3000 maintained by the Oregon State University Center for Genome Research and Biocomputing. Each sample was given roughly 1/10 of a 100-base paired-end lane. For lanes that included *ssDNA* libraries, which are lower diversity due to a low diversity molecule that is ligated to one end of the insert during library preparation,

we designed the lanes such that ~40% of the pool comprised higher diversity *dsDNA* libraries, and we spiked in 4% PhiX to further increase the diversity of the pool. These steps were taken because the HiSeq 3000 platform is known to show reduced sequencing success when a high percentage of the molecules being sequenced are of low diversity.

Trimming and assembly

Demultiplexing was performed using CASAVA version 1.8 (Illumina). Paired-end reads were imported into CLC Genomic Workbench version 8.5.1 (CLC Bio, referred to below as CLC GW), using default options except for the minimum and maximum paired-read distances, which we determined by analyzing a dilution of the enriched library on a Bioanalyzer 2100 (Agilent Technologies). Failed reads were removed during import.

We trimmed and excluded adapter sequences from reads in CLC GW. We set the quality score limit to 0.5, allowing for a maximum of two ambiguities per read, and searched on both strands to remove adapters and retained broken pairs. For all reads derived from *dsDNA* libraries, we trimmed 6 bases from 3' and 5' ends of all reads; we determined this as a reasonable trimming protocol based on visualizing read quality score graphs. We trimmed 10 bases from 3' and 5' ends of all *ssDNA* libraries following manufacturer recommendation in order to eliminate the low complexity bases that are ligated to the end of fragments as part of the library preparation. We explored a more aggressive trim approach to *ssDNA* library reads, as it is known that some ligated oligos may exceed 10 bases in length, but found that this decreased

assembly quality; however, the target sequences that were obtained appeared identical to sequences obtained from the less stringently trimmed assemblies.

We assembled trimmed reads with *de novo* and reference-based approaches in CLC GW. We generated *de novo* assemblies of paired, trimmed reads using an automatic word and bubble size, with the minimum contig length set to 200. We conducted reference-based assembly using sequences from a closely related species as a reference with the "Map Reads to Reference" tool in CLC GW. References used for each historical specimen are summarized in Table 2.5. Following read mapping, we removed duplicate mapped reads with a maximum representation of minority sequence threshold set to 20% and generated consensus sequences by assigning ambiguity codes at positions of discordance in CLC GW.

Assessing recovery of gene targets from sequenced libraries

Recovery of mtGenome and rDNA complex

We tested for recovery of the mitochondrial genome and ribosomal DNA gene complex from both *de novo* and reference-based assemblies.

We determined recovery success of *de novo* assemblies for these targets by creating BLAST databases from the assemblies of each library using NCBI's makeblastdb tool, and conducting BLAST searches (e-value cutoff of 1E-80, and word size=11) from within Mesquite v3.10 (Maddison and Maddison) using full mtGenome and rDNA complex query sequences derived from their respective reference specimens as listed in Table 2.5.

Many of these BLAST searches produced a single best hit. If this was the case, we verified that it closely matched other carabid beetle sequences (as opposed to it being a contaminant sequence) through a BLAST search of the NCBI databases before accepting it as our target sequence. In other cases, BLAST searches of historical specimen assemblies yielded multiple hits for a given target sequence. If this was the case, we excluded any hits that did not BLAST to beetles, as well as any contigs that were fully nested within a larger contig of a lower e-value. All other hits were merged into a single sequence in Mesquite, with any overlapping hits containing discordant characters at a given site being assigned ambiguity codes of the union of bases in the merged sequence. We then calculated the percent recovery of targets from *de novo* assemblies by aligning the bases recovered from historical specimen assemblies to the mtGenome (after removal of the control region) of a closely related species as listed in Table 2.5, and dividing the total number of bases recovered by the total number of bases in the reference sequence.

We determined recovery success from reference-based assemblies by employing the 'Map Reads to Reference' tool in CLC GW, and using reference sequences from the taxa listed in Table 2.5. Reads were mapped with a mismatch score 4, length fraction 0.85, similarity fraction 0.85. We removed duplicate mapped reads with a maximum representation of minority sequence threshold set to 20% and generated consensus sequences by assigning ambiguity codes at positions of discordance in CLC GW. We then calculated the percent recovery of targets from reference-based assembly as described above for *de novo* assembly targets.

We obtained the mtGenome and rDNA complex reference sequences used in the preceding analyses by preparing genomic DNA libraries with DNA from specimens preserved for DNA quality of the reference species using NEBNext® Ultra™ II DNA Library Prep Kits for Illumina® (New England BioLabs) following the manufacturer's protocol, and sequenced the libraries on an Illumina HiSeq 3000 maintained by the Oregon State University Center for Genome Research and Biocomputing. We trimmed sequences based on quality score and removed adapters, and conducted *de novo* assemblies in CLC GW. We used *Bembidion* sp.nr.*transversale* 3205 as a query in BLAST searches to identify contigs containing mtDNA and rRNA in the *de novo* assemblies. Because the mitochondrial genome is a circular molecule, it was necessary to re-arrange the linear sequences such that all began at the same point of reference. We aligned and re-arranged mitochondrial sequences in Mesquite using the mitochondrial genome of *Trachypachus holmbergi* (Sheffield *et al.* 2008) as a reference.

Recovery of 67 low-copy genes

We tested for recovery of 67 low-copy number nuclear protein-coding genes previously used in arthropod phylogeny (Regier *et al.* 2008) through a reference-based approach using the "Map Reads to Reference" tool in CLC GW. Reads were mapped with mismatch score 3, length fraction 0.8, and similarity fraction 0.85. We used reference sequences taken from *Bembidion* sp.nr. *transversale* in the read mapping of all historical specimens as in Kanda et al. (2015). Following read mapping, we removed duplicate mapped reads with a maximum representation of

minority sequence threshold set to 20% and generated consensus sequences by assigning ambiguity codes at positions of discordance in CLC GW.

Assessing the accuracy of sequences recovered from historical specimen libraries

Obtaining 7 gene fragments from context and historical specimens

We obtained context sequences from seven gene fragments from a combination of previously published data (Kanda et al. 2015; Maddison 2012; Regier et al. 2008), and sequencing (both Sanger and Illumina) of specimens new to this study. We targeted 7 gene fragments for analysis: 18S or 18S rDNA: approximately 2000 bases of 18S nuclear ribosomal DNA; 28S or 28S rDNA: approximately 1000–1100 bases of 28S nuclear ribosomal DNA; COI 5': 658 bases of the mitochondrial gene cytochrome oxidase I; this is the so-called barcode region; COI 3': approximately 820 bases of the mitochondrial gene cytochrome oxidase I; COII: approximately 745 bases of the mitochondrial gene cytochrome oxidase II, plus a small portion of tRNA-Leu. CAD: approximately 730 bases of the carbamoyl phosphate synthetase domain of the *rudimentary* gene; Topo: approximately 890 bases of topoisomerase I. Sequences generated via PCR and Sanger sequencing followed protocols used in Kanda et al. (2015). For COII and COI 3' (which were not included in Kanda et al 2015) the COI thermal profile was used with primers that were used for both PCR and cycle sequencing reactions: TCTAATATGGCAGATTAGTGC (COII, forward) and GTACTTGCTTTCAGTCATCTWATG (COII, reverse) (Liu et al., 1992), and

CAACATTTATTTTGATTTTTTG (COI 3', forward) and TCCAATGGACTAATCTGCCATATTA (COI 3', reverse) (Simon et al. 1994). For context sequences obtained from Illumina sequencing, we extracted target sequences from *de novo* assemblies using BLAST with a closely related species serving as a query sequence as listed in Table 2.5. A list of all context specimens and their sequences is provided in Tables 2.3 and 2.6.

We acquired the 7 gene fragments for historical specimens through the same approach as was used in testing the recovery of the mtGenome and rDNA complex from both *de novo* and ref-based assemblies, which used BLAST to extract target fragments from *de novo*, and read-mapping in CLC GW for ref-based assemblies. We trimmed 50 bases from the end of each *de novo* target recovered and retained both the untrimmed and trimmed sequences for phylogenetic analysis.

Phylogenetic Methods

We processed and aligned most target sequences and context sequences in Mesquite version 3.1 (Maddison & Maddison 2016), supplemented by MAFFT version 7.130b (Katoh & Standley 2013) for alignment of ribosomal genes. Sequences from historical specimens that were less than half the full length of the fragment were examined for their distinctiveness, and removed if they were identical at all overlapping bases to sequences in two or more of the context species. This resulted in removal of one CAD sequence from *Lionepha* (length 185 bases) and five sequences from *Ocydromus* (two 28S sequences, two 18S sequences, and one Topo sequence, ranging in length from 15 to 210 bases), and one Topo from *Trepanedoris*

(length 141 bases). We also removed the sequence for 28S from LIB0179 of specimen **7** due to an apparent assembly in which a bacterial contaminant assembled in the same contig as target sequence as evidenced through BLAST searches of specific regions of the contig.

We selected a model of evolution for each gene using jModelTest v2.1.10 (Darriba et al. 2012). We evaluated models across three substitution schemes and selected the best model among those supported by GARLI (Zwickl 2006) using the Bayesian information criterion. Models chosen are shown in Table 2.17. For the analysis of the full, concatenated Bembidiina gene set, models were chosen using PartitionFinder 1.1.1 (Lanfear *et al.* 2012); the optimal partition had COI 5' third positions in one part with model GTR+G, and all remaining sites in another part with model GTR+I+G. Searches for the maximum likelihood tree were conducted by GARLI, using 10 replicates for each search. We included in phylogenetic analyses three sequences from each historical specimen library (if recovered) which have standardized names in tree figures as follows:

- "DeNovo untrimmed ends" = sequences obtained from *de novo* assemblies.
- "DeNovo" = sequences obtained from *de novo* assemblies with 50 bases trimmed from each end.
- "Ref-Based" = sequences obtained from reference-based assemblies.

We assessed the accuracy of target sequences based on their consistent phylogenetic placement among candidate specimens at both broad and fine scales

(see Table 2.7 for predicted placement), and branch length relative to putative conspecifics.

Explanation of predicted candidate specimens

The predictions presented in Table 2.7 were made using morphological and geographic evidence, as summarized below:

Specimens **1**–**3**: These three specimens are part of the type series of *Bembidium erasum* LeConte. They are all teneral females, and cannot confidently be assigned to species based upon morphological evidence. The type locality is "Oregon". Extensive collecting by the authors indicates that there are only four species in what was Oregon in 1853–1857 to which the historical specimens might belong: the species now called *Lionepha erasa*, *L. chintimini*, and two undescribed species, *L.* "Bitterroots" and *L.* "Carson Spur", with the first two most likely given their distribution.

Specimen **4**: This specimen is a syntype of *Bembidion flohri* Bates from near Mexico City. This specimen is included in the present study to verify that it closely matches northern populations in the USA and Canada of what is also considered *Bembidion flohri*. Included are putative *B. flohri* from four northern populations and one specimen each of *B. obtusidens* and *B.* "Harney County", two close relatives of *B. flohri* with similar morphological characteristics.

Specimens **5** and **13**: Morphological data indicate these specimens are closely related to *Bembidion obscuripenne*, part of the "Nearctic Clade" of the *Ocydromus* complex of *Bembidion*. Specimen **5** is a paratype of *Bembidion ulkei* Lindroth, which, based upon genitalic structures, is distinct from *B. obscuripenne*, but very closely related. The genitalia of specimen **13** look identical to that of *B. obscuripenne*, but the external body form is distinct, as the specimen is wingless. We predict that this specimen represented a wingless, high-elevation form of *B. obscuripenne*.

Specimens **7** and **9**: These belong to the *breve* species group of *Bembidion* (*Plataphus*). Male genitalic characters are critical to diagnosing species in this group. Both historical specimens are female, and thus cannot be confidently identified to species using morphological characters. Based on external characters, specimen **7** could belong to either *Bembidion laxatum* or *B.* "Ebbets Pass". Included in the context specimens are three of each species from the central Sierra Nevadas as near the type locality of *Bembidion lividulum* Casey (of which specimen **7** is the lectotype) as possible. Specimen **9** could belong to *B.* "University Peak", *B. breve*, or possibly (though less likely) *B.* "Lily Lake Creek". Three specimens of *B.* "University Peak" and two specimens each of *B. breve* and *B.* "Lily Lake Creek" are included, with some specimens near the type locality of *B. saturatum* Casey (of which specimen **9** is the lectotype).

Specimen **11** and **15**: These two specimens are from SE Alaska and nearby British Columbia. The only species of *Lionepha* known from this area with similar morphological structures (especially microsculpture) is *Lionepha casta*.

Specimen **12**: This specimen, from California, belongs to the *Bembidion connivens* group of *Bembidion (Trepanedoris)*. The only species known from California or neighboring states to which it could belong based upon external characters are *B. remotum*, *B.* "Red Bluff", and *B.* "Lake Moreno". Included are one specimen of *B.* "Red Bluff", two *B. remotum* and three *B.* "Lake Moreno", the latter from as near to the type locality of *Bembidion disparile* Casey (of which specimen **12** is the lectotype) as available.

Specimen **16**: This specimen is the holotype of *Bembidion lindrothellus* Erwin & Kavanaugh, from SE Alaska. It is very teneral, and thus male genitalic characters are difficult to discern. However, its microsculpture indicates that it could only belong to the two species now known as *Lionepha chintimini* and *L. erasa*, or it could be a separate, closely related species.

Comparison of mtGenome and rDNA complex

For specimens **7** and **3**, we further explored sequence accuracy by aligning ref-based and *de novo* targets to the complete mtGenome and rDNA complex (*de novo* assembled) of several candidate taxa. We aligned ref-based and *de novo* targets from each library for both specimens to the complete mtGenome and rDNA complex

(*de novo* assembled) of candidate species to which the historical specimens could potentially belong (Table 2.7). Specimen **7** was aligned to *Bembidion* "Ebbets Pass" and *Bembidion laxatum*, and specimen **3** was aligned to *Lionepha erasa*, and *Lionepha chintimni*. We conducted sequence alignment in MAFFT using the G-INS-i strategy with the unalign level set to 0.6 using the "Leave gappy regions" selected (scoring matrix = 200PAM/k=2, Gap opening penality = 1.53, offset value=0.0). We removed invariant sites, condensed the matrix to a matrix of patterns in Mesquite, and tabulated the number of sites that followed each of three patterns:

1. Sites at which the historical sequence has a unique base (different from either of the two candidate species).
2. Sites at which the historical sequence matched candidate species A, and for which candidate species B had a different base at that site.
3. Sites at which the historical sequence matched candidate species B, and for which candidate species A had a different base at that site.

APPENDIX II

APPENDIX III

Obtaining rDNA reference sequence and parameter sensitivity analysis

We obtained this reference sequence through *de novo* assembly of *B. aeruginosum* reads in CLC GW with default settings, creating a BLAST database of the resulting contigs and using both 18S rRNA and 28S rRNA gene sequences obtained from *B. aeruginosum* as query sequences in a BLAST search against the database of contigs. The best scoring hit for both queries was a single contig ~14k bases in length. We then annotated boundaries of rRNA genes on the contig using RNAmmer 1.2 Server (Lagesen 2007).

Given the initial striking pattern of CNV variation observed in B. lividulum, we screened read mapping results for the presence of assembly artifacts and contamination. We performed initial validation of read mapping results in *B. lividulum* and *B. saturatum* through a qPCR assay that targeted both 18S and 28S for amplification in *B. lividulum*, in which 28S show dramatic increase in copy number relative to 18S. We used the ΔΔCT approach to calculate the difference in amplification success between 18S and 28S targets and observed a 120 fold increase in amplification product in 28S relative to 18S, which was comparable to the 156 fold increase in maximum copy number of 28S compared to 18S observed in initial read mapping results.

We screened for contamination by conducting *de novo* assembly of *B. lividulum* (DNA3486) and created a BLAST database of the resulting contings. We used the reference rDNA cistron sequence of *B. aeruginosum* as a query in a BLAST search of that database. We observed the region showing inflated CN had hundreds of BLAST hits. We selected the first 40 hits and used them as query sequences to NCBI databases to estimate the fraction of contaminants present in top hits, we determined contaminants to be any contigs that mapped to 18S or 28S regions that had top hits that were not congeners (i.e., *Bembidion*), or contigs that mapped to IGS regions that did not include insects in the top five hits. Relaxing the

definition of contaminants for contigs mapping to the IGS regions was necessary given that sequence variation of IGS regions can be high even among closely related taxa, and no IGS data are available for *Bembidion* in NCBI databases.

We conducted a parameter sensitivity analysis by selecting four specimens, two from the *breve* group (*B. lividulum* DNA3486 and *B. breve* DNA4187) and two other *Plataphus* (*B. gordoni* DNA2358 and *B. sp.nr.curtulatum* Idaho DNA3613) which were chosen to represent a diversity of rDNA profile shapes. We repeated read mapping nine times for each specimen across varying stringency values for each parameter (i.e., match score, mismatch score, insertion cost, deletion cost, length fraction, and similarity fraction) as listed in Table 4.1. We visualized rDNA profiles for all read mapping trials. Read mapping parameters were judged to be ideal if they produced a profile that had a stable shape for parameter settings one level of higher stringency, and one level of lesser stringency for at least three of four taxa studied. An example of the strategy parameter sensitivity analysis is provided in Fig. 4.4

Prior to settling on our approach to use *B. aeruginosum* as the reference for all taxa we explored the effect of reference choice on rDNA profiles ampliatum vs aerug.). Using a single reference was desirable so that rDNA profiles for various taxa would be all be generated from the same point of reference. This is particularly relevant given that indels can be common in rRNA genes and having a single reference as allows for a uniform point of view. An alternative approach would be to generate rDNA profiles by obtaining a reference sequence of the rDNA cistron for each species individually, and mapping reads to that reference. The advantage of this approach is that some artifacts of the read mapping observed (e.g., valleys and some minor peaks) (Fig. 4.5), would be eliminated, however this approach is infeasible given that for any specimens that have notably inflated rDNA regions will difficult/impossible to generate *de novo* assemblies of the entire rDNA cistron due to challenges associated with assembly of the high coverage regions, even using downsampling approaches (We made extensive effort to systematically downsample reads and repeat *de*

novo assembly in an attempt to improve contig lengths from *de novo* for several such species, but the contig length were only improved incrementally, and we still failed to recover a full length contig, or few contigs that collectively represented nearly all the target sequence, despite these efforts). Thus, for both efficiency and comparative value, we found the use of a single reference an effective approach with the closely related species of the *breve* group, and an adequate approach for the extended sampling of the subgenus. Had our sampling extended a further taxonomic breadth, use of additional references would have likely been necessary.

Testing for recovery of rDNA profiles from targeted sequencing approach

We enriched 500 ng of *B. oromaia* library LIB0308 in an individual capture reaction. For the remaining 8 libraries, we pooled 63 ng of each library and conducted enrichment in multiplexed capture reaction. We libraries incubated capture reactions with biotinylated RNA baits for 16–20 hours at 65° and captured the target-hybridized baits using streptavidin-coated beads. Following manufacturer recommended wash steps, we released the enriched bead-bound targets from the RNA baits via heat denaturation and amplified the enriched targets for 14 cycles using a Kapa Library Amplification Kit (Kapa Biosystems) and purified amplified products using Aline PCRClean DX beads (Aline Biosciences).

Because the baitset did not contain any probes for rDNA, and the target capture protocol we followed is expected to result in very low sequencing rates of non-target reads, we pooled unenriched library for eight of our nine samples and spiked that pool into the same lane that was used to sequence the enriched libraries, such that approximately half of the reads obtained for eight of the nine samples that received an unenriched spike would be reads from the enriched portion of the library, and half would be from the unenriched library. Because the enriched library is generated from an aliquot of the unenriched library, and

therefore have the same combination of dual indices, the reads derived from both sources would be grouped in the sequencing output. We desired to see if this strategy of spiking in unenriched library in the same sequencing lane would result in rDNA profiles that were similar to those from unenriched low-coverage sequencing, or whether the enrichment process would result in unexpected profile morphology. Enriched libraries and their unenriched counterparts were pooled and sequenced on the same 150 base PE Illumina HiSeq3000 lane at the Oregon State Center for Genome Research and Biocomputing.

Cytogenetic mapping of ribosomal DNA

Chromosome preparation

We fixed tissue for chromosome squashes following Larracuente and Ferree (2015). We dissected testes from freshly collected specimens of *Bembidion lividulum* and *B. testatum* in 1X PBT, transferred tissue to 0.5% sodium citrate for 10 minutes to promote chromosome spreading. We incubated tissue in 2.5% paraformaldehyde in 45% acetic acid for four minutes on a Sigmacote (Sigma-Aldrich) treated coverslip, and squashed chromosomes onto a polylysine-coated slide by folding the slide and coverslip in filter paper and applying firm downward pressure to the coverslip for 30 seconds. In most cases, individual testes were of sufficient size that we were able to prepare two slides from each specimen, each containing one testis.

Probe synthesis and fluorescence *in-situ* hybridization (FISH)

We generated probes for FISH using polymerase chain reaction (PCR) to amplify 18S and 28S targets with newly designed primers (Table 4.4). We designed primers to amplify two non-overlapping ~500 base amplicons in each target gene (Table 4.4 Figs. 4.27–28). We generated probes for *Bembidion lividulum* and *B. testatum* from PCR products amplified from the same species. We pooled PCR products by locus such that both 18S amplicons were in a pool and both 28S products were in pool. We then cleaned and

concentrated PCR products from each pool using Aline PCRClean DX beads (Aline Biosciences). We quantified total DNA of pools using a Qubit Fluorometer (Life Technologies) with a Quant-iT dsDNA HS Assay Kit with 1 µl of sample. We directly labeled 1µg of DNA from each pool using a ULYSIS® Alexa Fluor® 488 Nucleic Acid Labeling Kit (ThermoFisher) following the manufacturer's recommended protocol, and purified labeled probes with Centri-Sep spin columns (Princeton Separations).

We pretreated fixed chromosome squashes with RNase and pepsin prior to FISH following Symonova *et al*,. (2015). We incubated slides with 200 µl of RNase solution (200 µg of RNase in 1 ml 2X SSC buffer) for three hours at 37° and then for 3 minutes in pepisinization solution (0.005% pepsin in 0.01 N HCL) at 37°. We conducted hybridization and post-hybridization washes following Larracuente (2016). After dehydrating slides in ethanol and air drying, we incubated slides for five minutes at 95°with 20 µl of a hybridization solution containing 100 ng of fluorescently labeled probe, 10 µl formamide, 4 µl 50 % dextran sulfate, and 2 µl 20 SSC, 4 µl H2O. Slides were then cooled slightly, wrapped in parafilm, placed in a humidity chamber, and incubated for 14-20 hours at 37°. In early FISH experiments we compared two post-hybridization washing approaches, a simplified protocol (Larracuente 2015) in which slides were washed 3X for 15 min in 0.1X SSC at room temperature, and a more stringent approach in which slides were washed 3X for 5 minutes in 4X SSCT at 42°, and then 3X in 0.1X SSC at 60° for 5 minutes (Larracuente 2016). Although consistent patterns of FISH signals were observed regardless of wash strategy, we used the latter, more stringent protocol (Larracuente 2016) for results reported herein as it resulted in a better signal to noise ratio. We counterstained and mounted FISH slides using 11 µl SlowFade® Diamond Antifade Mountant with DAPI (ThermoFisher) and imaged FISH signals using a Zeiss LSM 780 NLO Confocal Microscope System. We verified FISH patterns resulting from the above protocol on dozens of interphase nuclei from five *B. lividulum* individuals, and three *B. testatum* individuals. We also verified *B. lividulum*-

specific FISH patterns on consensed chromosomes in at least 10 nuclei from five replicate individuals.

In addition to our preferred protocol described above we tested for the presence of probe-specific, and fluorophore-specific artifacts in FISH patterns by using alternative probe synthesis and FISH approaches. For *B. lividulum*, we used the above protocol but with an different pair of 28S primers (Table 4.4). Also, as an alternative to our direct probe labeling strategy, we generated biotin-labeled probes using Biotin-Nick Translation Mix (Roche) following the manufacturer's recommended protocol. We used as input DNA for nick-translation an approximately 1100 base fragment of 28S (Table 4.4). We verified that input DNA was fragmented between 200-700 bases using gel electrophoresis before completing probe synthesis. Sample preparation and hybridization of these probes followed our preferred protocol described above, except that following the post-hybridization washes, we incubated slides with 100 µl blocking solution for 30 minutes at 37°. Following blocking we detected probes by incubation with Streptavidin, Rhodamine Red™-X conjugate (ThermoFisher) for 30 minutes at 37°, and repeated the post-hybridization wash steps as described above before mounting slides following Larracuente (2016). Finally, we tested for non-specific binding of 28S probes generated from *B. testatum* and *B. testatum* PCR products by hybridizing probes generated from *B. lividulum* PCR products to *B. testatum* chromosomes, and hybridizing *B. testatum*-generated probes to chromosomes of *B. lividulum* (using our preferred protocol).

Testing for rDNA profile variation across the subgenus

We assembled chromatograms using Phred (Green & Ewing, 2002) and Phrap (Green, 1999) via the Chromaseq package in Mesquite v3.2 (Maddison & Maddison, 2016, 2017). Final sequence editing was conducted manually in Chromaseq. We aligned sequences

from protein-coding genes in Mesquite; no insertion or deletion events need be presumed in the history of the sequences examined. The ribosomal gene (28S) was aligned in MAFFT 7.130b (Katoh & Toh, 2008) with the G-INS-I algorithm as implemented in Mesquite. Following alignment, data matrices for each gene were prepared for downstream analysis using Mesquite.

We used IQ-TREE v1.6.5 (Nguyen et al 2014) to find optimal models of character evolution and partitioning of the data, and for tree inference. The beginning partition had each codon position in each gene as a separate partition, plus all of 28S in another partition. We searched for the Maximum Likelihood tree across 100 search replicates.

REFERENCES

Bates HW. 1878 On new genera and species of geodephagous Coleoptera from Central America. *Proceedings of the Scientific Meetings of the Zoological Society of London* 1878, 587-609.

de Bello Cioffi M, Bertollo LAC, Villa MA, de Oliveira EA, Tanomtong A, Yano CF, Supiwong W, Chaveerach A. 2015. Genomic organization of repetitive DNA elements and its implications for the chromosomal evolution of channid fishes (Actinopterygii, Perciformes). *PLoS One* 10: e0130199.

Bennett EA, Massilani D, Lizzo G, Daligault J, Geigl EM, Grange T. 2014. Library construction for ancient genomics: single strand or double strand. *BioTechniques* 56: 289–90, 292–6, 298.

Besnard G, Bertrand JA, Delahaie B, Bourgeois YX, Lhuillier E, Thébaud C. 2015. Valuing museum specimens: high-throughput DNA sequencing on historical collections of New Guinea crowned pigeons (Goura). *Biological Journal of the Linnean Society*.

Besnard G, Christin PA, Malé PJG, Lhuillier E, Lauzeral C, Coissac E, Vorontsova MS. 2014. From museums to genomics: old herbarium specimens shed light on a C3 to C4 transition. *Journal of Experimental Botany* 65: 6711–6721.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, Moritz C. 2013. Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* 22: 6018–6032.

Blaimer BB, Lloyd MW, Guillory WX, Brady SG. 2016. Sequence Capture and Phylogenetic Utility of Genomic Ultraconserved Elements Obtained from Pinned Insect Specimens. *PLoS One* 11: e0161531.

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008: P10008.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* 10: e1003537.

Brock TD, Freeze H. 1969. Thermus aquaticus gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of Bacteriology* 98: 289–297.

Burrell AS, Disotell TR, Bergey CM. 2015. The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution* 79: 35–44.

Carmi O, Witt CC, Jaramillo A, Dumbacher JP. 2016. Phylogeography of the Vermilion Flycatcher species complex: multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. *Molecular Phylogenetics and Evolution*.

Cabral-de-Mello DC, Cabrero J, López-León MD, Camacho JPM. 2011. Evolutionary dynamics of 5S rDNA location in acridid grasshoppers and its relationship with H3 histone gene and 45S rDNA location. *Genetica* 139: 921–931.

Cabral-de-Mello DC, Moura R, Martins C. 2010. Chromosomal mapping of repetitive DNAs in the beetle Dichotomius geminatus provides the first evidence for an association of 5S rRNA and histone H3 genes in insects, and repetitive DNA similarity between the B chromosome and A complement. *Heredity* 104: 393–400.

Casey TL. 1918. A review of the North American Bembidiinae. *Memoirs on the Coleoptera. VIII. The New Era Printing Company, Lancaster, Pennsylvania*: 1–223.

Carstens BC, Pelletier TA, Reid NM, Satler JD. 2013. How to fail at species delimitation. *Molecular Ecology* 22: 4369–4383.

Cioffi M, Bertollo L. 2012. Chromosomal distribution and evolution of repetitive DNAs in fish. *Repetitive DNA*. Karger Publishers, 197–221.

Consortium MGS. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420: 520.

Consortium ICGS. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.

Cruz-Dávalos DI, Llamas B, Gaunitz C, Fages A, Gamba C, Soubrier J, Librado P, Seguin-Orlando A, Pruvost M, Alfarhan AH. 2016. Experimental conditions improving in-solution target enrichment for ancient DNA. *Molecular Ecology Resources*.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.

Darwin C. 1859. *On the Origin of Species*. Routledge.

Da Silva E, Busso A, Parise-Maltempi PP. 2012. Characterization and Genome Organization of a Repetitive Element Associated with the Nucleolus Organizer Region in Leporinuselongatus (Anostomidae: Characiformes). *Cytogenetic and Genome Research* 139: 22–28.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

Denver DR, Brown AMV, Howe DK, Peetz AB, Zasada IA. 2016. Genome Skimming: A Rapid Approach to Gaining Diverse Biological Insights into Multicellular Pathogens. *PLOS Pathogens* 12: e1005713.

Dimitri P, Arcà B, Berghella L, Mei E. 1997. High genetic instability of heterochromatin after transposition of the LINE-like I factor in Drosophila melanogaster. *Proceedings of the National Academy of Sciences* 94: 8052–8057.

Dimitri P, Junakovic N. 1999. Revising the selfish DNA hypothesis: new evidence on accumulation of transposable elements in heterochromatin. *Trends in Genetics* 15: 123–124.

Ding XL, Xu TL, Wang J, Luo L, Yu C, Dong GM, Pan HT, Zhang QX. 2016. Distribution of 45S rDNA in Modern Rose Cultivars (Rosa hybrida), Rosa rugosa, and Their Interspecific Hybrids Revealed by Fluorescence in situ Hybridization. *Cytogenetic and Genome Research* 149: 226–235.

Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoël M, Weiss-Schneeweiss H, Leitch AR. 2015. Genomic Repeat Abundances Contain Phylogenetic Signal. *Systematic Biology* 64: 112–126.

Domingos FM, Bosque RJ, Cassimiro J, Colli GR, Rodrigues MT, Santos MG, Beheregaray LB. 2014. Out of the deep: cryptic speciation in a Neotropical gecko (Squamata, Phyllodactylidae) revealed by species delimitation methods. *Molecular Phylogenetics and Evolution* 80: 113–124.

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

Erwin TL, Kavanaugh DH. 1981. Systematics and zoogeography of Bembidion Latreille: 1. The carlhi and erasum groups of western North America (Coleoptera: Carabidae, Bembidiini). *Entomologica Scandinavica Supplement* 15: 33–72.

Erwin TL. 1984. Studies of the tribe Bembidiini (Coleoptera: Carabidae): lectotype designations and species group assignments for Bembidion species described by Thomas L. Casey and others. *The Pan-Pacific Entomologist* 60: 165–197.

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*: sys004.

Feschotte C, Pritham EJ. 2007. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annual Review of Genetics* 41: 331–368.

Firkowski CR, Bornschein MR, Ribeiro LF, Pie MR. 2016. Species delimitation, phylogeny and evolutionary demography of co-distributed, montane frogs in the southern Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution* 100: 345–360.

Ganglbauer L. 1891. Die Käfer von Mitteleuropa. Die Käfer der österreichisch-ungarischen Monarchie, Deutschlands, der Schweiz, sowie des französischen und italienischen Alpengebietes. Ester Band. Familienreihe Caraboidea. *Carl Gerold's Sohn*: iii, 557.

Gansauge MT, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols* 8: 737–748.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* 27: 182–189.

Gong J, Dong J, Liu X, Massana R. 2013. Extremely high copy numbers and polymorphisms of the rDNA operon estimated from single cell analysis of oligotrich and peritrich ciliates. *Protist* 164: 369–379.

Green P. 1999. Phrap. Version 0.990329. Available: http://phrap.org.

Green P, Ewing B. 2002. Phred. Version 0.020425c. Available: http://phrap.org.

Guschanski K, Krause J, Sawyer S, Valente LM, Bailey S, Finstermeier K, Sabin R, Gilissen E, Sonet G, Nagy ZT. 2013. Next-generation museomics disentangles one of the largest primate radiations. *Systematic Biology* 62: 539–554.

Hatch MH. 1950. Studies on the Coleoptera of the Pacific Northwest. II: Carabidae: Bembidiini. *The Pan-Pacific Entomologist* 26: 97–106.

Hawkins MT, Hofman CA, Callicrate T, McDonough MM, Tsuchiya MT, Gutiérrez V, Helgen KM, Maldonado JE. 2015. In-solution hybridization for mammalian mitogenome enrichment: pros, cons and challenges associated with multiplexing degraded DNA. *Molecular Ecology Resources*.

Heled J, Drummond AJ. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27: 570–580.

Hershler R, Sada DW. 2002. Biogeography of Great Basin aquatic snails of the genus Pyrgulopsis. *Smithsonian Contributions to the Earth Sciences* 33: 255–276.

Hind KR, Miller KA, Young M, Jensen C, Gabrielson PW, Martone PT. 2015. Resolving cryptic species of Bossiella (Corallinales, Rhodophyta) using contemporary and historical DNA. *American Journal of Botany* 102: 1912–1930.

Hofreiter M, Paijmans JL, Goodchild H, Speller CF, Barlow A, Fortes GG, Thomas JA, Ludwig A, Collins MJ. 2015. The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays* 37: 284–293.

Holland PM, Abramson RD, Watson R, Gelfand DH. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'—3'exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences* 88: 7276–7280.

Houston DD, Shiozawa DK, Riddle BR. 2010. Phylogenetic relationships of the western North American cyprinid genus Richardsonius, with an overview of phylogeographic structure. *Molecular Phylogenetics and Evolution* 55: 259–273.

Huang JP, Knowles LL. 2015. 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.

Iwata-Otsubo A, Radke B, Findley S, Abernathy B, Vallejos CE, Jackson SA. 2016. Fluorescence In Situ Hybridization (FISH)-Based Karyotyping Reveals Rapid Evolution of Centromeric and Subtelomeric Repeats in Common Bean (Phaseolus vulgaris) and Relatives. *G3: Genes, Genomes, Genetics* 6: 1013–1022.

James SA, West C, Davey RP, Dicks J, Roberts IN. 2016. Prevalence and Dynamics of Ribosomal DNA Micro-heterogeneity Are Linked to Population History in Two Contrasting Yeast Species. *Scientific Reports* 6: 28555.

Jiang J, Gill BS. 1994. New 18S- 26S ribosomal RNA gene loci: chromosomal landmarks for the evolution of polyploid wheats. *Chromosoma* 103: 179–185.

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

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research* 110: 462–467.

Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258: 818–821.

Kanda K, Pflug JM, Sproul JS, Dasenko MA, Maddison DR. 2015. Successful recovery of nuclear protein-coding genes from small insects in museums using Illumina sequencing. *PLoS One* 10: e0143929.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* 9: 286–298.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772–80.

Knowles LL, Carstens BC. 2007. Delimiting species without monophyletic gene trees. *Systematic Biology* 56: 887–895.

Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: An information aesthetic for comparative genomics. *Genome Research* 19: 1639–1645.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* 29: 1695–1701.

Larracuente AM. 2017. FISH in Drosophila. *Fluorescence In Situ Hybridization (FISH)*. Springer, 467–472.

Larracuente AM, Ferree PM. 2015. Simple method for fluorescence DNA in situ hybridization to squashed chromosomes. *Journal of visualized experiments: JoVE*.

Lemmon AR, Emme SA, Lemmon EM. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*: sys049.

Lim HC, Braun MJ. 2016. High-throughput SNP genotyping of historical and modern samples of five bird species via sequence capture of ultraconserved elements. *Molecular Ecology Resources* 16: 1204–1223.

Lindroth CH. 1963. The ground beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska, Part 3. *Opuscula Entomologica Supplementum*: 201–408.

Lindroth CH. 1969. The ground beetles (Carabidae, excl. Cicindelinae) of Canada and Alaska, Part 6. *Opuscula Entomologica Supplementum* 34: 945–1192.

Lindstrom SC, Gabrielson PW, Hughey JR, Macaya EC, Nelson WA. 2015. Sequencing of historic and modern specimens reveals cryptic diversity in Nothogenia (Scinaiaceae, Rhodophyta). *Phycologia* 54: 97–108.

Lower SS, Johnston JS, Stanger-Hall KF, Hjelmen CE, Hanrahan SJ, Korunes K, Hall D. 2017. Genome Size in North American Fireflies: Substantial Variation Likely Driven by Neutral Processes. *Genome Biology and Evolution* 9: 1499–1512.

Maddison D. 1985. Chromosomal diversity and evolution in the ground beetle genus Bembidion and related taxa (Coleoptera: Carabidae: Trechitae). *Genetica* 66: 93–114.

Maddison DR. 1993. Systematics of the holarctic beetle subgenus Bracteon and related Bembidion (Coleoptera: Carabidae). *Bulletin of the Museum of Comparative Zoology* 153: 143–299.

Maddison DR. 2008. Systematics of the North American beetle subgenus Pseudoperyphus (Coleoptera: Carabidae: Bembidion) based upon morphological, chromosomal, and molecular data. *Annals of Carnegie Museum* 77: 147–193.

Maddison DR. 2012. Phylogeny of Bembidion and related ground beetles (Coleoptera: Carabidae: Trechinae: Bembidiini: Bembidiina). *Molecular Phylogenetics and Evolution* 63: 533–576.

Maddison DR, Cooper KW. 2014. Species delimitation in the ground beetle subgenus Liocosmius (Coleoptera: Carabidae: Bembidion), including standard and next-generation sequencing of museum specimens. *Zoological Journal of the Linnean Society* 172: 741–770.

Maddison DR, Maddison WP. 2014. Chromaseq: a Mesquite module for analyzing sequence chromatograms. Version 1.1. Available: http://mesquiteproject.org/packages/chromaseq.

Maddison WP, Maddison DR. 2016. Mesquite: a modular system for evolutionary analysis. version 3.10. http://mesquiteproject.org

Maddison WP, Maddison DR. 2017. Mesquite: a modular system for evolutionary analysis. Version 3.2. Available: http://mesquiteproject.org.

Mannerheim CG. 1853. Dritter Nachtrag zur Kaefer-Fauna der nord-amerikanischen Laender des russischen Reiches. *Bulletin de la Société Impériale des Naturalistes de Moscou* 26: 95–273.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR. 2010. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* 39: D225–D229.

Mason VC, Li G, Helgen KM, Murphy WJ. 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. *Genome Research* 21: 1695–1704.

Martins C, Ferreira IA, Oliveira C, Foresti F, Galetti Jr PM. 2006. A tandemly repetitive centromeric DNA sequence of the fish Hoplias malabaricus (Characiformes: Erythrinidae) is derived from 5S rDNA. *Genetica* 127: 133–141.

Mayr E. 1982. *The growth of biological thought: Diversity, evolution, and inheritance*. Harvard University Press.

McClintock B. 1934. The relation of a particular chromosomal element to the development of the nucleoli in Zea mays. *Zeitschrift für Zellforschung und mikroskopische Anatomie* 21: 294–326.

McCormack JE, Tsai WL, Faircloth BC. 2015. Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources*.

McKee BD, Habera L, Vrana JA. 1992. Evidence that intergenic spacer repeats of Drosophila melanogaster rRNA genes function as X-Y pairing sites in male meiosis, and a general model for achiasmatic pairing. *Genetics* 132: 529–544.

Motschoulsky V. 1845. Remarques sur la collection de coléoptères Russes de Victor de Motschoulsky. *Bulletin de la Société Impériale des Naturalistes de Moscou* 18: 3–127.

Motschulsky V. 1864. Enumération des nouvelle espèces de coléoptères rapportés de ses voyages. 4-éme article. *Bulletin de la Société Impériale des Naturalistes de Moscou* 37: 171–240.

Mutanen M, Kekkonen M, Prosser SW, Hebert PD, Kaila L. 2015. One species in eight: DNA barcodes from type specimens resolve a taxonomic quagmire. *Molecular Ecology Resources* 15: 967–984.

Nguyen P, Sahara K, Yoshido A, Marec F. 2010. Evolutionary dynamics of rDNA clusters on chromosomes of moths and butterflies (Lepidoptera). *Genetica* 138: 343–354.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32: 268–274.

Novák P, Neumann P, Macas J. 2010. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* 11: 378.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* 29: 792–793.

Ogura Y, Ooka T, Iguchi A, Toh H, Asadulghani M, Oshima K, Kodama T, Abe H, Nakayama K, Kurokawa K. 2009. Comparative genomics reveal the mechanism of the parallel evolution of O157 and non-O157 enterohemorrhagic Escherichia coli. *Proceedings of the National Academy of Sciences* 106: 17939–17944.

Ojanguren-Affilastro AA, Mattoni CI, Ochoa JA, Ramírez MJ, Ceccarelli FS, Prendini L. 2016. Phylogeny, species delimitation and convergence in the South American bothriurid scorpion genus Brachistosternus Pocock 1893: integrating morphology, nuclear and mitochondrial DNA. *Molecular Phylogenetics and Evolution* 94: 159–170.

Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M. 2004. Genetic analyses from ancient DNA. *Annual Reviews in Genetics* 38: 645–679.

Padial JM, Miralles A, De la Riva I, Vences M. 2010. The integrative future of taxonomy. *Frontiers in Zoology* 7: 1.

Palacios-Gimenez OM, Cabral-de-Mello DC. 2015. Repetitive DNA chromosomal organization in the cricket Cycloptiloides americanus: a case of the unusual X1X20 sex chromosome system in Orthoptera. *Molecular Genetics and Genomics* 290: 623–631.

Panzera Y, Pita S, Ferreiro M, Ferrandis I, Lages C, Pérez R, Silva A, Guerra M, Panzera F. 2012. High dynamics of rDNA cluster location in kissing bug holocentric chromosomes (Triatominae, Heteroptera). *Cytogenetic and Genome Research* 138: 56–67.

Papakostas S, Michaloudi E, Proios K, Brehm M, Verhage L, Rota J, Pe√±a C, Stamou G, Pritchard VL, Fontaneto D. 2016. Integrative taxonomy recognizes evolutionary units despite widespread mitonuclear discordance: evidence from a rotifer cryptic species complex. *Systematic Biology* 65: 508–524.

Pérez-García C, Hurtado NS, Morán P, Pasantes JJ. 2014. Evolutionary dynamics of rDNA clusters in chromosomes of five clam species belonging to the family Veneridae (Mollusca, Bivalvia). *BioMed Research International* 2014.

Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, Gray J. 1988. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proceedings of the National Academy of Sciences* 85: 9138–9142.

Qi X, Zhang F, Guan Z, Wang H, Jiang J, Chen S, Chen F. 2015. Localization of 45S and 5S rDNA sites and karyotype of Chrysanthemum and its related genera by fluorescent in situ hybridization. *Biochemical Systematics and Ecology* 62: 164–172.

Prosser SW, deWaard JR, Miller SE, Hebert PD. 2016. DNA barcodes from century-old type specimens using next-generation sequencing. *Molecular Ecology Resources* 16: 487–497.

de Queiroz K. 2005. Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences* 102: 6600–6607.

de Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology* 56: 879–886.

Rannala B, Yang Z. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164: 1645–1656.

Raskina O, Belyayev A, Nevo E. 2004. Quantum speciation in Aegilops: molecular cytogenetic evidence from rDNA cluster variability in natural populations. *Proceedings of the National Academy of Sciences of the United States of America* 101: 14818–14823.

Raskina O, Barber J, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* 120: 351–357.

Regier JC, Shultz JW, Ganley ARD, Hussey A, Shi D, Ball B, Zwick A, Stajich JE, Cummings MP, Martin JW, Cunningham CW. 2008. Resolving Arthropod Phylogeny: Exploring Phylogenetic Signal within 41 kb of Protein-Coding Nuclear Gene Sequence. *Systematic Biology* 57: 920–938.

Richter S, Schwarz F, Hering L, Böggemann M, Bleidorn C. 2015. The Utility of Genome Skimming for Phylogenomic Analyses as Demonstrated for Glycerid Relationships (Annelida, Glyceridae). *Genome Biology and Evolution* 7: 3443–3462.

Schwarzacher HG, Wachtler F. 1993. The nucleolus. *Anatomy and Embryology* 188: 515–536.

Sember A, Bohlen J, Šlechtová V, Altmanová M, Symonová R, Ráb P. 2015. Karyotype differentiation in 19 species of river loach fishes (Nemacheilidae, Teleostei): extensive variability associated with rDNA and heterochromatin distribution and its phylogenetic and ecological interpretation. *BMC Evolutionary Biology* 15.

Sheffield N, Song H, Cameron S, Whiting M. 2008. A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. *Molecular Biology and Evolution* 25: 2499–2509.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. *science* 344: 738–742.

Sotero-Caio CG, Volleth M, Hoffmann FG, Scott L, Wichman HA, Yang F, Baker RJ. 2015. Integration of molecular cytogenetics, dated molecular phylogeny, and model-based predictions to understand the extreme chromosome reorganization in the Neotropical genus Tonatia (Chiroptera: Phyllostomidae). *BMC Evolutionary Biology* 15: 220.

Speicher MR, Carter NP. 2005. The new cytogenetics: blurring the boundaries with molecular biology. *Nature Reviews Genetics* 6: 782.

Sproul JS, Houston D, Davis N, Barrington E, Oh SY, Evans RP, Shiozawa DK. 2014. Comparative phylogeography of codistributed aquatic insects in western North America: insights into dispersal and regional patterns of genetic structure. *Freshwater Biology* 59: 2051–2063.

Sproul JS, Houston DD, Nelson CR, Evans RP, Crandall KA, Shiozawa DK. 2015. Climate oscillations, glacial refugia, and dispersal ability: factors influencing the genetic structure of the least salmonfly, Pteronarcella badia (Plecoptera), in Western North America. *BMC Evolutionary Biology* 15: 279.

Sproul JS, Maddison DR. 2017a. Cryptic species in the mountaintops: species delimitation and taxonomy of the Bembidion breve species group (Coleoptera: Carabidae) aided by genomic architecture of a century-old type specimen. *Zoological Journal of the Linnean Society*: zlx076.

Sproul JS, Maddison DR. 2017b. Sequencing historical specimens: successful preparation of small specimens with low amounts of degraded DNA. *Molecular Ecology Resources* 17: 1183–1201.

Staats M, Erkens RH, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One* 8: e69189.

Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99: 349–364.

Stukenbrock EH. 2013. Evolution, selection and isolation: a genomic view of speciation in fungal plant pathogens. *New Phytologist* 199: 895–907.

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

Symonová R, Majtánová Z, Sember A, Staaks GB, Bohlen J, Freyhof J, Rábová M, Ráb P. 2013. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evolutionary Biology* 13: 1.

Symonová R, Ocalewicz K, Kirtiklis L, Delmastro GB, Pelikánová Š, Garcia S, Kovařík A. 2017. Higher-order organisation of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (Esox sp.). *BMC Genomics* 18: 391.

Symonová R, Sember A, Majtánová Z, Ráb P. 2015. Characterization of fish genomes by GISH and CGH. *Fish Cytogenetic Techniques. Ray-Fin Fishes and Chondrichthyans. CCR Press: Boca Raton*: 118–131.

Tin MMY, Economo EP, Mikheyev AS. 2014. Sequencing degraded DNA from non-destructively sampled museum specimens for RAD-tagging and low-coverage shotgun phylogenetics. *PLoS One* 9: e96793.

Vitte C, Bennetzen JL. 2006. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proceedings of the National Academy of Sciences* 103: 17638–17643.

Wang W, Ma L, Becher H, Garcia S, Kovarikova A, Leitch IJ, Leitch AR, Kovarik A. 2016. Astonishing 35S rDNA diversity in the gymnosperm species Cycas revoluta Thunb. *Chromosoma* 125: 683–699.

West C, James SA, Davey RP, Dicks J, Roberts IN. 2014. Ribosomal DNA Sequence Heterogeneity Reflects Intraspecies Phylogenies and Predicts Genome Structure in Two Contrasting Yeast Species. *Systematic Biology* 63: 543–554.

White MJD. 1977. *Animal cytology and evolution*. Cambridge University Press, London.

Yang Z. 2002. Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics* 162: 1811–1823.

Zwickl, D.J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Unpublished thesis, The University of Texas.